

Educação em Investigação: DA MEDIÇÃO À INTERPRETAÇÃO- BIOESTATÍSTICA CIRÚRGICA

Organizadores

José María Jiménez Avila
Asdrubal Falavigna



Educação em Investigação:
**DA MEDIÇÃO À
INTERPRETAÇÃO–
BIOESTATÍSTICA CIRÚRGICA**

A decorative graphic consisting of multiple thin, light gray lines that flow from the left side of the page, curve downwards and then upwards, and extend towards the right side, creating a sense of movement and depth.

**José María Jiménez Avila
Asdrubal Falavigna**
Organizadores

Educação em Investigação:
**DA MEDIÇÃO À
INTERPRETAÇÃO-
BIOESTATÍSTICA CIRÚRGICA**



EDUCS

COMUNIDADE

© dos organizadores

Capa: André Secundino

Revisão: Carla Vinize Casagrande Gianini

Editoração: Traço Diferencial | (54) 3229 7740 | (54) 99901 3978

Dados Internacionais de Catalogação na Publicação (CIP)
Universidade de Caxias do Sul
UCS – BICE – Processamento Técnico

E24 Educação em investigação : da medição à interpretação – bioestatística cirúrgica / org. José María Jiménez Avila, Asdrubal Falavigna. – Caxias do Sul, RS : Educus, 2018.
210 p.: il.; 23 cm.
ISBN 978-85-7061-910-5
Apresenta bibliografia
Vários colaboradores

1. Bioestatística. 2. Coluna vertebral – Cirurgia. I. Falavigna, Asdrubal. II. Avila, José María Jiménez.

CDU 2. ed.: 57.087.1

Índice para o catálogo sistemático:

| | |
|---------------------|-------------|
| 1. Bioestatística | 57.087.1 |
| 2. Coluna vertebral | 616.711-089 |

Catalogação na fonte elaborada pela bibliotecária
Paula Fernanda Fedatto Leal – CRB 10/2291

Direitos reservados à:



ASSOCIAÇÃO AOSPINE LATIN AMERICA

Av. Silva Jardim 2042 Cj. 1505 | CEP 80250-200 | Curitiba – PR | Brasil

Telefone: 55 41 3016 4491 | E-mail: aosla@aospine.org

SUMÁRIO

Nota dos organizadores / 11

Colaboradores / 13

Prefácio / 17

UNIDADE 1 / 19

Capítulo 1

INTRODUÇÃO / 21

Asdrubal Falavigna e José María Jiménez Avila

Capítulo 2

MATEMÁTICAS: LINGUAGEM E INTERPRETAÇÃO / 23

Maurício Scussel

Capítulo 3

O ATOR PRINCIPAL: A VARIÁVEL / 29

José María Jiménez Avila

Capítulo 4

COMO PLANEJAR A ELABORAÇÃO DE UM BANCO DE DADOS? / 33

Pedro Luis Bazán

Capítulo 5

DETERMINANDO O TAMANHO DA AMOSTRA / 45

Emiliano Valle

Capítulo 6

COMO ESCOLHER OS TESTES ESTATÍSTICOS? / 53

Pedro Luis Bazán

Capítulo 7

BIOESTATÍSTICA DESCRITIVA / 63

Alberto Aceves Pérez

Capítulo 8

BIOESTATÍSTICA ANALÍTICA OU INFERENCIAL / 69

José María Jiménez Avila

Capítulo 9

MEDIDAS DE TENDÊNCIA CENTRAL / 75

Alberto Aceves Pérez

Capítulo 10**MEDIDAS DE DISPERSÃO (Amplitude, Variância, Desvio Padrão, Coeficiente de Variância) / 81***Delio Martins***Capítulo 11****APRESENTAÇÃO DOS DADOS / 87***José María Jiménez Avila***Capítulo 12****TESTE DE HIPÓTESE (IC) E VALOR DE “p” / 93***José María Jiménez Avila***Capítulo 13****TESTE DE KOLMOGOROV-SMIRNOV / 99***Nelson Astur Neto**Maximiliano Gimenez Gigón**Rodrigo Góes Medéa de Mendonça***Capítulo 14****TESTE DE SHAPIRO-WILK / 105***Nelson Astur Neto**Rodrigo Góes Medéa de Mendonça***UNIDADE 2 / 109****Capítulo 15****TESTE t DE STUDENT / 111***Barón Zarate**Francisco López Meléndez***Capítulo 16****TESTE DE ANOVA / 121***Juan Pablo Guyot***Capítulo 17****CORRELAÇÃO DE PEARSON / 125***Alisson R. Teles**Asdrubal Falavigna***UNIDADE 3 / 133****Capítulo 18****Chi² / 135***Alberto Aceves Pérez***Capítulo 19****TESTE DE POSTOS DE WILCOXON / 139***Baron Zarate*

UNIDADE 3 / 147**Capítulo 20****TESTE U DE MANN-WHITNEY / 149**

*José Manuel Pérez Atanasio
Victor Fernando Luján Celis
Maritza Belén Sandoval Rincón*

Capítulo 21**TESTE DE KRUSKAL-WALLIS / 153**

*Esteban Araya Ramírez
Fabián Víquez Monge*

Capítulo 22**COEFICIENTE DE CORRELAÇÃO DE SPEARMAN / 157**

*Fernando Alvarado
Miguel Farfán*

Capítulo 23**TESTE DE MCNEMAR / 163**

*Esteban Araya Ramírez
Fabián Víquez Monge*

Capítulo 24**REGRESSÃO LINEAR SIMPLES / 169**

*Alisson R. Teles
Asdrubal Falavigna*

Capítulo 25**DISTRIBUIÇÃO DE POISSON / 175**

*José Acuña Pinilla
Margarita Gómez Chantraine
Elizabeth Heins*

Capítulo 26**ODDS RATIO / 187**

*José Manuel Pérez Atanasio
Victor Fernando Luján Cellis
Maritza Belén Sandoval Rincón*

UNIDADE 4 / 193**Capítulo 27****O QUE DÁ VALOR E FORÇA ÀS EVIDÊNCIAS / 195**

*Mohsen Sadeghi Naini
Vafa Rahimi Movaghar*

AGRADECIMENTOS

Agradecemos aos autores por sua valiosa contribuição na elaboração dos capítulos, e os parabenizamos pela qualidade didática de simplificar o complexo mundo da bioestatística.

Nosso agradecimento à *AOSpine International Research Commision Members* pelo apoio incondicional na elaboração do livro.

Agradecemos também o suporte dos membros do *Board* da *AOSpine Latinoamérica* e do escritório pelo apoio logístico para a finalização deste grande projeto.

Publicar este 2º livro é motivo de grande satisfação, já que temos a certeza de que contribuirá para o fortalecimento diário do conhecimento da pesquisa clínica e experimental.

José María Jiménez Avila

Novamente, meu reconhecimento à Georgina, Zaira e Daniela, por terem dividido suas horas e, às vezes, seus dias de convívio a este livro. Sempre foram minha motivação e minha razão para continuar com este tipo de projetos.

Reconhecimento e agradecimento muito especial ao Dr. Asdrubal Falavigna, atual *Chairperson* da *AOSpine Latinoamérica*, que sempre confiou no projeto e o torna possível graças à sua visão e liderança.

Asdrubal Falavigna

Dedico este livro à minha esposa, Ana Maria, e a meus filhos Vincenzo e Rebecca, pelo apoio diário. Desejo que este livro potencialize o espírito científico dentro de cada estudante e profissional, a fim de capacitá-los a inovar e gerar conhecimento, visando buscar respostas aos nossos questionamentos diários.

NOTA DOS ORGANIZADORES

Existe uma demanda reprimida para qualificação em pesquisa, seja para o desenvolvimento de um pensamento reflexivo, seja para uma análise crítica dos fatos. É fundamental a geração de novos conhecimentos e produtos de inovação, que potencializem os avanços da ciência e permitam a aplicabilidade no desenvolvimento da sociedade, sob suas diversas dimensões.

A aprendizagem do conhecimento de bioestatístico normalmente é longa e difícil. Este livro aborda os princípios básicos da estatística aplicada ao especialista da área da saúde, sob a visão de especialistas e pesquisadores oriundos de diferentes áreas do conhecimento. Nessa jornada, o pesquisador encontrará dicas e exemplos que facilitarão o entendimento desta complexa área, dentre outros, dos problemas no entendimento do significado de $p < 0.05$, quando escolher um método estatístico sobre o outro, etc. Estes são os obstáculos a serem ultrapassados, a fim de se tornar um profissional crítico e reflexivo, capaz de intervir na realidade e produzir conhecimento específico da área, de modo a atender as demandas da população.

A leitura dos capítulos permite ao leitor adquirir conhecimento das dificuldades comumente encontradas, na área de bioestatística. No final, nada é mais gratificante do que ver o trabalho publicado, após vários esforços, disciplina e dedicação.

José María Jiménez Avila
Asdrubal Falavigna

COLABORADORES

Alberto Aceves Pérez

Ortopedista. Centro Médico Nacional de Occidente. Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México. Professor da disciplina na Universidad de Guadalajara. AOSpine Research Officer México.

Alisson R. Teles

Neurocirurgião da coluna. McGill Scoliosis & Spine Group. Universidade McGill, Montreal, Canadá.

Asdrubal Falavigna

AOSpine Latin America Chairperson. Diretor do Centro de Ciências da Vida da UCS. Coordenador do Programa de Pós-Graduação em Ciências da Saúde da UCS. Professor de Neurocirurgia da Universidade de Caxias do Sul. Coordenador do Grupo de Pesquisa Estudos Clínicos e Modelos Básicos de Patologias Raquimedulares e do Laboratório de Patologias Raquimedulares da UCS. Caxias do Sul, RS – Brasil

Barón Zárate Kalfópulos

Cirurgião de Columna. Instituto Nacional de Rehabilitación / Medica Sur Ciudad de México. Professor Alta Especialidade Cirurgia de Coluna Universidad Nacional Autónoma de México. Diretor do Centro AOSpine Cidade do México.

Delio Martins

Ortopedista. Hospital Israelita Albert Einstein. Doutor em Ciências pela Universidade Federal de Sao Paulo (UNIFESP); Coordenador do Curso de Aperfeiçoamento em Cirurgia de Coluna do Hospital Israelita Albert Einstein; Docente de Ortopedia da Universidade Anhembí-Morumbi.

Elizabeth Heins

Médica Cirurgiã Pontifícia Universidade Javeriana. Epidemiologista - Johannes Gutenberg Universität Mainz. Senior Manager do Medical Affairs Johnson & Johnson Medical Devices Northern Cluster.

Emiliano Vialle

Chefe do grupo de cirurgia de coluna, Hospital Universitário Cajuru – Banco de Tecidos Humanos, Pontifícia Universidade Católica do Paraná AOSpine Latin America Fellowship Officer. Brasil.

Esteban Araya

Ortopedista, Universidade de San Carlos de Guatemala. Cirurgião de Coluna, Universidade Militar Nueva Granada, Colombia. Professor no Departamento de cirurgia da Universidade de Ciências Médicas, Costa Rica. Coordenador na Clínica de Coluna Hospital do Trauma e Coordenador Médico Banco de Tecidos Hospital do Trauma, Ex Chefe de Serviços Cirúrgicos Hospital do Trauma, Costa Rica.

Fabian Vilquez Monge

Mestrado em Seguros e Gestão de Riscos pela Pontifícia Universidade de Salamanca, Espanha, Graduação em Risco Associado à Operação, Universidade da Costa Rica. Bacharel em Estatística pela Universidade da Costa Rica.

Fernando Alvarado Gomez

Cirurgião de Coluna - Hospital Universitário Fundación Santa Fe de Bogotá. Cirurgião de Coluna - Instituto Ortopédico Infantil de Roosevelt. Coluna de Cirurgia, Pelve e Acetábulo – Universidade Militar Nueva Granada. Ortopedia e Traumatologia – Pontifícia Universidade Javeriana. Epidemiologista Clínico – Universidade Nossa Senhora do Rosário. Professor clínico de Ortopedia e Traumatologia da Universidad de los Andes. Professor Associado de Ortopedia e Traumatologia da Universidade de Nuestra Señora del Rosario. Professor Ad Honorem do programa de cirurgia de coluna da Universidade Militar de Nova Granada. Pesquisa AOSpine Colombia. Secretário da coluna do capítulo SCCOT.

Francisco Manuel López Melendez

Cirurgião de Coluna. ISSSTE Hospital General Santiago Ramón e Cajal. Durango. México.

José Acuña Pinilla

Instrumentador Cirúrgico Profissional, Universidade de Santander UDES. Coordenador de Medicina Baseada em Evidências e Assuntos Médicos para a Região Norte da América Latina Divisão One Medical Devices da empresa Johnson & Johnson. Professor da Universidade de Santander - UDES, Bogotá.

José Manuel Pérez Atanasio

Médico ortopedista. Mestre em Ciências. Chefe da Divisão de Pesquisa em Saúde. UMAE “Dr. Víctorio de la Fuente Narváez” IMSS. Ciudad de México.

José María Jiménez Avila

Ortopedista, Instituto Tecnológico y de Estudios Superiores de Monterrey (TEC), campus Guadalajara, Professor no Departamento de Ciências Básicas e Coordenador de Clínica de Ortopedia e Traumatologia, Instituto Mexicano del Seguro Social, Clínica de Coluna, AOSpine Latinoamérica Research Officer. México..

Juan Pablo Guyot

Cirurgia de Coluna. Ortopedia e Traumatologia. Hospital Universitario Fundación Favaloro. Argentina.

Maritza Belén Sandoval Rincón

Médica psiquiatra. Especialista em transtornos do sono. Clínica de Trastornos do Movimento e Sono. Hospital General Ajusco Medio. Ciudad de México.

Maurício Scussel

Engenheiro de Controle e Automação pela Pontifícia Universidade Católica do Paraná (PUCPR). Doutorado em Neurologia / Neurociências pela Universidade Federal de São Paulo e Centro Nacional de Microeletrônica de Barcelona (UNIFESP/IMB-CNM). Pós-Doutorado no Instituto de Fisiologia e Fisiopatologia na Universidade de Heidelberg – Alemanha. Diretor da Simplifica –Pesquisa & Desenvolvimento para Life Sciences. Brasil.

Margarita Gómez Chantraine

Médica cirurgiã - Universidad Nacional de Colombia. Especialista em Obstetricia e Ginecologia - Universidad Nacional de Colombia. Magister (C) em Epidemiologia Clínica e Pesquisa - Pontificia Universidad Javeriana. Grouper Manager de Medical Affairs Johnson & Johnson Medical Devices Latin America. Colombia.

Maximiliano Gimenez Gigon

Ortopedia. Cirurgião de Coluna no Centro de Patologia Raquidea - Tucumán, Argentina. AOSpine Fellow na Santa Casa de la Misericórdia - São Paulo, Brasil.

Miguel Farfán

Médico cirurgião graduado pela Universidade Nacional da Colômbia, Especialista em Epidemiologia Clínica pela Universidade El Bosque, Ortopedista e traumatologista formado pela Fundação Santa Fé de Bogotá, Universidade del Rosario. Gerente Geral e Fundador Media For Health SAS, Gerente Geral e Fundador da Unidade Ortopédica da Colômbia UNOC. Experiência em inovação em saúde, otimização de processos através do uso de TIC (Apps) e melhoria da experiência do usuário e satisfação de pacientes e profissionais de saúde. Experiência em pesquisa

clínica, criação de registros clínicos e revisão sistemática da literatura. Desenvolvimento de projetos de inovação educacional em saúde para médicos e pacientes.

Mohsen Sadegi Naini

Neurocirurgião. Spine Trauma and Surgery Research Center, Tehran University of Medical sciences.

Nelson Astur Neto

Pesquisador de AOSpine Brasil. Cirurgião ortopédico de coluna no Hospital Albert Einstein - São Paulo, Brasil. Membro ativo da Sociedade Brasileira de Coluna (SBC), Sociedade Brasileira de Ortopedia e Traumatologia (SBOT), North American Spine Society (NASS), American Academy of Orthopedic Surgeons (AAOS), AOSpine. Brasil

Pedro Luis Bazán

Especialista em Ortopedia e Traumatologia, certificado em Cirurgia de Coluna Vertebral (SAPCV). Chefe da Unidades de Coluna HIGA San Martín e Hospital Italiano La Plata. Diretor do Curso de Capacitação Pós Básica em Patologia da Coluna Vertebral do Adulto (Departamento de Post Grado Facultad de Ciencias Médicas de la Universidad Nacional de La Plata). Diplomado por contrato em Ortopedia e Traumatología (FCM-UNLP). Argentina.

Rodrigo Góes Medéa de Mendonça

Ortopedia. Cirurgião de Coluna. Médico associado ao Grupo de Coluna do Hospital Santa Casa de São Paulo. Médico no Hospital Albert Einstein – São Paulo. Membro ativo da Sociedade Brasileira de Ortopedia e Trauma (SBOT), Sociedade Brasileira de Coluna (SBC), Sociedade Brasileira de Ortopedia Pediátrica (SBOP), North American Spine Society (NASS), AOSpine. Brasil.

Vafa Rhimi Movarghar

Professor de Neurocirurgia. Spine Trauma and Surgery Research Center, Tehran University of Medical Sciences

Víctor Fernando Luján Celis

Médico residente de pesquisa em serviço social. División de Investigación en Salud. UMAE “Dr. Víctorio de la Fuente Narváez” IMSS. Cidade do México.

PREFÁCIO

A *AOSpine* está se voltando para a ampliação de conhecimentos, além de ser uma transmissora de conhecimentos, nesse aspecto, no mundo. O presente livro encontra boa ressonância com o atual objetivo da Comissão Internacional de Pesquisas da *AOSpine*, de tornarem-se, cada um dos seus membros, Cirurgião Cientista. A cirurgia da coluna vertebral está, atualmente, em rápida evolução e, diariamente, somos confrontados por técnicas mais novas, implantes modernos e tecnologia de alto custo. É imperativo que cada cirurgião tenha um conhecimento completo da pesquisa clínica, o que não apenas lhe permitirá realizar pesquisas individualmente, mas também avaliar a qualidade do trabalho de pesquisa no material publicado.

A transformação de um cirurgião talentoso em um cirurgião cientista, contudo, requer um claro conhecimento de protocolos de pesquisa. E o presente livro é ideal para este fim. Os capítulos são delineados para guiar o leitor passo a passo no planejamento do estudo, na coleta de amostras, seleção de testes estatísticos apropriados, análise de dados e apresentação de resultados. Todos os autores são peritos nesse campo, e cada capítulo é escrito de maneira clara, de modo que o cirurgião não teria dificuldade em escolher o melhor protocolo de pesquisa. Eu fico especialmente feliz pelos cirurgiões de coluna vertebral, que estão atualmente em treinamento, *fellows* e residentes que estão ativamente envolvidos em pesquisa, pois este livro proporciona um entendimento claro da metodologia de pesquisa.

Congratulo-me com os autores que podem ter grande orgulho deste trabalho, que é uma compilação impressionante de informações valiosas para qualquer um que realize pesquisas clínicas.

Prof. S. Rajasekaran, PhD.
Chairman, Dept of Orthopaedics,
Ganga Hospital, Coimbatore, India
Chair, AOSpine International Research Commission

Unidade 1

CAPÍTULO 1

Introdução

José María Jiménez Ávila e Asdrubal Falavigna

CAPÍTULO 2

Matemáticas: linguagem e interpretação

Maurício Scussel

CAPÍTULO 3

O ator principal: A variável

José María Jiménez Avila

CAPÍTULO 4

Como planejar a elaboração de um banco de dados?

Pedro Luis Bazán

CAPÍTULO 5

Determinando o tamanho da amostra

Emiliano Vialle

CAPÍTULO 6

Como escolher os testes estatísticos?

Pedro Luis Bazán

CAPÍTULO 7

Bioestatística descritiva

Alberto Aceves Perez

CAPÍTULO 8

Bioestatística analítica ou inferencial

José María Jiménez Avila

CAPÍTULO 9

Medidas de tendência central

Alberto Aceves Pérez

CAPÍTULO 10

Medidas de dispersão (Amplitude, Variância, Desvio Padrão, Coeficiente de Variância)

Delio Martins

CAPÍTULO 11

Apresentação dos dados

José María Jiménez Avila

CAPÍTULO 12

Teste de Hipótese (ic) e valor de “p”

José María Jiménez Avila

CAPÍTULO 13

Teste de Kolmogorov-Smirnov

Nelson Astur Neto

Maximiliano Gimenez Gigón

Rodrigo Góes Medéa de Mendonça

CAPÍTULO 14

Teste de Shapiro-Wilk

Nelson Astur Neto,

Rodrigo Góes Medéa de Mendonça

CAPÍTULO 1

INTRODUÇÃO

*Asdrubal Falavigna
José María Jiménez Avila*

O processo de pesquisa é chave para a geração, amenização e renovação de ferramentas diagnósticas, preventivas, e de tratamento oportuno para os pacientes.

Um conflito paradoxal é que a pesquisa nesta área é gerada em países desenvolvidos que não são atingidos pelos problemas de saúde do terceiro mundo, além de vivenciarem uma realidade econômica e de desigualdades em saúde muito diferente das enfrentadas diariamente em nossos países.

Por esta razão, faz-se necessário gerar pesquisa em nível prático-secundário, menos direcionado às ciências básicas e mais às áreas de epidemiologia clínica, ciências sociais, economia em saúde e satisfação do usuário.

Como parte do processo de pesquisa, temos a análise de dados e, ainda mais importante, a interpretação da informação gerada pela totalidade dos detalhes revelados na coleta de informações em pesquisas de opinião.

Para isso, devemos conhecer fundamentos de estatística, área da disciplina de matemática aplicada, que visa a coleta, elaboração, análise e interpretação dos dados numéricos.

A estatística não é considerada uma ciência, mas um método que ensina os procedimentos lógicos de observação e análise a seguir, para obter o máximo proveito das experiências de outras áreas do conhecimento.

Este livro visa colaborar com todo cirurgião de coluna, estudante de medicina ou médico em geral que, embora contando com ampla experiência clínico-cirúrgica e vasta informação sobre seus casos, ainda não deu o grande passo de interpretar os dados. A interpretação não

apenas descritiva, mas do ponto de vista da análise, ou “inferência”, será o que lhe permitirá obter informações de grande utilidade para a comunidade científica.

O livro possui 27 capítulos, divididos em quatro unidades que abordam a estrutura dos bancos de dados, a interpretação dos testes paramétricos e não paramétricos na cirurgia de coluna e algo muito importante: a utilização dos dados obtidos, não apenas para apresentá-los num artigo original, mas para que essa informação seja útil na gestão e produção de uma mudança no atendimento médico-cirúrgico.

O livro pensa também naquelas pessoas que desejam continuar realizando sua atividade assistencial e cirúrgica: que gostam de pesquisa e que sentem a necessidade de saber interpretar os testes estatísticos que constam no parágrafo de material, métodos e resultados dos artigos originais. Muitas vezes, estes dados são considerados uma verdade absoluta, devido à falta de análise. E, em muitos casos, deixamos de perceber que suas conclusões carecem de congruência metodológica.

Este livro fortalece a análise crítica da literatura médica, para integrar o conhecimento ligado às diretrizes científicas.

Lembre-se: *“O que não é medido não é avaliado e, portanto, não é melhorado.”*

CAPÍTULO 2

MATEMÁTICAS: LINGUAGEM E INTERPRETAÇÃO*Maurício Scussel*

Matemática é a ciência do raciocínio lógico e abstrato, que investiga geometrias, proporções, variações, dentre outros temas. É uma ciência abrangente e inerente às outras ciências, que por diversas vezes impressiona por aparentes níveis elevados de complexidade.

Muitos veem nessa suposta complexidade um empecilho que dificulta a compreensão dos fenômenos, cuja representação é dada em linguagem matemática. Porém, questionamos: Estaria a dificuldade de interpretação de um dado problema numérico na linguagem simbólica da matemática aplicada, ou seria ela decorrente do não entendimento da razão básica, que está detrás dos símbolos e das notações matemáticas?

Muitas vezes, a avaliação estatística de um estudo fica restrita aos estatísticos responsáveis pela análise dos resultados. Contudo, a reflexão sobre a linguagem e a interpretação da matemática pode ser estimulada no pesquisador que pretende descrever a metodologia ou o resultado de um estudo. Desta forma, o intuito deste livro não é transformar o leitor num estatístico, mas, sim, despertar o interesse pela razão que constitui as bases deste conhecimento.

Desde a Antiguidade, a matemática sempre esteve presente no pensamento de filósofos, como Platão, Pitágoras, Aristóteles, entre outros. Isso é compreensível, uma vez que a matemática é uma ferramenta que possibilita demonstrar, através da lógica, a razão da mecânica de algum conhecimento, e ao mesmo tempo, ligar o real ao abstrato. É paradoxal uma ciência ser tão exata, lógica, definindo que dois mais dois são quatro, e também conectar o real ao imaginário onde $i^2 = -1$.

Com toda essa aparente complexidade, como é possível compreender e dominar esta linguagem?

A matemática é composta por um simbolismo extenso, para descrever diferentes funções, equações, variações... Porém, podemos dizer de forma resumida que esses símbolos, em essência, servem para transmitir algum conhecimento, que pode estar ao mesmo tempo em um nível objetivo ou subjetivo. E, para ilustrar esses aspectos objetivo e subjetivo dos símbolos, refletimos sobre os símbolos que se seguem:



Em uma análise simples e objetiva, percebe-se que cada elemento está relacionado a algo concreto. A *cruz*, o cristianismo. O *certo*, a uma marca de material esportivo. Os *anéis*, às Olimpíadas. Porém, em cada elemento apresentado, existe um conhecimento de ordem subjetiva, mascarado, que é associado à figura. Por exemplo, para a maioria dos leitores, provavelmente não seja necessário citar o nome da marca esportiva que a figura representa, nem mencionar o tamanho da influência que esta marca exerce no mundo esportivo, pois este é um conhecimento que está armazenado inconscientemente.

Para criar uma imagem sobre o objetivo e o subjetivo, poderíamos dizer que um símbolo é uma espécie de cartão de memória. O seu desenho seria como a unidade física, objetiva (objeto), e o conteúdo é o subjetivo (colocar sob), o que está em outra camada.

Mas, por que o subjetivo auxilia no entendimento da matemática?

Conseguir se aproximar mais do aspecto subjetivo, encontrado nos símbolos utilizados na matemática, facilita a compreensão da razão pela qual um elemento é aplicado em uma simples conta de soma ou subtração. Assim, percorreremos algumas operações aritméticas, com a mente aberta, para tentar observar esta propriedade dos símbolos de outro ângulo.

A aritmética é o ramo da matemática que trata dos números e das operações possíveis entre eles. Segmento mais antigo e mais elementar que compreende, dentre vários elementos, as quatro operações: adição, subtração, multiplicação e divisão. Para buscar uma aproximação da relação do objetivo *versus* subjetivo, analisaremos cada símbolo das quatro operações, buscando extrair uma *razão* pela qual o símbolo apresenta determinada forma:

- **Subtração:** operação representada por uma linha contínua. Segundo historiadores, este símbolo surge na literatura ocidental, a partir do século XIV e XV, com a metamorfose da palavra latina *minus* para \overline{mus} até chegar no $-$.

Analisando a função de uma subtração pela matemática ou pela linguística, chegamos à ideia de que subtrair representa o fato de tirar algo de. Assim, tendo em mente a necessidade de representar uma separação, uma ruptura, qual seria um elemento simples que representasse esta ação?

Imagine as seguintes situações: Dois vizinhos que necessitam delimitar suas propriedades: Qual elemento geométrico serviria pra representar essa separação de áreas? Quando é necessário separar a pele para uma incisão cirúrgica, qual é o desenho que se forma na pele? Uma linha, não?



É claro que esta aproximação abstrata da razão deve ser encarada de forma lúdica, não confrontando a explicação lógica da razão que deu origem ao símbolo. Porém, uma vez que o conhecimento é incorporado, tanto faz se a origem dele foi estritamente lógica ou lúdica; o importante é entender o conceito do elemento para poder melhor aplicá-lo.

- **Divisão:** é a operação atualmente representada por uma linha contínua separando dois pontos (\div). Segundo historiadores, este símbolo é mais moderno, pois apareceu pela primeira vez no século XVII.

De forma análoga à subtração, a divisão também se refere a uma ruptura entre duas partes. Porém, na divisão existe um elemento divisor que delimita em quantas porções **iguais** o dividendo será fracionado.



Dessa forma, para representar simbolicamente uma ruptura de algo em partes iguais, não seria necessário que o símbolo que representa esta operação contivesse a fragmentação com pesos iguais entre os dois lados?

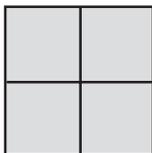
• **Adição:** operação representada por uma linha contínua sobreposta por outra linha contínua. O surgimento deste símbolo é concomitante ao da subtração, aparecendo nas mesmas publicações. A explicação histórica é que este elemento surge da palavra *et* (que corresponde à conjunção e no português: João e Maria; dois e dois são quatro, etc.), que, com o passar do tempo, começou a ser abreviada por *t*, até chegar ao símbolo atualmente utilizado (+).



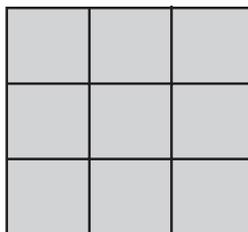
Para alcançar uma imagem lúdica deste elemento, podemos imaginar o resultado de alguma atividade que necessite ligar dois pontos separados por uma linha ou espaço. Como exemplo, podemos imaginar o cadarço de um tênis que junta duas partes, ou o ponto cirúrgico, que costura duas porções de pele separadas por um corte.

• **Multiplicação:** para demonstrar uma razão do porquê do símbolo da multiplicação apresentar o formato (x), é necessário o auxílio da geometria. Este recurso ajuda a entender que quando multiplicamos um número por ele mesmo, estamos levando este elemento ao quadrado (potenciação).

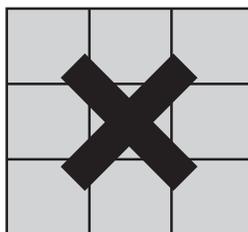
Vamos imaginar a multiplicação do número 2×2 . Também podemos dizer que estamos elevando o número 2 ao quadrado, pois $2 \times 2 = 4$, assim como $2^2 = 4$, portanto $2 \times 2 = 2^2$. Mas por que o uso da Geometria? Observe a imagem abaixo:



Podemos dizer que o quadrado desenhado acima está representando a equação 2^2 , pois temos quatro elementos dentro do quadrado. Ou, podemos dizer também que representa a multiplicação dos elementos 2 e 2, pois o seu resultado será 4. No caso do 3×3 , ou, 3^2 , temos uma situação análoga:



Representamos o 3^2 através da multiplicação do 3 por ele mesmo, resultando no número 9, que corresponde à quantidade de elementos dentro do quadrado. Observando pela geometria, percebe-se que, na multiplicação, os elementos não apenas crescem no comprimento, mas também na largura para atender o requisito da multiplicação, conseqüentemente da potenciação. Assim, um símbolo que possivelmente represente a expansão nas duas dimensões (largura e comprimento) poderia ser o X.



É claro que, para a apresentação de resultados científicos, a subjetividade do simbolismo dos elementos matemáticos não é evidenciada. Pois é necessária a fixação de conceitos e bases para garantir o avanço do conhecimento, e pelo objetivismo é mais sólido o caminhar. Assim, a proposição é apenas de despertar a curiosidade dos porquês de cada elemento matemático. Essa forma lúdica de encarar a matemática talvez não elucidie, em primeira instância, um problema, mas qualquer conhecimento, visto de outro ângulo, amplia a compreensão.

No decorrer da leitura deste livro, se tiver dificuldades para analisar ou entender alguma fórmula matemática, lembre-se que entender uma equação é análogo a entender o caso clínico de um paciente. Cada sinal, cada símbolo é equivalente a um sintoma, a um exame; e cada sintoma possui um histórico, uma razão. Quando você se deparar com algo complexo, lembre-se dos estudos dos casos clínicos. Comece a analisar ponto a ponto, e, através da razão de cada elemento, a equação começará a ficar mais compreensível.

LEITURA COMPLEMENTAR

1. Logic, Wesley C. Salmon. Prentice-Hall, 1963.
2. SUPPES, Patric. Introduction to Logic, Princeton, N.J., D Van Nostrand Company, Inc., 1957.

CAPÍTULO 3

O ATOR PRINCIPAL: A VARIÁVEL

José María Jiménez Avila

Na pesquisa e desde seu início, tudo gira em torno daquilo que se denomina *variável*. A variável pode ser definida como a qualidade, a propriedade ou característica das pessoas ou das coisas em estudo, que pode ser enumerada ou medida quantitativa ou qualitativamente e que varia de um sujeito a outro. Ela deve ser sempre levada em consideração, não apenas no momento de elaboração da pergunta de pesquisa (PICOT), mas durante todo o processo metodológico.

A função da variável é decompor a *hipótese* proposta em seus elementos mais simples, e é toda característica ou acontecimento de uma pessoa, pessoas ou grupo, que muda de uma situação para outra, ou de um momento para outro podendo, portanto, adquirir valores diversos.

As variáveis classificam-se em *independentes*, ou explicativas, exposição (causa), e dependentes, ou resposta (efeito ou doença). Além destas, existem aquelas que modificam algum acontecimento e que são denominadas *intervenientes*.

No momento em que se identifica um problema e procura-se uma explicação hipotética para a variável, é preciso testá-la, o que pode ser feito decompondo-a em pequenas variáveis que devem ser suscetíveis à avaliação empírica.

Se os procedimentos empíricos não refutarem a hipótese formulada, esta será aceita como provavelmente verdadeira.

Na maioria dos casos, é preciso medir as variáveis durante a comparação empírica da hipótese, e a mensuração torna-se um procedimento indispensável na prática científica.

Uma variável deve ter duas características indispensáveis: *mensurável e observável*.

Mensuração

A mensuração consiste em atribuir um número ou qualificação a alguma propriedade específica de um indivíduo, uma população ou um evento, usando certas regras. O que se mede é uma característica do indivíduo, e não o indivíduo em si. A mensuração é um processo de abstração.

Para realizar a mensuração, é preciso seguir um processo que consiste em passar de uma entidade teórica para uma escala conceitual e, posteriormente, para uma escala operacional.

Os passos seguintes são:

- a) delimitar a parte do evento que será medida;
- b) selecionar a escala que será usada para medir;
- c) comparar o atributo medido com a escala e;
- d) finalmente, emitir um juízo de valor sobre os resultados da comparação.

Principais escalas de mensuração (tipos de variáveis)

As escalas classificam-se em qualitativas (nominais e ordinais) e quantitativas (discretas e contínuas). Um requisito indispensável de todas as escalas é que as categorias sejam exaustivas e mutuamente excludentes.

• Escala nominal

É a característica que se define por um nome, e o fato de ser definida por um nome NÃO implica ser mais ou menos que a característica definida por outro nome diferente, por exemplo: Sexo: Masculino, Feminino; Ocupação; Tipo de parto; tabagismo, etc.

Na prática, algumas vezes opta-se por dar um número como codificação, visando facilitar o registro. Nestes casos, o número não possui significado nem valor algum, sua finalidade é apenas diferenciar um grupo de outro.

• Escala ordinal

É aquela cujas características podem receber alguma ordem subjetiva, pode acontecer que uma característica seja considerada com maior ou menor valor em relação a outras, mas desconhece-se a intensidade dessa diferença. Por exemplo, o bem e o mal.

A ordem é subjetiva, ou é uma escolha do pesquisador. Escolaridade (Ensino Fundamental, Ensino Médio, etc.)

- **Escala discreta**

Aquela cujos valores estão separados entre si por uma quantidade e é impossível fracionar a unidade.

No momento de dividir a unidade, esta desaparece. Exemplo: Idade (anos completos), número de filhos.

- **Escala contínua**

Aquela em que a escala de mensuração pode ser dividida numa quantidade infinita de valores entre dois pontos quaisquer.

O metro pode ser dividido em um número infinito de partes.

Todos os derivados da longitude ou do tempo, por exemplo; volume, área, peso, tamanho, temperatura.

Processo de Operacionalização das Variáveis

As variáveis pesquisadas num estudo são identificadas a partir do momento em que é definido o problema. No arcabouço teórico, são identificadas e conceituadas as variáveis secundárias. Este nível de definição é abstrato e complexo, e não permite a observação ou mensuração.

O processo de levar uma variável de um nível abstrato para um plano operacional é denominado *Operacionalização* e a função básica desse processo é precisar ou concretizar ao máximo o significado ou alcance, que se confere a uma variável em estudo.

Os conceitos representam o sistema teórico de qualquer ciência e são símbolos dos fenômenos estudados.

Um conceito é geral e abstrato, não sendo possível a observação dos fenômenos envolvidos; por isso, é necessário chegar à operacionalização das variáveis.

A operacionalização é obtida através de um processo que transforma uma variável em outras que tenham o mesmo significado e que sejam suscetíveis de mensuração empírica. Para isso, as variáveis principais são decompostas em outras, mais específicas, denominadas *dimensões*, que por sua vez precisam ser traduzidas para indicadores que permitam a observação direta.

Exemplo:

Variável: Evolução da fratura

Indicador: Estado de consolidação

Dimensão: Fases da consolidação

- Escalas:
1. Calo mole.
 2. Calo duro.
 3. Corticalização.
 4. Remodelação.

Após o final deste processo, podemos considerar que as variáveis, obtidas de maneira conceitualizada, podem ser medidas e observadas.

Ao mesmo tempo, isto permitirá estabelecer alguma associação que possibilite identificar fatores causais, que possam ser previsores da identificação de fatores de risco, que talvez também desencadeie algum tipo de enfermidade ou de resultado.

As variáveis devem ser definidas de maneira clara, com o objetivo de evitar confusões. Isto facilitará a busca e a análise posterior dos dados, e garantirá a comparação dos resultados obtidos com aqueles achados em estudos prévios.

LEITURA COMPLEMENTAR

1. Celis de la Rosa A. Bioestadística, en: Características de los datos y de su recolección, revisión y computo. 1. ed en español. Manual Moderno. 2004.
2. MacMahon B, Pugh TF. Epidemiology. Principles and methods. Boston: Little Brown & Co., 1970.
3. Moreno-Altamirano A. Salud Pública. Vol. 42. No 4 Julio/Agosto 2000, p. 337-348.

CAPÍTULO 4

COMO PLANEJAR A ELABORAÇÃO DE UM BANCO DE DADOS?

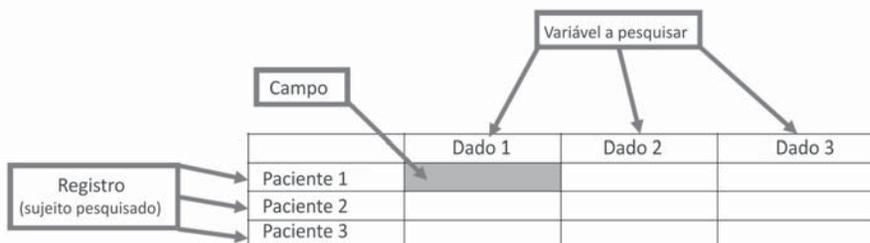
Pedro Luis Bazán

Introdução

O correto gerenciamento das informações é fundamental para qualquer tipo de atividade de trabalho do ser humano. A atividade médico-científica não foge a esta regra e, por isso, devemos conhecer e planejar, da melhor maneira possível, a coleta, o armazenamento e a análise das informações, para responder à pergunta que deu origem a nossa pesquisa.

Uma maneira simples de fazê-lo é criar arquivos em planilhas de cálculo (tabelas) que contendam um tópico por arquivo. A unidade funcional destas tabelas é a célula da folha, também chamada de campo. Na primeira célula de cada coluna, colocaremos o título da variável a avaliar, que será o mesmo para cada paciente registrado. Cada fila representa um registro, que contém a informação inserida de cada paciente (Fig. 1). É possível criar outra folha que reúna outras informações, ou outro documento.

Figura 1 – Exemplo de componentes de una planilha de cálculo



As planilhas de cálculo atuais permitem realizar diferentes funções para a análise estatística dos dados. Esta forma de gerenciar as informações pode apresentar algumas dificuldades, principalmente a complexidade do trabalho científico planejado, e a necessidade de contar com várias folhas. Assim, é possível que encontremos informações redundantes, inconsistentes e inúteis, somadas à dificuldade de manejo do erro de inserção da informação (um dos erros mais frequentes na pesquisa científica), e à perda de dados.

Outras dificuldades, ao usar arquivos para gerenciamento dos dados, são: modificação dinâmica da estrutura dos arquivos, consulta, atualização e proteção dos dados.

Outra forma de gerenciar as informações é criando bancos de dados que facilitem sua gestão. A Real Academia Espanhola define banco de dados como “um conjunto de dados organizado de tal modo que permita obter com rapidez diversos tipos de informação”.

Para sua criação e gestão, podemos utilizar diversos gerenciadores de bancos de dados (*software*) que, mesmo variando em nível de complexidade, tentam guardar os dados de forma estruturada, visando diminuir os erros de inserção e facilitar a análise das informações. Exemplos destes gerenciadores são Access, da Microsoft, EpiInfo, IBM DB2, etc.

Entre as características destes gerenciadores de bancos de dados, temos: a sua independência, sem nenhuma relação com os aplicativos utilizados ou onde estiverem armazenados; o acesso dos seus usuários mediante consultas; o armazenamento dos dados, de forma centralizada e independente. Isto torna os gerenciadores consistentes, completos, confiáveis e seguros.

Os bancos de dados são armazenados em arquivos que contêm dados (números e caracteres), que podem ser abertos, armazenados, combinados, editados, acrescentados, ou excluídos. A estrutura geral dos arquivos que contêm bancos de dados inclui tabelas que podem ser relacionadas entre si.

Podemos classificar os bancos de dados em:

Primários – projetados de acordo com as necessidades específicas de informações (estudo, registro, gerador de indicadores, etc.). Se projetados corretamente, ajustam-se à resposta feita inicialmente;

Secundários – bancos de dados usados para uma finalidade diferente daquela que motivou a coleta dos dados que contêm. Suas principais vantagens são a rapidez e a economia, mas os dados podem

ter baixa qualidade, o que, muitas vezes, poderá exigir uma adaptação dos objetivos aos dados;

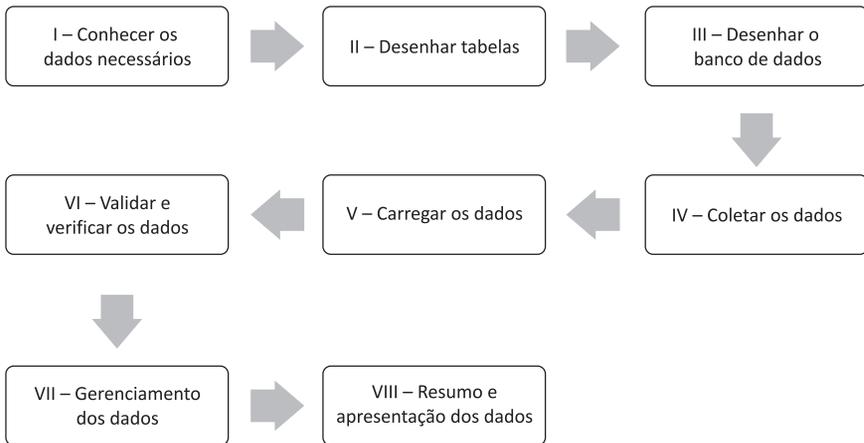
Terciários – são criados com dados provenientes de diferentes bancos de dados, combinações de dados relatados previamente.

Planejamento de um banco de dados

A origem de toda pesquisa científica é conhecer o motivo (pergunta) que a motivou. Após ter cumprido este requisito, o pesquisador deve planejar tudo o que se refere ao trabalho científico: bibliografia, tipo de estudo, critérios de inclusão e exclusão, dados necessários para resolver a questão a pesquisar, e definição de prazos para todas estas etapas.

No que se refere especificamente ao banco de dados, também é preciso seguir uma série de passos para a correta gestão da informação, além de conhecer a pessoa responsável por eles. (Figura 2).

Figura 2 – Diagrama de fluxo no planejamento e utilização de um banco de dados



Antes de começar a coleta de dados, é imprescindível determinar quais os dados que são importantes para a pesquisa. Isto permitirá coletar a totalidade das informações, e somente o necessário.

O não cumprimento desta premissa poderá causar insuficiência de informações para satisfazer os objetivos traçados. Poderemos perceber isso somente no momento de análise dos resultados, o que nos obrigaria a modificar o material avaliado ou o seu método de análise, evidenciando que houve perda de tempo ou a determinação do cancelamento da pesquisa.

Tampouco é sensato coletar dados desnecessários, que não servem para os objetivos da pesquisa e somente consomem tempo e recursos excessivos.

Por estes e outros motivos, podemos concluir que, como em toda atividade, um dos passos mais importantes é planejar a pesquisa, saber o que é necessário para realizá-la, como o faremos, com que ou com quem o faremos, etc. Em poucas palavras, o sucesso da pesquisa depende, em parte, do seu correto planejamento.

Projetar as tabelas

Como escrevemos anteriormente, um banco de dados é um conjunto de tabelas inter-relacionadas, que tem por finalidade facilitar o gerenciamento dos dados. Após termos conhecimento dos dados que precisaremos para nossa pesquisa, passaremos a decidir a estrutura de nosso banco de dados. Dependendo da complexidade dos dados a carregar, determina-se a complexidade do banco de dados a criar.

Os diferentes programas de gerenciamento de bancos de dados oferecem-nos a possibilidade de editar a inserção das informações e a maneira como se inter-relacionam as diferentes tabelas, de acordo com nossa necessidade.

Em cada campo selecionado (Figura 3), a variável receberá um nome, e os dados a carregar receberão características diferentes. Isto permitirá reduzir o erro de carga, além de garantir as mesmas características entre todos os elementos e, com isso, fazer uma análise correta.

Figura 3 – Projeto da tabela



Observe, no quadro superior, a seleção do tipo de dado para cada variável e, no quadro inferior, as validações realizadas em cada campo.

Existem diferentes tipos de **variáveis**:

- **Textos.** Contêm dados de texto e números, mas com estes últimos não é possível fazer operações matemáticas. São úteis para salvar informações como nomes e endereços.
- **Número.** São usados para armazenar informações numéricas. Podem conter dados contínuos ou categóricos, números inteiros ou decimais. Isto deve ser especificado ao montar a estrutura. Com eles, podem ser realizadas operações matemáticas.
- **Lógicas; Sim/Não; Verdadeiro/Falso.** São armazenados dados que só tem duas possibilidades de resposta, também denominados binários.
- **Data/Hora.** São armazenadas datas. Estas podem ser curtas (ano com as últimas duas cifras), longas (ano de quatro cifras), associadas ou não à hora. Com eles podem ser realizados cálculos entre datas e a maioria dos programas permite carregar somente datas válidas.

Cada um destes campos corresponderá à variável a estudar, e é representado por colunas na folha de dados. As filas correspondem a cada sujeito, e cada célula em particular representa o valor específico da variável a estudar em cada sujeito (Figura 4).

Figura 4 – Folha de dados da tabela projetada na figura anterior

| HISTORIA CL. | APELLIDO Y. | SEXO | EDAD | OCUPACIÓN | MAREOS | VÉRTICO | DERIVACIÓN | PTE | CM | ORL |
|--------------|-------------|------|------|----------------|--------|---------|------------|-----|----|-----|
| 539379 | | F | 44 | EMPLEADA DOM | | | | | | |
| 13027871 | | F | 56 | AUXILIAR EN IA | | | | | | |
| 24151992 | | F | 38 | | | | | | | |
| 27512676 | | M | 33 | Administrativo | | | | | | |
| 27742034 | | F | 34 | ADC | | | | | | |
| 92707115 | | F | 25 | ESTUDIANTE | | | | | | |

Tendo conhecimento das diferentes tabelas, das quais podemos precisar, podemos determinar como será nosso banco de dados.

Este passo pode ser subdividido em três etapas:

1. Projeto conceitual: nesta etapa planeja-se a estrutura do banco de dados, sem base na tecnologia nem no *software* que será utilizado. Faz-se um esboço do que será necessário para trabalhar, com base nos dados que precisamos avaliar. Neste ponto, decidiremos qual o tipo de banco de dados que será usado (planilhas, relacional, centrado em objetos, hierárquico, redes, relacionais, operacionais, etc.);

2. Projeto lógico: é o momento de realização da ideia com base na tecnologia que será usada, ou seja, com que banco de dados irá trabalhar e quais as chaves e variáveis hierarquizadas;

3. Projeto físico: consiste em relacionar a estrutura do banco de dados com o *hardware* necessário para obter maior eficácia na carga, armazenamento e consulta dos dados. Por exemplo, considerar a capacidade de memória interna ou externa, para o armazenamento dos dados a carregar, cópias de segurança, etc., principalmente se além dos dados houver armazenamento de imagens.

Ao finalizar esta etapa, é útil realizar vários testes de inserção, consulta e modificação de dados virtuais, para avaliar o rendimento do banco de dados e reconhecer falhas de programação, antes de começar a carregar os dados.

Coleta de dados

Antes de começar a inserção de pacientes, devemos realizar uma planilha que servirá para a coleta das informações necessárias. Estas informações devem ser coletadas na mesma ordem em que serão inseridas no banco de dados.

Cada planilha deve ser rotulada com a mesma chave do registro do banco de dados, o que possibilitará acesso rápido, no caso de haver algum erro de inserção ou falta de algum dado.

Carga de dados

Há duas maneiras de carregar as informações no banco de dados, tomando como base a tabela já carregada. Uma delas é vendo a tabela como a folha de dados que criamos no início (Figura 4). A dificuldade é que, ao carregar os dados, devemos manter o cursor na mesma fila (paciente) ao movimentá-lo na coluna (variável a avaliar), para não cometer o erro de carregar um dado em outro registro.

Outra maneira de registrar as informações num banco de dados é criando um formulário baseado no projeto da tabela (Figura 5).

Figura 5 – Formulário de entrada de dados com os campos a preencher

The image shows a screenshot of a data entry form titled "COLUMNA CERVICAL Y MAREOS". The form is displayed in a web browser window with the address bar showing "W MAREOS: Base de datos: C:\Users\Pedro Luis Bazán\Dropbox\Trabajo\PLB\En curso\Mareos\W MAREOS.acctb...". The browser's menu bar includes "Archivo", "Inicio", "Crear", "Datos externos", "Herramientas de base de datos", and "¿Qué desea hacer?". The form itself has a "Panel de navegación" on the left side. The main content area contains the following fields and options:

- HISTORIA CLINICA:
- APELLLIDO Y NOMBRE:
- SEXO: EDAD: OCUPACIÓN:
- CONSULTA: MAREOS: VÉRTIGO:
- DERIVACIÓN: PTE: CM: ORL: NEURO: OTROS:
- TEC: FECHA: TRAUMA CERVICA: FECHA:
- CERVICALGIA: BRAQUIALGIA: CEFALEA: ACUFENC: NISTAGMUS:
- ROOMBERG: ROMBERG SENSIBILIZAI: GIRO TOTAL: GIRO CABEZA FI:
- Rx:
- RMN:

At the bottom of the form, there is a status bar that reads "Registro: 4 de 7" and "Sin filtro - Buscar". The browser's status bar at the very bottom shows "Vista Formulario".

Desta maneira, vemos na tela somente os campos a preencher com dados de um único registro (paciente). Após aceitação do dado carregado, passa-se ao campo seguinte do mesmo paciente e, ao finalizar a carga, passa-se para outro registro.

Como existem opções de validação automática do formato de carga ou da obrigatoriedade da mesma, o programa alerta para a dificuldade de realizar a inserção, o que reduz erros de inserção.

Validar e verificar os dados

Este passo é de fundamental importância para a análise correta das informações e para a obtenção dos resultados do trabalho.

A validação dos dados deve ser realizada metodologicamente, a cada transferência de informações (coleta, carga e armazenamento). Se isso não for feito, na etapa final da pesquisa perceberemos que houve perda de dados ou inconsistência dos mesmos, o que fará a pesquisa fracassar ou exigirá investimento em tempo extra para controle do erro.

Nas pesquisas em grupo, recomenda-se que o carregamento dos dados tenha um único responsável, e que a mesma pessoa faça a validação e verificação do processo de carregamento.

Gerenciamento dos dados

Um ponto crucial no gerenciamento das informações é a análise dos dados, e será tema de outro capítulo.

Após carregamento dos dados, pode-se entrar no banco de dados para fazer consultas e relatórios, além de decidir qual a operação estatística que será utilizada.

Resumo e apresentação dos dados

Os sistemas gestores de banco de dados podem, após análise dos mesmos, entregar um relatório resumido, tabelas de dupla entrada e gráficos que ilustram os resultados.

Responsáveis

Para evitar erros na gestão dos dados (coleta, carga, armazenamento e análise) cada passo deve ter um responsável conhecido.

Todos os autores da pesquisa são responsáveis pela tomada de decisão, tanto no que se refere ao planejamento da pesquisa quanto ao reconhecimento dos dados necessários para a mesma.

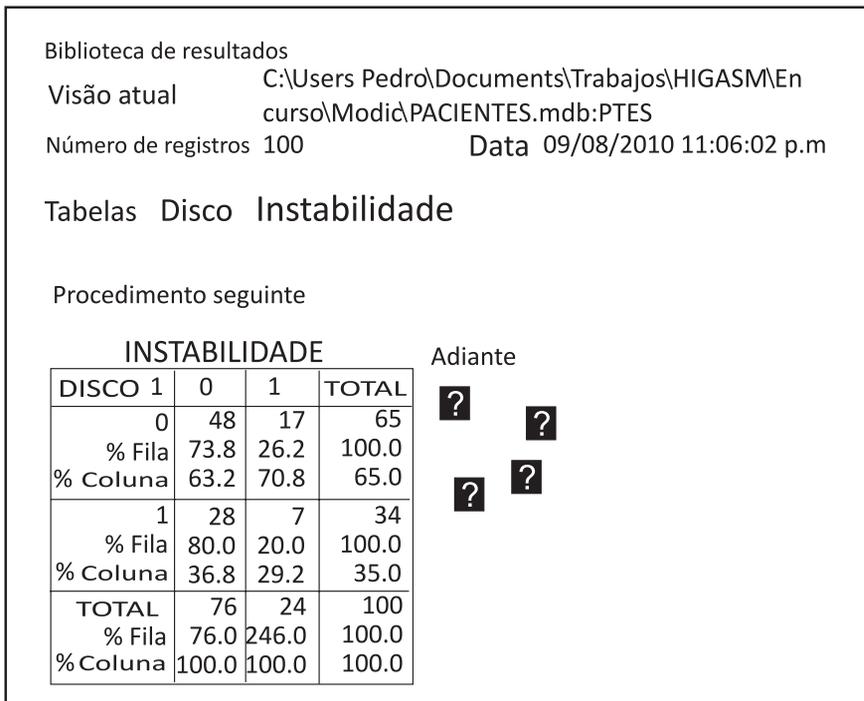
O planejamento do banco de dados e da planilha de coleta de informações deve contar com um ou dois responsáveis; e é aconselhável que um deles seja um técnico em estatística.

A coleta dos dados e a análise dos resultados requerem o envolvimento de todos os participantes.

O carregamento dos dados deve ser acompanhado por um único responsável que, ao mesmo tempo, garanta que todos os dados necessários tenham sido coletados.

O responsável pela seleção e análise dos testes estatísticos deve ser um técnico em estatística, ou um dos autores que esteja familiarizado com a análise dos dados. (Figura 6)

Figura 6 – Relatório de análise de dados realizado por EpiInfo, com tabela de dupla entrada e resultados



| Análise da Tabela Simples | | | |
|---|------------------|----------------------------|---------------------|
| | Ponto Estimativa | 95% Intervalo de confiança | |
| | | L. Inferior | L. Superior |
| PARÂMETROS: baseados em OR | | | |
| Odds Ratio (Produto Cruzado) | 0.7059 | 0.2607 | 1.9111 (T) |
| Odds Ratio (EMV-MLE) | 0.7083 | 0.2465 | 1.9004 (M) |
| | | 0.2201 | 2.0773 (F) |
| PARÂMETROS:Baseados no risco | | | |
| Razão de riscos (RR) | 0.9231 | 0.7408 | 1.1502 (T) |
| Diferença de riscos (DR) | -6.1538 | -23.1763 | 10.8686 (T) |
| (T-series Taylos; C-Cornfield, M-P-Media; F-Fisher) | | | |
| TESTES ESTATÍSTICOS | | | |
| | Chi ² | p. de 1 cola | p de 2 colas |
| Chi ² , sem corrigir | 0.4723 | | 0.4919156556 |
| Chi ² , Mantel-Haenszel | 0.4676 | | 0.4940887145 |
| Chi ² : Correção de Yates | 0.1952 | | 0.6586242660 |
| P-Media exata | | 0.2548047339 | |
| Teste exato de Fisher | | 0.3335295358 | |

Se o planejamento realizado foi bem-sucedido, o banco de dados criado não apenas permitirá inserir dados em um formulário gerenciado na tela, como permitirá também a realização de consultas sobre diversos parâmetros, que possibilitarão avaliações parciais a cada carga de dados, relacionando mais de uma tabela do total.

É possível obter relatórios individuais de cada registro e das consultas. A diferença é que os formulários só podem ser visualizados na tela, e os relatórios podem ser impressos.

Conclusões

A gestão dos dados inclui carregamento, armazenamento, consulta e análise dos mesmos, e é muito importante para a pesquisa médica.

Os bancos de dados são ferramentas imprescindíveis para o gerenciamento das informações da pesquisa; seu planejamento requer a análise e o conhecimento dos dados necessários para levar a cabo nosso trabalho, e devemos basear-nos neles.

As perguntas anteriores à criação de um banco de dados seriam as seguintes:

- Que dados desejo armazenar e qual é a melhor forma de fazê-lo? (Tabelas e suas relações)
- Como insiro os dados relativos à pesquisa? (Formulário)
- Como faço as avaliações parciais dos dados? (Consultas)
- Como avalio os dados? (Operações estatísticas)

LEITURA COMPLEMENTAR

1. Bleisch, S: Geographic Information Technology Training Alliance (GITTA) presents: Databases.

2. Deserno, TM; Haak, D; Brandenburg, V; Deserno, V; Classen, C; Specht, P: Integrated Image Data and Medical Record Management for Rare Disease Registries. A General Framework and its Instantiation to the German Calciphylaxis Registry. *J Digit Imaging* (2014) 27:702–713 DOI 10.1007/s10278-014-9698-8.

3. Gómez Ballester, E; Martínez Barco, P; Moreda Pozo, P; Suárez Cueto, A; Montoyo Guijarro, A; Saquete Boro, E: Bancos de dados I. Introducción a las bancos de dados. Fundamentos de diseño de bases de datos. Departamento de Lenguajes y Sistemas Informáticos Escuela Politécnica Superior. Universidad de Alicante.

4. Haak, D; Samsl, C; Gehlen, J; Jonas, S; Deserno, TM: simplifying electronic data capture in clinical trials: workflow embedded image and biosignal file integration and analysis via web services. *J Digit Imaging* (2014) 27:571-580. DOI 10.1007/s10278-014-9694-z.

5. López Martínez A, C; Sánchez Brea, LM; Gómez, O; González Del Campo Rodríguez Barbero, MAV. Editorial McGraw-Hill/Interamericana de España. 2002.

6. Marcus, DS; Harms, MP; Snyder, AZ; Jenkinson, M; Wilson, JA; Glasser, MF; et al. Human Connectome Project Informatics: quality control, database services, and data visualization *Neuroimage*. 2013; October 15; 80. .doi:10.1016/j. neuroimage. 2013.05.077.

7. Musick, BS; Robb, SL; Burns, DS; Stegenga, K; McCorkle, KJ: Development and Use of a Web-based Data Management System for a Randomized Clinical Trial of Adolescents and Young Adults. *Comput Inform Nurs*. 2011 June; 29(6): 337-343. doi:10.1097/NCN.0b013e3181fcbc95.

8. Quintana Rondón, Y; Dominguez, LC; Diaz Berenguer, A: Diseño de la Base de datos para Sistemas de Digitalización y Gestión de Medias. Revista de Informática Educativa y Medios Audiovisuales. Vol 8 (15). 17-25. 2011.
9. Robbins, RJ: Database Fundamentals. Johns Hopkins University.
10. Sanchez, J: Principios sobre Bancos de datos Relacionales.
11. Silberschatz, A; Korth, HF; Sudarshan, S; Fundamentos de bases de datos. Cuarta edición. Traducido de la cuarta edición en inglés por Sáenz Pérez, F; García Cordero.

CAPÍTULO 5

DETERMINANDO O TAMANHO DA AMOSTRA

Emiliano Vialle

Introdução

Ao ser convidado para escrever sobre este tema, tive a impressão de que finalmente iria desvendar uma pergunta crucial na elaboração de um estudo, que assola médicos e pesquisadores em todos os níveis de pesquisa e formação. A realidade que encontrei foi de uma infinidade de fórmulas com grande subjetividade na sua seleção. Longe de transmitir uma mensagem de desânimo ao leitor, sugiro uma leitura mais informativa sobre o tema, sem a preocupação com os cálculos em si, mas com as informações necessárias para o desenvolvimento destes, em conjunto com um estatístico.

Definir o número de participantes de uma pesquisa sempre foi um tema crucial e de difícil solução, e com certa frequência descobrimos que o número de participantes foi insuficiente, após o término do projeto. Parece óbvio que um maior número de sujeitos, incluídos numa pesquisa, produzirá resultados mais confiáveis e próximos da realidade. Tempo, recursos e frequência da questão a ser respondida fazem com que não se possa incluir pacientes num estudo indefinidamente, e a pergunta crucial passa a ser: Qual o número mínimo de pacientes em cada grupo de pesquisa? O cálculo amostral também tem sido solicitado por comitês de ética para aprovação de projetos de pesquisa.

Neste capítulo tentarei definir, através de exemplos práticos, as maneiras mais comuns e práticas de se chegar ao tal “número mágico”. É importante ressaltar que as fórmulas para cálculo da amostra são sensíveis a erros, e variações nos parâmetros podem levar a grandes diferenças no tamanho da amostra. Por isto, antes de começarmos a buscar o tamanho de nossa amostra, tendo o projeto de pesquisa elaborado, alguns itens devem ser definidos:

1) Erro do tipo I (alfa)

Em estudos clínicos avalia-se apenas uma parte da população, testando a hipótese de que há diferença entre amostras (taxa de infecção em diabéticos, por exemplo), ou no resultado de tratamento entre grupos (artrodese de coluna em diabéticos *versus* não diabéticos), ou no resultado de tratamentos em um grupo (artrodese *versus* tratamento conservador em pacientes diabéticos). A hipótese nula (H_0) parte do princípio de que os pacientes (amostras), que estão sendo comparados não são diferentes. A hipótese alternativa (H_1) assume que os grupos são diferentes. Nesse momento, o cálculo amostral define em que momento a inclusão de pacientes num estudo deixará de afetar o resultado.

Neste processo, dois tipos de erro podem ocorrer (tipo I e tipo II, Quadro 1)

| | | Population | |
|---------------------------|-----------------------|---------------------------|----------------------|
| | | Difference does not exist | |
| Difference does not exist | | False negative result | Type II error (beta) |
| Difference exists | False positive result | Power (1 beta) | Type I error (alpha) |

O erro do tipo 1 (alfa) mede a probabilidade de que, a partir da hipótese nula (H_0), as amostras vêm da mesma população, as diferenças encontradas no estudo são verdadeiras. Ele representa a chance de erroneamente rejeitarmos a H_0 e escolhermos um resultado falso/positivo. Alfa geralmente é fixado em 0.05, significando que o pesquisador deseja um risco inferior a 5%, ao apresentar resultados falsamente positivos.

2) Erro tipo II (beta)

Ao invés de chegar a uma conclusão falsamente positiva, o pesquisador pode chegar a um falso-negativo, concluindo que não há diferença num tratamento entre grupos, quando na verdade há diferença. Em outras palavras, aceitam a H_0 como verdadeira, e isto se denomina erro do tipo II (beta). Tradicionalmente, beta é definido como 0.2, significando que o pesquisador deseja uma chance inferior a 20%

de chegar a uma conclusão falsamente negativa. Para calcular o tamanho da amostra, deve-se determinar o poder do estudo. Este poder estatístico reflete a capacidade de o estudo encontrar um efeito na população, baseando-se numa amostra dela (verdadeiro-positivo). O poder estatístico é o complemento do erro tipo II (1-beta). Portanto, se o erro tipo II for 0.2, o poder do estudo será 0.8 ou 80%. Esta é a probabilidade de rejeitar a hipótese nula corretamente, ou evitar uma conclusão falsamente negativa. Sendo mais objetivo, seria a probabilidade de que, ao se apresentar uma diferença num estudo, esta seja verdadeira.

3) Diferença mínima relevante

É a diferença mínima que o pesquisador deseja identificar entre os grupos estudados. Deve ser uma diferença clinicamente relevante e biologicamente plausível.

4) Variabilidade

O cálculo da amostra se baseia na variação populacional para a variável estudada, através do desvio padrão. Este valor geralmente é obtido através de um estudo piloto, ou a partir de estudos prévios. A diferença mínima e a variabilidade podem, algumas vezes, ser representadas como um múltiplo do desvio padrão das observações. Fórmula:

$$\text{Standardized difference} = \frac{\text{difference between the means in the two treatment groups}}{\text{population standard deviation}}$$

5) Qual o tamanho da população?

Muitas vezes este dado é de difícil obtenção, principalmente se seu projeto envolve uma pesquisa muito específica. Imagine que seu estudo envolve uma comparação entre tratamento cirúrgico e conservador de fraturas da coluna toracolombar. Sua população seria obtida através de algum estudo epidemiológico, que nem sempre se aplica à sua realidade, ou através de uma revisão de prontuários de atendimento de seu hospital.

6) Qual a margem de erro (intervalo de confiança) desejada?

Partindo do princípio de que sua amostra não será perfeita, deve-se definir quanto erro será aceito, e geralmente o valor de 5% é utilizado.

7) Qual o nível de confiança desejado?

Os intervalos de confiança mais utilizados são os valores de 90%, 95%, e 99%. Este valor vai determinar a possibilidade de que seu resultado esteja dentro do intervalo de confiança.

8) Qual o desvio padrão?

Quanta variação você espera para seus resultados? Como esta é uma resposta com difícil solução, a maioria das pesquisas utiliza o valor de 0.5.

Tendo os dados acima em mãos, mas sem saber o tamanho da população em questão, passamos ao cálculo da amostra. O seu nível de confiança corresponde a um Z-score, valor constante utilizado no cálculo da mostra:

$$90\% - Z\text{-score} = 1.645$$

$$95\% - Z\text{-score} = 1.96$$

$$99\% - Z\text{-score} = 2.576$$

Existem tabelas para os valores do Z-score disponíveis na internet, mas os valores acima são os mais comuns para aplicação da fórmula:

$$\text{Tamanho da amostra} = (Z\text{-score})^2 \times DP \times (1-DP) / (\text{margem de erro})^2$$

Digamos que você escolheu um nível de confiança de 95%, um desvio padrão de .5, e um intervalo de confiança de +/- 5%. Vamos deixar o tamanho da amostra de lado, ou considerar que o estudo envolve uma parcela significativa da população.

$$\text{Tamanho da amostra} = ((1.96)^2 \times .5(.5)) / (.05)^2$$

$$(3.8416 \times .25) / .0025$$

$$.9604 / .0025$$

$$384.16$$

385 participantes seriam necessários.

Se este número não for animador, é possível reduzir o intervalo de confiança ou aumentar a margem de erro. A fórmula acima representa uma das maneiras mais simples de determinar amostras de pesquisa, e

é utilizada com frequência para pesquisas de opinião, por exemplo, nas quais a comparação entre grupos não é necessária. A Tabela abaixo ilustra como a definição dos valores de alfa e beta influenciam no tamanho da amostra.

Tabela 1 – Tamanho aproximado da amostra para diferentes níveis de alfa e potência

| | Alpha (type I error) | | |
|----------------|----------------------|------|-------|
| | 0.05 | 0.01 | 0.001 |
| Power (1-beta) | | | |
| 0.80 | 100 | 149 | 218 |
| 0.90 | 134 | 190 | 266 |
| 0.99 | 234 | 306 | 402 |

Quando tratamos de estudos clínicos, nos quais diferenças entre grupos de pacientes são pesquisadas, algumas outras informações são necessárias.

Vamos considerar um estudo comparando tratamento cirúrgico e conservador para fratura de coluna toracolombar, no qual os participantes serão distribuídos aleatoriamente em dois grupos homogêneos, e o resultado final seria o nível de dor seis meses após o início do tratamento. A partir de estudos prévios, sabe-se que fraturas tratadas conservadoramente (grupo A) apresentam melhora da incapacidade de 15 pontos, e que as cirúrgicas (grupo B) melhoram 20 pontos.

Estudos prévios mostram que a variação da incapacidade pode ser 5 pontos (desvio padrão da amostra).

Para esclarecer se o tratamento cirúrgico leva a uma melhora significativa em relação ao conservador, o teste-*t* de Student para mostras não pareadas pode ser utilizado, e para garantir que o estudo não terá um número insuficiente ou excessivo de participantes, é necessário um planejamento do tamanho da amostra. Isto requer uma especificação do força estatística (probabilidade de identificar uma diferença real com o teste, geralmente 80 ou 90%) e do nível de significância do teste estatístico (probabilidade de obter um resultado significativo, quando não há diferença real, geralmente de 2.5% para testes unicaudais). A partir destes dados, começam a surgir fórmulas estatísticas que não são familiares à maioria dos médicos, no caso do teste-*t* acima:

$$n \approx \left[\frac{2(z_{\text{Power}} + z_{1-\alpha})}{2(\mu_1 - \mu_2)/\delta} \right]^2$$

Na fórmula acima, os grupos ($n_1=n_2$) e os desvios padrão são iguais. A diferença média entre as duas populações é $\mu_1-\mu_2$. O poder (Z-power) é dado como 0.8 ou 80%, e o nível de significância é α . N é a soma dos dois grupos (n_1+n_2). Os valores a serem inseridos nas fórmulas são obtidos a partir de tabelas disponíveis na internet. O trabalho de Noordzij et al. (2010) fornece os valores mais frequentemente utilizados.

É evidente que esta fórmula é desanimadora, mas claramente não pode ser interpretada unicamente por um estatístico. O motivo de sua descrição é justamente informar o cirurgião sobre os dados necessários para a realização do cálculo de tamanho da amostra, e permitir um trabalho em conjunto para a escolha do melhor teste estatístico e da amostra mais viável à realidade clínica em que o estudo será realizado.

Para o teste- t demonstrado acima, é necessária uma estimativa de valores médios nos dois grupos comparados, assim como os desvios padrão.

Para o teste de Fisher, deve-se fornecer uma estimativa de proporção de eventos (por exemplo uma reação adversa) nos dois grupos.

O teste de Mann-Whitney requer uma estimativa da probabilidade de que a variável estudada será menor em um grupo do que no outro.

Estudos mais simples ou menos pretenciosos podem utilizar gráficos para estimar o tamanho da amostra, partindo de presunções sobre o tipo de dado e teste estatístico a ser usado, como o publicado por Altman (1980).

Exemplo de estudo experimental

Para determinar o tamanho da amostra de um estudo sobre o efeito do transplante celular da degeneração discal em coelhos, utilizou-se o valor obtido em estudo prévio comparando discos normais e degenerados.

O cálculo do tamanho da amostra foi realizado no programa WinPEPI (*Programs for Epidemiologists for Windows*) versão 11.43 e baseado no estudo de Vialle et al. (2012). Para um nível de significância

de 5%, poder de 80%, e um tamanho de efeito mínimo de 1,2 desvios padrão entre os grupos, obteve-se um total mínimo de 12 animais em cada grupo.

Infelizmente, muitas etapas da determinação são subjetivas, como a determinação do poder do teste estatístico, e outras são externas, como a frequência da doença estudada, tempo de recrutamento, recursos disponíveis. Adaptações podem ser feitas em decorrência destas limitações, reduzindo-se a força do teste, ou o intervalo de confiança.

Apesar das dificuldades impostas pelas etapas da determinação do tamanho da amostra, e da possibilidade de erro, ela é sempre melhor em relação a uma escolha arbitrária da amostra.

LEITURA COMPLEMENTAR

1. Florey CD. Sample size for beginners. *BMJ*. 1993;306(6886):1181-1184. doi:10.1136/bmj.306.6886.1181.
2. Gupta KK, Attri JP, Singh A, Kaur H, Kaur G. Basic concepts for sample size calculation: Critical step for any clinical trials! *Saudi J Anaesth*. 2016;10(3):328-331. doi:10.4103/1658-354X.174918.
3. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: Basic principles and common pitfalls. *Nephrol Dial Transplant*. 2010;25(5):1388-1393. doi:10.1093/ndt/gfp732.
4. Röhrig, B. et al., 2010. Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt international*, 107(31-32), p. 552-6. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2933537&tool=pmcentrez&rendertype=abstract>.
5. Tobergte DR, Curtis S. No Title No Title. *J Chem Inf Model*. 2013;53(9):1689-1699. doi:10.1017/CBO9781107415324.004.

CAPÍTULO 6

COMO ESCOLHER OS TESTES ESTATÍSTICOS?

Pedro Luis Bazán

Introdução

A estatística é a ciência que utiliza conjuntos de dados numéricos para obter, a partir deles, inferências baseadas no cálculo de probabilidades. Há dois tipos de inferências, as conhecidas como testes de hipóteses e as que se dedicam à estimativa de intervalos.

A importância da análise estatística fundamenta-se na decisão de aceitar ou rejeitar a hipótese do estudo que será realizado.

É possível dividi-la em duas grandes áreas: estatística descritiva, dedicada à descrição, visualização e ao resumo de dados originados, a partir dos fenômenos de estudo; e estatística inferencial, dedicada à geração de modelos, inferências e previsões associados aos fenômenos em questão, levando em consideração a aleatoriedade das observações.

Desta forma, podemos analisar diferentes variáveis, que se definem como: qualidades, traços, atributos, características ou propriedades de diferentes valores, magnitudes ou intensidades avaliadas em um grupo de elementos. Temos, então, as variáveis qualitativas, que podem ser: binárias (por exemplo, sim ou não; feminino ou masculino; verdadeiro ou falso); as variáveis ordinais (por exemplo, estadiamento de tumores), as variáveis nominais (por exemplo, tipo de fratura torácico-lombar), e as variáveis quantitativas, que conferem um valor, podendo ser contínuas (por exemplo, peso, idade, tamanho) ou descontínuas (por exemplo, número de infecção de local cirúrgico).

Outra forma de classificar as variáveis estatísticas é: variáveis independentes (antecedentes, causa, etc.) e variáveis dependentes (efeito, resultado, etc.).

Um erro frequente é identificar, no final da pesquisa, uma grande quantidade de informações e, nesse momento, ter que decidir como será feita a análise de todo esse material. É nessa hora que percebemos

que não sabemos o que fazer com esses dados e que, muitas vezes, eles estão errados ou incompletos, o que impede encontrar uma resposta para a pergunta que originou a pesquisa.

O planejamento conjunto do processo de pesquisa é fundamental para o sucesso de nosso trabalho. Deve ser realizado de forma completa, antes de começar a carga dos dados, e precisa considerar os passos seguintes:

- propor a pergunta que dá origem à pesquisa;
- decidir sobre os objetivos do trabalho que responderão à pergunta. Neste ponto são decididas as variáveis a pesquisar;
- definir a hipótese nula, com um índice de confiança entre 0.05 e 0.01. Este passo permite saber como serão interpretados os resultados;
- decidir o tipo de estudo e a amostra necessária. Neste ponto, inferem-se os testes estatísticos a usar;
- coletar os dados em planilhas que descrevam cada variável;
- introduzir os dados nos formulários eletrônicos;
- fazer uma lista codificada das variáveis a analisar;
- corrigir os bancos de dados usando gestores de dados e planilhas de cálculos;
- descrever os dados de forma gráfica e numérica (estatística descritiva). Esta é a última possibilidade para confirmar a qualidade dos dados;
- inferir as características da população baseando-se na amostra (estatístico-inferencial);
- interpretar os resultados de acordo com os objetivos e a hipótese nula.

Todos os passos são importantes, mas os primeiros sete são a base para o sucesso da análise estatística dos dados.

Formas de selecionar os diferentes testes estatísticos

Não existe uma receita infalível para selecionar o teste estatístico mais adequado à nossa pesquisa, mas há diferentes formas de escolher o que está mais de acordo. É o que tentaremos analisar.

A seleção do teste deve ser realizada antes de começar a carga dos dados.

O primeiro passo é o teste de hipótese, para sua confirmação ou rejeição e, por outro lado, o intervalo de confiança.

Tipos de testes estatísticos

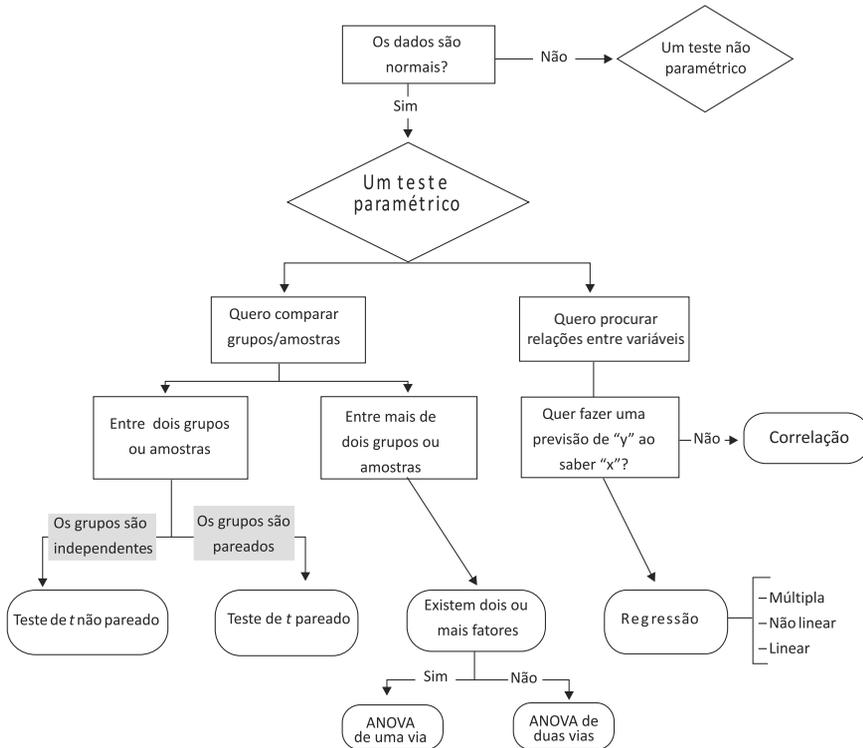
Os testes estatísticos podem ser classificados em dois grandes grupos: paramétricos e não paramétricos. A utilização destes testes depende de vários fatores que podem ser analisados, a partir de diferentes pontos de vista: distribuição dos dados, observações a estudar, suas variáveis, centralização, dispersão e contagem dos dados, tamanho da amostra (Quadro 1).

Quadro 1 – Diferentes características entre os testes paramétricos e não paramétricos

| Características | Testes paramétricos | Testes não paramétricos |
|--------------------|--------------------------------|---|
| Distribuição | Normal e variâncias homogêneas | Assimétrico e variâncias heterogêneas |
| Observações | Reais | Reais ou convertidos para amplitude |
| Variáveis | Proporcionais ou de intervalos | Nominais, ordinais escala de intervalos |
| Centralização | Média | Medianas, modas |
| Dispersão | Variâncias | Amplitudes |
| Contagem | Dados que mostram frequências | Devem ser transformados |
| Tamanho da amostra | $n > 30$ | $n < 30$ |

Entre os testes paramétricos podemos mencionar: teste do valor Z da distribuição normal, teste de t de Student para dados relacionados (amostras dependentes), teste de t de Student para dados não relacionados (amostras independentes), teste t de Student-Welch para duas amostras independentes, com variâncias não homogêneas; teste de Chi^2 (ji quadrado) de Bartlett para demonstrar homogeneidade de variância; teste F (análise de variância ou ANOVA). Na tomada de decisão sobre os testes paramétricos, podemos esquematizar com o diagrama de fluxo que se mostra na Figura 1.

Figura 1 – Diagrama de fluxo sobre testes paramétricos

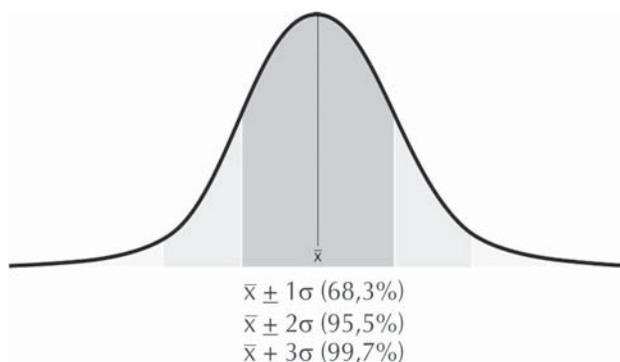


Fonte: *Guía de José Carlos Herrera F y Laura E, Carse.*

Quando os requisitos paramétricos não forem satisfeitos, deve-se, em primeiro lugar, transformar os dados para normalizar a distribuição e reduzir o viés. Para isso, podemos usar testes de raiz quadrada, logarítmicos, arco de seno ou Chi invertido.

Então, ao tomar a decisão sobre o teste a usar, a primeira pergunta a responder é se os dados seguem uma distribuição normal e se sua expressão é em média ou em desvio padrão (onde está 99,7% da população) seguindo, por exemplo, um gráfico como o sino de Gauss (Figura 2). Se esta resposta for positiva, o teste a usar será o paramétrico. A decisão seguinte é se quero comparar amostras diferentes ou se quero relacionar variáveis diferentes. Se a intenção for comparar duas amostras de dados independentes, usaremos o teste *t* de Student, não pareado; se os grupos forem dependentes, usaremos o teste *t* de Student pareado. Se forem mais de duas amostras, usaremos o ANOVA de uma ou duas vias, dependendo se forem dois ou mais fatores.

Figura 2 – Gráfico correspondente a um sino de Gauss, que mostra a distribuição normal dos dados com seus respectivos desvios padrão



Se nossa intenção for a de encontrar relações entre variáveis e quisermos conhecer um valor, já tendo conhecimento de outro, usaremos testes de regressão e, no caso contrário, usaremos testes de correlação.

Quando a distribuição dos dados não for normal, será expressa em termos de modo e amplitude, mediana, percentis, e serão usados os testes não paramétricos. Podemos mencionar: teste χ^2 de Pearson, teste binominal, teste de Anderson-Darling, teste de Cochran, teste de Cohen Kappa, teste de Fisher, teste de Friedman, teste de Kendall; teste de Kolmogórov-Smirnov, teste de Kruskal-Wallis, teste de Kuiper, teste de Mann-Whitney (Wilcoxon), teste de McNemar, teste da mediana, teste de Siegel-Tukey, teste dos sinais, coeficiente de correlação de Spearman, tabelas de contingência, teste de Wald-Wolfowitz e o teste das amplitudes com sinal de Wilcoxon.

As vantagens de usar testes não paramétricos é que são simples e fáceis de aplicar e usar; podem ser usados com grupos maiores de população; podem ser usados com dados ordinais ou nominais; podem ser usados em amostras pequenas.

Tamanho da amostra

Outro ponto importante, para decidir qual ou quais são as variáveis a usar, é o tamanho da amostra (n). Por isso, muitos pacotes estatísticos não permitem saber quantos dados devemos ter para que possamos extrapolar os resultados e para que seus relatórios sejam consistentes do ponto de vista da pesquisa.

Os autores não chegam a uma conclusão definitiva, mas existe consenso em relação ao tamanho da amostra; quanto maior a amostra, maior a possibilidade de diminuição do erro aleatório. O volume mínimo para considerar uma amostra como representativa poderia ser 20 ou 30.

Em suma, o tamanho da amostra é aquele necessário para a realização de um estudo com o menor número possível de elementos, que possibilite a obtenção dos resultados com a precisão estabelecida.

Tipo de variável

O primeiro passo para a análise dos dados é conhecer as variáveis com que se trabalha.

Os testes estatísticos têm maior potência se a variável for quantitativa, já que ela contém mais informações do que as variáveis qualitativas.

Segundo o tipo de variáveis e sua correlação, podemos selecionar diferentes testes. Se forem variáveis qualitativas, usaremos testes não paramétricos, como o teste χ^2 . Se uma for qualitativa e a outra quantitativa, poderemos selecionar entre alguns paramétricos (t de Student e análise de variância) e não paramétricos (U de Mann-Whitney; teste de amplitudes de Wilcoxon e o teste de Kruskal-Wallis). Se ambas as variáveis forem quantitativas, também poderemos selecionar entre paramétricos (coeficiente de correlação de Pearson) e não paramétricos (Coeficiente de Correlação de Spearman), Quadro 2.

Quadro 2 – Seleção de testes de acordo com diferentes variáveis

| Variável 1 | Variável 2 | Teste paramétrico | Teste não paramétrico |
|--------------|--------------|--|--|
| Qualitativa | Qualitativa | | Test de χ^2 |
| Quantitativa | Qualitativa | t de Student t de Student para dados pareados Análise de variância | U de Mann-Whitney Teste das amplitudes de Wilcoxon Teste de Kruskal-Wallis |
| Quantitativa | Quantitativa | Coeficiente de Correlação de Pearson | Coeficiente de Correlação de Spearman |

Outra maneira em que a relação entre as diferentes variáveis poderia indicar o tipo de teste a escolher seria:

- a. Duas variáveis contínuas com N grande, usa-se coeficiente de Pearson;
- b. Duas variáveis contínuas com n médio ou pequeno, usa-se coeficiente de correlação de Spearman ou Coeficiente de correlação de Kendall;
- c. Contínua *versus* ordinal; Coeficiente de correlação de Kendall ou o Teste de tendência de Cuzick;
- d. Contínua *versus* nominal; ANOVA paramétrico;
- e. Contínua *versus* dicotômica; teste *t* ou Teste Mann-Whitney;
- f. Duas variáveis ordinais; Coeficiente de correlação de Kendall; Teste de tendência de Cuzick ou Chi^2 ;
- g. Ordinal *versus* nominal; Teste de Chi^2 ;
- h. Ordinal *versus* dicotômica; Chi^2 de tendência linear;
- i. Duas variáveis nominais; Chi^2 ;
- j. Nominal *versus* dicotômica; Chi^2 sem tendência linear;
- k. Duas variáveis dicotômicas; Teste de Fisher; Risco Relativo ou Odds Ratio.

Pergunta

Como mostra a frase cunhada por Albert Einstein, “a formulação de um problema é mais importante que sua solução”, também a escolha do teste está relacionada com a pergunta que motivou o trabalho.

Se for só um grupo de análise, o teste a escolher pode ser o intervalo de confiança ou o teste *t*. Se estivermos observando duas amostras independentes, com uma variável quantitativa em escala contínua, e quisermos saber se existem diferenças estatisticamente significativas, devemos usar o teste de Student não pareado. Se, pelo contrário, forem dois grupos dependentes, com variáveis quantitativas, escalas descontínuas, os testes poderão ser: teste de Student pareado, Wilcoxon ou Correlação de Pearson. Se forem três ou mais grupos independentes, os testes podem ser ANOVA ou Chi^2 . Se forem três ou mais grupos dependentes, os testes de Cochran ou o ANOVA com observações repetidas. Se, ao contrário, forem grupos multivariados, os testes serão logarítmicos lineares ou regressão.

Tipo de estudo

Podemos selecionar o teste estatístico relacionando-o com o tipo de estudo que planejamos; por isso, poderemos saber quais os testes a usar, a partir do planejamento de nossa experiência.

Para os estudos de coorte, estudos transversais e ensaios clínicos, pode-se usar o risco relativo. Para os estudos de casos e controles, use-se o Odds Ratio.

Tipos de dados

Os grupos de dados podem ser independentes ou dependentes.

São independentes quando os grupos não têm relação e seus tamanhos são diferentes.

São dependentes quando existe relação direta entre eles, são grupos de causa e efeito, e o n de cada grupo é similar.

Comentários finais

O plano de análise estatístico é uma etapa definida antes da coleta dos dados.

É consequência da elaboração de um bom arcabouço teórico do assunto a pesquisar.

O projeto do formulário e o tamanho da amostra dependem da complexidade da análise estatística estabelecida.

Manuel Gomez Gomez e colaboradores ilustram este tema concluindo que a escolha do teste estatístico depende do(a): 1 – tipo de projeto de pesquisa; 2 – objetivo ou dos objetivos do trabalho; 3 – interesse particular dos autores; 4 – distribuição dos dados; 5 – maneira como foi feita a coleta dos dados; e 6 – potencial de viés e confusão dos dados obtidos.

LEITURA COMPLEMENTAR

1. Castañeda, MB; Cabrera, AF; Navarro, E; de Vries, W: Análisis de los datos (Cap 3; 26-54) e II (Cap 4; 55-70) en Procesamiento de datos y análisis estadísticas utilizando SPPS. Un libro práctico para investigadores y administrativos educativos. Ed. Edipruers, 2010.

2. Cazau, P: Fundamentos de estatísticas. UBA; 2006.
3. Gomez-Gomez, M; Danglot-Banck, C; Vega Franco, L: Cómo seleccionar una prueba estadística. Primera de dos partes. Rev Mex Pediatr 2013; 80 (1); 30-34.
4. Fernandez, P; Livacic-Rojas, P; Vallejo, G: Cómo elegir la mejor prueba estadística para analizar un diseño de medidas repetidas. Int J Clin Health Psychol; 7 (1): 154-175.
5. Herrera F, JC; Carse, LE: Guía de aplicación de pruebas estadísticas en el programa Systat 7.0 para ciencias biológicas e forestales. Sección 4.
6. Tetitla Munive, JM: Cómo elegir la prueba estadística. En Metodología de la pesquisa para el área de la salud. Ed McGraw Hill. 2013; Cap. 9: 71-79.
7. Universidad de Alcalá – Departamento de ecología: Métodos de análisis de datos en ecología.

CAPÍTULO 7

BIOESTATÍSTICA DESCRITIVA

Alberto Aceves Perez

Quando se fala em estatística, geralmente pensa-se numa organização de dados numéricos apresentados de forma ordenada e sistemática. Esta ideia é consequência de um conceito comum que existe sobre o termo e que se disseminou devido à influência do nosso meio, já que, atualmente, é quase impossível a um meio de comunicação não falar sobre algum tipo de informação estatística.

Desta maneira, e já falando especificamente no campo da pesquisa em medicina, a estatística começa a ser percebida como uma entidade diferente, como uma ferramenta que permite obter resultados e, portanto, benefícios em qualquer tipo de estudo, e cujos movimentos e relações, por sua natureza variável, não podem ser abordados, a partir de uma perspectiva determinante. A estatística pode ser definida como uma ciência que estuda como devem ser empregadas as informações e como dar diretrizes de ação em situações práticas, que têm como base uma pergunta determinada.

A estatística dedica-se aos métodos e procedimentos para coletar, classificar, resumir, identificar regularidades e analisar os dados, sempre e quando a variabilidade e a incerteza sejam causas intrínsecas dos mesmos. Além disso, faz inferências, a partir dos dados, com a finalidade de ajudar na tomada de decisões e, quando for o caso, a fazer previsões.

Também é considerada a arte de chegar a conclusões, a partir de uma série de dados imperfeitos. Geralmente, os dados são considerados imperfeitos porque, mesmo quando trazem informações úteis, não relatam a história completa. É necessário contar com métodos que permitam extrair informações, a partir dos dados observados, para compreender melhor as situações que eles representam.

Todos, até mesmo os especialistas em pesquisa, enfrentam problemas quando têm listas de dados. Existem diversos métodos estatísticos cujo propósito é ajudar a manifestar as características

proeminentes e interessantes dos dados, e que podem ser usados em quase todas as áreas do conhecimento.

Os métodos estatísticos podem e devem ser usados em todas as etapas de um processo de pesquisa, do início ao fim. Existe a convicção de que a estatística trata da análise dos dados, mas este ponto de vista exclui aspectos vitais relacionados aos projetos de pesquisa. É extremamente importante ter consciência de que a escolha do método de análise de um problema deve basear-se tanto no tipo de dados disponíveis quanto na forma como foram coletados esses dados.

Para que serve e por que estudar bioestatística?

A resposta poderia ser considerada simples: porque os dados estatísticos e as conclusões obtidas, aplicando o método científico, exercem uma influência profunda em praticamente todos os campos da atividade humana. Em particular, a estatística faz parte, cada vez mais, de qualquer pesquisa realizada no campo da medicina. Porém, este crescimento, provavelmente relacionado ao interesse em aumentar a credibilidade e confiabilidade das pesquisas, não garante que a metodologia estatística tenha sido usada corretamente, em todos os casos ou, ainda mais grave, que tenha validade.

Portanto, a importância da aplicação correta da bioestatística tem a ver com:

- a possibilidade de chegar a conclusões incorretas;
- a falta de consciência e de capacidade de muitos leitores de detectar o erro.

O estudo da bioestatística, assim como o modo de pensamento gerado a partir do mesmo, prepara as pessoas para avaliarem objetiva e efetivamente as informações que recebem (via tabelas, gráficos, percentuais, taxas, etc.) e confirmar se são ou não relevantes, adequadas, verdadeiras e objetivas. Logicamente, a interpretação de qualquer problema requer não apenas conhecimento metodológico, mas também conhecimento profundo sobre o assunto. Mesmo que uma pessoa não tenha interesse em se especializar em estatística, um treinamento básico no assunto permitirá sua melhor compreensão das informações.

Áreas da bioestatística

A seguir, uma breve descrição de cada uma das áreas nas quais pode ser dividida a bioestatística:

- projeto: planejamento e desenvolvimento de pesquisas;
- descrição: resumo e exploração dos dados;
- inferência: realização de previsões sobre as características específicas de uma amostra, com base nas suas informações.

Projeto

É uma atividade fundamental. Consiste em definir como será desenvolvida a pesquisa para dar resposta às perguntas que a motivaram. A coleta dos dados requer, geralmente, um grande esforço; e dedicar cuidado especial à etapa de planejamento da pesquisa; poupa trabalho nas etapas seguintes. Um estudo bem projetado fica simples de analisar e as conclusões costumam ser óbvias. Um experimento pobremente projetado, ou com dados coletados ou registrados de maneira inapropriada pode ser incapaz de dar respostas às perguntas que motivaram a pesquisa, por maior que seja a sofisticação da análise estatística.

Mesmo nos casos em que se estudam dados já registrados e nos restringimos a informações existentes, os princípios do bom projeto de experimentos podem ser úteis para ajudar a selecionar um conjunto razoável de dados relacionados ao problema de interesse.

Descrição

Os métodos da estatística descritiva ou análise exploratória de dados ajudam a apresentar os dados de maneira a destacar sua estrutura. Há várias formas simples e interessantes de organizar os dados em gráficos, que permitem detectar tanto suas características proeminentes como aquelas características inesperadas. Outro modo de descrever os dados é resumi-los em um ou dois números que pretendem caracterizar o conjunto, com a menor distorção ou perda de informação possível.

Explorar os dados deve ser a primeira etapa de toda análise de dados. Por que não analisá-los diretamente? Em primeiro lugar, porque os computadores não são muito hábeis (só são rápidos); fazem aquilo para o que foram programados e agem sobre os dados que nós lhes fornecemos. Dados errados ou inesperados serão processados de modo

inapropriado e nem você nem o computador o perceberão. Uma análise exploratória prévia dos dados é fundamental para a obtenção de resultados confiáveis.

Inferência

Refere-se a um conjunto de métodos que permitem fazer previsões acerca das características de um fenômeno com base em informações parciais do mesmo.

Os métodos da inferência permitem-nos propor o valor de uma quantidade desconhecida, ou decidir, entre duas teorias contrapostas qual delas explica melhor os dados observados. A finalidade principal de qualquer estudo é aprender sobre as populações, mas, geralmente, é necessário e mais prático estudar somente uma amostra de cada população.

Portanto, a estatística pode ser classificada como inferencial ou descritiva. Na primeira, o objeto de estudo consiste em inferir conclusões obtidas para um conjunto de dados mais amplo. Na segunda – que abordamos neste capítulo –, os resultados da análise pretendem ir além de um conjunto de dados.

Estatística descritiva: Descreve, analisa e representa um grupo de dados utilizando métodos numéricos e gráficos que resumem e apresentam as informações contidas neles.

Estatística inferencial: Apoia-se no cálculo de probabilidades e a partir de dados amostrais; toma decisões; faz estimativas, previsões ou outras generalizações sobre um conjunto maior de dados.

LEITURA COMPLEMENTAR

1. A. Martín Andrés, J.D. Luna del Castillo, Bioestadística para las Ciencias de la salud. Norma, Granada, 1994.
2. D. Peña Sánchez de Rivera, Estadística: Modelos y Métodos, Alianza Universidad Textos, Madrid, 1994.

3. E. Cobo, P. Muñoz y J.A. González (2007). Bioestadística para no estadísticos. Ed. Elsevier/Masson.
4. P. Armitage, G. Berry, Estadística para la Investigación Biomédica. Doyma, Barcelona, 1992.
5. J. Montanero Fernández, Manual de bioestadística. Universidad de Extremadura. Departamento de matemáticas. 2001.
6. J. S. Milton. Estadística para Biología y Ciencias de la Salud. Ed. Interamericana. McGraw-Hill.
7. Peña Sánchez de Rivera, D. (1987). "Estadística. Modelos y Métodos. Volumen 2". Alianza Editorial. Madrid.

CAPÍTULO 8

BIOESTATÍSTICA ANALÍTICA OU INFERENCIAL

José María Jiménez Avila

Muitas vezes, ao iniciar uma pesquisa, não é possível abranger toda a população, razão pela qual geralmente seleciona-se uma amostra, com a esperança de que ela seja representativa da população a estudar. Isso nem sempre acontece, já que elas podem não refletir a realidade do universo de onde surgem. Há condições do acaso que tem um papel muito importante no processo de pesquisa e, conseqüentemente, efeitos no seu resultado.

A área da matemática que estuda os resultados do acaso é a teoria das probabilidades, o núcleo da inferência estatística.

A análise inferencial ou analítica realiza-se após a análise descritiva que evidencia como é e quais são as características do grupo da população e as características particulares de nossa amostra.

Como definição, diríamos que a inferência estatística é o procedimento estatístico que permite estabelecer conclusões, considerando os efeitos da probabilidade (resultados), e geralmente é nesta etapa que são aceitas ou rejeitadas as hipóteses. É também nesta etapa que se estimam os padrões e a intensidade da associação entre as variáveis que vão emoldurar o resultado dentro do conceito: existem ou não diferenças entre os grupos.

A realização de experimentos dá origem a uma série de probabilidades, com resultados possíveis. Mediante a probabilidade do evento, pode-se expressar a confiança de que o evento ocorra ao observar o experimento, representando-se isso de "0" a "1". "1" indica que o evento ocorrerá com toda certeza, enquanto "0" corresponde a um evento que, com toda certeza, NÃO ocorrerá.

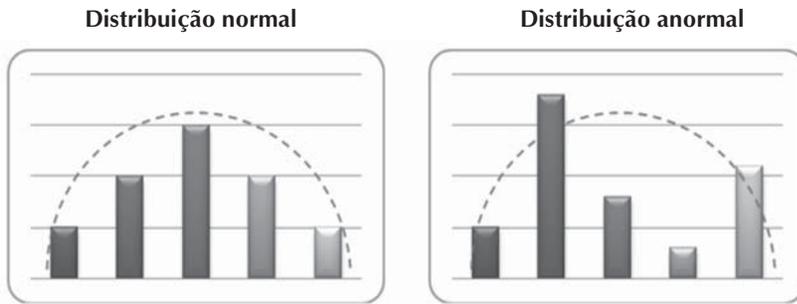
Existem dois tipos de inferência estatística: o Teste de Hipóteses e a Estimativa de Intervalo. (Serão revisados em capítulos posteriores.).

Tipos de testes

Os procedimentos estatísticos podem ser classificados em três grupos:

- aqueles que consideram a Distribuição Normal que, em conjunto, são reconhecidos como testes paramétricos;
- aqueles que Não consideram a distribuição normal dos dados, que são identificados como procedimentos livres de distribuição (distribuição binomial e de Poisson);
- aqueles que não se referem a parâmetros da população, conhecidos como testes não paramétricos. (χ^2 , teste exato de Fisher), Figura 1.

Figura 1 – Distribuição normal e anormal dos dados



Nota: Alguns textos de estatística agrupam os procedimentos não paramétricos e os livres de distribuição. Todos são apresentados com o título de não paramétricos.

Existe uma grande quantidade de testes que podemos classificar em dois tipos: Testes paramétricos e não paramétricos, Figura 2.

Figura 2 – Tipos de testes estatísticos

Testes Paramétricos em Cirurgia

t de Student
ANOVA
Correlação de Pearson

Testes Não Paramétricos em Cirurgia

χ^2
U de Mann-Whitney
Kruskal-Wallis
Correlação de Spearman
McNemar
Regressão Linear
Probabilidade de Poisson
Odds Ratio

Paramétricos

Os testes paramétricos são os considerados com maior potência estatística e que proporcionam mais informação, o que permite estabelecer estimativas dos parâmetros da população, através de amostras estatísticas.

Requisitos:

- variáveis com distribuição normal, que podem ser determinadas com o teste de Kolmogorov-Smirnov e com o teste de Shapiro-Wil;
- não usados para variáveis ordinais;
- necessitam de um número mínimo de 30 pessoas por grupo na amostra, já que quanto maior a amostra, mais exata será a estimativa probabilística;
- a hipótese é formulada sobre os valores numéricos, especialmente a média da população.

Não paramétricos

Os testes não paramétricos, ou de distribuição livre, são de fácil aplicação e podem ser usados para análise de variáveis ordinais ou normais, isto quando não for necessário fazer uma inferência sobre quais são os parâmetros da população.

Requisitos:

- variáveis de tipo ordinal ou nominal;
- quando duas séries de observações são provenientes de diferentes populações;
- a amostra pode ser menor ou igual a 20 pessoas;
- a hipótese é formulada sobre as amplitudes, a mediana ou a frequência dos dados.

É importante ter conhecimento de todos estes requisitos e familiarizar-se com todas as variáveis, já que a seleção do teste estatístico vai depender do tipo de variável, da presença ou ausência de normalidade da distribuição dos grupos que serão analisados, e de como foi formulada a pergunta de pesquisa (PICOT), Figura 3.

Figura 3 – Classificação de acordo com o objetivo e a variável

| Teste Estatístico segundo Objetivo e Variável | | | | | |
|---|-----------------|----------------------------|----------------------------------|------------------|---|
| Tipo de variável | Tipo de amostra | Diferença | | Mostrar relação | Prever 1 variável*** |
| | | 2 Grupos | 3 Grupos | 2 Variáveis | Variável |
| Quantitativa (distribuição normal) | NR R | t Student* t Student** | Anova 1 Factor Anova 1 Factor | Pearson | Regressão linear |
| Qualitativa ordinal | NR R | U Mann-Whitney Wilcixon | Krusj=kal-Wallis Friedman | Spearman | |
| Qualitativa dicotômica | NR R | X2 (Fischer) McNemar | X2 | Coefficiente phi | Regressão lógica Curvas de sobrevivência |

NR = Não relacionada R= Relacionada
 *t de Student para amostras independentes
 **t de Student para amostras relacionadas
 *** A variável preditiva pode ser quantitativa, dicotômica ou ordinal (transformar a variáveis tipo dummy)

Interpretação

Este tipo de teste permite ao cirurgião contar com ferramentas para fazer prognósticos sobre seus pacientes, como sobrevida ou possibilidades de infecção ou complicações de uma cirurgia. O cirurgião fará sua estimativa com base no seu conhecimento da literatura, das características do paciente, de suas expectativas e evolução, considerando sua própria experiência e os recursos à sua disposição para atendimento ao doente, além do possível desenlace.

Isto permite aplicar os conceitos fundamentais da Medicina baseada em evidências, considerada como o uso consciente, explícito e sensato da melhor evidencia científica disponível, para tomar decisões sobre os pacientes, Figura 4.

Figura 4 – O que significa fazer Medicina Baseada em Evidências?

LEITURA COMPLEMENTAR

1. Celis de la Rosa A. Bioestadística. 1. ed en español. Manual Moderno. 2004.
2. Daniel WW: Estadística con aplicaciones a las Ciencias Sociales y a la Educación. 1. ed. en español. McGraw-Hill, México. 1988.
3. Gómez Gómez M, Danglot Banck C, Velásquez Jones L. Sinopsis de pruebas estadísticas no paramétricas. Cuando usarlas. Revista Mexicana de Pediatría. Vol. 70 N.2 2003
4. Gómez Gómez M, Danglot Banck C, Velásquez Jones L. Bases para la revisión de artículos médicos. Revista Mexicana de Pediatría. Vol. 68 No. 4. Julio – Agosto 2001.
5. Pita Fernández, S. Tipos de estudios clínico epidemiológicos. Atención Primaria de la Red. 2001.
6. Siegel S, Castellan NJ: Nonparametric statistics for the behavioral sciences. 2. ed. McGraw-Hill, New York, 1988.
7. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992; 268: 2420-5.

CAPÍTULO 9

MEDIDAS DE TENDÊNCIA CENTRAL

Alberto Aceves Pérez

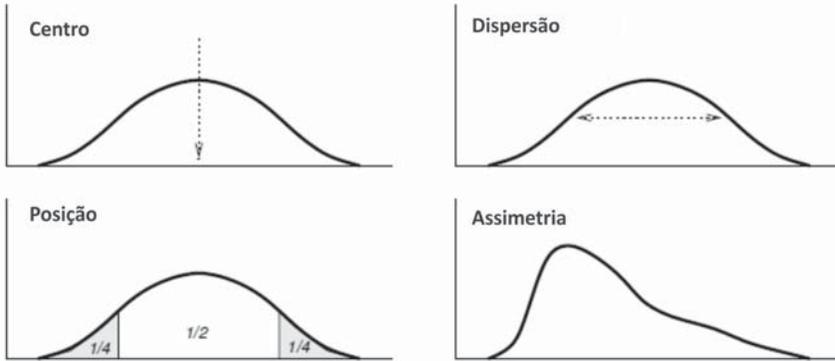
Após a elaboração da tabela e sua representação gráfica, na maioria das vezes, é mais eficaz “condensar” essa informação em alguns números que a expressem de forma clara e concisa.

Os fenômenos biológicos não costumam ser constantes, por isso será necessário associar a uma medida que indique o valor em torno do qual se agrupam os dados, uma medida que faça referência à variabilidade que reflita essa flutuação.

Portanto, o próximo passo e objeto deste capítulo será definir alguns tipos de medidas (estatísticas ou parâmetros) que os sintetizam ainda mais. Ou seja, diante de um grupo de dados organizados numa distribuição de frequências (ou uma série de observações sem ordenar), pretendemos descrevê-los mediante duas ou três quantidades sintéticas.

Neste sentido, podem ser examinadas diversas características, sendo que as mais comuns são:

- a tendência central dos dados;
- os dados ocupam certas posições;
- a simetria dos dados;
- a forma como os dados estão agrupados;
- a dispersão dos dados em relação ao centro.

Figura 1 – Medidas representativas de um conjunto de dados estatísticos

Neste capítulo, e seguindo esta ordem, iremos estudando a estatística que nos orientará sobre cada um destes níveis de informação: valores em torno dos quais a amostra está agrupada, a maior ou menor flutuação desses valores, e nosso interesse estará focado em certos valores que marcam posições características de uma distribuição de frequências, assim como sua simetria e sua forma.

É fácil identificar o centro, quando a distribuição é simétrica, mas quando é assimétrica, é difícil. Por este motivo, não existe uma única medida de posição para resumir uma distribuição. Se a distribuição for simétrica, diferentes medidas conduzirão a resultados similares. Se a distribuição for claramente assimétrica, diferentes propostas indicarão diferentes conceitos de “centro” e, portanto, os valores serão diferentes.

As três medidas mais comuns de tendência central são: média, mediana, moda.

Em certas ocasiões, estes três valores estatísticos coincidem, embora geralmente não seja assim. Cada um deles apresenta vantagens e inconvenientes que detalharemos mais adiante. Em primeiro lugar, vamos definir os conceitos anteriores.

A média

É a medida de posição mais frequentemente usada. Para calcular a média aritmética de um conjunto de observações, somam-se todos os valores e divide-se pelo número total de observações.

Se nossa amostra tem n observações, designadas por X_1, X_2, \dots, X_n , definimos a média amostral \bar{X} da seguinte maneira:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

Exemplo:

$$X_1 = 10 \quad X_2 = 14 \quad X_3 = 12 \quad X_4 = 11 \quad X_5 = 12 \quad X_6 = 13$$

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_6}{n} = \frac{10 + 14 + 12 + 11 + 12 + 13}{6} = \frac{72}{6} = 12$$

Características da média:

- usa-se para dados numéricos;
- representa o centro de gravidade ou o ponto de equilíbrio dos dados;
- a soma da distância dos dados até a média é zero;
- é muito sensível à presença de dados atípicos.

Embora a média seja a medida mais simples de tendência central, outras medidas oferecem-nos mais informações e, às vezes, são mais apropriadas.

A mediana

A *mediana* é o dado que ocupa a posição central na amostra ordenada de maneira crescente.

$$\left(\frac{n+1}{2} \right)$$

Exemplo:

$$X_1 = 10 \quad X_2 = 14 \quad X_3 = 12 \quad X_4 = 18 \quad X_5 = 11 \quad X_6 = 23$$

Ordenamos os dados:

$$10 \quad 11 \quad 12 \quad 14 \quad 18 \quad 23$$

$$\text{Posição da mediana} \Rightarrow \frac{6+1}{2} = 3.5$$

Obtemos a mediana fazendo a média do terceiro e quarto dado:

$$\tilde{X} = \frac{12+14}{2} = 13.$$

Características da mediana:

- como medida descritiva, tem a vantagem de não ser afetada pelas observações extremas, já que não depende dos valores que a variável adquire, mas da ordem das mesmas. Por isso, seu uso é adequado em distribuições assimétricas;
- seu cálculo é rápido e a interpretação é simples;
- diferentemente da média, a mediana é sempre um valor da variável que estudamos.

A moda

A moda é o dado que ocorre com maior frequência no conjunto.

É uma medida de pouca utilidade, exceto para dados categóricos em que o interesse é identificar a categoria com maior quantidade de dados. Em uma amostra de dados numéricos, pode ocorrer que a moda seja um valor que se repete um certo número de vezes, mas que não é típico.

Características da moda:

- é muito fácil de calcular;
- pode não ser única.

Figura 2 – Resumo das medidas de tendência central

| | | MEDIDAS DE TENDÊNCIA CENTRAL | | | | |
|---------|--|--|---|---|---|---|
| | | DADOS SEM AGRUPAR | DADOS AGRUPADOS | | | |
| | | (ordenados) x_1, x_2, \dots, x_N | Interv. l_0-l_1 l_1-l_2 \dots $l_{k-1}-l_k$ | x_i x_1 x_2 \dots x_k | n_i n_1 n_2 \dots n_k | N_i N_1 N_2 \dots N_k |
| MÉDIA | | $\bar{x} = \frac{x_1 + \dots + x_n}{N}$ | $\bar{x} = \frac{n_1 x_1 + \dots + n_k x_k}{N}$ | | | |
| MEDIANA | | Primeira observação que deixa abaixo de si estritamente as $[N/2]$ observações menores $x_{[N/2]+1}$ | $M_{ed} = l_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} \cdot a_i$ | | | |
| MODA | | $M_{oda} = x_i$ de maior frequência | $M_{oda} = l_{i-1} + \frac{n'_i - n'_{i-1}}{(n'_i - n'_{i-1}) + (n'_i - n'_{i+1})} a_i$ | | | |

Relação entre a média, a mediana e a moda

No caso de distribuições unimodais, a mediana está, frequentemente, compreendida entre a média e a moda (inclusive, mais próxima da média).

Em distribuições que apresentam alguma inclinação, é mais aconselhável o uso da mediana. No entanto, em estudos relacionados com propósitos estatísticos e de inferência, a média costuma ser mais apta.

LEITURA COMPLEMENTAR

1. A. Martín Andrés, J.D. Luna del Castillo, Bioestadística para las Ciencias de la salud. Norma, Granada, 1994.
 2. E. Cobo, P. Muñoz y J.A. González.(2007) Bioestadística para no estadísticos. Ed. Elsevier/Masson.
 3. D. Peña Sánchez de Rivera, Estadística: Modelos y Métodos, Alianza Universidad Textos, Madrid, 1994.
 4. J. Montanero Fernández, Manual de bioestadística. Universidad de Extremadura. Departamento de matemáticas. 2001.
 5. J. S. Milton. Estadística para Biología y Ciencias de la Salud. Ed. Interamericana. McGraw-Hill.
 6. P. Armitage, G. Berry, Estadística para la Investigación Biomédica. Doyma, Barcelona, 1992.
- Peña Sánchez de Rivera, D. (1987): "Estadística. Modelos y Métodos. Volumen 2". Alianza Editorial. Madrid.

CAPÍTULO 10

MEDIDAS DE DISPERSÃO

(Amplitude, Variância, Desvio Padrão, Coeficiente de Variâncias)

Delio Martins

Um aspecto importante no estudo descritivo dos dados de um trabalho é a determinação da variabilidade ou dispersão desses dados em relação à medida de localização do centro da amostra.

Por exemplo, em um estudo sobre escoliose, em que foi avaliado o efeito de um determinado tratamento cirúrgico A, comparado com outro tratamento cirúrgico B, em crianças com graus diferentes de curvatura. A média da magnitude das curvas foi de 60 graus Cobb tanto no grupo operado quanto no de controle, e ambos os grupos apresentaram moda e mediana iguais. Será que isso é suficiente para que eu possa dizer que esses grupos são similares?

Vejamos alguns conceitos importantes...

Amplitude

Essa é uma medida em função do maior e menor valor em um conjunto de dados.

Assim, se considerarmos o exemplo acima do grau de curvatura da escoliose das crianças e observarmos os dois grupos com as seguintes magnitudes de curvas:

Tratamento A: 40, 42, 44, 50, 65, 65, 65, 68, 70, 70, 81

Tratamento B: 48, 48, 50, 54, 65, 65, 65, 65, 66, 66, 68

A amplitude corresponde à diferença entre a observação mais elevada e a mais baixa. Nesse caso observamos que, no tratamento A, a amplitude foi de 41 graus enquanto no tratamento B foi de apenas 20 graus, demonstrando uma grande variação entre os grupos.

Sem dúvida nenhuma a amplitude é a medida de dispersão mais facilmente visualizada; no entanto, baseia-se apenas em dois valores e não considera a distribuição dos dados.

Variância

Vejamos outra medida de dispersão, a Variância

Para simplificar nosso entendimento, vamos diminuir a nossa amostra:

Tratamento A: 40, 50, 65, 81

Tratamento B: 48, 54, 66, 68

Em ambos os casos, temos uma média de 59 graus Cobb:

$$A: (40 + 50 + 65 + 81 / 4) = 59$$

$$B: (48 + 54 + 66 + 68 / 4) = 59$$

As amplitudes continuam diferentes:

$$A = 81 - 40 = 41$$

$$B = 68 - 48 = 20$$

Notamos que o grupo do tratamento A é mais disperso do que o grupo do tratamento B.

Para iniciarmos o entendimento do desvio-padrão, faremos, primeiramente, o cálculo do grupo A da distância que cada observação tem da média do grupo. Assim teremos:

$$(40 - 59, 50 - 59, 65 - 59, 81 - 59)$$

Dessa forma, temos os valores:

$$(-19, -9, 6, 22)$$

Agora um teste: calcule a média aritmética desses números...

Calculou???

Observou que a soma de todos os desvios em relação à média aritmética é igual a zero? Isso é uma propriedade aritmética que sempre vai se repetir!

Dessa forma, os valores negativos vão sempre anular os positivos e **não** é possível obter uma medida de dispersão utilizando este cálculo algébrico!

Vamos tentar de outra forma então...

Como no cálculo o que importa é apenas a magnitude do desvio e não se são positivos ou negativos, vamos fazer uma forma em que todos os números fiquem positivos!!

Eleve todos os números ao quadrado!

$$(-19)^2, (-9)^2, (6)^2, (22)^2$$

ou:

$$361, 81, 36, 484$$

Dessa forma, a média dos quadrados dos números torna-se uma medida razoável de variabilidade.

No entanto, para calcular a média desses valores, de tal forma que tenha propriedades que funcionem com todas as teorias (aqui acredite: isso é mais do que você vai querer saber sobre esse assunto), não dividimos pelo número total de casos (aqui: $n=4$), mas sim dividimos pelo número total - 1 ($n-1$) que, nesse caso, é ($4-1=3$).

Dessa forma:

$$361 + 81 + 36 + 484 / 3 = 320,6$$

Assim calculamos a variância da amostra. Vamos ao próximo passo...

Desvio padrão

O desvio padrão talvez seja a média de dispersão mais empregada em estatística, mas também talvez a menos entendida. No entanto, agora que entendemos a variância, isso vai ficar mais fácil para a gente.

Como o próprio nome diz, é um padrão de desvio da média. Representa a distância normal de qualquer ponto no conjunto de dados até a média.

O desvio padrão nada mais é que a raiz quadrada da variância. Assim evitamos ficar trabalhando com quadrados de valores.

No nosso exemplo, então, o desvio padrão é $\sqrt{320,6} = 17,90$.

Desvio padrão da população inteira é representado pela letra grega σ , mas como raramente conhecemos a população inteira, e normalmente trabalhamos com uma amostra, o desvio padrão de uma amostra da população é designado pela letra S .

Mais importante do que calcular o desvio padrão é entender o seu significado. Um desvio padrão pequeno significa que os valores do conjunto de dados estão, na média, próximos do centro do conjunto. O desvio padrão alto significa que os valores do conjunto de dados estão, na média, distantes do centro.

Um desvio padrão alto não significa, necessariamente, algo ruim. Ele apenas reflete uma grande variabilidade dentro do grupo avaliado.

Outro cuidado é observar sempre a unidade em que está o desvio padrão. Por exemplo, se considerar um desvio padrão de dois anos, em uma determinada amostra, isso pode mudar para um desvio padrão de 24, caso os dados estejam em meses!

Em uma curva de distribuição normal, presente na maioria dos trabalhos científicos, temos por padrão que 68% da amostra está distante até um desvio padrão da média, e que 95% da amostra está distante até dois desvios padrão da média. Assim, por exemplo, se temos um cálculo da quantidade de proteoglicano presente no disco intervertebral, e nos é informado que a média de proteoglicano foi de 250 mg/g e que o desvio-padrão foi de 8,75 mg/g nessa amostra, sabemos que 68% dos discos apresentam dosagem de proteoglicanos dentro de um desvio padrão (8,75 mg/g) da média (250 mg/g). Portanto, 68% dos discos continuam entre $250 - 8,75$ mg/g e $250 + 8,75$ mg/g de proteoglicano, ou seja, 68% dos discos da amostra continuam entre 241,25 e 258,75 mg/g de proteoglicano. Da mesma forma, 95% das amostras medidas têm entre $250 - 2(8,75$ mg/g) e $250 + 2(8,75$ mg/g), ou seja, 95% das amostras contêm entre 232,5 e 267,5 mg/g de proteoglicanos.

Algumas propriedades importantes do desvio padrão:

– nunca pode ser um número negativo (devido à forma como é calculado, e como mede uma distância; e distâncias nunca são números negativos);

- o menor valor possível é zero, e isso acontece somente em situações planejadas, em que cada número do conjunto de dados é igual! – Não há variação;
- é influenciado por valores discrepantes dentro do conjunto (ou muito altos ou muito baixos). Isso porque, lembre-se: o desvio padrão baseia-se na distância dos dados com relação à média;
- o desvio padrão tem a mesma unidade dos dados originais.

Sem o desvio padrão, não se consegue saber se os dados estão próximos da média ou se estão muito espalhados. Por exemplo, se o reembolso de uma cirurgia que dure 1h, em média é de \$ 7000,00, pode-se pensar: Uau!! Que maravilha! Mas, e se o desvio padrão para os reembolsos da cirurgia foi de \$ 2000,00 e utilizarmos a regra empírica de distribuição normal, essa mesma cirurgia poderia ser reembolsada entre \$ 3000,00 e \$ 11.000,00 (ou seja, \$ 7000,00 mais ou menos dois desvios padrão). Dessa forma, você nota que existe uma variabilidade muito grande no reembolso dessa cirurgia, assim apenas a média que foi informada não reflete tão bem o que esperar desse procedimento.

Coeficiente de variância

Por fim, vamos entender o coeficiente de variância. Quando comparamos a variabilidade de uma medida entre dois grupos, o desvio padrão pode ser empregado adequadamente. No entanto, se os grupos tiverem médias muito discrepantes, a comparação direta dos desvios padrão pode induzir a conclusões erradas, principalmente quando se espera que o grupo com maior média tenha maior variabilidade.

Por exemplo, se compararmos, hipoteticamente, a altura de vértebras de adultos que suponhamos que tenha um desvio padrão de 5,9 mm, com a altura de vértebras de recém-nascidos, que têm um desvio padrão de 4,0 mm. O desvio padrão é maior nos adultos, o que sugere que sua variabilidade é superior a dos recém-nascidos. No entanto, os adultos têm vértebras bem maiores, suponhamos em torno de 50 mm enquanto os recém-nascidos têm vértebras bem menores, suponhamos em torno de 25 mm. Dessa forma, será melhor comparar a variação de altura das vértebras em relação à média. Assim, pode ser aplicado o coeficiente de variação, que é, simplesmente, o desvio padrão expresso como uma porcentagem da média.

Assim a fórmula do coeficiente de variância (CV) é:

$$CV = \frac{\text{Desvio padrão}}{\text{Média}} \times 100$$

O CV independe de unidade de medida e é expresso em porcentagem. Nesse exemplo, o CV dos adultos é de 11,8% e o CV dos recém-nascidos é de 16%. O que demonstra que a altura da vértebra do recém-nascido é mais variável que a de um adulto, ainda que menos variável em termos absolutos.

Em suma, o estudo dos dados descritivos de uma amostra precisa ser avaliado em conjunto com as medidas de tendência central com as medidas de dispersão, para que se possa ter uma análise real dos dados.

LEITURA COMPLEMENTAR

1. Blair, RC; Taylor, RA. Bioestatística para ciências da saúde. Pearson Education do Brasil, 2013.
2. Daly, LE; Bourke, GJ. Interpretação e aplicações da estatística em medicina. Instituto Piaget. 2000.
3. Griffiths, D. Use a cabeça! Estatística. Alta Books Editora. Rio de Janeiro. 2008.
4. Levine, DM; Berenson, ML; Stephan, D. Estatística: Teoria e Aplicações usando Microsoft Excel em Português. Rio de Janeiro: LTC, 2000.
5. Rumsey, DJ. Estatística para leigos. Alta Books Editora. Rio de Janeiro. 2012.
6. Triola, MF. Introdução à Estatística. 7. ed. Rio de Janeiro: LTC, 1999.

CAPÍTULO 11

APRESENTAÇÃO DOS DADOS

José María Jiménez Avila

A apresentação dos dados é uma das partes do manuscrito mais usadas na estatística, já que a partir de diferentes tipos de gráficos é possível evidenciar e analisar os resultados com rapidez.

Um quadro ou uma figura bem desenhada mostra, com ordem, o maior número de ideias na menor quantidade de espaço. É importante que imite o formato de quadros ou figuras da revista que tenha escolhido para sua próxima publicação (títulos, linhas, cores, variáveis e números decimais). É importante também numerar os quadros e as figuras de maneira consecutiva, e redigir o título de maneira breve, informativa e precisa. Algo que não pode ser esquecido é a elaboração de notas sob os quadros, e figuras com observações e abreviaturas.

Na análise dos resultados, um detalhe que não deve ser esquecido é manter o objetivo da pesquisa, razão pela qual devem ser selecionados os dados que reflitam, por si só, o resultado da mensagem que se quer passar, que seja fácil de entender e permita compreender a relação que os dados guardam entre si.

Deve incluir: título, corpo da figura, notas explicativas.

Geralmente, o número total recomendado para uso de quadros e figuras em artigos científicos é de 6, por exemplo, 2 e 4, 3 e 3, etc.

Não seja reiterativo com as informações que apresenta, ou seja, não repita no texto o que descreve nos quadros e/ou figuras.

A qualidade da apresentação dos resultados é muito importante para a adequada compreensão e interpretação dos resultados, que serão obtidos da pesquisa.

Tipos de representação gráfica

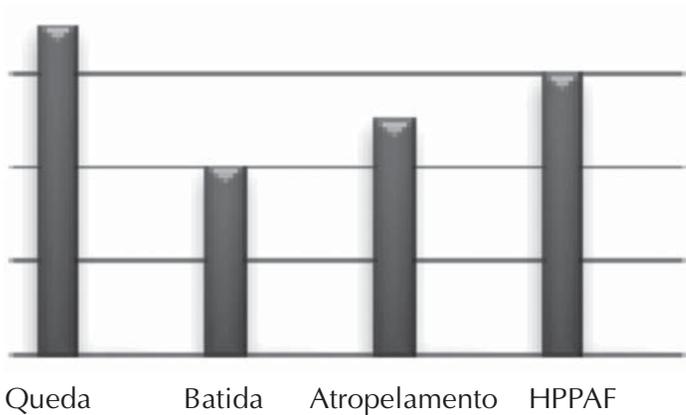
Diagrama de barras

Este tipo de representação gráfica é muito útil para evidenciar variáveis de tipo quantitativas discretas e qualitativas em escala nominal.

Sugere-se que as barras tenham a mesma largura e que estejam separadas por um espaço que não deve ser maior que a espessura das barras.

A sugestão é apresentá-las de maneira decrescente, a não ser que exista uma sequência predeterminada que deva ser seguida.

Exemplo:



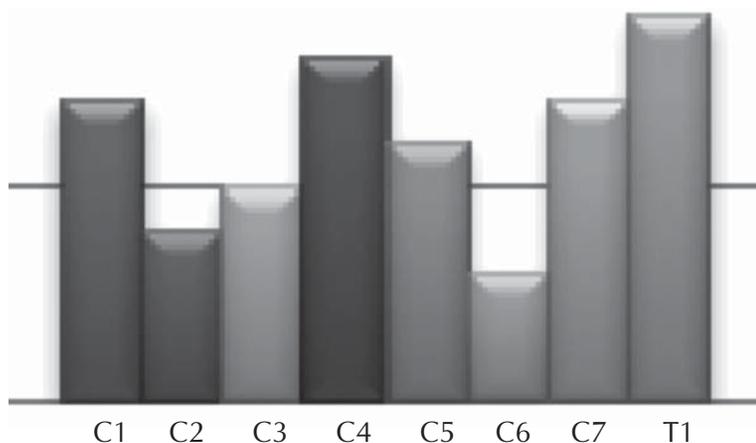
Aplicação na cirurgia:

Distribuição de 60 pacientes que tenham sofrido fratura de coluna lombar, conforme o mecanismo da lesão. Jalisco, 2014-2015.

Histograma

Este tipo é adequado para variáveis quantitativas contínuas, porque pode agrupar as frequências de apresentação das variáveis em intervalos.

Exemplo:



Aplicação na cirurgia:

Distribuição de 100 pacientes com fratura cervical e torácica alta. Jalisco. 2014-2015.

Diagrama de pizza

Este tipo de representação é adequado para variáveis qualitativas em escala nominal.

Exemplo:



Aplicação na cirurgia

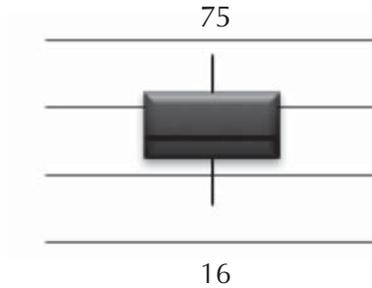
Distribuição de 60 pacientes com fratura de coluna, conforme sua localização anatômica. Jalisco. 2014-2015.

Observação: Prestar atenção ao usar este tipo de gráfico, já que pode mostrar evidência muito óbvia que, em algum momento, poderia ser omitida dentro do manuscrito, quando não faz parte da mensagem principal.

Diagrama de caixa

Este é adequado para variáveis quantitativas nas quais queremos demonstrar o desvio da distribuição normal. Geralmente, o extremo superior da caixa corresponde ao chamado percentil 75, o extremo inferior corresponde ao percentil 25, e a linha horizontal que dividirá a caixa corresponderá à média.

Exemplo:



Aplicação na cirurgia

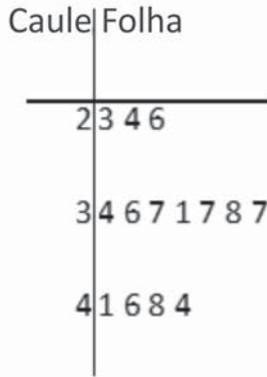
Distribuição das idades em um grupo de 60 pacientes com fratura de coluna. Jalisco, 2014-2015,

Diagrama de caule e folhas

Este diagrama é útil para variáveis quantitativas que têm poucos valores, já que neste tipo de apresentação cada um dos valores das variáveis representadas agrupa-se em relação à sua distribuição. É possível que tenha semelhança com o histograma.

Os valores aleatórios são ordenados de maneira decrescente, e a folha de cada coluna corresponde ao último valor de cada leitura, anotando-se o número de repetições e formando o equivalente à coluna do histograma. O caule corresponde ao restante da leitura e escreve-se somente uma vez para cada região.

Exemplo:



Aplicação na cirurgia

Distribuição de tipo de fratura (subescala) em um grupo de 14 pacientes. Jalisco, 2014-2015.

Geralmente, a primeira representação deve incluir os dados demográficos do grupo de estudo, nos quais as características gerais possam ser observadas objetivamente, ou seja, a análise descritiva da amostra.

LEITURA COMPLEMENTAR

1. Camel F: Estadística Médica y Planificación de la Salud. Tomo I. 1ra edición Universidad de los Andes. Consejo de Publicaciones. Mérida, Venezuela, 1991.
2. Gómez Gómez M, Danglot Banck C, Velásquez Jones L. Sinopsis de pruebas estadísticas no paramétricas. Cuando usarlas. Revista Mexicana de Pediatría. Vol. 70 No.2, 2003.
3. Talavera J. O, Rivas Ruiz R. Pertinencia de la prueba estadística seleccionada. Revista Médica del Instituto Mexicano del Seguro Social. Vol. 49 No. 2. 2011.

CAPÍTULO 12

TESTE DE HIPÓTESE (IC) E VALOR DE “p”

José María Jiménez Avila

Descrição do teste

Um teste de hipótese é definido como um teste estatístico, que pode ser usado para determinar se existe *evidência suficiente* numa **amostra** de dados para fazer a inferência de que uma determinada condição pode ser válida para toda a **população**.

Ou seja, um teste pode examinar duas hipóteses opostas: a hipótese nula (H_0) e a hipótese alternativa (H_a). A primeira é a que será testada. Geralmente, a hipótese nula estabelece um enunciado de que “não há efeito” ou “não há diferença”, e a hipótese alternativa é aquela que se deseja poder concluir que é verdadeira, estabelecendo que “há diferenças.”

Em virtude dos dados da amostra, o teste poderá determinar se a hipótese nula deve ser rejeitada. Para tomar essa decisão, usa-se um valor denominado “p”, que se for $<$ (menor) ou igual ao nível de significância (o ponto de corte definido), indicará que a hipótese nula poderá ser rejeitada.

Um erro comum de percepção é que os testes estatísticos de hipótese são desenhados para selecionar o que é mais provável em duas hipóteses, o que significa que um teste manterá a validade da hipótese nula, até que exista evidência suficiente (resultados) a favor da hipótese alternativa.

Exemplo:

A cirurgia minimamente invasiva tem melhores resultados que a cirurgia convencional aberta?

Isso pode ser verificado mediante dois tipos de testes: Teste de hipótese; Estimativa (Intervalo de confiança).

Teste de hipótese

Formula-se uma hipótese com a intenção de comparar duas populações. Quando não existe a possibilidade de estudar a totalidade do universo, seleciona-se uma amostra de cada grupo inferindo os parâmetros dos dois grupos, com a finalidade de obter informações suficientes para tomar a decisão correta sobre a “equidade” ou a “diferença” dos dois grupos.

Para isso, devem ser contemplados vários aspectos:

– ambas as populações são idênticas e não há diferenças entre os grupos;

– selecionar do nível de significância, mediante um teste estatístico que pode ser o valor de “p”.

a) $p > 0.05$: Aceita-se a hipótese nula. Ou seja, “não há diferença”, é igualmente efetiva (Efeito 0).

b) $p < 0.05$: Rejeita-se a hipótese nula, ou seja, “há diferenças”, e define-se como a probabilidade de que as diferenças sejam significativas, e não devidas ao acaso (Efeito 1).

Observações interessantes

Ao assumir a hipótese nula, a reflexão é que não existem mais diferenças do que aquelas devidas ao acaso.

A probabilidade implica incerteza, por menor que seja o valor de “p”, nunca podemos ter certeza de contar com a verdade absoluta. Ter obtido um $p < 0.0000001$, apenas indica que a probabilidade de que H_0 esteja certa é mínima, mas jamais será nula.

Do ponto de vista médico-cirúrgico, um teste estatisticamente significativo, não é sinônimo de clinicamente importante.

Estimativa (intervalo de confiança)

Intervalo de confiança (IC) é o intervalo no qual sabemos que se encontra um parâmetro, com um nível de confiança específico.

Traduz-se como a probabilidade de que o parâmetro a estimar esteja no intervalo de confiança, sendo os mais utilizados de 95% ou 99%.

Exemplo de interpretação: se tivéssemos uma população cuja média de idade está entre 40.2 e 53.6 e selecionarmos 100 amostras; 95 vão cobrir o valor verdadeiro da média da população e cinco não o farão.

Ou seja, estabelece-se uma estimativa de aproximação do parâmetro correspondente à população da qual a amostra foi retirada.

Tipos de erro

Ao realizar e identificar as inferências, há riscos que podem levar a certos tipos de erros, os quais são denominados: erro tipo I e erro tipo II.

- a. Erro tipo I: Rejeitar a hipótese nula quando a hipótese nula é verdadeira: a probabilidade de cometer este tipo de erro é justamente α . Por isso, este erro denomina-se erro α .
- b. Erro tipo II: Aceitar a hipótese nula quando a hipótese nula é falsa. Este é conhecido como erro de aceitação, ou erro β :

Exemplos:

1. O acusado era inocente e o júri declarou sua inocência: OK! Está tudo certo: bom trabalho: a pessoa sai em liberdade, e não há problema algum.
2. O acusado era culpado (ele a matou!!), e o júri (muito esperto!), declara que ele é culpado: prisão perpétua: Tudo OK! (era o que merecia!).

Mas, o júri pode cometer dois tipos de erros:

3. Declará-lo culpado e o pobre coitado ser inocente!!!: Erro tipo I,
4. Declará-lo inocente e o desgraçado ser culpado!!!: Erro tipo II.

1. Se o valor de “p” mostrar que a probabilidade de ter encontrado esse valor por acaso é muito baixa, rejeitaremos a hipótese nula: diremos que a diferença é estatisticamente significativa: Se nos enganarmos, estaremos cometendo um erro tipo I.

2. Se o valor de p for muito alto, diremos que a probabilidade de ter encontrado esse valor por acaso é alta. Portanto, aceitaremos a hipótese nula: diremos que as diferenças encontradas não eram estatisticamente significativas: se nos enganarmos (e SIM, havia diferenças), estaremos cometendo um erro de tipo II.

O valor de “p” e o nível de significância

É a probabilidade de que o acaso por si só possa produzir uma diferença entre os dois grupos comparados.

Se o valor de “p” for grande ($p > 0.10$ ou $p > 0.25$), a diferença observada entre as incidências poderá ser devida ao acaso e, portanto, a associação será considerada como *não significativa*.

Se o valor de “p” for pequeno ($p < 0.05$ ou $p < 0.01$), a diferença observada terá uma probabilidade muito pequena de ter ocorrido por acaso e é quase certo que se deve a uma diferença real entre as incidências das populações em estudo.

Exemplo na cirurgia de coluna

Você gostaria de determinar se a cirurgia minimamente invasiva tem um efeito benéfico sobre a cirurgia convencional aberta, nos pacientes que sofrem de dor lombar provocada por uma hérnia de disco?

1º Formular a pergunta de pesquisa (PICOT).

2º Estabelecer uma postura com base na possível comparação das duas opções.

3º Transformar a pergunta em duas hipóteses

Hipótese nula: H_0 : Os resultados são iguais.

Hipótese alternativa: H_a : Os resultados são diferentes.

4º Obter o valor de “p” mediante testes estatísticos.

Este mostrará duas possibilidades: $p > 0.05$ ou $p < 0.05$.

Interpretação

A maneira mais comum é comparar o valor de “p” com o nível de significância (α); α é a probabilidade de rejeitar H_0 , quando H_0 for verdadeira.

O valor de “p” indica o grau em que a evidência da amostra apoia a rejeição de H_0 . Geralmente, quanto menor for o valor de “p”, maior será a influência da evidência da amostra para rejeitar H_0 .

Mais especificamente, o valor de “p” é o menor valor de α que leva à rejeição de H_0 .

Para qualquer valor de $\alpha >$ ao valor “p”, H_0 não pode ser rejeitada, e para qualquer valor de $\alpha \leq$ ao valor de “p”, rejeita-se H_0 .

Consideremos o exemplo da cirurgia minimamente invasiva, na qual são comparados alguns aspectos relacionados à melhora, em comparação com a cirurgia convencional aberta. No caso hipotético de que, em média, 25 pacientes tenham apresentado melhores resultados, o estatístico de teste vai depender da média. Se o valor de “p” fosse, por exemplo, .026, indicaria que 2.6% das amostras com melhora no resultado, extraídas da população com $\mu = 25$, produzirá uma média com uma evidência tão (ou mais) forte que a amostra atual onde μ não é igual a 25. Agora, questionemos o que é mais provável: que a $\mu = 25$ e você simplesmente selecionou uma amostra muito pouco comum, ou que a μ não seja igual a 25?

Tradicionalmente, o valor de “p” é comparado a valores de α menores, que .05 ou .01, dependendo do campo de estudo.

No exemplo, suponhamos que o valor de α seja .05. O valor de “p” de .026 indica que a média (não apenas a média dos demais pacientes incluídos no estudo) provavelmente não é igual a 25. Uma forma mais correta de expressar isto, em termos estatísticos seria: “a um nível de significância de .05, a média de melhora em decorrência deste tipo de cirurgia parece ser significativamente diferente de 25”.

Isto é interpretado da seguinte maneira:

“Melhora de resultados com o uso da cirurgia minimamente invasiva” --- $p < 0.05$.

O uso da cirurgia minimamente invasiva tem uma probabilidade maior de obter melhores resultados do que a cirurgia convencional aberta.

Para isso, podemos assumir que poderia existir uma associação que não se deve ao acaso e que, provavelmente, gere o resultado mencionado.

Usar os valores de “p” é simples, se tivermos conhecimento dos dados-chave: os valores de α que são aceitáveis no seu campo, e as hipóteses nula e alternativa para os testes que serão usados.

LEITURA COMPLEMENTAR

1. Celis de la Rosa A. Bioestadística. 1. ed en español. Manual Moderno. 2004.
2. Daniel WW: Estadística con aplicaciones a las Ciencias Sociales y a la Educación. 1ª ed en español. McGraw-Hill, México. 1988.
3. Gómez Gómez M, Danglot Banck C, Velásquez Jones L. Sinopsis de pruebas estadísticas no paramétricas. Cuando usarlas. Revista Mexicana de Pediatría. Vol. 70, N. 2 2003.
4. Talavera J. O, Rivas Ruiz R. Pertinencia de la prueba estadística seleccionada. Revista Médica del Instituto Mexicano del Seguro Social. Vol. 49 No. 2. 2011.

CAPÍTULO 13

TESTE DE KOLMOGOROV-SMIRNOV

*Nelson Astur Neto
Maximiliano Gimenez Gigón
Rodrigo Góes Medéa de Mendonça*

Descrição

O teste de Kolmogorov-Smirnov (KS) tem por objetivo determinar se dois conjuntos de dados diferem significativamente.

A vantagem do teste de KS é a de não fazer suposições sobre a distribuição dos dados. Tecnicamente, é um teste não paramétrico e sem distribuição. Por outro lado, tem a desvantagem de que outros testes podem ser mais sensíveis, quando os dados contemplam os requisitos dos mesmos.

No âmbito da estatística, os testes aplicados na pesquisa são os paramétricos e não paramétricos. Entre os testes não paramétricos – geralmente usados para verificar se uma distribuição ajusta-se ou não a uma distribuição esperada, particularmente à distribuição normal –, temos o teste de Kolmogorov-Smirnov. O teste de Kolmogorov-Smirnov é bem potente com amostras grandes. O nível de medição da variável e sua distribuição são elementos que intervêm na seleção do teste que será usado, no processamento posterior. Na verdade, se a variável for contínua com distribuição normal, poderão ser aplicadas técnicas paramétricas. Se for uma variável discreta ou contínua não normal, só poderão ser aplicadas técnicas não paramétricas, pois a aplicação das primeiras traria resultados de validade duvidosa.^{1,4,3,5}

Fórmula

Hipóteses a comparar:

H_0 : Os dados analisados seguem uma distribuição M .

H_1 : Os dados analisados não seguem uma distribuição M .

Comparar distribuição empírica da amostra com a distribuição proposta sob H_0 .

Se esta comparação revelar diferenças significativas, rejeita-se H_0 .

Estatística de contraste:^{2,4}

$$D = \sup_{1 \leq i \leq n} \left| \hat{F}_n(x_i) - F_0(x_i) \right|$$

Onde:

- é i -ésimo o valor observado na amostra (cujos valores foram previamente ordenados de menor a maior).
- é uma estimativa da probabilidade de observar valores menores ou iguais a x_i .
- é a probabilidade de observar valores menores ou iguais a x_i , quando H_0 for verdadeiro.

Então, D é a maior diferença absoluta observada entre a frequência acumulada observada $F_n(x)$ e a frequência acumulada teórica $F_0(x)$, obtida a partir da distribuição de probabilidade, que se especifica como hipótese nula.

Se os valores observados $F_n(x)$ forem similares aos esperados $F_0(x)$, o valor de D será pequeno.

Quanto maior for a discrepância entre a distribuição empírica $F_n(x)$ e a distribuição teórica, maior será o valor de D .

Portanto, o critério para tomada da decisão entre as duas hipóteses será da forma₃:

| |
|--|
| <p>Se $D \leq D_\alpha \Rightarrow$ Aceitar H_0 Se $D > D_\alpha \Rightarrow$ Rejeitar H_0</p> |
|--|

Aplicação

Aplicações do teste:^{4,5}

Comparar se um conjunto de dados amostrais pode ser considerado procedente de uma distribuição determinada.

Alternativa para o teste Chi^2 , quando o modelo proposto sob a hipótese nula é do tipo contínuo e o tamanho amostral é pequeno.

Vantagens do teste Kolmogorov-Smirnov diante do teste Chi²:

- Não requer agrupação dos dados em classes
- É aplicável a amostras pequenas.

Inconvenientes do teste Kolmogorov-Smirnov diante do teste qui-quadrado: Só é válido para modelos de tipo contínuo.

Desenvolvimento do teste:

Seja X_1, X_2, \dots, X_n uma m.a.s. de uma v.a. X com distribuição de tipo contínuo.

Compare-se:

H_0 : X segue a distribuição F .

H_1 : X não segue a distribuição F .

Exemplo:

Realizar um teste de *Kolmogorov-Smirnov*, em nível $\alpha=0.1$, para comparar se é possível supor que os 10 dados: **10.5, 8, 15, 12.1, 4.1, 12.1, 6, 10.5, 15** procedem de uma distribuição normal $N(10.84, 3.5)$

1. Ordenados os dados da amostra, construímos a tabela com os valores D_i

| $x_{(i)}$ | $\hat{F}_n(x_{(i)})$ | $F(x_{(i)})$ | D_i |
|-----------|----------------------|--------------|-------|
| 4.1 | 0.1 | 0.027 | 0.073 |
| 8 | 0.3 | 0.209 | 0.109 |
| 10.5 | 0.5 | 0.641 | 0.161 |
| 12.1 | 0.8 | 0.640 | 0.160 |
| 15 | 0.9 | 0.882 | 0.082 |
| 16 | 1 | 0.930 | 0.070 |

$$\max\{0.027, |0.1 - 0.027|\}$$

$$\max\left\{ \begin{array}{l} |0.1 - 0.209| \\ |0.3 - 0.209| \end{array} \right\}$$

2. $D_{exp} = \max\{D_i, i = 1, 2, \dots, n\} = 0.161$

3. Região crítica $C = [D_{1-\alpha}, +\infty[= [0.368, +\infty[$

4. Conclusão: $0.161 < 0.368$, portanto, não se rejeita que os dados procedam de uma distribuição $N(10.84; 3.5)$

Aplicação na cirurgia de coluna

Em trabalho recente, de 2009, publicado pela revista *Spine*, foi possível verificar o uso deste teste KS.

Foi estudada a variação entre a forma anatômica das facetas articulares e suas variações, durante o crescimento de crianças eutróficas dinamarquesas.

Masharawi e seus colaboradores, da Universidade de Tel Aviv, juntamente com a Universidade do Sul da Dinamarca, realizaram medições de todas as facetas lombares (L1-S1) de 100 crianças (51 meninos e 49 meninas), usando ressonância magnética, tanto na largura como no comprimento das mesmas.

Essas medidas foram tomadas aos 13 anos e 3 anos mais tarde, com um *software* específico para esses casos.

Foram tomadas mais de 4.400 medidas. Depreende-se disto a importância de um teste que possa verificar se estamos diante de uma amostra normal. Aqui ficaria clara a importância do teste KS.

Esse trabalho concluiu que as facetas daquelas crianças continuavam crescendo após os 12 anos. Qualquer assimetria deveria ser considerada fora da normalidade, e deixou clara a necessidade de fabricação de implantes mais personalizados, considerando as faixas etárias dessa população.

LEITURA COMPLEMENTAR

1. Daniel WW. Biostatistics. A foundation for analysis in the health sciences. 7th ed. New York: John Wiley and Sons Inc, 1999: 658- 736.
2. Kolmogorov AN. Sulla determinazione empirical di una legge di distribuzione. Giornale Inst Ital Altuari 1933; 4: 83-91. 18.
3. Masharawi YM, Kjaer P, Bendix T, Manniche C, May H, Mirovsky Y, Anekshtein Y, Jensen TS, Hershkovitz I. Lumbar facet and interfacet shape variation during growth in children from the general population: a three-year follow-up MRI study. *Spine* (Phila Pa 1976). 2009 Feb 15;34(4):408-12.
4. Smirnov NV. Estimate of deviation between empirical distribution functions in two independent samples. *Bull Moscow University* 1939; 2: 3-16.

5. Siegel S, Castellan NJ. Estadística no paramétrica aplicada a las ciencias de la conducta. 4. ed. México: Editorial Trillas, 1995: 151-7.
6. Song-Hee Kim and Ward Whitt, 2013. The Power of Alternative Kolmogorov-Smirnov Tests Based on Transformations of the Data. ACM Trans. Model. Comput. Simul. V, N, Article A (January YYYY), 18 pages. DOI:<http://dx.doi.org/10.1145/0000000.0000000>.

TESTE DE SHAPIRO-WILK

Nelson Astur Neto

Rodrigo Góes Medéa de Mendonça

Definição do teste

O teste de Shapiro-Wilk tem uma normalidade nas estatísticas de frequência. Foi publicado em 1965 por Samuel Shapiro Sanford e Martin Wilk.¹ A distribuição normal é uma das mais importantes distribuições de probabilidades da estatística, conhecida também como Distribuição de Gauss ou gaussiana.² O teste rejeita a hipótese de normalidade, quando o valor de p é menor ou igual a 0,05.

Para a verificação da Normalidade, por meio do teste de *Shapiro-Wilk (S-W)*, existem diversas ferramentas e programas estatísticos, dentre as quais o *Statistical Package for Social Sciences (SPSS)* versão 23.0 (SPSS, 2016). O SPSS é usado para a análise de dados e permite manipular, transformar, criar tabelas e gráficos que resumam as informações obtidas. As suas potencialidades vão além da análise descritiva de um conjunto de dados. É possível realizar com este *software* procedimentos mais avançados como inferência estatística, testes de hipóteses e estatísticas multivariadas para dados qualitativos e quantitativos.²

Fórmula

O teste Shapiro-Wilk, proposto em 1965, é baseado na estatística W dada por:

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

em que $\text{an-i}+1$ são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n de uma distribuição Normal.³

Exemplos na cirurgia da coluna

Como exemplo do teste de Shapiro-Wilks aplicado à coluna, temos o trabalho desenvolvido na universidade da Pensilvânia, em conjunto com os grupos de ortopedia e radiologia, além do departamento de estatística e publicado em 2014 pela Academia Americana de Fisiatria e Reabilitação.⁴ Billy e colaboradores tiveram como proposta estudar se poderiam existir diferenças entre as alturas e os diâmetros dos discos lombares, antes e depois de períodos prolongados, na posição sentada e sem intervalos, durante quatro horas consecutivas. O desenho do estudo foi prospectivo, em pacientes ambulatoriais, utilizando Ressonância Magnética para as medidas. No primeiro dia, os pacientes foram analisados no início do dia de trabalho e sentados, após 4 horas; já no segundo dia, os mesmos 12 pacientes foram estudados também no início de suas atividades, porém foram submetidos a mudanças e protocolo de alongamento a cada 15 minutos até o término das mesmas 4h.

Neste exemplo, o teste de Shapiro-Wilks foi decisivo para saber se a amostra se encontrava na normalidade, pois, sendo normal, o próximo passo seria realizar um Teste T pareado e se a normalidade não fosse satisfeita, usar-se-iam os testes de Wilcoxon e Rank.

Ainda sobre os resultados deste trabalho, foi encontrada uma diferença estatística entre as alturas dos discos naqueles que ficaram todo tempo sentados. A aplicabilidade clínica desta pesquisa reside no fato de que pequenas mudanças na altura discal podem estar correlacionadas com melhora da lombalgia e conseqüentemente das suas conseqüências.⁴

Aplicação em cirurgia

Quando escutamos o nome específico de algum teste estatístico, como Shapiro-Wilk, é difícil imaginarmos como poderíamos utilizá-lo ao nosso favor, ainda mais quando fazemos pesquisas com populações submetidas a alguma forma de intervenção, como cirurgia da coluna.

Como exemplo ilustrativo dessa aplicação, temos a pesquisa realizada por Farjoodi et al., no *Hospital Johns Hopkins*, nos Estados Unidos, onde o efeito do volume cirúrgico do hospital e do cirurgião

foi avaliado nas complicações após cirurgia na coluna lombar. O artigo foi publicado no periódico *Spine* em 2011.⁵

Nesse caso, a função do teste de Shapiro-Wilks foi verificar a normalidade das variáveis, e separar aquelas que precisaram ser ajustadas para não se tornarem fatores de confusão, como as variáveis demográficas. Foram estudados 232.668 procedimentos hospitalares listados como descompressão pósterio-lateral com artrodese e/ou exploração/descompressão do canal vertebral.

Após a aplicação dessa ferramenta estatística, os autores concluíram que a mortalidade e a taxa de complicações, associadas à cirurgia da coluna lombar, foram menores quando pacientes foram tratados por cirurgiões e em hospitais com alto volume.⁵

Interpretação

O teste de *S-W* fornece o parâmetro-valor de prova (*valor-p*, *p-value* ou *significância*), que pode ser interpretado como a medida do grau de concordância entre os dados e a hipótese nula (H_0), sendo H_0 correspondente à distribuição Normal. Quanto menor for o valor-p, menor é a consistência entre os dados e a hipótese nula. Logo, a regra de decisão adotada para saber se a distribuição é Normal ou não é rejeitar H_0 : (i) se, rejeita-se H_0 , ou seja, não se pode admitir que o conjunto de dados em questão tenha distribuição Normal; (ii) se, não se rejeita H_0 , ou seja, a distribuição Normal é uma distribuição possível para o conjunto de dados em questão.²

Falhando, o teste de normalidade permite afirmar com 95% de confiança que os dados não se encaixam na distribuição normal. Passar no teste de normalidade só lhe permite declarar que não foi encontrado desvio significativo da normalidade.⁴

LEITURA COMPLEMENTAR

1. Shapiro, S. S.; Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". *Biometrika* 52 (3-4): 591-611. p. 593.
2. Lopes M. M, Castelo Branco V. T. F. Soares J. B. Utilização dos testes estatísticos de Kolmogorov- Smirnov e Shapiro Wilk para verificação da normalidade de pavimentação. *TRANSPORTES* v. 21, n. 1 (2013) p. 59-66.

3. <http://www.portalaction.com.br/inferencia/64-teste-de-shapiro-wilk>
4. Billy GG, Lemieux SK, Chow MX. Changes in lumbar disk morphology associated with prolonged sitting assessed by magnetic resonance imaging. *PM R*. 2014 Sep;6(9):790-5.
5. Farjoodi P, Skolasky RL, Riley LH. The effects of hospital and surgeon volume on postoperative complications after Lumbar Spine surgery. *Spine (Phila Pa 1976)*. 2011 Nov 15;36(24):2069-75.
6. <http://www.variation.com/da/help/hs141.htm>

Unidade 2

CAPÍTULO 15

Teste t de Student

Barón Zarate

Francisco López Meléndez

CAPÍTULO 16

Teste de ANOVA

Juan Pablo Guyot

CAPÍTULO 17

Correlação de Pearson

Alisson R. Teles

Asdrubal Falavigna

CAPÍTULO 15

TESTE t DE STUDENT

*Barón Zarate
Francisco López Meléndez*

A primeira explicação para este assunto é que existem dois grandes grupos de testes estatísticos, baseados na distribuição da amostra ou da população, sendo eles os testes paramétricos e não paramétricos. A diferença entre testes PARAMÉTRICOS e NÃO PARAMÉTRICOS é que os testes paramétricos assumem os parâmetros da distribuição de uma variável (média e variância), e sua distribuição é Normal; e os testes não paramétricos NÃO assumem nada sobre a distribuição, nem se preocupam com o tipo de distribuição da amostra, somente trabalham com o ordenamento e a contagem dos valores da variável, sem se importarem com a distribuição (razão pela qual também são chamados de distribuição livre).

Características da amostra para saber que tipo de teste pode ser realizado.

Paramétricas:

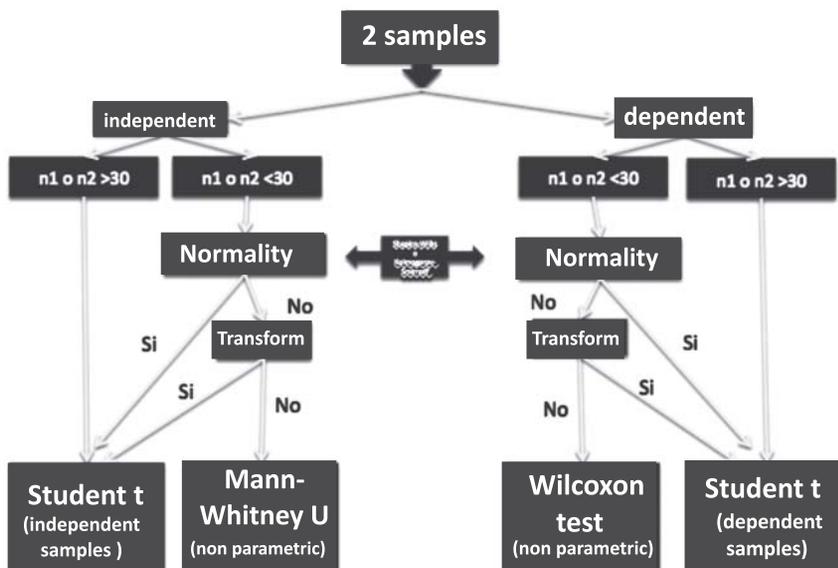
- as variáveis têm que ser quantitativas;
- usar escalas de intervalo ou razão;
- as variáveis devem seguir uma distribuição normal;
- as variâncias devem ser iguais;
- amostras grandes ($n > 30$).

Não paramétricas:

- variáveis de distribuição livre (não Normal);
- se forem variáveis quantitativas, ordinais ou nominais;
- grupos de variâncias grandes;
- amostras pequenas ($n < 30$ o < 10).

Em bioestatística, uma das questões mais frequentes é avaliar a significância estatística das diferenças entre amostras de indivíduos, sejam de um mesmo grupo (ex. *Antes e depois*), ou de diferentes grupos (ex. *Tratamento 1 Vs Tratamento 2*). São estas as amostras chamadas de emparelhadas ou independentes, respectivamente. (Figura 1).

Figura 1 – Diagrama de fluxo para decidir o tipo de teste estatístico para fazer comparação de hipóteses em duas amostras (dependentes ou independentes)



Student é o pseudônimo de William Sealy Gossett, estadista e matemático inglês que trabalhava para a cervejaria Guinness. Ele foi o criador do teste *t* ou *t* de Student, em 1908, para avaliar a distribuição de amostras pequenas baseando-se no teorema de pequenas amostras e no teorema do limite central.

O teste de Student ou *t* de Student é geralmente usado para comparar amostras contínuas, distribuídas de forma *normal* (teste paramétrico) entre dois grupos ou entre grupos. Por isso, existem dois tipos de *t* de Student, um para amostras independentes e outro para amostras dependentes ou emparelhadas.

Nesta ocasião, falaremos do *t* de Student para amostras emparelhadas (*amostras relacionadas*). Diz-se que há emparelhamento quando:

- as mensurações são feitas em um mesmo indivíduo (ex.: antes e depois de uma intervenção) e chama-se auto-emparelhamento;
- são comparadas entre si duas duplas naturais (ex.: gêmeos), e chama-se emparelhamento natural;
- procura-se um controle para cada caso, o que é chamado de Casos e Controles, nestes, procuram-se características iguais (sexo, idade, etc.) e chama-se emparelhamento artificial.

NOTA: A seleção de um teste estatístico para comparar hipóteses é diferente quando se trata de amostras independentes (ex.: dois tipos de indivíduos) e de amostras emparelhadas (ex.: mesmos indivíduos observados duas vezes), devido a que a variabilidade aleatória entre indivíduos (interindividual) não existe ou é menor nas amostras relacionadas (variabilidade intraindividual), o mesmo ocorre se a distribuição for normal e não normal.

A fórmula de *t* de Student é a seguinte:

$$T_{n-1} = \mu \text{ DIF} / \text{EEMDIF}$$

Ou

$$t = \frac{\bar{X}_D - \mu_0}{s_D / \sqrt{n}}$$

A seguir, desenvolve-se, com um exemplo, a aplicação prática de um teste *t* de Student para dados relacionados.

Interpretação na cirurgia de coluna

Um exemplo cotidiano no qual poderíamos usar um teste *t* de Student, para amostras relacionadas, seria avaliar a qualidade de vida (em uma escala contínua validada de 0 a 100, em que o valor mais alto corresponde a melhor qualidade de vida), de um número *n* de pacientes que se submeteram a um procedimento cirúrgico (por exemplo: uma artrodese L4-L5), onde se realiza uma medição pré-

operatória e outra medição pós-operatória, para depois avaliar se existe mudança pós-cirúrgica. (Objetivo: avaliar se existe uma mudança estatisticamente significativa na qualidade de vida, após o tratamento cirúrgico). (Tabela 1).

Tabela 1

| Paciente | Pré-operatório (SF-36) | Pós-Operatório (SF-36) |
|----------|------------------------|------------------------|
| 1 | 43.2 | 56.3 |
| 2 | 46.2 | 46.4 |
| 3 | 38.6 | 65.3 |
| 4 | 34 | 43 |
| 5 | 36.4 | 39.2 |
| 6 | 16.9 | 49.5 |
| 7 | 43 | 56 |
| 8 | 18 | 45.3 |
| 9 | 36.4 | 62.1 |
| 10 | 49.1 | 63.2 |
| 11 | 38.6 | 49 |
| 12 | 43.2 | 45 |
| 13 | 42 | 55 |
| 14 | 38.6 | 56 |

Nossa hipótese nula (H_0) seria a de que não existe diferença significativa entre a medição pré e pós-operatória.

$$H_0 \cong \mu \text{ antes} = \mu \text{ depois}$$

E nossa hipótese alternativa (H_1) seria a de que existe diferença estatisticamente significativa entre ambas as medições.

$$H_1 \cong \mu \text{ antes} \neq \mu \text{ depois}$$

Passos a seguir para realizar um teste *t* Student para amostras relacionadas:

1. A primeira coisa a fazer será criar uma nova variável (DIF), que é a diferença entre o valor de “antes” e “depois”, e esta nova variável mostrará quanto mudou um indivíduo entre as duas medições. Depois, calcula-se a média (μ) da variável *DIF*. (Tabela 2).

Tabela 2

| Paciente | Pré-operatório | Pós-operatório | DIF. |
|----------|----------------|----------------|--------------|
| 1 | 42 | 50 | 8,00 |
| 2 | 37.3 | 52.3 | 15,00 |
| 3 | 26.4 | 38 | 11,60 |
| 4 | 29.7 | 48 | 18,30 |
| 5 | 24.9 | 49.1 | 24,20 |
| 6 | 44.6 | 55 | 10,40 |
| 7 | 33.5 | 49 | 15,50 |
| 8 | 44.6 | 67 | 22,40 |
| 9 | 26.5 | 39 | 12,50 |
| 10 | 18 | 56 | 38,00 |
| 11 | 33.4 | 47 | 13,60 |
| 12 | 42 | 65 | 23,00 |
| 13 | 37.5 | 57 | 19,50 |
| 14 | 24.9 | 67 | 42,10 |

Média:19.58

2. Se a amostra for menor de 30, precisamos comprovar se a nova variável segue uma *distribuição normal*. Se não tiver distribuição normal, pode-se transformar de maneira logarítmica, ou fazer um teste não paramétrico (*teste de Wilcoxon*). No programa SPSS são realizados os dois testes de normalidade Kolmogorov-Smirnov e Shapiro-Wilks.

NOTA: O teste de Shapiro-Wilks é considerado o teste mais potente para amostras inferiores a 30 casos.

3. O passo seguinte, após calcular a média da variável (*DIF*), é calcular o erro padrão da mesma ($EEMDIF =$ erro padrão da média da variável “diferença”).

A fórmula para calcular o $EEMDIF$ é a divisão do desvio padrão da população (s) pela raiz quadrada do tamanho da amostra (h).

$$EEMDIF = \sigma / \sqrt{h}$$

Exemplo:

$$\begin{aligned} EEMDIF &= \sigma / \sqrt{h} \\ &= 9.98 / \sqrt{14} \\ &= 9.98 / 3.74 \\ EEMDIF &= 2.66 \end{aligned}$$

4. Depois, calcula-se os graus de liberdade ($GL = \eta - 1$) onde h é o número total de indivíduos da amostra, e 1 é constante para teste t de Student de amostras relacionadas.

$$\begin{aligned} GL &= \eta - 1 \\ GL &= 14 - 1 = 13 \end{aligned}$$

5. Calcular o t de Student para amostras relacionadas; este é o resultado obtido após dividir a média (μ) da variável DIF pelo seu erro padrão ($EEMDIF$).

$$T_{n-1} = \mu DIF / EEMDIF$$

EXEMPLO:

$$\begin{aligned} t_{13} &= \mu DIF / EEMDIF = m DIF / (\sigma / \sqrt{\eta}) \\ t_{13} &= 19.58 / 2.66 \\ \mathbf{t}_{13} &= \mathbf{7.36} \end{aligned}$$

6. O passo seguinte é comparar, e procurar nas tabelas, o valor obtido de t e seus graus de liberdade, para obter o valor de p ; também pode-se usar EXCEL em vez de consultar as tabelas, com a fórmula seguinte:

$$= \text{DISTR.T.2 (valor de } t; GL; 2)$$

O primeiro número será o valor obtido de *t*, o segundo são os graus de liberdade, e o terceiro pode ser 2, se for um teste de duas caudas, ou 1, se for um teste de uma cauda.

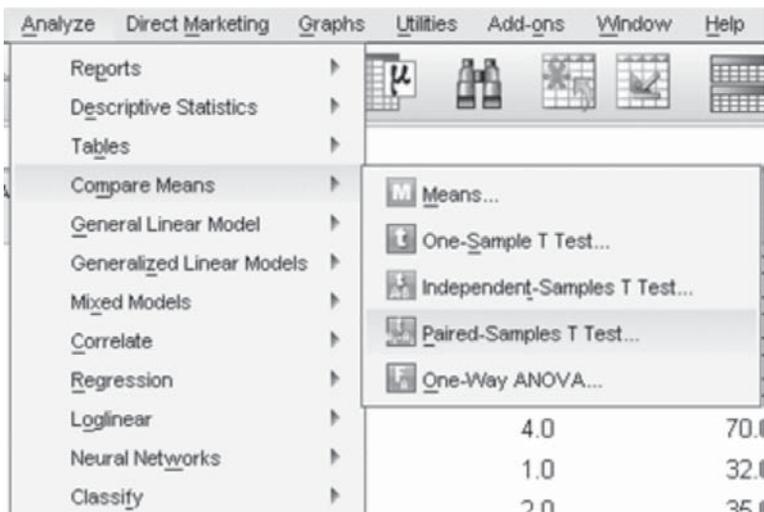
7. O resultado de *p* obtido deve ser sempre acompanhado, juntamente com o intervalo de confiança para a diferença de médias.

Resultado do exemplo: neste exemplo, obtemos um *t* de 7.36, com 13 GL, o que equivale a um $p < 0.0000$, com um IC de 95% de duas caudas. Isto significa que existe evidência suficiente para rejeitar a hipótese nula; ou seja, existe um aumento significativo na escala de qualidade de vida dos pacientes pós-operados, o aumento médio foi de 19.58 pontos na escala, com uma confiança de 95% de que o aumento estará entre 13.8 e 25.3 pontos.

Exemplo e forma de fazê-lo com SPSS:

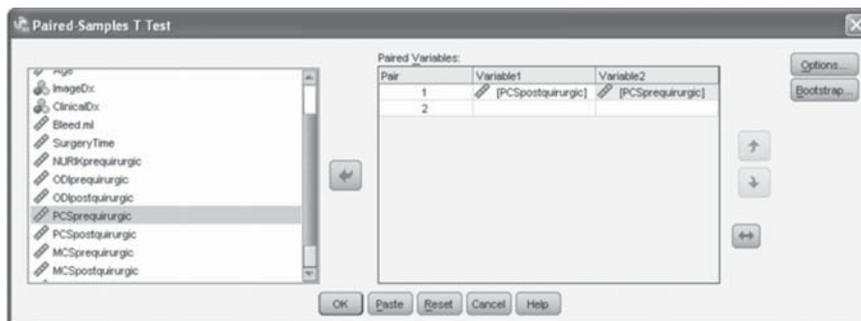
1. A primeira coisa que temos que fazer no SPSS é ordenar as variáveis que queremos comparar; depois, ir até o menu em “analisar” ⇒ comparar médias ⇒ teste de T para amostras relacionadas...” (Figura 2).

Figura 2 – Teste T Amostras relacionadas



2. Aparecerá um quadro onde aparece introduzir as duas variáveis a comparar. Nesse caso, vamos introduzir primeiro a variável “depois”, e após a variável “antes” ⇒ aceitar (Figura 3).

Figura 3 – Teste T Amostras relacionadas



3. Resultados: O computador mostrará três tabelas (Figura 4). A primeira tabela apresenta os dados estatísticos descritivos das variáveis a comparar. A segunda tabela mostra um coeficiente de correlação (*Pearson*) entre as variáveis. A *terceira tabela*, e a mais importante para interpretar, mostra, na primeira coluna, a média da *diferença* (a mudança antes e depois); na segunda coluna, o Desvio padrão da média da diferença e, depois, o erro padrão da média (EEMDIF); a coluna seguinte mostra o intervalo de confiança (IC) com o limite inferior e o limite superior, a seguir, o valor de *t*, depois, os graus de liberdade e, finalmente, a significância estatística.

Figura 4 – Tabelas de resultados de SPSS (V. 19)

| | Mean | N | Std. Deviation | Std. Error Mean |
|------------------|--------|----|----------------|-----------------|
| PCSPreoperative | 52.814 | 14 | 9.14506 | 2.44412 |
| Pair 1 | 3 | | | |
| PCSPostoperative | 33.235 | 14 | 8.42347 | 2.25127 |
| | 7 | | | |

| | N | Correlation | Sig. |
|------------------------|----|-------------|------|
| Pair 1 PCSPreoperative | 14 | .357 | .216 |
| PCSPostoperative | | | |

| | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|--|--------------------|----------------|-----------------|---|---------|-------|----|-----------------|
| | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | Lower | Upper | | | |
| Pair 1 PCSpstquirurgic-PCSprequirurgic | 19.5786 | 9.9794 | 2.6671 | 13.8167 | 25.3405 | 7.341 | 13 | .000 |

NOTA: Foram utilizados os mesmos dados para ambos os exemplos: a forma manual e com SPSS. A decisão de rejeitar a hipótese nula é quando $p = < 0.05$.

Tabela 1 – Tabela do *t* de Student

Tabla de la *t* de Student.

Contiene los valores *t* tales que $P(|T| > t) = \alpha$,
donde *n* son los grados de libertad.



| <i>n</i> \ α | 0,90 | 0,80 | 0,70 | 0,50 | 0,30 | 0,20 | 0,10 | 0,05 | 0,02 | 0,01 | 0,001 |
|---------------------|--------|--------|--------|--------|--------|--------|--------|---------|---------|---------|----------|
| 1 | 0,1584 | 0,3249 | 0,5095 | 1,0000 | 1,9626 | 3,0777 | 6,3137 | 12,7062 | 31,8210 | 63,6559 | 636,5776 |
| 2 | 0,1421 | 0,2887 | 0,4447 | 0,8165 | 1,3862 | 1,8856 | 2,9200 | 4,3027 | 6,9645 | 9,9250 | 31,5998 |
| 3 | 0,1366 | 0,2767 | 0,4242 | 0,7649 | 1,2498 | 1,6377 | 2,3534 | 3,1824 | 4,5407 | 5,8408 | 12,9244 |
| 4 | 0,1338 | 0,2707 | 0,4142 | 0,7407 | 1,1896 | 1,5332 | 2,1318 | 2,7765 | 3,7469 | 4,6041 | 8,6101 |
| 5 | 0,1322 | 0,2672 | 0,4082 | 0,7267 | 1,1558 | 1,4759 | 2,0150 | 2,5706 | 3,3649 | 4,0321 | 6,8685 |
| 6 | 0,1311 | 0,2648 | 0,4043 | 0,7176 | 1,1342 | 1,4398 | 1,9432 | 2,4469 | 3,1427 | 3,7074 | 5,9587 |
| 7 | 0,1303 | 0,2632 | 0,4015 | 0,7111 | 1,1192 | 1,4148 | 1,8946 | 2,3946 | 2,9979 | 3,4995 | 5,4081 |
| 8 | 0,1297 | 0,2619 | 0,3995 | 0,7064 | 1,1081 | 1,3968 | 1,8595 | 2,3060 | 2,8965 | 3,3554 | 5,0414 |
| 9 | 0,1293 | 0,2610 | 0,3979 | 0,7027 | 1,0997 | 1,3830 | 1,8331 | 2,2622 | 2,8214 | 3,2498 | 4,7809 |
| 10 | 0,1289 | 0,2602 | 0,3966 | 0,6998 | 1,0931 | 1,3722 | 1,8125 | 2,2281 | 2,7638 | 3,1693 | 4,5868 |
| 11 | 0,1286 | 0,2596 | 0,3956 | 0,6974 | 1,0877 | 1,3634 | 1,7959 | 2,2010 | 2,7181 | 3,1058 | 4,4369 |
| 12 | 0,1283 | 0,2590 | 0,3947 | 0,6955 | 1,0832 | 1,3562 | 1,7823 | 2,1788 | 2,6810 | 3,0545 | 4,3178 |
| 13 | 0,1281 | 0,2586 | 0,3940 | 0,6938 | 1,0795 | 1,3502 | 1,7709 | 2,1604 | 2,6503 | 3,0123 | 4,2209 |
| 14 | 0,1280 | 0,2582 | 0,3933 | 0,6924 | 1,0763 | 1,3450 | 1,7613 | 2,1448 | 2,6245 | 2,9768 | 4,1483 |
| 15 | 0,1278 | 0,2579 | 0,3928 | 0,6912 | 1,0735 | 1,3406 | 1,7531 | 2,1315 | 2,6025 | 2,9467 | 4,0728 |
| 16 | 0,1277 | 0,2576 | 0,3923 | 0,6901 | 1,0711 | 1,3368 | 1,7459 | 2,1199 | 2,5835 | 2,9208 | 4,0149 |
| 17 | 0,1276 | 0,2573 | 0,3919 | 0,6892 | 1,0690 | 1,3334 | 1,7396 | 2,1098 | 2,5669 | 2,8982 | 3,9651 |
| 18 | 0,1274 | 0,2571 | 0,3915 | 0,6884 | 1,0672 | 1,3304 | 1,7341 | 2,1009 | 2,5524 | 2,8784 | 3,9217 |
| 19 | 0,1274 | 0,2569 | 0,3912 | 0,6876 | 1,0655 | 1,3277 | 1,7291 | 2,0930 | 2,5395 | 2,8609 | 3,8833 |
| 20 | 0,1273 | 0,2567 | 0,3909 | 0,6870 | 1,0640 | 1,3253 | 1,7247 | 2,0860 | 2,5280 | 2,8453 | 3,8496 |
| 21 | 0,1272 | 0,2566 | 0,3906 | 0,6864 | 1,0627 | 1,3232 | 1,7207 | 2,0796 | 2,5176 | 2,8314 | 3,8193 |
| 22 | 0,1271 | 0,2564 | 0,3904 | 0,6858 | 1,0614 | 1,3212 | 1,7171 | 2,0739 | 2,5083 | 2,8188 | 3,7922 |
| 23 | 0,1271 | 0,2563 | 0,3902 | 0,6853 | 1,0603 | 1,3195 | 1,7139 | 2,0687 | 2,4999 | 2,8073 | 3,7676 |
| 24 | 0,1270 | 0,2562 | 0,3900 | 0,6848 | 1,0593 | 1,3178 | 1,7109 | 2,0639 | 2,4922 | 2,7970 | 3,7454 |
| 25 | 0,1269 | 0,2561 | 0,3898 | 0,6844 | 1,0584 | 1,3163 | 1,7081 | 2,0595 | 2,4851 | 2,7874 | 3,7251 |
| 26 | 0,1269 | 0,2560 | 0,3896 | 0,6840 | 1,0575 | 1,3150 | 1,7056 | 2,0555 | 2,4786 | 2,7787 | 3,7067 |
| 27 | 0,1268 | 0,2559 | 0,3894 | 0,6837 | 1,0567 | 1,3137 | 1,7033 | 2,0518 | 2,4727 | 2,7707 | 3,6895 |
| 28 | 0,1268 | 0,2558 | 0,3893 | 0,6834 | 1,0560 | 1,3125 | 1,7011 | 2,0484 | 2,4671 | 2,7633 | 3,6739 |
| 29 | 0,1268 | 0,2557 | 0,3892 | 0,6830 | 1,0553 | 1,3114 | 1,6991 | 2,0452 | 2,4620 | 2,7564 | 3,6595 |
| 30 | 0,1267 | 0,2556 | 0,3890 | 0,6828 | 1,0547 | 1,3104 | 1,6973 | 2,0423 | 2,4573 | 2,7500 | 3,6460 |
| 40 | 0,1265 | 0,2550 | 0,3881 | 0,6807 | 1,0500 | 1,3031 | 1,6839 | 2,0211 | 2,4233 | 2,7045 | 3,5510 |
| 80 | 0,1261 | 0,2542 | 0,3867 | 0,6776 | 1,0432 | 1,2922 | 1,6641 | 1,9901 | 2,3739 | 2,6387 | 3,4164 |
| 120 | 0,1259 | 0,2539 | 0,3862 | 0,6765 | 1,0409 | 1,2886 | 1,6576 | 1,9799 | 2,3578 | 2,6174 | 3,3734 |
| ∞ | 0,126 | 0,253 | 0,385 | 0,674 | 1,036 | 1,282 | 1,645 | 1,96 | 2,326 | 2,576 | 3,291 |

LEITURA COMPLEMENTAR

1. Altman DG, Bland JM. Statistics Notes: Transforming data. *BMJ* 1996; 312: 770.
2. Altman DG. Practical statistics for medical research. Londres: Chapman and Hall, 1991.
3. Altman DG, Bland JM. Detecting skewness from summary information, *BMJ* 1996; 313:1200.
4. Altman DG, Bland JM. Statistics Notes: The use of transformations when comparing two means. *BMJ* 1996; 312:1153.
5. Altman DG. Preparing to analyse data. En: Practical statistics for medical research. London: Chapman and Hall; 1991. p.132-145.
6. Berry G., Armitage P. Statistical Methods in Medical Research. 3rd. ed. Oxford: Blackwell Science; 1994.
7. Braitman LE. Confidence intervals asses both clinical significance and statistical significance. *Ann Intern Med* 1991; 114 (6): 515-517.
8. Gehlbach SH. Interpreting the medical literature: practical epidemiology for clinicians. 5th ed. USA: McGraw-Hill. Medical; 2006; p. 166.
9. Livingston EH. Who was student and why do we care so much about his t-test. *J Surg Research* 2004;118:58.
10. Moreno V, Vallescar R, Martín M. Las pruebas no paramétricas en el análisis estadístico de datos. *Aten Primaria* 1991; 8 (1): 58-60.
11. Moser CJ, Stevens MB, Homogeneity of variance in the two-sample means test, *The American Statistician* 46 (1992) 19-21.
12. Newcombe, RG. Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: Asymptotic methods and evaluation. *Statistics in Medicine*, (2006) 25, 559-573.
13. Rajiv Gandhi, Holly N. Smith, Nizar N. Mahomed, Randy Rizek, MD, Mohit Bhandari. Incorrect Use of the Student t Test in Randomized Trials of Bilateral Hip and Knee Arthroplasty Patients. *The Journal of Arthroplasty* Vol. 26 No. 5 2011.
14. Riegelman, RK. Studying a Study and Testing a Test: How to Read the Medical Evidence. Philadelphia: Lippincott Williams & Wilkins, 1999.

CAPÍTULO 16

TESTE DE ANOVA

Juan Pablo Guyot

O Teste de ANOVA é um teste estatístico usado para analisar duas variâncias, daí o termo proveniente do inglês (*ANalysis Of VAriance*). Seu desenho inicial tem origem no ano 1925, quando foi aplicado em trabalhos agropecuários. No entanto, sua difusão gerou seu emprego em muitos campos. O objetivo inicial consistia em observar a influência de diferentes fatores, no crescimento de uma semente determinada, que iria crescer da mesma maneira, se não fossem usados diferentes métodos de cultura. Em outras palavras, procura avaliar a importância de um ou mais fatores, observando o seu impacto sobre uma ou mais populações determinadas, com variâncias constantes.

No desenho experimental de um teste ANOVA, procura-se identificar “fatores fixos” e “fatores randomizados”. Fala-se dos primeiros quando se tenta estudar um grupo que não exposto apresentaria resultados inalterados, e o interesse é o de analisar a resposta diante de diferentes fatores específicos. Por outro lado, define-se o teste com fatores randomizados, escolhendo ao acaso fatores com diferentes níveis de exposição, com um número de níveis múltiplos tendendo ao infinito. Neste caso, repetindo as provas, os fatores devidos ao acaso devem surgir da escolha sobre diferentes níveis de escolha, e não seguindo sempre um mesmo padrão.

Se quiséssemos representá-lo com um exemplo gráfico, estaríamos definindo fatores fixos na comparação de médias entre diferentes populações, enquanto os randomizados seriam análises de diferenças devidas ao acaso, entre a mesma população. Um exemplo prático de fatores randomizados na cirurgia de coluna seria o estudo da evolução do balanço sagital da população latino-americana. O pesquisador escolherá fatores ao acaso, mas sempre baseado numa distribuição simétrica: amplitude de idades, diferenças por raças, etc.

Existem testes de ANOVA de um fator e de dois fatores em relação aos fatores a analisar. Determina-se que é ANOVA de um fator a análise que envolve o estudo sobre uma única variável, e ANOVA de duplo fator a análise que envolve variáveis provenientes de duas variâncias diferentes.

ANOVA de um fator

Procura-se determinar a influência apresentada pelo resultado de uma variável Y relacionada diretamente aos diferentes fatores X (X1, X2, X3.....). É usada quando tentamos analisar a variância de diversas variáveis independentes, ou vários níveis de uma só variável independente, mantendo um mesmo nível de confiança.

Ao desenhar um modelo experimental que use o teste de ANOVA, deve-se certificar-se de que cumpram quatro variáveis:

- os resultados das variáveis estudadas devem ser absolutamente independentes. Ou seja, o valor de uma observação não deve influenciar o resultado de outra variável;
- os dados são registrados seguindo um modelo aditivo e incorporando todas as informações, tanto dos efeitos fixos quanto dos erros devidos ao acaso;
- assume-se que a distribuição dos erros segue um padrão ordenado;
- os erros ao acaso seguem uma variância homogênea.

Um exemplo deste tipo de teste seria o estudo de diferentes tipos de tratamentos para pacientes com estenose degenerativa do canal lombar. Neste caso, a população a estudar seria o fator fixo. Espera-se que os pacientes com estenose do conduto lombar evoluam de uma maneira similar, e estudar-se-á o impacto de diferentes procedimentos (fatores) sobre seu seguimento. Será estudada a resposta conforme os tratamentos aplicados (descompressão simples, descompressão + fusão por via posterior e descompressão indireta por dispositivos intersomáticos colocados por via lateral), usando o estudo de ANOVA, ao comparar diferentes aspectos dos mesmos (tempo cirúrgico, sangramento, correção da deformidade, Oswestry pré e pós-operatório).

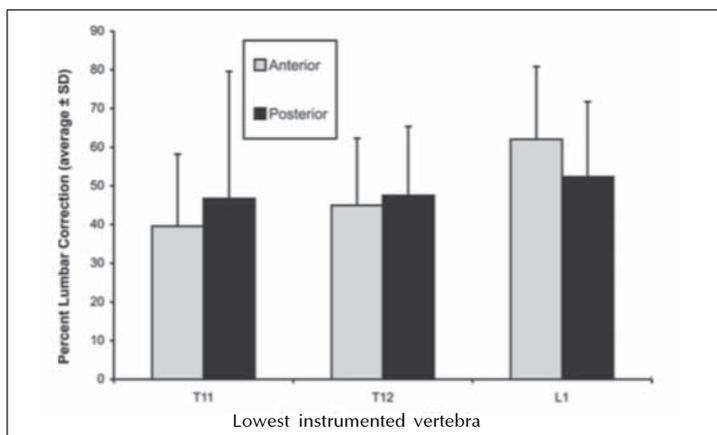
ANOVA de duplo fator

O desenho deste tipo de experimento envolve o estudo de duas ou mais variáveis fixas (Y e Z) em relação a diferentes graus de exposição aos fatores X (X1, X2, X3.....). A inclusão de outro fator não é inocente, já que este pode gerar alguns problemas associados, devido a que ambos os fatores podem ser fixos, aleatórios ou um fixo e outro aleatório. Ao mesmo tempo, pode ocorrer que tais fatores atuem de maneira independente, ou que os mesmos interajam entre si potencializando ou inibindo sua potência final. Finalmente, o desenho pode determinar dois grupos de fatores que sejam equilibrados ou não.

Interpretação

Um exemplo deste tipo de desenho experimental, vinculado à cirurgia de coluna, seria a avaliação do grau de correção espontânea das curvaturas lombares em instrumentações seletivas de escoliose torácica tipo Lenke 1, comparando diferentes vias de abordagem.

Gráfico 1



Neste caso, a variável a estudar seria a correção espontânea da coluna lombar. No entanto, diferentemente do exemplo apresentado na ocorrência de um fator, daí tem-se duas variáveis fixas independentes entre si (via anterior *versus* via posterior). Finalmente, estuda-se qual é a influencia da vértebra caudal instrumentada (T11, T12 o L1) sobre cada um dos grupos.

LEITURA COMPLEMENTAR

1. Tejada Barreras M. Estadística Básica. En: Falavigna A, Jimenez Avila JM. Educaciòn en investigaciòn AOSPine Latin America : de la idea a la publicaciòn. Caxias do Sul, RS: Educ. 2014: 149-165.
2. Larson MG. Analysis of variance. *Circulation*. 2008 Jan 1;117(1):115-21.
3. Dangelmajer S, Zadnik PL, Rodriguez ST, Gokaslan ZL, Sciubba DM. Minimally invasive spine surgery for adult degenerative lumbar scoliosis. *Neurosurg Focus*.2014 May;36(5):E7.
4. Sit V. Analyzing ANOVA Designs: Biometrics Information Handbook No. 5. Province of British Columbia, Ministry of Forests Research Program. Working paper 07/1995. Available at: <http://www.for.gov.bc.ca/hfd/pubs/docs/Wp/Wp07.pdf>. Accessed September 1, 2016.
5. Patel PN, Upasani VV, Bastrom TP, Marks MC, Pawelek JB, Betz RR, Lenke LG, Newton PO. Spontaneous lumbar curve correction in selective thoracic fusions of idiopathic scoliosis: a comparison of anterior and posterior approaches. *Spine(Phila Pa 1976)*. 2008 May 1;33(10):1068-73.

CORRELAÇÃO DE PEARSON

*Alisson R. Teles
Asdrubal Falavigna*

Descrição do teste

Correlação é uma medida do grau de dependência linear entre duas variáveis ou entre uma medida da intensidade de associação dessas variáveis. Em outras palavras, o teste de correlação avalia a existência e a força de associação entre duas variáveis quantitativas. Por exemplo, a probabilidade de existir uma correlação entre o grau de desequilíbrio sagital (medido em centímetros) e a incapacidade funcional dos pacientes com escoliose degenerativa (medida pela escala de incapacidade Oswestry). Outro exemplo seria a correlação entre a incidência pélvica e a lordose lombar. Diz-se que duas variáveis são correlacionadas (variam juntas), quando se consegue demonstrar que existe associação linear entre estas variáveis.

Existem basicamente duas formas de verificar correlação. A mais simples é através de um gráfico denominado diagrama de dispersão, também conhecido como diagrama de pontos (Figura 1). Por convenção, denomina-se x a variável independente ou preditora, que é descrita no eixo horizontal, e y a variável dependente ou desfecho, descrita no eixo vertical. A outra maneira de avaliar correlação é por um coeficiente. O coeficiente é superior ao diagrama de dispersão, pois representa um valor absoluto, o que facilita a interpretação da correlação entre duas variáveis de medidas diferentes (ex.: desequilíbrio sagital, medido em centímetros, e índice de incapacidade Oswestry, medido na escala entre 0 e 100 pontos).

O coeficiente de correlação de Pearson é a medida mais conhecida de correlação entre duas variáveis quantitativas. Ele deve ser aplicado para avaliar correlação de variáveis contínuas (não categóricas, no mínimo com mensuração intervalar) que apresentem distribuição normal (distribuição gaussiana). Ou seja, o coeficiente de correlação

de Pearson não deve ser utilizado para avaliar correlação de variáveis assimétricas, que não tenham distribuição normal. Nesse caso, deve-se aplicar o teste de correlação de Spearman.

É importante lembrar que o coeficiente de correlação mede uma associação, não uma relação de causa e efeito. Por exemplo, em um estudo envolvendo pacientes com doenças degenerativo-lombares, identificou-se que existe correlação entre perda sanguínea intraoperatória e qualidade de vida medidas pelo componente físico do SF36, um ano após a cirurgia. Neste exemplo, a correlação estatisticamente significativa não indica que a perda sanguínea seja a causa de pior qualidade de vida, um ano após o procedimento. Outro conceito é que em amostras grandes, mesmo coeficientes de correlação baixos podem ser estatisticamente significativos. Isso corrobora a importância da avaliação do grau de correlação (ex.: $n = 1300$, $r = 0,15$, $p = 0,001$, correlação fraca).

Fórmula

A fórmula de cálculo do coeficiente de correlação (r) foi proposta por Karl Pearson em 1896. O coeficiente de correlação é um valor absoluto, independente das unidades usadas para medir as variáveis x e y .

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}$$

Atualmente, diversos *softwares* de análise de dados geram o coeficiente de correlação, nível de significância da correlação, grau da correlação e coeficiente de determinação (Figura 2).

Interpretação

A interpretação do coeficiente de correlação (r) envolve basicamente quatro passos (Figura 1). Primeiramente, verifica-se se existe de fato a correlação através do teste de hipótese. O raciocínio do teste de hipótese é basicamente o mesmo das comparações de médias. Quando se calcula o r em uma amostra, extraída de uma população particular, é necessário ter em mente que se está estimando a associação verdadeira entre as variáveis x e y na população; portanto, é necessário verificar se a correlação identificada existe de fato ou se

está relacionada a um erro de amostragem. Através do teste de hipótese, obtém-se o valor p , que indica a probabilidade de a correlação realmente existir. (Em termos estatísticos, é aceitar a hipótese alternativa.). Geralmente, aceita-se a existência de correlação verdadeira quando $p < 0.05$; isto é: há menos que 5% de chance de a correlação identificada não existir de fato (aceitar hipótese nula, rejeitar a hipótese alternativa), portanto aceita-se que ela é de fato verdadeira.

O segundo passo é determinar a força da correlação entre as variáveis. Usualmente, correlação com coeficiente entre 0 – 0,3 é considerada fraca, entre 0,3 – 0,6 regular, 0,6 – 0,9 forte, e 0,9 – 1 muito forte, e 1 correlação perfeita.

O terceiro passo é analisar o sinal da correlação. O coeficiente de correlação (r) pode apresentar uma variação entre -1 e +1. Valores negativos expressam uma correlação indireta, isto é, quando x aumenta, y em média diminui. Valores positivos indicam correlação direta, quando x e y variam no mesmo sentido. Por exemplo, em indivíduos normais, os valores de incidência pélvica apresentam correlação direta com a lordose lombar. Em contrapartida, quanto maior é a idade menor é lordose lombar (correlação indireta).

Outra maneira de avaliar o valor da medida de associação linear entre duas variáveis é pelo coeficiente de determinação (r^2). Elevando-se o coeficiente de correlação de Pearson ao quadrado, obtém-se o coeficiente de determinação, que representa a fração da variabilidade que é compartilhada entre as duas variáveis. Em outras palavras, representa a percentagem de variação explicada por uma das variáveis em relação à outra. O valor r^2 pode variar entre 0 e 1. Por exemplo, se o coeficiente de correlação de Pearson (r) é 0,734, o coeficiente de determinação (r^2) é $0,834 \times 0,834 = 0,69$. Ou seja, 69% da variação de y é explicada pela variável x . O coeficiente de determinação não pode ser obtido através do coeficiente de correlação de Spearman (testes não paramétricos), somente do coeficiente de correlação de Pearson.

Exemplo na área de cirurgia de coluna

A correlação de Pearson é um dos testes mais utilizados em pesquisa médica. Para sua adequada aplicação, entretanto, é necessário que sejam seguidos os requisitos básicos: as variáveis devem ser contínuas (ou no mínimo com distribuição intervalar), e elas devem apresentar distribuição normal (gaussiana).

Um dos exemplos em cirurgia de coluna é a correlação entre o eixo sagital vertical cervical, medido através da distância entre a linha de prumo do ponto médio de C2 e C7. Diversos estudos têm demonstrado que existe correlação negativa entre o eixo sagital cervical e a qualidade de vida no pós-operatório de pacientes, que são submetidos a três ou mais níveis: laminectomia cervical e fusão posterior. Tang et al. demonstraram que o coeficiente de correlação entre eixo sagital vertical C2-C7 e o Índice de Incapacidade Cervical (NDI – *Neck Disability Index*) é 0.2957 ($p = 0.024$). O coeficiente de correlação entre o eixo e o componente físico do SF36 (PC-SF36) - 0.4262 ($p = 0.001$). Neste exemplo, verifica-se que (1) existe correlação entre a medida radiográfica e os instrumentos de avaliação NDI e PC-SF36 (ambos os testes de hipótese demonstram $p < 0.05$); (2) a correlação entre eixo C2-C7 com NDI é fraca ($r = 0.2957$) e com PC-SF36 é regular ($r = -0.4262$); (3) a correlação é positiva entre o eixo cervical e o NDI (quanto maior é a distância entre o eixo sagital vertical cervical maior é a incapacidade medida pelo NDI), e negativa entre o eixo e o PC-SF36 (quanto maior o eixo, menor é qualidade de vida aferida pelo PC-SF36).

Figura 1 – Exemplos de diagramas de dispersão e os respectivos coeficientes de correlação (r)

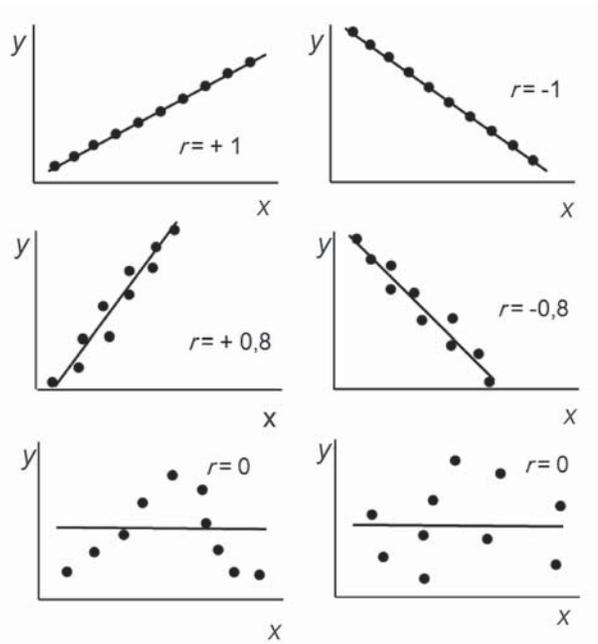
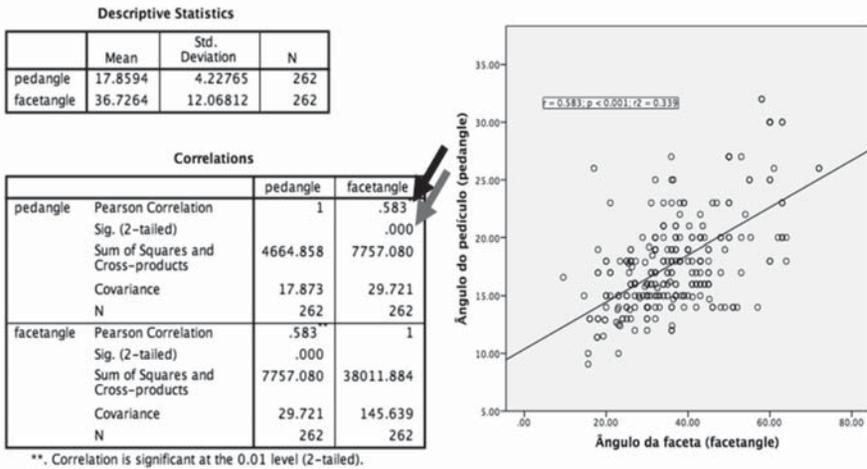


Figura 2 – Exemplo de teste de correlação entre duas variáveis, usando o *software* SPSS



Nota: No exemplo acima, demonstra-se o teste de correlação entre o ângulo do pedículo (pedangle) e o ângulo da articulação interapofisária (facetangle) de vértebras lombares aferidos em tomografia computadorizada. Observa-se que quanto maior é o ângulo pedicular em relação a linha média, maior é o ângulo da articulação interapofisária ($r = 0.583$, $p < 0.0001$). O *software* também gera o diagrama de dispersão entre as duas variáveis.

Tabela 1 – Interpretação do teste de correlação de Pearson (r)

-
1. Existe correlação entre as variáveis?
- Quando $P < 0,05$ considera-se que a correlação é estatisticamente significativa, ou seja, não ocorreu ao acaso. Se $P > 0,05$, não existe correlação.

2. Qual é o grau da correlação?
- Avaliar o valor do r (avaliação qualitativa de r quanto à intensidade):

| | |
|------------|------------------------------|
| 0 | Correlação nula |
| 0 – 0,3 | Correlação fraca |
| 0,31 – 0,6 | Correlação regular |
| 0,61 – 0,9 | Correlação forte |
| 0,91 – 1 | Correlação muito forte |
| 1 | Correlação plena ou perfeita |

3. Qual o sinal da correlação?
- Valores positivos indicam correlação direta, valores de x aumentam à medida que os valores de y aumentam, ou vice-versa.
 - Valores negativos indicam correlação indireta, quando os valores de x diminuem à medida que os valores de y aumentam, ou vice-versa.
4. Qual é o coeficiente de determinação (r^2)?
- O r^2 indica a percentagem de variação de x que é explicada pela variável y .
-

LEITURA COMPLEMENTAR

1. Peat J, Barton B. Medical Statistics: a guide to data analysis and critical appraisal. BMJ Books, 2008. ISBN: 978-0-7279-1812-3.
2. Peacock JL, Peacock PL. Oxford Handbook of Medical Statistics. Oxford University Press, 2011. ISBN: 978-0-19-9555128-6.
3. Tang JA, Scheer JK, Smith JS, Deviren V, Bess S, Hart RA, Lafage V, Shaffrey CI, Schwab F, Ames CP; ISSG. The impact of standing regional cervical sagittal alignment on outcomes in posterior cervical fusion surgery. *Neurosurgery*, 2015, 76 (Suppl 1): S14-21. DOI 10.1227/01.neu.0000462074.66077.2b.
4. Teles AR. Basic statistical shortcomings can induce mistaken interpretations of neurosurgical articles. *World Neurosurgery*, 2016, 92 (577-578) DOI: 10.1016/j.wneu.2015.12.090.

Unidade 3

CAPÍTULO 18

Chi²

Alberto Aceves Pérez

CAPÍTULO 19

Teste de postos de Wilcoxon

Baron Zarate

CAPÍTULO 20

Teste u de Mann-Whitney

José Manuel Pérez Atanasio

Victor Fernando Luján Celis

Maritza Belén Sandoval Rincón

CAPÍTULO 21

Teste de Kruskal-Wallis

Esteban Araya Ramírez

Fabián Víquez Monge

CAPÍTULO 22

Coefficiente de correlação de Spearman

Fernando Alvarado

Miguel Farfán

CAPÍTULO 23

Teste de McNemar

Esteban Araya Ramírez

Fabián Víquez Monge

CAPÍTULO 24

Regressão linear simples

Alisson R. Teles

Asdrubal Falavigna

CAPÍTULO 25

Distribuição de Poisson

José Acuña Pinilla

Margarita Gómez Chantraine

Elizabeth Heins

CAPÍTULO 26

Odds Ratio

José Manuel Pérez Atanasio

Victor Fernando Luján Celis

Maritza Belén Sandoval Rincón

CAPÍTULO 18

CHI²

Alberto Aceves Pérez

Existem muitas situações no âmbito da saúde, nas quais as variáveis de interesse não podem ser quantificadas em quantidades numéricas. Entre elas, algumas nas quais o pesquisador tem interesse em determinar possíveis relações. Exemplos deste tipo de variáveis seriam as complicações após uma intervenção cirúrgica, o gênero, o nível sociocultural, etc. Neste caso, teríamos, no máximo, as observações agrupadas em forma de frequência, dependendo das modalidades apresentadas pelos pacientes, em cada uma das variáveis.

O objetivo deste tema é estudar estas questões em relação às variáveis qualitativas. São estas as comparações associadas à estatística. Geralmente, estes testes consistem em tomar uma amostra e observar se há diferença significativa entre as frequências observadas e aquelas especificadas pela lei teórica do modelo de comparação, também conhecidos como “testes de frequências esperadas”.

No entanto, embora este seja o aspecto mais conhecido, o uso deste teste não se limita ao estudo de variáveis qualitativas. Poderíamos dizer que existem três aplicações básicas deste teste, cujo desenvolvimento abordaremos no decorrer deste capítulo:

Abordaremos três temas:

1. Teste de bondade de ajuste: consiste em determinar se os dados de certa amostra correspondem a uma determinada distribuição populacional. Neste caso, é necessário que os valores da variável na amostra, e sobre a qual queremos fazer a inferência, estejam divididos em tipos de ocorrência, ou seu equivalente. Seja qual for a variável em estudo, deveremos categorizar os dados atribuindo seus valores a diferentes classes ou grupos;

2. Teste de homogeneidade de várias amostras qualitativas: consiste em comprovar se várias amostras de caráter qualitativo procedem da mesma população (por exemplo: estas três amostras de pacientes provêm de populações com a mesma distribuição de doenças?). É necessário que as duas variáveis mensuráveis estejam representadas mediante categorias com as quais construiremos uma tabela de contingência;

3. Teste de independência: comprova se duas características qualitativas estão relacionadas entre si (por exemplo: a presença de destruição vertebral está relacionada à presença de um tumor?). Embora, conceitualmente seja diferente do anterior, operacionalmente proporciona os mesmos resultados. Este tipo de comparações aplica-se quando desejamos comparar uma variável em duas situações diferentes.

Fórmula:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \approx \chi_{k-p-h}^2$$

Exemplo:

Procura-se saber se uma determinada doença afeta do mesmo modo homens e mulheres. Para isso, considera-se uma amostra de $n = 618$ indivíduos que padecem da mesma doença, e observa-se que 341 são homens e o restante, mulheres. Qual a conclusão?

Se p for a porcentagem de homens na população de doentes, podemos considerar a comparação:

$$\begin{cases} H_0 : p = 1/2 \\ H_1 : p \neq 1/2 \end{cases}$$

Da amostra, obtivemos a seguinte estimativa pontual da porcentagem de doentes do sexo masculino: $p = 341/618 = 0.55178$.

LEITURA COMPLEMENTAR

1. P. Armitage, G. Berry, Estadística para la Investigación Biomédica. Doyma, Barcelona, 1992.
2. A. Martín Andrés, J.D. Luna del Castillo, Bioestadística para las Ciencias de la salud. Norma, Granada, 1994.
3. D. Peña Sánchez de Rivera, Estadística: Modelos y Métodos, Alianza Universidad Textos, Madrid, 1994.
4. J. Montanero Fernández, Manual de bioestadística. Universidad de Extremadura. Departamento de matemáticas. 2001.
5. J. S. Milton. Estadística para Biología y Ciencias de la Salud. Ed. Interamericana. McGraw-Hill.
6. E. Cobo, P. Muñoz y J.A. González.(2007) Bioestadística para no estadísticos. Ed. Elsevier/Masson.
7. Peña Sánchez de Rivera, D. (1987): "Estadística. Modelos y Métodos. Volumen 2". Alianza Editorial. Madrid.
8. J. Monge, Estadística no paramétrica. Secretaria de Educación y Universidades. www.uoc.edu. Consultado 03/09/2016.

TESTE DE POSTOS DE WILCOXON

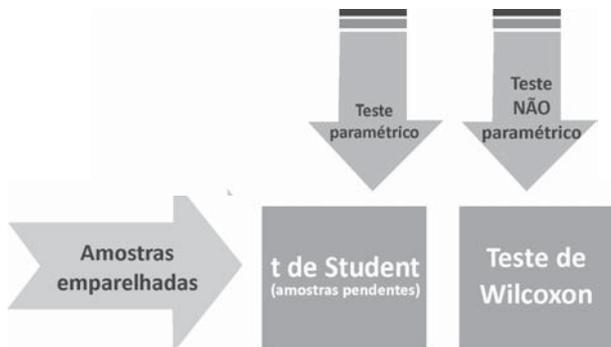
Baron Zarate

O teste de Wilcoxon é um teste de comparação de hipóteses paralelo (similar) ao teste paramétrico de t de Student para amostras pareadas. É usado para comparar variáveis de dois grupos relacionados que NÃO têm uma distribuição normal ou que são variáveis *ordinais*; ou quando as variáveis não cumprem os requisitos básicos dos testes paramétricos.

Deve ser usado de maneira *direta*, sempre em estudos com menos de 10 observações ($n < 10$).

Neste teste, quando a variável DIF não cumpre os critérios de normalidade, trabalha-se com a diferença de medianas. (Primeiro, faz-se os testes de normalidade K-S ou S-W, se não houver normalidade, pode-se tentar transformá-la de maneira logarítmica e, se não se obtiver, continuar com o teste). (Figura 1).

Figura 1 – Testes de comparação de hipóteses para amostras relacionadas



Os passos a seguir no teste de Postos de Wilcoxon são os seguintes:

1. ajustar as observações pareadas e obter as diferenças de cada par;
2. ajustar as diferenças em função de postos como valores absolutos, sem importar o sinal, mas de maneira que estas conservem o sinal correspondente à diferença;
3. obter o somatório dos postos cujo sinal é o menos frequente, por exemplo: se o sinal for +, será considerado para efetuar somatórios; não obstante, o somatório mencionado acaba perdendo o sinal;
4. se forem amostras pequenas, comparar o valor obtido com os valores críticos da tabela de Wilcoxon;
5. distribuir as amostras maiores que 25 sob a curva normal e, portanto, calcular o valor Z, consultando a probabilidade de diferir em relação à média na tabela de probabilidades associadas;
6. decidir se aceitar ou rejeitar a hipótese.

Exemplo: Neste exemplo, tenta-se verificar se existe significância estatística na mudança da escala de mielopatia cervical de Nurick, em pacientes operados de uma descompressão cervical. Usou-se uma escala ordinal de 0 a 5, e avaliou-se mediante o teste de Wilcoxon.

Nossa hipótese nula (H_0) é de que não existe diferença no grau da mielopatia dos pacientes, antes e depois do tratamento cirúrgico.

$$H_0 \cong \mu \text{ antes} = \mu \text{ depois}$$

A hipótese alternativa (H_1) é de que existe uma diferença estatisticamente significativa (positiva), na melhora da disfunção medular.

$$H_1 \cong \mu \text{ antes} \neq \mu \text{ depois}$$

$$\text{Soma de postos: } n(n+1) / 2$$

1. O primeiro passo seria verificar a diferença dos valores do antes e do depois, mas neste caso o n só seria considerado nos valores com diferença DIFERENTE de 0. (Tabela 1).

Tabela 1

| Caso | Nurick pre | Nurick post | DIF | DIF. Absoluta | Ordem | Rangos | Soma Postos + | Soma Postos - |
|------|------------|-------------|-----|---------------|-------|--------|---------------|---------------|
| 1 | 4 | 4 | 0 | | 0 | | | |
| 2 | 3 | 1 | 2 | 1 | 1 | 3 | 3 | |
| 3 | 4 | 1 | 3 | 1 | 2 | 3 | 3 | |
| 4 | 3 | 3 | 0 | 1 | 3 | 3 | 3 | |
| 5 | 1 | 1 | 0 | 1 | 4 | 3 | 3 | |
| 6 | 5 | 3 | 2 | 1 | 5 | 3 | 3 | |
| 7 | 4 | 3 | 1 | 2 | 6 | 7 | 7 | |
| 8 | 3 | 3 | 0 | 2 | 7 | 7 | 7 | |
| 9 | 3 | 2 | 1 | 2 | 8 | 7 | 7 | |
| 10 | 3 | 3 | 0 | 3 | 9 | 9 | 9 | |
| 11 | 4 | 2 | 2 | | | | | |
| 12 | 3 | 2 | 1 | | | | | |
| 13 | 3 | 2 | 1 | | | | | |
| 14 | 5 | 4 | 1 | | | | | |
| 15 | 2 | 2 | 0 | | | | | |
| | | | | | | | 45 | 0 |

Neste caso, n seria 9. Consideram-se somente os valores da DIF diferentes de 0.

2. O passo seguinte seria ordenar o valor absoluto (não se considera o valor do sinal +ou -) em ordem crescente e, posteriormente, atribuir o posto médio de cada intervalo. (Tabela 2).

Tabela 2

| Caso | Nurick pre | Nurick post | DIF | DIF. Absoluta | Ordem | Rangos | Soma Postos + | Soma Postos - |
|------|------------|-------------|-----|---------------|-------|--------|---------------|---------------|
| 1 | 4 | 4 | 0 | | 0 | | | |
| 2 | 3 | 1 | 2 | 1 | 1 | 3 | 3 | |
| 3 | 4 | 1 | 3 | 1 | 2 | 3 | 3 | |
| 4 | 3 | 3 | 0 | 1 | 3 | 3 | 3 | |
| 5 | 1 | 1 | 0 | 1 | 4 | 3 | 3 | |
| 6 | 5 | 3 | 2 | 1 | 5 | 3 | 3 | |
| 7 | 4 | 3 | 1 | 2 | 6 | 7 | 7 | |
| 8 | 3 | 3 | 0 | 2 | 7 | 7 | 7 | |
| 9 | 3 | 2 | 1 | 2 | 8 | 7 | 7 | |
| 10 | 3 | 3 | 0 | 3 | 9 | 9 | 9 | |
| 11 | 4 | 2 | 2 | | | | | |
| 12 | 3 | 2 | 1 | | | | | |
| 13 | 3 | 2 | 1 | | | | | |
| 14 | 5 | 4 | 1 | | | | | |
| 15 | 2 | 2 | 0 | | | | | |
| | | | | | | | 45 | 0 |

3. Ordenar e somar os postos positivos e negativos. (Tabela 3).

Tabela 3

| Caso | Nurick pre | Nurick post | DIF | DIF. Absoluta | Ordem | Rangos | Soma Postos + | Soma Postos - |
|------|------------|-------------|-----|---------------|-------|--------|---------------|---------------|
| 1 | 4 | 4 | 0 | | 0 | | | |
| 2 | 3 | 1 | 2 | 1 | 1 | 3 | 3 | |
| 3 | 4 | 1 | 3 | 1 | 2 | 3 | 3 | |
| 4 | 3 | 3 | 0 | 1 | 3 | 3 | 3 | |
| 5 | 1 | 1 | 0 | 1 | 4 | 3 | 3 | |
| 6 | 5 | 3 | 2 | 1 | 5 | 3 | 3 | |
| 7 | 4 | 3 | 1 | 2 | 6 | 7 | 7 | |
| 8 | 3 | 3 | 0 | 2 | 7 | 7 | 7 | |
| 9 | 3 | 2 | 1 | 2 | 8 | 7 | 7 | |
| 10 | 3 | 3 | 0 | 3 | 9 | 9 | 9 | |
| 11 | 4 | 2 | 2 | | | | | |
| 12 | 3 | 2 | 1 | | | | | |
| 13 | 3 | 2 | 1 | | | | | |
| 14 | 5 | 4 | 1 | | | | | |
| 15 | 2 | 2 | 0 | | | | | |
| | | | | | | | 45 | 0 |

Neste caso, o valor obtido da maior soma de postos (45) é comparado aos valores críticos da tabela T, em testes de postos com sinais, em pares iguais de Wilcoxon. Podemos ver que para ser significativo (ou seja, abaixo de 0.05, que foi o nível de significância) deve ser menor de 0.05. (Tabela 4).

Tabela 4 – Valores críticos da Tabela de Wilcoxon

A hipótese nula será rejeitada se o maior da soma de postos for \geq que:

| n | Com $\alpha = 0.01$ | Com $\alpha = 0.05$ |
|----|---------------------|---------------------|
| 6 | | 21 |
| 7 | | 26 |
| 8 | 36 | 33 |
| 9 | 44 | 40 |
| 10 | 52 | 47 |
| 11 | 61 | 56 |
| 12 | 71 | 65 |
| 13 | 82 | 74 |
| 14 | 93 | 84 |
| 15 | 105 | 95 |
| 16 | 117 | 107 |
| 17 | 130 | 119 |
| 18 | 144 | 131 |
| 19 | 158 | 144 |
| 20 | 173 | 158 |
| 21 | 189 | 173 |
| 22 | 205 | 187 |
| 23 | 222 | 203 |
| 24 | 239 | 219 |

Neste caso, o valor é de 45 com um n de 9. O valor é menor que 0.01, portanto rejeita-se a Hipótese Nula (H_0).

Nota: Quando o valor de n for igual ou maior que 25, deve-se fazer o cálculo manual de Z . (Tabela 5).

$$Z = \frac{\sum (R+) - (n(n+1)/4)}{\sqrt{n(n+1)(n2 + 1)/24}}$$

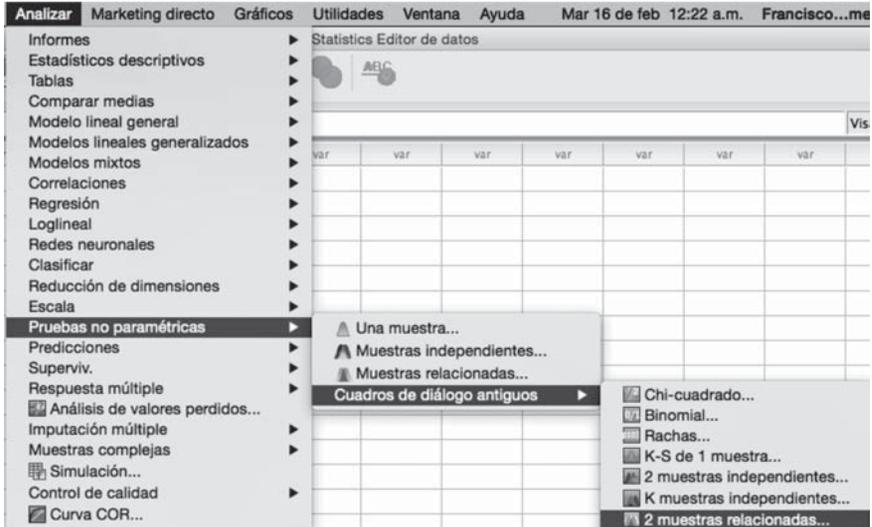
Tabela 5 – Valores críticos de Z e sua significância (uma ou duas caudas)

| Percentis da distribuição normal | | | | | | |
|----------------------------------|------------------|--------|--------|-----------------|--------|--------|
| | Teste Unilateral | | | Teste bilateral | | |
| Significância | 0.05 | 0.01 | 0.001 | 0.05 | 0.01 | 0.001 |
| Percentil | 95 | 99 | 99.9 | 97.5 | 99.5 | 99.95 |
| z | 1.6449 | 2.3263 | 3.0902 | 1.9600 | 2.5758 | 3.2905 |

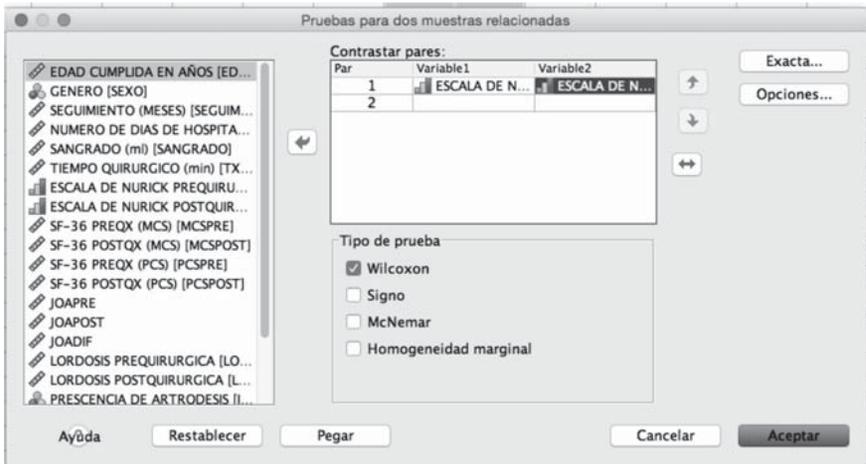
Para realizá-lo com SPSS:

1. Ter as variáveis ordenadas ao introduzir os dados, depois, no menu *analisar* \Rightarrow *testes não paramétricos* \Rightarrow *quadros de diálogo antigos* \Rightarrow *duas amostras relacionadas* \downarrow . (Figura 2).

Figura 2



2. Aparecerá um quadro de diálogo onde é preciso introduzir as duas variáveis a comparar, introduzir a variável antes e, posteriormente, a variável depois. Neste teste, a célula Wilcoxon aparece automaticamente; depois, clicar em aceitar. (Figura 3).



3. a seguir, aparecerão duas tabelas. (Interpretação)

A primeira mostra as variáveis introduzidas (h) e a soma de postos (negativos, positivos, nulos); a segunda (denominada dados estatísticos de comparação), mostra o valor de z e a significância com duas caudas. (Figura 4).

Figura 4 – Teste dos postos com sinal de Wilcoxon

| | | Postos | | |
|---------------------------------------|-------------------------|----------------|-------------|----------------|
| | | N | Posto médio | Soma de postos |
| <i>Escala de Nurick pós-cirúrgica</i> | <i>Postos negativos</i> | 9 ^a | 5,00 | 45,00 |
| <i>Escala de Nurick pré-cirúrgica</i> | <i>Postos negativos</i> | 0 ^b | ,00 | ,00 |
| | <i>Empates</i> | 6 ^c | | |
| | <i>Total</i> | 15 | | |

a. ESCALA DE NURICK PÓS-CIRÚRGICA < ESCALA DE NURICK PRÉ-CIRÚRGICA

b. ESCALA DE NURICK PÓS-CIRÚRGICA > ESCALA DE NURICK PRÉ-CIRÚRGICA

c. ESCALA DE NURICK PÓS-CIRÚRGICA = ESCALA DE NURICK PRÉ-CIRÚRGICA

Dados estatísticos de comparação^a

| | ESCALA DE NURICK PÓS-CIRÚRGICA ESCALA DE NURICK PRÉ-CIRÚRGICA |
|----------------------------|--|
| Z | -2,724 |
| Sinal assintót (bilateral) | ,006 |

a. Teste dos postos com sinal de Wilcoxon

b. Baseado nos postos positivos

A interpretação seria com um p de 0.006, o que indica que há evidência suficiente para rejeitar a hipótese nula H_0 ($p < 0.05$), e aceitar a hipótese alternativa (H_1). Neste caso, a hipótese alternativa (H_1) é direcional para a diminuição da escala de Nurick no pós-operatório, o que indicaria significância estatística na melhora da função com a cirurgia de descompressão.

Nota: Foram usados os mesmos dados para ambos os exemplos: com a forma manual e com SPSS. A decisão de rejeitar a hipótese nula ocorre quando $p = < 0.05$.

LEITURA COMPLEMENTAR

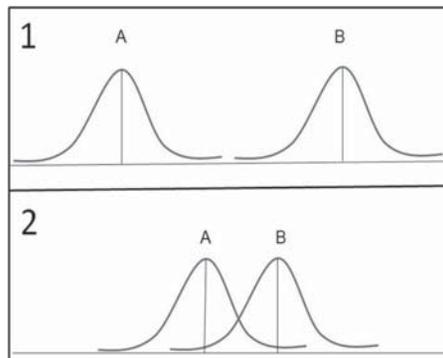
1. Altman DG, Bland JM. Statistics Notes: Transforming data. *BMJ* 1996; 312: 770.
2. Altman DG. Practical statistics for medical research. Londres: Chapman and Hall, 1991.
3. Altman DG, Bland JM. Detecting skewness from summary information, *BMJ* 1996; 313:1200.
4. Altman DG, Bland JM. Statistics Notes: The use of transformations when comparing two means. *BMJ* 1996; 312:1153.
5. Altman DG. Preparing to analyse data. En: Practical statistics for medical research. London: Chapman and Hall; 1991. p.132-145.
6. Berry G., Armitage P. Statistical Methods in Medical Research. 3rd. ed. Oxford: Blackwell Science; 1994.
7. Braitman LE. Confidence intervals asses both clinical significance and statistical significance. *Ann Intern Med* 1991; 114 (6): 515-517.
8. Gehlbach SH. Interpreting the medical literature: practical epidemiology for clinicians. 5th ed. USA: McGraw-Hill. Medical; 2006; p. 166.
9. Livingston EH. Who was student and why do we care so much about his t-test. *J Surg Research* 2004;118:58.
10. Moreno V, Vallescar R, Martín M. Las pruebas no paramétricas en el análisis estadístico de datos. *Aten Primaria* 1991; 8 (1): 58-60.
11. Moser CJ, Stevens MB, Homogeneity of variance in the two-sample means test, *The American Statistician* 46 (1992) 19-21.
12. Newcombe, RG. Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: Asymptotic methods and evaluation. *Statistics in Medicine*, (2006) 25, 559-573.
13. Rajiv Gandhi, Holly N. Smith, Nizar N. Mahomed, Randy Rizek, MD, Mohit Bhandari. Incorrect Use of the Student t Test in Randomized Trials of Bilateral Hip and Knee Arthroplasty Patients. *The Journal of Arthroplasty* Vol. 26 No. 5 2011.
14. Riegelman, RK. Studying a Study and Testing a Test: How to Read the Medical Evidence. Philadelphia: Lippincott Williams & Wilkins, 1999.

TESTE U DE MANN-WHITNEY

*José Manuel Pérez Atanasio
Victor Fernando Luján Celis
Maritza Belén Sandoval Rincón*

Para compreender o teste U de Mann Whitney, precisamos lembrar que os testes de hipóteses têm o objetivo de identificar se a diferença entre os dados de dois grupos é estatisticamente significativa, o que pode ser claramente identificado na seguinte Figura.

Figura 1 – Diferença entre os grupos A e B. 1. a diferença entre os grupos é estatisticamente significativa. 2. A diferença entre os grupos não é estatisticamente significativa



Descrição do teste

O teste de U de Mann-Whitney permite-nos identificar se existe diferença entre duas amostras independentes, a variável de interesse deve apresentar uma distribuição não normal. Este teste consiste em comparar as medianas de ambos os grupos e determinar se existe uma diferença estatisticamente significativa. É preciso enfatizar que o U de Mann Whitney é específico para comparar somente duas amostras

independentes. Se nossa intenção for comparar mais de duas amostras, seriam usados outros testes.

Fórmula

$$U_1 = \frac{R_1 - n_1(n_1 + 1)}{2}$$

$$U_2 = \frac{R_2 - n_2(n_2 + 1)}{2}$$

Onde U_1 = Amostra com o menor número de pacientes

U_2 = Tamanho da amostra maior

R_1 y R_2 = Soma de postos para cada grupo

Exemplo na área de cirurgia de coluna

Para entender mais facilmente este teste, apresentamos o seguinte exemplo:

No Centro Médico Nacional de Coluna, as patologias degenerativas da coluna cervical são tratadas com dois tipos de abordagens cirúrgicas: a abordagem tipo A, defendido pelo Dr. Juan como o melhor, e a abordagem tipo B, defendida pelo Dr. José. Para identificar se existe diferença estatisticamente significativa nos dias que demora a cicatrizar, formulamos a hipótese seguinte:

- Existe uma diferença estatisticamente significativa entre o tempo de cicatrização da abordagem A e da abordagem B.

Em primeiro lugar, registram-se, em uma folha de cálculo, os valores do tempo de cicatrização em ambos os grupos, e faz-se o ordenamento “ranqueado” de todos os dados em ordem crescente, como mostramos no quadro seguinte:

Quadro 1 – Ordenamento dos dados

| Abordagem tipo A n1=6 | Ordem Global (Rank) | Abordagem tipo B n2=5 | Ordem Global (Rank) |
|------------------------|---------------------|-----------------------|---------------------|
| 15 | 5 | 11 | 1 |
| 14 | 4 | 17 | 7 |
| 12 | 2 | 19 | 9 |
| 20 | 10 | 13 | 3 |
| 18 | 8 | 21 | 11 |
| 16 | 6 | | |
| Soma dos postos | 35 | | 31 |

Posteriormente, calcula-se a mediana de cada grupo: a mediana da abordagem tipo A é: 15.5 e a mediana da abordagem tipo B é: 17, agora somamos os valores obtidos na ordem global de cada um dos grupos (soma dos postos), Grupo A: 35 Grupo B: 31, já com estes valores, aplicamos a fórmula.

$$U_1 = \frac{31 - 5(5+1)}{2} = 16$$

$$U_2 = \frac{35 - 6(6+1)}{2} = 14$$

Finalmente, o menor valor de U , que neste caso é 14, é comparado numa tabela de valores críticos de alfa (α) de 0.05 para o teste U de Mann Whitney, visando conhecer o valor de p que, neste caso, é de 0.567.

Interpretação

Como podemos ver o p não é significativo, ou seja, não é <0.05 , motivo pelo qual a interpretação é de que não existe uma diferença estatisticamente significativa entre o tempo de cicatrização com a abordagem A e com a abordagem B, pelo que descartamos a hipótese formulada.

Para realizar este teste nos *softwares* de análise estatística, como o programa SPSS, é necessário selecionar o capítulo de testes não

paramétricos e usar a comparação de medianas entre amostras independentes.

LEITURA COMPLEMENTAR

1. Clifford B.R., Taylor R.A., Bioestadística, México: Pearson Educación; 2008.
2. Dawson B., Trapp R.G., Bioestadística médica. 4ª ed., México: Manual Moderno; 2005.
3. Nachar N., The Mann-Whitney U: A test for Assessing Whether two Independent Samples Come from the Same Distribution, *Tutorials in Quantitative Methods for Psychology*, 2008;vol 4(1):13-20.
4. Rivas. R.R., Moreno P.J., Talavera O.J., Diferenças de medianas con la U de Mann-Whitney, *Rev Med Inst Mex Seguro Soc.* 2013; 51(4):414-9.
5. Rivas. R.R., Moreno P.J., Talavera O.J., Pertinencia de la teste estadística, *Rev Med Inst Mex Seguro Soc.* 2013;51(Supl):S30-34.
6. Wayne W.D. Bioestadística Base para el análisis de las ciencias de la salud. 4ª ed., México: Limusa Wiley; 2011.

TESTE DE KRUSKAL-WALLIS

*Esteban Araya Ramírez
Fabián Víquez Monge*

Esta metodologia é usada quando se tem dois ou mais grupos de dados e a intenção é verificar se eles são provenientes de uma mesma população, ou de populações diferentes, mas com as mesmas características entre si. Estatisticamente, trata-se de comprovar a hipótese de que k amostras independentes provêm de uma população com a mesma mediana.

Este teste de Kruskal é catalogado como um teste não paramétrico, ou seja, não é necessário que os dados analisados possuam uma distribuição com parâmetros específicos. Além disso, tampouco é necessário confirmar a hipótese de igualdade de variâncias (homogeneidade de variâncias).

Matematicamente, o teste é definido da seguinte forma:

$$H = \frac{\frac{12}{N(N+1)} \sum \frac{\sum Rc^2}{ni} - 3(N+1)}{L}$$

Onde:

H = Valor estatístico do teste Kruskal-Wallis

N = tamanho total da amostra

Rc^2 = somatório dos postos elevados ao quadrado

ni = tamanho da amostra de cada grupo

L = ajuste dado pelo ajuste de empate dos postos

O ajuste de L calcula-se da maneira seguinte:

$$L = 1 - \frac{\sum(Li^3 - Li)}{N^3 - N}$$

No estudo, comparam-se três grupos independentes, além disso, o tamanho da amostra é pequeno e a variável resultante é uma variável ordinal. Para análise estatística, o teste escolhido foi o de **Kruskal-Wallis**.

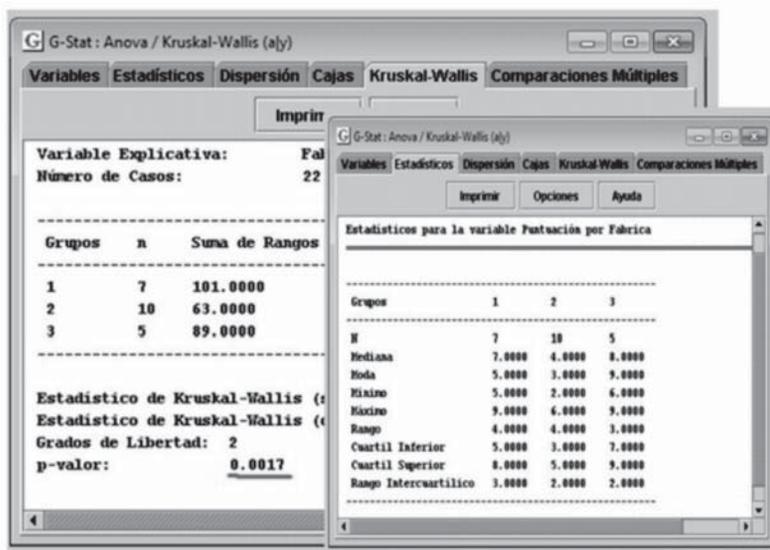
Para isso, mediante ajuda de um programa estatístico, no cabeçalho *Anova*, seleciona-se a opção *Kruskal-Wallis*.

Posteriormente, coloca-se na variável resposta (y) a variável ordinal (variável pontuação), a variável explicativa (b), e a variável qualitativa (variável planta).

A seguir, seleciona-se o cabeçalho de *Kruskal-Wallis* para ver os resultados do teste e comprovar se existem diferenças, além disso, é possível ver, no cabeçalho de estatísticas, as medianas dos três grupos.

No caso do teste de *Kruskal-Wallis*, é preciso prestar atenção no valor de p que é de 0,0017.

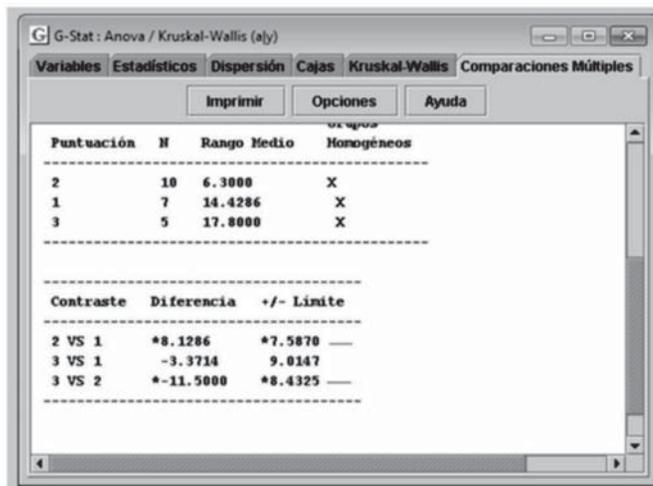
Figura 2



Finalmente, sendo o valor menor de 0,05, rejeita-se a hipótese nula (H_0) de igualdade e aceita-se a hipótese alternativa (H_a) de diferença de medianas com um intervalo de confiança de 95%.

A posteriori, são feitas, mediante o teste de Dunn, as comparações entre os diferentes grupos de estudo. Desta forma, constata-se que existem diferenças entre a planta A e B, e entre a planta B e C.

Figura 3



| Puntuación | N | Rango Medio | Homogéneos |
|------------|----|-------------|------------|
| 2 | 10 | 6.3000 | X |
| 1 | 7 | 14.4286 | X |
| 3 | 5 | 17.8000 | X |

| Contraste | Diferencia | +/- Limite |
|-----------|------------|------------|
| 2 VS 1 | *8.1286 | *7.5870 — |
| 3 VS 1 | -3.3714 | 9.0147 |
| 3 VS 2 | *-11.5000 | *8.4325 — |

Portanto, pode-se dizer que existem diferenças significativas nas medianas da percepção de mal-estar dos trabalhadores no ambiente de trabalho, entre a fábrica A e a fábrica B e a fábrica C, com um intervalo de confiança de 95%.

Não podemos afirmar que existam diferenças significativas entre a fábrica A e a fábrica C.

LEITURA COMPLEMENTAR

1. Altman DG: Practical statistics for medical research. 1ra edición. Chapman and Hall, Londres, 1991.
2. Lopez, JP. Contraste de hipótesis. Comparación de más de dos medias independientes mediante pruebas no paramétricas: Prueba de Kruskal-Wallis. Enfermería del Trabajo 2013; III: 116-117.
3. Siegel S, Castellan NJ Jr: Nonparametric statistics for the behavioral sciences. 2. edición. McGraw-Hill, New York, 1988.

COEFICIENTE DE CORRELAÇÃO DE SPEARMAN

*Fernando Alvarado
Miguel Farfán*

O conjunto de técnicas estatísticas que permitem medir a intensidade de associação ou relação que pode existir entre duas ou mais variáveis é conhecida como correlação. Esta correlação pode existir entre no mínimo duas variáveis (uma dependente e outra independente) ou entre mais de duas variáveis (uma dependente e duas ou mais independentes). No último caso, fala-se de correlação múltipla.

Os tipos de associação (correlação) entre duas variáveis podem ser lineares, positivos, negativos ou nulos.

O coeficiente de correlação de Spearman (Rho de Spearman) é um teste de natureza não paramétrica, útil, para determinar (quantificar) se existe uma correlação entre duas variáveis de natureza numérica (ordinais), cuja distribuição seja não normal ou não paramétrica.

Esta mensuração da correlação tem três objetivos fundamentais: o primeiro é determinar se tal correlação realmente existe (os valores de uma variável tendem a ser mais altos ou mais baixos para os valores mais altos ou mais baixos da outra variável); o segundo, é poder prever o valor de uma variável, conhecendo o valor determinado da outra variável e, finalmente, avaliar o nível de concordância ou a relação entre os valores das variáveis.

O coeficiente de correlação de Spearman tem grande utilidade em variáveis que, dado seu comportamento não normal, apresentam valores extremos em seus dados, que afetariam estatísticas como o coeficiente de correlação de Pearson. (Ver capítulo respectivo).

Cabe ressaltar que o Rho de Spearman mede a existência e a intensidade de associação de duas quantidades e não detecta o nível de acordo ou concordância destas; razão pela qual não deve ser usado para comparar valores procedentes de instrumentos que mensurem o mesmo evento.

Hipótese do coeficiente de correlação de Spearman

H_0 : Não existe associação entre as duas variáveis

H_a : Existe uma relação ou associação entre as duas variáveis (Positiva ou negativa).

Equação

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Onde $d_i = r_{xi} - r_{yi}$ é a diferença entre os postos de X e Y.

Interpretação

Para a interpretação, os autores, antes de realizar o teste, devem definir qual o objetivo da pesquisa, além de determinar a importância das relações das variáveis que estão envolvidas no teste para, no caso de presença de causalidade, não deixar as conclusões da correlação exclusivamente com um estatístico matemático.

A correlação de Spearman mostra um valor numérico da intensidade de associação (positiva ou negativa) de uma tendência de duas variáveis que crescem ou decrescem, e não indica causalidade entre elas.

Segundo a literatura, existem, no mínimo, quatro maneiras de interpretar o Rho de Spearman em escalas numéricas, mas em termos gerais, todas as escalas determinam amplitudes do teste entre -1 e +1. Aqueles valores que se aproximam de 1, indicam uma correlação forte e positiva; aqueles valores que se aproximam de -1, indicam uma correlação forte e negativa. Os valores próximos de 0 indicam a não existência de correlação **linear**. O sinal positivo ou negativo prévio ao valor numérico indica o aumento ou a diminuição de uma variável em relação à outra. Ou seja, um sinal negativo reflete o aumento de uma variável, à medida que a outra diminui ou vice-versa; um sinal positivo indica o aumento de uma variável à medida que a outra o faz, ou a diminuição de uma medida que a outra faz simultaneamente.

É pouco frequente encontrar em estudos clínicos, variáveis com valores aproximados de +0,95. Sugere-se a revisão de dados e cálculos à procura de erros ou inconsistências, com valores maiores ou menores de 1.

Martínez Ortega, do *Instituto Superior de Ciências Médicas de Havana* (ver referência na bibliografia) realiza, juntamente com seus colaboradores, uma revisão da bibliografia em relação à estatística Rho de Spearman e resume 4 diferentes escalas para interpretar o teste, a saber:

Escala 1:

O coeficiente de correlação oscila entre -1 e $+1$, o valor 0 indica que não existe associação linear entre as duas variáveis em estudo.

Escala 2:

| | |
|-------------------------------------|------|
| Correlação negativa perfeita | -1 |
| Correlação negativa moderada frágil | -0,5 |
| Nenhuma correlação | 0 |
| Correlação positiva moderada forte | +0,5 |
| Correlação positiva perfeita | +1 |

Escala 3:

- 1) Perfeita $R = 1$
- 2) Excelente $R = 0.9 \leq R < 1$
- 3) Boa $R = 0.8 \leq R < 0.9$
- 4) Regular $R = 0.5 \leq R < 0.8$
- 5) Ruim $R < 0.5$

Escala 4:

| Posto | Relação |
|------------------|------------------------|
| 0 -0,25 | Escassa ou nula |
| 0,26-0,50 | Frágil |
| 0,51-0,75 | Entre moderada e forte |
| 0,76-1,00 | Entre forte e perfeita |

Exemplo da aplicação da estatística Rho de Spearman

Feise et. al. (ver referência na bibliografia) em seu estudo de construção e validação de uma *escala de qualidade de vida para escoliose idiopática* (SQLI na sigla em inglês) de 22 itens, usam a estatística de correlação Rho de Spearman, ao comparar sua nova escala, com o já conhecido *Perfil de qualidade de vida para deformidades da coluna* (QLPSD na sigla em inglês).

Em um grupo de 84 pacientes, 74 são pacientes-adolescentes operados de escoliose idiopática (manejo cirúrgico, com *brace* ou sem *brace*), e 14 são controles de (irmãos), foram aplicadas as duas escalas (SQLI e QLPSD). Os resultados numéricos globais das escalas foram comparados com a finalidade de determinar se existia ou não uma correlação positiva entre elas.

Feise descreve uma classificação adicional para a interpretação da estatística Rho de Spearman, como mostra a tabela 2.

Ao comparar os valores numéricos globais das duas escalas, obtém-se uma estatística de correlação Rho de Spearman de +0.79, assumindo uma correlação muito boa entre elas (Figura 1).

Tabela 1 – Interpretação da Rho de Spearman de acordo com Feise et al.

| Posto | Relação |
|-----------|-----------|
| 0- 0,2 | Pobre |
| 0,2-0,4 | Frágil |
| 0,41-0,6 | Boa |
| 0,61- 0,8 | Muito boa |
| 0,81-1 | Excelente |

Figura 1 – Tabela de correlação de escalas SQLI y QLPSD em escoliose idiopática do adolescente. Estudo de Feise et al.

| | SQLI Autoestima | SQLI Dor nas costas | SQLI Atividade física | SQLI Humore e Sentimentos | SQLI Global | QLPSD Funcionamento Psicossocial | QLPSD Distúrbios do Sono | QLPSD Dor nas Costas | QLPSD Imagem Corporal | QLPSD Flexibilidade das Costas |
|----------------------------------|-----------------|---------------------|-----------------------|---------------------------|-------------|----------------------------------|--------------------------|----------------------|-----------------------|--------------------------------|
| SQLI dor nas costas | 0.200 | | | | | | | | | |
| SQLI atividade física | 0.264 | 0.283 | | | | | | | | |
| SQLI humor e sentimentos | 0.377 | 0.498 | 0.208 | | | | | | | |
| SQLI global | 0.642 | 0.659 | 0.552 | 0.782 | | | | | | |
| QLPSD Funcionamento Psicossocial | 0.345 | 0.410 | 0.364 | 0.332 | 0.477 | | | | | |
| QLPSD distúrbio do sono | 0.301 | 0.476 | 0.107 | 0.576 | 0.525 | 0.550 | | | | |
| QLPSD dor nas costas | 0.293 | 0.813 | 0.354 | 0.523 | 0.709 | 0.636 | 0.564 | | | |
| QLPSD imagem corporal | 0.545 | 0.462 | 0.220 | 0.465 | 0.624 | 0.415 | 0.517 | 0.538 | | |
| QLPSD flexibilidade das costas | 0.343 | 0.466 | 0.457 | 0.351 | 0.549 | 0.545 | 0.313 | 0.562 | 0.334 | |
| QLPSD global | 0.478 | 0.697 | 0.381 | 0.611 | 0.788 | 0.723 | 0.730 | 0.861 | 0.753 | 0.706 |

LEITURA COMPLEMENTAR

1. Abraira. V, Pérez de Vargas A. Métodos Multivariantes en Bioestadística. ED. Centro de Estudios Ramón Areces. España; 1996.
2. Feise RJ, Donaldson S, Crowther ER, Menke JM, Wright JG. Construction and validation of the scoliosis quality of life index in adolescent idiopathic scoliosis. *Spine (Phila Pa 1976)*. 2005 Jun 1;30(11):1310-5. PubMed PMID:15928558.
3. Martínez Ortega, R. M. (2009). El coeficiente de correlación de los rangos de Spearman caracterización. *Rev. Haban Cienc Méd La Habana, VIII(2)*, 1-19. Retrieved from <http://scielo.sld.cu/pdf/rhcm/v8n2/rhcm17209.pdf>
4. Pita Fernández S, Pértega Díaz S. Unidad de Epidemiología Clínica y Bioestadística. Utilización e Interpretación de las Técnicas de Correlación disponibles. Complejo Hospitalario Juan Canalejo. A Coruña (España): Cad Aten Primaria; 4: 141.
5. Santander Montes AJ, Ruiz Vaquero R. Relación entre variables cuantitativas. Cuba. Informática Médica II. Editorial ECIMED; 2004, p.1-30.

TESTES DE MCNEMAR

*Esteban Araya Ramírez
Fabián Víquez Monge*

Em certas ocasiões, várias medidas da mesma característica são feitas para cada um dos indivíduos incluídos em uma pesquisa. Nesses casos, o foco é comparar se há ou não mudanças significativas entre as medidas feitas em dois momentos diferentes, geralmente antes e depois da aplicação de um tratamento ou do desempenho de uma intervenção.

O teste de McNemar é usado para decidir se é possível ou não aceitar que determinado *tratamento* induza a uma mudança na resposta dos elementos submetidos ao mesmo. Em certas ocasiões, por exemplo, são realizadas várias mensurações da mesma característica para cada um dos indivíduos incluídos numa pesquisa, sendo, nestes casos, o interesse tem o objetivo de comparar se há ou não mudanças significativas entre as mensurações efetuadas em dois momentos diferentes, normalmente, antes e depois da aplicação de um tratamento ou da realização de alguma intervenção. O teste é aplicável aos desenhos *“antes – depois” nos quais cada elemento atua como seu próprio controle e consiste num teste não paramétrico para duas variáveis binárias relacionadas, que compara as mudanças nas respostas usando a distribuição do $\text{Chi}^2 \text{ X}^2$.*

Não é útil somente para detectar as mudanças nas respostas de estudos experimentais antes e depois, mas também para comparar dois tipos de tratamento.

Os resultados correspondentes a uma amostra de **N** elementos são dispostos em uma tabela de frequências de 2x2, para coletar o conjunto das respostas dos mesmos elementos, antes e depois. A escala de mensuração para X e Y é nominal com duas categorias tais como: positivo ou negativo, fêmea ou macho, presença ou ausência, que podem ser denominadas 0 e 1.

Quadro 1

| | | Classificação Y_i | |
|---------------------|---------------|---------------------|-------------------|
| | | (+) $Y_i = 0$ | (-) $Y_i = 1$ |
| Classificação X_i | (+) $X_i = 0$ | A (0,0) | B (0,1) |
| | (-) $X_i = 1$ | C (1,0) | D (1,1) |

Para exemplificar ainda melhor o quadro anterior, pode-se simplificar para:

| Antes / Depois | - | + |
|----------------|---|---|
| - | a | b |
| + | c | d |

Nas células da tabela, **a** é o número de elementos cuja resposta é a mesma (-); **b** é o número de elementos cuja resposta é (-) antes do *tratamento* e (+) depois deste; **c** é o número de elementos que mudaram de (+) para (-); e **d** é o número de elementos que mantém a resposta (+).

Portanto, $b + c$ é o número total de elementos cujas respostas mudaram e são os únicos que vão intervir na comparação. A hipótese nula é que o *tratamento* não induz a mudanças significativas nas respostas, ou seja, as mudanças observadas na amostra são devidas ao acaso, de forma que é igualmente provável uma mudança de $+ a -$ que uma mudança de $- para +$. Então, se H_0 estiver certa, dos $b+c$ elementos cuja resposta mudou é de se esperar que $(b+c)/2$ tenham passado de $+ para -$, e $(b+c)/2$ tenham passado de $- para +$. Em outras palavras, se H_0 estiver certa, a frequência esperada nas células correspondentes é de $(a+b)/2$.

A hipótese alternativa pode ser não direcional, quando postula que a probabilidade de uma mudança de $+ a -$ tem probabilidade diferente do que uma mudança de $- a$

$+$, ou direcional, quando prevê que uma mudança de $- a +$ é mais (ou menos) provável que uma mudança de $+ a -$.

A estatística de teste que permite comparar se existem diferenças significativas entre as frequências esperadas e as frequências observadas é:

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

O_i = frequência observada i -ésima célula

E_i = frequência esperada na i -ésima célula se H_0 estiver certa

K = número de células

Para comparar a significação das mudanças interessam somente aquelas células que mostram mudanças, portanto, a estatística pode ser expressa como:

$$X^2 = \frac{\left[b - \frac{b+c}{2}\right]^2}{\frac{b+c}{2}} + \frac{\left[c - \frac{b+c}{2}\right]^2}{\frac{b+c}{2}} = \frac{(b-c)^2}{b+c}$$

Se H_0 estiver certa, a estatística X^2 tem distribuição aproximadamente Chi^2 com um grau de liberdade. A aproximação será mais precisa se se fizer a correção de continuidade de Yates e, nesse caso, a estatística ficaria assim:

$$X^2 = \frac{(|b-c| - 1)^2}{b+c}$$

Exemplo

Um pesquisador em medicina preventiva observa que os empregados de uma fábrica apresentam frequentemente um quadro diarreico, motivo de grande ausência ao trabalho.

Todos os empregados comem no refeitório da fábrica, como benefício do trabalho.

O pesquisador pressupõe que o denominador comum da causa da diarreia é o local da ingestão de alimentos, ou seja, existe uma higiene inadequada na preparação dos alimentos; no entanto, a higiene pessoal dos empregados não é suficiente para atribuir toda a culpa ao pessoal da cozinha.

Portanto, ele é selecionado uma amostra ao acaso, de 50 indivíduos, constatando sendo que 34 deles apresentam um quadro diarreico frequente e 16 não. Sugere, então, a aplicação de medidas de higiene pessoal, supervisionadas, que consistem em lavagem exaustiva das mãos antes de ingerir alimentos, durante um período de duas semanas.

Ao finalizar o tratamento, obtém os seguintes resultados: dos 34 sujeitos com um quadro diarreico frequente, após o tratamento de lavagem de mãos, 16 conseguiram fazer desaparecer o processo intestinal e 18 persistiram com evacuações diarreicas; por sua vez, no grupo de 16 pessoas assintomáticas, 4 delas apresentaram diarreia apesar da lavagem de mãos e 12 se mantiveram nas mesmas condições.

Escolha do teste estatístico

O modelo experimental tem duas amostras dependentes.

Elaboração da hipótese

- Hipótese alternativa (H_a) A lavagem de mãos, como medida preventiva e fator de higiene pessoal, apresenta mudanças significativas de doença diarreica nos empregados que frequentam o refeitório da fábrica em estudo.
- Hipótese nula (H_0): as diferenças que se observam nas frequências de alteração, devido à lavagem de mãos, são obra do acaso.

Nível de significância

Para todo valor de probabilidade igual ou menor que 0.05, aceita-se H_a e rejeita-se H_0 .

Zona de rejeição

Para todo valor de probabilidade $>$ que 0.05, aceita-se H_0 e rejeita-se H_a .

Resultado da contingência 2×2 .

| | | Depois da lavagem de mãos | |
|--------------------------|---|---------------------------|----|
| | | - | + |
| Antes da lavagem de mãos | - | 16 | 18 |
| | + | 12 | 4 |

Aplicação do teste estatístico

$$X^2 = \frac{((A - D) - 1)^2}{A + B} = \frac{(16 - 4 - 1)^2}{16 + 4} = \frac{121}{20} = 6.05$$

Cálculo dos graus de liberdade (gl)

$$Gl = 1$$

O teste estatístico X^2 de McNemar é comparado com os valores críticos do Chi^2 , com o grau de liberdade, observando-se que 3.84 é correspondente a uma probabilidade de 0.05, enquanto o cálculo corresponde a uma probabilidade < que 0.05.

Decisão

Devido a que o valor de X^2 calculado tem probabilidade < que 0.05, cai no nível de significância, portanto, aceita-se H_a e rejeita-se H_o .

Interpretação

A lavagem de mãos exerceu alterações significativas na doença diarreica dos empregados da fábrica, o que significa que a higiene pessoal deficiente contribui de maneira marcante para a frequência do processo intestinal, causa do ausentismo.

A aplicação do teste estatístico não cancela a participação da higiene defeituosa na preparação dos alimentos pelo pessoal da cozinha, já que 4/16 empregados que eram assintomáticos e sob exaustivo processo de lavagem de mãos antes da refeição, manifestaram diarreia; no entanto, o teste estatístico revela que a higiene pessoal participa de maneira mais significativa que o defeito na preparação da refeição.

LEITURA COMPLEMENTAR

1. Daniel WW: Estadística con aplicaciones a las Ciencias Sociales y a la educación. 1ra, edición en español. McGraw-Hill, México, 1988.
2. Manuel G.M, Danglot B.C, Vega F. L. Sinopsis de pruebas estadísticas no paramétricas. Cuándo usarlas. *Rev Mex Pediatr* 2003; 70(2); 91-99.
3. Martínez, M.A, Sánchez, A.V, Faulin F.J. Bioestadística Amigable. Segunda edición. España: Editorial Diaz de Santos; 2009.
4. Siegel. S, Castellan N.J. Estadística no paramétrica. Segunda edición. Méjico: Editorial Trillas; 1995.
5. Sprent, P. Applied nonparametric statistical methods. Tercera edición. Boca Ratón; 2001.

REGRESSÃO LINEAR SIMPLES

*Alisson R. Teles
Asdrubal Falavigna*

Descrição do teste

A regressão linear simples é um modelo matemático usado para descrever a relação entre duas variáveis, com o objetivo de utilizar uma delas para se prever o valor da outra. Em outras palavras, é uma técnica estatística usada quando há razões para supor uma relação de causa-efeito entre duas variáveis quantitativas e se deseja expressar matematicamente essa relação.

Os objetivos de um teste de regressão linear simples são, portanto: (1) avaliar uma possível dependência entre duas variáveis quantitativas (x = variável preditora ou independente / y = variável dependente ou desfecho), e (2) expressar matematicamente essa resposta por meio de uma equação.

Por exemplo, pode-se calcular a dependência linear entre a lordose lombar e a incidência pélvica de uma determinada população. Outro exemplo muito utilizado em pesquisa médica é quando se quer saber a relação dose-resposta de uma determinada droga, para que se possa prever a reação fisiológica de um indivíduo em função do uso de uma determinada quantidade de medicamento.

Os requisitos para o estudo da regressão linear são: a variável dependente (y) deve ser aleatória, isto é, obtida por amostragem; as variáveis independente (x) e dependente (y) devem estar associadas linearmente (o gráfico entre as duas variáveis deve apresentar uma reta); e as variâncias da variável dependente (y), dados diferentes valores da variável independente (x) são todas iguais.

Fórmula e interpretação

A equação apresenta as variáveis dependente (y) a independente (x) e um coeficiente, que reflete a intensidade da relação entre as variáveis x e y .

A equação da análise de regressão é definida por: $y = a + bx$, onde y é a variável dependente (desfecho), x a variável independente (preditora), a é o coeficiente linear (valor de y quando $x = 0$), e b é o coeficiente angular (inclinação da reta; acréscimo ou decréscimo em y para cada acréscimo de uma unidade em x). Graficamente, a regressão linear simples é representada pelo diagrama de dispersão, também chamado de diagrama de ponto. A equação representa-se pela linha reta que expressa a relação linear entre o valor médio de y para os distintos valores correspondentes de x .

Para determinar-se a reta da regressão linear simples, deve-se identificar o valor de a (intercepto vertical) e o valor de b (coeficiente angular). A estimação desses parâmetros é feita por uma técnica conhecida como método de mínimos quadrados. Calculam-se os valores de a e b , de modo que a soma dos quadrados dos desvios dos pontos observados y e a própria reta seja um mínimo. A distância vertical entre o ponto e a reta é chamada erro ou resíduo da reta em relação ao ponto. Os valores de a e b são calculados pela seguinte fórmula:

$$b = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

Uma vez calculado o valor de b (coeficiente angular), o valor de a (intercepto vertical) é dado por:

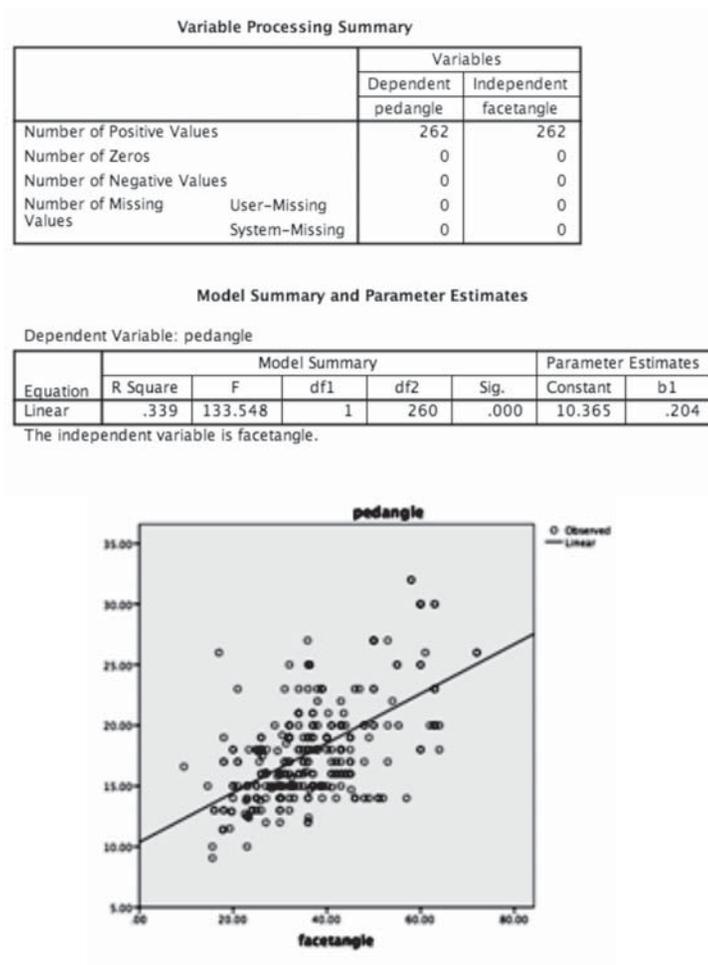
$$a = \bar{y} - b\bar{x}$$

Onde \bar{x} e \bar{y} são, respectivamente, as médias dos dados relativos às variáveis x e y . A reta estimada de regressão linear é: $\hat{y} = a + bx$, onde \hat{y} é a estimativa da observação y para um valor específico da variável x .

Na prática, a equação da regressão linear simples pode ser gerada por diversos *softwares* de análises de dados. A análise gera primeiramente dados que identificam a existência ou não de correlação significativa entre as variáveis (pré-requisito para obtenção da equação da regressão); fornece o gráfico de dispersão com a reta plotada e

identifica os valores de a e b . Na figura, apresenta-se um exemplo de regressão linear simples utilizando-se o *software* SPSS®.

Figura 1 – Exemplo de análise de regressão linear simples utilizando-se o SPSS®



Nota: No exemplo acima, demonstra-se a relação entre o ângulo do pedículo (pedangle) e o ângulo da articulação interapofisária (facetangle) de vértebras lombares aferidos em tomografia computadorizada. Observa-se (1) que a correlação entre as duas variáveis é significativa (Sig. .000, representa o valor $P < 0.0001$), (2) identifica-se o gráfico de dispersão com a reta da regressão plotada, (3) identificam-se os valores de a (constant) e b (b_1). Portanto, pode-se dizer que, na amostra estudada, a relação entre o ângulo do pedículo pode ser definido pela seguinte equação: $\text{ângulo do pedículo} = 10.365 + (0.204 \times \text{ângulo da faceta})$.

Exemplo na área de cirurgia de coluna

A análise de regressão linear simples é utilizada quando se pretende prever uma variável em relação à variação da outra. Na prática, primeiramente identifica-se a existência de correlação significativa entre duas variáveis contínuas, após realiza-se análise de regressão linear para gerar a equação que determina a variável desfecho (y), em relação aos valores da variável preditora (x).

Diversas são as aplicações na área de coluna vertebral. Por exemplo, Hasegawa e colaboradores identificaram a relação entre a incapacidade funcional medida pelo índice de incapacidade Oswestry (ODI: *Oswestry Disability Index*) e (1) idade do voluntário de pesquisa e (2) *mismatch* entre o índice pélvico e lordose lombar. Eles estudaram 136 voluntários sem doença da coluna vertebral com radiografia e avaliação clínica. No estudo, eles identificaram a seguinte equação entre ODI e idade: $\text{ODI} = 0.36 + 0.12 \times \text{idade}$. Nessa análise, eles obtiveram um valor r (coeficiente de correlação) de 0.2143 (Capítulo de Correlação de Pearson) e um valor de $p = 0.01664$. A equação entre ODI e *mismatch* incidência pélvica (IP) e lordose lombar (LL) identificada foi: $\text{ODI} = 5.62 + 0.12 \times (\text{IP} - \text{LL})$, com valor $r = 0.1969$ e $P = 0.0277$. Com essa equação, pode-se prever o valor de ODI levando-se em consideração a idade e também o mismatch IP-LL.

LEITURA COMPLEMENTAR

1. Hasegawa K, Okamoto M, Hatsushikano S, Shimoda H, Ono M, Watanabe K. Normative values of spino-pelvic sagittal alignment, balance, age, and health-related quality of life in a cohort of healthy adult subjects. *Eur Spine J* 2016, DOI 10.1007/s00586-016-4702-2.
2. Peat J, Barton B. *Medical Statistics: a guide to data analysis and critical appraisal*. BMJ Books, 2008. ISBN: 978-0-7279-1812-3.
3. Peacock JL, Peacock PL. *Oxford Handbook of Medical Statistics*. Oxford University Express, 2011. ISBN: 978-0-19-9555128-6.

DISTRIBUIÇÃO DE POISSON

*José Acuña Pinilla
Margarita Gómez Chantraine
Elisabeth Heins*

Descrição do teste

O teste é assim denominado em honra do físico e matemático de origem francesa Simeón-Denis Poisson (1781-1840), que o divulgou em 1837 por meio de seu trabalho: *Recherches sur la probabilité des jugements en matières criminelles et matière civile* (Investigação sobre a probabilidade dos julgamentos em matérias criminais e civis).

A distribuição de Poisson é uma das funções de probabilidade mais importantes e faz parte do grupo de funções para a distribuição da probabilidade para variáveis aleatórias discretas. Ou seja, os eventos só podem ter valores inteiros (ex.: 0,1,2,3,4...n), e os eventos aleatórios que aparecem são independentes entre si, e se apresentam durante um intervalo de tempo o espaço.

Pontualmente, o teste é especializado na frequência de eventos com probabilidades muito baixas em relação à frequência da não ocorrência e, por essa razão, são considerados *eventos raros*. Sua aproximação da binomial ocorrerá quando a probabilidade de apresentação do evento (**p**) estiver muito próxima de zero, ou o tamanho da amostra (**n**) for superior a 30 ($p < 0,1, n > 30$).

Esta função parte de uma frequência de ocorrência média (λ) conhecida também como parâmetro positivo para a esperança de ocorrência, e que identifica o número de ocorrências ou eventos médios que se apresentam dentro de um intervalo de tempo ou espaço. Além disso, envolve o número de ocorrências do evento ou fenômeno (**x**) que se espera identificar (a função nos mostra a ocorrência precisa de x vezes). Por exemplo, se o evento estudado ocorre, em média, cinco vezes por minuto e estamos interessados na probabilidade de que ocorra x vezes, dentro de um intervalo de 10 minutos, usaremos um modelo de distribuição de Poisson com $\lambda = 10 \times 5 = 50$.

Características da distribuição de Poisson

A distribuição é resultado de um conjunto de suposições sobre um processo implícito, para formar um conjunto de observações numéricas. As afirmações seguintes descrevem o que se conhece como processo de Poisson:

- a) a ocorrência dos eventos é independente. O aparecimento de um evento (x) em um intervalo de tempo ou espaço não tem efeito sobre a probabilidade de ocorrer um seguinte aparecimento do evento estudado no mesmo intervalo, ou em algum outro;
- b) teoricamente, é possível o aparecimento do evento em um número infinito de vezes, dentro do intervalo de tempo ou espaço;
- c) a probabilidade de um só aparecimento do evento estudado num intervalo determinado é proporcional à dimensão do intervalo de tempo ou espaço (ou algo de volume da matéria);
- d) a probabilidade de mais de uma ocorrência do evento em estudo, em qualquer fração infinitesimal do intervalo de tempo ou espaço, é insignificante.

Uma característica interessante da distribuição de Poisson é que a média e a variância são iguais.^{1,6,10,11}

Aplicação

A distribuição de Poisson é usada para descrever vários processos, entre outros:

- a) o número de lesões vasculares de vasos principais na cirurgia de coluna via anterior, um ano depois;
- b) o número de lesões radiculares em dissectomias minimamente invasivas do total de pacientes operados por esta técnica em um período determinado;
- c) o número de parafusos pediculares que apresentaram *pull-out*, na redução do total de níveis funcionais abordados na cirurgia de escoliose;
- d) o número de parafusos pediculares malposicionados, em cirurgia fusão lombar após um ano;
- e) o número de casos com pseudoartrose posterior a uma cirurgia de fusão cervical antes de um mês.

Cada uma destas variáveis aleatórias representa o número total de ocorrências de um fenômeno durante um período de tempo fixo ou em uma região fixa do espaço. Expressa a probabilidade de um número (x) de aparecimentos do evento, em um tempo determinado, se estes eventos ocorrerem com uma frequência média conhecida e forem independentes do tempo transcorrido desde a última ocorrência ou evento.

A probabilidade do aparecimento do número de eventos pode ser calculada de maneira pontual ou acumulada, ou seja, uma quantidade agrupada de possíveis eventos no intervalo de tempo ou espaço.

Função da Densidade da Probabilidade:

$$f(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{sí } x = 0, 1, 2, 3, \dots, n$$

É possível demonstrar que $f(x) \geq 0$ para cada (x) e que $\sum_x f(x) = 1$, de modo que a distribuição cumpre os requisitos para uma distribuição de probabilidade.

Onde:

λ = Parâmetro esperança de ocorrência (# ocorrências/Unidade de tempo ou espaço)

x = # de eventos do fenômeno esperados (0, 1, 2, 3, 4...n)

e = Base dos logaritmos naturais ($e = 2,71828\dots$)

Função da Probabilidade Acumulada:

$$f(X \leq x) = \sum_0^x \frac{e^{-\lambda} \lambda^x}{x!}$$

Variância σ^2 = Média μ = Parâmetro esperança de ocorrência λ

$$\text{Desvio padrão } \sigma = \sqrt{\sigma^2} = \sqrt{\mu}$$

Gráficos

Figura 1 – A função densidade de probabilidade (FDP) para toda ocorrência em valores inteiros (x), com uma expectativa de ocorrência (λ). A linha contínua é somente um guia visual e não representa continuidade

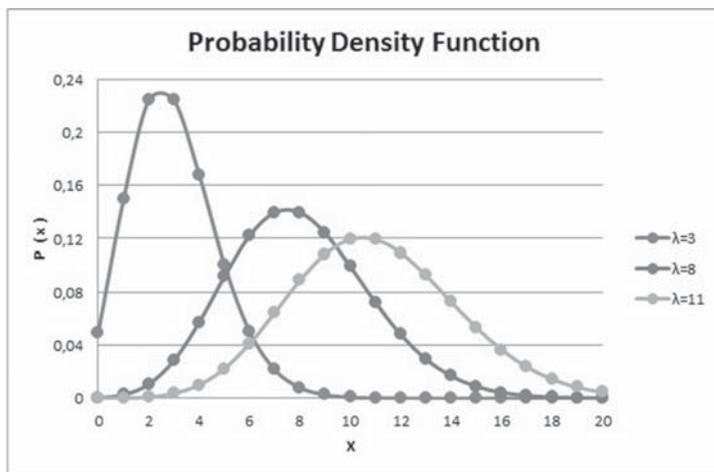
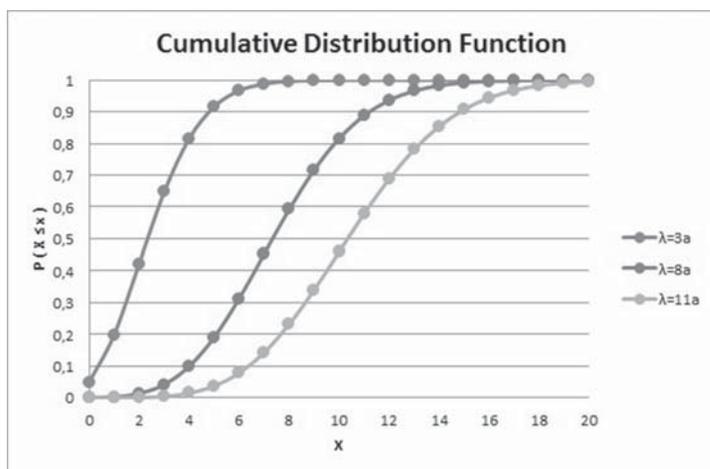


Figura 2 – A função de distribuição acumulada (FDA). A FDA é descontínua nos inteiros de x e plana em todos os lugares devido à variável, que é a distribuição de Poisson, e que considera somente valores inteiros



Exemplos na área de cirurgia de coluna

1. Em uma meta-análise sobre a presença de pseudoartrose em discectomia cervical anterior e fusão como o uso de enxertos, durante o período de seguimento, Shriver MF. et al. (2015), nos 17 estudos finalmente incluídos com revisão até 10 de novembro de 2014, encontraram uma ocorrência de pseudoartrose em 35 casos, durante o pós-cirúrgico de pacientes tratados com aloenxertos (n:736).⁸ Qual a probabilidade de que, durante o ano seguinte, entre os pacientes que passarem por discectomia cervical via anterior e fusão com aloenxerto, apareçam exatamente dois casos com pseudoartrose no seguimento pós-operatório?

Solução:

$$p(x = 2)$$

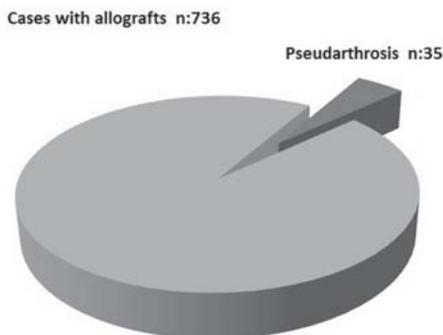
$$\lambda = \frac{35 (\text{Casos de pseudoartrose})}{736 (\text{Casos con Alo - enxertos})} = 0.048$$

$$x = 2 \text{ (casos esperados com pseudo - artrose)}$$

$$f(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$f(x = 2) = \frac{e^{-0.048} * 0.048^2}{2!} = \frac{0.9531 * 0.0023}{2} = 0,0011$$

A probabilidade de aparecimento de exatamente dois casos com pseudoartrose no seguimento de discectomia cervical, via anterior e fusão com alo-enxerto é de 0.11%.

Figura 3 – Frequência de casos de pseudoartrose

2. Em uma meta-análise sobre o correto posicionamento dos parafusos pediculares, com o uso de navegação, comparado com o posicionamento sem navegação, Kosmopoulos V, et al. (2007), encontraram nos estudos incluídos com revisão entre 1966 até 14 de março de 2006, uma mediana de malposicionamento de parafusos de 9,7% para o grupo sem uso de navegação em casos *in vivo*, equivalentes a uma ocorrência mediana mensal de oito parafusos malposicionados.⁵ Qual é a probabilidade de que num mês selecionado aleatoriamente seja aquele em que apareçam três ou quatro casos com parafusos pediculares malposicionados na cirurgia?

Solução:

$$p(x = 3 \text{ o } 4) = p(x = 3) + p(x = 4)$$

$$\lambda = \frac{8 \text{ (Casos de parafusos mal posicionados)}}{1 \text{ (Unidade de tempo em meses)}} = 8$$

$$x_1 = 3 \text{ (Casos esperados de parafusos mal posicionados)}$$

$$x_2 = 4 \text{ (Casos esperados de parafusos mal posicionados)}$$

$$f(X = x_1) = \frac{e^{-\lambda} \lambda^{x_1}}{x_1!} \rightarrow f(x_1 = 3) = \frac{e^{-8} * 8^3}{3!} = \frac{0.00034 * 512}{6} = 0,029$$

$$f(X = x_2) = \frac{e^{-\lambda} \lambda^{x_2}}{x_2!} \rightarrow f(x_2 = 4) = \frac{e^{-8} * 8^4}{4!} = \frac{0.00034 * 4096}{24} = 0,057$$

$$f(x_1 = 3 \text{ o } x_2 = 4) = f(x_1 = 3) + f(x_2 = 4)$$

$$f(x_1 = 3) + f(x_2 = 4) = 0,029 + 0,057 = 0,086$$

A probabilidade de se apresentarem três ou quatro casos com parafusos pediculares malposicionados em cirurgia é de 8,6%.

3. Em um estudo prospectivo multicêntrico sobre o correto posicionamento dos parafusos pediculares, com uso de sistema de imagens com braço em O e navegação Stealth Station, nos níveis torácico, lombar e sacro, Van de Kelft E et al. (2012), nos três centros de neurocirurgia de dois países entre novembro de 2009 e abril de 2011 encontraram uma ocorrência de malposicionamento de parafusos em 2,5%, relatando uma ocorrência anual de cinco parafusos com comprometimento da cortical medial do pedículo.⁹ Qual a probabilidade de que, durante o ano seguinte, entre os pacientes submetidos à fixação via posterior com parafusos pediculares apareçam, na cirurgia, pelo menos quatro parafusos pediculares malposicionados e compromisso da cortical medial?

Solução:

A possibilidade de aparecimento de eventos é infinita, e o somatório de todos os possíveis eventos é igual a 1. Então, podemos usar o conceito de complementariedade ao subtrair de um a probabilidade de três ou menos parafusos malposicionados. Ou seja:

$$p(x \geq 4) = 1 - p(x \leq 3)$$

Pelo exposto, deve-se verificar primeiro a probabilidade acumulada de $p(x \leq 3)$

$$p(x \geq 4) = 1 - [p(x = 0) + p(x = 1) + p(x = 2) + p(x = 3)]$$

$$p(X \leq x) = \sum_0^x \frac{e^{-\lambda} \lambda^x}{x!} \quad \rightarrow \quad p(X \leq 3) = \sum_0^3 \frac{e^{-5} 5^3}{3!}$$

$$= p(x = 0) + p(x = 1) + p(x = 2) + p(x = 3)$$

Então:

$$\lambda = \frac{5 \text{ (Casos de parafusos mal posicionados)}}{1 \text{ (Unidade de tempo em anos)}} = 5$$

$x = 0, 1, 2$ y 3 (Casos esperados de parafusos mal posicionados)

Tabela 1 – Cálculo distribuição de Poisson para densidade, e acumulada para $x = 0$ até $x = 3$ com um $\lambda = 5$

| X | P(X=x) FDP | P(X≤x) FDA |
|---|------------|------------|
| 0 | 0,0067 | 0,0067 |
| 1 | 0,0337 | 0,0404 |
| 2 | 0,0842 | 0,1247 |
| 3 | 0,1404 | 0,2650 |

$$f(X \geq x) = 1 - \sum_0^x \frac{e^{-\lambda} \lambda^x}{x!}$$

$$f(X \geq 4) = 1 - \sum_0^3 \frac{e^{-5} 5^x}{x!}$$

$$p(x \geq 4) = 1 - [p(x = 0) + p(x = 1) + p(x = 2) + p(x = 3)]$$

$$p(x \geq 4) = 1 - \left[\frac{e^{-5} * 5^0}{0!} + \frac{e^{-5} * 5^1}{1!} + \frac{e^{-5} * 5^2}{2!} + \frac{e^{-5} * 5^3}{3!} \right]$$

$$p(x \geq 4) = 1 - [0,0067 + 0,0337 + 0,0842 + 0,1404]$$

$$p(x \geq 4) = 1 - 0,2650 = 0,7350$$

A probabilidade de ocorrerem quatro ou mais casos por ano com parafusos pediculares malposicionados e comprometimento da cortical medial, durante a cirurgia é de 73.50%.

Tabela 2 – Transformação da fórmula de Poisson para o cálculo em Excel

Aplicação

Como calcular distribuição de probabilidade de Poisson em Excel.

| | Função | Fórmula em Excel |
|--------------------------------|--|-------------------------|
| Probabilidade densidade | $f(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$ | = POISSON.DIST(x; λ; 0) |
| Probabilidade acumulada | $f(X \leq x) = \sum_0^x \frac{e^{-\lambda} \lambda^x}{x!}$ | = POISSON.DIST(x; λ; 1) |

Figura 4 – Os componentes da fórmula são:

The screenshot shows an Excel spreadsheet titled "The Poisson Distribution". In cell B3, the parameter "Landa (λ)" is set to 5. A table below shows the probability density function (p(x)) and cumulative distribution function (pi(X≤x)) for x values from 0 to 20. Annotations explain the arguments of the =POISSON.DIST function: x (values 0, 1, 2, ...), lambda (average number of events per interval), and a logical value (FALSE for density, TRUE for cumulative).

| x | p(x) | pi(X≤x) |
|----|-----------|-----------|
| 0 | 0,0067379 | 0,0067379 |
| 1 | 0,0336897 | 0,0404277 |
| 2 | 0,0842243 | 0,1246520 |
| 3 | 0,1403739 | 0,2650259 |
| 4 | 0,1754674 | 0,4404933 |
| 5 | 0,1754674 | 0,6159607 |
| 6 | 0,1462228 | 0,7621835 |
| 7 | 0,1044449 | 0,8666283 |
| 8 | 0,0652780 | 0,9319064 |
| 9 | 0,0362656 | 0,9681719 |
| 10 | 0,0181328 | 0,9863047 |
| 11 | 0,0082422 | 0,9945469 |
| 12 | 0,0034342 | 0,9979811 |
| 13 | 0,0013209 | 0,9993020 |
| 14 | 0,0004717 | 0,9997737 |
| 15 | 0,0001572 | 0,9999310 |
| 16 | 0,0000491 | 0,9999801 |
| 17 | 0,0000145 | 0,9999946 |
| 18 | 0,0000040 | 0,9999986 |
| 19 | 0,0000011 | 0,9999997 |
| 20 | 0,0000003 | 0,9999999 |

(x) número de eventos do fenômeno esperados.

(λ) Parâmetro esperança de ocorrência.

(Acumulado) é um valor lógico, para a função densidade de probabilidade, usar "FALSO" ou "0"; para a função de distribuição acumulada, usar "VERDADEIRO" ou "1".

Tabela 3 – Transformação da fórmula de Poisson para o cálculo em SPSS

| | Função | Fórmula em SPSS |
|--------------------------------|---------------------------------|---|
| Probabilidade densidade | Transformar → Calcular variável | |
| | $f(X = x)$ | Expressão numérica = PDF. POISSON(x, λ) |
| Probabilidade acumulada | Transformar → Calcular variável | |
| | $f(X \leq x)$ | Expressão numérica = CDF. POISSON(x, λ) |

Figura 5 – Introduzir os valores do número de eventos do fenômeno esperado (X), Selecionar em transformar a função calcular variável (1). Nomear a variável para a função densidade de probabilidade (FDP) (2). Selecionar no grupo de funciones “FDP & FDP não centrada” (3). Em funções e variáveis especiais, selecionar clicando duas vezes “Pdf. Poisson” (4). Completar a fórmula substituindo em “(?)” primeiro pela variável X, depois pelo valor do parâmetro esperança de ocorrência (λ) (5). Finalizar com “Aceitar” para executar o cálculo

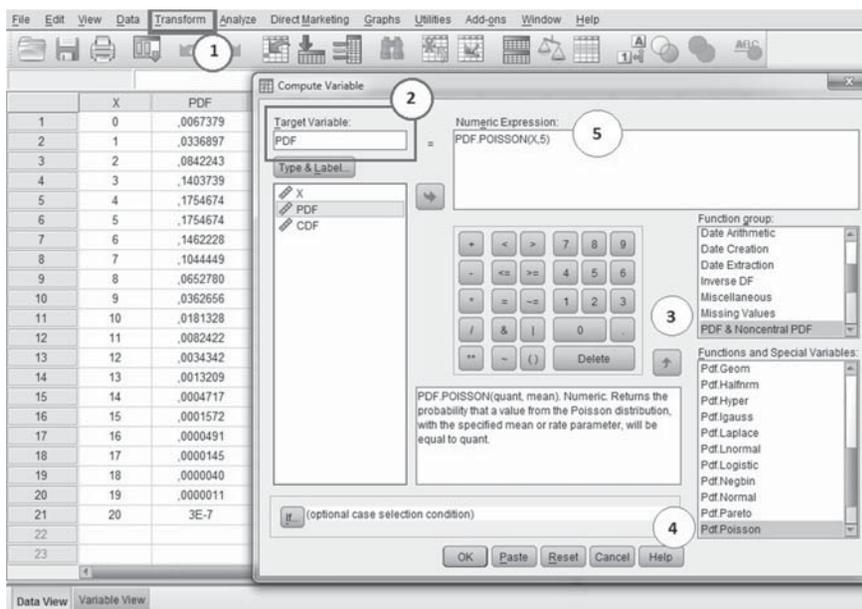


Figura 6 – Introduzir os valores esperados do número de eventos do fenômeno (X), Selecionar em transformar a função calcular variável (1). Nomear a variável para a função de distribuição acumulada (FDA) (2). Selecionar no grupo de funções “FDA & FDA não centrada” (3). Em funções e variáveis especiais, selecionar clicando duas vezes em “Cdf. Poisson” (4). Completar a fórmula substituindo em “(?,?)” primeiro pela variável X seguida pelo valor do parâmetro esperança de ocorrência (λ) (5). Finalizar com “Aceitar” para executar o cálculo

| | X | PDF | CDF |
|----|----|----------|----------|
| 1 | 0 | .0067379 | .0067379 |
| 2 | 1 | .0336897 | .0404277 |
| 3 | 2 | .0842243 | .1246520 |
| 4 | 3 | .1403739 | .2650259 |
| 5 | 4 | .1754674 | .4404933 |
| 6 | 5 | .1754674 | .6159607 |
| 7 | 6 | .1462228 | .7621835 |
| 8 | 7 | .1044449 | .8666283 |
| 9 | 8 | .0652780 | .9319064 |
| 10 | 9 | .0362656 | .9681719 |
| 11 | 10 | .0181328 | .9863047 |
| 12 | 11 | .0082422 | .9945469 |
| 13 | 12 | .0034342 | .9979811 |
| 14 | 13 | .0013209 | .9993020 |
| 15 | 14 | .0004717 | .9997737 |
| 16 | 15 | .0001572 | .9999310 |
| 17 | 16 | .0000491 | .9999801 |
| 18 | 17 | .0000145 | .9999946 |
| 19 | 18 | .0000040 | .9999986 |
| 20 | 19 | .0000011 | .9999997 |
| 21 | 20 | 3E-7 | .9999999 |
| 22 | | | |
| 23 | | | |

Interpretação

A distribuição de Poisson é de muita utilidade quando se deseja conhecer as possibilidades de aparecimento de um número de eventos pouco frequentes e/ou seu impacto no desenvolvimento de uma cirurgia de coluna. Podemos saber a possibilidade de aparecimento de uma quantidade específica de eventos nos quais estamos interessados (exemplo 1), ou o conjunto de possíveis aparecimentos num período de tempo determinado (exemplos 2,3).

Por outro lado, a distribuição de Poisson permite fazer avaliações sobre estratégias para minimizar o aparecimento de eventos não desejados na cirurgia de coluna, como nos dois casos de mau posicionamento de parafusos pediculares (Exemplos 2 e 3). Fazendo estes cálculos, poderíamos prever mudanças nas possibilidades de aparecimento de eventos não desejados no futuro, em períodos de tempo definidos, e assim permitir o controle do número de eventos

neste tipo de cirurgia. O método é aplicável a qualquer evento pouco frequente.

LEITURA COMPLEMENTAR

1. Ahlbom A. *Biostatistic for Epidemiologists*. Florida: Lewis. Publishers, 2010.
2. Everitt B.S., Skrondal A. *The Cambridge Dictionary of Statistics*. 4th Ed. Cambridge, 2010.
3. Glover T, Mitchell K. *An Introduction to Biostatistics*. 3th Ed. Illinois, Waveland Press Ind, 2016.
4. Haight F. *Handbook of the Poisson Distribution*, Wiley, New York, 1967.
5. Kosmopoulos V, Schizas C. Pedicle screw placement accuracy: a meta-analysis . *Spine (Phila Pa 1976)* 2007; 32 : E111-20.
6. Levine D.M., Stephan D.F. *Even You Can Learn Statistics. A Guide for Everyone Who Has Ever Been Afraid of Statistics*. Second Edition. New Jersey, Pearson, 2010.
7. Rosner B. *Fundamentals of Biostatistics*. 7th Ed. Boston, Books/Cole, 2011.
8. Shriver MF, Lewis DJ, Kshetry VR, Rosenbaum BP, Benzel EC, Mroz TE. Pseudoarthrosis rates in anterior cervical discectomy and fusion: a meta-analysis. *Spine J.* 2015; 15(9):2016-27.
9. Van de Kelft E, Costa F, Van der Planken D, Schils F. A prospective multicenter registry on the accuracy of pedicle screw placement in the thoracic, lumbar, and sacral levels with the use of the O-arm imaging system and StealthStation Navigation. *Spine (Phila Pa 1976)*. 2012; 37(25): E1580-7.
10. Walpole R. E., Myers R. H., Myers S. L. *Probability & Statistics for Engineers & Scientists*. 9th Ed. Boston, Pearson, 2012.
11. Wayne WD. *BIOSTATISTICS A Foundation for Analysis in the Health Sciences*. 10th Ed. New York, Willey, 2013.

ODDS RATIO

*José Manuel Pérez Atanasio
Víctor Fernando Luján Celis
Maritza Belén Sandoval Rincón*

Descrição do teste

Odds ratio ou razão de chances ou de possibilidades é um teste estatístico definido como a razão entre a probabilidade de um evento ocorrer e a probabilidade de não ocorrer, expressa o risco de apresentar uma condição ou doença se houver algum fator de risco.

Fórmula

Figura 1 – Fórmula do *odds ratio*

$$OR = \frac{(a \times d)}{(b \times c)}$$

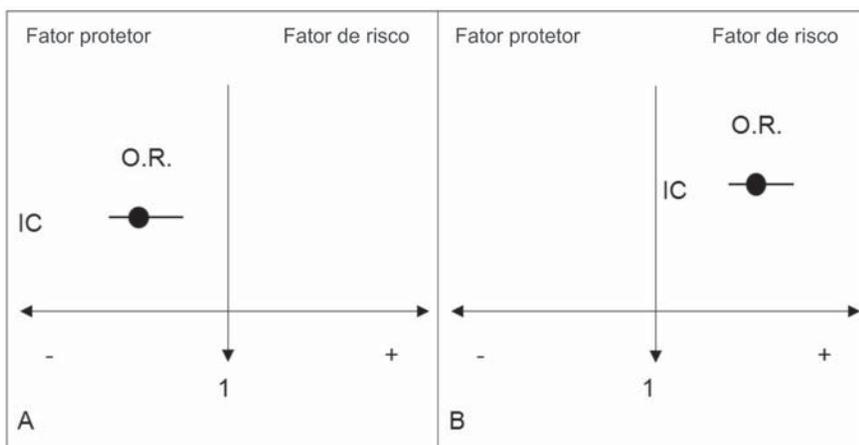
Para obter o *odds ratio*, em primeiro lugar, é preciso elaborar uma tabela de contingência, ou tabela 2x2, que contém os espaços a, b, c, d.

Figura 2 – Tabela de contingência ou de 2 x 2

| | Pacientes que desenvolveram a doença | Pacientes que não desenvolveram a doença |
|--------------------------------|--------------------------------------|--|
| Pacientes com o fator de risco | a | b |
| Pacientes sem o fator de risco | c | d |

O valor obtido do *odds ratio* indica se o fator estudado é um fator protetor (<1) ou um fator de risco (>1) para desenvolver a doença, isto pode ser exemplificado na imagem seguinte:

Figura 3 – Representação gráfica do *odds ratio* como fator protetor (A) e como fator de risco (B)



Exemplo na área de cirurgia de coluna

O Dr. Juan, cirurgião de coluna, trabalha há mais de 20 anos com o Hospital X. Recentemente, observou um aumento das infecções do sítio cirúrgico nos pacientes pós-operados de instrumentação na coluna lombar. Além disso, observou um aumento da obesidade nos pacientes, razão pela qual formula a seguinte pergunta de pesquisa:

A obesidade é um fator de risco para infecção do sítio cirúrgico, em pacientes pós-operados de instrumentação na coluna lombar?

Para responder a esta pergunta, elaborou um protocolo de pesquisa com desenho metodológico de casos e controles; revisou os expedientes de 269 pacientes e identificou aqueles que apresentaram infecção do sítio cirúrgico (casos) e os pacientes que não apresentaram (controles). Além disso, identificou os que tinham obesidade, considerando-o como o fator de risco. Após obter os dados, elaborou a tabela seguinte.

Tabela 1 – Tabela de contingência para calcular o *odds ratio*

| | Pacientes que desenvolveram infecção (casos) | Pacientes que não desenvolveram infecção (controles) | |
|-------------------------|--|--|-----|
| Pacientes com obesidade | a 42 | b 55 | 97 |
| Pacientes sem obesidade | c 28 | d 144 | 172 |
| | 70 | 199 | 269 |

a = pacientes com obesidade que desenvolveram infecção

b = pacientes com obesidade que não desenvolveram infecção

c = pacientes sem obesidade que desenvolveram infecção

d = pacientes sem obesidade que não desenvolveram infecção

Com estes dados, aplicou a fórmula mencionada anteriormente.

$$Odds\ ratio = \frac{(a) 42 \times (d) 144}{(b) 55 \times (c) 28}$$

Para complementar os dados, usou um *software* de análise estatístico para obter o intervalo de confiança de 95% e o valor de *p* obtendo os seguintes resultados:

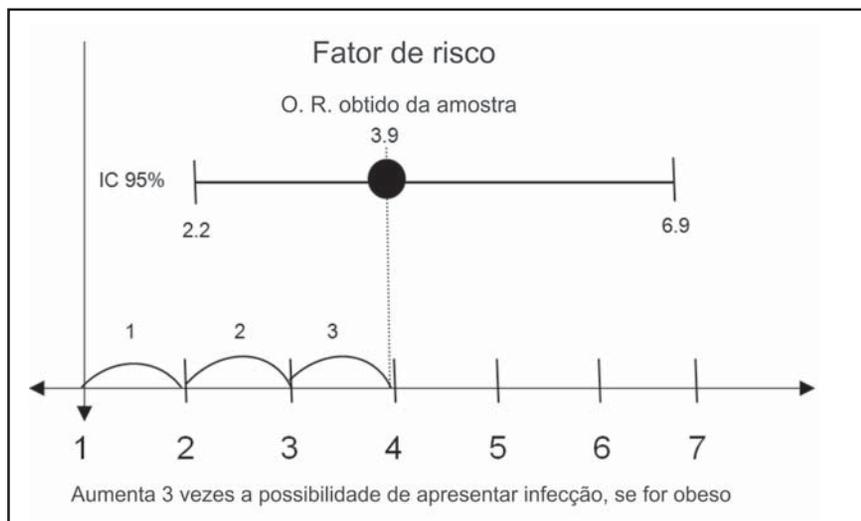
Odds ratio = 3.9

Intervalo de confiança (95%): 2.2 a 6.9

p: <0.0001

Interpretação

O *odds ratio* e seu intervalo de confiança podem ser localizados graficamente sobre uma linha horizontal, que tem como ponto central o número 1, como valor de referência. (Figura 4)

Figura 4 – Representação gráfica do *odds ratio* obtido

Com estes resultados, interpreta-se que os pacientes obesos têm três vezes mais probabilidades de apresentar infecção no sítio cirúrgico, que os que não apresentam obesidade.

O *odds ratio* sempre é expresso em conjunto com seu intervalo de confiança, normalmente de 95%. Este intervalo de confiança mostra, então, que se fossem escolhidos 100 pacientes com obesidade, 95% das vezes teriam um *odds ratio* entre 2.2 e 6.9. Intervalos de confiança com valores próximos do *odds ratio* refletem análise suficiente de indivíduos e, ao contrário, se o resultado mostrar valores afastados, indicará que o número de indivíduos estudado foi muito pequeno.

O valor de *p* orienta-nos sobre a probabilidade de que os dados obtidos são devidos ao acaso. Neste caso, interpretamos que existe uma probabilidade muito baixa (0.001) de que estes dados sejam devidos ao acaso.

LEITURA COMPLEMENTAR

1. Aedo.S.M., Pavlov. S.D, Clavero.F.Ch., Riesgo Relativo y odds ratio ¿Qué son y cómo se interpretan? Rev. Obstétrica Ginecológica: 2010; vol 5: 51-54.
2. Cerda J., Vera C., Rada G., Odds ratio: aspectos teóricos y prácticos, Rev Med Chile 2013; 141: 1329-1335.
3. Clifford B.R., Taylor R.A., Bioestadística, México: Pearson Educación; 2008.
4. Dawson B., Trapp R.G., Bioestadística médica. 4. ed. México: Manual Moderno; 2005.
5. Pérez A. J.M., Principios de bioestadística para comprender los resultados reportados en los artículos científicos, Medigraphic, v. 11 no 2; 2015.
6. Talavera J.O., Rivas.R.R., Pérez M.R., Relevancia Clínica. Rev Med Inst Mex Seguro Soc. 2013; 51 (Supl):S42-S46.
7. Wayne W.D. Bioestadística Base para el análisis de las ciencias de la salud. 4. ed. México: Limusa Wiley; 2011.

Unidade 4

CAPÍTULO 27

O que dá valor e forças às evidências

Mohsen Sadeghi Naini

Vafa Rahimi Movaghar

CAPÍTULO 27

O QUE DÁ VALOR E FORÇA ÀS EVIDÊNCIAS

*Mohsen Sadeghi Naini
Vafa Rahimi Movaghar*

Nível de evidência (NDE ou LoE -*Level of evidence*) foi primeiro usada pela força-tarefa canadense do exame periódico de saúde em 1979, ao realizarem um levantamento na literatura relevante, a fim de identificar 138 condições médicas potencialmente evitáveis (1). Naquele relatório, a eficácia das intervenções médicas foi classificada em três categorias principais, com base na qualidade das evidências (Quadro 1).

Quadro 1 – Nível de evidências, de acordo com a força tarefa canadense

| Eficácia da intervenção | |
|--------------------------------|--|
| I | Evidências obtidas de pelo menos um estudo controlado randomizado |
| II-a | Evidências obtidas de uma coorte bem delineada ou de estudos analíticos de casos e controles, preferivelmente provenientes de mais de um centro ou grupo de pesquisas |
| II-b | Evidências obtidas de comparações entre tempos ou lugares, com ou sem intervenção. Resultados dramáticos em experimentos não controlados (tais como os resultados da introdução da penicilina na década de 1940) também poderiam ser considerados como esse tipo de evidências |
| III | Opiniões de autoridades respeitadas, baseadas em experiência clínica, estudos descritivos, ou relatórios de comitês de especialistas |

Aqui, a qualidade das evidências foi definida em relação a três conceitos principais:

1. Estudos controlados randomizados (ECRs), ou revisões sistemáticas de muitos ECRs, geralmente trazendo um apoio evidenciário maior do que estudos observacionais
2. Estudos clínicos comparativos oferecem apoio evidenciário maior do que o raciocínio “mecanístico” (“fundamentação fisiopatológica”) de ciências mais básicas;
3. Estudos clínicos comparativos oferecem apoio mais forte de evidências do que uma opinião especializada.

As evidências foram então levadas em conta para pontuar recomendações de 1 a 5 (Quadro 2).

Quadro 2 – Pontuação de recomendações pela força-tarefa canadense

| Classificação de recomendações | |
|---------------------------------------|---|
| A | Há boas evidências para apoiar a recomendação de que a condição deve ser especificamente considerada, em um exame de saúde periódico |
| B | Há evidências regulares para apoiar a recomendação de que a condição deve ser especificamente considerada, em um exame de saúde periódico |
| C | Há escassas evidências sobre a inclusão da condição em um exame de saúde periódico, e podem ser feitas recomendações com base em outros motivos |
| D | Há evidências regulares para apoiar a recomendação de que a condição deve ser excluída de consideração em um exame de saúde periódico |
| E | Há boas evidências para apoiar a recomendação de que a condição seja excluída de consideração em um exame de saúde periódico |

Este relatório foi atualizado com pequenas modificações nos anos seguintes (2,3). Desde então, o NDE tem sido amplamente usado e criticado por diferentes grupos de trabalho, para construir o seu próprio esquema (4–6). Cada esquema foi delineado para responder a diferentes perguntas para diferentes interessados, inclusive médicos clínicos, pesquisadores e pessoas que estabelecem as políticas. Entre

os processos evolucionários de NDE, uma melhoria importante no NDE foi feita pelo centro de medicina baseada em evidências *Center for Evidence Based Medicine* (CEBM), em setembro de 2000, que acelerou o processo do encontro de evidências para o médico clínico.

Naquela época, o problema comum de muitos dos esquemas de ordenamento de evidências era o fato de que as evidências eram pontuadas somente para terapia e prevenção, sem enfatizar os testes diagnósticos, marcadores prognósticos ou danos. Portanto, a equipe do CEBM atualizou o NDE que propunha, para que incluísse diferentes aspectos de problemas de saúde, inclusive terapia, prevenção, etiologia, prognóstico, diagnóstico, diagnóstico diferencial e análises econômicas e de decisão (7) (Quadro 3, nas páginas seguintes).

As primeiras hierarquias de evidências (inclusive o relatório da CEBM em 2000) (2,5, 8) foram introduzidas principalmente para avaliar a qualidade das evidências para terapêutica. Essas hierarquias colocavam os ECRs categoricamente acima de estudos observacionais e julgamento especializado abaixo de observações não controladas, que tornavam a classificação de evidências simples e fácil de usar.

Em anos recentes, esta abordagem foi criticada por sua simplicidade (9). Por exemplo, algumas reações adversas a drogas são tão convincentes que um relatório subjetivo bem documentado pode fornecer evidências definitivas de uma associação causal e não se necessita de mais confirmação (10). Além disso, a capacidade de mostrar que pacientes com uma doença, que era antigamente fatal, podem sobreviver após um novo tratamento, estabelece evidências suficientes para a eficácia (11).

Quadro 3 – Nível de evidências proposto pelo CEBM 2009

| Nível | Terapia / Prevenção, Etiologia / Danos | Prognóstico | Diagnóstico | Diagnóstico diferencial/estudo de prevalência de sintomas | Análises econômicas e de decisão |
|-------|--|---|--|---|--|
| 1a | Revisão sistemática (com homogeneidade) de ECRs | Revisão sistemática (com homogeneidade) de estudos de coorte de injeção; "CDR (<i>critical design review</i>)" validada em diferentes populações | Revisão sistemática (com homogeneidade) de estudos diagnósticos de Nível 1; CDR" com estudos 1b de diferentes centros clínicos | Revisão sistemática (com homogeneidade) de estudos de coorte prospectivos | Revisão sistemática (com homogeneidade) de estudos econômicos de Nível 1 |
| 1b | ERC individual (com Intervalo de Confiança estreito) | Estudos de coorte de injeção individuais com um > 80% de seguimento; CDR validado em uma única população. | Validação de estudos de coorte com bons padrões de referência; ou CDR testado em um centro clínico | Estudos de coorte prospectivos com bom seguimento | Análise baseada em custos clinicamente sensíveis ou alternativas; revisões sistemáticas(s) da evidência; inclusive análises de sensibilidade multidirecionais |
| 1c | Todos ou nenhum | Séries de casos com todos ou nenhum | SpPins e SnNouts absolutos | Séries de casos com todos ou nenhum | Análises absolutas de melhor valor ou pior valor |
| 2a | RS (com homogeneidade) de estudos de coorte | RS (com homogeneidade) de ou estudos de coorte retrospectivos ou grupos de controle não tratados em ECRs | RS (com homogeneidade) de estudos diagnósticos de Nível >2 | RS (com homogeneidade) de estudos de 2b e melhores | RS (com homogeneidade) de estudos econômicos de Nível >2 |
| 2b | Estudo de coorte individuais (inclusive ERC de baixa qualidade; p. ex; p. ex., <80% de seguimento) | Estudo de coorte retrospectivo ou seguimento de pacientes de controle não tratados em um ERC; derivação de CDR" ou validado na amostra dividida somente | Estudo de coorte exploratório com bons padrões de referência; CDR" após derivação ou validado apenas com amostra dividida ou bancos de dados | Estudo de coorte retrospectivo, ou seguimento pobre | Análise baseada em custos clinicamente sensíveis ou alternativas; revisões limitadas(s) das evidências ou estudos únicos; e inclusão de análises de sensibilidade multidirecionais |

Continua...

| 2c | Pesquisa de “desfechos”, estudos ecológicos | Pesquisa de “desfechos” | Estudos ecológicos | Auditoria ou desfechos de pesquisa |
|----|--|--|--|--|
| 3a | RS (com homogeneidade) de estudos de casos e controles | RS (com homogeneidade) de estudos 3b e melhor | RS (com homogeneidade) com estudos 3b e melhor | RS (com homogeneidade) de estudos 3b e melhor |
| 3b | Estudo individual de casos e controles | Estudo não consecutivo; ou sem padrões de referência aplicados consistentemente | Estudo de coorte não consecutivo ou população muito limitada | Análise baseada em alternativas limitadas ou custos, estimativas de baixa qualidade dados, mas incluindo análises de sensibilidade incorporando variações clinicamente sensíveis |
| 4 | Séries de casos (e estudos e coorte e casos-controle de baixa qualidade) | Séries de casos (e estudos de coorte de baixa qualidade prognóstica) | Séries de casos ou padrões de referência suplantados | Análise sem análise de sensibilidade |
| 5 | Opinião especializada sem avaliação crítica explícita, ou baseada em fisiologia, pesquisa de bancada ou “primeiros princípios” | Opinião especializada sem avaliação crítica explícita, ou baseada em fisiologia, pesquisa de bancada ou “primeiros princípios” | Opinião especializada sem avaliação crítica explícita, ou baseada em fisiologia, pesquisa de bancada ou “primeiros princípios” | Opinião especializada sem avaliação crítica explícita, ou baseada em fisiologia, pesquisa de bancada ou “primeiros princípios” |

Alguns exemplos históricos de tratamentos com efeitos dramáticos, que vêm de estudos observacionais, incluem:

- insulina para diabetes;
- transfusão de sangue para choque hemorrágico grave;
- desfibrilação para fibrilação ventricular;
- redução fechada e colocação de talas para fratura de ossos longos com deslocamento;
- neostigmina para miastenia grave;
- traqueostomia para obstrução da traqueia;
- suturas para a reparação de grandes ferimentos;
- estreptomicina para meningite tuberculosa;
- drenagem para dor associada com abscessos;
- pressão ou sutura para interromper uma hemorragia;
- éter para anestesia;
- válvula unidirecional ou drenagem de vedação subaquática para pneumotórax e hemotórax.

Por outro lado, ninguém pode saltar de um avião com e sem paraquedas para alcançar evidências de nível 1, a respeito do papel exercido pelo paraquedas (12). Embora a atual opinião seja a favor de considerar todos os tipos de evidências disponíveis (inclusive estudos observacionais), ao realizar recomendações, também deve ser lembrado que já que prognóstico e o efeito de tratamento interagem como ruído e sinal, nem todos os efeitos de tratamento são tão fáceis de demonstrar (13).

Considerando estas limitações, esforços mais recentes evitam a intenção de fazer um *upgrade* de evidências simplesmente através do delineamento do estudo (ensaio clínico *versus* estudos observacionais, ao permitir que seja aumentado o NDE de estudos observacionais com efeitos dramáticos (4). Essa mudança na interpretação da qualidade das evidências foi outra evolução no NDE.

Em 2006, o Grupo de Trabalho de Pontuação de Recomendações, Avaliação e Desenvolvimento (GRADE - *Grading of Recommendations, Assessment, Development and Evaluations*) propôs uma abordagem clara e sistemática à realização de julgamentos sobre a qualidade de evidências e força de recomendações. GRADE classifica a qualidade de um corpo de evidências em quatro níveis: alta qualidade, qualidade moderada, baixa qualidade e qualidade muito baixa.

A qualidade de evidências na abordagem GRADE pode mudar quanto ao delineamento do estudo e risco de viés, inconsistência dos resultados, fato de ser indireto (não generalizável), imprecisão (dados esparsos), e outro fatores (por exemplo viés de relatório (14). Esta abordagem dentro de pouquíssimo tempo torna-se popular por uma revisão sistemática, e agora é vista de modo generalizado como o método mais efetivo de ligar avaliações de qualidade de evidências a recomendações clínicas (4). A abordagem GRADE avalia as evidências, não estudo por estudo, mas transversalmente aos estudos para desfechos clínicos específicos. Esta abordagem especificamente avalia quatro domínios nos estudos incluídos das revisões sistemáticas, avaliações e diretrizes de tecnologia da saúde inclusive (4):

- falhas metodológicas dentro dos estudos componentes;
- consistência de resultados transversalmente a diferentes estudos;
- generalizabilidade de resultados de pesquisa à base de pacientes mais ampla;
- eficácia apresentada pelos tratamentos.

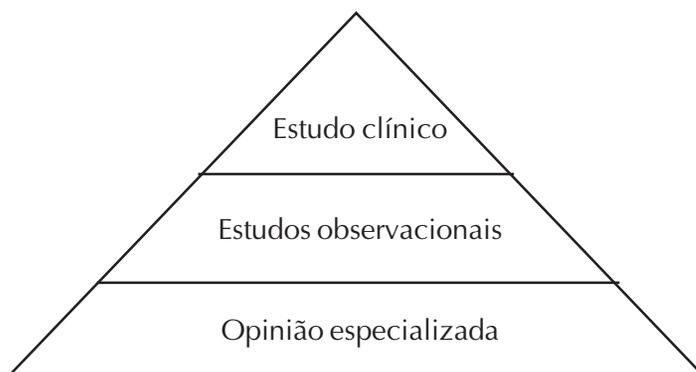
Contudo, o que GRADE acrescentou, em termos de precisão, pode ter perdido em simplicidade. Portanto, poderá ser impossível de usar por clínicos atarefados ao buscar as melhores evidências disponíveis para um problema clínico específico (7). Um conflito importante entre GRADE e OCEBM é que a abordagem GRADE avalia a força das evidências, em vez de estudos isolados, mas os Níveis de OCEBM enfocam o seu público específico (clínicos ou pacientes), e tornam sua ferramenta adequada para avaliar estudos individuais, caso não haja uma revisão sistemática disponível.

NDE

Níveis de evidência explicam a ideia de que há graus ou níveis variáveis de evidências para um relacionamento específico de “causa e efeito”, em determinada era científica, em que diferentes tipos de delineamentos de estudos podem ser proporcionados). Em outras palavras, ao falar sobre NDE, enfocamos a classificação do delineamento metodológico de estudos, que avaliam os efeitos de intervenções no atendimento à saúde. No mundo real, é raro que um único estudo demonstre uma hipótese de maneira tão irrefutável e que fique impossível de questioná-la. Geralmente, diferentes estudos e diferentes níveis e/ou tipos de evidências em última análise, constroem o corpo de evidências.

Tradicionalmente, o NDE é apresentado em uma hierarquia de peso com referência à força da associação entre a causa e efeito incorporados por aquele estudo especificamente (Figura 1).

Figura 1 – Hierarquia do nível de evidências



Estudos com animais constituem o primeiro passo para proporcionar evidências médicas que irão gerar e testar hipóteses. Isso também proporcionará a infraestrutura para desenvolver um raciocínio mecanístico para as regras comumente aceitas.

Opiniões especializadas sem avaliação crítica ou baseadas em raciocínio clínico, pesquisas de bancada ou “primeiros princípios” fornecem o segundo nível de evidências.

Estudos não controlados são divididos em observacionais (por exemplo, um relatório de caso e séries de casos) e grupos experimentais (por exemplo, estudos de antes/depois e estudo randomizado N-de-1), conforme o tipo de intervenção.

Estudos observacionais com grupos de controle: estes incluem estudos de coorte, estudos com casos e controles e estudos transversais, onde o investigador está descrevendo o que está ocorrendo naturalmente, sem perturbar o cenário natural. Este tipo de estudo é mais comumente usado para anotar associações entre eventos e gerar hipóteses.

Estudos experimentais com grupos de controle mais comumente conhecidos como estudo (ou ensaio) clínico (inclusive randomizado ou não randomizado) intervêm ativamente ao designar alguns participantes para receber um tratamento e outros para receber uma intervenção diferente (ou nenhuma intervenção). Obviamente, estudos

experimentais agregaram um valor maior ao processo de construir o corpo de evidências.

Meta-análises e revisões sistemáticas, por outro lado, resumem grupos de estudos em uma questão de pesquisa específica, para avaliar o peso das evidências sobre aquela questão (Quadro 4).

Quadro 4 – Taxonomia do delineamento de estudos que avaliam os efeitos de intervenções de atendimento à saúde

| | |
|--|--|
| Revisão sistemática | <ul style="list-style-type: none"> • Revisão sistemática com meta-análise • Revisão sistemática sem meta-análise |
| Estudos experimentais com grupo de controle (“estudos clínicos” ou “estudos”): O pesquisador controla a decisão relativa à alocação de participantes aos diferentes grupos de intervenção | <ul style="list-style-type: none"> • Estudo Controlado Randomizado (ECR) • Estudo controlado quase randomizado (ECR-Q) • Estudo clínico controlado (ECC) |
| Estudos observacionais com grupo de controle: A intenção do pesquisador é observar e não interferir nos cuidados rotineiros | <ul style="list-style-type: none"> • Estudo de coorte • Caso de estudo de coorte de sobrevivência • Estudo de casos e controles • Estudo transversal |
| Estudos não controlados (sem um grupo de controle separado): podem ser de natureza experimental ou observacional | <ul style="list-style-type: none"> • Estudos de antes/depois • Estudo randomizado N-de-1 • Série de casos • Relatórios de caso |
| Opinião especializada | |
| Estudos com animais | |

Além de vários pontos positivos sob o aspecto hierárquico de OCEBM–NDE, faz algumas suposições importantes que não são geralmente válidas (p.ex.: em intervenções complexas, mas que são raramente debatidas (15):

- **Equilíbrio (Equipose):** paciente e clínico não têm preferência por um tratamento;
- **Falta de conhecimentos:** verdadeiramente não se sabe qual das duas escolhas é “melhor” e há evidências inadequadas sobre os efeitos de tratamento provenientes de outras fontes;
- **Preferência por especificidade:** apenas efeitos específicos atribuíveis à intervenção são terapeuticamente válidos;
- **Independência de contexto:** há uma extensão “verdadeira” de eficácia, ou um efeito de tamanho constante independente do contexto;
- **Validade ecológica e externa:** conhecimentos sobre um efeito terapêutico extraído de um ECR são prontamente atribuíveis à prática clínica, se os critérios de elegibilidade do estudo estão de acordo com as características de determinado paciente.

A alternativa à visão hierárquica do nível das evidências é circular. Deriva da experiência de que todo método de pesquisas tem pontos fortes e limitações que não podem ser fixadas por aquele método em si. Portanto, triangular um resultado alcançado com um método por meio da repetição do mesmo com outros métodos pode levar a uma conclusão potente (15).

Tipos de viés e ferramentas para avaliar o viés

A força das evidências pode ser violada por vieses (ou erro sistemático) e fatores de confusão.

Os vieses podem desviar o resultado em qualquer uma de duas direções (subestimativa ou sobre estimativa do verdadeiro efeito da intervenção), com diferenças de grandeza. É geralmente impossível medir o impacto de vieses sobre os resultados de determinado estudo (16).

O viés pode ser confundido com imprecisão. Viés refere-se a um erro sistemático que não muda por meio de replicações múltiplas da mesma metodologia. A imprecisão se eleva a um erro aleatório, o que

significa uma variação de amostragem e pode levar a oscilações dos resultados em múltiplas replicações da mesma metodologia, mesmo se todas essas replicações dessem a resposta certa em média. Podemos tratar da imprecisão aumentando o tamanho da amostra, mas os vieses não têm tal efeito. A imprecisão é espelhada no intervalo de confiança em torno do resultado analisado. Assim, resultados com maior precisão levam a um intervalo de confiança mais estreito e vice-versa.

Os vieses são categorizados segundo cinco classes principais (Quadro 5).

Quadro 5 – O esquema de classificação de Cochrane para vies

| Tipo de vies | Descrição | Domínios relevantes no instrumento Cochrane ‘Risco de vies’ |
|---------------------|--|---|
| Viés de seleção | Diferenças sistemáticas entre características dos grupos que são comparados | <ul style="list-style-type: none"> • Geração de seqüências • Ocultamento de alocação |
| Viés de desempenho | Diferenças sistemáticas entre grupos quanto aos cuidados proporcionados ou na exposição a fatores outros do que as intervenções de interesse | <ul style="list-style-type: none"> • Cegamento dos participantes e do pessoal. • Outras ameaças potenciais à validade |
| Viés de detecção | Diferenças sistemáticas entre grupos na maneira pela qual são determinados os desfechos | <ul style="list-style-type: none"> • Cegamento de avaliação de desfecho • Outras ameaças potenciais à validade |
| Viés de atrição | Diferenças sistemáticas entre grupos em retiradas de um estudo | <ul style="list-style-type: none"> • Dados de desfecho incompletos |
| Viés de relatório | Diferenças sistemáticas entre achados relatados e não relatados | <ul style="list-style-type: none"> • Relatório seletivo de desfechos |

Até agora três tipos de ferramentas são propostos para avaliar vies na pesquisa clínica, especificamente estudos clínicos: escalas, lista de verificação e avaliações baseadas em domínios. Moher e colegas identificaram 34 ferramentas que haviam sido usadas para avaliar vieses em estudos randomizados (17). A ferramenta recomendada pela Colaboração Cochrane, para avaliar o risco de vies, é uma avaliação baseada em domínios, na qual as avaliações críticas são separadas para diferentes domínios (16). É difícil demonstrar a extensão do vies

em dado estudo, portanto a possibilidade de validar qualquer ferramenta proposta é limitada. (Quadro 6).

Quadro 6 – A ferramenta da colaboração Cochrane para avaliar o risco de viés

| Domínio | Suporte para julgamento | Revisão do julgamento dos autores |
|--|--|--|
| <i>Viés de seleção</i> | | |
| Geração de sequência aleatória | Descreva o método usado para gerar a sequência de alocação com suficientes detalhes para permitir uma avaliação mostrando se devem ser produzidos grupos comparáveis | Viés de seleção (alocação enviesada às intervenções) devido à geração inadequada de uma sequência randomizada |
| Ocultação de alocação | Descreva o método usado para ocultar a sequência de alocação com detalhes suficientes para determinar se alocações de intervenção poderiam ter sido previstas antes de ou durante o recrutamento | Viés de seleção (alocação enviesada às intervenções) devido à ocultação inadequada de alocações antes de selecionar os participantes do estudo |
| <i>Viés de desempenho</i> | | |
| Cegamento dos participantes e de pessoal. Devem ser feitas avaliações para cada desfecho principal (ou classe de desfechos) | Descreva todas as medidas usada, se houver, para cegar os participantes e pessoal envolvido no estudo a respeito do conhecimento da intervenção aplicada ao participante. Forneça todas as informações que informam se o cegamento objetivado foi eficaz | Viés de desempenho devido ao conhecimento de intervenções alocadas por participantes e pessoal durante o estudo |
| <i>Viés de detecção</i> | | |
| Cegamento da avaliação de desfecho. Devem ser feitas avaliações para cada desfecho principal (ou classe de desfechos) | Descreva todas as medidas usadas, se houver, para cegar os avaliadores do desfecho ao conhecimento de que intervenção foi recebida por um participante. Forneça todas as informações que informam se o cegamento objetivado foi eficaz | Viés de detecção devido ao conhecimento das intervenções alocadas por avaliadores de desfecho |
| <i>Viés de atribuição</i> | | |
| Dados incompletos sobre desfechos. Devem ser feitas avaliações para cada desfecho principal (ou classes de desfechos) | Descreva quão completos são os dados de desfechos para cada desfecho principal, inclusive atribuição e exclusões da análise. Diga se a atribuição e as exclusões foram relatadas, números em cada grupo de intervenção (comparado com número total de participantes randomizados), motivos pelos quais foram relatadas atribuições/exclusões, e qualquer reinclusões nas análises realizadas pelos autores de revisões | Viés de atribuição devido à quantidade, natureza ou manuseio de dados incompletos sobre desfechos |
| <i>Viés de relatório</i> | | |
| Relatório seletivo | Diga como foi examinada a possibilidade de relatório de desfecho seletivo pelos autores de revisões e o que foi encontrado | Viés de relatório devido a relatório seletivo de desfecho |
| <i>Outros vieses</i> | | |
| Outras fontes de viés | Diga quaisquer preocupações importantes sobre viés não discutidas nos outros domínios da ferramenta. Se determinadas perguntas/ anotações foram pre-especificadas no protocolo da revisão, devem ser fornecidas respostas para cada pergunta/anotação | Viés devido a problemas não tratados em outra parte da tabela |

O que pode violar a força das evidências?

Há fatores em potencial que podem diminuir a qualidade do corpo de evidências derivadas da análise de dados reunidos (*pooled*) de um ECR, conforme proposto na abordagem Quadro (4).

Esses pontos avaliam diferentes tipos de viés neste delineamento específico de estudo e então julgam com base na força global de evidências derivadas de estudos clínicos incluídos. Portanto, essa abordagem é descrita para ECR, como o bloco fundamental do processo de produção de evidências.

1. Limitações de delineamento e de implementação: cada delineamento de estudo tem algumas limitações potenciais no delineamento que deveriam ser abordadas ao incorporar seu resultado a um ambiente clínico ou pergunta de pesquisa. Para estudos randomizados, as questões potenciais a serem abordadas incluem:

- ocultação de alocação;
- cegamento (especialmente quando os desfechos são avaliados subjetivamente);
- perda do seguimento;
- parar cedo para benefício ou relatos seletivos dos desfechos.

2. Evidências indiretas: ao usar os resultados do estudo, os critérios de elegibilidade preenchidos naquele estudo devem ser avaliados e comparados ao atual *setting* clínico. Por exemplo, uma versão restrita de pergunta principal em termos de população, intervenção, comparador ou desfechos pode produzir evidências indiretas. Outras fontes de informações indiretas podem surgir de intervenções estudadas (por exemplo, se num estudo relevante uma intervenção cirúrgica foi implementada por um cirurgião sênior, em centros especializados, então as evidências sobre os efeitos da intervenção fora desses centros poderá ser indireta); comparadores usados (por exemplo, se a intervenção nos grupos de controle é menos eficaz do que o tratamento padrão), e desfechos avaliados (por exemplo indireto, devido a desfechos alternativos).

3. Heterogeneidade inexplicada ou inconsistência de resultados: estimativas com grandes diferenças sobre o efeito (heterogeneidade, em outras palavras), sem explicação razoável, poderiam diminuir a qualidade das evidências.

4. Imprecisão de resultados: intervalos de confiança largos, independentes da heterogeneidade, podem levar à diminuição da qualidade das evidências. Isso pode ser tratado aumentando o tamanho da amostra.

5. Alta probabilidade de viés de publicação: a qualidade do nível de evidências pode ser rebaixada (*downgraded*), se os investigadores não relatarem estudos (classicamente aqueles que não apresentam algum efeito: viés de publicação), ou desfechos (classicamente aqueles que podem ser prejudiciais ou para os quais não foi relatado nenhum efeito: viés de relatório de desfecho seletivo), como base em resultados. A grande influência de patrocinadores sobre um estudo traz à baila perguntas sobre se existem estudos não publicados que não sugerem nenhum benefício.

Agradecimento

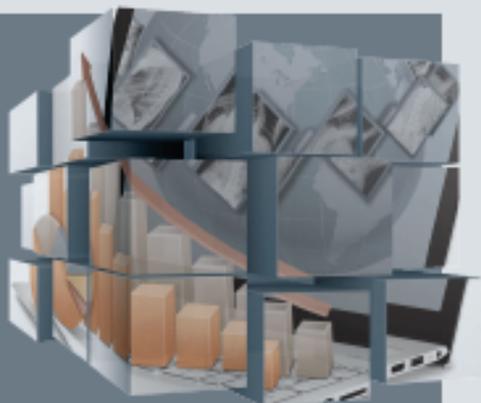
Agradeço o trabalho de editoração da língua inglesa realizado pela Sra. Bitá Pourmand, afiliada ao *Research Development Center, Sina Hospital, Tehran University of Medical Sciences*.

LEITURA COMPLEMENTAR

1. Hill N, Frappier-Davignon L, Morrison B. The periodic health examination. *Can Med Assoc J.* 1979;121:1193-254.
2. Force CT, Examination PH. The periodic health examination 2. 1984 update. 1984;130.
3. Examination CTF on the PH. The periodic health examination: 2. 1987 update. *TT – C Can Med Assoc J.* 1988;138(7):618-26.
4. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. Vol. 9, *Chinese Journal of Evidence-Based Medicine.* 2009. p. 8-11.
5. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest. American College of Chest Physicians;* 1986;89(2):25-35.
6. Judith Welsh, BSN M. Levels of evidence and analyzing the literature. National Institutes of Health Library.
7. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, et al. Explanation of the 2011 Oxford Center for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document). *Oxford Cent Evidence-Based Med.* 2011;(12):1-8.
8. Force USPST. Guide to clinical preventive services. DIANE Publishing; 1989.
9. Howick J. The Philosophy of Evidence-Based Medicine. *The Philosophy of Evidence-Based Medicine.* 2011.
10. Aronson JK, Hauben M. Anecdotes that provide definitive evidence. *BMJ Br Med J.* 2006;333(7581):1267-9.
11. How to read clinical journals: V: To distinguish useful from useless or even harmful therapy. *Can Med Assoc J.* 1981;124(9):1156-62.
12. Lee MJ, Norvell DC, Dettori JR. *SMART Approach to Spine Clinical Research.* Thieme; 2013.
13. Glasziou P. When are randomised trials unnecessary/ ? Picking signal from noise Paul Glasziou and colleagues When are randomised trials unnecessary/ ? Picking signal from noise. 2016;(May).
14. Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P GG. Interpreting results and drawing conclusions. In: Higgins JPT GS, editor. *Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011).* The Cochrane Collaboration; 2011.
15. Walach H, Falkenberg T, Fønnebø V, Lewith G, Jonas WB. Circular instead of hierarchical/ : methodological principles for the evaluation of complex interventions. 2006;9:1-9.

16. Higgins JPT, Altman DG SJ (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT GS, editor. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.10 (updated March 2011) [Internet]. The Cochrane Collaboration; 2011. Available from: Available from <www.cochrane-handbook.org>.
17. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*. Elsevier; 1995;16(1):62-73.

Quando temos uma ideia clara da pergunta que queremos fazer e que nos leva a realizar nossa investigação, teremos resultados que nos permitirão fortalecer nossas conclusões mediante inferência. Para isso devemos conhecer os aspectos teóricos relacionados à Bioestatística Clínica, de onde derivaremos conceitos que nos permitirão tomar decisões e aplicar o conhecimento obtido nos nossos pacientes e cumprir com excelência o papel de prevenir e tratar enfermidades, incorporando os conceitos à prática clínica.



Da Medição à Interpretação, um processo essencial que vincula a investigação ao mundo cotidiano e à realidade, em que nossas hipóteses, rodeadas de probabilidades, buscam encontrar o valor que estabeleça se existem ou não diferenças entre as opções.

Não devemos esquecer que os números são crus e não entendem nada sobre razões, e é aí que o nosso conhecimento e a teoria darão valor real aos dados, já que esses não nos dão a certeza de qualquer evento, simplesmente nos indicam a probabilidade de que algo possa acontecer.



José María Jiménez Avila
Regional Research Officer—AOSpine Latin America



Asdrubal Falavigna
Chairperson—AOSpine Latin America

ISBN 978-85-7061-910-5



9 788570 619105