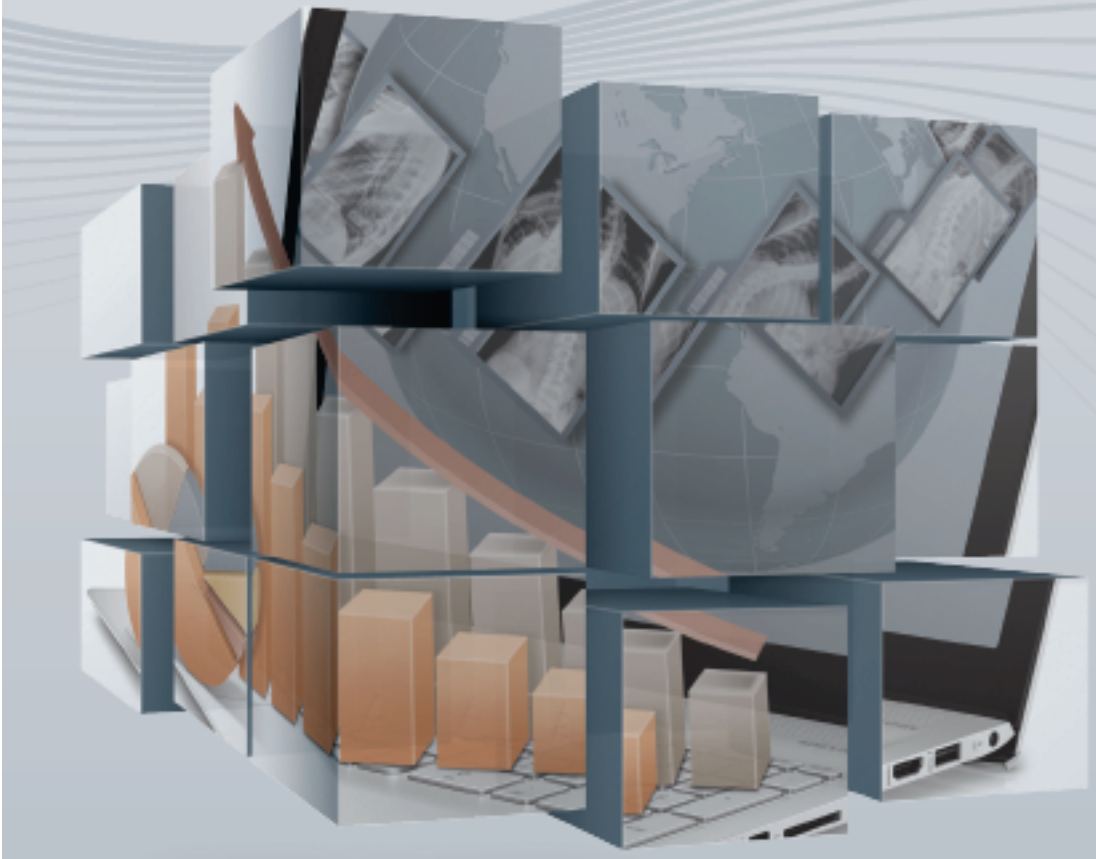
AOSPINE
LATIN AMERICA

Education in Research:
FROM MEASUREMENT TO INTERPRETATION– SURGICAL BIOSTATISTIC

Organizers
José María Jiménez Avila
Asdrubal Falavigna



EDUCS

Education in Research:

FROM MEASUREMENT TO INTERPRETATION– SURGICAL BIOSTATISTIC

José María Jiménez Avila
Asdrubal Falavigna
Organizers

Education in Research:
FROM MEASUREMENT TO INTERPRETATION– SURGICAL BIOSTATISTIC

AOSPINE

EDUCS
COMMUNITY

Cover: André Secundino
Review: Carla Vinize Casagrande Gianini

Publishing: Traço Diferencial | + 55 (54) 3229 7740 | (54) 99901 3978

Cataloguing in Publication Data (CIP)
Caxias do Sul University
UCS – BICE – Tecnical Processing

E24 Education in research : from measurement to interpretation-surgical biostatistic / org. José María Jiménez Avila, Asdrubal Falavigna . – Caxias do Sul, RS : Educs, 2018.
204 p.: il.; 23 cm.

ISBN: 978-85-7061-912-9
Presenta bibliografia
Varios colaboradores

1. Biostatistics. 2. Surgery – Biosyatistics. 3. Medicine – Research. I. Avila, José María Jiménez. II. Falavigna, Asdrubal.

CDU 2. ed.: 57.087.1

Index for the systematic catalogue:

1. Biostatistics	57.087:1
2. Surgery – Biosyatistics	617:57.087.1
3. Medicine – Research	61:001.8

Cataloging at source by:
Ana Guimarães Pereira – CRB 10/1460

ASSOCIATION AOSPINE LATIN AMERICA
Av. Silva Jardim 2042 Cj. 1505 | CEP 80250-200 | Curitiba – PR | Brazil
Phone: + 55 41 3016 4491 | *E-mail:* aosla@aospine.org

SUMMARY

Acknowledgments / 9
Note organizers / 11
Collaborators / 13
Foreword / 17

UNIT 1 / 19

Chapter 1
INTRODUCTION / 21
José María Jiménez Avila e *Asdrubal Falavigna*

Chapter 2
MATHEMATICS: LANGUAGE AND INTERPRETATION / 23
Maurício Scussel

Chapter 3
THE MAIN ACTOR: THE VARIABLE / 29
José María Jiménez Avila

Chapter 4
HOW DOES ONE PLAN A DATABASE? / 33
Pedro Luis Bazán

Chapter 5
SAMPLE SIZE CALCULATION / 45
Emiliano Vialle

Chapter 6
HOW TO CHOOSE STATISTICAL TESTS / 53
Pedro Luis Bazán

Chapter 7
DESCRIPTIVE BIOSTATISTICS / 63
Alberto Aceves Pérez

Chapter 8
ANALYTIC OR INFERENTIAL BIOSTATISTICS / 67
José María Jiménez Avila

Chapter 9
CENTRAL TENDENCY MEASURES / 73
Alberto Aceves Pérez

Chapter 10
MEASURE for dispersion (Range, Variance, Standard Deviation, Coefficient of Variance) **/ 79**
Delio Martins

Chapter 11
DATA PRESENTATION / 85
José María Jiménez Avila

Chapter 12
HYPOTHESIS TEST (CI) AND VALUE OF "p" / 91
José María Jiménez Avila

Chapter 13
KOLMOGOROV-SMIRNOV TEST / 97
Nelson Astur Neto
Maximiliano Gimenez Gigón
Rodrigo Góes Medéa de Mendonça

Chapter 14
SHAPIRO-WILK TEST / 101
Nelson Astur Neto
Rodrigo Góes Medéa de Mendonça

UNIT 2 / 105

Chapter 15
STUDENT *t* TEST / 107
Barón Zarate
Francisco López Meléndez

Chapter 16
ANOVA TEST / 117
Juan Pablo Guyot

Chapter 17
PEARSON CORRELATION / 121
Alisson R. Teles
Asdrubal Falavigna

UNIT 3 / 127

Chapter 18
CHI² / 129
Alberto Aceves Pérez

Chapter 19
WILCOXON RANK SUM TEST / 133
Baron Zarate

Chapter 20
MANN-WHITNEY U TEST / 143
José Manuel Pérez Atanasio
Victor Fernando Luján Celis
Maritza Belén Sandoval Rincón

Chapter 21
KRUSKAL-WALLIS TEST / 147
Esteban Araya Ramírez
Fabián Víquez Monge

Chapter 22
SPEARMAN CORRELATION COEFFICIENT / 151
Fernando Alvarado
Miguel Farfán

Chapter 23
MC NEMAR´S TEST / 157
Esteban Araya Ramírez
Fabián Víquez Monge

Chapter 24
SIMPLE LINEAR REGRESSION / 163
Alisson R. Teles
Asdrubal Falavigna

Chapter 25
POISSON DISTRIBUTION / 169
José Acuña Pinilla
Margarita Gómez Chantraine
Elizabeth Heins

Chapter 26
ODDS RATIO / 181
José Manuel Pérez Atanasio
Víctor Fernando Luján Celis
Maritza Belén Sandoval Rincón

UNIT 4 / 187

Chapter 27
WHAT GIVES VALUE AND FORCE TO THE EVIDENCE / 189
Mohsen Sadeghi Naini
Vafa Rahimi Movaghar

ACKNOWLEDGMENTS

We thank the authors for their invaluable contribution in writing the chapters, and congratulate them for the quality of the didactic presentation simplifying the complex world of biostatistics.

Our thanks to the AOSpine International Research Commission Members for their unconditional support to the elaboration of the book.

We also thank the members of the Board of AOSpine Latinoamerica for their support and the office for logistic support to finalize this great project.

We are very pleased to publish this second book, since we are certain that it will contribute to daily strengthening the knowledge of clinical and experimental research.

José María Jiménez Avila

Again, my thanks to Georgina, Zaira and Daniela for having shared hours and sometimes days of our time together with this book. They have always been my motivation and my reason for continuing to work with this type of project.

Very special acknowledgments and thanks to Dr. Asdrubal Falavigna, the current Chairperson of AOSpine Latinoamerica, who always placed his trust in the project and makes it possible thanks to his vision and leadership.

Asdrubal Falavigna

I dedicate this book to my wife, Ana Maria, and to my children Vincenzo and Rebecca, for their daily support. I hope that this book will further the scientific spirit of every student and professional, aiming to give them the ability to innovate and generate knowledge in order to seek answers to our everyday concerns.

NOTE ORGANIZERS

There is a repressed demand for further training in research, whether it be for developing reflexive thinking or for the critical analysis of facts. It is essential to generate new knowledge and products of innovation that will potentiate the advances of science and allow them to be applied to the development of different dimensions of society.

Learning biostatistics is usually a long and difficult process. This book focuses on the basic principles of statistics applied to specialists in the field of health from the point of view of specialists and researchers in different fields of knowledge. In this process the researcher will find suggestions and examples that will make it easier to understand this complex field, including problems in understanding the meaning of $p<0.05$, when to choose one statistical method over another, and so on. These are the obstacles to be overcome in order to be a critical and reflexive professional who knows how to intervene in reality and produce specific knowledge of the specific field, so as to respond to the demands of the population.

The chapters enable the reader to acquire knowledge regarding the difficulties commonly encountered in the field of biostatistics. Ultimately, nothing is more gratifying than seeing one's work published after several efforts, discipline and dedication.

José María Jiménez Avila
Asdrubal Falavigna

COLLABORATORS

Alberto Aceves Pérez
Orthopedist. Centro Médico Nacional de Occidente. Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, Mexico. Professor at Universidad de Guadalajara. AOSpine Research Officer Mexico.

Alisson R Teles
Spine Neurosurgeon. McGill Scoliosis & Spine Group. McGill University, Montreal, Canada.

Asdrubal Falavigna
Chairperson of AOSpine Latin America. Director of the Life Science Center at UCS. Coordinator of Health Science Post-Graduation Program. Professor of Neurosurgery at Universidade de Caxias do Sul. Coordinator of the Research Group on Clinical Studies and Basic Models of Spinal Cord Pathologies and of the Laboratório de Patologias Raquimedulares at UCS. Caxias do Sul, RS – Brazil.

Barón Zárate Kalfópulos
Spine Surgeon. Instituto Nacional de Rehabilitación / Medica Sur Ciudad de México. Professor of Highly Specialized Spine Surgery at Universidad Nacional Autónoma de México. Director of Centro AOSpine Mexico City.

Delio Martins
Orthopedist. Israelita Albert Einstein Hospital . Doctor of Science from Federal University of Sao Paulo (UNIFESP); Coordinator of the Specialization course in Spine Surgery at Israelita Albert Einstein Hospital; Professor of Orthopedics at Anhembi-Morumbi University.

Elizabeth Heins
Surgeon at Pontificia Universidad Javeriana. Epidemiólogy- Johannes Gutenberg Universität Mainz. Senior Manager of Medical Affairs Johnson & Johnson Medical Devices Northern Cluster.

Emiliano Vialle
Head of the spine surgery group, Universitário Cajuru Hospital, Human Tissue Bank, Pontifícia Universidade Católica do Paraná, AOSpine Fellowship Officer for Latin America. Brazil.

Esteban Araya
Orthopedist, University of San Carlos, Guatemala. Column Surgeon, Nueva Granada Military University, Colombia. Professor at the Department of Surgery, University of Medical Sciences, Costa Rica. Coordinator at the Clinic of Hospital Trauma Hospital and Medical Coordinator Tissue Bank Trauma Hospital, Former Head of Surgical Services Trauma Hospital, Costa Rica.

Fabián Viquez Monge
Master's Degree in Insurance and Risk Management, from the Pontifical University of Salamanca, Spain. Degree in Risk Associated with Operation, University of Costa Rica. Bachelor in Statistics, University of Costa Rica.

Fernando Alvarado Gomez
Column Surgeon - University Hospital Fundación Santa Fe de Bogotá. Column Surgeon - Roosevelt Children's Orthopedic Institute. Surgery Column, Pelvis and Acetabulum - Military University Nueva Granada. Orthopedics and Traumatology - Pontificia Universidad Javeriana. Clinical Epidemiologist - University of Nuestra Señora del Rosario. Clinical professor of Orthopedics and Traumatology of the Universidad de los Andes. Associate Professor of Orthopedics and Traumatology at the University of Nuestra Señora del Rosario. Ad Honorem professor of the column surgery program of the Nueva Granada Military University. Research AOSpine Colombia. Secretary of Column Chapter SCCOT.

Francisco Manuel López Melendez
Spine Surgeon. ISSSTE General Santiago Ramón y Cajal Hospital. Durango. Mexico.

José Acuña Pinilla
Professional Surgical Instrument, University of Santander UDES. Coordinator of Evidence-Based Medicine and Medical Affairs for the North Region of Latin America One Medical Devices division of the Johnson & Johnson company. Professor at the University of Santander - UDES, Bogotá.

José Manuel Pérez Atanasio
Orthopedist, MSc, Head of the Health Research Division UMAE "Dr. Victorio de la Fuente Narváez" IMSS. Mexico City.

José María Jiménez Avila
Orthopedist, Instituto Tecnológico y de Estudios Superiores de Monterrey (TEC), Guadalajara campus, Professor in Departamento de Ciencias Básicas and Coordinator of Orthopedics and Traumatology, Instituto Mexicano del Seguro Social, Clínica de Columna, Regional Research Officer of AOSpine Latinoamérica. México.

Juan Pablo Guyot
Spine, Orthopedic and Trauma Surgery, Fundación Favaloro University Hospital. Argentina.

Maritza Belén Sandoval Rincón
Psychiatrist, Specialist in Sleep Disorders, Clínica de Trastornos del Movimiento y Sueño. Hospital General Ajusco Medio. Mexico City.

Maurício Scussel
Graduated as Control and Automation Engineer from Pontifícia, Universidade Católica do Paraná (PUCPR), Doctorate in Neurology/ Neurosciences from Universidade Federal de São Paulo and Centro Nacional de Microeletrônica de Barcelona (UNIFESP/IMB-CNM), Post Doctorate at the Institute of Physiology and Pathophysiology at the University of Heidelberg -Germany, Director of Simplifica – Research and Development for Life Sciences. Brazil.

Margarita Gómez Chantraine
Surgeon - Universidad Nacional de Colombia. Specialist in Obstetrics and Gynaecology - Universidad Nacional de Colombia. M.Sc in Clinical Epidemoilogy and Research - Pontificia Universidad Javeriana. Grouper Manager of Medical Affairs Johnson & Johnson Medical Devices LatinoAmérica. Colombia.

Maximiliano Gimenez Gigon
Orthopedic Spine Surgeon at Centro de Patologia Raquidea - Tucumán, Argentina AOSpine Fellow at "Santa Casa de la Misericórdia" - São Paulo, Brazil.

Miguel Farfán
Medical surgeon graduated from the National University of Colombia, Specialist in Clinical Epidemiology at the El Bosque University, Orthopedist and traumatologist graduated from the Santa Fé de Bogotá Foundation, Universidad del Rosario. General Manager and Founder Media For Health SAS, General Manager and Founder Orthopedic Unit of Colombia UNOC.co. Experience in health innovation, process optimization through the use of ICT (Apps) and improvement of the user experience and satisfaction of patients and health personnel. Experience in clinical research, creation of clinical records and systematic review of the literature. Development of educational innovation projects in health for doctors and patients.

Mohsen Sadegi Naini
Neurosurgeon. Spine Trauma and Surgery Research Center, Tehran University of Medical sciences.

Nelson Astur Neto
Associate Physician of the Spine Group of the Department of Orthopaedics and Traumatology at Santa Casa de São Paulo - Pavilhão "Fernandinho Simonsen" Research Officer AOSpine Brazil. Orthopaedic spine surgeon at Hospital Albert Einstein - São Paulo, Brazil. Active member of Sociedade Brasileira de Coluna (SBC), Sociedade Brasileira de Ortopedia e Traumatologia (SBOT), North American Spine Society (NASS), American Academy of Orthopedic Surgeons (AAOS), AOSpine. Brazil.

Pedro Luis Bazán
Specialist in Orthopedics and Traumatology, certified in Spine Surgery ((SAPCV). Head of the HIGA San Martin Spine Unit and Hospital Italiano La Plata. Director of the Post-Basic Training Course in Spine Pathology in Adults (Postgraduate Department Facultad de Ciencias Médicas de la Universidad Nacional de La Plata). Graduate assistant under contract for Orthopedics and Traumatology (FCM-UNLP). Argentina.

Rodrigo Góes Medéa de Mendonça
Orthopedic Spine Surgeon. Associate Physician of Spine Group at Santa Casa of São Paulo. Physician at Albert Einstein Hospital - São Paulo. Active Member of Sociedade Brasileira de Ortopedia e Traumatologia (SBOT), Brazilian Spine Society (SBC), Brazilian Pediatric Orthopedic Society (SBOP), North American Spine Society (NASS), AOSpine. Brazil.

Vafa Rhimi Movarghar
Professor of Neurosurgery. Spine Trauma and Surgery Research Center, Tehran University of Medical Sciences.

Víctor Fernando Luján Celis
Medical resident for research in social service División de Investigación en Salud. UMAE "Dr. Victorio de la Fuente Narváez" IMSS. Mexico City.

FOREWORD

AOSpine is focusing itself to be a Knowledge Creator in addition to being a world-class knowledge provider. This book resonates well with the current aim of the International Research Commission of AOSpine to make each one of its members into 'Surgeon Scientists'. Spine Surgery is undergoing rapid evolution and every day we are faced with newer techniques, modern implants and expensive technology. It becomes imperative that every surgeon has a thorough knowledge of clinical research, which will not only allow him to perform research individually but also evaluate the quality of the research work in the published material.

The conversion of a talented surgeon to a surgeon scientist however requires a clear knowledge of research protocols and this book is ideally suited for the same. The chapters are designed to guide the reader step by step on the planning of the study, collection of samples, selection of appropriate statistical tests, analysis of data and presentation of results. All the authors are experts in the field and each chapter is clearly written so that the surgeon would have no difficulty in choosing the best research protocol. I am particularly happy for the spine surgeons in training, fellows and residents who are actively involved in research as this book provides a clear understanding of the research methodology.

I congratulate the authors who can be very proud of this work that is an impressive compilation of valuable information for anyone performing clinical research.

Prof S Rajasekaran, Ph.D.

Chairman, Dept of Orthopaedics, Ganga Hospital, Coimbatore,
India Chair, AOSpine International Research Commission

Unit 1

Chapter 1
Introduction
Asdrubal Falavigna y José María Jiménez Avila

Chapter 2
Mathematics: language and interpretation
Maurício Scussel

Chapter 3
The main actor: the variable
José María Jiménez Avila

Chapter 4
How does one plan a database?
Pedro Luis Bazán

Chapter 5
Sample size calculation
Emiliano Vialle

Chapter 6
How to choose statistical tests
Pedro Luis Bazán

Chapter 7
Descriptive biostatistics
Alberto Aceves Perez

Chapter 8
Analytic or inferential biostatistics
José María Jiménez Avila

Chapter 9
Central tendency measures
Alberto Aceves Pérez

Chapter 10
Measures of dispersion (Range, Variance, Standard Deviation, Coefficient of Variance)
Delio Martins

Chapter 11
Data presentation
José María Jiménez Avila

Chapter 12
Hypothesis test (CI) and value of "p"
José María Jiménez Avila

Chapter 13
Kolmogorov-Smirnov test
Nelson Astur Neto
Maximiliano Gimenez Gigón
Rodrigo Góes Medéa de Mendonça

Chapter 14
Shapiro-Wilk test
Nelson Astur Neto
Rodrigo Góes Medéa de Mendonça

Chapter 1

INTRODUCTION

Asdrubal Falavigna
José María Jiménez Avila

The process of investigation is an essential element for the generation, palliation and renewal of diagnostic and preventive tools and for the timely treatment of patients.

A paradoxical conflict is that research in these fields is done in developed countries that are very far from having the health problems of the Third World, and also the experience of an economic reality and of inequalities in health, which are very different from their everyday experience.

This is the reason why it is necessary to generate research at a secondary practical level that is not so much oriented to basic sciences, but rather tending to the fields of clinical epidemiology, social sciences, health economics and user satisfaction.

A part of the research process is data analysis and something even more important, the interpretation of the information that generates the sum of all the details revealed in collecting information from the surveys.

For this we must know the fundamentals associated with statistics, which is a discipline of applied mathematics that is dedicated to the management of numerical data, ie., it is the method used to collect, elaborate, analyze and interpret numerical data.

Statistics is not considered a science but a method that teaches the logical procedures of observation and analysis, for which it is necessary to take into account and get the most profit out of the experiences in other fields of knowledge.

This book is aimed at all spine surgeons, medical students or physicians in general, who have broad surgical experience and have much information on their computer, but who have not taken the big step forward of interpreting the data, not only from the descriptive

standpoint, but from analysis or 'inference", which will allow them to obtain information that will be very useful for the scientific community.

The book has 28 chapters, divided into 4 units, which go from how the databases are structured, how the parametric and non parametric tests are interpreted in spine surgery, and very importantly, how to use the data obtained, not only to show them in an original article, but to know how to use the information to manage and produce a change in surgical medical care.

This book is aimed at people who wish to continue providing assistential and surgical care, who like research and feel the need to know how to interpret the statistical tests that appear in the section on material and methods and results, in the original articles, and which are often not analyzed. The data are taken as an absolute truth without realizing that in many cases their conclusions lack "Methodological Congruence".

This book provides support to Critical Analysis of Medical Literature to integrate the knowledge attached to the scientific criterion.

Remember: What is not measured is not evaluated, and therefore does not improve.

CHAPTER 2

MATHEMATICS: LANGUAGE AND INTERPRETATION

Maurício Scussel

Mathematics is the science of logical and abstract reasoning, which investigates geometries, proportions, variations and other topics. It is a wide-ranging science and inherent to the other sciences which it occasionally impresses due to apparently high levels of complexity.

Many see in this supposed complexity a barrier that makes it difficult to understand the phenomena that are represented in mathematical language. However, we ask: Could the difficulty in interpreting a given numerical problem lie in the symbolic language of applied mathematics, or could it be the result of not understanding the basic reason which is behind the symbols and mathematical notations?

Often the statistical evaluation of a study is restricted to the statisticians responsible for analyzing the results. However, the reflection on the language and interpretation of mathematics may be stimulated in the researcher who intends to describe the methodology or the result of a study. Thus the purpose of this book is not to transform the reader into a statistician, but rather to arouse interest in the reason that is at the foundation of this knowledge.

Since Antiquity, mathematics has always been present in the thinking of philosophers like Plato, Pythagoras, Aristotle and others. This is understandable, since mathematics is a tool that enables to use logic to demonstrate the reason for the mechanics of certain knowledge, and at the same time connect the real to the abstract. It is paradoxical that a science is so exact, logical, defining that two plus two are four, and yet also connects what is real to the imaginary, where $i^2 = -1$.

Considering all of this apparent complexity, how can one understand and dominate this language?

Mathematics is composed of an extensive symbolism to describe different functions, equations, and variations. However, in brief we

can say that these symbols, essentially, are used to transmit some knowledge that may be at the same time at an objective or subjective level. In order to illustrate these objective and subjective aspects of the symbols, let us think about the following symbols:

In a simple and objective analysis, each element is seen to be related to something concrete. The "cross", Christianity. The "correct" sign, to a brand of sports material. The "rings" to the Olympics. However, in each element presented there is masked subjective knowledge that is associated with the figure. For instance, for the vast majority of readers it is probably unnecessary to cite the name of the sports brand represented by the figure, nor to mention the great influence exerted by this brand on the sports world, since this is knowledge that is stored in the unconscious.

In order to create an image of the objective and subjective, we might say that a symbol is a sort of memory card. Its design would be like the physical unit, objective (object), and the content is the subjective one ("place under") what is in another layer.

But why does the subjective help understand mathematics?

If one manages to get closer to the subjective aspect found in the symbols used in mathematics, it is easier to understand the reason why an element is applied in a simple addition or subtraction. Thus, we will look at a few arithmetic operations with an open mind, to try to observe this property of the symbols from another angle.

Arithmetic is the field of mathematics that deals with the numbers and the operations that can be performed between them. The oldest and most elementary segment comprises, among various elements, the four operations: addition, subtraction, multiplication and division. To seek a closer approach between between "objective x subjective", we will analyze each symbol of the four operations, attempting to extract a "reason" why the symbol presents a given shape:

• **Subtraction:** this operation is represented by a continuous line (–). According to historians, this symbol appears in Western literature beginning in the 14th and 15th century with the metamorphosis of the Latin word *minus* to $\overline{mus}$ until it becomes –.

Analyzing the function of a subtraction, using mathematics or linguistics, we come to the idea that subtracting is the fact of taking something from. Thus, considering the need to represent a separation, a break, what would be a simple element to represent this action?

Imagine the following situations: Two neighbors who need to demarcate their properties: what geometrical element could be used to represent this separation between areas? When it is necessary to separate the skin for a surgical incision, what design is formed on the skin? It is a line.

Obviously this abstract approach to reason must be considered in a playful manner, not confronting the logical explanation for the reason that gave rise to the symbol. However once its knowledge is incorporated, it does not matter whether its origin was strictly logical or playful, it is simply important to understand the concept of element in order to be able to better apply it.

• **Division:** this is the operation currently represented by a continuous line separating two points (÷). According to historians, this symbol is more modern, and appeared for the first time in the 17th century.

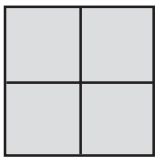
Analogously to subtraction, division also refers to a break between two parts. However, in division there is a dividing element that demarcates into how many **equal** portions the dividend will be fractionated.

Thus, to symbolically represent breaking something into equal parts, wouldn't it be necessary for the symbol that represents this operation to contain the fragmentation with equal weights between the two sides?

• **Addition:** this operation is represented by a continuous line with another continuous line overlaid. This symbol arose concomitantly with that of subtraction, and appeared in the same publications. The historical explanation is that this element came from the word "*et*" (which corresponds to the conjunction "and" (*e* in Portuguese): John and Mary: two plus two are four, etc.), which over time began to be abbreviated as "*t*", until it reached the currently used symbol (+).

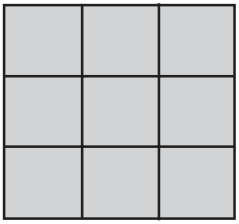
In order to attain a playful image of this element, we can imagine the result of some activity that needs to connect two points separated by a line or space. For instance, we can imagine a tennis shoelace that joints two parts, or a surgical suture that sews together two portions of skin separated by an incision.

• **Multiplication:** to demonstrate a reason why for the symbol of multiplication to present the (x) format, it is necessary to use geometry. This resource helps understand that when we multiply a number by itself, we are raising this element to the square (potentiation).

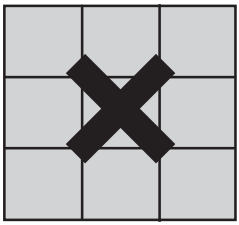


Let us imagine multiplying number 2 x 2. We can also say that we are raising number 2 to the square, since 2 x 2 = 4, just as $2^2 = 4$, therefore 2 x 2 = 2^2. But why use geometry? Observe the image below:

We can say that the square drawn above represents equation 2^2, since we have 4 elements in the square. Or, we can also say that it represents the multiplication of elements 2 and 2, since the result will be 4. In the case of 3 x 3, or, 3^2, we have an analogous situation:



We represent the 3^2 by multiplying the 3 by itself, resulting in number 9, which corresponds to the quantity of elements inside the square. Observing it from the standpoint of geometry, it is noted that, in multiplication, the elements not only grow in length, but also in width to fulfill the requirement of multiplication, consequently of potentiation. Thus, a symbol that possibly represents expanion in the two dimensions (width and length) could be X.



Of course, to present the scientific results, the subjectivity of the symbolism of the mathematical elements is not evidenced, because it is necessary to fix the concepts and bases to ensure the advance of knowledge, and the objectivism leads to a sounder path. Thus, the proposition is only to arouse curiosity regarding the whys of each mathematical element. This playful form of dealing with mathematics may not elucidate a problem to begin with, but any knowledge seen from another angle broadens understanding.

While reading this book, if you find it difficult to analyze or to understand a mathematical formula, remember that understanding an equation is analogous to understanding the clinical case of a patient. Each sign, each symbol is equivalent to a symptom, an examination, and each symptom has a history, a reason. When you encounter something complex remember the studies of clinical cases. Begin to analyze point by point, and through the reason of each element the equatlon will begin to become easier to understand.

Good Reading!

COMPLEMENTARY READING

1. Logic, Wesley C. Salmon. Prentice-Hall, 1963.

2. SUPPES, Patric. Introduction to Logic, Princeton, N.J., D Van Nostrand Company, Inc., 1957.

Chapter 3

THE MAIN ACTOR: THE VARIABLE

José María Jiménez Avila

In research, since its very beginning, everything turns on what has been called "variable" and this can be defined as the quality, property or characteristic of people or things being studied, which can be numbered or measured quantitatively or qualitatively, and that varies from one subject to another.

It should always be taken into account, from the time the research question is elaborated (PICOT), and also throughout the methodological process.

The function of the variable is to decompose the hypothesis presented into its simplest elements, and it is any characteristic or event that people or study groups, change from one situation to another or from one time to another, and therefore can take on various values.

In the literature we find several concepts involving the definition of the variables, and therefore each of the terms should be analyzed. Variables are classified as independent or explanatory variables, exposure (**cause**) and dependent or response variables (**effect** or **illness**) and there are others that modify some event, which are called intervening variables.

When a problem is identified and one attempt to provide a hypothetical explanation of the variable, it is necessary to submit it to a test, and this can be done by decomposing it into small variables with empirical evaluation.

If the empirical procedures do not refute the hypothesis presented, this is accepted as probably true. In most cases it is necessary to measure the variables during the empirical testing of the hypothesis and the measurement is an essential procedure in scientific practice.

A variable must have 2 essential characteristics. It must be "Measurable and Observable".

Measurement

Measuring consists of assigning a number or qualification to some specific property of an individual, a population, or an event, using certain rules. The individual is not measured, but a certain characteristic of the individual; however, measurement is an abstraction process.

In order to measure, a process must be followed. It consists of going from a theoretical entity to a concpetual scale and later to an operational scale.

The steps followed are:

a) To demarcate the part of the event that will be measured.

b) To select the scale to be used for measuring.

c) To compare the attribute measured with the scale.

d) To analyze and promte a judgment of the results of the comparison.

Main measurement scales (types of variables)

The scales are classified as qualitative, nominal and ordinal, and quantitative, continuous and discrete. An essential requirement for all scales is that the categories must be exhaustive and mutually exclusive.

- ***Nominal scale***

The characteristic of this scale is defined by a name, and when it is defined, this does not imply that it is more or less than the characteristic defined by a different name. For instance: Sex: Male, Female, Occupation, Type of childbirth smoking, etc. For practical purposes, occasionally a number is given for coding. This is to make it easier to capture. In such cases the number is meaningless, it is only to distinguish one group from another, without higher or lower value.

- ***Ordinal scale***

The characteristic of this scale can receive some subjective order, it can be assumed that a characteristic is more or less than the others, but it is not known how much more or how much less. Example: Good or Bad. The order is subjective or rather chosen by the researcher, schooling: elementary, secondary, etc.

• ***Discrete Scale***

In this scale the values are separated from each other by a determinate quantity and the unit cannot be fractionated. When the unit is divided, it disappears: For instance: age (number of years), Number of children.

• ***Continuous Scale***

In this one, the measurement scale can be divided into an infinite quantity of values between any two points. A meter can be divided into an infinite number of parts. All the derivatives with length or time, for instance: volume, area, weight , size, temperature.

Process of Operationalization of Variables

The variables that are researched in a study are identified as soon as the problem is defined. Within the theoretical framework the secondary variables are identified and conceptualized. This level of definition is abstract and complex and does not allow observation or measurement.

The process of taking a variable from an abstract level to an operational level is called "Operacionalization" and the basic function of tis process is to define or implement as much as possible the significance or outreach granted to a variable that is being studied.

The concepts represent the theoretical system of any science and are symbols of the phenomena that are studied.

A concept is general and abstract, and it is not possible to observe the phenomena involved, therefore it is necessary to operationalize the variables.

Operationalization is done through a process that transforms a variable into others that have the same meaning and can be measured empirically to achieve this. The main variables are decomposed into other more specific ones called dimensions, which must in their turn be translated into indicators to allow direct observation.

Example:

Variable: Evolucion of the fracture

Indicator: State of consolidation

Dimension: Phases of consolidation

Scales: 1. Soft callus.
2. Hard callus.
3. Corticalization.
4. Remodeling.

When this process ends, we can tell that the variables obtained in a conceptualized manner can be measured and observed. At the same time this will allow establishing an association that will enable the identification of causal factors that can be predictors of the identification of risks factors, which can trigger some kind of illeness or result.

Variables must be clearly defined in order to avoid confusion. This will make it easier to search for and later analyze the data, which will ensure the compatibility of the results obtained with some previously performed studies.

COMPLEMENTARY READING

1. Celis de la Rosa A. Bioestadística, en: Características de los datos y de su recolección, revisión y computo. 1ª 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. N. 4, julio/agosto 2000 p. 337-348.

CHAPTER 4

HOW DOES ONE PLAN A DATABASE?

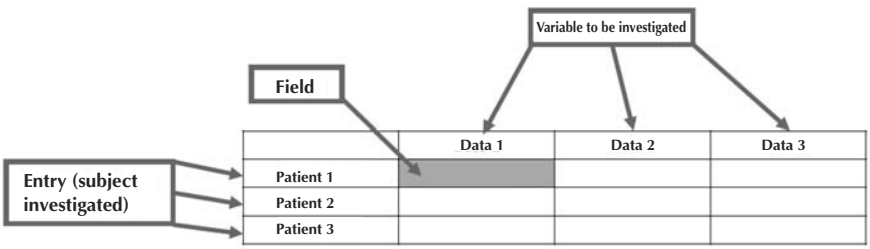
Pedro Luis Bazán

Introduction

The correct management of information is essential for any kind of work done by human beings Scientific medical work follows the same rule and that is why we should know and plan the best way possible of collecting, storing and analyzing information, in order to answer the question that gave rise to our research.

A simple way is to create files as calculation spreadsheets (tables) that contain a topic in each file. The functional unit of these tables are the cells of the sheet, also called fields. In the first cell of each column will be the title of the variable to be evaluated, which will be the same for each patient entered. Each line represents a registration, which represents the information entered for each patient (Fig.1). Another sheet can be created that receives other information and another document.

Figure 1 – Example of components of a calculation spreadsheet



The current calculation spreadsheets allow performing different functions for the statistical analysis of data. Managing the information in this way may present some problems mainly involving the complexity of the scientific work planned and the need to have several sheets. In this way we may find information to be redundant, inconsistent and unusable. This makes it difficult to manage the error of the input of information, one of the most frequent errors in scientific research and the loss of data.

Other difficulties in using files for data management are the complexity of dynamically modifying the structure of the files, consulting, updating and protecting the data.

Another way to manage information is by creating databases that will make it easier to manage them. The Royal Academy of Spain defines the database as a set of data organized in such a manner that it will allow obtaining various types of information speedily".

To create and manage them we can use various database managers (Software). Although they are of different degrees of complexity, they all try to store the data in a structured form in order to reduce input errors and make it easier to analyze the information. Examples of thse managers are Microsoft Access, EpiInfo, IBM, etc.

These database managers are characterized by (1) the independence of the data without being related to the applications that use them or where they are stored; (2) the users of these databases can access by consulting them; and (3) the data are stored in a centralized and independent manner, making them consistent, complete, reliable and safe.

The databases are stored in files that contain data, numbers and characters, that can be opened, saved, combined, edited, added or removed. The general structure of the files that contain databases includes tables that can be related to each other.

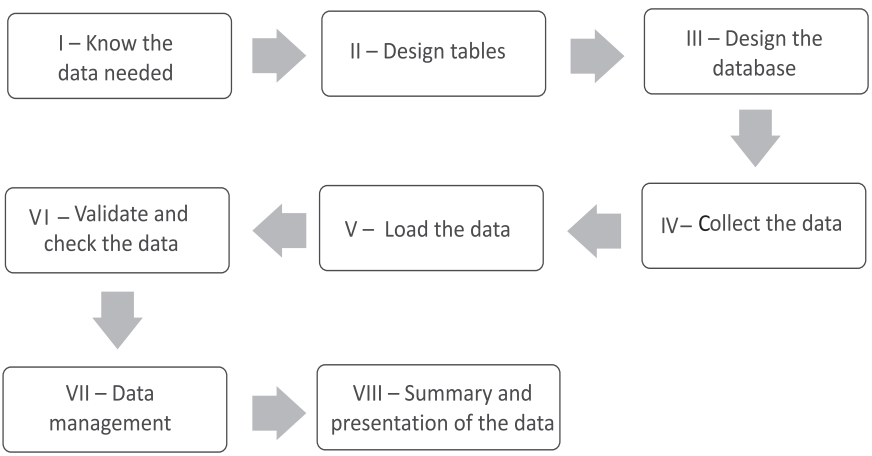
We can classify the databases as: **primary**, which are designed according to the specific needs for information (study, recording, generating indicators, etc.) and therefore, with the appropriate design, they adjust the answer to our initial question; **secondary**, this is a database used for a different purpose from that which motivated the collection of the data it contains. The speediness and economy of its use are its main advantages, but the data may be of a low quality, often requiring that the objectives be adapted to the data; **tertiary**, these are created using the data from different databases combined from previously reported data.

Planning a database

All scientific research originates in getting to know the reason (question), which stimulated it. Once this requirement has been met, the researcher must plan everything concerning scientific work, bibliographic support, type of study, inclusion and exclusion criteria, the data needed to decide on the topic to be researched and to establish a time line for all these moments.

Concerning the database proper, it is also necessary to complete a number of steps for the appropriate management of information, and in each of them we must know who is responsible for it (Figure 2).

Figure 2 – Flow chart in planning and using a database



Getting to know the necessary data

Before beginning to collect data it is essential to determine which data are important for the research. This will allow collecting all of the information and only what is needed.

If this premise is not followed, it may lead to the information not being sufficient to meet the objectives established. Analyzing the results we can realize that this is happening and be obliged to modify the material evaluated or the method for their analysis, having lost time in vain or determining that the research must be abandoned.

It is also not reasonable to collect unnecessary data that are of no use to answer the objectives of the research study and only take up

excess time and resources. For these and other reasons, it can be concluded that, like in any activity, one of the most important steps is planning what we are going to research, what we need to do this, how we shall do it, with what or with whom we will do it, etc.; in brief, the success of research depends partly on correct planning.

Designing the tables

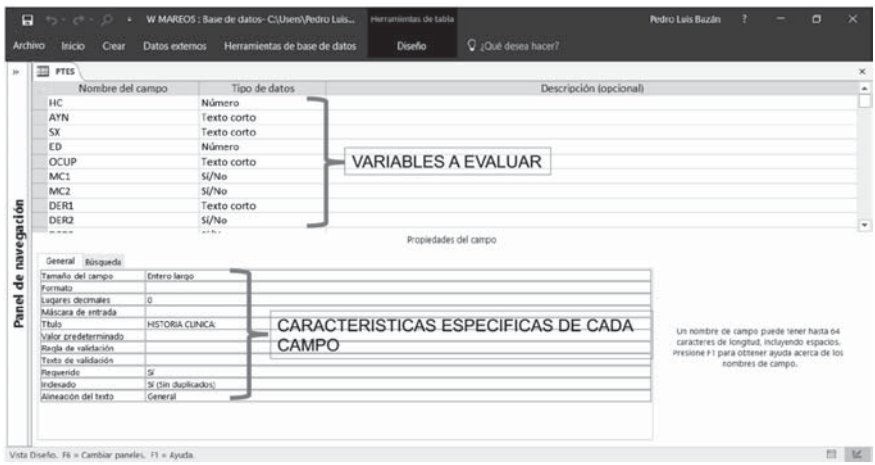
As we mentioned previously, a database is a set of interrelated tables to make data management easier.

When we know which data we shall need for our research, we will begin to decide on the structure of our database. Depending on the complexity of the data to be loaded, the complexity of the database to be created will be determined.

The different database management programs enable us to edit the way information is entered and how the different tables are interrelated according to our needs.

In each field selected (Figure 3), a number is given to this variable and different characteristics are given to the data to be loaded, which will enable us to reduce the loading error and have all elements share the same characteristics, thus allowing the correct analysis.

Figure 3 – Design of the table. Observe the selection of the type of data for each variable in the upper table and the validations performed for each field in the lower one



1. Text. They contain data of text and numbers, but in the latter case no mathematical operations can be performed. They are useful to save information such as names and addresses.

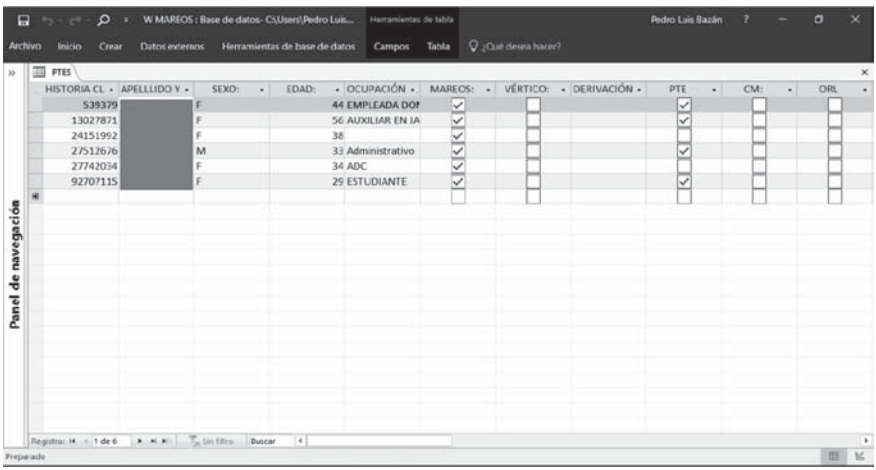
2. Number. They are used to store numerical information. They can contain continuous or categorical data, integers or decimals. This must be specified when the structure is set up. They can be used to perform mathematical operations.

3. Logics; Yes/No; True/False. Data are stored that have only two possible answers, and they are called binary.

4. Date/Hour. Dates are stored. They can be short (year with the last two digits), long (year with four digits), associated or not with the time of day. They can be used to perform calculations between dates, and most of the programs only allow loading valid dates.

Each of these fields will correspond to the variable to be studied, and is represented at a glance on the data sheet, with columns, the rows corresponding to each subject, and each particular box represents the specific value of the variable to be studied in each subject (Figure 4).

Figure 4 – Data sheet of the table designed in the previous figure



Designing the database

Knowing the different tables required it is necessary to determine how our database would be.

This step can be broken down into three stages.

1. Conceptual design: during this stage the structure of the database is planned, without defining what technology or software is to be used. A sketch is made of what is needed to work, based on the data that we need to evaluate. At this point we shall decide what type of database is to be used (spreadsheet, relational, object-centered, hierarchical, networks, operational relations, etc.)

2. Logical design: is the moment when the idea is applied to reality based on the technology to be used, in other words with which database manager we shall work and which keys and variables we hierarchize.

3. Physical design: this is to relate the structure of the database to the hardware needed to obtain greater efficacy of the load, storage and consultation of the data, for instance, take into account the internal or external memory, storage capacity for storage of the data to be loaded, safety copies, etc. This occurs especially if besides image data are also stored.

It is useful to end this stage by performing several tests of input, consulting and modification of virtual data to evaluate the database yield, and to recognize flaws in programming before beginning to load the data.

Collecting the data

Before we begin entering patients, we must create a spreadsheet that will serve to collect the information required and must be collected in the same order as it will be loaded. Each spreadsheet must be labelled with the same key of the database record, to be able to access it rapidly if there is an input errror or if some intem of information is missing.

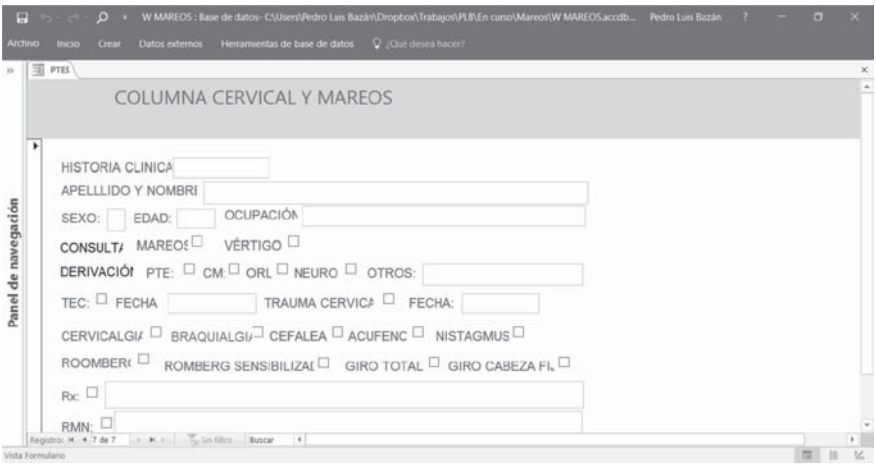
Loading the data

The information can be loaded into our database in two possible ways, based on the table that has already been loaded. One of them is seeing the same table as a data sheet that we created in the beginning.

(Figure 4). The problem is that when loading the data, the cursor must kept on the same row (patient) and move from the column (variable to be evaluated) in order not to maket the mistake of loading an item of data in another record.

Another way of recording the information in a database is by creating a form based on the table design (Figure 5). In this way we see on the screen only the field to be completed with data from a single record (patient) and accept the item of data loaded, and then move on to the next field of the same patient. When the loading is ended one goes on to another record.

Figure 5 – Data input form with the fields to be completed



Since there are options for the automatic validation of the load or its mandatoriness, the program lets us know the problem in performing the input, and thus we reduce the input error.

Validate and check the data

This step is essential for the correct analysis of the information and to obtain the results of the study.

The data should be validated methodologically, every time information is transferred (collection, loading and storage). If this is not done, we will realize that there are lost or inconsistent data in a final stage of the research, and this will make it fail or require additional time to control the error.

When the research is done as a group it is advisable that a single person be responsible for loading the data, and that this person, whoever they are, validate and check the loading process.

Data management

A crucial point in managing information is to analyze the data, and this will be the subject of another chapter.

When the data have been loaded, it is possible to enter the database for consulting or reporting, and that is where one decides which statistical operation will be used.

Summary and presentation of the data

The database management systems can analyze the database and then deliver a summary report, double entry tables and graphs that illustrate the results.

Persons responsible

In order to avoid data management errors (collection, loading, storage and analysis) a know person should be responsible for each step. All the authors of the research are responsible for the decision making, both in planning the research and in recognizing the data necessary for it.

There should be one or two persons responsible for planning the database and the spreadsheet to collect the information: if one of them is a specialist in statistics so much the better. All the researchers must be involved in collecting the data and analyzing the results.

Only one person must be responsible for loading the data. This person should be the same one who evaluates that all necessary data have been collected.

A specialist in statistics or one of the authors who is more familiar with data analysis should be responsible for selecting and analyzing the statistical tests (Figure 6).

Figure 6 – Report on the data analisys performed by Epi Info, with a double entry table and results

Epi Info

Library of results

Current view C: C:\Users\Pedro\Documents\Trabajos\HIGA SM\En curso\Modic\PACIENTES.mdb:PTES

Number of records: 100 Date: 09/08/2010 11:06:02 p.m.

Tablas Disc I Instability

Following procedure

INSTABILITY

DISC _1	0	1	TOTAL
0	48	17	65
% Fila	73.8	26.2	100.0
% Column	63.2	70.8	65.0
1	28	7	35
% Fila	80.0	20.0	100.0
% Column	36.8	29.2	35.0
TOTAL	76	24	100
% Fila	76.0	24.0	100.0
% Column	100.0	100.0	100.0

Continue

Analysis of a simple table

	Point Estimation	95% Confidence interval lower L.	Upper L.
PARAMETERS: basead on OR			
Odds Ratio (cross product)	0.7059	0.2607	1.9111 (T)
Odds Ratio (EMV-MLE)	0.7083	0.2465	1.9004 (M)
		0.2201	2.0773 (F)
PARAMETERS: Based on risck			
Risk Ratio (RR)	0.9231	0.7408	1.1502 (T)
Risk Difference (DR)	-6.1538	-23.1763	10.8686 (T)

(T-series Taylos; C-Cornfield, M-P-Media; F-Fisher)

Statistical Tests	Chi-Square	p. de 1 tail	p of 2 tails
Chi-square, sem corrigir	0.4723		0.4919156556
Chi-square, Mantel-Haenszel	0.4676		0.4940887145
Chi-square: Yates correction	0.1952		0.6586242660
Exact p–mean		0.2548047339	
		0.3335295358	

Utilities of a database

If the planning has been successfully performed, the database created not only allows us to input data through a form which has the management utility on the screen. It also is possible to consult using various parameters that allow us to perform partial evaluations every time the data are loaded, relating more than one table of the total.

Reports of each entry can be performed individually, and also of the consultations. The difference between the reports and the forms in viewing each entry is that the forms are viewed only on the screen, on the other hand the report is the form of printed visualization that can be used.

Conclusions

Data management includes loading, storage, consultation and analysis of the data and it is very important in medical research. The databases are essential tools to manage research information. In order to plan them it is necessary to perform analysis and know the data needed to carry out our work, and we must base it on them.

Before creating a database, the following questions must be asked:

What data do I wish to store, and what is the best way to do it (Tables and their relations).

How do I introduce data concerning the research? (Form).

How do I perform partial evaluations of the data? (Consultations).

How do I evaluate the data? (statistical operations).

COMPLEMENTARY READING

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: Bases de Datos I. Introducción a las bases de datos. fundamento 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: worklflow 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; at col. 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 Bases 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.

Chapter 5

SAMPLE SIZE CALCULATION

Emiliano Vialle

Introduction

When I was invited to write about this topic, I had the feeling that I would finally solve a crucial question that arises when a study is elaborated. This question troubles physicians and researchers at all levels of research and training. The reality that I encountered was an infinity of fomulas that are chosen very subjectively. Far from transmitting a message of discouragement to the reader, I suggest a more informative reading on the subject, without concerning oneself with the calculations proper, but with the information needed to develop them together with a statistician.

Defining the number of participants in a research study has always been a crucial issue that is difficult to solve, and often, after the end of the project, we find that the number of participants was insufficient. It appears obvious that a larger number of subjects included in a research study will produce more reliable results that are closer to reality. Due to time, resources and frequency of the question to be answered it is not possible to include patients in a study indefinitely, and the crucial question is then what is the minimum number of patients in each research group. The calculation of the sample also has been requested by ethics committees to approve research projects.

In this chapter I shall attempt to use practical examples to define the most common ways and practices to attain that "magic number". It should be undescored that the formulas to calculate the sample are sensitive to errors, and variations in the parameters may lead to great differences in sample size. Therefore, before we begin to establish the size of our sample, once the research project has been elaborated, a few items must be defined:

1) Type I error (alpha)

In clinical studies only a part of the population is evaluated, testing the hypothesis that there is a difference between samples (rate of infection in diabetics, for instance), or in the result of treatment among groups (arthrodesis of the spine in diabetics, versus non-diabetics), or in the results of treatments in a group (arthrodesis versus conservative treatment in diabetic patients). The null hypothesis (H_0) is based on the principle that the patients (samples) that are being compared are not different. The alternative hypothesis (H_1) assumes that the groups are different. At ths time, the sample collection defines when the inclusion of patients in a study will no longer affect the result.

Two types of error can occur in this process (type I and type II, Table 1).

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

Type I error (alpha) measures the probability that based on the null hypothesis (H_0) that the samples come from the same population, the differences found in the study are true. It represents the chance of erroneously rejecting H_0 and choosing a false positive result. Alpha is generally set as 0.05, meaning that the researcher desires a risk of less than 5% of presenting false positive results.

2) Ttype II error (beta)

Instead of reaching a falsely positive conclusion, the researcher may reach a false negative one, concluding that there is no difference in a treatment between groups when in truth there is a difference. In other words, they accept H_0 as true, and this is called type II error (Beta). Traditionally, Beta is defined as 0.2, meaning that the researcher wishes for a chance of less than 20% of arriving at a fase negative conclusion. To calculate the sample size one must determine the power

of the study. This statistical power reflects the capacity of the study to find an effect on the population, based on a sample of this population (true positive). The statistical power is the complement of type II error (1-beta). Thus, if the type II error is 0.2, the power of the study will be 0.8 or 80%. This is the probability of rejecting the null hypothesis correctly, or avoiding a falsely negative conclusion. To be more objective, it is the probability that, when a difference is presented in a study, it is true.

3) Relevant minimum difference

It is the minimum difference that the researcher wishes to identify among the groups studied. It must be a clinically relevant and biologically plausible difference.

4) Variability

The calculation of the sample is based on the population variation for the variable studied through the standard deviation. This value is generally obtained by a pilot study or from previous studies. The minimum difference and the variability can sometimes be represented as a multiple of the standard deviation of the observations. Formula:

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

5) What is the size of the population?

Often this information is difficult to obtain especially if the project involves a very specific study.Imagine that your study involves a comparison between surgical and conservative treatment of fractures in the thoracolumbar spine. The study population will be obtained through some epidemiological study that does not always apply to its reality, or through a review of the hospital patients' charts.

6) What is the margin for error (confidence interval) desired?

Based on the principle that the sample will not be perfect, it must be defined how much error will be accepted, and generally 5% is the value used.

7) What is the level of confidence desired?

The most widely used confidence intervals are the values of 90%, 95%, and 99%. This value will determine the possibility that its result is within the confidence interval.

8) What is the standard deviation?

How much variation do you expect for your results? Since it is difficult to answer this, most studies use the value of 0.5.

Having the data above, but not knowing the size of the population involved, we begin to calculate the sample. Its level of confidence corresponds to a Z-score, a constant value used to calculate the sample:

90% – Z-score = 1.645

95% – Z-score = 1.96

99% – Z-score = 2.576

There are tables for the Z-score values available on internet, but the values above are the most common to apply the formula:

Sample size = (Z-score)^2 x SD x (1-SD) / (margin of error)^2

Let us say that you chose a level of confidence of 95%, a standard deviation of 5, and a confidence interval of +/- 5%. Let's ignore the sample size or consider that the study involves a significant part of the population.

Sample size = ((1.96)^2 x .5(.5)) / (.05)^2

(3.8416 x .25) / .0025

.9604 / .0025

384.16

385 participants would be necessary.

If this number is not encouraging, it is possible to reduce the confidence interval or increase the margin of error. The formula above is one of the simplest ways of determining reseach samples, and it is often used for opinion surveys, for instance where it is not necessay to

compare groups. The Table below shows how the definition of the values alpha and beta influences the sample size.

Table 1 – Approximate relative sample size for different levels of alpha and power

	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

When we perform clinical studies where differences between groups of patients are surveyed, some further information is needed.

Let us consider a study comparing surgical and conservative treatment for thoracolumbar spine fracture, where the participants will be randomly distributed into two homogeneous groups and the final result would be the level of pain six months after the beginning of the treatment. Based on previous studies, it is known that fractures treated conservatively (group A) present an improvement of 15 points of disability, and that surgical procedures (Group B) improved 20 points.

Previous studies show that the variation of disability can be 5 points (standard deviation of the sample).

In order to find out whether the surgical treatment leads to a significant improvement compared to the conservative one, the Student t test for unpaired samples can be used, and to ensure that the study will not have an insufficient or excessive number of participants, it is necessary to plan the sample size. This requires a specification of the statistical power (probability of identifying a real difference with the test, generally 80 or 90%) and of the level of significance of the statistical test (probability of obtaining a significant result when there is no real difference, generally 2.5% for one-tailed tests). Based on these results, statistical formulas begin to appear that are not familiar to most physiciains, in the case of the t-test above:

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

In the formula above, the groups (n1=n2) and standard deviations are the same. The mean difference between the two populations is u1-u2. Power (Z-power) is given as 0.8 or 80% and the level of significance is alpha. N is the sum of the two groups (n1+n2). The values to be inserted into the formulas are obtained from tables available on the internet. The work by Noordzij et al (2010) supplies the most frequently used values.

It is obvious that this formula is discouraging, but clearly it cannot be interpreted only by a statistician. The reason it is described is precisely to inform the surgeon about the data needed to calculate the sample size and allow joint work to choose the best statistical test and the sample that is most feasible for the clinical reality in which the study will be performed.

For the t test shown above, it is necessary to estimate the mean values in both groups compared, and also the standard deviations.

For the Fisher test, an estimation of the proportion of events must be supplied (for instance an adverse reaction) in both groups.

The Mann-Whitney test requires an estimation of the probability that the variable studied will be smaller in one group than in the other.

Simpler or less pretentious studies can use graphs to estimate the sample size based on assumptions about the type of data and statistical test to be used, as the one published by Altman (1980).

Example of experimental study

To determine the sample size of a study on the effect of cell transplants on disc degeneration in rabbits, the value obtained in a previous study comparing normal and degenerated discs was used.

The samples size was calculated in program WinPEPI (*Programs for Epidemiologists for Windows*) version 11.43 and based on the study of Vialle et al (2012). For a 5% level of significance, power of 80% and a minimum effect size of 1.2 standard deviations among the groups, a minimum total of 12 animals in each group was obtained.

Unfortunately many stages of the determination are subjective, such as the determination of the statistical test power, and others are external, such as the frequency of the disease studied, time of recruitment, resources available. Adaptations can be made as a result of these limitations, reducing the test power or the confidence interval.

Despite the difficulties imposed by the stages of the determination of sample size and the possibility of error, it is always better than an arbitrary choice of sample.

COMPLEMENTARY READING

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.

Chapter 6

HOW TO CHOOSE STATISTICAL TESTS

Pedro Luis Bazán

Introduction

Statistics is the science that uses sets of numerical data as a base to obtain inferences based on the calculation of probabilities. Two types of inferences can be distinguished, those that are known as hypothesis tests and those that estimate intervals.

The importance of statistical analysis lies in the decision to accept or reject the hypothesis of the study that is to be performed.

It can be divided into two large areas: descriptive statistics, which dedicates itself to the description, visualization and summary of data originating in the phenomena studied; and inferential statistics, that is dedicated to generating the models, inferences and predictions associated with the phenomena involved taking into account the randomness of the observations.

Thus we can analyze distinct variables that are defined as the qualities, traits, attributes, characteristics or properties of different values, magnitudes or intensities valuated in a group of elements. In this way we have the qualitative variables which can be binary (e.g. yes or no; feminine or masculine, true or false), ordinal (eg. staging of tumors) and nominal (e.g. type of thoracolumbar fracture); and the quantitative ones, which assign a value, and can be continuous (e.g. weight, age, size) or discontinuous (e.g. number of infections at the surgical site).

Another way of classifying the statistical variables is: independent variables (background, cause, etc) and dependent variables (effect, result, etc.).

A frequent error is to have a large quantity of information at the end of our research and to decide at that time how we shall analyze it; we then realize that we do not know what to do with those data and that often they are wrong or incomplete, and do not allow us to answer the question that led to our research.

The right planning of the research process is essential for the success of our work. It should be completed before beginning to load the data and there are different steps to be taken into account:

- Plan the question that originates the research.
- Decide the objectives of the work that will answer the question.
- Here one decides which variables are to be researched.
- Define the null hypothesis, with a confidence index between 0.05 and 0.01. This step shows how the results are to be interpreted.
- Decide the type of study and the sample needed. From this the statistical tests to be used are inferred.
- Collect the data on spreadsheets that describe each variable.
- Introduce those data into the electronic spreadsheets.
- Make coded lists of the variables to be analyzed.
- Correct the databases using data managers and calculation spreadsheets.
- Describe the data graphically and numerically (descriptive statistics). This is the last possibility to confirm data quality.
- Infer the characteristics of the population based on the sample (inferential statistics).
- Interpret the results according to the objectives and the null hypothesis.

All the steps are important, but the first seven steps are the foundation for success of the statistical analysis of data.

Forms of selecting the different statistical tests

There is no master recipe to select which is the most appropriate statistical test for our research, but there are different ways of choosing the appropriate one, and this is what we shall analyze. The test to be used must be selected before beginning to load the data. The first step for this is the hypothesis test, to confirm or rule it out, and on the other hand, the confidence interval.

Types of statistical tests

Statistical tests can be classified in two large groups, the parametric tests and the non-parametric tests. The use of these tests depends on

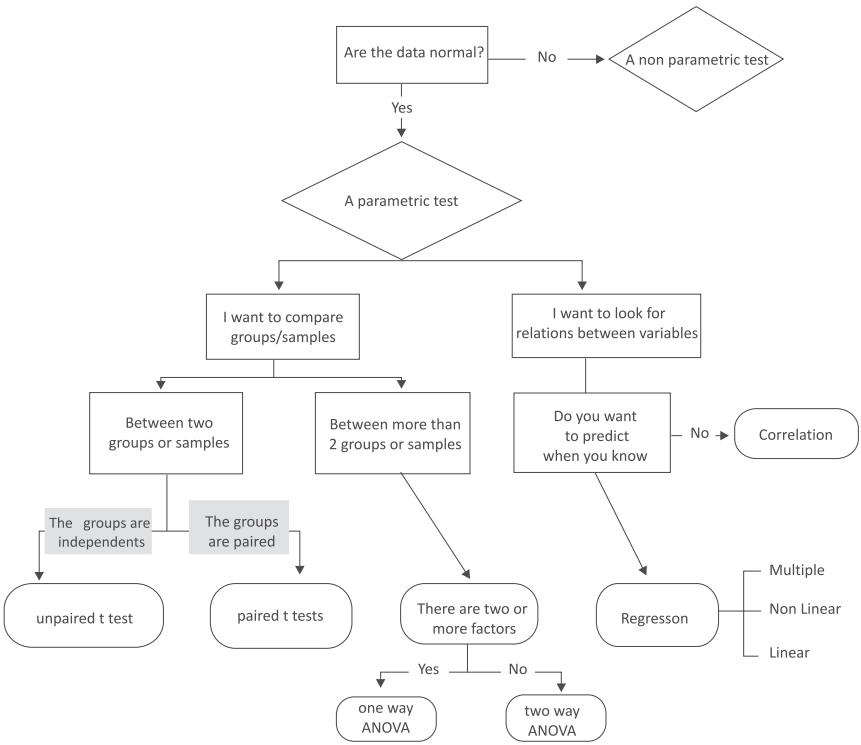
various factors that can be analyzed from different points of view: data distribution, observations to be studied, their variables, centralization, dispersion, content of the data, and sample size (Table 1).

Table 1 – Different characteristics between the parametric and non-parametric tests

Caracteristics	Parametric test	Non-parametric tests
Distribution	Normal and homogeneous variances	Asymmetrical heterogeneous variances
Observations	Real	Real or converted to ranks
Variables	Proportional or at intervals	Nominal, ordinal, scale of intervals
Centralization	Mean	Medians, modes
Dispersion	Variances	Ranges
Counts	Dados that sample frequencies	Must be transformed
Sample size	n>30	n<30

Among the parametric tests are: Z-value test of normal distribution, Student t test for related data (dependent samples), Student t test for non-related data (independent samples), Student-Welch t test for two independent samples with non-homogeneous variables, Chi(ji) – square tests of Bartlett to show the homogeneity of variance t, F-test (analysis of variance or ANOVA). We can create a decision-making schematic for the parametric tests with the flow diagram shown in Figure 1.

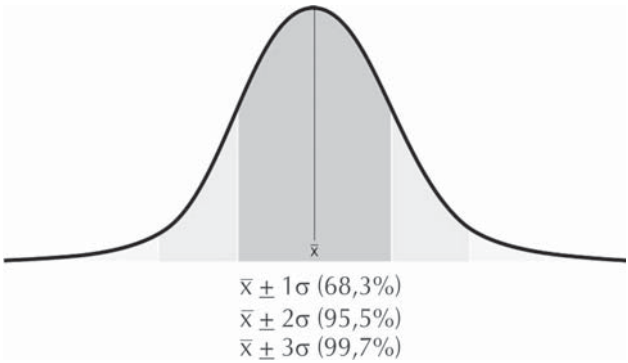
Table 1 – Flow diagram of parametric tests



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

When the parametric requirements are not met, first of all it is necessary to transform the data in order to normalize the distribution and diminish the bias, and for this we can use square root, logarithmic tests arc sine or inverted chi^2, if in spite of this transformation it is necessary to perform non parametric. Therefore, in deciding which test to use, the first question that must be answered is whether the data are distributed normally and expressed as mean and standard deviation (where 99.7% of the population is located), for instance following a Gauss bell shaped graph (Figue 2) and, if the answer is positive the test that should be used is parametric. The next decision is do I wish to compare different samples, or do I want to relate different variables. If I wish to compare two samples of independent data, I shall use the unpaired Student test; if the groups are dependent the paired student t test is used; and If there are more than two samples we shall use the one-way or two-way ANOVA, depending on whether there are two or more factors.

Figure 2 – Graph corresponding to a Gauss bell that shows the normal distribution of the data with their respective standard deviations



If we want to find relations between variables and if I want to know a value, if I know another, we will use regression tests; otherwise a correlation test is used.

When the data distribution is not normal, it is expressed in terms of mode and the range, median and percentiles; and if non parametric tests were used, one can mention: Pearson x^2 test, binomial test, Anderson Darling test, Cochran test, Cohen-Kappa test, Fisher test, Friedman test, Kendall test; Kolmogórov-Smirnov test, Kruskal-Wallis test, Kuiper test, Mann-Whitney (Wilcoxon) test, McNemar test, median test, Siegel-Tukey test, sign test, Spearman correlation coefficient, Contingency tables, Wald-Wolfowitz test and rank test with the Wilcoxon sign.

The advantages of using non-parametric tests is that they are simple, easy to apply and use; they utilize larger population groups; they can be used with ordinal or nominal data; and they can be used on small samples.

Sample size

Another important point to decide which variables should be used is sample size (n). Therefore, many statistical packages do not allow one to know how much data we need so that the results can be extrapolated, and their reports consistent from the point of view of research. Different authors do not come to a definitive conclusion, but there is a consensus that the larger the sample, the more the random error can be diminished. The minimum volume to consider a sample as representative could be 20 or 30.

In brief, the size of the sample is necessary to perform a study with the least possible number of elements that will be sufficient to obtain the results with the established precision.

Type of variable

A first step to analyze the data is to get to know the variables with which one is working.

The statistical tests are more powerful if the variable is quantitative, since it contains more information than if it were qualitative. According to the type of variable and its correlation, we can select different tests. If there are two qualitative variables we will use non-parametric tests such as x^2. If one is qualitative and the other quantitative, we will be able to choose between a few parametric ones (Student t and analysis of variance) and non-parametric ones (Mann-Whitney U; Wilcoxon rank test and Kruskal Wallis test). If both variables are quantitative, we will also be able to choose between parametric (Pearson correlation coefficient) and non-parametric (Spearman correlation coefficient), Table 2.

Table 2 – Selection of tests according to different variables

Variable 1	Variable 2	Parametric test	Non parametric test
Qualitative	Qualitative		x^2 Test
Quantitative	Qualitative	Student t Student t for paired data Analysis of variance	Mann-Whitney U Wilcoxon rank test Kruskal-Wallis test
Quantitative	Quantitative	Pearson correlation coefficient	Spearman correlation coefficient

Another way in which the relation between the different variables can indicate the type of test to be chosen could be:

a. Two continuous variables with large n, the Pearson coefficient is used.

b.Two continuous with median or small n; Spearman correlation coefficient or Kendal correlation coefficient.

c. Continuous vs ordinal; Kendall correlation coefficient or Cuzick's tendency test.

d. Continuous Vs nominal; parametric ANOVA.

e. Continuous Vs dichotomic; t test or Mann-Whitney test.

f . Two ordinal variables; Kendal correlation coefficient; Cuzick tendency or Chi-Square.

g. Ordinal vs nominal; Chi square test.

h. Ordinal vs dichotomic; Chi-square with a linear tendency.

i. Two nominal variables; Chi square.

j. Nominal Vs dichotomic; Chi square without a linear tendency.

k. Two dichotomic variables; Fisher test, Relative Risk or Odds Ratio.

Question

According to the phrase coined by Albert Einstein: "The formulation of a problem is often more essential that its solution", the choice of the test is also related to the question that motivated the study.

If it is only one group of analysis, the test chosen could be the confidence interval or the t test. If we have two independent samples, with a quantitative variable on a continuous scale, and we wish to know whether there are statistically significant differences, we must use the unpaired Student t test. If on the other hand, they are two dependent groups, with quantitative variables, discontinuous scales, the tests can be paired Student test, Wilcoxon or Pearson correlation. If there are three or more independent groups, the tests can be ANOVA or Chi-Square. If there are three or more dependent groups the tests used will be Cochran test or ANOVA with repeated observations. If on the contrary they are multivariate groups, the tests will be linear logarithmic or regression.

Type of study

We can select the statistical test to be used by relating it to the type of study we are planning, therefore, when we begin planning our experiment we can know what tests we are going to request. For cohort studies, cross-sectional studies and clinical trials Relative Risk can be used, and for cases and controls the Odds Ratio.

Types of data

The groups of data can be independent or dependent. They are independent when the groups are not related and each of them is of a different size. They are dependent when they are directly related, they are cause-effect groups. The n of each group is similar.

Final comments

The statistical analysis plan is a defined stage before collecting the data. It is the consequence of a good theoretical framework of the topic to be researched. The design of the form and the sample size depend on the complexity of the statistical analysis established.

Manuel Gomez Gomez and collaborators explain this topic concluding that the choice of statistical test depends on: (1) the type of design of the research; (2) the objective or objectives of the study; (3) the specific interest of the authors; (4) the data distribution; (5) the way the data are compiled; and (6) the potential for bias and confusion of the data obtained.

COMPLEMENTARY READING

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

2. Cazau, P: Fundamentos de estadisticas. 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 reperidas. Int J Clin Health Psychol; 7 (1): 154-175.

5. Herrera F, JC; Carse, LE: Guia de aplicación de pruebas estadísticasnen el programa Systat 7.0 para ciencias biológicas y forestales. Seccion 4.

6. Tetitla Munive, JM: Cómo elegir la prueba estadísticca. En Metodología de la investigación para el área de la salud. Ed McGraw Hill. 2013; Cap 9: 71-79.

7. Universidad de Alcalá – Departamente de ecología: Métodos de análisis de datos en ecología.

CHAPTER 7

DESCRIPTIVE BIOSTATISTICS

Alberto Aceves Perez

Biostatistics are usually thought of as an organization of numerical data presented in an ordered and systematic form This idea is the consequence of a concept of common use regarding the term, and that has been increasingly extended due to the influence of our surroundings, since currently it is practically impossible that any diffusion media will not tell us some kind of statistical information.

Thus, already specifically in the field of research such as medicine, one begins to perceive statistics as a different entity that is converted into a tool which allows obtaining results and hence benefits in any kind of study whose movements and relations due to its variable nature cannot be approached from a determinant perspective. Statistics can be defined as a science that studies how information should be used and how to provide a guideline for practical situations based on a question.

Statistics deals with the methods and procedures to collect, classify, summarize, and find regularities and analyze data, always and whenever variability and uncertainty are their intrinsic cause; and also to make inferences based on them, aiming to help make decisions and in this case to formulate predictions.

It is also considered the art of reaching conclusions based on a series of imperfect data. Data are commonly imperfect in the sense that even when they have useful information they do not tell us the whole story. It is necessary to have methods that will allow us to extract information based on the data observed to gain a better understanding of the situations they represent.

Everyone, even researchers, has problems in dealing with lists of data. There are various statistical methods whose purpose is to help highlight the outstanding and interesting characteristics of the data that can be used in almost all fields of knowledge.

The statistical method can and should be used at all stages of a research process, from the beginning to the end. There is a conviction that statistics deal with data analysis, however, this point of view excludes vital aspects related to research design. It is very important to be aware that the choice of method of analysis for a problem is based both on the type of data available and on the way in which they were collected.

What are biostatistics used for and why study it?

The answer is simple: because the statistical data and the conclusions obtained applying the scientific method exert a profound influence on practically all fields of human activity. Statistics, in particular, are increasingly invading any research perfomed in the field of medicine. This growth, probably related to the interest in increasing the credibility and reliability of the research, does not ensure that in all cases the statistical methodology has been correctly used, or even worse, that it is valid. Therefore, the importance of the correct application of biostatistics is a matter of concern because:

1. The conclusions may be wrong.
2. Not all readers have the capacity and the awareness to detect a mistake.

The study of biostatistics and the way of thinking generated from it, enables the person to evaluate objectively and effectively whether the information they receive (via tables, graphs, percentages, rates, etc.) is relevant, adequate, true and objective. The interpretation of any problem requires not only methodological knowledge but also profound knowledge of the subject. Even when a person is not interested in specializing in statistics, basic training on the subject enables a better understanding of the information.

Areas of biostatistics

Below is a brief description of each of the areas into which biostatistics can be divided:

1. Design: Planning and development of research.
2. Description: Summary and use of the data.
3. Inference: Predicting the specific characteristics of a sample based on its information.

Design

This is an essential. It consists of defining how the research will be developed in order to answer the questions that motivated it. Data collection usually requires a great effort, and therefore dedicating special care to the planning stage of the research saves work in the following stages. A well designed study is simple to analyze and the conclusions are usually obvious. A poorly designed experiment, or one with improperly collected or recorded data may be unable to answer the questions that motivated the research, no matter how sophisticated the statistical analysis. Even in cases in which already recorded data are studied, where we are limited to the existing information, the principles of good experiment design may be useful to help select a reasonable set of data related to the problem of interest.

Description

The methods of descriptive statistics or exploratory analysis of data help present the data in such a way that their structure is highlighted. There are several simple and interesting ways of organizing the data in graphs that will allow detecting both the outstanding characteristics and the unexpected characteristics. The other way of describing the data is to summarize them in one or two of the numbers that intend to characterize the set with the smallest distortion or loss of information possible. Exploring the data should be the first stage of any data analysis. Why not analyze them directly? First of all because computers are not very skillful (they are only fast), they do what are programmed for and act on the data we offer them. Wrong or unexpected data will be processed improperly and neither you nor the computer will realize this unless an exploratory analysis of the data is previously performed.

Inference

This refers to a set of methods that allow performing predictions about characteristics of a pheonmenon based on partial information about it. The inference methods allow us to propose the value of an unknown quantity or decide between two opposite theories which of them provides a better explanation of the data observed. The ultimate aim of any study is to learn about the populations. However it is usually necessary and more practical to study only a sample of each of the populations.

Therefore statistics can be classified as inferential and descriptive, the first when the object of study consist of deriving the conclusions obtained to a set of broader data and the second which we discuss in this chapter is when the results of analysis do not intend to go beyond a data set.

Descriptive statistics: This describes, analyzes and represents a group of data using numerical methods and graphs that summarize and present the information contained in them.

Inferential statistics: Based on the calculation of probabilities and using sample data it performs estimates, decisions, predictions and other generalizations of a larger set of data.

COMPLEMENTARY READING

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. Elsewier/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. Vol. 2". Alianza Editorial. Madrid.

CHAPTER 8

ANALYTIC OR INFERENTIAL BIOSTATISTICS

José María Jiménez Avila

When a research study begins, sometimes it is not possible to cover the entire population, and that is why a sample is usually selected, hoping that it will be representative of the population that is to be studied, although this is not always so, since sometimes it does not reflect the reality of the universe whence it comes. This is because there are conditions such as *chance*, which play a very important role in the process of research and consequently affect its results.

The branch of mathematics that studies the results of chance is the theory of probabiilty and this is the core of statistical *inference*.

Inferential or analytic analysis is performed after the descriptive analysis, where it is evidenced how and which are the characteristics of the population group and the specific characteristics of our sample.

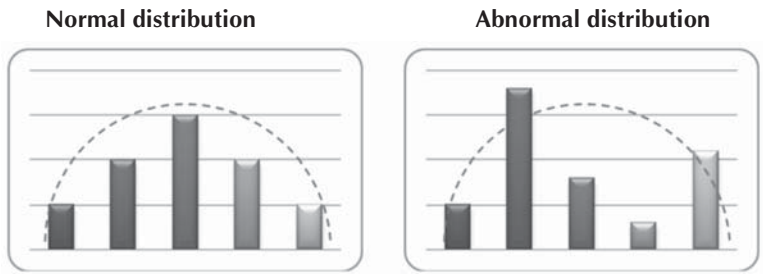
As a definition, *statistical inference* is the statistical procedure that allows establishing conclusions, taking into account the effects of probability (results) and it is usually in this stage that the hypotheses are accepted or rejected, and the estimates of patterns and intensity of the association between the variables, framing the result within the concept: "*are there differences or not*" among the groups.

When experiments are performed, a series of probabilities are derived, with possible results, and through the probability of the event it is possible to express confidence that the event will occur when the experiment is observed, and one can represent from "0" to "1" in such as way that "1" indicates that the event will certainly occur, while "0" corresponds to an event that will certainly NOT occur. There are 2 types of statistical inference: Hypothesis Testing and the Interval Estimation (they will be reviewed in later chapter).

Types of tests

Statistical procedures can be classified in 3 groups:
a) Those that assume Normal Distribution, which together are recognized as parametric tests.
b) Those that do not consider normal distribution of the data, which are identified as free distribution procedures (binomial and Poisson distribution).
c) Those that do not refer to population parameters that are known as non-parametric tests (Chi^2, Fisher's exact test). Figure 1.

Figure 1 – Normal and abnormal distribution of the data



Note: In some statistical tests both the non-parametric procedures and those of free distribution are grouped and presented under the title of "non-parametric".

There are a great number of tests that can be classified as 2 types: Parametric tests and non-parametric tests. Figure 2.

Figure 2 – Types of statistical tests

Parametric tests in surgery	**Non parametric test in surgery**
Student t	Chi^2
ANOVA	Mann-Whitney U
Pearson corrlelation	Kruskal-Wallis
	Spearman correlation
	McNemar
	Linear regression
	Poisson probability
	Odds Ratio

Parametric

Parametric tests are those that are considered statistically most powerful and provide more information, which allows estimating the population parameters using statistical samples.

Requirements:

- Variables with a normal distribution, that can be determined using the Kolmogorov-Smirnov test and the Shapiro-Wilk test.
- Not used for ordinal variables.
- A minimum of 30 persons is required per group in the sample, since the larger the sample, the more precise will be the probabilistic estimation.
- The hypothesis is based on numerical values, especially the mean population.

Non-parametric

The non-parametric or free distribution tests are easy to apply and can be used to analyze ordinal or normal variables, that is when it is not necessary to establish or propose an inference regarding which are the population parameters.

Requirements:

- Ordinal or nominal variables.
- When 2 series of observations are derived from distinct populations.
- The sample can be smaller than or equal to 20 persons.
- The hypothesis is made on ranks, median or data frequency.

It is important to know all these requirements and become familiar with all and every one of them, since the selection of the statistical test will depend on the type of variable, the presence or absence of normality of the distribution of the groups that are to be analyzed and how the research question was formulated (PICOT). Figure 3.

Figure 3 – Classification according to objective and variable

Statistical test according to objective and variable					
		Difference		Show relation	Outcome variables***
Type variable	**Type of sample**	**2Groups**	**3Groups**	**2 Variables**	**Variables**
Quantitative (normal distribution)	NR R	t Student* t Student**	Anova 1 Factor Anova 1 Factor	Pearson	Linear regression
Ordinal qualitative	NR R	U Mann-Whitney Wilcixon	Krusj=kal-Wallis Friedman	Spearman	
Dichotomic qualitative	NR R	X2 (Fischer) McNemar	X2	Coeficiente phi	Survival Curves Phi coefficient

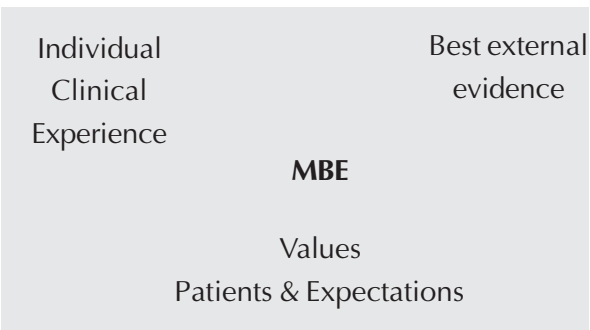
NR = not related R = related
* Student t for independent samples
** Student t for related samples
***The predictive variable may be quantitative, dichotomic or ordinal (transform dummy type variables)

Interpretation

This type of test gives one the tools with which the surgeon who sees the patient can prognosticate survival or the possibility that a surgery will become infected or the probability that some type of complication will occur. In this case the surgeon will base his estimation on his knowledge of what has been written in the literature, the characteristics of the patient himself, as well as his expectations and evolution, considering his own experience and the resources available to him for the care of the patient and possible outcome.

This allows applying the fundamental concepts of "*Evidence based Medicine"* which is considered the conscious, explicit and judicious use of the "*best scientific evidence*" available to make decisions about the patients (Figure 4).

Figure 4 – ¿How is Evidence Based Medicine performed?



COMPLEMENTARY READING

1. Celis de la Rosa A. Bioestadística.1ª 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. Gómez Gómez M, Danglot Banck C, Velásquez Jones L. Bases para la revisión de artículos médicos. Revista Mexicana de Pediatría. Vol. 68 N. 4. Julio/agos 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. 2ª 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.

Chapter 9

CENTRAL TENDENCY MEASURES

Alberto Aceves Pérez

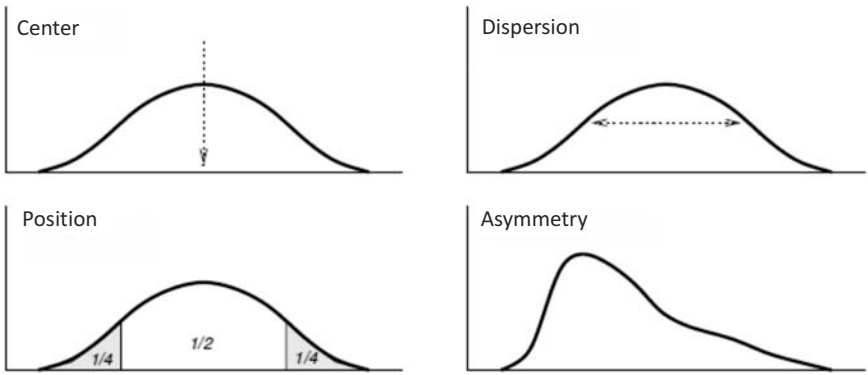
After elaborating the table and its graphic representation, it is mostly more effective to "condense" this information into a few number that will express it clearly and concisely.

Biological phenomena are usually not constant, and that is why it will be necessary, together with a measure that indicates the value around which the data are grouped, to associate a measure that will provide a reference to the varialbility that reflects this fluctuation. Therefore, the next step and objective of this chapter will consist in defining a few types of measures, statistical or parameters, that will summarize them even further. In other words, given a group of data organized as a distribution of frequencies (or a series of observations that have not been placed in order), we intend to describe them by two or three synthetic quantities.

For this several characteristics can be examined, the most common of them being:

1. The central tendency of the data.
2. The data that occupy certain positions.
3. Data symmetry.
4. The form in which the data are grouped.
5. The dispersion of data in relation to the center.

Figure 1 – Measures representing a set of statistical data



In this chapter and following this order, we will study the statistics which will guide us about each of these levels of information, values around which is grouped the sample, the greatest or smallest fluctuation around these values. We will look at certain values that mark positions characteristic of a frequency distribution, as well as their symmetry and form.

The center is easy to identify if the distribution is symmetrical, but it is difficult if the distribution is asymmetrical. For this reason, there is no single measure of position to summarize a distribution. If the distribution is symmetrical different measures will lead to similar results. If the distribution is clearly asymmetrical different proposals will point to distinct concepts of "center" and therefore the values will be different.

The three most usual measures of the central tendency are:

1. Mean
2. Median
3. Mode

On certain occasions, these three statistics agree, although this does not generally happen. Each of them presents advantages and disadvantages that we will need further on. First of all lets define the previousl concepts.

The mean

This is the most frequently used measure of position. In order to calculate the arithmetic mean or average of a set of observations all values are added up and divided by the total number of observations.

If we have a sample of n observations and denoted by X1, X2, ..., Xn, we define the sample mean X as follows:

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

Example:

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

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

Characteristics of the mean

1. It is used for numerical data.
2. It represents the center of gravity or the point of equilibrium of the data.
3. The sum of the distance from the data to the mean is zero.
4. It is very sensitive to the presence of atypical data.

Although the mean is the simplest measure with a central tendency, other measures offer us more information and are occasionally more appropriate.

The median

The median is the datum that occupies the central position of the sample arranged from smallest to largest..

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

Example:

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

We place the data in order

10 11 12 14 18 23

Position of the median $\Rightarrow \frac{6+1}{2} = 3.5$

We obtain the median averaging the third and fourth datum:

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

Characteristics of the median:

1. As a descriptive measure, it has the advantage of not being affected by the extreme observations, since it does not depend on the values taken on by the variable, but instead from the order in which they are arranged. Therefore it is adequate to use in asymmetrical distributions.

2. It is quick to calculate and simple to interpret.

3. Differently from the mean, the median is always a value of the variable that we study.

The mode

The mode is the datum that occurs most frequently in the set.

It is a measure that is not very useful except for categorical data for which it is of interest to identify the category with the greatest amount of data. In a sample of numerical data, it may occur that the mode is a value that is repeated a certain number of times, but that is not typical.

Characteristics of the mode:

1. It is very easy to calculate.

2. It may not be single.

Relation between mean, median and mode

In the case of unimodal distributions, the median often lies betwen the mean and the mode (in fact, closer to the mean). In distributions that present a certain inclination it is more advisable to use the median. However, in studies related to statistical purposes and inference it may be more appropriate to use the mean.

Figure 2 – Summary of the measures with a central tendency

	CENTRAL TENDENCY MEASURES	
	NON GROUPED DATA	GROUPED DATA
	(ordered) $x_1, x_2, \ldots, x_N$	$\begin{array}{cccc} \text{Interv.} & x_i & n_i & N_i \\ \hline l_0—l_1 & x_1 & n_1 & N_1 \\ l_1—l_2 & x_2 & n_2 & N_2 \\ \ldots & \ldots & \ldots & \ldots \\ l_{k-1}—l_k & x_k & n_k & N_k \end{array}$
MEAN	$\overline{x} = \frac{x_1 + \cdots + x_n}{N}$	$\overline{x} = \frac{n_1 x_1 + \cdots + n_k x_k}{N}$
MEDIAN	First observation that leaves below itself strictly the [N/2] smaller observation $x_{[N/2]+1}$	$M_{ed} = l_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} \cdot a_i$
MODE	$M_{oda} = x_i$ with greater frequency	$M_{oda} == l_{i-1} + \frac{n'_i - n'_{i-1}}{(n'_i - n'_{i-1}) + (n'_i - n'_{i+1})} a_i$

COMPLEMENTARY READING

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. Elsewier/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.

7. Peña Sánchez de Rivera, D. (1987): "Estadística. Modelos y Métodos. Volumen 2". Alianza Editorial. Madrid.

CHAPTER 10

MEASURES OF DISPERSION

(Range, Variance, Standard Deviation, Coefficient of Variance)

Delio Martins

An important aspect of the descriptive study of data in a study is to determine the variability or dispersion of these data in relation to the measure of location of the center of the sample.

For instance, in a study about scoliosis, in which the effect of a given surgical treatment A was evaluated, compared to another surgical treatment B in children with different degrees of curvature. The mean of the magnitude of the curves was 60 Cobb degrees, both in the group submitted to surgery and in the control, and both groups presented an equal mode and median. Could this be sufficient for me to be able to say that these groups are similar? Let us see a few important concepts.

Range

This is a measure based on the largest and smallest value in a set of data. Thus, if we consider the example above regarding the degree or curvature of scoliosis in children and we observe the 2 groups with the following magnitudes of curves:

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

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

Amplitude is the difference between the highest and the lowest observation. In this case we observe that in treatment A the amplitude was 41 degrees, while in treatment B it was only 29 degrees showing a great variation among the groups. Certainly amplitude is the most easily visualized measure of dispersion, however it is based only on two values and does not take data distribution into account.

Variance

Let us see another measure of dispersion, Variance. In order to simplify our understanding, we will reduce our sample:

Treatment A: 40, 50, 65, 81

Treatment B: 48, 54, 66, 68

In both cases we have a mean of 59 Cobb degrees.

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

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

The amplitudes remain different:

A = 81 – 40 = 41

B= 68 – 48 = 20

We note that the treatment A group is more dispersed that the treatment B group. In order to begin understanding the standard deviation, we first calculate from Group A, the distance of each observation is from the group mean. Thus we will have:

(40 – 59, 50 – 59, 65 – 59, 81 – 59)

In this way we have the values:

(-19, -9, 6, 22)

Now a test: calculate the arithmetic mean of these numbers....

Have you finished calculating it?

Have you observed that the sum of all deviations in relation to the arithmetic mean is equal to zero? This is an arithmetic property that will always be repeated!

In this way the negative values will always annull the positive one, and it is **not** possible to obtain a measure of dispersion using this algebraic calculation!

Let's try another way then.

Since in calculating we only care about the magnitude of the deviation and not if they are positive or negative, we shall do so in a way that all numbers remain positive! Raise all the numbers to the square!

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

or:

361, 81, 36, 484

In this way the mean of squares of the numbers becomes a reasonable measure of variability. However, to calculate the mean of these values in such a way that there are properties that function with all the theories (believe me here: this is more than you are going to want to know on this subject), we do not divide by the total number of cases (here: n=4), but rather divide by the total number – 1 (n-1) which in this case is: (4-1=3). Thus:

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

Thus we calculate the sample variance. Let's go on to the next step.

Standard Deviation

Standard deviation may be the most used dispersion mean in statistics, but it is also the least understood. However, now that we understand variance, we will find it easier. As its name says, it is a standard of deviation from the mean. It represents the normal distance from any point in the set of data to the mean.

Standard deviation is simply the square root of variance. Thus we avoid having to work with the squares of values. In our example, then, the standard deviation is $\sqrt[2]{320,6} = 17,90$.

Standard deviation of the entire population is represented by the Greek letter σ, but since we rarely know the entire population, and normally we work with a sample, the standard deviation of a sample of the population is designated by the letter S.

More important than calculating the standard deviation is to understand its significance. A small standard deviation means that the values of the set of data are, on average, close to the center of the set. A high standard deviation means that the values of the set of data are on average distant from the center.

A high standard deviation does not necessarily means something bad. It only reflects a great variability in the group evaluated. Another thing that one must look out for is to always observe the unit used for the standard deviation. For instance, if one considers a standard deviation of 2 years in a given sample, this could change to a standard deviation of 24 if the data are in months!

On a normal distribution curve present in most scientific work, we have as a standard that 68% of the sample is up to 1 standard deviation away from the mean and that 95% of the sample is up to 2 standard deviations from the mean. Thus, for instance, if we have a calculation of the quantity of proteoglycans present in the intervertebral disc, and we are informed that the mean of proteoglycans was 250 mg/g and that the standard deviation was 8.75 mg/g in this sample, we know that 68% of the discs present a proteoglycan dosage within 1 standard deviation (8.75 mg/g) from the mean (250 mg/g). Thus, 68% of the discs contained between 250 – 8.75 mg/g and 250 + 8,75 m/g of proteoglycan, ie, 68% of the discs in the sample contained between 241.25 and 258.75 mg/g of proteoglycans. Likewise, 95% of the samples measured have between 250 – 2(8,75 mg/g) and 250 + 2(8.75 mg/g), in other words 95% of the samples contain between 232.5 and 267.5 mg/g of proteoglycans.

A few important properties of the standard deviation:

– It can never be a negative number (due to the way it is calculated, and since it measures a distance; and distances are never negative numbers).

– The smallest possible value is zero, and this happens only in planned situations, where each number in the set of data is equal – There is no variation.

– It is influenced by discrepant values within the set (either very high or very low). Remember, this is because the standard deviation is based on the distance of the data from the mean.

– The standard deviation is in the same unit as the original data.

Without the standard deviation you cannot know whether the data are close to the mean, or whether they are very spread out. For instance, if the reimbursement for a surgery that lasts 1 h on average is $ 7000.00, you may think: Wow! Wonderful! But what if the standard deviation to reimburse the surgery was $ 2000.00 and we use the empirical rule of normal distribution? This surgery can be reimbursed between $ 3000.00 and $ 11,000.00 (i.e., $ 7000.00 more or less two standard deviations). In this way you notice that there is a very great variability in the reimbursement of this surgery, thus, the mean that was informed alone does not really reflect well what is to be expected from this procedure.

Coefficient of Variance

Finally, we will understand the coefficient of variance. When we compare the variability of a measure between two groups, the standard deviation can be employed adequately. However, if the groups have very discrepant means, the direct comparison of the standard deviations can induce erroneous conclusions, especially when one expects the group with the largest mean to have the greatest variability.

For instance, if we hypothetically compare the height of the vertebrae of adults which, we assume, have a standard deviation of 5.9 mm, to the height of vertebrae of newborns who have a standard deviation of 4.0 mm. The standard deviation is larger in adults, which suggests that its variability is greater than that of newborns. However, adults have much larger vertebrae, we suppose that they are around 50 mm, whereas the newborn have much smaller vertebrae, we assume that around 25 mm. Thus, it will be better to compare the variation of the height of the vertebrae in relation to the mean. In this way, the coefficient of variation can be used, which is simply the standard deviation expressed as a percentage of the mean.

Therefore the formula of the coefficient of variance (CV) is:

$$CV = \frac{\text{Standard Deviation X } 100}{\text{Mean}}$$

The CV is independent of a unit of measure and it is expressed in percentage. In this example, the CV of adults is 11.8% and the CV of newborns is 16%. This shows that the height of the vertebra in the newborn is more variable than that of an adult, even if less variable in absolute terms.

In brief, in the study of the descriptive data of a sample it is necessary to evaluate together the measures of the central tendency and the measures of dispersion, in order to have a real analysis of the data.

COMPLEMENTARY READING

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. Rio de Janeiro: Alta Books Editora. 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. Rio de Janeiro: Alta Books Editora. 2012.

6. Triola, MF. Introdução à Estatística. 7ª ed. Rio de Janeiro: LTC, 1999.

CHAPTER11

DATA PRESENTATION

José María Jiménez Avila

Data presentation is one of the parts of a manuscript most used in statistics, since based on different types of graphics one can evidence and analyze results speedily.

A table or a well-designed figure clearly shows the greatest number of ideas in the smallest amount of space. It should mimic the format of tables or figures used by the journal that has been chosen for the next publication (titles, lines, colors, variables and decimal numbers). Also number the tables and figures consecutively and draft a brief, informative and precise title. Do not forget to elaborate footnotes to the tables and figures with notes and abbreviations.

When analyzing the results one should remember the objective of the research. Therefore data should be selected that by themselves reflect the result of the message that is to be imparted, that are easy to understand and will show the relationship between the data.

It should include:

a)Title.

b) The figure itself.

c) Explanatory notes.

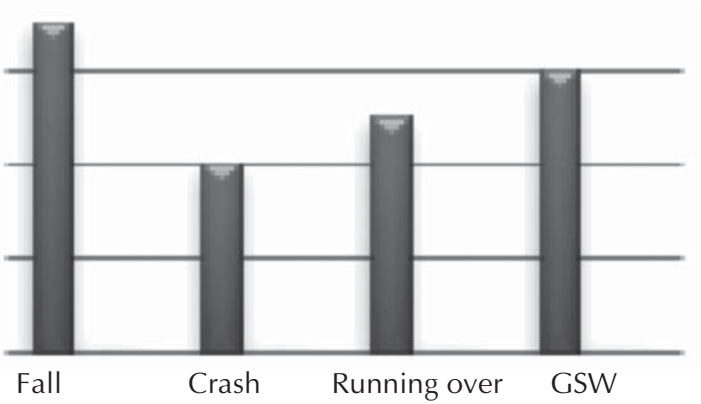
Usually the total number of tables and figures recommended for a scientific article is 6, for instance 2 and 4, 3 and 3, etc. Do not repeat the information that you present, ie., do not repeat in the text what is described in the tables and/or figures. The quality of the presentation of results is very important for adequate understanding and interpretation of the results that are obtained in research.

Type of Graphic Representation

Bar graph

This type of graphic representation is very useful to evidence discrete and qualitative variables on a nominal scale. It is suggested that the bars have the same width and they must be separated by a space that should not be greater than the thickness of the bars. The suggestion is that they should be presented from largest to smallest unless there is a predetermined sequence in which they are to be presented.

Example:



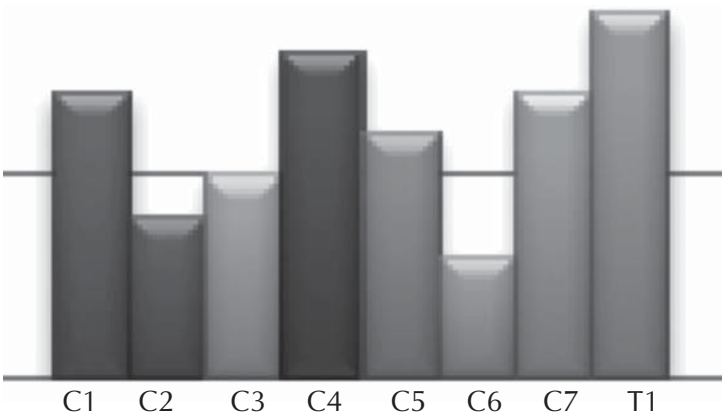
Application in surgery

Distribution of 60 patients who have suffered a lumbar spine fracture according to the injury mechanism. Jalisco, 2014-2015.

Histogram

This type is appropriate for continuous quantitative variables, because it can group the frequencies of presentation of the variables into intervals.

Example:



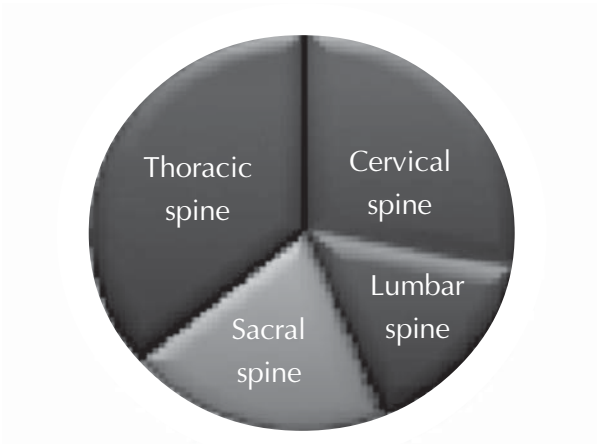
Application in surgery

Distribution of 100 patients with cervical and high thoracic fracture. Jalisco. 2014-2015.

Pie graph

This type of representation is adequate for qualitative variables on a nominal scale.

Example:



Application in surgery

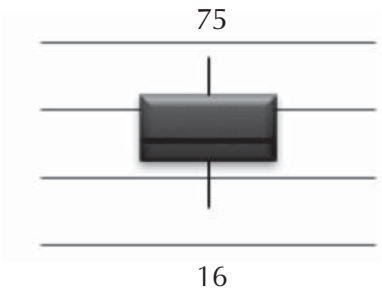
Distribution of 60 patients with spine fracture according to anatomical site. Jalisco. 2014-2015.

Note: Be careful when you use this type of graph since it may show very obvious evidence, which at some point might be left out of the manuscript when it is not part of the main message.

Box plot

This is appropriate for quantitative variables in which we want to demonstrate the deviation from the normal distribution. Usually the superior extreme of the box corresponds to what is called percentile 75, the inferior extreme to percentile 25, and the horizontal line that will divide the box will correspond to the mean or average.

Example:



Application in surgery

Distribution of ages in a group of 60 patients with spine fracture. Jalisco, 2014-2015.

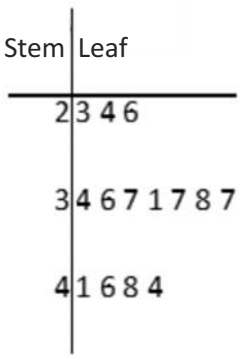
Stem and Leaf Diagram

This is useful for quantitative variables that have few values, because in this type of presentation each of the values of the variables represented

is grouped in relation to its distribution, and may be similar to the histogram.

Random values are ranked from largest to smallest and the leaf of each column corresponds to the last value of each reading and every time it is repeated a note is made, forming the equivalent of the histogram column and the stem corresponds to the rest of the reading and is only written once for each region.

Example:



Application in surgery

Distribution of type of fracture (subscale) in a group of 14 patients. Jalisco, 2014-2015.

The first presentation should usually include the demographic data of the group studied, where the general characteristics are observed objectively, that is, the descriptive analysis of the sample.

COMPLEMENTARY READING

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, N. 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, N. 2. 2011.

Chapter 12

HYPOTHESIS TEST (CI) AND VALUE OF "p"

José María Jiménez Avila

Description of the test

A hypothesis test is defined as a statistical test that can be used to determine whether there is sufficient evidence in a data **sample** to infer that a certain condition can be valid for the entire **population**.

In other words, a test can examine 2 opposing hypotheses: the null hypothesis (H_o) and the alternative hypothesis (H_a). The first is the one that will be tested. Generally the null hypothesis establishes an enunciation that "there is no effect", or "there is no difference", and the alternative hypothesis is the enunciation that one wishes to be able to conclude is true, "there are differences".

Because of the data of the sample, the test can determine whether the null hypothesis should be rejected, and in order to make the decision a value called "p" is used, if this is < (smaller) than or equal to the level of significance, which is a cutoff point that is defined and then the null hypothesis can be rejected.

A common error of perception is that the statistical hypothesis tests are designed to select which is the more likely of 2 hypotheses. This means that a test will maintain the validity of the null hypothesis until there is sufficient evidence (results) in favor of the alternative hypothesis.

Example:

Does minimally invasive surgery have better results than open conventional surgery?

Two types of tests can be used for this:

a) Hypothesis Test .

b) Estimation (Confidence interval).

Hypothesis Test

When a hypothesis is formulated in which one wishes to compare 2 populations and it is not possible to study the total universe, a sample of each group is taken based on inferring the parameters of both groups. This is done so that the information will be sufficient to make a correct decision about the "equality", or "difference" of the 2 groups.

Several aspects should be assumed for this:

1. Both populations are identical and there are no differences between the groups.

2. Selection of the level of significance, using a statistical test that can be the value of "p".

a) $p>0.05$: The null hypothesis is accepted. That is "*there is no difference* (Effect 0).

b) $p<0.05$: The null hypothesis is rejected, that is there "*are differences*", and one define, as the probability that the differences are significant and that they are not due to chance (Effect 1).

Notes of interest

When one assumes the null hypothesis, the idea is that there are no more differences except those due to chance.

Probability implies uncertainty, no matter how small the value of "p", we will never be sure of counting on the absolute truth, having obtained $p<0.0000001$, only tells us that the probability that Ho is right is minimal, but will never be null.

From the surgical medical perspective, that a test is "statistically significant does not mean the same as being clinically important".

Estimation (Confidence interval)

It is associated with what is known as confidence interval (CI), which is the interval in which we know that there is a parameter with a specific level of confidence. It is translated as the probability that the parameter to be estimated is found in the confidence interval (the most used are 95% or 99%). For instance, it is interpreted that if one had a population with an average age between 40.2 and 53.6 and we selected 100 samples, 95 will cover the true value of the mean of the population, and 5 will not. That is, the approximation to the parameter

corresponding to the population from which the sample was extracted is estimated.

Types of error

There are certain risks that can provoke some errors when the inferences are made and identified. These are called type I and II error.

a. Type I error: Rejecting the null hypothesis when the null hypothesis is true: The probability of making this type of error is precisely á. Therefore the error is called error á.

b. Type II error: Accepting the null hypothesis when the null hypothesis is false. Known as acceptance error or â error:

Example:

1. The defendant was innocent and the jury declared him innocent: OK, all is well: good job: the individual is released and there is no problem.

2. The defendant was guilty (he killed her!!), and the jury (they are very clever!!) ddeclares him guilty: life sentence; and everything is OK (he deserved it!).

However, the jury can make two types of error:

3. To declare the defendant guilty and poor guy was innocent! Type I.

4. The jury declares the defendant innocent and the nasty wretch was guilty: Type II error I.

1. If the value of "p" tells us that the probability of having found this value by chance is very low, we shall reject the null hypothesis: we shall say that the difference is statistically significant: *If we are wrong we shall be committing a type I error.*

2. If the value of p is high: we shall say that the probability of having found this value by chance is high. Therefore, we shall accept the null hypothesis: we shall say that the differences found were not statistically significant: if we made a mistake (and IF there were differences), we shall be committing a type II error.

The value of "p" and the level of significance

This is the probabiility that chance alone can produce a difference between the groups compared. If the value of "p" is large ($p > 0.10$ or $p > 0.25$), the difference observed between the incidences *may well be due to chance,* and therefore the association will be considered "*non significant*". If the value of "p" is small ($p < 0.05$ or $p < 0.01$) the difference observed has a very small probability of being due to chance and is almost certainly due to a real difference between the incidences of the populations that are being studied.

Example in spine surgery

You wish to determine whether minimally invasive surgery has a beneficial effect as opposed to open conventional surgery, in patients who have low back pain caused by a disc herniation.

1º Formulate the research question (PICOT).

2º Establish a position based on the possible comparison of the 2 options.

3º Transform the question into 2 hypotheses.

Null hypothesis – H_o: The results are the same.

Alternative hypothesis – H_a: The results are different.

4º Obtain the value of "p" using statistical tests.

This will give 2 possibilities $p>0.05$.

$p<0.05$.

Interpretation

The most common way is to compare the value of "p" to the level of significance (α). α is the probability of rejecting H_o, when H_o is true.

The value of "p", indicates the degree to which the evidence of the sample suports the rejection of H_o. Generally the smaller the value of "p", the greater will be the influence of the evidence of the sample to reject Ho. More specifically, the value of "p", is the smallest value of α which leads to the rejection of Ho. For any value of $\alpha >$ the value of "p", cannot reject Ho and for any value of $\alpha \leq$ the value of "p", rejects H_o.

In the example of minimally invasive surgery, where one attempts to compare a few aspects related to the improvement compared to open conventional surgery, in the hypothetical case in which an average

of 25 patients presented better results, the statistical test will depend on the mean and if the value of "p" were, for instance .026, this indicates that 2.6% of the samples with improvement in the result extracted from the population where $\mu = 25$, will produce a mean that provides evidence that is as strong as, or stronger than, the current sample μ is not equal to 25. Now ask yourself what is most probable: that $\mu = 25$ and you simply have selected a very unusual sample or that μ is not equal to 25?

Traditionally, the value of "p" is compared to values of α smaller than .05 or .01, depending on the field of study.

In the example, let us assume that the value of α is .05. The value of "p" is .026, This indicates that the mean (not only the mean of the other patients included in the study) probably is not equal to 25. A more correct way of signalling this in statistical terms is: "at a level of significance of .05, the mean improvement through this type of surgery appears to be significantly different from 25".

This is interpreted as follows: "Improvement of results using minimally invasive surgery" $---p < 0.05$.

The use of minimally invasive surgery has a greater probability of achieving better results that open conventional surgery. For this reason it can be assumed that there may be an association that is not due to chance and that will probably generate the result mentioned. Using the values of "p" is easy if one knows the key data: the values of α that are acceptable in their field and the null and alternative hypothesis for the tests that will be used.

COMPLEMENTARY READING

1. Celis de la Rosa A. Bioestadística.1ª 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 N. 2, 2011.

CHAPTER 13

KOLMOGOROV-SMIRNOV TEST

Nelson Astur Neto
Maximiliano Gimenez Gigón
Rodrigo Góes Medéa de Mendonça

Description

The purpose of the Kolmogorov-Smirnov test (KS) is to determine whether two sets of data differ significantly. The advantage of the KS test is that it does not make assumptions about the data distribution. Technically it is a non-parametric test without distribution. On the other hand the disadvantage is that other tests may be more sensitive if the data take the test requirements into account.

In statistical research the parametric and non-parametric tests or hypothesis tests are applied. Among the non-parametric tests that are commonly used to verify whether a distribution is or not adjusted to an expected distribution, particularly the normal distribution is the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test is very powerful with large samples. The level of measurement of the variable and its distribution are elements that intervene in the selection of the test, which is to be used in later processing. In fact, if the variable is continuous with normal distribution, parametric techniques can be applied. If it is a non-normal discrete or continuous variable only non parametric techniques can be used, because using the former would lead to results with doubtful validity.[1,4,3,5]

Formula

Test Hypotheses:

H_0: The data analyzed follow an M distribution.

H_1: The data analzyed do not follow an M distribution.

Compare the empirical distribution of the sample to the distribution proposed in H_0

If this comparison shows significant differences, H_0. Is rejected.

Test statistics: [2,4]

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

Where:

- is the i[th] value observed in the sample (whose values have been ordered previously from smallest to largest).
- is an estimator of the probability of observing values smaller than or equal to xi.
- is the probability of observing values smaller than or equal to xi when H_0 is right.

Thus, *D* is the largest absolute difference observed between the accumulated frequency observed $F_0(x)$ and the theoretical accumulated frequency $F_0(x)$, obtained from the probability distribution which is specified as the null hypothesis. If the values observed $F_N(x)$ are similar to those expected $F_0(x)$, the value of *D* will be small. The larger the discrepancy between the empirical distribution $F_n(x)$ and the theoretical distribution, the greater will be the value of *D*. Therefore, the criterion for decision-making between the two hypotheses will be in the form[3]:

Si $D \le D_\alpha \Rightarrow$ Accept H_0
Si $D > D_\alpha \Rightarrow$ Reject H_0

Application

Applications of the test [4,5]: Test whether a set of sample data can be considered as coming from a given distribution. An alternative to the Chi[2] test when the model proposed according to the null hypothesis is of the continuous type and the sample size is small.

Advantages of the Kolmogorov-Smirnov test compared to the Chi[2] test:

- It does not require grouping the data into classes.
- It is applicable to small samples.

Inconvenient aspects of the Kolmogorov-Smirnov test compared to the Chi –square test:

- It is only valid for models of the continuous type.

Development of the test:

Let $X_1, X_2, ..., X_n$ be an s.r.s. of an r.v. X with a continuous type distribution.

Contraste:

H_0:X follows the distribution F

H_1:X does not follow the distribution F

Perform a Kolmogorov-Smirnov test at the level of α =0.1 to test whether one can assume that the 10 data: **10.5, 8, 15, 12.1, 4.1,12.1, 6, 10,5, 16** are from a normal distribution N(10.84, 3.5)

1.Once the data of the sample are ordered, we construct the table with values D_1

$x_{(i)}$	$\hat{F}_n(x_{(i)})$	$F(x_{(i)})$	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

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

$\max\{|0.1-0.209|, |0.3-0.209|\}$

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

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

4. Conclusion: 0.161<0.368, therefore it is not rejected that the data come from an N distribution (10.84;3.5).

Application to spine surgery

In a recent study, published by the Spine journal in 2009, we were able to verify the use of the aforementioned KS test.

The variation between the anatomical form of the articular facets and their variation during the growth of Danish eutrophic children was studied.

The author, Masharawi and his collaborators at the University of Tel Aviv, together with the University of South Denmark, measured all lumbar facets (L1-S1) of 100 children (51 boys and 49 girls) using magnetic resonance, both for their length and their width. These measurements were performed at the age of 13 years and 3 years later, using software appropriate for these cases. More than 4400 measurements were performed. From this the importance of a test that can verify whether we are looking at a normal sample is inferred. This shows the importance of the KS test. The aforementioned study concluded that these children's facets continued to grow after the age of 12 years. Any asymmetry should be considered non-normal and clearly shows the need to manufacture more personalized implants, taking into account the age ranges for this population.

COMPLEMENTARY READING

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.

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. 4ª 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>.

Chapter 14

SHAPIRO-WILK TEST

Nelson Astur Neto
Rodrigo Góes Medéa de Mendonça

Test definition

The Shapiro-Wilk test is a normality test in frequentist statistics. It was published in 1965 by Samuel Shapiro Sanford and Martin Wilk.[1] Normal distribution is one of the most important probability distributions in statistics, known also as the Gauss or Gaussian Distribution.[2] The test rejects the hypothesis of normality when the value of p is less than or equal to 0.05.

There are several tools and statistical programs to verify Normality using the *Shapiro-Wilk (S-W) test,* including the Statistical Package for Social Sciences (SPSS) version 23.0 (SPSS, 2016). SPSS is used to analyze the data and allows manipulating, transforming, creating tables and graphics that summarize the information obtained. Its potentials go beyond the descriptive analysis of a set of data. Using this software it is possible to perforrm more advanced procedures such as statistical inference, hypothesis tests and multivariate statistics for qualitative and quantitative data.[2]

Formula

The Shapiro-Wilk test proposed in 1965 is based on the W statistic given by:

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

Where an-i+1 are constants generated by the means, variances and covariances of the order statistics of an n size sample of a Normal distribution.[3]

Examples of spine surgery

As an example of the Shapiro-Wilks test applied to the spine we have the work developed at the University of Pennsylvania together with the orthopedics and radiology groups, as well as the department of statistics, and published in 2014 by the American Academy of Physiatry and Rehabilitation.[4] Billy and collaborators proposed to study whether there could be differences between the heights and diameters of the lumbar discs, before and after prolonged periods in the sitting position and without breaks during four consecutive hours. The study design was prospective in ambulatory patients using Magnetic Resonance for the measurements. On the first day, the patients were analyzed at the beginning of the work day and seated, after 4 hours; on the second day, already, the same 12 patients were studied, also at the beginning of their activities, but they were submitted to changes and a stretching protocol very 15 minutes until the end of the same 4 h.

In this example the Shapiro-Wilk test was decisive to know whether the sample was in normality, because being normal the next step would be to perform a paired T test and if normality were not satified, the tests of Wilcoxon and Rank would be used. Also about the results of this work, a statistical difference was found between the heights of the discs in those who remained seated all the time. The clinical applicability of this research lies in the fact that small changes in the disc height can be correlated with improved lumbalgia and therefore with its consequences.[4]

Application in surgery

When we hear the specific name of some statistical test such as Shapiro-Wilk, it is difficult to imagine how we could use it in our favor, so much the more when we do research with populations submitted to some form of intervention, such as spine surgery. As an example to illustrate this, we have the research performed by Farjoodi et al at Johns Hopkins Hospital in the United States, where the effect of the surgical volume of the hospital and the surgeon were evaluated in complications after lumbar spine surgery. The article was published in the periodical Spine in 2011.[5]

In this case, the function of the Shapiro-Wilk test was to verify the normality of the variables and to separate those that need to be adjusted in order not to become confounding factors, such as the demographic variables. The number of hospital procedures studied was 232,668, listed as posterolateral decompression with arthrodesis and/or exploration/decompression of the vertebral canal. After applying this statistical tool, the authors concluded that the mortality and rate of complications associated with lumbar spine surgery were lower when patients were treated by surgeons and hospitals working with a high volume.[5]

Interpretation

The S-W test supplies the parameter test value (p value or significance) which can be interpreted as the measure of the degree of concordance beween the data and the null hypothesis (H_0), H_0 corresponding to the Normal Distribution. The lower the p value, the smaller is the consistency between the data and the null hypothesis. Thus, the decision rule adopted to know whether the distribution is Normal or not is to reject H_0: (i) if, one rejects H_0, in other words, one cannot accept that the set of data inolved has a normal distgribuiton; (ii) if, one does not reject H_0, ie, the Normal distribution is a possible distribution for the set of data involved.[2] If the normality test fails, it can be stated with 95% confidence that the data do not fit the normal distribution. Passing the normality test only allows one to declare that no signfiicant deviation from normality was found.[4]

COMPLEMENTARY READING

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 LumbarSpine surgery. Spine (Phila Pa 1976). 2011 Nov 15;36(24):2069-75.

6. <http://www.variation.com/da/help/hs141.htm>.

Unit 2

Chapter 15
Student *t* test
Barón Zarate
Francisco López Meléndez

Chapter 16
ANOVA test
Juan Pablo Guyot

Chapter 17
Pearson correlation
Alisson R. Teles
Asdrubal Falavigna

Chapter 15

STUDENT *t* TEST

Barón Zarate
Francisco López Meléndez

First we should explain that there are two large groups of statistical tests, which are based on the distribution of the sample or of the population. There are the parametric and non parametric tests; the difference between parametric and non parametric tests is that parametric tests take on the parameters of the distribution of a variable (mean and variance) and their distribution is normal. The non-parametric tests on the other hand do not assume anything about the distribution and also do not concern themselves with the type of sample distribution, they only work with organization and recounting the values of the variable, without concern about the distribution (*and therefore they are also called of free distribution).*

Sample characteristics to know what kind of test can be performed.

Parametric:

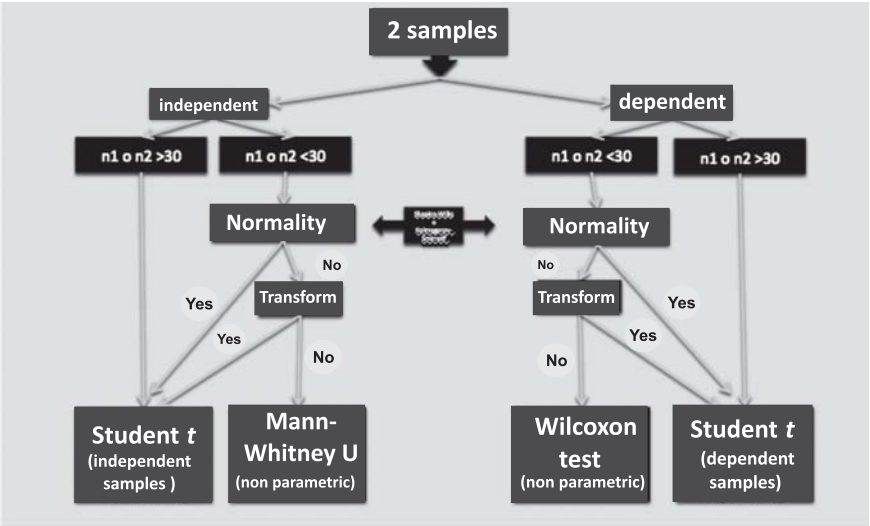
- The variables must be *quantitative*
- Use interval scales or ratio
- The variables must follow a **normal distribution**
- The variances must be equal
- **Large samples (n>30)**

Non-parametric:

- Variables with free distribution (**not Normal**)
- If they are quantitative, ordinal or nominal variables
- Groups with large variances
- **Small samples (n<30 or <10)**

In biostatistics one of the most frequent question is to evaluate the statistical significance of the differences between samples of individuals, whether they be from a same group (*e.g. Before and After*), or from different groups (*e.g. Treatment 1 vs. Treatment 2*), and this we will call paired samples or independent samples, respectively *(Figure 1).*

Figure 1 – Flow diagram to decide the type of statistical test in order to contrast the hypothesis in two samples (dependent or independent)



Student is the pseudonym of William Sealy Gossett, an English statistician and mathematician who worked for Guiness Brewery and devised the t test or Student *t* test in 1908 to value the distribution of small samples. It is based on the small sample theorem and on the central limit theorem.

The Student t test is generally used to compare *normally* distributed continuous samples (parametric test) between two groups or between groups; for this reason there are two types of Student *t* , one for independent samples and another for dependent or paired samples.

We will now discuss the Student t test for paired samples (**related samples**). It is said that there is pairing when:

• Measurements are performed in a same individual (e.g. Before and after an intervention) and it is called auto-pairing.

• Two natural pairs are compared to each other(e.g. twins) and this is natural pairing.

• A control is sought for each case, which is called Cases and Controls, looking for equal characteristics (sex, age, etc), and this is called artificial pairing.

Note: Selecting a statistical test to contrast hypotheses is different when dealing with independent samples (e.g. Two types of individuals) and paired samples (e.g Same individuals observed twice), due to the fact that random variability between individuals (inter individual) does not exist or is smaller in related samples (intra individual variability), and the same occurs in normal and non normal distribution.

The Student t formula is:

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

Or

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

Next it will be developed with an example of a practical application of a Student t test for related data.

Interpretation in spine surgery

An everyday example where a Student t test can be used for related samples would be to evaluate the quality of life (on a validated continuous scale from 0 to 100, where the highest is the value concerning better quality of life), in an **n** number of patients who were submitted to a surgical procedures (for instance, L4-L5 arthrodesis), where preoperative measurement will be performed and another postoperatively, to then evaluate whether there is a postsurgical change. **Objective**: to evaluate whether there is a statistically significant change in quality of life after the surgical treatment. (**Table 1**).

Patient	Preoperatively (SF-36)	Posoperatively (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

Table 1

Our null hypothesis **(H_0)** would be that there is no significant difference between the preoperative measurement and the postoperative measurement.

$$H_0 \cong \mu \text{ before } = \text{ after}$$

And our alternative null hypothesis **(H_1)** would be that there is a statistically significant difference between both measurements.

$$H_1 \cong \mu \text{ before } \neq \mu \text{ after}$$

The following steps are taken to perform a Student t test for related samples:

1. The first step will be to create a new variable (***DIF***), which is the difference between the value of "before" and "after", and this new variable tells us how much an individual has changed between the two measurements, and then the mean (***μ***) of variable ¨*DIF*¨ must be calculated .(Table 2).

Patient	Preoperatively	Postoperatively	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**
			Mean:19.58

Table 2

2. If the sample is smaller than 30, we should test whether the new variable follows a *normal distribution*. If there is no normal distribution it can be transformed logarithmically or a non-parametric test (**Wilcoxon test**) can be performed. In the SPSS program both normality tests are performed, Kolmogorov-Smirnov and Shapiro-Wilks.

> **NOTE:** The Shapiro-Wilks test is considered the most powerful test for samples of less than 30 cases.

3. The next step after calculating the mean of the variable (DIF), is that it is necessary to calculate its standard error (**EEMDIF=** standard error of the mean of the variable "difference").

The formula to calculate EEMDIF is by dividing the standard deviation of the population (s) by the square root of sample size (η).

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

EXAMPLE:

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

$= 9.98/\sqrt{14}$

$= 9.98/3.74$

EEMDIF = 2.66

4. Then it is necessary to calculate the degrees of freedom ***(Df= η-1)*** where h is the total number of individuals of the sample, and 1 is constant for the Student t test of related samples.

Df $= \eta$ -1

Df = 14-1= **13**

5. Calculate the Student t test for related samples; this is the result obtained by dividing the mean (μ) of variable DIF by its standard error (**EEMDIF**).

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

EXAMPLES:

$t_{13} = \mu$ *DIF / EEMDIF* $= \mu$ *DIF* / ($\sigma / \sqrt{\eta}$)

$t_{13} =$ *19.58/2.66*

$t_{13} =$ ***7.36***

6. The next step is to compare and seek in the tables the value obtained from *t* and its degrees of freedom to obtain the value of *p*; **EXCEL** can also be used instead of looking at the tables, with the following formula:

= DISTR.T.2 (value of t;Df;2)

The first number will be the value of t obtained, the second is the degrees of freedom , and the third can be 2, which is a two-tailed test or 1 if it is a one-tailed test.

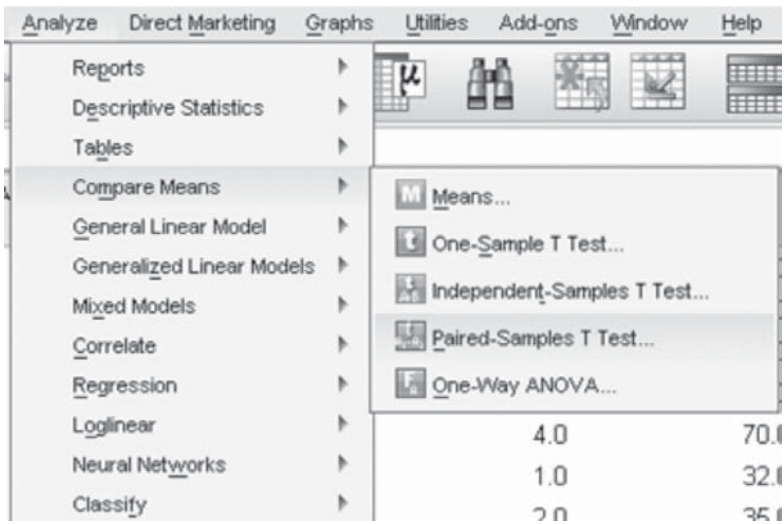
7. One should always follow the result of *p* obtained, together with the confidence interval for the difference in means.

Result of the example: in this example we obtain a t of 7.36, with 13 Df, which is the equivalent of p <0.0000, with a CI of 95% with two tails. This means that there is sufficient evidence to reject the null hypothesis; in other words, there is a significant increase in the scale of the quality of life of the patients postoperatively, and that there was an average increase of 19.58 points on the scale, with a 95% confidence that the increase will be between 13 and 25.3 points.

Example and how to perform it with SPSS:

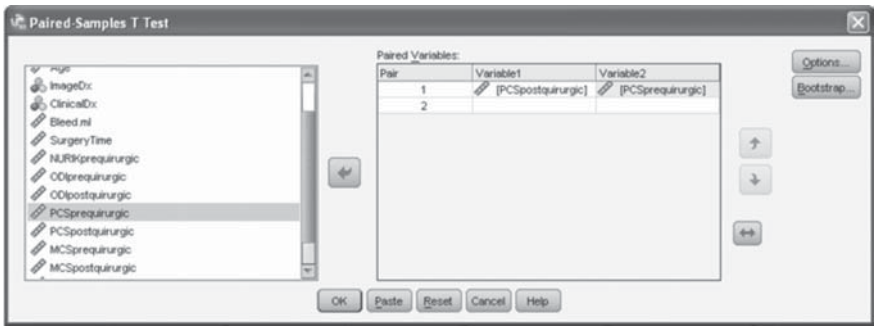
1. The first thing to do in SPSS is to rank our variables that are to be matched, and then go to the menu "analyze" ⇒ compare means ⇒ T test for related samples…¨↵. (Figure 2).

Figure 2 – T Test Related samples



2. A box will appear introducing the two variables to be matched. In this case we will first introduce the variable "after" and then the variable "before" ⇒ accept ↵. (Figure 3).

Figure 3 – T test for related samples



3. Results: The computer will give us three tables (Figure 4). The first table shows the descriptive statistics of the variables to be matched, the second table gives us a correlation coeffcient *(Pearson),* betwen the variables; and the most important to interpret is the **third table**, which shows us in the first column the mean of the "¨*difference*¨ (the change that occurred before and after). In the second column is the standard deviation of the mean of the difference and later the standard error of the mean (**EEMDIF**). The next column is the confidence interval (CI) with the lower limit and the upper limit, which follows in the value of ***t***, then the degrees of freedom, and finally the statistical significance.

Figure 4 – Tables of SPSS results. (V. 19)

Statistics of related samples

		Mean	N	typ deviation	Type erro f of the mean
Par 1	PCSPOST	52,8143	14	9,14506	2,44412
	PCSPRE	33,2357	14	8,42347	2,25127

Correlations of related samples

		N	Correlation	Sig.
Par 1	PCSPOST y PCSPRE	14	,357	,210

Test of related samples

		Related differences							
		Mean	Correlations of	Deviation Typ error of the mean	95% confidence interval for the difference		t	gl	
					Lower	Upper			
Par 1	PCSPOST - PCSPRE	19,57857	9,97938	2,66710	13,81665	25,34049	7,341	13	,000

NOTE: the same data were used for both examples: the manual form and using SPSS. The decision to reject the null hypothesis is when p = <0.05.

Table 3 – Student *t* table

Tabla de la t de Student.
Contiene los valores t tales que $p[|T| > t] = \alpha$,
donde n son los grados de libertad.

$n \setminus \alpha$	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	1,4149	1,8946	2,3646	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	4,1403
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

COMPLEMENTARY READING

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 C. J., Stevens M. B., Homogeneity of variance in the two-sample means test, The American Statistician 46 (1992) 19-21.

12. Newcombe, R. G. 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 N. 5, 2011.

14. Riegelman, R. K. Studying a Study and Testing a Test: How to Read the Medical Evidence. Philadelphia: Lippincott Williams & Wilkins, 1999.

Chapter 16

ANOVA TEST

Juan Pablo Guyot

The ANOVA test is a statistical test used to analyze 2 variances and that is why it is called by the English-language term (**AN**alysis **O**f **VA**riance). It was initially designed in 1925 when it was applied to agricultural work, however its dissemination led to using it in many different fields. The initial objective consisted of observing the influence of different factors on the growth of a given seed, which if different methods of cultivation were not used would all grow the same way. In other words, it aims at evaluating the importance of one or more factors by observing their impact on one or more given populations with constant variances.

In the experimental design of an ANOVA test, an attempt is made to identify "fixed factors" and "randomized factors". The first are when one attempts to study a group that if not exposed would present as unaltered and our interest is to analyze the response to different specific factors. On the other hand, the test is defined with randomized factors to the chance choice of factors with different levels of exposure with a number of multiple levels tending to the infinitive. If the tests are repeated, the factors of chance must arise from the choice on the different levels of choice, and not always following the same pattern.

If we were to make a graph with an example, we would be defining fixed factors in a comparison of means between different populations while the randomized ones would be the analysis of differences by chance among the same population. A practical example of randomized factors in spine surgery would be guided in the study of the evolution of sagittal balance of the Latin American population. The researcher will take factors by chance, but always based on a symmetrical distribution: age range, differences by race, etc.

There are one-way and two-way ANOVA tests related to the factors to be analyzed. One determines one-way ANOVA, or ANOVA with one factor to be analyzed, that involves the study of a single variable,

while two-way or two factor ANOVA is when the variables involved come from different variances.

One-way ANOVA

An attempt is made to determine the influence presented by the result of a Y variable directly related by the distinct factors X (X1, X2, X3, ...). It is used when one intends to analyze the variance of several independent variables or several levels of a single independent variable, maintaining a same degree of confidence.

When an experimental model is designed that uses the one-way ANOVA test we must consider that four variables must be accomplished:

1 – The results of the variables studied should be complementely independent. That is, the value of an observation should not influence the result of another variable.

2 – The data are recorded following an additive model incorporating all of the information, both regarding the fixed effects and the chance errors.

3 – It is assumed that the error distribution follows an organized pattern.

4 – The chance errors follow a homogeneous variance.

An example of this kind of test would be the study of different types of treatment for patients with degenerative stenosis of the lumbar canal. In this case the population to be studied is the fixed factor. The patients with stenosis of the narrow lumbar canal are expected to have a similar outcome, and we shall study the impact of different procedures (factors) on its follow up The response will be studied according to the different treatments applied (simple decompression), decompression + fusion by posterior approach and indirect decompression by intersomatic devices placed by lateral approach) using the ANOVA study when comparing different aspects pertaining to them (surgical time, bleeding, correction of deformity, pre and postoperative Oswestry).

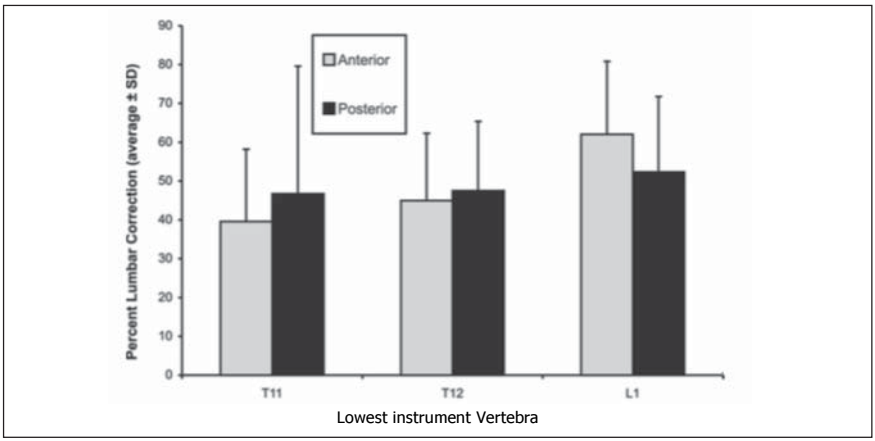
Two way ANOVA

The design of this type of experiment involves the study of 2 or more fixed variables (Y and Z) in relation to different degrees of exposure to the X factors (X1, X2, X3,...). The inclusion of another factor is not

naive since this can generate a few other associated problems, because both factors can be fixed, random, or one fixed and the other random. At the same time, it may happen that said factors act independently, or that they interact among themselves potentiating or inhibiting their final power. Finally, the design can determine 2 groups of factors that are balanced or not.

Interpretation

An example of this kind of experimental design connected for spine surgery would be the evaluation of the degree of spontaneous correction of the lumbar curvatures in selective instrumentations of Lenke type 1 thoracic scoliosis comparing different approaches.



In this case the variable to be studied is the spontaneous correction of the lumbar spine. However, different from the example presented in the one-way case, in this one we have 2 fixed variables independent from each other (anterior vs posterior approach). Finally we shall study the influence on each of the groups, of the instrumented caudal vertebra (T11, T12 or L1).

COMPLEMENTARY READING

1. Tejeda Barreras M. Estadística Básica. En: Falavigna A, Jimenez Avila JM. Educación en investigación AOSPine Latin America : de la idea a la publicación. Caxias do Sul, RS: Educs. 2014, p. 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 , 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.

Chapter 17

PEARSON CORRELATION

Alisson R. Teles
Asdrubal Falavigna

Test description

Correlation is a measure of the degree of linear dependence between two variables or between a measure of intensity of association of these variables. In other words the correlation test evaluates the existence and the force of an association between two quantitative variables. For instance, is there a correlation between the degree of sagittal imbalance (measured in centimeters) and the functional disability of the patients with degenerative scoliosis (measured by the Oswestry)? Another example would be the correlation between the pelvic incidence and lumbar lordosis. Two variables are said to be correlated (vary together) when it is possible to demonstrate that there is a linear association between these variables.

There are basically two ways of verifying a correlation. The simplest is by a graph called dispersion diagram, also known as point diagram (Figure 1). By convention, x is called the independent or predictive variable, which is described on the horizontal axis, and y is the dependent variable or outcome, described on the vertical axis. The other way of evaluating correlation is by a coefficient. The coefficient is superior to the dispersion diagram because it represents an absolute value, which makes it easier to interpret the correlation between two variables with different measures (e.g. sagittal imbalance, measured in centimeters and Oswestry disability index, measured on the scale between 0 and 100 points).

The Pearson correlation coefficient is the best-known measure of correlation between two quantitative variables. It must be applied to evaluate the correlation of continuous variables (non categorical, with at least an interval measurement) which present normal distribution (Gaussian distribution). In other words, the Pearson correlation coefficient should not be used to evaluate the correlation of asymmetrical

variables that does not have a normal distribution. In this case, the Spearman correlation test should be applied.

It is important to remember that the correlation coefficient measures an association, not a cause and effect relationship. For instance, in a study involving patients with lumbar degenerative diseases, it was found that there is a correlation between intraoperative blood loss and quality of life measured by the physical component of SF36 one year after surgery. In this example, the statistically significant correlation does not indicate that blood loss is the cause of worse quality of life one year after the procedure. Another concept is that in large samples, even low correlation coefficients can be statistically significant. This corroborates the importance of the evaluation of degree of correlation (e.g.: $n = 1300$, $r = 0.15$, $p = 0.001$, weak correlation).

Formula

The formula to calculate the correlation coefficient (r) was proposed by Karl Pearson in 1896. The correlation coefficient is an absolute value, independent of the units used to measure the variables x and y.

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

Currently, various data analysis softwares generate the correlation coefficient, level of significance of the correlation, degree of correlation and determination coefficient (Figure 2).

Interpretation

The interpretation of the correlation coefficient (r) basically involves 4 steps (Table 1). First the hypothesis test is used to verify whether there is really a correlation. The rationale of the hypothesis test is basically the same as the comparisons of means. When r is calculated in a sample extracted from a particular population, it is necessary to bear in mind that one is estimating the true association between variables x and y in the population, and therefore it is necessary to verify whether the correlation identified really exists or whether it is related to a sampling error. Through the hypothesis test the p value is obtained, which indicates the probability that the correlation really exists (in statistical terms,

accepting the alternative hypothesis). Generally the existence of a true correlation is accepted when $p < 0.05$; ie., there is less than a 5% chance that the correlation identified does not really exist (accept null hypothesis, reject the alternative hypothesis), therefore it is accepted that it really true.

The second step is to determine the force of the correlation between the variables. Usually, a correlation with a coefficient between 0 – 0.3 is considered weak, between 0.3 – 0.6 regular, 0.6 – 0.9 strong,), 0.9 – 1 very strong, and 1 perfect correlation.

The third step is to analyze the sign of the correlation. The correlation coefficient (r) can present a variation between -1 and +1. Negative values express an indirect correlation, ie., when x increases, y on average diminishes. Positive values indicate a direct correlation, when x and y vary in the same direction. For instance, in normal individuals, the values of the pelvic incidence present a direct correlation with lumbar lordosis. On the other hand, the greater the age, the smaller the lumbar lordosis (indirect correlation).

Another way to evaluate the value of the measure of linear association between two variables is using the determination coefficient (r^2). Raising the Pearson correlation coefficient to the square the determination coefficient is obtained, which represents the fraction of the variability that is shared between the two variables. In other words, it represents the percentage of variation explained by one of the variables in relation to the other. The value r^2 can vary between 0 and 1. For instance, if the Pearson correlation coefficient (r) is 0.734, the determination coefficient (r^2) is 0.834 x 0.834 = 0.69. In other words, 69% of the variation of y is explained by the variable x. The determination coefficient cannot be obtained through the Spearman correlation coefficient (non parametric tests), only through the Pearson correlation coefficient

Example in the field of spine surgery

The Pearson correlation is one of the most widely used tests in medical research. To apply it adequately, however, it is necessary to obey the basic requirements: the variables must be continuous (or at least have an interval distribution) and they must present a normal distribution (Gaussian).

One of the examples in spine surgery is the correlation between the cervical vertical sagittal axis, measured by the distance between the plumb line of the middle point of C2 and C7. Several studies have shown that there is a negative correlation between the cervical sagittal axis and the postoperative quality of life of patients who are submitted to 3 or more levels of cervical laminectomy and posterior fusion.

Tang et al. demonstrated that the correlation coefficient between the vertical sagittal axis C2-C7 and the Neck Disability Index (NDI) is 0.2957 ($p = 0.024$). The correlation coefficient between the axis and the physical component of SF36 (PC-SF36) -0.4262 ($p = 0.001$). In this example, it is found that (1) there is a correlation between the radiographic measure and the NDI and PC-SF36 evaluation instruments (both hypothesis tests demonstrate $p < 0.05$); (2) the correlations between axis C2-C7 and NDI is weak ($r = 0.2957$) and with PC-SF36 is regular ($r = -0.4262$); (3) the correlation is positive between the neck axis and the NDI (the greater the distance between the neck vertical sagittal axis, the greater is the disability measured by NDI), and (4)negative beween the axis and the PC-SF36 (the greater the axis, the lower is the quality of life measured by PC-SF36).

Figure 1 – Examples of disperaion diagrams and the respective correlation coefficients (*r*)

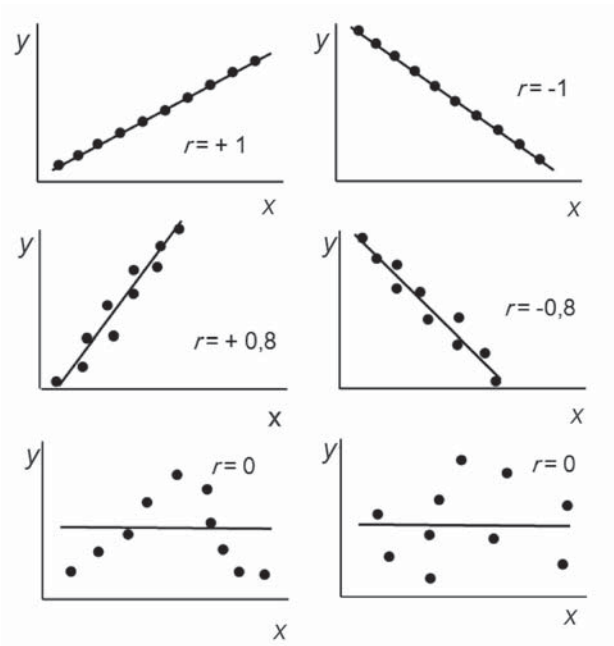


Figure 2 – Example of correlation test between two variables using the SPSS software

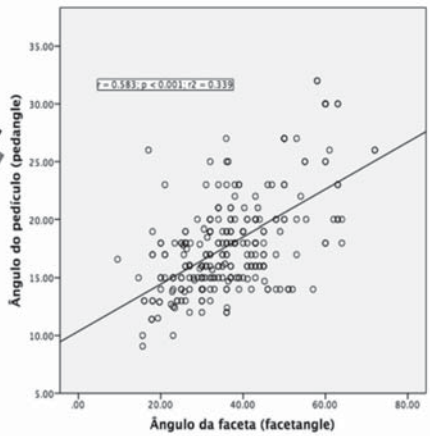
Descriptive Statistics

	Mean	Std. Deviation	N
pedangle	17.8594	4.22765	262
facetangle	36.7264	12.06812	262

Correlations

		pedangle	facetangle
pedangle	Pearson Correlation	1	.583**
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	4664.858	7757.080
	Covariance	17.873	29.721
	N	262	262
facetangle	Pearson Correlation	.583**	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	7757.080	38011.884
	Covariance	29.721	145.639
	N	262	262

**. Correlation is significant at the 0.01 level (2-tailed).



Note: In the example above, the correlation test between the angle of the pedicle (pedangle) and the angle of the interhypophyseal articulation (facetangle) of lumbar vertebrae measured in a computed tomography is demonstrated. It is observed that the greater the pedicular angle in relation to the midline, the greater is the angle of interhypophyseal articulation ($r = 0.583$, $p < 0.0001$). The software also generates the dispersion diagram between the two variables.

Table – Interpretation of the Pearson correlation test (r)

1. Is there a correlation between the variables?

- When $P < 0.05$ does one consider that the correlation is statistically significant, ie., that it did not occur by chance. If $P > 0.05$ there is no correlation.

2. What is the degree of the correlation?

• Evaluate the value of r (qualitative evaluation of r regarding intensity):

0	Null correlation
0 – 0,3	Weak correlation
0,3| – 0,6	Regular correlation
0,6| – 0,9	Strong correlation
0,9| – 1	Very strong correlation
1	Full or perfect correlation

3. What is the sign of the correlation?

• Positive values indicate a direct correlation, vales of x increase as the values of y increase, or vice-versa.

• Negative values indicate an indirect correlation when the values of x diminish as the values of y increase, or vice-versa.

4. What is the determination coefficient (r^2)?

• r^2 indicates the percentage of variation of x that is explained by variable y.

COMPLEMENTARY READING

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. Orxford Handbook of Medical Statistics. Oxford University Express, 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 impacto f standing regional cervical sagital 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.

Unit 3

Chapter 18
Chi²
Alberto Aceves Pérez

Chapter 19
Wilcoxon rank sum test
Baron Zarate

Chapter 20
Mann-Whitney u test
José Manuel Pérez Atanasio
Victor Fernando Luján Celis
Maritza Belén Sandoval Rincón

Chapter 21
Kruskal-Wallis test
Esteban Araya Ramírez
Fabián Víquez Monge

Chapter 22
Spearman correlation coefficient
Fernando Alvarado
Miguel Farfán

Chapter 23
MC Nemar test
Esteban Araya Ramírez
Fabián Víquez Monge

Chapter 24
Simple linear regression
Alisson R. Teles
Asdrubal Falavigna

Chapter 25
Poisson distribution
José Acuña Pinilla
Margarita Gómez Chantraine
Elizabeth Heins

Chapter 26
Odds Ratio
José Manuel Pérez Atanasio
Víctor Fernando Luján Celis
Maritza Belén Sandoval Rincón

Chapter 18

CHI^2

Alberto Aceves Pérez

There are many situations in the field of health, in which the variables of interest, which cannot be quantified by numerical quantities, include those in which the researcher is interested in determining possible relations. Examples of this kind of variable may be the complications after a surgical procedure, sex, sociocultural level, etc. In this case at most the observations would be grouped in the form of frequency, depending on the modalities that are presented by each patient in each of the variables.

The objective of this topic is the study of this type of questions as related to the qualitative variables. These are the contrasts associated with statistics. Generally this type of test consists of taking a sample and observing whether there is a significant difference between the frequencies observed and those specified by the theoretical law of the model which is compared, also called "expected frequencies". However, although this is the best-known aspect, using this test is not limited to the study of qualitative variables. We might say that there are three basic applications to use this test, which we shall develop during the course of this chapter.

We will now discuss three topics in this way:

1. Goodness of fit test: it consists of determining whether the data of a given sample correspond to a given population distribution. In this case it is necessary that the values of the variable in the sample on which we wish to perform an inference be divided into classes of occurrence, or equivalent. Whatever the study variable we will have to categorize the data assigning their values to different classes or groups.

2. Homogeneity test, of various qualitative samples: this consist of testing whether several qualitative samples come from the same

population (for instance: do these three patient samples come from populations with equal distribution of illnesses? The two measurable variables must be represented by categories with which we willl construct a contingency table.

3. Independence test: consists of testing whether two qualitative characteristics are related to each other. For instance, is the presence of vertebral destruction related to the presence of a tumor? Although conceptually it is difference from the previous one, operatively the results are the same. This type of contrast is applied when we wish to compare a variable in two different situations or populations.

Formula:

$$\chi^2 = \sum_{i=1}^{k} \frac{(\mathcal{O}_i - \mathcal{E}_i)^2}{\mathcal{E}_i} \overset{\approx}{\rightsquigarrow} \chi^2_{k-p-h}$$

Example:

We wish to know whether a given illness affects men and women in the same way. In this case, a sample of n=618 individuals who suffer from the disease is taken, and it is observed that 341 are male and the rest are female. What conclusions can be drawn from this?

If p is the percentage of males in the sick population, we can consider the contrast:

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

From the sample we obtain the following point estimate of the percentage of male patients: p = 341/618 = 0.55178.

COMPLEMENTARY READING

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. Elsewier/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>. 3/9/2016.

CHAPTER 19

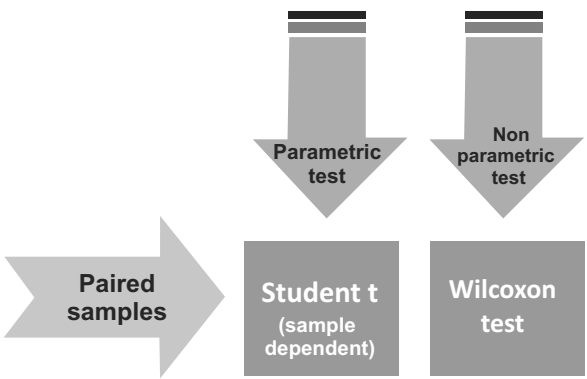
WILCOXON RANK SUM TEST

Baron Zarate

The Wilcoxon test is a hypothesis comparison test parallel (similar) to the Student *t* parametric test for paired samples. It is used to compare variables from 2 related groups that do NOT have a normal distribution, or that are **ordinal** variables, or when the variables do not comply with the basic assumptions of the parametric tests.

It should be used *directly* always in studies with less than 10 observations (n<10). In this statistical test, when the variable "DIF" does not meet the criteria of normality, and one works with the difference in medians. First perform K-A or S-W normality tests, if it is not normal, an attempt can be made to transform it logarithmically, if no normality is achieved, continue the test (Figure 1).

Figure 1 – 5 Tests to compare hypothesis for related samples



The next steps in the Wilcoxon Rank-Sum Test are as follows:

1. Get the paired observations in order and obtain the differences of each pair

2. Organize the differences based on ranks as absolute values, without giving attention to the sign, but in such a way that the ranks will keep the sign corresponding to the difference.

3. Obtain the sum total of the ranks whose sign is less frequent, for instance: if the sign is +, it will be considered to perform sum totals; however, the sum total mentioned finally loses its sign.

4. If they are small samples, compare the value obtained to the critical values of the Wilcoxon table.

5. Distribute the samples larger than 25 under the normal curve and, then calculate value Z, referring to which it is necessary to look at the probability of differing from the average in the table of associated probabilities.

6. Decide whether to accept or reject the hypothesis.

Example: In this example an attempt is made to decide whether the change in the scale of Nurick's cervical myelopathy in patients operated on for cervical decompression is statistically significant. An ordinal scale was used from 0 to 5, and it was evaluated using the Wilcoxon test. Our null hypothesis ($\mathbf{H_0}$) is that there is no difference in the degree of myelopathy of patients before and after surgical treatment.

$$H_0 \cong \mu \text{ before} = \mu \text{ after}$$

The alternative hypothesis ($\mathbf{H_1}$) is that there is a statistically significant (positive) difference in the improved medullary function.

$$H_1 \cong m \text{ before} \neq \mu \text{ after}$$

$$\text{Rank sum} : n(n+1) / 2$$

1– The first step is to take the difference of the values of "before" and "after", but in this case ***n*** would be taken into account only in the values where the difference is DIFFERENT from 0. (Table 1).

Table 1

Case	Nurick pre	Nurick post	DIF	Absolute Dif.	Order	Ranks	Rank Sum +	Rank Sum -
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

In this case n would be 9. Only the values of "DIF" that are different from 0 are taken into account.

Table 2

Case	Nurick pre	Nurick post	DIF	Absolute Dif.	Order	Ranks	Rank Sum +	Rank Sum -
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

The next step would be to rank the absolute value (the sign + or – is not taken into account) from smaller to larger, and then assign the middle rank of each interval. (Table 2) .

Table 3 – Rank and sum up the positive and negative ranks (Table 3)

Case	Nurick pre	Nurick post	DIF	Absolute Dif.	Order	Ranks	Rank Sum +	Rank Sum -
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

In this case the value obtained for the higher rank sum (45) is compared to the critical values of table T in rank tests signalled for equal pairs of Wilcoxon, and it can be assumed that in order to be significant (ie., less than 0.05, which was the level of significance), it should be less than 0.05. (Table 4).

Table 4 – Critical values of the Wilcoxon Table

The null hypothesis is rejected if the largest of the rank sums is >= than:

n	With α = 0.01	With α = 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

In this case the value is 45 with an n of 9, and therefore the value is less than 0.01, which is why the Null Hypothesis (H_o) must be rejected.

Note: When the value of n is equal to or greater than 25, Z must be calculated manually. (Table 5).

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

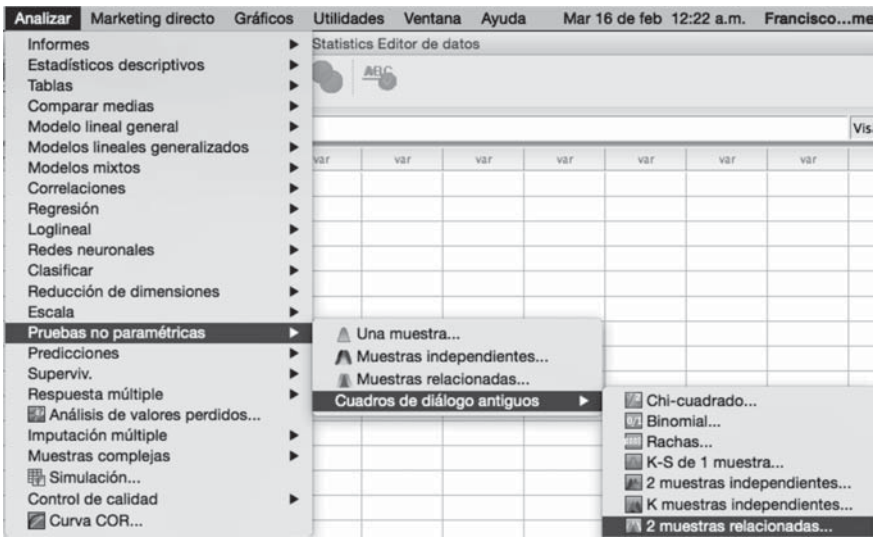
Table 6 – Critical values of Z and their significance (one tailed or two tailed)

Percentiles of the normal distribution						
	Unilateral test			*Bilateral test*		
Significance	0.05	0.01	0.001	0.05	0.01	0.001
Percentile	95	99	99.9	97.5	99.5	99.95
z	1.6449	2.3263	3.0902	1.9600	2.5758	3.2905

To perform it with SPSS:

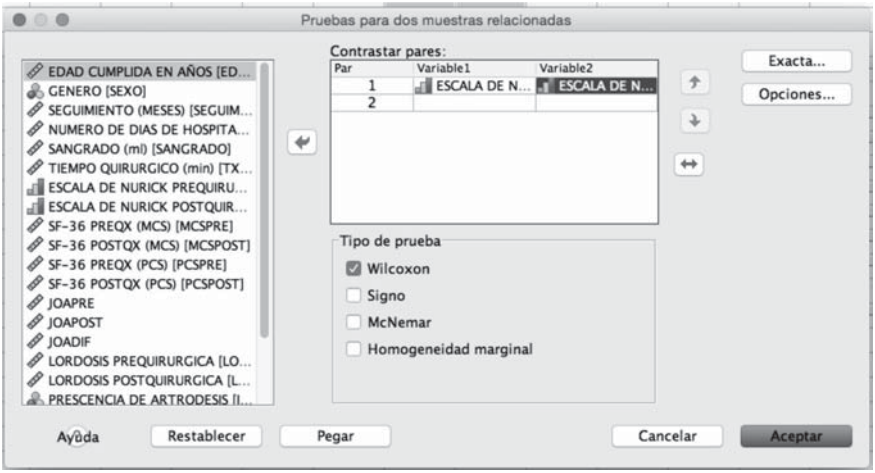
1. Maintain the variables ranked when introducing. Later in the menu *¨analyze¨⇒ non parametric tests¨⇒ ¨old dialogue boxes¨⇒¨two related samples ¨↵ (Figure 2).*

Figure 2



2. A dialogue box will appear where it is necessary to introduce the two variables that are to be compared; introduce the variable "before" and then the variable "after", in this test. Auutomatically the "Wilcoxon" field is selected, then press accept. (Figure 3).

Figure 3



3. Later two tables will appear (Interpretation). The first gives us the variables introduced (***η***) and the rank sum (negative , positive, null). And the second table (called comparative statistics), gives us the value of z and the two tailed significance. (Figure 4).

Figure 4 – The rank test with the Wilcoxon sign

Ranks

		N	Rank sum	Postsurgical
Postsurgical Nurick scale – Presurgical Nurick scale	Negative ranks	9[a]	5,00	45,00
	Positive ranks	0[b]	,00	,00
	Ties	6[c]		
	Total	15		
	Average Rank			

a. Postsurgical Nurick scale < Presurgical Nurick scale
b. Postsurgical Nurick scale > Presurgical Nurick scale
c. Postsurgical Nurick scale = Presurgical Nurick scale

Comparative statistics*

	Nurick scale Postsurgical Nurick scale Presurgical
Z	-2,724
Asymptotic sign (bilateral)	,006

a. Rank tests with Wilcoxon sign
b. Based on the positive ranks

Interpretation would be with a p of 0.006 which indicates that there is sufficient evidence to eject the null hypothesis H_o, ($p<0.05$), and accept our alternative hypothesis (H_1):, which in this case is directional towards diminishing the Nurick scale postoperatively. This would indicate statistical significance of improving the function by decompression surgery.

> **Note**: The same data will be used for both elements: the manual form and with SPSS. The decision to reject the null hypothesis is when $p = <0.05$.

COMPLEMENTARY READING

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 N. 5, 2011.

14. Riegelman, RK. Studying a Study and Testing a Test: How to Read the Medical Evidence. Philadelphia: Lippincott Williams & Wilkins, 1999.

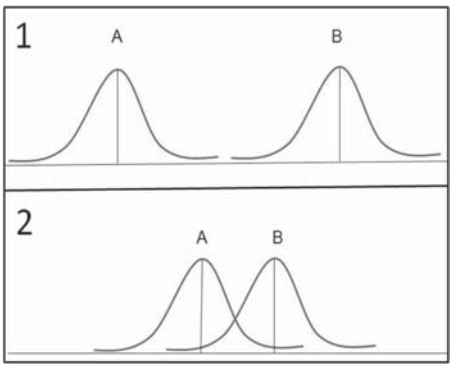
Chapter 20

MANN-WHITNEY U TEST

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

To understand the Mann Whitney U test we must remember that the objective of the hypothesis tests is to identify whether the difference between the data of the groups is statistically significant, which we can identify graphically and clearly in the following figure.

Figure 1– Difference betwen groups A and B. 1 The difference between the groups is statistically significant. 2. The difference between the groups is not statistically significant



Test description

The Mann-Whitney U test allows us to identify whether there is a difference between two independent samples. The variable of interest should present a non normal distribution. This test consists of comparing the medians of both groups and determining if there is a statistically significant difference. It is necessary to emphasize that the Mann

Whitney U test specifically compares only two independent samples. If we wish to compare more than 2 samples other tests would be used.

Formula

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

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

Where U_1 = Sample with the smallest number of patients

U_2 = Larger sample size

R_1 y R_2 = Rank sum for each group

Example in the field of spine surgery

In order to make it easier to understand this test, we present the following example:

At the National Medical Center for the Spine, degenerative pathologies of the cervical spine are treated with two types of surgical approaches: type A advocated by Dr. Juan as the best, and type B advocated by Dr. José. In order to identify whether there is a statistically significant difference in the days it takes to heal, we suggest the following hypothesis:

- There is a statistically significant diffeence between the healing time in approach A and approach B.

First, the values of healing time in both groups are recorded on a calculation sheet and they are ranked with all data in an order from smallest to largest, as shown in the following table:

Tabla 1 – Ranking the data

Type A approach n1=6	Global Order (Rank)	Type B pproach n2=5	Global Order (Rank)
15	5	11	1
14	4	17	7
12	2	19	9
20	10	13	3
18	8	21	11
16	6		
Rank sum	**35**		**31**

We then calculate the median of each group. The median of the type A approach is 15.5 and the median of the type B approach is 17. Now we add up the values obtained in the global order of each of the groups (rank sum), Group A: 35 Group B: 31, and with these values we isolate the formula.

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

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

Finally, the lowest value of U which, in this case is 14, is collated in a table of critical values of alpha (α) of 0.05 for the Mann-Whitney U test to get to know the value of p which in this case is 0.567.

Interpretation

As we can see, p is not significant, in other words it is not <0.05 and therefore it is interpreted that there is no statistically significant difference between the time of healing using the approaches A and B, and therefore we rule out the hypothesis proposed.

In order to perform this test using the statistical analysis softwares, such as the SPSS program, it is necessary to select the paragraph of non-parametric tests and use the comparison of medians between independent samples.

COMPLEMENTARY READING

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 fos 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., Diferencias 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 prueba 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.

CHAPTER 21

KRUSKAL-WALLIS TEST

Esteban Araya Ramírez
Fabián Víquez Monge

This methodology is used in the cases where there are two or more groups of data and it is wished to find out whether these are from a same population or from distinct populations with the same characteristics.

Statistically, it is a matter of proving the hypothesis that ***k*** independent samples are from a population with the same median. It is catalogued as a **non-parametric** test, in other words, it is necessary for the data analyzed to have a distribution with specific parameters. Furthermore, it is also not necessary to accomplish the assumption of equality of variances among the groups of data (homogeneous variances).

The test is mathematically defined as follows:

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

Where:

H = Statistical value of the Kruskal-Wallis test.

N = total size of the sample.

Rc^2 = sum total of the ranks raised to the square.

ni = size of the sample of each group.

L = adjustment given by the adjustment of rank leagues or draws.

The adjustment of L is calculated as follows:

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

Where:

Li= value of the number of draws in a rank.

N= total size of the sample.

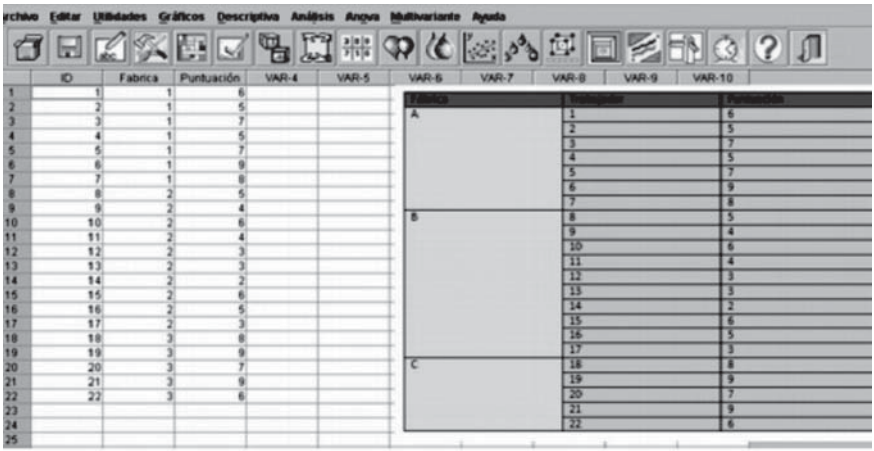
Example

A nurse performs a multicenter cross-sectional study to find out the perception of discomfort of workers in their workplace and wishes to find out whether there are significant differences between the three power plants.

The study was performed at three electrical power plants with a total of 22 subjects previously diagnosed as presenting a Burnout syndrome. The first plant (A) had 7 subjects, the second (B) 10, and the last (C) 5 subjects.

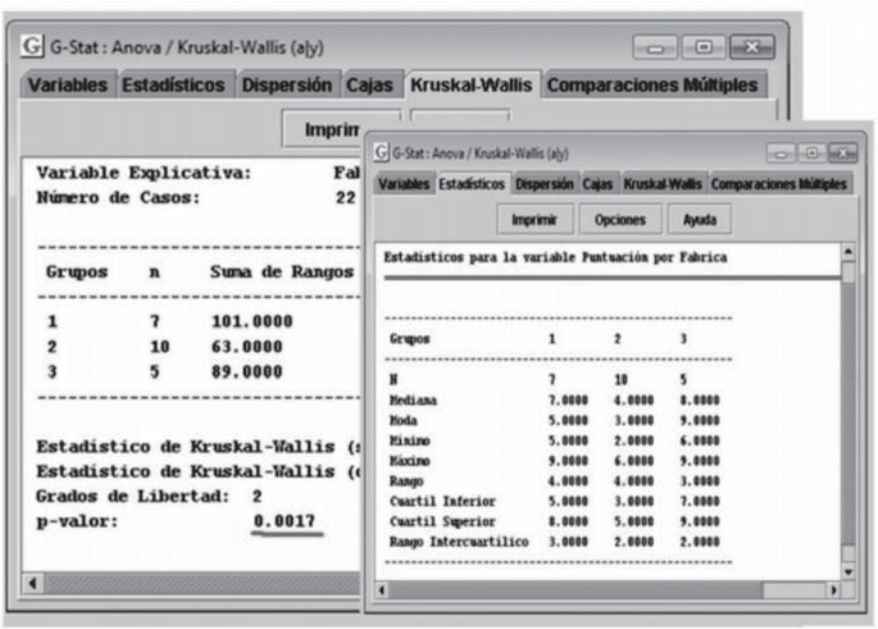
In order to hear the opinion of each worker, a questionnaire was applied that includes a number of questions evaluated according to a scale from 1 to 10, the highest values being related to negative feelings. The study variable has an ordinal character, therefore the central tendency measure to be used is the median.

In the study the null hypothesis (H_o), will be equality of the medians in the different groups, the alternate hypothesis (H_a) in turn being the existence of significant differences between the different study groups. Firstly, the study database is developed and a description of the study is obtained:



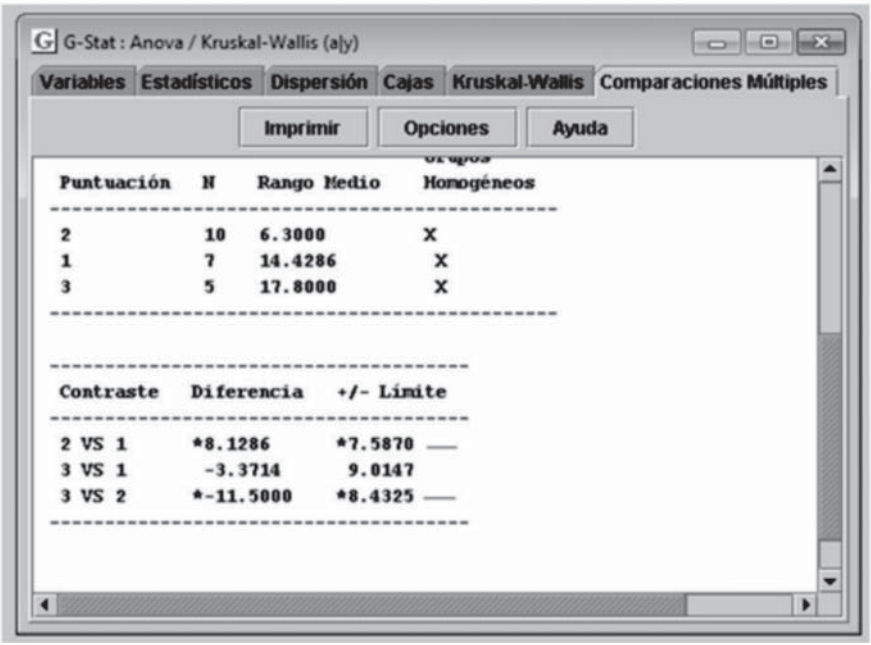
In this study three independent groups are compared. Moreover, the sample size is small and the variable result is an ordinal variable. For statistical analysis the statistical test of choice is the **Kruskal-Wallis** test**.**

For this, using a statistical program under the ANOVA heading the *Kruskal-Wallis* option is selected. Later, the ordinal variable (score variable), the explanatory variable(b) and qualitative variable (plant variable) are placed in the response variable (y). Next the *Kruskal-Wallis* heading is selected to see the results of the test and prove whether there are differences. Further, under the statistics heading the medians of the three groups can be seen. In the case of the Kruskall-Wallis test, it is necessary to pay attention to the value of p, which is 0.0017.



Finally, if the value is less than 0.05 the null hypothesis (H_o) of equality is rejected and the alternate hypothesis (H_a) of difference of medians is accepted with a confidence interval of 95%.

Later, using the Dunn test, the different study groups are compared. This shows a difference between plant A and B and between plant B and C.



Thus, it can be said that there are significant differences in the medians of the perception of discomfort by the workers in the workplace between plant A and plant B and plant C with a confidence interval of 95%. We cannot say that there are significant differences between plant A and C.

COMPLEMENTARY READING

1. Altman DG: Practical statistics for medical research. 1ª 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. 2ª edición. McGraw-Hill, New York, 1988.

CHAPTER 22

SPEARMAN CORRELATION COEFFICIENT

Fernando Alvarado
Miguel Farfán

The ensemble of statistical techniques that allow measuring the intensity of association or relationship that can exist between two or more variables is known as correlation. This correlation may exist between at least two variables (one dependent and the other independent), or between more than two variables (one dependent and two or more independent). In the latter case it is called a multiple correlation.

The types of association (correlation) between two variables can be linear, positive, negative or null. The Spearman correlation coefficient (Spearman Rho) is a non-parametric test that is useful to determine (quantify) whether there is a correlation between two numerical (ordinal) variables, whose distribution is likewise non normal or non parametric.

This correlation measurement has three basic goals: the first is to determine whether said correlation really exists (the values of a variable tend to be higher or lower for the higher or lower values of the other variable); the second is to be able to predict the value of a variable knowing the value determined for the other variable and, finally, to evaluate the level of concordance or relationship between the values of the variables.

The Spearman correlation coefficient is very useful in variables that, given their non normal behavior, present extreme values in their data that would affect statistics such as the Pearson correlation coefficient (See respective chapter).

It should be highlighted that Spearman Rho measures the existence and intensity of association of two quantities and does not detect their level of agreement or concordance; accordingly, it should not be used to compare values from instruments that measure the same event.

Spearman correlation coefficient hypothesis

H_o: There is no association between the two variables

H_a: There is a relationship or association between the two variables (positive or negative).

Equation

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

Where $d_i = r_{xi} - r_{yi}$ is the difference between the ranks of X and Y.

Interpretation

To perform the interpretation, before running the test the authors should define what is the objective of the research, and also determine the importance of the relations of the variables that are involved in the test, so that, in case there is chance involved, the conclusions of the correlation are not left exclusively to a mathematical statistic.

The Spearman correlation produces a numerical value of intensity of association (positive or negative) of a tendency of two variables to increase or decrease together and does not indicate causality between them.

According to the literature there are at least four ways of interpreting the Spearman Rho in numerical scales, but in general all the scales determine ranks of the test between -1 y +1. The values that approach 1 indicate a strong and positive correlation. The values that are close to -1 indicate a strong and negative correlation. The values close to 0 indicate the non-existence of a **linear** correlation. The positive or negative sign before the numerical value indicates the increase or decrease of a variable in relation to the other. That is, a negative sign reflects the increase of a variable as the other decreases or viceversa, and a positive one, the increase of a variable as the other does, or the decrease of one as the other decreases at the same time.

Variables with values close to + 0.95 are not often found in clinical studies. The revision of data and calculations in search of errors or inconsistencies is suggested, with values greater than or smaller than 1.

Martínez Ortega of the higher institute of Medical Sciences of Havana, together with collaborators, does a literature search regarding the Spearman Rho statistic and summarizes 4 different scales to interpret the test, namely:

Scale 1:

The correlation coefficient oscillates between –1 and +1, the value 0 that indicates that there is no linear association between the two variables studied.

Scale 2:

Perfect negative correlation	-1
Moderate to weak negative correlation	-0,5
No correlation	0
Strong moderate positive correlation	+0,5
Perfect positive correlation	+1

Scale 3:

1) Perfect R = 1
2) Excellent R = 0.9 < = R<1
3) Good R = 0.8 < = R < 0.9
4) Fair = R = 0.5 < = R < 0.8
5) Bad = R < 0.5

Scale 4: Rank

Rank	Relationship
0-0,25	Scarce or Null
0,26-0,50	Weak
0,51-0,75	Between moderate and strong
0,76-1,00	Between strong and perfect

Example of application of the Spearman Rho statistic

Feise et al. in their study on the construction and validation of a *quality of life scale for idiopathic scoliosis* (SQLI, according to its acronym in English) of 22 items, help the Spearman Rho correlation statistic, comparing the new scale with the already known *quality of life profile for spinal deformities* (QLPSD according to its acronym in English). In a group of 84 patients, 74 adolescents in whom an intervention was performed for idiopathic scoliosis (surgical management, with brace or without brace) and 14 controls (siblings) both scales were applied (SQLI and QLPSD). The global numerical results of the scales were compared for the purpose of determining whether there was or not a positive correlation between them. Feise describes an additional classification for the interpretation of the Spearman Rho statistic as shown in Table 2. Comparing the global numerical values of the two scales, a Spearman Rho correlation statistic of +0.79 is obtained, assuming a very good correlation between them (Figure 1).

Table 2 – Interpretation of Spearman Rho according to Feise et al.

Rank	Relatioship
0 – 0,2	Poor
0,2-0,4	Weak
0,41- 0,6	Good
0,61- 0,8	Very good
0,81- 1	Excellent

Figure 1 – Correlation table for scales SQLI and QLPSD in idiopathic scoliosis in adolescents in the study by Feise et al.

	SQLI Self-esteem	SQLI Back Pain	SQLI Physical Activity	SQLI Moods & Feelings	SQLI Global	QLPSD Psychosocial Functioning	QLPSD Sleep Disturbance	QLPSD Back Pain	QLPSD Body Image	QLPSD Back Flexibility
SQLI back pain	0.200									
SQLI physical activity	0.264	0.283								
SQLI moods & feelings	0.377	0.498	0.208							
SQLI global	0.642	0.659	0.552	0.782						
QLPSD psychosocial functioning	0.345	0.410	0.364	0.332	0.477					
QLPSD sleep disturbance	0.301	0.476	0.107	0.576	0.525	0.550				
QLPSD back pain	0.293	0.813	0.354	0.523	0.709	0.636	0.564			
QLPSD body image	0.545	0.462	0.220	0.465	0.624	0.415	0.517	0.538		
QLPSD back flexibility	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

COMPLEMENTARY READING

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 disponible. Complexo 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.

CHAPTER 23

MCNEMAR'S TEST

Esteban Araya Ramírez
Fabián Víquez Monge

Sometimes several measurements of the same characteristic are performed for each of the individuals included in a research study. In such cases, the focus of interest is to compare whether any significant changes have occurred between the measurements performed at two different times, usually before and after a treatment is applied or an intervention carried out.

Mc Nemar's test is used to decide whether it is or not acceptable for a given "treatment" to induce a change in the response of the elements submitted to it and is applicable to the "*before-after*" designs in which each element acts as its own control.

It consists of a non parametric test for two related binary variables that compares the changes in the responses using the Chi^2 X^2 distribution. It is not only useful to detect the changes in the responses in the before and after experimental studies, but also to compare two types of treatment.

This procedure is useful when there are dependent samples and the type of scale is nominal. The results corresponding to a sample of ***N*** elements are shown in a 2x2 frequency table to collect the set of responses of the same elements before and after.

The measurement scale for X and Y is nominal with 2 categories such as: positive or negative, female or male, presence or absence, which can be called 0 and 1.

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

To provide an even clearer example of the table above, it can be simplified as:

Before/After	–	+
–	a	b
+	c	d

In the table cells, ***a*** is the number of elements whose response is the same , (–); ***b*** is the number of elements whose response is (–) before the *treatment y* and (+) after it; ***c*** is the number of elements that have changed from (+) to (–); and ***d*** is the number of elements that maintain the response (+).

Therefore, b + c is the total number of elements whose responses have changed, and they are the only ones that intervene in the comparison.

The null hypothesis (H_o) is that the *treatment* does not induce significant changes in the responses, ie., the changes observed in the sample are due to chance, so that it is just as likely to have a change from + to -+ as a change from – to +. Thus if H_o is right, of the b+c elements whose response has changed, it is to be expected that (b+c)/2 have gone from + to -, and (b+c)/2 have gone from – to +. In other words, if Ho is right, the expected frequency in the corresponding cells is (a+b)/2.

The alternative hypothesis (H_a) can be non-directional when it postulates that the probability of a change from + to – has a distinct probability of a change from –to +, or direction when it predicts that a change from – to + is more (or less) probable than a change from + to-. The statistical test that allows comparing whether there are significant differences between the expected and observed frquencies is:

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

Oi= frequency observed in the i-th cell

Ei= frequency expected in the i-th cell if H0 is right

K= number of cells

To compare the significance of the changes only the cells that show changes are of interest, therefore the statistic can be expressed as:

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

If H_o is right, the X^2 statistic has a distribution of approximately chi-square with 1 degree of freedom. The approach is more precise if the Yates continuity correction is performed and the statistic becomes:

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

	Probability of a higher value - *Alfa* (α)				
Degrees of Freedom	**0 1**	**0 05**	**0 025**	**0 01**	**0 005**
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	9.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55

Example

A professional of preventive medicine observes that employees in a factory often suffer from diarrhea, which is why there is so much absenteism. All of the employees eat in the factory canteen as one of their benefits.

He assumes that the common denominator of the cause of diarrhea is the place where they eat the food, in other words, the food is not prepared with appropriate hygiene. However, the personal hygiene of

the employees is not enough to blame everyhing on the kitchen staff. He selects a random sample of 50 individuals, 34 of who often have diarrhea and 16 do not have any diarrhea. The researcher suggests that surveillance be carried out, and that personal hygiene measures be applied consisting of exhaustively washing their hands before eating during a two-week period. At the end of the treatment the results were as follows: of the 34 subjects who often presented with diarrhea, after the treatment by washing hands, in 16 the bowel problems had disappeared and 18 continued to have diarrhea; on the other hand, in the group of 16 asymptomatic individuals, 4 of them had diarrhea despite washing their hands, and in 12 the situation remained unchanged.

Choice of statistical test

The experimental model has two dependent samples.

Hypothesis presented

• Alternate hypothesis (H_a). Washing hands as a preventive measure and factor of personal hygiene shows significant changes in the diarrhea suffered by the employees who use the factory canteen studied.

• Null hypothesis (H_o) The differences that are observed in the frequencies of change due to washing hands are due to chance.

Level of significance

For every probability value equal to or less than 0.05, H_a is accepted and H_o rejected.

Rejection zone

For every probability value > than 0.05, H_o is accepted and H_a rejected. Result of the 2x2 contingency:

		After washing hands	
		-	+
Before washing hands	-	16	18
	+	12	4

Statistical test applied

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

Calculation of the degrees of freedom (df)

df = 1

McNemar' X^2 statistical test is compared to the critical values of chi^2, with the degree of freedom. It is observed that 3.84 corresponds to a probability of 0.05, so that the calculation corresponds to a probability < than 0.05.

Decision

Since the value of X^2 calculated has a probability < than 0.05, it falls into the level of significance, and therefore Ha is accepted and H_o rejected.

Interpretation

Washing hands led to significant changes in the diarrhea among the factory employees, which means that deficient personal hygiene clearly contributes to the frequency of the bowel problem which causes absenteism.

Applying the statistical test does not cancel the participaton of deficient hygiene among the kitchen staff who prepare the food, since 4/16 employees who had been asymptomatic and washed their hands exhaustively before eating presented diarrhea; however, the statistical test reveals that personal hygiene has a more significant participation than the failure in preparing food.

COMPLEMENTARY READING

1. Daniel WW: Estadística con aplicaciones a las Ciencias Sociales y a la educación. 1ª 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. Estadistica no paramétrica. 2ª edición. Méjico: Editorial Trillas, 1995.

5. Sprent, P. Applied nonparametric statistical methods. 3ª edición. Boca Ratón, 2001.

CHAPTER 24

SIMPLE LINEAR REGRESSION

Alisson R. Teles
Asdrubal Falavigna

Test description

Simple linear regression is a mathematical model used to describe the relationship between two variables, aiming to utilize one of them to predict the value of the other. In other words, it is a statistical technique used when there are reasons to assume that there is a cause-effect relationship between the two quantitative variables and that one wishes to express this relation mathematically.

The objectives of a simple linear regression test are: (1) to evaluate a possible dependence between two quantitative variables (x= predictive or independent variable / y= dependent variable or outcome), and (2) to express this response mathematically by means of an equation.

For instance, it is possible to calculate the linear dependence between lumbar lordosis and the pelvic incidence of a given population. Another much used example in medical research is when one wants to know the dose-response relation of a given drug in order to be able to predict the physiological reaction of an individual as a function of the use of given quantity of medication.

The requirements for the study of linear regression are: the dependent variable (y) must be random, ie., obtained by sampling; the independent variable (x) and the dependent one (y) should be associated linearly (the graph between the two variables must present a straight line); and the variances of the dependent variable (y), given different values of the independent variable (x), are all equal.

Formula and interpretation

The equation presents the dependent variable (y), the independent variable (x) and a coefficient that reflects the intensity of the relationship between variable x and y.

The regression analysis equation is defined by: $y = a + bx$, where y is the dependent variable (outcome), x is the independent variable (predictive), a is the linear coefficient (value of y when $x = 0$), and b is the angular coefficient (slope of the straight line: addition or reduction in y for each addition of a unit in x). Graphically, the simple linear regression is represented by the scatterplot, also called scatter diagram. The equation is represented by the straight line that expresses the linear relation between the mean value of y for the different corresponding values of x.

In order to determine the simple linear regression line, it is necessary to identify the value of a (vertical intercept), and the value of b (angular coefficient). These parameters are estimated using a technique known as the least squares method. The values of a and b are calculated so that the sum of the squares of the deviations from the observed points y and the straight line itself are minimal. The vertical distance between the point and the straight line is called error or residualsof the straight line in relation to the point. The values of a and b are calculated using the following formula:

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

Once the value of b (angular coefficient) has been calculated, the value of a (vertical intercept) is given by:

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

Where $\bar{x}$ and $\bar{y}$, are, respectively, the means of the data relating to variables x and y. The estimated straight line of linear regression is: $\hat{y} = a + bx$, where $\hat{y}$ is the estimate of observation y for a specific value of variable x.

In practice, the simple linear regression equation can be generated by various data analysis softwares. The analysis first generates data that identify the existence or not of a significant correlation between the

variables (prerequisite for obtaining the regression equation), supplies the scatterplot with a plotted straight line and identifies the values of *a* and *b*. In the figure, there is an example of simple linear regression using SPSS® software®.

Figure 1 – Example of simple linear regression analysis using SPSS®

Variable Processing Summary

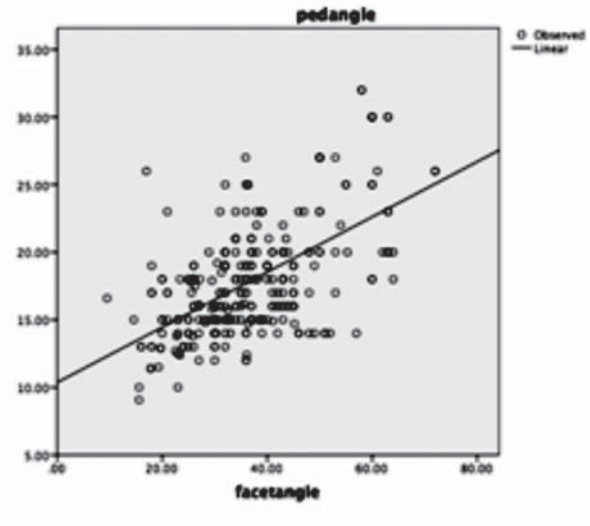
		Variables	
		Dependent	Independent
		pedangle	facetangle
Number of Positive Values		262	262
Number of Zeros		0	0
Number of Negative Values		0	0
Number of Missing Values	User-Missing	0	0
	System-Missing	0	0

Model Summary and Parameter Estimates

Dependent Variable: pedangle

	Model Summary					Parameter Estimates	
Equation	R Square	F	df1	df2	Sig.	Constant	b1
Linear	.339	133.548	1	260	.000	10.365	.204

The independent variable is facetangle.



Note: In the example above, the relation between the pedicle angle (pedangle) and the interapophyseal joint angle (facetangle) of lower back vertebrae measured by a CT scan is demonstrated. It is observed (1) that the correlation between the two variables is significant (Sig. .000, represents the value of $P < 0.0001$), (2) the scatterplot is identified with the plotted regression straight line, (3) the values of a (constant) and b (b1) are identified. Therefore it can be said that in the sample studied the relation between the pedicle angle can be identified by the following equation: pedicle angle = 10.365 + (0.204 x facet angle).

Example in spine surgery

Simple linear regression analysis is used when it is intended to predict a variable in relation to the variation of the other. In practice, first one identifies the existence of a significant correlation between the two continuous variables, and then a linear regression analysis is performed to generate the equation that determines the outcome variable (*y*) in relation to the values of the predictive variable (*x*).

There are several applications in the field of the spine. For instance, Hasegawa and collaborators identified the relationship betwen functional incapacity measured by the Oswestry Disability Index (ODI) and (1) age of the research volunteer and (2) mismatch betwen the pelvic index and lumbar lordosis. They studied 136 volunteers without spine disease with an X ray and clinical evaluation. In the study they identified the following equation between ODI and age: ODI= 0.36 + 0.12xAge. In this analysis they obtained value r (correlation coefficient) of 0.2143 (Pearson Correlation Chapter) and a value of p= 0.01664. The equation between ODI and mismatch pelvic incidence (PI) and lumbar lordosis (LL) identified was: ODI= 5.62 + 0.12x(IP-LL), with value r= 0.1969 and P= 0.0277. Using this equation it is possible to predict the value of ODI taking into account age and also the PI-LL mismatch.

COMPLEMENTARY READING

1. Hasegawa K, Okamoto M, Hatsushikano S, Shimoda H, Ono M, Watanabe K. Normative values of spino-pelvic sagital 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. Orxford Handbook of Medical Statistics. Oxford University Express, 2011. ISBN: 978-0-19-9555128-6.

Chapter 25

POISSON DISTRIBUTION

José Acuña Pinilla
Margarita Gómez Chantraine
Elizabeth Heins

Test description

The test received its name in honor of the French physicist and mathematician Simeón-Denis Poisson, (1781-1840) who disseminated it in 1837 through his paper: *Recherches sur la probabilité des jugements en matières criminelles et matière civile* (Research on the probability of judgments in criminal and civil matters).[2,3,11]

The Poisson distribution is one of the most important probability functions and it is part of the group of functions for the probability distribution of discrete random variables. In other words, events can only have integer values (e.g.: 0,1,2,3,4…n) and the random events that appear are independent from each other, presenting during a time or space interval.

It is punctually specialized in the frequency of events with very low probabilities compared to the frequency of the non-occurrence, and therefore they are considered as "rare events". Its approach to the binomial will occur when the probability that the event (p) take place is very close to zero, or the sample size (n) is greater than 30 ($p < 0,1$, $n > 30$).

This function is based on a mean frequency of occurrence **(λ)** also known as a positive parameter for the hope of occurrence, which identifies the average number of occurrences or events that present within a time or space interval. Furthermore, it involves the number of times the event or phenomenon **(x)** that one expects to identify (the function gives us the probability that the event will occur precisely x times). For instance, if the event studied occurs on average 5 times per minute and we are interested in the probability that it will occur x times within a 10-minute interval, we will use a Poisson distribution model with $\lambda = 10x5 = 50$.[4,7]

Characteristics of the Poisson distribution

The distribution results from a set of assumptions about an implicit process to form a set of numerical observations. The following statements describe what is known as the Poisson process.

a) The events have independent occurrences. The appearance of an event (x) at a time or space interval does not have any effect on the probability that there may be a following appearance of the event studied at a same interval, or at some other.

b) Theoretically it should be possible for the event studied to appear a finite number of times within the time or space interval.

c) The probability that a single appearance of the event studied at a given time interval is proportional to the dimension of the time or space interval (or some volume of the matter).

d) The probability of more than one occurrence of the event studied at any infinitesimal fraction of the time or space interval is insignificant.

An interesting characteristic of th Poisson distribution is that the mean and variance are the same.[1,6,10,11]

Application

The Poisson distribution is used to describe several processes, including:

a) The number of vascular lesions of the main vessels in spine surgery using an anterior approach per year.

b) The number of radicular lesions in minimally invasive discectomies out of the total of patients operated using this technique in a given period.

c) The number of pedicular screws that presented pullout in the reduction of the total of functional levels that underwent intervention in scoliosis surgery.

d) The number of pedicular screws that are not well positioned in posterior lumbar fusion surgery per year.

e) The number of cases of pseudoarthrosis after surgery for anterior cervical fusion per month.

Each of these random variables represents the total number of occurrences of a phenomenon during a fixed time period in a fixed region of the space. It expresses the probabiity of a number (x) of appearances of the event in a given time, if these events occur with a known mean frequency and are independent of the time elapsed since the last occurrence or event.

The probability of the appearance of the number of events can be calculated puntually or in an accumulated manner, ie, a grouped quantity of positive events in the time or space inteval.

Probability Density Function:

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

It can be demonstrated that $f(x) \geq 0$ for each (x) and that $\sum_x f(x) = 1$ so that the distribution meets the requirements for a probability distribution.

Where:

$\boldsymbol{\lambda}$ = Parameter hope of occurrence (# ocurrences / Unit of time or space)

x = # of expected events of the phenomenon (0,1,2,3,4…n)

e = Base of the natural logarithms (e = 2.71828...)

Accumulated Probability Function:

$$f(X \leq x) = \sum_{0}^{x} \frac{e^{-\lambda}\,\lambda^{x}}{x!}$$

Variance σ^2 = Mean μ = Paramenter hope of occurrence λ

Standart deviation $\sigma = \sqrt{\sigma^2} = \sqrt{\mu}$

Figure 1 – The probability density function for the entire occurrence in integer values(x) with an expectation of occurrence (λ). The continuous line is only a visual guide, it does not mean continuity

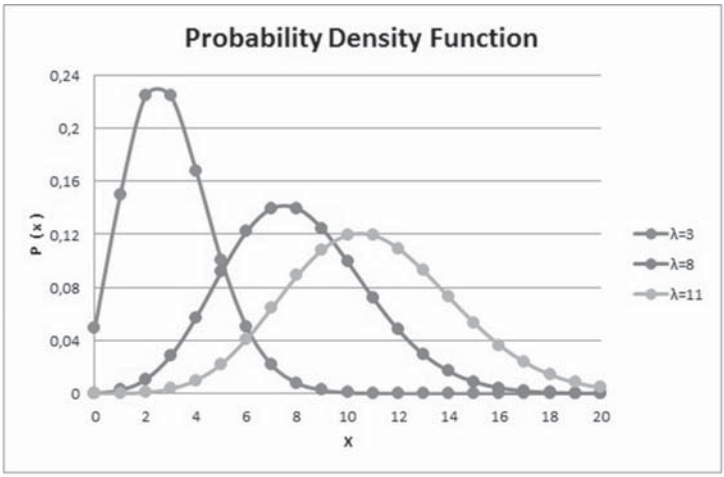
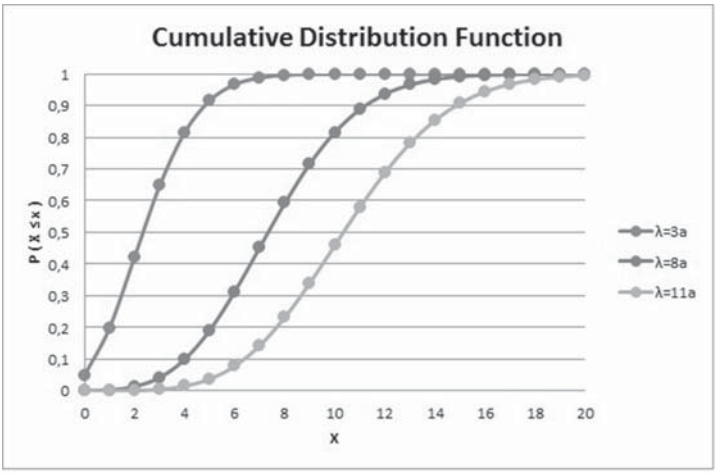


Figure 2 –The accumulated distribution function (ADF) is discontinuous in the integers of x and plane in all parts because a variable that is the Poisson distribution takes only integer values



Examples in spine surgery

1. In a meta-analysis on the presence of pseudoarthrosis in an anterior cervical discectomy and fusion with the use of grafts during the follow up period, Shriver MF. et al. (2015), in the 17 studies finally

included with a review until November 10, 2014, found an occurrence of pseudoarthrosis in 35 cases during the postoperative period of patients who were treated with allografts (n: 736).8 What is the probability that during the following year, among the patients who underwent cervical discectomy with an anterior approach and fusion with allograft, that precise case will present with pseudoarthrosis in the postoperative follow up?

Solution:

$$\lambda = \frac{35(Cases\ of\ \ pseudoarthrosis)}{736\ (Cases\ with\ allografts)} = 0.048$$

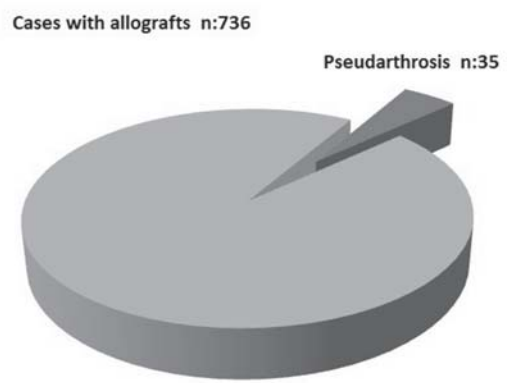
$$x = 2\ (\text{cases expected with pseudoarthrosis})$$

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

$$f(x = 2) = \frac{e^{-0.048} * 0.048^{2}}{2!} = \frac{0.9531 * 0.0023}{2} = 0{,}0011$$

The probability that exactly 2 cases will present with pseudoarthrosis in the follow up of cervical discectomy with an anterior approach and fusion with allograft is 0.11%.

Figure 3 – Frequency of cases of pseudoarthrosis. $p(x = 2)$



2. In a meta-analysis of the correct positioning of pedicular screws using navigation compared to positioning without navigation, Kosmopoulos V, et al. (2007), found in the studies included with revision between 1966 and March 14, 2006 a median of malpositioned screws of 9.7% for the group that did not use navigation in in vivo cases, equivalent to a median monthly occurrence of 8 malpositioned screws.[5] What is the probability that in a randomly selected month it will be one in which 3 or 4 cases with pedicular screws that have been malpositioned in surgery will present themselves?

Solution:

$$p(x = 3 \text{ o } 4) = p(x = 3) + p(x = 4)$$

$$\lambda = \frac{8\ (Cases\ of\ malpositioned\ screws)}{1\ (Unit\ of\ time\ in\ months)} = 8$$

$$x_1 = 3\ (Cases\ expected\ of\ malpositioned\ screws)$$

$$x_2 = 4\ (Cases\ expected\ of\ malpositioned\ screws.)$$

$$f(X = x_1) = \frac{e^{-\lambda}\,\lambda^x}{x!} \rightarrow f(x_1 = 3) = \frac{e^{-8} * 8^3}{3!} = \frac{0.00034 * 512}{6} = 0{,}029$$

$$f(X = x_2) = \frac{e^{-\lambda}\,\lambda^x}{x!} \rightarrow f(x_2 = 4) = \frac{e^{-8} * 8^4}{4!} = \frac{0.00034 * 4096}{24} = 0{,}057$$

$$f(x_1 = 3 \text{ o } x_2 = 4) = f(x_1 = 3) + f(x_2 = 4)$$

$$f(x_1 = 3) + f(x_2 = 4) = 0{,}029 + 0{,}057 = 0{,}086$$

The probability that 3 or 4 cases with surgically malpositioned pedicular screws will appear is 8.6%.

3. In a multi-center prospective study on the correct positioning of the pedicular screws using the O-arm imaging system and Stealth Station navigtion at the thoracic, lumbar and sacral levels, Van de Kelft E, et al. (2012), at 3 neurosurgery centers in 2 countries, between November 2009 and April 2011, found an occurrence of screw malpositioning at around 2.5% on which they report an annual occurrence of 5 screws compromising the medial cortical of the pedicle.[9] What is the probability that during the following year among the patients in whom posterior fixation is performed with pedicular screws, at least 4 malpositioned pedicular screws will present with a compromised medial cortical?

Solution:

The possiblity that events will occur is infinite and the sum total of all possible events is equal to 1. Therefore we can use the concept of complementarity when we subtract from 1 the probability of 3 or less malpositioned screws. In other words:

$$p(x \geq 4) = 1 - p(x \leq 3)$$

According to this, one should first look at the accumulated probability of $p(x \leq 3)$.

$$p(x \geq 4) = 1 - [p(x = 0) + p(x = 1) + p(x = 2) + p(x = 3)]$$

$$p(X \leq x) = \sum_{0}^{x} \frac{e^{-\lambda} \lambda^{x}}{x!} \quad \rightarrow \quad p(X \leq 3) = \sum_{0}^{3} \frac{e^{-5} 5^{3}}{3!}$$

$$= p(x = 0) + p(x = 1) + p(x = 2) + p(x = 3)$$

Then:

$$\lambda = \frac{5\ (Cases\ of\ malpositioned\ screws)}{1\ (Unit\ of\ time\ in\ years)} = 5$$

$$x = 0, 1, 2 \text{ y } 3\ (Cases\ expected\ of\ malpositioned\ screws)$$

$$f(X \geq x) = 1 - \sum_{0}^{x} \frac{e^{-\lambda} \lambda^{x}}{x!}$$

$$f(X \geq 4) = 1 - \sum_{0}^{3} \frac{e^{-5} 5^{3}}{3!}$$

$$p(x \geq 4) = 1 - [p(x = 0) + p(x$$

$$p(x \geq 4) = 1 - \left[\frac{e^{-5} * 5^{0}}{0!} + \frac{e^{-5}}{}\right.$$

$$p(x \geq 4) = 1 - [0{,}0067 + 0{,}0337 + 0{,}0842 + 0{,}1404]$$

$$p(x \geq 4) = 1 - 0{,}2650 = 0.7350$$

Table 1 – Calculation of Poisson distribution for density and accumulated for x= 0 to x=3 with λ=5.

X	P(X=x) FDP	P(Xd"x) FDA
0	0.0067	0.0067
1	0.0337	0.0404
2	0.0842	0.1247
3	0.1404	0.2650

The probability that 4 or more cases will occur per year of malpositioned pedicular screws and medial cortical copromised during surgery is 73.50%.

Table 2 – Transformation of the Poisson formula for calculation using Excel

Application
How to calculate the Poisson probability distribution using Excel.

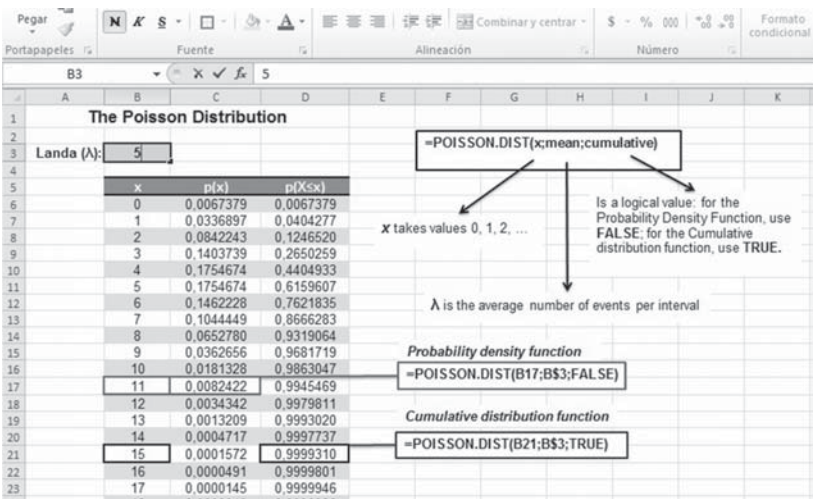
	Function	Formula in Excel
Density probability	$f(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!}$	→ = POISSON.DIST(x; λ; 0)
Accumulated probability	$f(X\leq x)=\sum_{0}^{x}\frac{e^{-\lambda}\lambda^{x}}{x!}$	→ = POISSON.DIST(x; λ; 1)

Table 3 – Transformation of the Poisson formula for calculating in SPSS

Application
How to calculate the Poisson probability distribution using Excel.

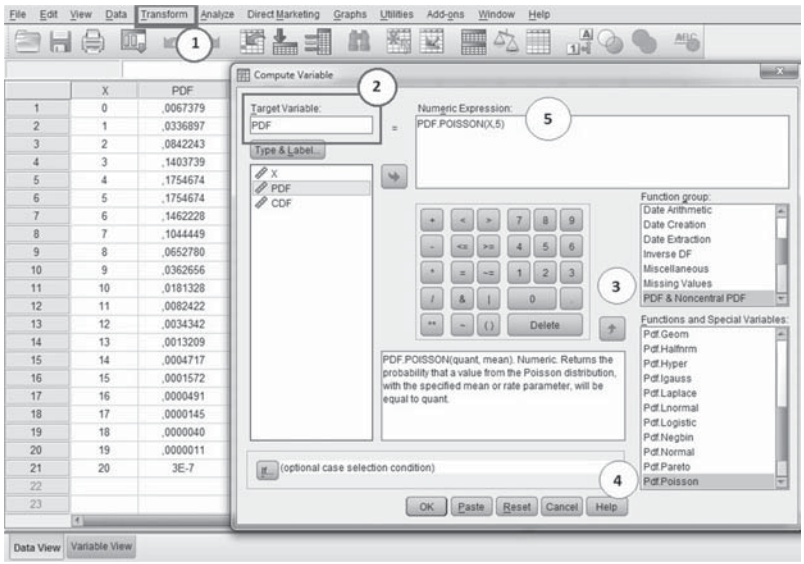
	Function	Formula in Excel
Density probability	$f(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!}$	→ = POISSON.DIST(x; λ; 0)
Accumulated probability	$f(X\leq x)=\sum_{0}^{x}\frac{e^{-\lambda}\lambda^{x}}{x!}$	→ = POISSON.DIST(x; λ; 1)

Figure 4 – The components of the formula are



x	p(x)	p(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

Figure 5 – Enter the values of the expected number of events of the phenomenon (X). Select in transform the function calculate variable (1). Name the variable for the probability density function (FDP) (2). Select in the group of functions "FDP & FDP not centered" (3). In special functions and variables double click to select "Pdf. Poisson" (4). Complete the formula substituting in "(?,?)" first by variable X followed by the value of the parameter hope of occurrence (λ) (5). Finalize with "Accept" to execute the calculation



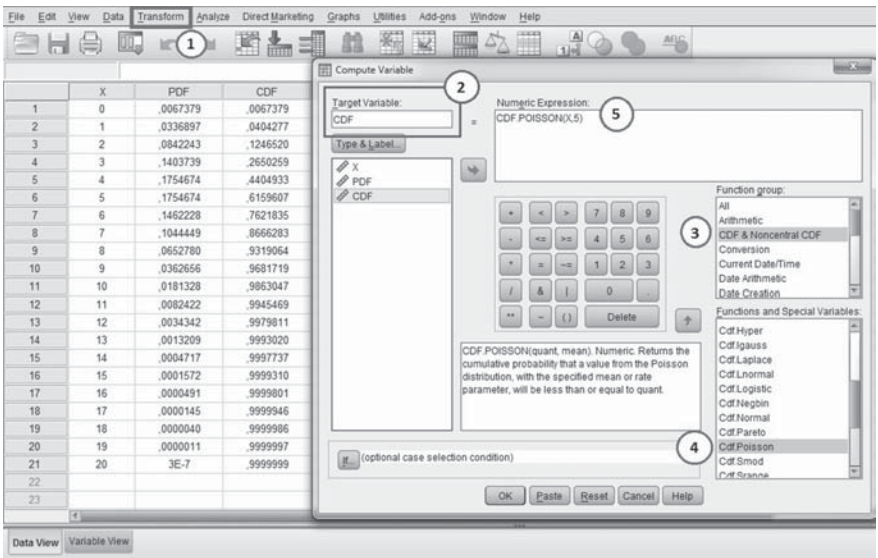
	X	PDF
1	0	,0067379
2	1	,0336897
3	2	,0842243
4	3	,1403739
5	4	,1754674
6	5	,1754674
7	6	,1462228
8	7	,1044449
9	8	,0652780
10	9	,0362656
11	10	,0181328
12	11	,0082422
13	12	,0034342
14	13	,0013209
15	14	,0004717
16	15	,0001572
17	16	,0000491
18	17	,0000145
19	18	,0000040
20	19	,0000011
21	20	3E-7
22		
23		

Interpretation

The Poisson distribution is very useful when one wishes to know the possibility that a number of not very frequent events will appear and/or their impact on the development of a surgical procedure involving the spine. We can get to know the possibility that a specific number of events in which we are interested will appear (example 1), and also a set of possible appearances in a given period of time (examples 2, 3).

On the other hand, the Poisson distribution allows evaluations of strategies to minimize the appearance of undesired events in spine surgery, as in the cases of pedicular screw malpositioning (Ex. 2 and 3). By calculating this we could predict changes in the possibilties of the appearance of undesable events in the future, during defined periods of time, and thus enable controlling the number of events that will present in this type of surgery. The method is applicable to any less frequent event.

Figure 6 –Enter the values of the expected number of events of the phenomenon (X), Select in transform the function calculate variable (1). Name the variable for the accumulated distribution function (FDA) (2). Select in the group of functions "FDA & FDA not centered" (3). In special functions and variables double click to select "Cdf. Poisson" (4). Complete the formula substituting in "(?,?)" first by variable X followed by the value of the parameter hope of occurrence (λ) (5). Finalize with "Accept" to execute the calculation



COMPLEMENTARY READING

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, Kshettry 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.

Chapter 26

ODDS RATIO

José Manuel Pérez Atanasio
Víctor Fernando Luján Celis
Maritza Belén Sandoval Rincón

Test description

The odds ratio is a statistical test that corresponds to the ratio between the probability that an event will occur and the probability that it will not occur, and expresses the risk of presenting a condition or disease if there any risk factor present.

Formula

$$OR=\frac{(a \times d)}{(b \times c)}$$

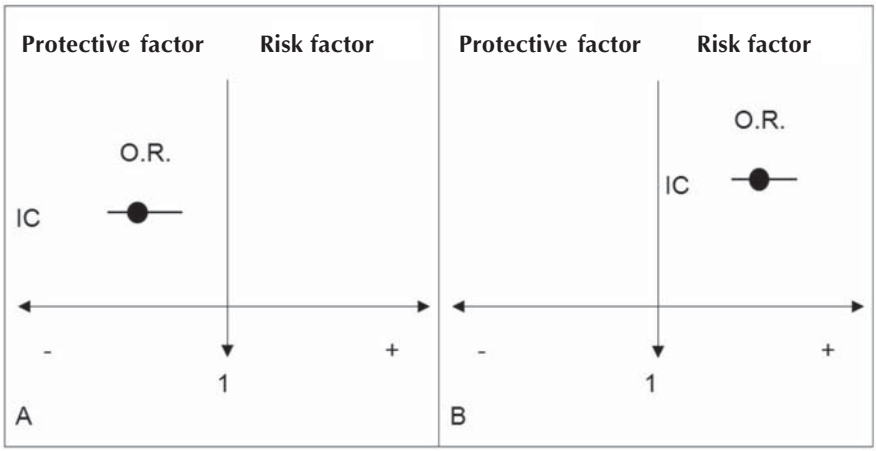
In order to obtain the odds ratio first it is necessary to elaborate a contingency table or 2x2 table, containing the spaces a, b, c, d.

Figure 1 – Contingency or 2x2 table

	Patients who developed the disease	Patients who did not develop the disease
Patients with the risk factor	**a**	**b**
Patients without the risk factor	**c**	**d**

The value obtained from the odds ratio shows us whether the factor studied is a protective factor (<1) or a risk factor (>1) to develop the disease, and this can be exemplified by the following image.

Figure 2 – Graphic representation of the odds ratio as a protective factor (A) and as a risk factor (B)



Example in spine surgery

Dr. Juan, a spine surgeon, has worked for over 20 years at X Hospital, and of late he has seen an increase in surgical site infections in postoperative patients who had undergone surgery to place instrumentation in the lumbar spine. He also observed an increase obesity in patients, and therefore he asks himself the following research question: Is obesity a risk factor for postoperative surgical site infection in patients who underwent instrumentation of the lumbar spine?

To answer this question he elaborated a research protocol with a cases-controls methodological design. He reviewed the charts of 269 patients and identified those who presented surgical site infection (cases) and the patients who did not (controls), and also identified those that were obese, considering this as a risk factor. Once he had obtained the data he elaborated the following table.

Table 1 – Contingency table to calculate the odds ratio

	Patients who developed infection (cases)	Patients who did not develop infection(controls)	
Obese patients	**a**42	**b**55	97
Patients who were not obese	**c**28	**d**144	172
	70	199	269

a = obese patients who developed an infection

b = obese patients who did not develop an infection

c = patients who were not obese and developed an infection

d = patients who were not obese and did not develop an infection

With these data, he set up the formula mentioned previously.

$$\text{Odds ratio} = \frac{(\mathbf{a})\ 42 \times (\mathbf{d})\ 144}{(\mathbf{b})\ 55 \times (\mathbf{c})\ 28}$$

In order to complement the data he used a statistical analysis software to obtain a 95% confidence interval, and the value of ***p*** obtaining the following results:

Odds ratio =3.9

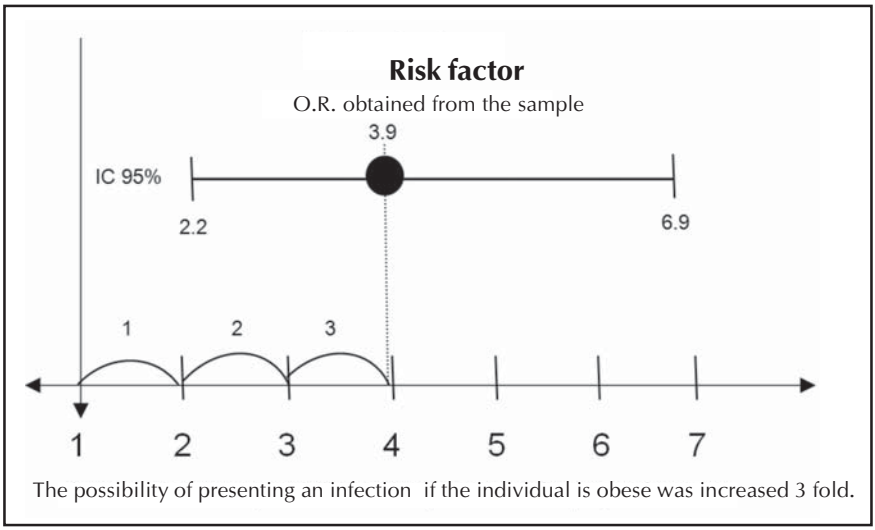
Confidence interval (95%): 2.2 to 6.9

p: <0.0001

Interpretation

The odds ratio and its confidence interval obtained can be graphically located on a horizontal line whose central point in number 1 as a reference value. (Figure 4).

Figure 4 – Graphic representation of the Odds Ratio obtained



The interpretation of these results is that obese patients have a 3 fold greater probability of presenting an infection at the sugical site, than those that are not obese.

The odds ratio is always expressed together with its confidence interval, usually 95%. This confidence interval means, then, that if 100 obese patients were chosen by chance, 95% of the time they would have an odds ratio between 2.2 nd 6.9. Confidence intervals with values close to the odds ratio show that a sufficient number of individuals have been analyzed. On the other hand if the result has distant values, this indicates that the number of subjects studied was very small.

The value of p guides us regarding the probability that the data obtained are due to chance, and in that case we interpret that there is a very small probability (0.001) that these data are due to chance.

COMPLEMENTARY READING

1. Aedo.S.M., Pavlov. S.D, Clavero.F.Ch., Riesgo Relativo y odss 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. 4ª ed., México: Manual Moderno; 2005.

5. Pérez A. J.M., Principios de bioestadística para comprender los resultados reportados en los artículos científicos, Medigraphic, vol.11 n, 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. 4ª ed., México: Limusa Wiley, 2011.

Unit 4

Chapter 27
What gives value and force to the evidence
Mohsen Sadeghi Naini
Vafa Rahimi Movaghar

CHAPTER 27

WHAT GIVES VALUE AND FORCE TO THE EVIDENCE

Mohsen Sadeghi Naini
Vafa Rahimi Movaghar

Level of evidence (LoE) was first used by Canadian task force of the periodic health examination in 1979 when they surveyed the relevant world literature to identify 128 potentially preventable medical conditions(1). In that report, the effectiveness of medical interventions was classified into three main categories based on the quality of the evidence. (Table 1)

Table 1 – Level of evidence by Canadian task force

	Effectiveness of intervention
I	Evidence obtained from at least one properly randomized controlled trial.
II-a	Evidence obtained from well designed cohort or case-control analytic studies, preferably from more than one center or research group.
II-b	Evidence obtained from comparisons between times or places with or without intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin in the 1940s) could also be regarded as this type of evidence.
III	Opinions of respected authorities, based on clinical experience, descriptive studies or reports of expert committees.

In there, quality of the evidence was defined with regard to three main concepts:

1. Randomized controlled trials (RCTs), or systematic reviews of many RCTs, generally bringing stronger evidential support than observational studies.

2. Comparative clinical studies offer stronger evidential support than "mechanistic" reasoning ("pathophysiologic rationale") from more basic sciences.

3. Comparative clinical studies offer stronger evidential support than expert opinion.

The evidence was then taken into account to score recommendations from 1 to 5 (Table 2).

Table 2 – Grading of recommendation by Canadian task force

	Classification of recommendations
A	There is good evidence to support the recommendation that the condition be specifically considered in a periodic health examination.
B	There is fair evidence to support the recommendation that the condition be specifically considered in a periodic health examination.
C	There is poor evidence regarding the inclusion of the condition in a periodic health examination, and recommendations may be made on other grounds.
D	There is fair evidence to support the recommendation that the condition be excluded from consideration in a periodic health examination.
E	There is good evidence to support the recommendation that the condition be excluded from consideration in a periodic health examination.

This report was updated with minor modifications in the following years (2,3). Since then, LoE has been widely used and criticized by different working groups to construct their own scheme (4–6). Each scheme was designed to answer different questions for different stakeholders including clinical physicians, researchers and policy makers. Among evolutionary processes of LoE, one important improvement to the LoE was made by the center for evidence-based medicine (CEBM) in September 2000, which accelerated the process of evidence finding for clinical physician.

At the time, the common problem with many of the evidence ranking schemes was the fact that they graded evidence just for therapy and prevention without any focus on diagnostic tests, prognostic markers, or harm. Therefore, CEBM team updated their proposed LoE to include different aspects of health problems including therapy, prevention, etiology, prognosis, diagnosis, differential diagnosis and economic and decision analyses (7) (Table 3).

Table – 3 Level of evidence proposed by CEBM 2009

Level	Therapy / Prevention, Etiology / Harm	Prognosis	Diagnosis	Differential diagnosis / symptom prevalence study	Economic and decision analyses
1a	Systematic review (with homogeneity) of RCTs	Systematic review (with homogeneity) of inception cohort studies; CDR″ validated in different populations	Systematic review (with homogeneity) of Level 1 diagnostic studies; CDR″ with 1b studies from different clinical centers	Systematic review (with homogeneity) of prospective cohort studies	Systematic review (with homogeneity) of Level 1 economic studies
1b	Individual RCT (with narrow Confidence Interval)	Individual inception cohort study with > 80% follow-up; CDR validated in a single population	Validating cohort study with good reference standards; or CDR tested within one clinical centre	Prospective cohort study with good follow-up	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses
1c	All or none§	All or none case-series	Absolute SpPins and SnNouts	All or none case-series	Absolute better-value or worse-value analyses
2a	SR (with homogeneity) of cohort studies	SR (with homogeneity) of either retrospective cohort studies or untreated control groups in RCTs	SR (with homogeneity) of Level >2 diagnostic studies	SR (with homogeneity) of 2b and better studies	SR (with homogeneity) of Level >2 economic studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR″ or validated on split-sample only	Exploratory cohort study with good reference standards; CDR″ after derivation, or validated only on split-sample or databasesetter studies	Retrospective cohort study, or poor follow-up	Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses

Level	Therapy / Prevention, Etiology / Harm	Prognosis	Diagnosis	Differential diagnosis / symptom prevalence study	Economic and decision analyses
2c	"Outcomes" Research; Ecological studies	"Outcomes" Research		Ecological studies	Audit or outcomes of research
3a	SR (with homogeneity) of case-control studies		SR (with homogeneity) of 3b and better studies	SR (with homogeneity) of 3b and better studies	SR (with homogeneity) of 3b and better studies
3b	Individual Case-Control Study		Non-consecutive study; or without consistently applied reference standards	Non-consecutive cohort study, or very limited population	Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations
4	Case-series (and poor quality cohort and case-control studies)	Case-series (and poor quality prognostic cohort studies)	Case-control study, poor or non-independent reference standard	Case-series or superseded reference standards	Analysis with no sensitivity analysis
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on economic theory or "first principles"

Early evidence hierarchies (including CEBM report in 2000) (2,5,8) were introduced primarily to appraise the quality of evidence for therapeutics. These hierarchies placed RCTs categorically above observational studies and expert judgment below uncontrolled observations that made evidence classification simple and easy to use.

In the recent years, this approach was criticized due to simplicity (9). For example, some adverse drug reactions are so convincing that a well documented subjective report can provide definitive evidence of a causal association and further verification is not needed (10). In addition, the ability to show that patients with a previously universally fatal disease can survive following a new treatment establishes sufficient evidence for efficacy (11). Some historical examples of treatments with dramatic effects that come from observational studies include:

- Insulin for diabetes;
- Blood transfusion for severe hemorrhagic shock;
- Defibrillation for ventricular fibrillation;
- Closed reduction and splinting for fracture of long bones with displacement;
- Neostigmine for myasthenia gravis;
- Tracheostomy for tracheal obstruction;
- Suturing for repairing large wounds;
- Streptomycin for tuberculous meningitis;
- Drainage for pain associated with abscesses;
- Pressure or suturing for arresting hemorrhage;
- Ether for anesthesia;
- One-way valve or underwater seal drainage for pneumothorax and haemothorax.

On the other hand, no one can perform falling from airplane with and without parachute to achieve level 1 evidence for the role of parachute (12). Although the current opinion is on considering all types of available evidences (including observational studies) when making recommendations, it should also be kept in mind that since the prognosis and the treatment effect interact as noise and signal, not all treatment effects are so easy to demonstrate (13).

Considering these foible points, more recent efforts avoid the common abjection of upgrading evidence just by study design (clinical

trial versus observational studies) by allowing observational studies with dramatic effects to be upgraded the LoE (4). This change in the interpretation of the evidence quality was another evolution in the LoE.

In 2006, Grading of Recommendations, Assessment, Development and Evaluations (GRADE) Working Group proposed a systematic and clear approach to making judgments on the quality of evidence and strength of recommendations. GRADE classifies the quality of a body of evidence into four levels: high quality, moderate quality, low quality and very low quality.

Evidence quality in GRADE approach may change with regard to the study design and risk of bias, inconsistency of results, indirectness (not generalizable), imprecision (sparse data) and other factors (e.g. reporting bias) (14). This approach becomes popular by a systematic review very soon and is now widely seen as the most effective method of linking evidence-quality evaluations to clinical recommendations (4). GRADE approach rates the evidence not study-by-study, but across studies for specific clinical outcomes.

This approach specifically assesses four domains in the included studies of the systematic reviews, health technology assessments and guidelines including (4):

- Methodological flaws within the component studies
- Consistency of results across different studies
- Generalizability of research results to the wider patient base
- How effective treatments have been shown to be.

However, what GRADE has added in accuracy, it might have lost in simplicity. Therefore, it may be unusable for busy clinicians when searching to find the best available evidence for a specific clinical problem (7). One important conflict between GRADE and OCEBM is that GRADE approach assesses the strength of evidence instead of single studies, but OCEBM Levels focus on its specific audience (clinicians or patients) and make its tool-suitable to evaluate individual studies in case a systematic review is unavailable.

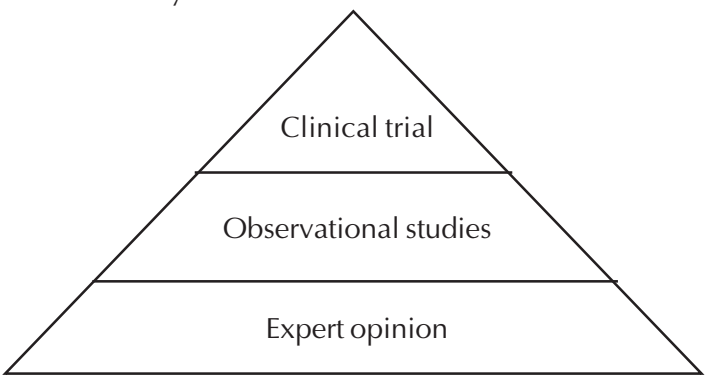
LOE

Levels of evidence explain the idea that there are varying degrees or levels of evidence for a specific "cause and effect" relationship in a particular scientific era that different kinds of **study design** can be

delivered. In other words, when talking about LOE, we focus on the classification of methodological design of studies assessing the effects of health-care interventions. In real world, it is a rare case when a single study demonstrates a hypothesis so irrefutably that it is beyond query. Usually, different studies add different levels and/or kinds of evidence that finally construct the body of evidence.

Traditionally, LOE is presented in a hierarchy of weight with reference to the strength of the association between the cause and effect that particular study designs incorporate (Figure 1).

Figure 1 – Hierarchy of level of evidence



Animal studies constitute the first step in providing medical evidence to generate and test hypotheses. This will also provide the infrastructure to develop mechanistic reasoning for the commonly accepted rules.

Expert opinions without critical appraisal or based on clinical reasoning, bench research or "first principles" provide the second level of evidence.

Uncontrolled studies are divided in observational (e.g. case report and case series) and experimental (e.g. before/after trials and N-of-1 randomized trial) groups according to the type of intervention.

Observational studies with control groups: These include cohort study, case control study and cross-sectional study where the investigator is describing what is naturally occurring without disrupting the natural scenario. This type of study is most commonly used to note associations between events and to generate hypotheses.

Experimental studies with control groups most commonly known as clinical trial (including randomized or non-randomized) actively

intervene by assigning some participants to receive a treatment and others to receive a different intervention (or no intervention). Obviously, experimental studies added a stronger value to the process of building the body of evidence.

Meta-analyses and systematic reviews, on the other hand, summarize groups of studies on a specific research question to evaluate the weight of the evidence on that question (Table 4).

Table 4 – Taxonomy of Study Design of Studies Assessing the Effects of Health-Care Interventions

Systematic review	Systematic review with meta-analysis
Experimental studies with control group ("clinical trials" or "trials"): The investigator has control over the decision concerning the allocation of participants to different intervention groups	• Systematic review without meta-analysis • Randomized controlled trial (RCT) • Quasi-randomized controlled trial (Q-RCT) • Controlled clinical trial (CCT)
Observational studies with control group: The investigator's intention is to observe and not to interfere with routine care	• Cohort study • Survival cohort study Case • Case control study • Cross-sectional study
Uncontrolled studies (without a separate control group): can be experimental or observational in nature	• Before/after trials • N-of-1 randomized trial • Case series • Case reports
Expert opinion	
Animal studies	

Besides various positive points in the hierarchical view of OCEBM LOE, it makes some important assumptions which are not generally valid (e.g. in complex interventions) but are rarely debated (15):

• **Equipoise**: Patient and clinician do not have a preference for a treatment.

• **Lack of knowledge**: It is truly unknown which of the two choices is "better" and there is inadequate evidence on treatment effects from other sources.

• **Preference for specificity**: Only specific effects attributable to the intervention are therapeutically valid.

• **Context independence**: There is a "true" extent of efficacy or a constant effect size independent of context.

• **Ecological and external validity**: Knowledge on a therapeutic effect extracted from an RCT is readily assignable into the clinical practice, if eligibility criteria of the trial match the characteristics of a given patient.

The alternative to the hierarchical view of the level of the evidence is a circular one. It's derived from the experience that every research method has strengths and limitations that cannot be fixed by that method itself. Therefore, triangulating a result achieved with one method by repeating it with other methods may provide a more powerful conclusion (15).

Type of biases and tools for assessing bias

Strength of evidence may be violated by biases (or systematic error) and confounders.

Biases can deviate the result in either direction (underestimation or overestimation of the true intervention effect) with different in magnitude. It is usually impossible to measure the impact of bias on the results of a particular study (16).

Bias may be confused with imprecision. Bias refers to a systematic error, which do not change by multiple replications of the same methodology. Imprecision raises to random error, which mean sampling variation, can lead to result fluctuation in multiple replications of the same methodology even if all those replications would give the right answer on average. We can treat the imprecision by increasing the sample size but biases do not have such effect. Imprecision is mirrored in the confidence interval around the analyzed result, so more precise results lead to more narrow confidence interval and vise versa.

Biases are categorized into five main classes. (Table 5).

Table 5 – Cochrane classification scheme for bias

Type of bias	Description	Relevant domains in the Cochrane 'Risk of bias' tool
Selection bias	Systematic differences between baseline characteristics of the groups that are compared	• Sequence generation. • Allocation concealment
Performance bias	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest	• Blinding of participants and personnel • Other potential threats to validity
Detection bias	Systematic differences between groups in how outcomes are determined	• Blinding of outcome assessment • Other potential threats to validity
Attrition bias	Systematic differences between groups in withdrawals from a study	• Incomplete outcome data
Reporting bias	Systematic differences between reported and unreported findings	• Selective outcome reporting

Three types of tools are proposed to evaluate bias in clinical research, specifically clinical trials: scales, checklist and domain-based evaluations. Moher and colleagues identified 34 tools that had been used to assess biases in randomized trials (17). The Cochrane Collaboration's recommended tool for assessing risk of bias is a domain-based evaluation, in which critical assessments are separated for different domains (16). It is difficult to demonstrate the extent of bias in a given study, therefore, the possibility of validating any proposed tool is limited. (Table 6).

Table 6 – The Cochrane Collaboration's tool for assessing risk of bias.

Domain	Support for judgment	Review authors' judgement
Selection bias		
Random sequence generation	Describe the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups.	Selection bias (biased allocation to interventions) due to inadequate generation of a randomized sequence.
Allocation concealment	Describe the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen in advance of, or during, enrolment.	Selection bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment.
Performance bias		
Blinding of participants and personnel Assessments should be made for each main outcome (or class of outcomes)	Describe all measures used, if any, to blind study participants and personnel from knowledge of which a participant receives. Provide any information relating to whether the intended blinding was effective.	Performance bias due to knowledge of allocated interventions by participants and personnel during the study.
Detection bias		
Blinding of outcome assessment *Assessments should be made for each main outcome (or class of outcomes)*	Describe all measures used, if any, to blind outcome assessors from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.	Detection bias due to knowledge of the allocated interventions by outcome assessors.
Attrition bias		
Incomplete outcome data *Assessments should be made for each main outcome (or class of outcomes)*	Describe the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. State whether attrition and exclusions were reported, numbers in each intervention group (compared with total randomized participants), reasons for attrition/exclusions were reported, and any re-inclusions in analyses were performed by the review authors.	Attrition bias due to amount, nature or handling of incomplete outcome data.
Reporting bias		
Selective reporting	State how the possibility of selective outcome reporting was examined by the review authors, and what was found.	Reporting bias due to selective outcome reporting.
Other bias		
Other sources of bias	State any important concerns about bias not addressed in the other domains in the tool If particular questions/entries were pre-specified in the review's protocol, responses should be provided for each question/entry.	Bias due to problems not covered elsewhere in the table.

What may violate the strength of evidence?

There are potential factors that may decrease the quality of the body of the evidence derived from pooled data analysis from RCT, as proposed in GRADE approach(4).

These points evaluate different kinds of biases in this particular study design and then judge on the overall strength of evidence derive from included clinical trials. Therefore, this approach is described for RCT as the fundamental block in the process of evidence production.

1. Limitations in the design and implementation: Each study design has some potential limitations in the design that should be addressed when incorporating its result into a clinical setting or research query. For randomized trials, the potential issue to be addressed include:

- allocation concealment;
- blinding (particularly when outcomes are assessed subjectively);
- loss to follow-up;
- stopped early for benefit or selective reporting of outcomes.

2. Indirectness of evidence: When using study results, the eligibility criteria met in that study should be evaluated and compared to the present clinical setting. For example, a restricted version of the main question in terms of population, intervention, comparator or outcomes may introduce indirectness of evidence. Other sources of indirectness may arise from interventions studied (e.g. if in relevant study a surgical intervention was implemented by a senior surgeon in specialist centers, then evidence on the effects of the intervention outside these centers may be indirect), comparators used (e.g. if the intervention in the control groups is less effective than the standard treatment) and outcomes assessed (e.g. indirectness due to alternate outcomes).

3. Unexplained heterogeneity or inconsistency of results: Widely differing estimates of effect (heterogeneity in other words) without reasonable explanation could decrease the quality of evidence.

4. Imprecision of results: Wide confidence intervals independent of the heterogeneity can lead to decreasing the quality of the evidence. This can be treated by increasing the sample size.

5. High probability of publication bias: The quality of evidence level may be downgraded if investigators fail to report studies (classically those that show no effect: publication bias) or outcomes (classically those that may be harmful or for which no effect was observed: selective outcome reporting bias) on the basis of results. The heavy influence of sponsors in a trial raises questions of whether unpublished trials suggesting no benefit exist.

Acknowledgement

We would like to appreciate English editing by Mrs. Bita Pourmand, affiliated to Research Development Center, Sina Hospital, Tehran University of Medical sciences.

COMPLEMENTARY READING

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 perodic health examina 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 Evicence-BAsed Medicine (OCEBM) Levels of Evicende (Backgroud 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.

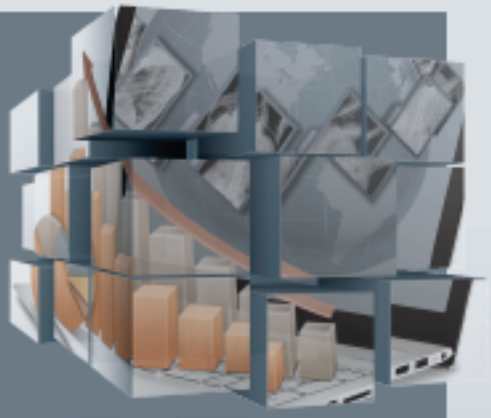
13. Glasziou P. When are randomised trials unnecessary/ ? Picking signal from noise Paul Glasziou and colleagues When are randomised trials unnecessary/ ? Picking signal from noise. 2016; (May).

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, Fonnebo 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 510 (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.



When we have a clear understanding of our question, which will lead us to perform our research, we will have the result that allows us to strengthen our conclusion by Inference and for this we need to know the theoretical aspects involved in Clinical Biostatistics. From these we will derive concepts that will allow us to make decisions and apply the knowledge obtained to our patients and thus accomplish with excellence our role in preventing and treating diseases, incorporating the concepts to clinical practice.

From Measurement to Interpretation, an essential process that connects research to the everyday world and to reality, where our hypotheses, surrounded by probabilities, seek to find the value that will establish whether there are or not differences between the options.

It should not be forgotten that the numbers are raw, and know nothing of reasons, and that is where our knowledge and theory will give true value to the data, since these do not make us certain of any event, they simply show us the probability that something may occur.

José María Jiménez Avila
Regional Research Officer–AOSpine Latin America

Asdrubal Falavigna
Chairperson–AOSpine Latin America

ISBN 978-85-7061-912-9
9 788570 619129