






## Métodos de Análisis Estadísticos / Methods of Statistical Analyses

## Métodos de Análisis Estadísticos / Methods of Statistical Analyses

### Definiciones Importantes

- **Variable:** Un símbolo al que se asignan números o valores.
  - **Definición constitutiva:** Define una construcción con otras construcciones.
  - **Definición operacional:** Se mide:
    - Describe cómo se medirá una variable.
    - Exponen los detalles de las manipulaciones del investigador de la variable.

### Important Definitions

- **Variable:** A symbol to which numerals or values are assigned.
  - **Constitutive definition:** Defines a construct with other constructs.
  - **Operational definition:** It is measured:
    - Describes how a variable will be measured.
    - Spells out the details of the investigator's manipulations of the variable.

### Definiciones Importantes

- **Variable independiente:** una variable manipulada en un experimento o estudio cuya presencia o grado determina el cambio en la variable dependiente. (predicho de)
- **Variable dependiente:** La variable observada en un experimento o estudio cuyos cambios están determinados por la presencia o el grado de una o más variables independientes. (previsto para)
- **Variable Moderadora:** una variable cualitativa o cuantitativa independiente que afecta a la relación de las variables dependientes e independientes. O bien, puede afectar a la correlación de las variables.

### Important Definitions

- **Independent Variable:** A manipulated variable in an experiment or study whose presence or degree determines the change in the dependent variable. (predicted from)
- **Dependent Variable:** The observed variable in an experiment or study whose changes are determined by the presence or degree of one or more independent variables. (predicted to)
- **Moderator Variable:** an independent qualitative or quantitative variable that affects the relationship of the dependent and independent variables. OR, it can affect the correlation of the variables.

## Escalas de Medición de las Variables

- Nominal, Ordinal, Intervalo y la Proporción
- **Nominal:** Ofrece nombres o etiquetas para determinadas características (ex: las rocas se pueden clasificar como ígneas, sedimentarias, etc.)
  - Tendencia central puede describirse por modo.
  - No pueden definirse la media y la mediana.
- **Ordinal:** Clasifica los datos.
  - Tendencia central puede describirse por el modo o la mediana, pero no la media.

## Scales of Measurement of Variables

- Nominal, Ordinal, Interval, and Ratio
- **Nominal:** Offers names or labels for certain characteristics (ex: rocks can be categorized as igneous, sedimentary, etc.)
  - Central tendency can be described by mode.
  - Mean and median cannot be defined.
- **Ordinal:** Ranks the data.
  - Central tendency can be described by mode or median, but not mean.

## Escalas de Medición de las Variables

- Nominal, Ordinal, Intervalo y la Proporción
- **Intervalo:** Se utilizan variables escaladas o de intervalo.
  - Mide los atributos cuantitativos (ex: temperatura con grados Celsius o escala de Likert).
  - La tendencia central puede medirse a nivel de intervalo por modo, mediana o media aritmética.
- **Relación:** La mayoría de ciencia física se mide así: masa, longitud, tiempo, ángulo, energía, etc.
  - Realmente no existe en ciencias del comportamiento.
  - Puede medir media, mediana y modo.

## Scales of Measurement of Variables

- Nominal, Ordinal, Interval, and Ratio
- **Interval:** Interval or scaled variables are used.
  - Measures quantitative attributes (ex: temperature w/Celsius scale or the Likert scale).
  - Central tendency can be measured at the interval level by mode, median, or arithmetic mean.
- **Ratio:** Most physical science is measured this way: mass, length, time, angle, energy, etc.
  - Doesn't really exist in behavioral science.
  - Can measure mean, median, and mode.

## Otras Medidas

- **Medidas de Tendencia Central:** El centro de la distribución, usualmente medida como media, mediana o modo.
  - Se utiliza para medir valores "normales" o "promedio" de un conjunto de datos
- **Medida de dispersión:** Describe la propagación de los datos, o su variación alrededor de un valor central.
  - Se mide por rango, rango intercuartil, varianza, desviación estándar, media cuadrática, desviación absoluta de la media o variación recortada.

## Other Measures

- **Measures of Central Tendency:** The center of the distribution, usually measured as the mean, median, or mode.
  - It is used to measure "normal" or "average" values of a dataset.
- **Measure of Dispersion:** Describes the spread of the data, or its variation around a central value.
  - It is measured by range, interquartile range, variance, standard deviation, root mean square, median absolute deviation, or trimmed variance.

## Variables

- **Variables cuantitativas:** medidas en una escala numérica o cuantitativa.
  - La escala ordinal, de intervalo y de proporción son cuantitativas.
  - Es normalmente más **generalizable**. (ex: población del país, zapato de tamaño, velocidad del coche, etc..)
  - Generalmente implica el uso del **método científico** convencional, que incluye:
    - Generar una hipótesis
    - Desarrollar instrumentos y métodos
    - Control experimental y manipulación de las variables
    - Recopilar datos
    - Análisis de datos
    - Evaluar resultados

## Variables

- **Quantitative Variables:** measured on a numeric or quantitative scale.
  - Ordinal, interval, and ratio scales are quantitative.
  - It is typically more **generalizable**. (ex: a country's population, shoe size, car speed, etc.)
  - It generally involves use of the conventional **scientific method**, which includes:
    - Generate a hypothesis
    - Develop instruments and methods
    - Experimental control and manipulation of variables
    - Collect data
    - Analyze data
    - Evaluate results

## Variables

- **Variables cualitativas:** El investigador puede describir el interés con gran detalle.
  - Las variables nominales y ordinales pueden ser cualitativas.
  - Medición viene más de entrevistas y datos "sentidos"
  - No se preocupa mucho de ser generalizables.

## Variables

- **Qualitative Variables:** The researcher can describe the interest in great detail.
  - Nominal and ordinal variables can be qualitative.
  - Measurement comes more from interviews and "touchy-feely" data.
  - Not concerned very much with being generalizable.

## Tipos de Análisis

- **Análisis uni variado:** Método para analizar los datos de una sola variable a la vez. (ex: reportar el número de embarazos de adolescentes en nuestro condado desde 1995).
- **Análisis multi variante:** Método para realizar la observación simultánea y el análisis de más de una variable. Identifica las relaciones estadísticas entre variables (ej. relación entre el embarazo adolescente y el uso intermitente de control de la natalidad).

## Types of Analyses

- **Univariate Analysis:** Method for analyzing data on a single variable at a time. (ex: report the number of teenage pregnancies in our county since 1995).
- **Multivariate Analysis:** Method for performing simultaneous observation and analysis of more than one variable. It identifies statistical relationships between variables (ex. Relationship between teen pregnancy and intermittent use of birth control).

## Escalas

- **Escala de respuesta:** Normalmente se utiliza en la investigación en ciencias sociales (investigación con la participación de los encuestados humanos), el rango, grados o dimensiones por los cuales usted quiere que los encuestados midan el concepto de la encuesta presentado
- **Escala de medida:** Sistema de clasificación que se utiliza para describir la medición de conceptos o variables que se utilizan en la investigación.
  - El sistema de clasificación es nominal, ordinal, de intervalo o relación.

## Scales

- **Response Scale:** Typically used in social science research (research involving human respondents), the range, degrees, or dimensions by which you want your survey respondents to measure the survey item concept you've presented.
- **Measurement Scale:** Classification system that is used to describe the measurement of concepts or variables that are used in research.
  - The classification system is nominal, ordinal, interval or ratio.

## Otros Términos

- **Población:** Conjunto de datos a de los cuales se puede sacar medidas *potenciales*
  - No incluye sólo los casos observables.
  - Estamos interesados en todo el grupo del cual queremos describir o hacer conclusiones. (ex: todas las mujeres mayores de 40 en Perú)
- **Muestra:** Una muestra es un grupo de unidades seleccionada de un grupo más grande (la población).
  - Mediante el estudio de la muestra se espera sacar conclusiones válidas sobre el grupo más grande. (ex: mujeres mayores de 40 elegidas al azar en los 3 ciudades)
- **Censo:** datos que se recolectan de todos los miembros de una población.

## Other Terms

- **Population:** A data set from which *potential* measurements can be drawn.
  - Doesn't include only observable cases.
  - It is the entire group we are interested in, which we wish to describe or draw conclusions about. (ex: all females over the age of 40 in Peru)
- **Sample:** A sample is a group of units selected from a larger group (the population).
  - By studying the sample it is hoped to draw valid conclusions about the larger group. (ex: females over the age of 40 randomly chosen in 3 cities)
- **Census:** data that is collected from all members of a population.

## Otros Términos

- **Parámetro:** Un parámetro un valor, usualmente desconocido (y que por tanto tiene que ser estimado), utilizado para representar determinada característica de la población.
  - Por ejemplo, la media de la población es un parámetro que se utiliza a menudo para indicar el valor promedio de una cantidad.
- **Estadística:** Es un valor numérico que caracteriza la muestra de la que se deriva.
  - O bien, una cantidad que se calcula de una muestra de datos.
  - Se utiliza para dar información acerca de valores desconocidos en la población correspondiente.

## Other Terms

- **Parameter:** A parameter is a value, usually unknown (and which therefore has to be estimated), used to represent a certain population characteristic.
  - For example, the population mean is a parameter that is often used to indicate the average value of a quantity.
- **Statistic:** It is a numerical value that characterizes the sample from which it was derived.
  - Or, a quantity that is calculated from a sample of data.
  - It is used to give information about unknown values in the corresponding population.

## Otros Términos

- **Estimado:** Una indicación del valor de una cantidad desconocida basada en datos observados.
  - Usualmente está representado por un sombrero (^) sobre el símbolo. (Cuando una estadística — que representa un valor de una muestra— se utiliza para indicar el valor de un parámetro, el valor de una población — entonces esa estadística se dice que es una estimado — es decir, un estimado del parámetro de la población.
- **Estadísticas aceptables en lo mínimo:** El lector debe tener suficiente información en el artículo para verificar los cálculos.
  - Estadística descriptiva debe incluir el tamaño de muestra, media y SD.

## Other Terms

- **Estimate:** An indication of the value of an unknown quantity based on observed data.
  - It is usually represented by a hat (^) over the symbol. (When a statistic—representing a value from a sample—is used to indicate the value of a parameter—the value of a population—then that statistic is said to be an estimate—that is, an estimate of the population parameter.
- **Minimally acceptable statistics:** The reader should have enough information in the article to verify the computations.
  - Descriptive statistics should include sample size, mean, and SD.

## Estadística Paramétrica

- **Estadística paramétrica:** Supone que los datos proceden de un tipo de distribución de probabilidad y hacen inferencias acerca de los parámetros de la distribución.
  - La estadística paramétrica hace más suposiciones acerca de los datos que los métodos no paramétricos.
  - La estadística paramétrica es más precisa.
  - Requieren de una distribución de probabilidad y por lo general no son tan sólidas como la estadística no paramétrica.

## Parametric Statistics

- **Parametric Statistics:** Assumes data have come from a type of probability distribution and makes inferences about the parameters of the distribution.
  - Parametric statistics make more assumptions about the data than do non-parametric methods.
  - Parametric statistics are more precise.
  - They require a probability distribution and are not usually considered as robust as nonparametric statistics.

## Estadística Paramétrica

- Correlación de momento de producto de Pearson
- **t-test**
  - muestras independientes t-prueba, separar la estimación de la variación
  - muestras independientes t-test, estimación combinada de variación
  - muestras emparejadas t-test (también llamadas t-test dependientes o t-test correlacionadas)
- Análisis de varianza unidireccional
- Análisis de varianza de dos vías o dirección n
- Análisis de covarianza
- Análisis de varianza de medidas repetidas

## Parametric Statistics

- Pearson product moment correlation
- **t-test**
  - independent samples t-test, separate estimate of variance
  - independent samples t-test, pooled estimate of variance
  - paired samples t-test (also called dependent t-test or correlated t-test)
- One-way analysis of variance
- Two-way or n-way analysis of variance
- Analysis of covariance
- Repeated measures analysis of variance

## Estadística Paramétrica

- Análisis mixto de diseño de la varianza
- Regresión simple (análisis de regresión)
- Regresión múltiple
- Análisis de ruta
- Análisis discriminante
- Análisis factorial
  - Análisis factorial exploratorio
  - Análisis factorial de confirmación
- Análisis multi variante

## Parametric Statistics

- Mixed design analysis of variance
- Simple regression (analysis of regression)
- Multiple regression
- Path analysis
- Discriminant analysis
- Factor analysis
  - Exploratory factor analysis
  - Confirmatory factor analysis
- Multivariate analysis

## Estadística no Paramétrica

- **Estadística no paramétrica:** No se basan en la hipótesis de que los datos provienen de una distribución de probabilidad dada.
  - Estadísticas basadas en el rango de observaciones son comunes.
  - Hacer suposiciones menos y son más robustas.
  - Son más simples y fáciles de usar. Por ejemplo:
    - Prueba de Anderson–Darling
    - Métodos estadísticos de Bootstrap
    - Q de Cochran
    - Kappa de Cohen
    - Friedman bidireccional análisis de varianza por filas
    - Kaplan–Meier
    - Tau de Kendall
    - W de Kendall

## Non-Parametric Statistics

- **Non-Parametric Statistics:** Do not rely on assumptions that the data are drawn from a given probability distribution.
  - Statistics based on the rank of observations are common.
  - They make fewer assumptions and are more robust.
  - They are more simple, and easier to use. Examples include:
    - Anderson–Darling test
    - Statistical Bootstrap Methods
    - Cochran's Q
    - Cohen's kappa
    - Friedman two-way analysis of variance by ranks
    - Kaplan–Meier
    - Kendall's tau
    - Kendall's W

## Estadística no Paramétrica

- Prueba de Kolmogorov–Smirnov
- Un análisis de varianza de Kruskal-Wallis por filas
- Prueba de Kuiper
- Prueba de Logrank
- Prueba de rango de suma Mann–Whitney U o Wilcoxon
- Prueba mediana
- Prueba de permutación de Pitman
- Productos de rango
- Prueba de Siegel–Tukey
- Coeficiente de correlación de Spearman
- Wald–Wolfowitz se ejecuta la prueba
- Prueba de Wilcoxon rango firmado

## Non-Parametric Statistics

- Kolmogorov–Smirnov test
- Kruskal-Wallis one-way analysis of variance by ranks
- Kuiper's test
- Logrank Test
- Mann–Whitney U or Wilcoxon rank sum test
- median test
- Pitman's permutation test
- Rank products
- Siegel–Tukey test
- Spearman's rank correlation coefficient
- Wald–Wolfowitz runs test
- Wilcoxon signed-rank test

## Estadísticas de Relaciones

- Relación de dos estadísticas: Estadísticas correlacionadas miden el grado en que dos variables están relacionadas.
  - Estadísticas correlacionales producen un número llamado coeficiente de correlación.
- Estadística inferencial: El investigador está tratando de llegar a conclusiones que van más allá de los datos inmediatos solamente. Utilizamos las estadísticas inferenciales para intentar deducir de los datos de ejemplo de lo que la población podría pensar.
  - O, utilizamos la estadística inferencial para hacer juicios de la probabilidad de que una diferencia observada entre los grupos es confiable o que podría haber ocurrido por casualidad en este estudio.

## Relational Statistics

- Relation of two statistics: Correlational statistics measure the extent to which two variables are related.
  - Correlational statistics yield a number called the coefficient of correlation.
- Inferential statistics: Researcher is trying to reach conclusions that extend beyond the immediate data alone. We use inferential statistics to try to infer from the sample data what the population might think.
  - Or, we use inferential statistics to make judgments of the probability that an observed difference between groups is a dependable one or one that might have happened by chance in this study.

## Hipótesis

- Hipótesis de NULL: La hipótesis de que no hay ninguna validez a la pretensión específica.
  - Quiere decir que la correlación de las poblaciones entre dos variables es igual a cero que la diferencia en la media de dos grupos de la población es igual a cero.
  - En general la declaración null se expresa como ninguna relación entre dos variables o ninguna diferencia entre los dos grupos.
- Hipótesis alternativa: lo contrario del null es una declaración expresando una relación entre dos variables o que indica las diferencias entre los grupos.

## Hypothesis

- Null hypothesis: The hypothesis that there is no validity to the specific claim made.
  - That it states that the populations correlation between two variables is equal to zero that the difference in the means of two groups in the population is equal to zero.
  - In general the null statement is expressed as no relationship between two variables or no difference between two groups.
- Alternative hypothesis: the opposite of the null is a statement expressing a relationship between two variables or indicating differences between groups.