IG: @saulo_pv
IG: @motorpsico.uba

Estadística - Final.

UNIDAD I: datos estadísticos y su distribución de frecuencia.

La **estadística** tiene sentido porque existe **variabilidad** en las características de los objetos de estudio.

La **variabilidad** de las características en estudio lleva a la definición de un concepto clave del curso, que es el de **variable**: una variable es una característica que puede tomar distintas modalidades en los individuos de una población que son objeto de estudio.

Las variables pueden ser:

- **Directamente observables.**
- **No directamente observables:** en este caso hablamos de variable latente, que forma parte de un **constructo**.

Un **constructo** es una construcción teórica sobre un rasgo psicológico (depresión, motivación, ansiedad). Para que el **constructo** pueda ser abordado estadísticamente, es necesario registrar sus manifestaciones observables a partir de su definición operacional; es decir, se lo operacionaliza (por ejemplo: mediante test) para tener una **variable observable**. De esa manera puede ser medido.

Las variables se clasifican en **cualitativas** o **cuantitativas** (estas son **discretas** o **continuas**). Las variables están asociadas con un nivel de **medición** el cual refiere a un proceso en el cual a las características de estudio se les puede asignar un valor numérico (edad) o un valor simbólico (sexo, nacionalidad).

SEGÚN VALORES	**SEGÚN NIVEL DE MEDICIÓN**	**EJEMPLOS**
CUALITATIVAS: sus valores son atributos o cualidades (valor en símbolos). Son exhaustivos (contempla todas las categorías de la variable) y excluyentes (el individuo puede pertenecer a una categoría sola).	NOMINAL —--- ORDINAL	Profesión, nacionalidad. —--- Grado de satisfacción, nivel socioeconómico
CUANTITATIVAS: sus valores son cantidades numéricas. Pueden ser: **discretas** (no hay valores intermedios) o **continuas** (hay valores intermedios)	INTERVALAR (cero relativo): conserva la distancia entre los valores. —--- RAZÓN (cero absoluto): indica la ausencia de la característica.	Temperatura, puntaje en un test —--- Tiempo, cantidad de palabras memorizadas.

@motorpsico uba

Por la **medición** se obtienen los **datos** con los que trabaja la **estadística**. Estos datos se guardan en una base que se organiza en una matriz, llamada **matriz de datos**, la cual está conformada por columnas y filas tabla donde cada fila representa un individuo que posee la información de interés, y cada columna es el aspecto del individuo que se ha seleccionado para estudiar.

La **estadística** se subdivide en **descriptiva** e **inferencial**. La descriptiva provee una serie de procedimientos que permiten resumir la información para poder interpretarla y brindar información importante sobre un conjunto de datos.

La manera de organizar los **datos** puede ser mediante una **distribución de frecuencia** o en distintos gráficos.

- **Frecuencias absolutas (F):** es la cantidad de veces que el valor está en el conjunto de datos.
- **Frecuencia relativa (Fr):** es la proporción de veces que el valor está en el conjunto.
- **Frecuencia porcentual (F%):** es el porcentaje de veces que ese valor está en el conjunto.
- **Frecuencia acumulada (FA):** es la acumulación de las frecuencias de cada fila (desde el nivel ordinal en adelante).

Ejemplo:

DESTINO	**F**	**Fr**	**F%**
Salta	9	0.36 (9/25=) (25=n)	36 % (F. R. * 100)

Gráficos:

TIPOS DE GRÁFICOS	**TIPOS DE VARIABLE**
Circular o torta	Cualitativa nominal
Barras anchas o rectángulos	Cualitativa nominal —---------- Cualitativa ordinal
Bastones o barras delgadas	Cuantitativa discreta
Tallo y hoja	Cuantitativa discreta
Caja y bigote	Cuantitativa discreta —---------- Cuantitativa continua
Histograma y ojiva de Galton	Cuantitativas continuas

UNIDAD II: resúmenes estadísticos.

La información contenida en una **distribución de frecuencias** se resume en **medidas** (resúmenes estadísticos) que se refieren a diversas características de la distribución:

MEDIDAS DE	**MEDIDAS DE**	**MEDIDAS DE LA**	**MEDIDAS DE**

@motorpsico.uba

POSICIÓN	**TENDENCIA CENTRAL**	**FORMA**	**VARIABILIDAD**
CENTILES/ PERCENTILES	MEDIA	ASIMETRÍA: POSITIVA (DERECHA) O NEGATIVA (IZQUIERDA). SIMÉTRICA.	ENTROPÍA
DECILES	MEDIANA	CURTOSIS: LEPTOCÚRTICA, PLATICÚRTICA, MESOCÚRTICA.	VARIANZA
QUARTILES	MODA		DESVÍO
			COEFICIENTE DE VARIACIÓN

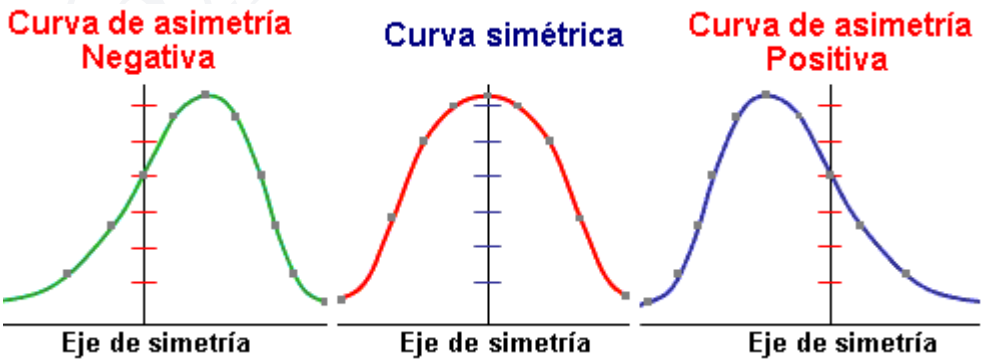
Las **medidas de posición** son medidas en las que se compara una puntuación en particular en relación a un grupo de referencia. Se pueden utilizar en las variables cuantitativas continuas. Por ejemplo: bebé con P80, cuando lo esperable es P50. Supera al 80% y es superado por el 20%. Gráfico: caja y bigote.

Las **medidas de tendencia central** son medidas que nos ayudan a comparar grupos y poder brindar información importante sobre ellos.

- **Media (promedio):** es la suma de los valores observados dividido el total de ellos. Sirven para variables cuantitativas.
- **Mediana:** es el valor que supera y es superado por el 50%. Sirven para variables ordinales y cuantitativas.
- **Moda:** es la variable con mayor frecuencia absoluta. Pueden ser *bimodales* (2 modas) o *amodal* (ninguna se destaca). Sirven para las variables cualitativas.

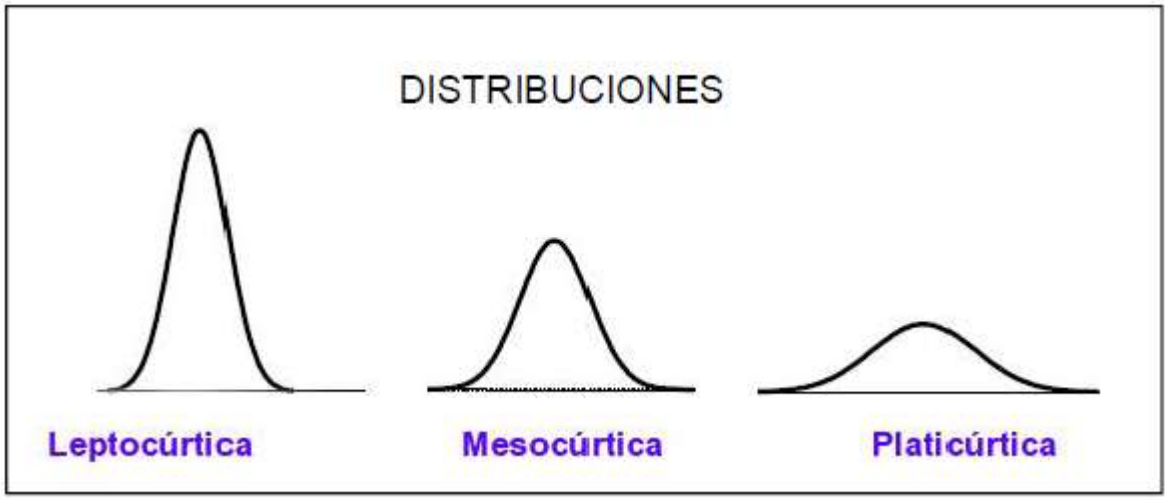
Las **medidas de las formas** pueden ser asimétricas/métricas o por curtosis:

- **Asimetría negativa:** media<mediana<moda.
- **Simétrica:** media=mediana=moda.
- **Asimetría positiva:** media>mediana.



- **Leptocúrtica:** predominan valores positivos.
- **Mesocúrtica:** similar a la curva simétrica.
- **Platicúrtica:** predominan valores negativos.

motorpsico.uba



Las **medidas de variabilidad** sirven para comparar conjuntos y saber si los datos son más homogéneos o heterogéneos.

- **Entropía:** mide el grado de incertidumbre o desorden de una distribución. A > entropía, > incertidumbre Variables cualitativas.
- **Varianza:** permite comparar grupos que tienen medias iguales o similares y se midan con la misma variable. Variables cuantitativas.
- **Desvío:** sirve para armar puntuaciones a partir de la media. Variables cuantitativas.
- **Coeficiente de variación:** permite comparar los grupos cuando las medias son diferentes. Cuando su nivel es bajo, el grupo es más homogéneo (+ representativa) y viceversa. Variables cuantitativas de razón.

Las variables puede ser:

- **Dependientes:** es el objeto de estudio. Ej: color de buzo de egresados.
- **Independientes:** es la que modifica los valores de la variable dependiente. Ej: turnos escolares.

Puntuaciones típicas y escalas derivadas: el concepto de puntuaciones típicas proviene de la conversión de puntajes *x* (directos) a puntajes transformados *(z, t, CI)*. Esto se realiza debido a que los puntajes directos no tienen interpretación. Dándole mayor rendimiento al puntaje más cerca de la media.

- ***Z (media = 0, desvío = 1)*: (**x - media)/desvío.
  Resultados entre -3 y 3.
- **T *(media = 50, desvío = 10)*:** 10*Z + 50.
  Resultados positivos.
- **CI *(media = 100, desvío = 15)*:** 15*Z + 100.
  Resultados positivos.

El **baremos** es una tabla de valores transformados que permiten ubicar a un sujeto en relación a un grupo de referencia.

UNIDAD III: relación entre variables.

Para la relación entre variables cualitativas se utilizan **tablas de contingencia**, a partir de estas se analiza la independencia entre las variables. Cuando son solamente dos variables la tabla se denomina

bivariada. En el caso de variables cuantitativas se utilizan **diagramas de dispersión**. En los mismos se puede ajustar una recta de regresión para presidir el valor de una nota a partir de la otra y calcular la intensidad, el sentido de las mismas.

Tabla de contingencia bivariada:

- X_1: variable vertical: lugares de residencia. Valores: Avellaneda, Lugano.
- X_2: variable horizontal: sexo: femenino/masculino.

X_1: ZONA DE RESIDENCIA	**X_2: SEXO FEMENINO**	**X_2: SEXO MASCULINO**	**DISTRIBUCIONES MARGINALES**
AVELLANEDA	10	25	**35**
LUGANO	5	5	**10**
DISTRIBUCIONES MARGINALES	**15**	**30**	**45 (MUESTRA)**

Los valores se denominan **frecuencia conjunta**, ya que se puede leer las dos variables al mismo tiempo.

La **distribución marginal** es la suma de las **distribuciones condicionadas**.

Q de Kendall: evalúa la intensidad entre variables dicotómicas (dos opciones).

A*D - C*B / A*B + C*B=

- Q = 0: las variables son independientes y no existe relación.
- Q = -1;1: relación perfecta.
- Q = cercano a -1;1: relación moderada.

Riesgo relativo (rr): cociente entre la proporción de casos que se encuentran de una categoría de una variable consecuente bajo determinado escenario y la proporción de casos que se hallan de la misma.

Ej: *"el riesgo relativo de cardiopatía isquémica es de 2,5 veces mayor en pacientes hipertensos"*. Se interpreta que las personas hipertensas tienen dos veces y media mayores chances de sufrir cardiopatía isquémica que quienes no lo son. En este ejemplo las variables son: ser hipertenso y haber sufrido cardiopatía. La primera es antecedente y la segunda consecuente, porque interesa plantear el efecto de la hipertensión sobre la cardiopatía isquémica.

Chi cuadrado (cualitativas): es una medida de distancia a la que se encuentran las frecuencias observadas de las que se esperaría encontrar si las variables fueran independientes.

- No puede ser negativo.
- Puede ser que sea 0 si la frecuencia observada es igual a la frecuencia esperada.
- Puede ser un número grande.
- Solo se puede comparar la intensidad de la asociación entre variables si tienen las tablas de la misma dimensión y el mismo número de casos.

V de Cramer: está basado en el puntaje *chi cuadrado* y tiene un valor máximo de 1. Variables nominales.

Para las variables cuantitativas:

R de Pearson: es un índice que mide la relación entre variables cuantitativas. Mide el sentido y la intensidad de las mismas. No la causalidad.

@motorpsico.uba

- -1 y 1 = relación perfecta.
- Cercano a 1= directa y creciente.
- Cercano a -1: indirecta y decreciente.
- 0 = variables independientes.
- Cercano a 0 = no hay relación entre las variables.

Coeficiente de determinación de R^2: mide la parte de la varianza compartida por ambas variables. Coincide con el R de Pearson.

Diagrama de dispersión:

Recta de regresión: predecir el valor de una variable a partir de la otra.

UNIDAD IV: inferencia estadística y modelos de probabilidad.

La **estadística inferencial** permite extraer, generalizar conclusiones de la población a partir de las muestras. La **población** es un conjunto de unidades de análisis (individuos, entes, hogar, etc.) que comparten una o más características en común en tiempo y espacio. Por ejemplo: *"todos los estudiantes de psicoanálisis de la Facultad de Psicología de la UBA en el 1er cuatrimestre"*.

La **muestra** es una parte de la población, es un subconjunto de unidades de análisis, y ésta debe ser representativa de la población.
Hay dos tipos de muestreo:

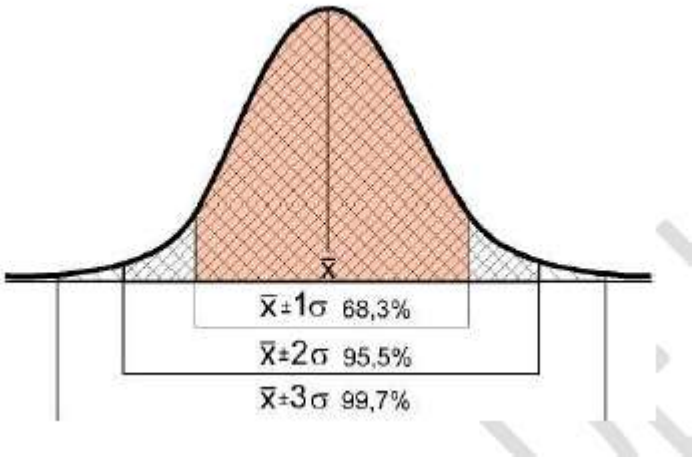
- **Probabilísticos:** la muestra es elegida al azar y todos tienen la misma posibilidad de ser elegidos.
- **No probabilísticos:** no es al azar y por lo tanto, no todos tienen posibilidad de ser elegidos.

A partir de la elección de la técnica de muestreo se proporcionan **estadísticos** (características numéricas y descriptivas muestrales) que permiten hacer inferencias sobre los **parámetros** (características descriptivas numéricas poblacionales) (n = muestra, p = probabilidad).

Dentro de la estadística inferencial se trabaja con dos modelos:

- **Distribución binomial:** es un modelo para variables cuantitativas discretas y está compuesta por dos elementos.
    - **Variable dicotómica:** variable con dos valores.
    - **Modelo Bernoulli:** los dos valores se subdividen en dos.
        - **Éxito (P):** una condición se cumple.
        - **Fracaso (1-p):** la condición no se cumple.
- **Distribución normal:** es un modelo para variables cuantitativas continuas y sus parámetros son la media y el desvío poblacional. Se grafica por medio de la *campana de Gauss*.
    - La mayor parte de los individuos se encuentran en el centro de la distribución. Cuando más se alejan de ella, más serán valores menos probables.
    - Es simétrica.
    - Es asintótica.
    - Los puntos de la media serán +- 1 desvío de la media.
    - Maneja la distribución normal estándar: media = 0, desvío = 1.

@motorpsico.uba



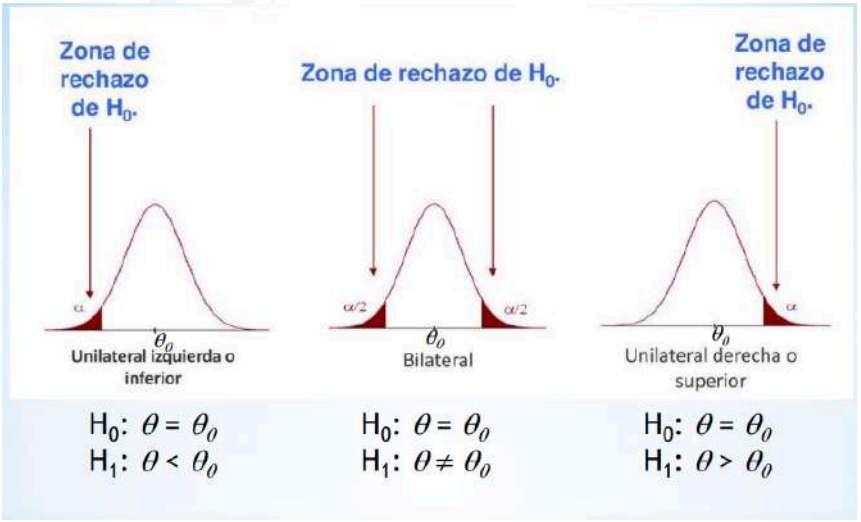
El **teorema central del límite (TCL)** plantea que si la muestra es mayor a 30 se considera una muestra grande y debe utilizarse la distribución normal, en cambio, si la muestra es menor a 30 se toma como aproximadamente normal.

UNIDAD V: prueba de hipótesis.

El **contraste de hipótesis** es un método para tomar decisiones, es un proceso en el cual una hipótesis formulada estadísticamente es puesta a prueba para saber si es compatible o no con datos empíricos. Se contrasta una hipótesis histórica con evidencia muestral.

La **hipótesis estadística** tiene aspectos de probabilidad y se hace preguntas con respecto a lo estudiado, lo pone a prueba. Siempre mantiene un margen de error. Hay dos tipos:

- **H_0 / nula:** es la hipótesis histórica, la que siempre se pone a prueba. Es exacta porque la media poblacional no sufre ningún cambio a lo largo de la distribución.
- **H_1 / alternativa:** propone 3 tipos de contraste.
    - **Contraste unilateral a la derecha:** media >.
    - **Contraste unilateral a la izquierda:** media <.
    - **Contraste bilateral:** =/= (cambiar).



La **lógica del contraste** consta de tres pasos:

- **Proceso de verificación de una hipótesis:** transformar una hipótesis científica en una hipótesis estadística.
- **Buscar evidencia empírica** relevante capaz de informar sobre si la hipótesis estadística es sostenible o no.
- **Regla de decisión:** siempre se establece en términos de probabilidad. Si el resultado muestral observado es muy poco probable se dice que hipótesis es incompatible con los datos y si es muy probable la hipótesis será compatible con los datos. Hay 2 tipos de errores.

	H_O VERDADERA	**H_O FALSA**
SE RECHAZA H_O	Error tipo I	Decisión correcta
SE MANTIENE H_O	Decisión correcta	Error tipo II

Para tomar una decisión se compara el valor de P (probabilidad mínima para rechazar H_0) con el nivel de significación (alfa). Si P es < alfa se rechaza H_0. Caso contrario, no se rechaza.

EXPLICACIÓN DEL MAPA:

La estadística está relacionada con el concepto de variabilidad el cual está anudado al concepto de variable que refiere a una característica que puede tomar diferentes valores o modalidades en los individuos o unidades de análisis que son objeto de estudio. Las variables pueden ser observables o no observables, en este último caso se plantea el concepto de constructo el cual es una construcción teórica de rasgos psicológicos que para poder ser observables se deben operacionalizar. Las variables se pueden clasificar en dos grandes grupos: cualitativas y cuantitativas. Y a cada una le corresponde un nivel o escala de medición que puede ser: nominal, ordinal, intervalar, cociente o razón. A través de la medición se obtienen datos que se pueden organizar en: matriz de datos, gráficos y tablas de distribución de frecuencias (absolutas, relativas, porcentuales).

Esta parte de la materia está abocada a la estadística descriptiva la cual resume, exhibe y organiza datos. A partir de la distribución de frecuencias se pueden calcular resúmenes estadísticos: medidas de posición (percentiles, deciles, cuartiles), medidas de tendencia central (media, mediana y moda) en relación a las medidas de la forma (asimetría y curtosis) y por último las medidas de variabilidad o dispersión (entropía, varianza, Desvío, coeficiente de variación,etc).

En cuanto a los test psicológicos se utilizan puntajes típicos o estandarizados para poder interpretar los resultados y generalizarlos en baremos en grupos normativos.

En la distribución de frecuencias podemos relacionar variables cualitativas en tablas de contingencia, bivariadas, donde también se pueden estudiar la independencia entre ellas. Si son variables cuantitativas en diagramas de dispersión, en este se pueden visualizar la relación entre variables (correlación) y utilizando una recta de regresión poder predecir el valor de una variable a partir de la otra.

En la estadística Inferencial la cual utiliza métodos para resumir, generalizar o inferir datos de la población a partir de las muestras. Tenemos el concepto de población que es totalidad de individuos o unidades de análisis que se desea estudiar y que posean una o más características en común. A raíz de ahí podemos trabajar el concepto de muestra que es una selección de la población, la cual como

característica principal es ser representativa, a su vez podemos realizar diferentes tipos de muestreo de selección para poder estimar valores poblacionales a partir de estadísticos.

La distribución de frecuencias trabaja con dos modelos: distribución binomial y distribución normal, estos permiten controlar el error de muestreo de los métodos de inferencia estadística. Estos métodos pueden ser las pruebas de hipótesis donde la distribución de probabilidades proporciona un nivel de significación (probabilidad de error tipo I) y por otro lado la estimación de intervalos de confianza que proviene de la probabilidad.

@motorpsico.uba