Estadística Descriptiva

lunes, 1 de diciembre de 2025 16:18

Definición: Se refiere a los métodos de recolección, organización, resumen y presentación de un conjunto de datos. Se trata principalmente de describir las características fundamentales de los datos y para ellos se suelen utilizar indicadores, gráficos y tablas.

Objetivo Básico: Describir lo más simplemente posible los resultados obtenidos en un experimento/encuesta/censo, etc.

u = Unidad Elemental: Es cualquier objeto real o ideal sobre el cual pueden hacerse mediciones.

N = Población: Conjunto de unidades elementales que satisfacen una definición común. Debe estar bien definida en tiempo y espacio.

puede ser

- Finita: Puede ser listada
- Infinita: No puede ser listada

n = Muestra:
$n \leq N; n \subseteq N$; Es un subconjunto de la población

Esta debe de ser:

- **Representativa**: Debe representar los aspectos más importantes de la población de la cual se extrae.
- **Aleatoria**: Todos los elementos de la población deben tener la misma probabilidad de ser extraídos de la fuente.

Diseño del Estudio

- Censo: Analiza todas las unidades de la población
- Muestreo: Analiza una parte de una muestra

Factores:

- Exactitud y Precisión
- Costo-Tiempo
- Imposibilidad
- Destrucción

Variable: Cualquier característica susceptible de tomar distintos estados entre unidades elementales, o que varía dentro de una misma unidad elemental a través del tiempo.

Pueden ser:

- **Cualitativas:** Expresan características que el objeto en estudio

tiene o no, o bien lo tiene en distinto grado.
  - Dicotómicas: 2 categorías o clases
  - Policotómicas: >2 categorías o clases

- **Cuantitativas:** Expresan cantidad
  - Discretas: Surgen de contar (solo toman valores discretos dentro de su campo de variación //Enteros). Ej.: Cantidad de focos, de alumnos, etc.
  - Continuas: Surgen de medir (toman cualquier valor dentro de su rango de variación //Decimales). Ej.: Altura de un alumno, tiempo transcurrido, etc.

Escalas de Medición:

Cualitativas:
- **Nominal o Clasificatoria:**
  - Excluyente y exhaustivo
  - Contar cuantos elementos pertenecen a cada categoría
  - Constituye el nivel de medición más bajo
- **Ordinal o Jerárquico:**
  - Es posible ordenar los datos de acuerdo a una jerarquía
  - Relaciones lógicas propias de la escala:
    - Relación de equivalencia (dentro de cada categoría)
    - Relación de orden estricto (entre categorías)

Cuantitativas:
- **Intervalos:**
  - Deben ser números.
  - El punto de origen es llamado cero arbitrario.
  - Admite negativos.
- **Razón o Proporción:**
  - Deben ser números.
  - El punto de origen es realmente cero, es llamado cero real o cero absoluto.
  - No admite negativos.
  - Se puede establecer una distancia o una proporcionalidad entre dos entes cualesquiera.
  - Constituye el nivel de medición más alto.

Aclaración: La presentación de los datos no tiene que ver con la escala de medida

Por ejemplo en el caso del análisis de bombilla:
Población: Las bombillas de una fábrica de Rosario en el año 2014
Muestra: 200 bombillas de la población
Unidad Elemental: Cada bombilla
Variable: Duración en horas de cada bombilla
Clasificación: Cuantitativa Continua
Escala de Medición: Razón o Proporción

Datos en Agrupación Simple

lunes, 1 de diciembre de 2025 16:18

Datos en Agrupación Simple

1 2

Supondremos que X es una variable cualitativa o cuantitativa discreta. Sean $x_1; \ldots; x_k; k \leq n$ los distintos valores que adopta la variable.

Frecuencias Simples:

- f_i: frecuencia absoluta simple de x_i. Es el número de veces que se repite ese valor en las n observaciones.
- $r_i = \frac{f_i}{n}$: frecuencia relativa simple de x_i.
- $p_i = 100 \times r_i$: frecuencia porcentual simple de x_i.

Frecuencias Acumuladas

Solo para 2

Supongamos X es cuantitativa discreta, y $x_1 < x_2 < \cdots < x_k$.

- $F_i = \sum_{j=1}^{i} f_j$: frecuencia absoluta acumulada, suma de las frecuencias absolutas anteriores hasta el dato actual.
- $R_i = \frac{F_i}{n}$: frecuencia relativa acumulada.
- $P_i = 100 \times R_i$: frecuencia porcentual acumulada.

Representación Tabular

Tablas de frecuencias para datos en Agrupación Simple

- Representación tabular de los datos correspondientes a una variable.
- Variable Cualitativa: En 1^{er} columna modalidades que adopta la variable, en las siguientes columnas: frecuencias absolutas, relativas y porcentuales SIMPLES.
- Variable Cuantitativa Discreta: En 1^{er} columna los k distintos valores de la variable ordenados en forma creciente, siguientes columnas: frecuencias absolutas, relativas y porcentuales SIMPLES y ACUMULADAS.

EJEMPLO: Ej. 2

solo para ②

Nro. de Hermanos	Cantidad	r_i	p_i	F_i	R_i	P_i
0	740	0,247	24,7	740	0,247	24,7
1	1097	0,366	36,7	1837	0,613	61,3
2	658	0,219	21,9	2495	0,832	83,2
3	345	0,115	11,5	2840	0,947	94,7
4	126	0,042	4,2	2966	0,989	98,9
5	34	0,011	1,1	3000	1	100
Total	3000	1	100			

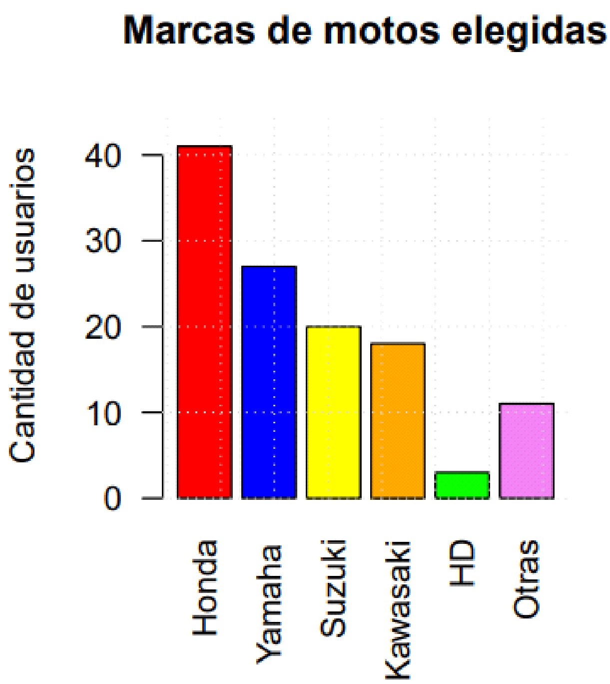
La tabla nos dice, por ejemplo, que:

- 1097 estudiantes tienen 1 sólo hermano.
- Aproximadamente el 4 % de los estudiantes tienen 4 hermanos.
- 2840 estudiantes tienen 3 hermanos o menos.
- Aproximadamente el 5 % de los estudiantes tienen más de 3 hermanos.

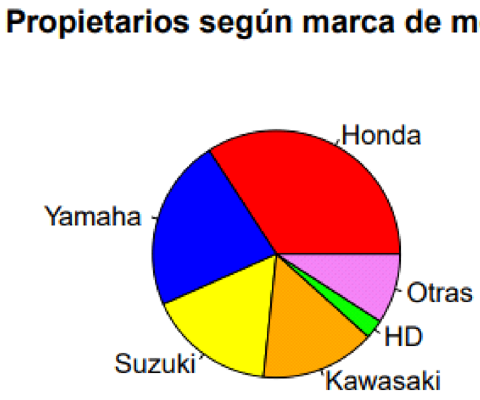
Representación Gráfica

- **Variables Cualitativas Nominales:**
  - Diagrama de barras.

Marca	f_i	f_{r_i}
Honda	41	0.33
Yamaha	27	0.23
Suzuki	20	0.17
Kawasaki	18	0.15
Harley-Davidson	3	0.03
Otra	11	0.09

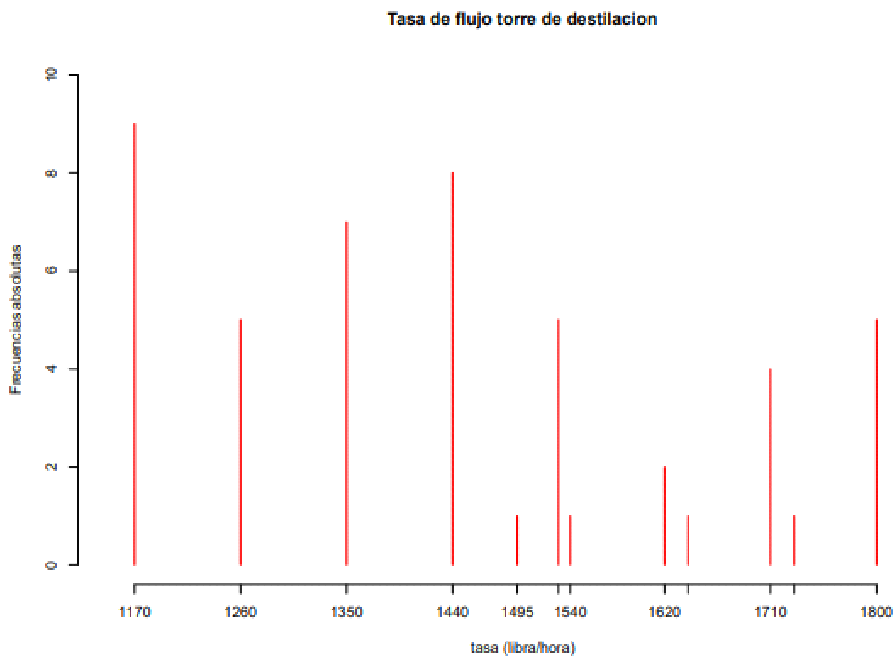


  - Diagrama de sectores.



<u>**Los datos en agrupación simple se representan de manera gráfica de la siguiente forma:**</u>

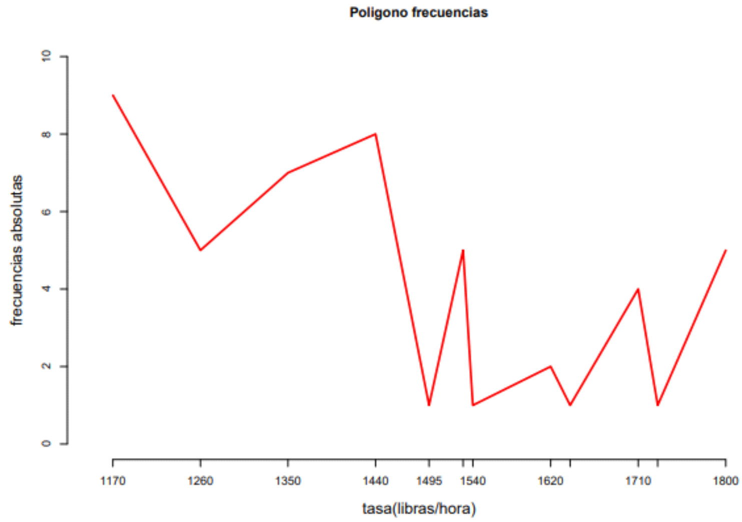
- **Frecuencias simples (fi, ri, pi):**
  - Bastones



  - Polígonos de frecuencia

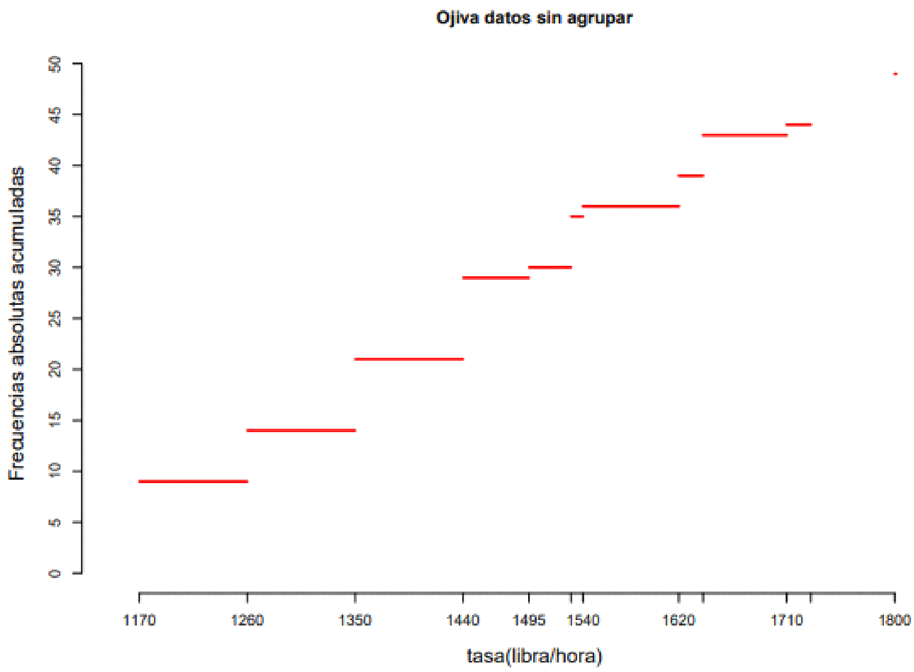
Datos Agrupación Simple

1. Eje de las abscisas: x_i
2. Eje de las ordenadas: frecuencia simple
3. Se unen esos puntos con segmentos de recta



- Frecuencias acumuladas (Fi, Ri, Pi):
  - Función/Gráfico Escalonado (Ojiva)

1. Datos en Agrupación Simple
   - Función escalonada.
   - Eje de las abscisas: x_i.
   - Eje de las ordenadas: Frecuencia acumulada.



Datos Agrupados en Intervalos de Clase

lunes, 1 de diciembre de 2025 18:33

Datos Agrupados en Intervalos de Clase

X variable cuantitativa continua, o cuantitativa discreta pero:

- Demasiados datos;
- Pocos datos, pero muy dispersos (muy diferentes entre sí).
- Interesa una clasificación particular de los resultados.

Se agrupan los valores en intervalos de clase:

- Intervalos deben ser disjuntos: cada observación debe estar contenida en un, y sólo un intervalo de clase.
- Se pierde información sobre las observaciones, pero la variable es más "manejable".

¿Cuántos intervalos?

- Algunos consideran $\sqrt{n}$ como primera aproximación.
- Recomendación: no inferior a 5, no superior a 20.
- Puede definirse previamente de acuerdo al criterio de los investigadores.

Amplitud

- Observaciones ordenadas: $x_1 \leq x_2 \leq \ldots \leq x_n$. Rango de la información: $R = x_n - x_1$.
- I: cantidad de intervalos. $\rightarrow$ Amplitud de los intervalos: $A = R/I$ (redondeando, si es necesario, al entero inmediato superior).
- El primer intervalo de clase debe contener el valor mínimo (x_1) y el último al valor máximo (x_n).

Tener en cuenta:

- NO pueden existir intervalos con frecuencia cero (0). En tal caso, definir intervalos de distinta amplitud.
- La agrupación en intervalos no debe modificar la forma de la distribución original de los datos.

Sean $L_{i\,inf}$ y $L_{i\,sup}$ los límites inferior y superior del intervalo $i-$ésimo, $i = 1; \ldots; k$, k : número de intervalos.

- Los intervalos, que deben ser disjuntos, pueden ser del tipo:
  - $L_{i\,inf} \leq x < L_{i\,sup}$ (incluyen al límite inferior),
  - $L_{i\,inf} < x \leq L_{i\,sup}$ (incluyen al límite superior).
- Se define la Marca de clase del intervalo i a

$$M_{c\,i} = \frac{L_{i\,inf} + L_{i\,sup}}{2}$$

- $M_{c\,i}$ es el punto medio del intervalo i y es el "representante"de dicho intervalo.

Representación Tabular

- 1^{er} columna: Intervalos de clase.
- 2^{da} columna: Marca de clase.
- El resto de las columnas se construye de igual manera que la tabla de frecuencias para datos en agrupación simple:
  - f_i : frecuencia absoluta simple del intervalo i , cantidad de observaciones que caen en dicho intervalo.
  - $r_i = \frac{f_i}{n}$: frecuencia relativa simple del intervalo i.
  - $p_i = 100 \times r_i$: frecuencia porcentual simple del intervalo i.
  - $F_i = \sum_{j=1}^{i} f_j$: frecuencia absoluta acumulada, suma de las frecuencias absolutas anteriores hasta el intervalo i.
  - $R_i = \frac{F_i}{n}$: frecuencia relativa acumulada.
  - $P_i = 100 \times R_i$: frecuencia porcentual acumulada.

Ejemplo 5, agrupando los datos en intervalos de amplitud 100, comenzando en 1100

Tasa	M_{C_i}	f_i	f_{r_i}	p_i	F_i	F_{r_i}	P_i
[1100,1200)	1150	9	0.19	19	9	0.19	19
[1200,1300)	1250	5	0.10	10	14	0.29	29
[1300,1400)	1350	7	0.14	14	21	0.43	43
[1400,1500)	1450	9	0.19	19	30	0.62	62
[1500,1600)	1550	6	0.12	12	36	0.74	74
[1600,1700)	1650	3	0.06	6	39	0.80	80
[1700,1800)	1750	5	0.10	10	44	0.90	90
[1800,1900)	1850	5	0.10	10	49	1.00	100

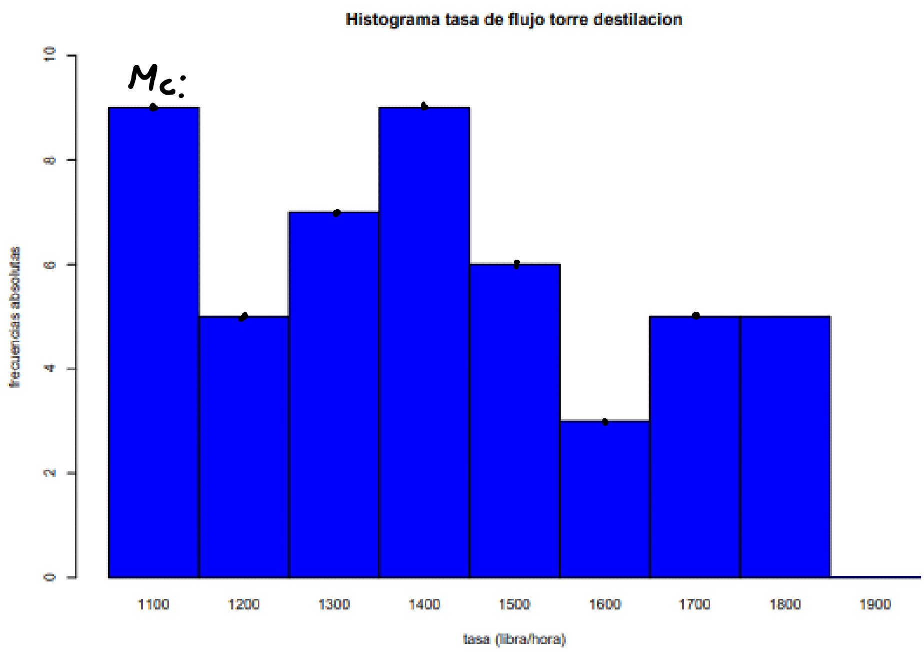
La tabla nos dice, por ejemplo, que:

- 7 observaciones midieron entre 1300 y 1400 libras/hora
- 36 observaciones midieron una tasa menor a 1600 libras/hora
- Un 20 % de las observaciones midió 1700 libras/hora o más.

Representación Gráfica

Los datos agrupados en intervalos de clase se representan de manera gráfica de la siguiente forma:

- Frecuencias simples (fi, ri, pi)
  - Histogramas

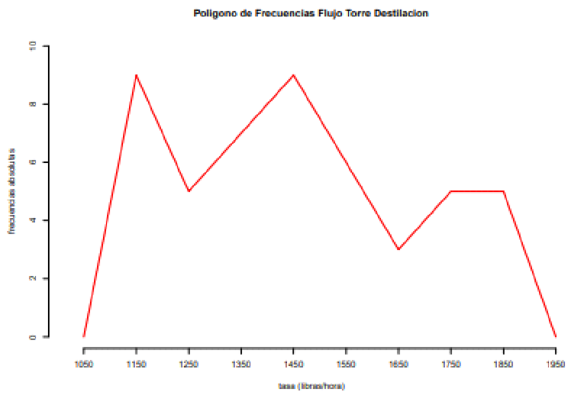


  - Polígonos de Frecuencias

Polígonos de Frecuencias-Datos Agrupados en Intervalos

Datos Agrupados en Intervalos de Clase

1. Eje de las abscisas: marcas de clase de los intervalos.
2. Eje de las ordenadas: frecuencia simple
3. Se unen esos puntos con segmentos de recta

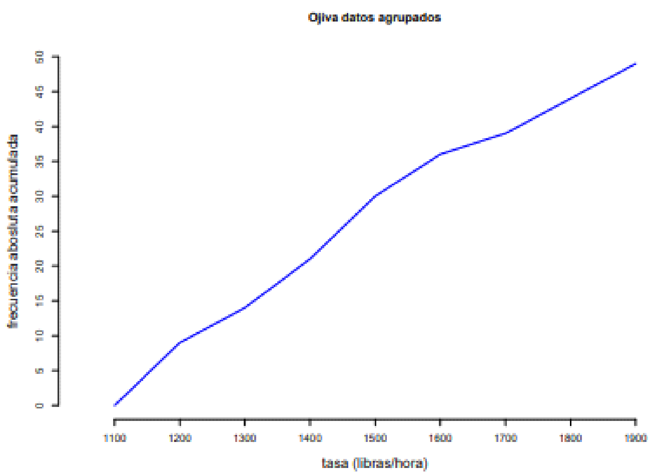


- Frecuencias Acumuladas (Fi, Ri, Pi)
    - Ojiva

Datos Agrupados en Intervalos

- Eje de las abscisas: límites de los intervalos.
- Eje de las ordenadas: frecuencias acumuladas sobre el límite superior del intervalo.
- Al límite inferior del primer intervalo se le asigna el valor 0.
- Se unen esos puntos con segmentos de recta.

Datos Agrupados en Intervalos de clase



Medidas Descriptivas Numéricas

lunes, 1 de diciembre de 2025 18:53

El **objetivo** de las medidas descriptivas es **caracterizar una distribución de frecuencias por medio de un número reducido de medidas numéricas, las cuales complementan la información aportada por tablas de frecuencia y gráficos.**

Estas medidas **brindan en forma resumida información del conjunto de datos y una idea del comportamiento global de la población o muestra en estudio.**

Las Medidas Descriptivas Numéricas se clasifican en:

- **Medidas de Tendencia Central (MTC)**: Valores numéricos que se obtienen de variables cuantitativas y cuyos resultados se localizan por el centro de la distribución.
  - **Media (**promedio aritmético - $\bar{X}$**),**
  - **Mediana (Med),**
  - **Moda (Mo).**

- **Medidas de Posición:** Valores numéricos que permiten dividir la distribución de datos en partes iguales.
  - **Cuartiles,**
  - **Deciles,**
  - **Percentiles.**

- **Medidas de Dispersión o Variación:** Valores numéricos que proporcionan una idea sobre cuan esparcidos o concentrados están los datos correspondientes a una variable.
  - **Rango (R),**
  - **Rango Intercuartílico (RIC),**
  - **Varianza (S^2),**
  - **Desvío o Desviación Estándar (S),**
  - **Coeficiente de Variación (CV).**
- **Medidas de Forma:** Dan una idea de la forma de la distribución de la variable
  - **Coeficiente de Asimetria de Pearson (a_s)**
  - **Coeficiente de Kurtosis (K)**

Medidas de Tendencia Central

martes, 2 de diciembre de 2025 18:44

MTC

Media (promedio aritmético - $\overline{X}$)

La media aritmética es el momento central de primer orden.
En su cálculo se emplea toda la información disponible.
Se expresa en la misma unidad que la variable en estudio.
Es el centro de gravedad de toda la distribución, representando a todos los valores observados.
Es muy sensible a los valores extremos de la variable.

Datos sin agrupar

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i}{n}$$

Datos en agrupación simple

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n}$$

Datos Agrupados en Intervalos

$$\bar{x} = \frac{\sum_{i=1}^{k} M_{ci} f_i}{n}$$

Mediana (Med)

La **Mediana** es aquel valor de la variable que divide al conjunto de valores observados en dos partes, de modo que el 50% de los valores son $\geq$ que la Mediana y el otro 50% son $\leq$ que ella.
Ocupa el lugar central del conjunto de datos (ordenados en forma creciente), dejando a su izquierda y derecha la misma cantidad de observaciones.

Es el 2do cuartil, el 5to decil y el 50vo percentil.

No es afectada por los valores extremos. Por lo tanto su uso es adecuado en distribuciones asimétricas.
En su cálculo no interviene toda la información disponible.
Hay que ordenar los datos antes de determinarla.
Para poder calcularla, el nivel de medición debe ser al menos jerárquica

DAS:

- n impar: Med = dato que ocupa la posición $\frac{n+1}{2}$, esto es, si los datos son $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$, $Med = x_{\left(\frac{n+1}{2}\right)}$.
- n par: $Med = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}$

DAIC

- Calcular $n/2$.
- Buscar en la tabla de frecuencias absolutas acumuladas el intervalo de clase que contenga a la frecuencia $n/2$. Llamaremos a dicho intervalo "intervalo Mediana", y denotaremos por $L_{Med\,inf}$ al límite inferior de ese intervalo, por f_{Med} a la frecuencia absoluta simple del mismo y por F_{Med-1} a la frecuencia absoluta acumulada de la clase inmediata anterior. Sea A_{Med} la amplitud del intervalo mediana.
- $Med = L_{Med\,inf} + \frac{n/2 - F_{Med-1}}{f_{Med}} A_{Med}$.

Moda (Mod)

No siempre existe

Medidas de Tendencia Central- Moda

Datos en agrupación simple: Valor que aparece más frecuentemente que cualquier otro. Puede haber más de una moda (distribución bimodal, trimodal, multimodal).

Datos agrupados en intervalos:

1. Determinar la clase modal (la de mayor frecuencia absoluta).
2. Determinar valor de la moda dentro de la clase modal:

$$Mo = L_{Mod\,inf} + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) A_{Mod}$$

donde $L_{Mod\,inf}$ es el límite inferior de la clase modal,
$\Delta_1 = f_{Mod} - f_{preMod}$ y $\Delta_2 = f_{Mod} - f_{postMod}$.
f_{Mod}; f_{preMod} y $f_{postMod}$ son las frecuencias absolutas simples de las clases Modal, pre Modal y post Modal, respectivamente.
A_{Mod} es la amplitud del intervalo modal.

Medidas de Posición

Cuantiles: Son valores del conjunto de observaciones que permiten dividirlo en partes iguales.

Cuantiles más usados:

Cuartiles (Q)
Dividen el conjunto en cuatro (4) partes iguales, con un cuarto (25 %) de la información. Se denotan Q1, Q2, Q3, Q4.

Deciles (D)
Dividen el conjunto de observaciones en diez (10) partes iguales, son 10 Deciles denotados como: D1; . . . ; D10.

Percentiles (P)
Dividen el conjunto de observaciones en cien (100) partes iguales cada una de las cuales contiene un 1 % de las observaciones, denotados por:

P1; P2; . . . ; Pk ; . . . ; P100

Donde k denota el porcentaje de observaciones que quedan a la izquierda del percentil Pk

Cálculo de Cuantiles

Datos en agrupación simple:

- Se ordenan los datos de menor a mayor.
- Se calcula la posición del cuantil $I_k = \frac{k \cdot n}{NoC}$, luego se busca el valor correspondiente del cuantil con la ayuda de la tabla de frecuencias acumuladas.
- C_k: =Cuantil buscado; $1 \leq k \leq NoC$, $n =$total de observaciones, $NoC = 4$ Para Cuartiles; 10 para Deciles y 100 para Percentiles.

Datos agrupados en intervalos:

- Se identifica el intervalo que contenga el cuantil buscado con la fórmula $I_{Ck} = \frac{k \cdot n}{NoC}$.
- Se calcula:

$$C_k = L_{inf} + \frac{\frac{k \cdot n}{NoC} - \sum f_{ant}}{f_{I_{Ck}}} A_{I_{Ck}}$$

Medidas de Dispersión

martes, 2 de diciembre de 2025 18:46

Medidas de Dispersión o Variabilidad

Las medidas de tendencia central describen el valor representativo de un conjunto de datos, pero por sí solas no informan sobre cómo se distribuyen los valores. Para eso se utilizan las medidas de dispersión o variabilidad. Cuando la dispersión es baja, los datos están agrupados cerca de la MTC y esta resulta representativa; en cambio, una dispersión alta indica que la MTC no es confiable. Por lo tanto, toda medida de tendencia central debe ir acompañada de una medida de variabilidad para que la información sea adecuada.

(más corta)

Las medidas de tendencia central indican el valor representativo de un conjunto de datos, pero no informan sobre su dispersión. Por ello deben complementarse con medidas de variabilidad: si la dispersión es baja, la MTC es representativa; si es alta, la MTC no resulta confiable.

Estas medidas se dividen en 2 grupos

- Absolutas:
  - Rango o Amplitud Máxima (R)
  - Rango Intercuartílico (RIC)
  - Varianza (S^2)
  - Desviación Estándar (S)
- Relativas:
  - Coeficiente de Variación (CV)

Rango

Es la diferencia entre el máximo y el mínimo valor del conjunto de datos

$$R = x_{(n)} - x_{(1)}$$

No se basa en ninguna MTC

Es afectado por valores OUTLIERS

Es afectado por el tamaño de la muestra

Se desaprovecha mucha información ya que solo se usan 2 valores

Rango Intercuartílico (RIC)

Indica la variación máxima que sufre el 50% central de los valores de la variable. Este desvío deja mucho a cada lado (el 25% de la información). La mediana parte a la distribución en dos partes iguales, pero a veces es más significativo el 50% entre Q3 y Q1; porque es un 50 % más puro, más homogéneo.

$$RIC = Q_3 - Q_1$$

El intervalo $[MTC - RIC/2; MTC + RIC/2]$ concentra, aproximadamente, el 50 % de los datos centrales.
Problema: Sigue sin estar basado en una MTC.

Varianza

Es el promedio de los cuadrados de las desviaciones de los valores muestrales respecto de la media aritmética $\bar{X}$. Se representa por S^2

representa por S^2. D.S.A: $S^2 = \dfrac{\sum_i^n (x_i - \bar{X})^2}{n-1}$

D.A.S.: $S^2 = \dfrac{\sum_i^k (x_i - \bar{X})^2 f_i}{n-1}$

D.A.I.C.: $S^2 = \dfrac{\sum_i^k (M_{Ci} - \bar{X})^2 f_i}{n-1}$

Fórmula de Trabajo: $S^2 = \bar{X^2} - (\bar{X})^2$.

El hecho de dividir por n — 1 en lugar de n es apenas apreciable cuando n es grande.
Es el momento central de segundo orden.
En su cálculo intervienen todos los datos observados.
Es una medida de variabilidad promedio respecto de una MTC.
Se pierde la unidad de medida original (queda afectada por el cuadrado).

PROPIEDADES DE LA VARIANZA TECSUP

1) V(X)>0 (la varianza siempre es positiva)
2) Si c es una constante entonces: V(c)=0
3) Si Y=X+c entonces: V(Y) = V(X+c) = V(X)
4) Si Y=bX entonces: V(Y) = V(bX) = b^2 V(X)
5) Si Y=bX+c entonces: V(Y) = V(bX+c) = b^2V(X)

$$V[X] = E[X^2] - (E[X])^2 \qquad \boxed{E(X)=M(X)}$$

EJEMPLO: Si V(X)= 2 , C = 5 y b = 3

- Si Y = X+5 entonces V(Y) = V(X+5) = V(X) = 2
- Si Y = 3X entonces V(Y) = V(3X) = 3^2 V(X) = 9x2 = 18
- si Y = 3X+5 entonces V(Y) = V(3X+5) = 3^2 V(X) = 9x2 =18

$\bar{X}$

$\underline{800m} \pm 10m$

Desviación Estándar (S)

Es la raíz cuadrada de la Varianza, y se representa por S.
Expresa la dispersión de la distribución y se expresa en las mismas unidades de medida de la variable. La Desviación Estándar es la medida de dispersión más utilizada en Estadística.

$790 \leq x \leq 810$

Algunas **características de la Varianza y el Desvío Estándar**:

- Son índices que describen la variabilidad o dispersión y por tanto cuando los datos están muy alejados de la media, el numerador de sus formulas será grande y la Varianza y la Desviación Estándar también lo serán.
- Al aumentar el tamaño de la muestra, disminuye la Varianza y el Desvío Estándar.
- Cuando todos los datos de la distribución son iguales, la Varianza y el Desvío Estándar son iguales a cero.
- Para su cálculo se utilizan todos los datos de la distribución; por tanto, cualquier cambio de valor será detectado.

Distribución Normal - Campana de Gauss

Para distribuciones de datos que se aproxima a la distribución normal podemos también obtener fracciones de datos que caen dentro de ciertos límites. La más usada es la regla (68 - 95 - 99).

- Aproximadamente, el 68,27 % de los casos están entre $\bar{X} - S$ y $\bar{X}S$.
- Aproximadamente, el 95,45 % de los casos están entre $\bar{X} - 2S$ y $\bar{X} + 2S$.
- Aproximadamente, el 99,73 % de los casos están entre $\bar{X} - 3S$ y $\bar{X} + 3S$.

Coeficiente de Variación (CV)

El Coeficiente de Variación es una medida de dispersión relativa que se expresa generalmente en porcentajes.
Si dos conjuntos van a ser comparados, los valores absolutos son convenientes para este fin, únicamente si los promedios de dichos conjuntos son más o menos iguales y si se refieren a un mismo fenómeno.

$$CV = \frac{S}{\bar{X}} \cdot 100$$

Medidas de Forma

martes, 2 de diciembre de 2025 18:47

Medidas de Forma

Coeficiente de asimetría de Pearson

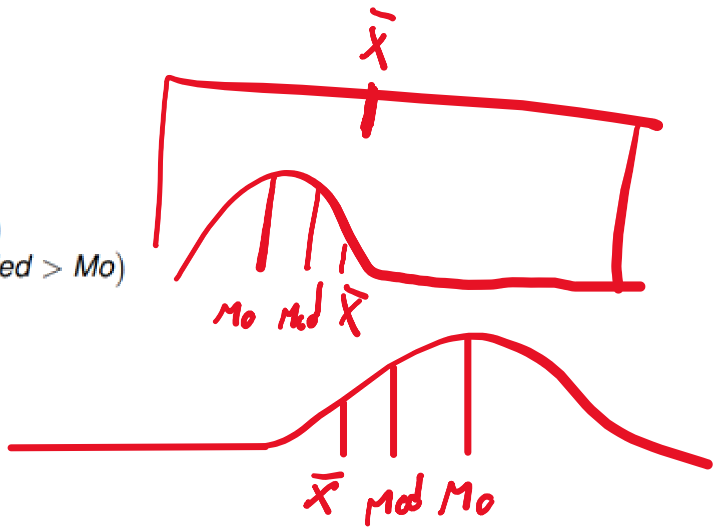
Grado de simetría de una distribución respecto a su media

Una distribución puede ser:

- **Simétrica**: los valores equidistantes de una posición central tienen la misma frecuencia
- **Asimétrica positiva**: las frecuencias más altas corresponden a valores que se encuentran al lado izquierdo de esa posición central (cola a la derecha)
- **Asimétrica negativa**: distribuciones con cola a la izquierda.

$$a_s = \frac{3\left(\bar{X} - Med\right)}{S}$$

- $a_s = 0 \Rightarrow$ distribución simétrica $\left(\bar{X} = Med = Mo\right)$
- $a_s > 0 \Rightarrow$ distribución asimétrica positiva $\left(\bar{X} > Med > Mo\right)$
- $a_s < 0 \Rightarrow$ distribución asimétrica negativa $\left(\bar{X} < Med < Mo\right)$



Curtosis

Se aplica a distribuciones unimodales simétricas o apenas asimétricas, ya que permite ver si la distribución está más alta o más achatada en comparación con la distribución normal

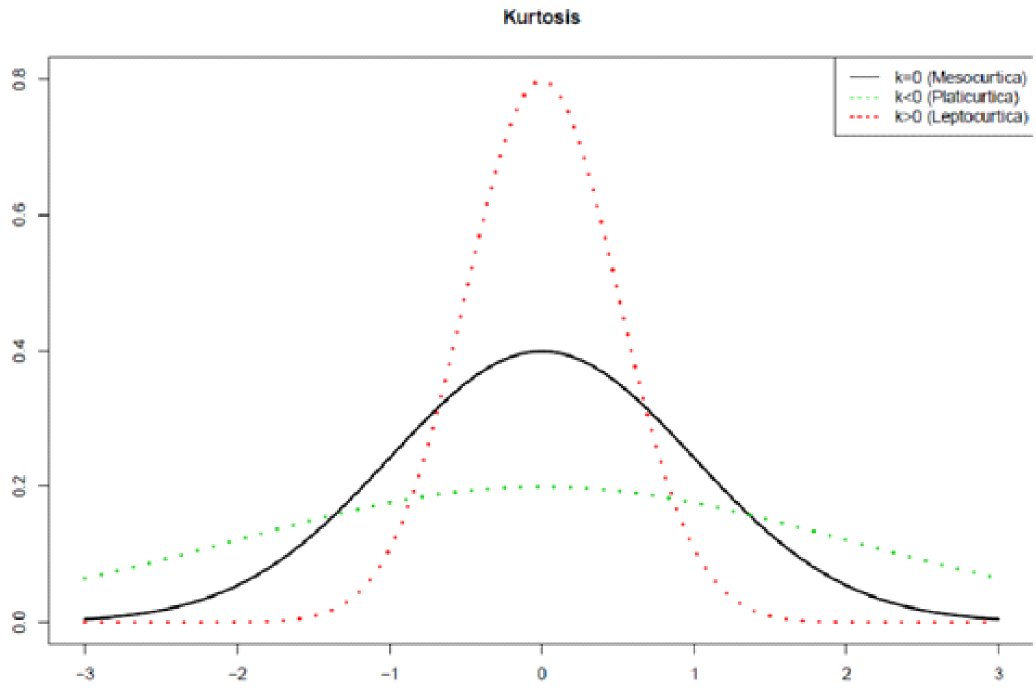
Se calcula como:

$$\kappa = \frac{m_4}{S^4} - 3$$

Siendo k el Coeficiente de Curtosis

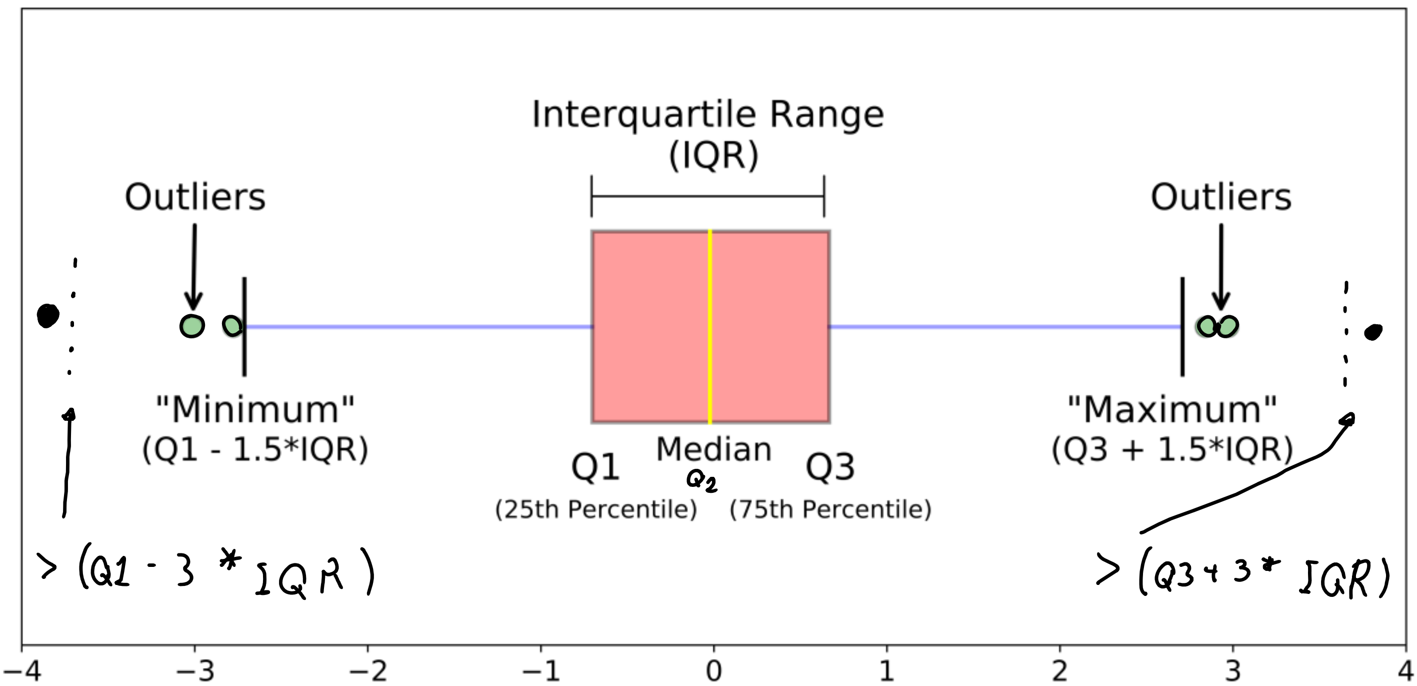
siendo $m_4 = \dfrac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^4}{n}$ el momento central de orden 4.

- $\kappa = 0 \Rightarrow$ mismo grado de elevación que distribución normal (Mesocúrtica).
- $\kappa > 0 \Rightarrow$ más apuntamiento que distribución normal (Leptocúrtica)
- $\kappa < 0 \Rightarrow$ menor grado de elevación que distribución normal (Platicúrtica)


Kurtosis
k=0 (Mesocurtica)
k<0 (Platicurtica)
k>0 (Leptocurtica)
0.0
0.2
0.4
0.6
0.8
-3
-2
-1
0
1
2
3

Boxplot o Diagrama de Cajas y Bigotes

martes, 2 de diciembre de 2025 13:14



Preguntas Final

jueves, 27 de noviembre de 2025 11:26

Final 16/10 Pregunta 1

En una empresa se registran datos sobre la duración (en minutos) de las llamadas de atención al cliente realizadas por los empleados. El análisis de estos tiempos permite evaluar la eficiencia del servicio y detectar posibles irregularidades en la atención.

1.5p

- (a) Suponga que, al analizar los tiempos de atención, se obtiene: Empresa A: Media = 8.5 min, CV = 20% y Empresa B: Media = 9.0 min, CV = 35%. Explique cual empresa muestra una mayor homogeneidad en los tiempos de atención.
- (b) ¿Cómo ayudarían las medidas de variabilidad a complementar el análisis de los tiempos de atención de las llamadas?

Datos

Empresa A

$\bar{X}_A = 8,5$ min

$CV_A = 20\%$

Empresa B

$\bar{X}_B = 9$ min

$CV_B = 35\%$

$CV_A < CV_B$

a) La Empresa A (CV = 20%) al tener un coeficiente de variación menor al de la Empresa B (CV = 35%) muestra una mayor homogeneidad en los tiempos de atención. Esto debido a que el CV es una medida de dispersión relativa.

b) Las medidas de variabilidad ayudan a complementar el análisis de los tiempos de atención de las llamadas puesto que estas nos muestran la dispersión de los valores observados.
Si el valor de estas medidas de dispersión o variabilidad es pequeño, los datos estarán estrechamente agrupados alrededor de la MTC (en este caso la media de los tiempos de atención), entonces dicha MTC se considera representativa. Caso contrario la MTC no es confiable.
En este caso el CV nos ayudaría a encontrar irregularidades en la duración de las llamadas, puesto que si este fuera muy alto, con una media de unos minutos, nos indicaría que hay llamadas que duran segundos y otras que duran horas.

Final 11/09 Pregunta 1

En el análisis de tráfico de red y ciberseguridad, se recopilan estadísticas sobre la cantidad de paquetes enviados por los usuarios, los tiempos entre conexión o los tamaños de los mensajes transmitidos. Estos datos permiten identificar patrones normales y detectar anomalías. 2p

a) Explique brevemente las diferentes medidas estadísticas, indicando qué información aportan en este contexto.
b) Imagine que se registraron los tiempos de conexión (en ms) de un conjunto de usuarios ¿Porque sería importante analizar no solo la media sino también la dispersión y la asimetría de los tiempos para avalar la estadística de la red?
c) Suponga que al analizar los tiempos de conexión de los usuarios se obtuvieron Media = 250 ms y coeficiente de asimetría = 2.5. ¿Qué puede decir en este caso? Justifique su respuesta.
d) ¿Cómo ayudaría las medidas de variabilidad a completar el análisis?

a) Las medidas estadísticas son las siguientes:
   a. Medidas de Tendencia Central: Nos dicen cual es el comportamiento "normal" de la red
      i. La media no es dice el tamaño promedio de un paquete o el ping promedio
      ii. La moda nos dice cual es el tamaño de paquete o el tiempo de conexión que más se repite
      iii. La mediana nos muestra que medida ocupa el lugar central de nuestros datos
   b. Medidas de Posición: Estas sirven para dividir los datos en partes iguales
      i. En este caso por ejemplo podríamos ver el percentil 99 para ver donde cae el 99% de los datos de tiempos entre conexión y así ver la calidad del servicio prestado
   c. Medidas de Dispersión o Variabilidad: Nos dicen que tan homogéneos o estables son los datos
      i. Un desvío estándar o un CV bajo significaría una conexión estable. Pero un desvío estándar o un CV grande indicaría que a veces el tiempo entre conexiones es rápido y otras se cae.
   d. Medidas de Forma
      i. Nos sirven para entender de forma más clara el como se distribuirían los datos, si es una distribución normal, asimétrica o si la distribución de los datos es achatada, normal o elevada. Nos permitiría visualizar gráficamente donde caen o el como están ordenados los datos. En este contexto nos sirve para visualizar mejor la variación de los datos o el como se concentran, ya sea el tamaño de los paquetes, el tiempo de conexión, o la cantidad de paquetes enviados por los usuarios.

b) Es necesario analizar la dispersión y la asimetría de los tiempos de conexión para avalar la estadística ya que la media es muy susceptible a los valores extremos recolectados.

   La dispersión indica la confiabilidad de la media. Una red con media de 200ms y desvío de 5ms es mejor que una con media de 150ms y desvío de 100ms. Una media baja no sirve de nada si la desviación estándar es alta, ya que implica un servicio impredecible.

   La importancia de la Asimetría: Lo ideal en una red es tener muchos usuarios con una buena conexión y muy pocos o ninguno con una conexión lenta o con problemas, esto se podría ver mediante la asimetría, donde en una asimetría positiva (con cola a la derecha) se podría validar como la se cumple esta característica, puesto que caerían la mayoría de los datos en un tiempo de conexión menor, pero habrían algunos con un tiempo de conexión muy grande (conexión más lenta o problemas de red) que estarían desplazando la media con estos valores extremos.
   Si la asimetría fuera negativa, indicaría que la mayoría de los usuarios tiene problemas, lo cual sería crítico.

c) Media = 250 ms; Coeficiente de Asimetria: 2.5
   Un coeficiente de asimetría de 2.5 es positivo y bastante alto (mayor a 0).
   Esto indica una Asimetría Positiva, es decir que la cola de la distribución se alarga hacia la derecha (hacia los valores altos).

   Lo cual nos deja los datos de la siguiente manera:
   $\bar{X} > Med > Mo$

   En este contexto se podría interpretar que la gran mayoría de los usuarios tiene tiempos de conexión bajos (menores a la media de 250ms). Probablemente, la mediana sea mucho menor a 250ms (quizás 50ms o 100ms).

   Sin embargo, hay un pequeño grupo de usuarios (la cola derecha) que tiene tiempos de conexión extremadamente altos (quizás 2000ms, 5000ms, etc).

   Conclusión: Esos pocos usuarios con lag extremo están "arrastrando" el promedio hacia arriba hasta 250ms. La red funciona bien para la mayoría, pero hay casos puntuales graves que investigar.

d) Las medidas de variabilidad ayudan a complementar el análisis de las estadísticas recolectadas puesto que estas nos muestran la dispersión de los valores observados.

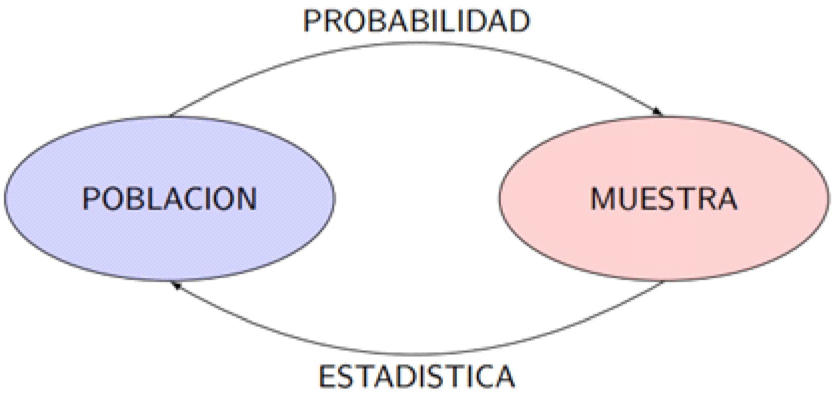
Si el valor de estas medidas de dispersión o variabilidad es pequeño, los datos estarán estrechamente agrupados alrededor de la MTC, entonces dicha MTC se considera representativa. Caso contrario la MTC no es confiable.

En este caso por ejemplo, si calculáramos la desviación estándar, se podría establecer un tiempo entre conexiones normal, por ejemplo: Media $\pm$ 2 Desviación Estándar. Y cualquier tiempo de conexión fuera de este rango se tomaría como una anomalía.
Y el Coeficiente de Variación nos permitiría comparar la estabilidad entre diferentes servidores, independientemente de si uno es rápido y el otro lento. Nos diría cual es más "predecible".

Introducción

miércoles, 3 de diciembre de 2025 09:57

Introducción



Un **experimento** es cualquier acción o procedimiento que, generalmente y bajo ciertas condiciones y reglas rigurosamente controladas, genere observaciones.

- **Experimento Determinístico**: Si se repite bajo las mismas condiciones, el resultado es siempre idéntico (ej. soltar un objeto y que caiga).
- **Experimento Aleatorio**: Aunque se repita en condiciones idénticas, los resultados varían. Tiene una componente de azar (ej. lanzar un dado, medir voltaje).

Modelo Probabilístico: Es una simplificación matemática de la realidad, que involucra operaciones y funciones matemáticas, **variables y parámetros.**

El Espacio Muestral (S)

Es el conjunto de todos los resultados posibles de un experimento aleatorio.

Se dice que el espacio muestral es:

- **Discreto**: Si consiste de un número finito, o infinitamente numerable de resultados posibles. Cada uno de estos posibles resultados se denomina punto muestral (ej. {1, 2, 3...} o {Cara, Cruz}).
- **Continuo**: Si contiene al menos un intervalo (acotado o no) de números reales (ej. voltaje entre 0V y 13V, temperatura).

Eventos o Sucesos

Un **Evento aleatorio** (o suceso aleatorio) es cualquier subconjunto del espacio muestral S. Se denotan por letras mayúsculas (A, E, etc.)

Evento imposible: no ocurre nunca. Se lo denota por $\emptyset$

Evento seguro: ocurre siempre, coincide con el espacio muestral.

Si el espacio muestral es **discreto**:

Un evento es **simple** (elemental) si consiste solamente de un punto muestral. Lo denotaremos E_i

Un evento es **compuesto** si consiste de más de un punto muestral.

Si los eventos A y B pueden ocurrir simultanemante, decimos que son eventos compatibles.

Si los eventos A y B no pueden presentarse simultanemante, decimos que son eventos incompatibles, excluyentes o disjuntos.

Decimos que dos eventos son opuestos si:

- no pueden presentarse simultaneamente
- la NO ocurrencia de uno implica la ocurrencia del otro.

El **opuesto** o **complemento** de un evento A, denotado por A' es el conjunto formado por todos los elementos de **S** que no están en A.

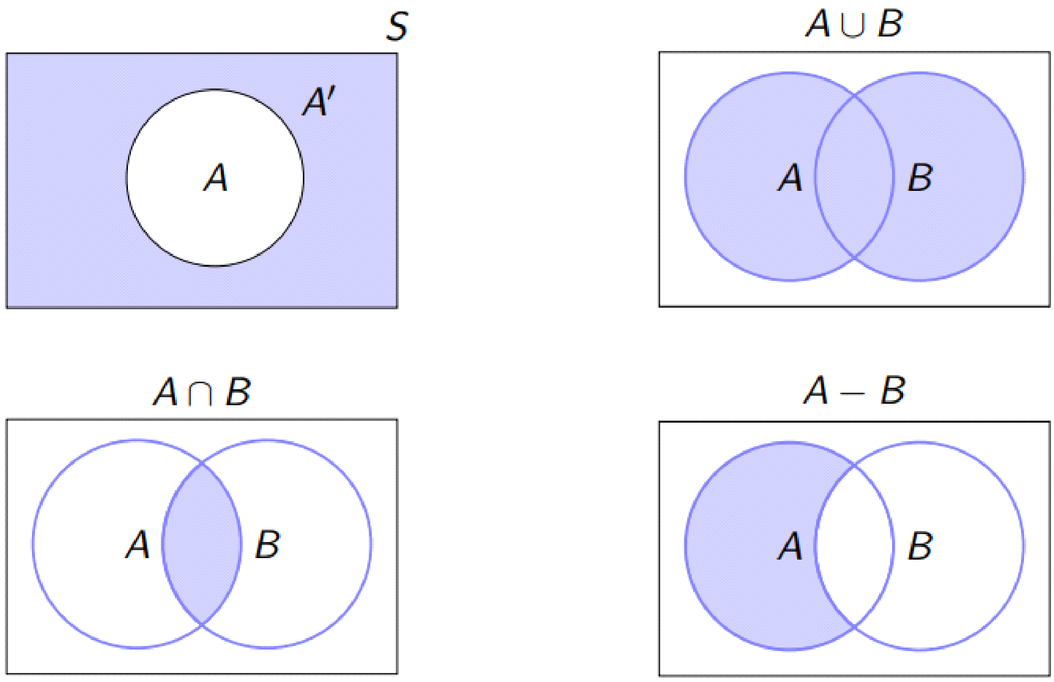
- **Mutuamente Excluyentes (Incompatibles):** No pueden ocurrir al mismo tiempo ($A \cap B = \emptyset$).
- **Opuestos (Complementarios A'):** La no ocurrencia de uno implica la ocurrencia del otro.

Operaciones entre conjuntos

El evento $A \cup B$ (**suma** o **unión**) es el evento que está formado por todos los resultados que están en A ó en B.

El evento $A \cap B$ (**producto** o **intersección**) es el evento que está formado por todos los resultados que están contenidos en ambos eventos.

El evento $A - B$ (**diferencia**) entre A y B (en ese orden) es el evento formado por todos los resultados presentes en A que no estén presentes en B



Clases de Conjuntos y $\mathcal{A}$

Un conjunto $\mathcal{A}$ cuyos elementos son conjuntos, es una clase o familia de conjuntos.

Ejemplo: Supongamos $S = \{a, b, c\}$.

$P(S) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, S\}$

es el conjunto de partes de S y es una clase, o familia, de conjuntos.

Definición 2.6

Una clase $\mathcal{A} \neq \emptyset$ formada por subconjuntos de un conjunto $\mathbf{S}$ que es *cerrada* para la *unión* y *complemento* de conjuntos (si $B \in \mathcal{A}, C \in \mathcal{A} \Rightarrow B \cup C \in \mathcal{A}, B' \in \mathcal{A}$) se denomina Algebra Booleana de Conjuntos (álgebra de subconjuntos de $\mathbf{S}$).
Si además se satisface que $A_i \in \mathcal{A}, \forall i = 1, 2, 3, \ldots \Rightarrow \bigcup_{i=1}^{\infty} A_i$ entonces $\mathcal{A}$ es una σ-álgebra de conjuntos

Espacio de Probabilidad

La descripción de los experimentos aleatorios se representa por la terna $(S, P(S), P)$ donde:

S: espacio muestral asociado al experimento

$P(S)$: algebra (o $\sigma - algebra$) de conjuntos

P: función de probabilidad

La función de probabilidad P es una función $P: P(S) \rightarrow [0, 1]$ que cuantifica la posibilidad de ocurrencia de un evento aleatorio perteneciente a $P(S)$.

Propiedades Resumidas

miércoles, 3 de diciembre de 2025 10:53

Faltan Algunas

- Ley de Laplace: $P_{(A)} = \frac{n^0 \text{ de } Casos\ Favorables}{n^0 \text{ de } Casos\ Posibles}$
- Suceso Seguro: $P(E) = 1$
- Suceso Imposible: $P(\emptyset) = 0$
- Sucesos Incompatibles $P(A \cap B) = 0$

Tema 2

- Sucesos Complementarios/Opuestos: $P(A') = 1 - P(A)$
- Si $A \subseteq B \Rightarrow P(B-A) = P(B) - P(A)$
- $P(A \cap B') = P(A) - P(A \cap B) = P(A-B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Sucesos Independientes $P(A/B) = P(A) \wedge P(B/A) = P(B)$
- Union de Sucesos Incompatibles $P(A \cup B) = P(A) + P(B)$
- Intersección de Sucesos Independientes $P(A \cap B) = P(A) \cdot P(B)$
- Intersección de Sucesos Dependientes $P(A \cap B) = P(A) \cdot P(B/A) = P(B) \cdot P(B/A)$

Tema 3

- Leyes de De Morgan $P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B}) \wedge P(\overline{A \cap B}) = P(\bar{A} \cup \bar{B})$
- Condicional $P(A/B) = \frac{P(A \cap B)}{P(B)}$
- Teorema de Bayes $\frac{P(A_j \cap B)}{P(B)} = \frac{P(B/A_j) \cdot P(A_j)}{\sum_{i=1}^{n} P(B/A_i) \cdot P(A_i)} = P(A_j/B)$

- Función de Distribución $F(x) = P(X \leq x) = \sum_{\{x_i \leq x\}} P(X = x_i) = \sum_{\{x_i \leq x\}} p(x_i)$

Tema 4...

- $E(x) = \sum_{x \in R_x} x \cdot p(x)$
- $Var(x) = \sigma_x^2 = E(x^2) - (E(x))^2$

- $Cov(x,y) = E(xy) - E(x) \cdot E(y)$
- $E(xy) = \sum_{i=1} \sum_{j=1} x_i \cdot y_j \cdot p_{ij}$
- $E(g(x,y)) = \sum_i \sum_j g(x_i, y_j)\, p_{ij}$

  si son independientes

  $\rightarrow E(x) \cdot E(y) = E(xy)$

Tema 5

- $m_{rs} = \sum_i \sum_j x_i^r \cdot y_j^s \cdot p_{ij}$
- $\mu_{rs} = E((X - E(x))^r \cdot (Y - E(Y))^s)$

  $= \sum_i \sum_j (x_i - \mu_x)^r \cdot (y_j - \mu_y)^s \cdot p_{ij}$

- $f(y/x) = \frac{f(x,y)}{f(x)}$

Variables Aleatorias Unidimensionales

martes, 3 de febrero de 2026 21:29

Informalmente una variable aleatoria es una característica numérica de un evento aleatorio

Variables Aleatorias Bidimensionales

viernes, 13 de febrero de 2026 06:44

$$f(x,y) = \begin{cases} \frac{1}{3}xy + y^2 & \text{si } 0<x<2;\ 0<y<1 \\ 0 & \text{c.c.} \end{cases}$$

$$F(x,y) = \begin{cases} 0 & \text{si } x \le 0 \wedge y \le 0 \\ \frac{1}{12}x^2y^2 + \frac{1}{3}x.y^3 & \text{si } 0<x<2 \wedge 0<y<1 \\ \frac{1}{3}y^2 + \frac{2}{3}y^3 & \text{si } x \ge 2 \wedge 0<y<1 \\ \frac{1}{12}x^2 + \frac{1}{3}x & \text{si } 0<x<2 \wedge y \ge 1 \\ 1 & \text{si } x \ge 2 \wedge y \ge 1 \end{cases}$$

$F_Y(y)$

$F_X(x)$

$$f_X(x) = \frac{x}{6} + \frac{1}{3} \quad \text{si: } 0<x<2$$

$$f_Y(y) = \frac{2}{3}y + 2y^2 \quad \text{si: } 0<y<1$$

$$f(x) = u^m \Rightarrow f'(x) = m.u^{m-1}(u')$$

$$\left[\left(\frac{x}{3}\right)^3\right]' = 3.u^2.u'$$

$$\frac{x}{3} = u$$

$$3 \cdot \left(\frac{x}{3}\right)^2 \cdot \frac{1}{3}$$

$$\cancel{3} \cdot \frac{x^2}{9} \cdot \frac{1}{\cancel{3}} = \frac{x^2}{9}$$

$$\frac{x^3}{27} \qquad \frac{1}{27} \cdot x^3 = \frac{1}{\cancel{27}} \cancel{3} x^2 = \frac{1}{9} x^2$$

Modelos Probabilísticos Discretos

lunes, 9 de febrero de 2026 16:03

Bernoulli: $X \sim B(\alpha)$

· éxito = 1
· fracaso = 0

· caso especial de la binomial

$p(1) = P(X=1) = \alpha$ $\Rightarrow$ $p(x) \begin{cases} \alpha & x=1 \\ 1-\alpha & x=0 \\ 0 & \text{caso contrario} \end{cases}$

· casi no se usa equis de

Sea X v.a. que puede tomar solo 2 valores

$E(X) = \mu = \alpha \rightarrow 0.q + 1.\alpha = \alpha$

$Var(X) = \sigma_x^2 = \alpha.(1-\alpha)$

Deducción Varianza

La varianza, por definición (Unidad 1), es la media de los cuadrados de la diferencia entre los valores muestrales y la media

Unidad 1 : $\dfrac{\sum_i^n (x_i - \bar{x})^2}{n}$ (en la práctica usamos $n-1$ porque da mejores aproximaciones)

$E(X) = \mu = \sum_i^n x_i . p(x_i)$

media de los cuadrados de la diferencia entre los valores muestrales y la media

$$E(X-\mu)^2 = E(x - E(x))^2 = E(X^2) - (E(x))^2$$

$$= E(X^2 - 2.X.E(X) + (E(x))^2) = \text{Linealidad}$$

$$= E(X^2) - E(2.X.E(x)) + E((E(X))^2) = \text{Los Cte. se pueden sacar}$$

$$= E(x^2) - 2.E(x).E(x) + \underline{E((E(X))^2)} \quad \text{Esp. de una Cte.} = \text{Cte}$$

$$= E(X^2) - 2(E(x))^2 + (E(x))^2$$

$$\boxed{= E(X^2) - (E(X))^2}$$

$\sigma_x^2 = \alpha.(1-\alpha) \wedge \mu = \alpha$

$\sigma_x^2 = \alpha \cdot (1-\alpha) \quad \wedge \quad \mu = \alpha$

$x = \{0,\ 1\}$

$$E(x) = \sum_i^n x_i \cdot p(x_i) = 0 \cdot 1-\alpha + 1 \cdot \alpha = \alpha$$

$$E(x^2) = \sum_i^n (x_i)^2 \cdot p(x_i) = (0)^2 \cdot 1-\alpha + (1)^2 \cdot \alpha = \alpha$$

$$E(x^2) - (E(x))^2$$

$$\alpha - \alpha^2 = \text{factor común } \alpha$$

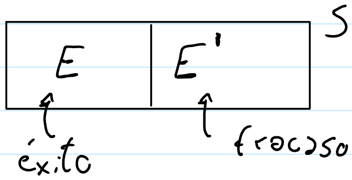
$$\alpha \cdot (1-\alpha)$$

Distribución Binomial $\quad X \sim \mathcal{B}(n, p)$

1) **n** repeticiones, de una misma prueba simple
↳ estas son independientes entre sí

2) Los resultados de las pruebas se clasifican en dos **eventos opuestos**, los cuales serán llamados arbitrariamente: (porque nos conviene básicamente)

- éxito
- fracaso



S: espacio muestral
$S = \{E, E'\}$
$S = E \cup E'$
$E' = 1 - E$
$\emptyset = E \cap E'$

3) La probabilidad de Éxito permanece cte. $P(E) = p$
La probabilidad de fracaso es $P(E') = q$

- Se verifica en tanto que $p + q = 1$
- $\Rightarrow q = 1 - p$

Si ambas son verdaderas, se supone: $0 < p < 1$

4) Interesa saber cuantos éxitos **p** hay en una cierta cantidad **n** de experimentos

Definición

Sea X una v.a. sobre un esp. de prob. $(S, \mathcal{A}, P)$, diremos que X tiene Distr. Bi. de parámetros n y p, $n \in \mathbb{N}$ y $p \in (0;1)$ si su función de probabilidad puntual es:

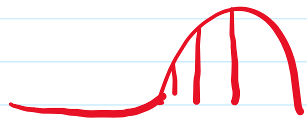
$$p(x) = P(X=x) = \binom{n}{x} p^{x} \cdot (1-p)^{n-x}$$

$$E(X) = \mu = np$$

$$Var(X) = \sigma_x^2 = npq$$

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{npq}$$

$\bar{x} > Med > Mo$



Cola izq

Distribución Multinomial

Resumencito Discretos

lunes, 9 de febrero de 2026 17:47

- **Bernoulli** $X \sim Bernoulli(p)$

↳ Discreto
↳ Casi no se usa
↳ Caso especial de Binomial

$p(x) = P(X=x) = \underbrace{p^x (1-p)^{n-x}}$

Binomial sin Combinatoria y un solo intento

$E(X) = \mu = \alpha = p$

$Var(x) = \sigma_x^2 = \alpha \cdot (1-\alpha)$
$= p \cdot q$

$\sigma_x = \sqrt{p \cdot q}$

con $x \in \{0, 1\}$

↳ $p(1) = P(X=1) = \alpha$

$$p(x) = \begin{cases} \alpha & x = 1 \\ 1-\alpha & x = 0 \\ 0 & c.c. \end{cases}$$

- **Binomial** $X \sim Bi(n, p)$

↳ Discreta
↳ n = número de exp.
↳ p : prob. de éxito (de)

$$P(X=x) = p(x) = \binom{n}{x} p^x \cdot (1-p)^{n-x}$$

con

$p \in (0, 1) \wedge n \in N$

es decir: $p(x) \begin{cases} \binom{n}{x} p^x \cdot (1-p)^{n-x} & x \in \{0, 1, 2, \dots, n\} \\ 0 & c.c. \end{cases}$

$E(X) = \mu = n \cdot p$

$Var(X) = \sigma_x^2 = n \cdot p \cdot q$

$\sigma_x = \sqrt{n \cdot p \cdot q}$

- **Multinomial**

Generalización de la binomial para experimentos con k resultados posibles

↳ Discreto
↳ n: número de exp.
↳ p_i: prob. de éxito
↳ r_i: X_i

$X \sim M_k(p_1, p_2, \dots, p_k, n)$

$$\sum_{i=1}^{k} r_i = n$$

$\to r_i = X_i$

$$\sum_{i=1}^{k} r_i = n$$

$$P(X_1 = r_1, \ldots, X_k = r_k) = \frac{n!}{r_1! r_2! \ldots r_k!} \cdot p_1^{r_1} \cdot p_2^{r_2} \cdot \ldots \cdot p_k^{r_k}$$

$E(X_i) = n \cdot p_i$ $\quad \sigma^2_{X_i} = n \cdot p_i \cdot (1 - p_i)$ $\quad Cov(X_i, X_j) = -n \cdot p_i \cdot p_j$

- **Geometrica**

$\hookrightarrow$ Discreta

$\hookrightarrow$ P hasta 1er éxito

$\hookrightarrow$ No tiene memoria

$P(X \geq s+t / X > s) = P(X \geq t)$

$X \sim G(p)$

$p(x) = (1-p)^{x-1} \cdot p$

$E(x) = 1/p$ $\quad Var(x) = \dfrac{1-p}{p^2}$

Binomial Negativa

$\hookrightarrow$ Discreta

$\hookrightarrow$ Número de exp. hasta obtener r éxitos

$\hookrightarrow$ r: cant. éxitos

$X \sim Bn(p, r)$

$p(X) = \binom{x-1}{r-1} (1-p)^{x-r} \cdot p^r$, $x = r, r+1, \ldots$

$\mu = E(x) = r/p$ $\quad Var(x) = r \cdot \dfrac{1-p}{p^2}$

Hipergeometrica

$\hookrightarrow$ Discreta

$\hookrightarrow$ Se tiene una población N, una muestra n, y una cantidad de éxitos n_1

$X \sim h(N, n_1, n)$

$$P(X = r) = \frac{\binom{n_1}{r} \cdot \binom{N-n_1}{n-r}}{\binom{N}{n}}$$

, Con $r \in \mathbb{N}_0$ sujeto a ① $r \leq n_1$ ② $(n-r) \leq (N - n_1)$

$$P(X=r) = \frac{\binom{n_1}{r}\binom{N-n_1}{n-r}}{\binom{N}{n}}, \quad \text{Con } r \in \mathbb{N}_0 \text{ sujeto a } \begin{cases} (1)\ r \leq n_1 \\ (2)\ (n-r) \leq (N-n_1) \end{cases}$$

$$E(X) = n \cdot \frac{n_1}{N} \qquad \sigma_X^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{n_1}{N} \cdot \left(1 - \frac{n_1}{N}\right)$$

Poisson

↳ Discreta

↳ Cuenta el número de eventos que ocurre en un intervalo de tiempo (volumen, espacio)

$$X \sim P(\lambda) \qquad \lambda = \mu \cdot \Delta t$$

$$P(X=r) = \frac{\lambda^r}{r!} \cdot e^{-\lambda}, \quad r \in \mathbb{N}_0$$

$$E(X) = \lambda, \quad \sigma_X^2 = \lambda$$

Relaciones entre modelos

Binomial y Poisson

Supongamos $X \sim B(n, p)$

$$(n \to \infty) \wedge (p \to 0) \Rightarrow \binom{n}{r} p^r \cdot (1-p)^{n-r} \cong \frac{(n \cdot p)^r}{r!} \cdot e^{-n \cdot p}$$

Regla Práctica: $n > 100$, $p < 0{,}05$ y $n \cdot p \leq 20$

- Dada p pequeña, cuanto mayor sea n mejor aproximación
- Dada n, cuanto menor sea p mejor aproximación

Hipergeométrica y Binomial

$$B(n,p) = \lim_{N\to\infty} h(N, n_1, n)$$

Si $N > 50$ y $\frac{n}{N} \leq 0,1$ da igual tomar observaciones con reemplazo ($X \sim B(n,p)$) o sin reemplazo ($X \sim h(N, n_1, n)$)

Binomial Ejemplos

lunes, 9 de febrero de 2026 18:34

Ejemplo. *La probabilidad de que un satélite, después de colocado en órbita, funcione de manera adecuada es 0,9. Supóngase que 5 de éstos se colocan en órbita y operan de manera independiente. Calcular la probabilidad de que:*

a) *El 80% de los satélites funcionen adecuadamente.*
b) *El número de satélites que no funcionan adecuadamente es 2.*
c) *Funcionen adecuadamente a lo sumo 4 satélites.*
d) *Funcionen adecuadamente por lo menos 2.*

$X \sim Bi(n, p)$

$n = 5$ X = v.a. que indica el núm. de satelites en órbita que funcionan correct.

$p = 0,9$

$\therefore X \sim Bi(5, 0,9)$

a) 80% de 5 = 4 $\Rightarrow P(X=4) = ?$

$$p(4) = \binom{5}{4} 0,9^4 \cdot (1-0,9)^{5-4} = 5 \cdot \underbrace{0,9^4}_{0,6561} \cdot 0,1 = 0,32805$$

b) Fracasos = 2 $\Rightarrow$ 5 - 2 = 3

Éxitos = X = 3

$$p(3) = \binom{5}{3} 0,9^3 \cdot (1-0,9)^{5-3} = \cancel{10} \cdot \underbrace{0,9^3}_{0,729} \cdot \cancel{0},01 = 0,0729$$

c) A lo sumo 4 satélites $\Rightarrow$ Como máximo 4 satélites

$$P(X \leq 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$$

$$= 1 - P(X=5) \qquad S = \{0, 1, 2, 3, 4, 5\}$$

$$= 1 - p(5)$$

$$= 1 - \binom{5}{5} 0,9^5 \cdot (1-0,9)^{5-5}$$

$= 1 - \binom{5}{5} 0{,}9^{5} \cdot (1-0{,}9)$

$= 1 - 1 \cdot 0{,}9^{5} \cdot 1 = 1 - 0{,}5905 = 0{,}4095$

d) Por lo menos 2 $\Rightarrow X \geq 2$

$P(X \geq 2) = P((X < 2)')$

$P((X<2)') = 1 - (P(X=0) + P(X=1))$

$= 1 - (p(0) + p(1))$

$= 1 - 0{,}000451$

$= 0{,}999541$

$p(0) = \binom{5}{0} 0{,}9^{0} \cdot (1-0{,}9)^{5} = 0{,}00001$

$p(1) = \binom{5}{1} 0{,}9^{1} \cdot (1-0{,}9)^{5-1} = 0{,}00045$

$p(2) = \binom{5}{2} 0{,}9^{2} \cdot (1-0{,}9)^{5-2} = 0{,}0081$

$p(3) = 0{,}0729$

$p(4) = 0{,}32805$

$p(5) = 0{,}5905$

$\sum_{i=2}^{5} = 0{,}99955$

$$\sigma_X^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{n_1}{N} \cdot (1 - n_1/N)$$

$$D^2(X) = n\left(\frac{N_1}{N}\right)\left(\frac{N_2}{N}\right)\underbrace{\left(\frac{N-n}{N-1}\right)}_{\text{factor de corrección}}$$

$$n \cdot \left(\frac{N-n}{N-1}\right) \cdot \left(\frac{n_1}{N}\right) \cdot$$

$$N = n_1 + n_2$$
$$n_2 = N - n_1$$

$$1 - \frac{n_1}{N} = \frac{n_2}{N}$$

$$\frac{n^2}{N} = 1 - \frac{n^1}{N}$$

Resumencito Continuos

martes, 24 de febrero de 2026 16:59

Uniforme

↳ Continua
↳ a : fin inf
↳ b : fin sup

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \le x \le b \\ 0 & \text{c.c.} \end{cases}$$

↳ Describe una prob. cte. en [a, b]

$X \sim U(a, b)$

$$F(x) = \begin{cases} 0 & x < a \\ \frac{(x-a)}{(b-a)} & a \le x \le b \\ 1 & x > b \end{cases}$$

$$E(X) = \frac{(a+b)}{2} \qquad \sigma_x^2 = \frac{(b-a)^2}{12}$$

Exponencial $X \sim \mathcal{E}(\alpha)$

$$f(x) = \alpha \cdot e^{-\alpha \cdot x}, \quad x \ge 0$$

$$F(x) = 1 - e^{-\alpha \cdot x}$$

$$E(X) = 1/\alpha \qquad \sigma_x = 1/\alpha$$

↳ Continua
↳ No tiene memoria
↳ $\alpha = 1/\mu$
↳ Describe tiempos entre eventos de Poisson

Normal $X \sim N(\mu, \sigma^2)$

↳ Continua

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

$$E(X) = \mu \qquad Var(X) = \sigma^2$$

• Para estandarizar: $Z = \frac{x-\mu}{\sigma} \Rightarrow Z \sim N(0,1)$

Relaciones entre modelos

Binomial y normal

- Se debe cumplir: $n.p > 5 \wedge n.(1-p) > 5$
- Se corrige con $\pm 0{,}5$

$$X \sim B(n,p) \approx N(\mu = n.p,\ \sigma^2 = n.p.(1-p))$$

Exponencial y Poisson

$X \sim P(\lambda)$, $\lambda = \mu.t \Rightarrow X \sim \mathcal{E}(\mu)$

- Cuenta cuanto tiempo hay entre dos eventos sucesivos de Poisson

Demostración

Sean: X v.a. que cuenta el tiempo entre dos eventos sucesivos

Y v.a. que cuenta cuantos eventos se dan en un intervalo de tiempo t

y además $Y \sim P(\mu.t)$

$\Rightarrow P(X \leq t) = 1 - P(X > t) = 1 - P($No ocurrió ningún evento en el intervalo $(0,t])$

$= 1 - P(Y=0) = 1 - \frac{\lambda^0}{0!} . e^{-\mu.t} = 1 - e^{-\mu.t}$

Poisson y normal

- Se debe cumplir $\lambda \geq 30$

- Se recomienda aplicar corrección $\pm 0{,}5$

- Se recomienda aplicar corrección $\pm 0,5$

$X \sim P(\lambda) \approx N(\mu = \lambda, \sigma^2 = \lambda)$

TCL

lunes, 16 de febrero de 2026 15:47

Teorema 9.4: TCL de Lindenberg-Levy

Sean $X_1, X_2, \ldots$ variables aleatorias independientes e indénticamente distribuidas, con $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$. Sea $S_n = X_1 + X_2 + \ldots + X_n$. Entonces

$$\frac{Sn - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$$

$X \sim Bern(n, p)$

Teorema 9.6 (DeMoivre- Laplace):

Sea S_n el número de éxitos en n ensayos Bernoulli independientes, con probabilidad p de éxito en cada ensayo. Entonces

$$\frac{Sn - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0,1)$$

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

$$\frac{Sn - n\mu}{\sigma . \sqrt{n}} \Rightarrow \frac{Sn - n.p}{\sqrt{p.(1-p)} . \sqrt{n}} = \boxed{\frac{Sn - n.p}{\sqrt{n.p.(1-p)}}}$$

$Z \sim N(0, 1)$ $\quad Z = \dfrac{X - \mu}{\sigma}$ para una v.a

Sean $X_1, X_2, \ldots, X_n$ v.a. i.i.d. con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$

Sea $S_n = \sum_{i=1}^{n} X_i$. Entonces:

$$\frac{S_n - n.\mu}{\sigma . \sqrt{n}} \xrightarrow{D} N(0, 1)$$

converge en distribución

T.C.L.

Lindenverg - Levy

- Para el TCL de DeMoivre - LaPlace se toma a Sn como el número de éxitos en n ensayos de Bernoulli

$X \sim Bernoulli\,(p)$

$$P(x) = \begin{cases} p & x=1 \\ 1-p & x=0 \\ 0 & C.C \end{cases}$$

$E(X) = p = \mu \qquad Var(X) = p\cdot(1-p) \qquad \sigma = \sqrt{Var(x)} = \sqrt{p\cdot(1-p)}$

$$\Rightarrow \frac{S_n - n\cdot\mu}{\sigma\cdot\sqrt{n}} = \boxed{\frac{S_n - n\cdot p}{\sqrt{p\cdot(1-p)\cdot n}}}$$ T.C.L. DeMoivre-LaPlace

En I.C. usamos los T.C.L

$$X \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right) \Rightarrow Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$Z = \dfrac{X-\mu}{\sigma} \qquad \bar{X} = \dfrac{S_n}{n}$

TCL Lindeberg-Levy

$$\Rightarrow \frac{S_n/n - \mu}{\sigma/\sqrt{n}} \cdot \frac{n}{n} = \frac{S_n/\cancel{n}\cdot\cancel{n} - n\cdot\mu}{\frac{\sigma}{\sqrt{n}}\cdot n} \overset{(x)}{=} \boxed{\frac{S_n - n\cdot\mu}{\sigma\cdot\sqrt{n}}}$$

(x) $n = (\sqrt{n})^2 \Rightarrow \dfrac{n}{\sqrt{n}} = \dfrac{\cancel{\sqrt{n}}\cdot\sqrt{n}}{\cancel{\sqrt{n}}} = \sqrt{n}$

$$\frac{n}{\sqrt{n}}\cdot 1 = \frac{n}{\sqrt{n}}\cdot\frac{\sqrt{n}}{\sqrt{n}} = \frac{\cancel{n}\cdot\sqrt{n}}{\cancel{(\sqrt{n})^2}} = \sqrt{n}$$

Nivel de confianza $1-\alpha$ donde α es el nivel de significancia

$\theta \in (\hat{\theta}_i, \hat{\theta}_s)$ anotar N.C.

¿n es grande?

- **No** → Obligatorio que la muestra provenga de una N (población) de distribución normal
  - μ
    - σ conocida → $Z \sim N(0,1)$
    - σ desconocida → $T \sim t_{n-1}$
  - σ^2
- Si:
  - μ (TCL)
    - σ conocida → usar σ
    - σ desconocida → usar S
    - } $Z \approx N(0,1)$
  - p → $\hat{p}$ → $Z \approx N(0,1)$

No se ponen condiciones respecto al tipo de distribución si la muestra es grande ($n \geq 30$)

$$\begin{array}{ccccc} \bar{X} & & t_{n-1,\,1-\alpha/2} & & S/\sqrt{n} \\ \text{IC} = \text{Estimador} & \pm & \text{Valor Crítico} & \cdot & \text{Error Estimado} \end{array}$$

Dato ± Estadístico · Desvío Estándar

#• Solvo σ^2, donde se aplica:

$$T = \frac{Z}{\sqrt{\frac{\chi^2}{v}}} \Rightarrow \sqrt{\frac{\chi^2}{v}} = \frac{Z}{T} \Rightarrow \sqrt{\frac{\chi^2}{v}} = \frac{S}{\sigma} \qquad v = n-1$$

$\cancel{\bar{X}-\mu}$

$\sqrt{\frac{X^2}{v}}$ $\quad v \quad v \quad T \quad v \quad v \quad \sigma$

$$= \frac{\cancel{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}}{\cancel{\frac{\bar{x}-\mu}{s/\sqrt{n}}}} \qquad \frac{X^2}{v} = \frac{s^2}{\sigma^2}$$

$$X^2 = \frac{s^2 \cdot v}{\sigma^2}$$

μ y σ desconocidas
=> la $\gamma_{v.a.}$

$$\sigma^2 = \frac{s^2 \cdot v}{X^2}$$

$\frac{v \cdot s^2}{\sigma^2} \sim \chi^2$, con $n-1$ g.l., siendo $v = n-1$

y un intervalo del $100 \cdot (1-\alpha)\%$ de confianza para σ^2 de una distribución normal es:

$$\left(\frac{v \cdot s^2}{\chi^2_{n-1,\,\alpha/2}} , \frac{v \cdot s^2}{\chi^2_{n-1,\,1-\alpha/2}} \right)$$

Test de Hipótesis: $\frac{\text{Estimador} - \text{Valor Hipotetizado}}{\text{Error Estimado}}$ Error Rel

$$\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}}$$:

$X \sim B(n, p)$

$$Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2} \cdot Var(X)$$

$$= \frac{1}{n^{\cancel{2}}} \cancel{n} \cdot p \cdot q$$

$$= \frac{p \cdot q}{n}$$

$$= \frac{\hat{p} \cdot (1-\hat{p})}{n}$$

T.H.

viernes, 13 de febrero de 2026 06:44

b) Datos

$X \sim N(52, \sigma^2)$

$\alpha = 0{,}05$ $\qquad \bar{X} = 49{,}8$ min

$\mu_0 = 52$ min ← param. nulo

μ ← param. de interés

Paso a Paso

1. Definir el N.S. α

$\alpha = 0{,}05$

2. Identificar el parámetro de interes

μ

3. Definir param nulo y proponer hip. nula.

$H_0: \mu = \mu_0$

$\mu = 52$ min

4. Plantear hip. alternativa

$H_1: \mu < \mu_0$

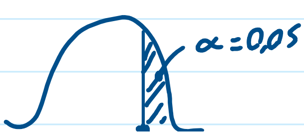
$\mu < 52$ min

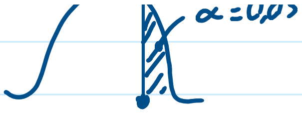
5. Identificar Est. de Prueba

$$t_0 = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{\underbrace{n-1}_{24}}$$

6. Establer R.R.

$R.R. = \{t \,/\, t \leq -1{,}711\}$





$t_{1-\alpha} = -1,711$

7 Calcular cantidades necesarias

$$t_0 = \frac{49,8 - 52}{4,5 / \sqrt{25}}$$

$t_0 = -2,4444$

$t \le -t_{\alpha, n-1}$

$2,4444 \le -1,711$ ✓

8 Se toma una desición (se rechaza o no H_0)

$t_0 \in R.R. \Rightarrow$ Se rechaza H_0 en favor de H_1

La muestra presenta evidencia suficiente como para rechazar la hipótesis nula, con un N.S. del 5% a favor de la alternativa que indica que el tiempo promedio de entrega actual es menor a 52 min.

Errores T.H.

lunes, 23 de febrero de 2026 18:33

$Z \geq 1 \rightarrow 1 - P(Z < 1)$

$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

.

$$\beta(185) = P(\bar{x} \geq 190/\mu = 185) = P\left(\frac{\bar{x} - 185}{15/3} \geq \frac{190 - 185}{15/3}\right)$$

$$= 1 - P(Z < 1) = 1 - 0.84 = 0.16$$

$\phi((\mu_0 + Z_c \cdot \sigma/\sqrt{n} - \mu_1)/\sigma/\sqrt{n})$

Acepto H0/ H0 es falso

$\mu = 132$

$n = 9 \Rightarrow \sqrt{n} = 3$

$B(13\ ) = P(\bar{x} > 130/\mu = 132) =$ x $-$ 1 -1

$1,5$

$44\ )$

$-$ $Z < 0,444)$

p- valor

miércoles, 25 de febrero de 2026 10:07

No sé la fecha de este

viernes, 6 de febrero de 2026 16:11

Regulares, sin tablas porque la profesora magdalena confeccionó el examen para que a priori no necesitemos tablas:

1. Estadística Descriptiva. Tipos de variables. Escalas de Medición. Cuando usar intervalos y supuestos. Como definir número de intervalos a usar. Definir fi y ri. 2,5 pts

2. Enunciar y ejemplo de Ley de Probabilidad Total. 1 punto

3. Definir espacio muestral y variable aleatoria. Caso continuo ejemplificar. Definir esperanza matemática. Hallar P (a <X <=b) conociendo sólo f (x). 1 punto.

4. Elegir el modelo de distribución adecuado para 2 escenarios dados. Seleccionar uno de ellos y desarrollarlos exhaustivamente. 2,5 puntos

5. Enunciar al menos un TCL y donde se uso en la materia. 1 punto

6. Prueba de Hipótesis unilateral. Casos y supuestos. Nivel de significancia. Forma de la hipótesis nula y alternativa en unilaterales. Cuando usar Student y cuando usar z. Desarrollar ejemplo con pasos en el contexto del ejemplo. No calcular. 2 puntos.

Cuando desarrollen binomial pongan que las n repeticiones son repeticiones del experimento aleatorio
A mi me desconto medio punto por eso kjjj

1. Estadística Descriptiva: Es la rama de la estadística que se encarga de recolectar, analizar y describir una muestra/población de datos.

   Tipos de Variables:
   a. Cualitativas: Expresan cualidades o características que puedan ser agrupadas o tomadas por la variable en estudio. Pueden ser:
      i. Dicotómicas: 2 categorías o clases
      ii. Policotómicas: 2 < categorías o clases
   b. Cuantitativas: Expresan cantidades, es decir, tienen un valor numérico (?. Pueden ser:
      i. Discretas: Surgen de contar (solo pueden tomar valores discretos dentro de su campo de variación, es decir, solo son enteros)
      ii. Continuas: Surgen de medir (toman cualquier valor dentro de su rango de variación)

   Escalas de Medición:
   a. Cualitativas:
      i. Nominal o Clasificatoria
         - Excluyente y exhaustivo
         - Cuenta cuantos elementos pertenecen a c/u
         - Es el nivel de medición más bajo
      ii. Ordinal o Jerárquica
         - Es posible ordenar los datos en jerarquías
         - Se pueden establecer relaciones lógicas:
            - Equivalencia (dentro de cada categoría)
            - Orden Estricto (entre categorías)
   b. Cuantitativas
      i. Intervalos
         - Deben ser números
         - El punto de origen es llamado cero arbitrario
            - Admite negativos
      ii. Razón o Proporción
         - Deben ser números
         - El punto de origen es realmente 0, y es llamado cero absoluto
         - No admite negativos
         - Se puede establecer una distancia o proporcionalidad entre dos entes cualesquiera
         - Es el nivel de medición más alto

   Se utilizan intervalos cuando X es una variable cuantitativa pero:
   - Hay demasiados datos
   - Hay pocos datos pero muy dispersos
   - Interesa una clasificación específica de los resultados

   Para definir cuantos intervalos se deben usar:
   - Se considera raíz de n como una primera aproximación
   - Se recomienda 5 <= Cant Int <= 20
   - Puede definirse previamente de acuerdo al criterio de los investigadores.
   - No pueden existir intervalos vacíos (frecuencia 0). En tal caso, definir intervalos de distinta amplitud
   - La agrupación en intervalos de no debe modificar la forma de la distribución original de los datos

   Definición de:
   - .fi = frecuencia absoluta simple de Xi.
      - DAS: Es el número de veces que se repite ese valor x
      - DAIC: Es el número de veces que un valor cae dentro del íntervalo i
   - .ri = frecuencia relativa simple de Xi. .fi/n

2.

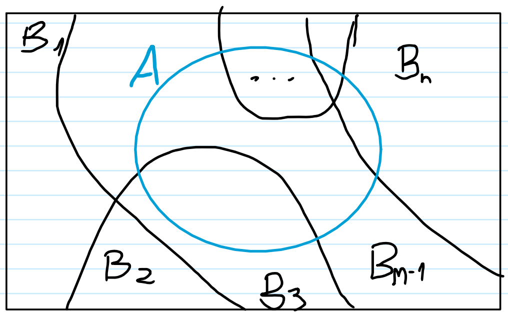
Sea $(S, \mathcal{A}, P)$ un espacio de probabilidad. Donde {B1, B2, ..., Bn} es ⊆ S

Y se cumple lo siguiente:

- $\bigcup_{i}^{n} Bi = S$ (1)
- $Bi \cap Bj = \emptyset, \forall i \neq j$

Entonces, la colección de eventos {B1, B2, ..., Bn} es una partición aleatoria de S. Y con A siendo un evento de S, *entonces la Ley de la Probabilidad Total se define como:*

$$\sum P(A/B_i) \cdot P(B_i) = P(A)$$



Demostración:

$P(A) = P(A \cap S) \overset{(1)}{=} P(A \cap (\bigcup_{i=1}^{n} B_i)) =$

$= P(\bigcup_{i=1}^{n} B_i \cap A) = \sum_{i=1}^{n} P(B_i \cap A) =$

$= \sum_{i=1}^{n} P(A/B_i) \cdot P(B_i)$

Regla de la multip.

Ejemplo:

Una empresa proveedora de internet ofrece 3 planes diferentes de conexión, PC1, PC2 y PC3. Del total de los usuarios, se sabe que el 50% tiene el PC1, 20% el PC2 y 30% el PC3. Ultimamente hay problemas de conexión debido a unas reparaciones en las líneas de fibra óptica. Se sabe que el 20% de los usuarios del PC1 tienen problemas de conexión, mientras que este problema lo tienen el 40% de PC2, y el 2% de PC3.

¿Cuál es la probabilidad de que un usuario tenga problemas de conexión?

Sea $(S, \mathcal{A}, P)$ un espacio de prob

$S = \bigcup_{i=1}^{3} PC_i$

PC_i = Plan de conexión a internet i, $i = 1, 2, 3$

$PC_i \cap PC_j = \emptyset \quad \forall i \neq j$

EC = Se tiene Errores de conexión a internet

$\{PC_i\}_{i=1}^{3}$ es una partición aleatoria de S

$P(PC_1) = 0,5$ $\quad P(EC/PC_1) = 0,2$

$P(PC_2) = 0,2$ $\quad P(EC/PC_2) = 0,4$

$P(PC_3) = 0,3$ $\quad P(EC/PC_3) = 0,02$

$P(EC) = \sum_{i=1}^{3} (EC/PC_i) \cdot PC_i = 0,186$

Rta: La probabilidad de que un usuario tenga problemas de conexión es de 0,186

3. Definir espacio muestral y variable aleatoria. Caso continuo ejemplificar. Definir esperanza matemática. Hallar P (a <X <=b) conociendo sólo f (x). 1 punto.

El espacio muestral S es el conjunto donde están todos los resultados posibles de un experimento aleatorio. La definición del S dependerá del objetivo del análisis, donde S puede ser

- Discreto: Si consiste de un número finito o infinitamente numerable de resultados posibles. Cada uno de estos resultados se denomina punto muestral.
- Continuo: Si contiene al menos un intervalo (acotado o no) de números reales.

Una variable aleatoria unidimensional es una regla o función X, que asigna a cada elemento de S un número real, esto es:

- Continuo: Si contiene al menos un intervalo (acotado o no) de números reales.

Una variable aleatoria unidimensional es una regla o función X, que asigna a cada elemento de S un número real, esto es:

Sea $(S, \mathcal{A}, P)$ un E.P.

$X : S \rightarrow \mathbb{R}$

Además:

$$\forall x \in \mathbb{R}, \{\omega \in S : X(\omega) \leq x\} \in \mathcal{A}$$

Una variable aleatoria continua puede definirse como una v.a. X que puede tomar cualquier valor en un intervalo, es decir, su rango de valores incluye un intervalo (a,b) $\subseteq \mathbb{R}$

Ejemplo: X = la temperatura máxima diaria en Corrientes en el mes de febrero (X asume valores en (20, 45))

La esperanza matemática de una v.a. continua se define:

$$\mu = E(x) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx = \int_{-\infty}^{\infty} x \cdot dF(x)$$

Hallar P(a<X<=b), conociendo solo f(x)

Primero se necesita encontrar la F(x) a partir de integrar f(x)

Luego dependiendo del tramo en el que estén a y b, se usaran esos valores de la función de distribución para calcular la probabilidad.

Como ejemplo tendriamos:

P(x<=b) = F(b)
P(x<a) = F(a)

$$P(a < x \leq b) = \int_a^b f(s) \cdot ds = F(b) - F(a)$$

P(a<x<=b) = F(b)-F(a)

- Observación: al tratarse de v.a. continuas, la probabilidad puntual es 0, ya que sería el área de un punto. Por lo tanto:

$$P(a < x \leq b) = P(a \leq x \leq b) = P(a \leq x < b) = P(a < x < b)$$

11/09/2025

viernes, 6 de febrero de 2026 16:15

1) En el análisis de tráfico de red y ciberseguridad, se recopilan estadísticas sobre la cantidad de paquetes enviados por los usuarios, los tiempos entre conexión o los tamaños de los mensajes transmitidos. Estos datos permiten identificar patrones normales y detectar anomalías. **2p**

   a)Explique brevemente las diferentes medidas estadísticas, indicando qué información aportan en este contexto.

   b) Imagine que se registraron los tiempos de conexión (en ms) de un conjunto de usuarios ¿Porque sería importante analizar no solo la media sino también la dispersión y la asimetría de los tiempos para avalar la estadística de la red?

   c) Suponga que al analizar los tiempos de conexión de los usuarios se obtuvieron Media = 250 ms y coeficiente de asimetría = 2.5. ¿Qué puede decir en este caso? Justifique su respuesta.

   d) ¿Cómo ayudaría las medidas de variabilidad a completar el análisis?

2) Da dos definiciones diferentes de la probabilidad de eventos. **1.p**

3) Defina probabilidad condicional y de un ejemplo referido a sistemas de información. **1.p**

4) Defina independencia de eventos y nombrar al menos dos propiedades. **1.p**

5) Sea (X, Y) un vector aleatorio con probabilidad de masa:

$$P_{ij} = P\left(X = x_1\, Y = y_j\right)$$

   a)¿Qué tipo de variable aleatoria es? justifique su respuesta

   b)Determine el valor de verdad de la proposición:

$$P\left(X = x_i, Y = y_j\right) = P_{ij} = p_i \cdot p_j$$

   Justifique su respuesta. **1.5p**

6) En un servidor de atención al cliente, cada llamada que ingresa puede ser un cliente nuevo o de un cliente que ya llamó anteriormente. La probabilidad de que una llamada corresponda a un cliente nuevo es del 40%. Se analizaron 20 llamadas recibidas en un día y se desea estudiar la probabilidad de que cierto números de ellas provengan de clientes nuevos.
   a) Definir la variable en estudio, modelo de distribución de la variable y los parámetros. justifique.**1.p**
   b) Calcula la esperanza de la variable definida.**0.5p**

7) Un investigador desea analizar el tiempo promedio de ejecución de un algoritmo en un determinado servidor. Se supone que los tiempos de ejecución siguen una distribución normal con desvío estándar de 8 segundos. Se toman 15 ejecuciones al azar, obteniendo como media muestral 52 segundos. Se quiere determinar si el tiempo promedio de ejecución es significativamente mayor a 50 segundos. El valor de p obtenido en el test de hipótesis es de p = 0.166. Nivel de significancia es de 0.10%.

a)Detalle los pasos de un test de hipótesis en el contexto del problema. **0.75**

b)Utilizando el p-valor, de una conclusión con un Nivel de confianza del 90%. **0.5**

c)Compare la conclusión en el intervalo de confianza del 9'% para el tiempo promedio de ejecución (47,95 , 56.05). **0.75**.

1) En el análisis de tráfico de red y ciberseguridad, se recopilan estadísticas sobre la cantidad de paquetes enviados por los usuarios, los tiempos entre conexión o los tamaños de los mensajes transmitidos. Estos datos permiten identificar patrones normales y detectar anomalías. **2p**

a)Explique brevemente las diferentes medidas estadísticas, indicando qué información aportan en este contexto.

b) Imagine que se registraron los tiempos de conexión (en ms) de un conjunto de usuarios ¿Porque sería importante analizar no solo la media sino también la dispersión y la asimetría de los tiempos para avalar la estadística de la red?

c) Suponga que al analizar los tiempos de conexión de los usuarios se obtuvieron Media = 250 ms y coeficiente de asimetría = 2.5. ¿Qué puede decir en este caso? Justifique su respuesta.

d) ¿Cómo ayudaría las medidas de variabilidad a completar el análisis?

1. Las medidas estadísticas son las siguientes:
    a. Medidas de Tendencia Central: Nos dicen cual es el comportamiento "normal" de la red
        i. La media no es dice el tamaño promedio de un paquete o el ping promedio
        ii. La moda nos dice cual es el tamaño de paquete o el tiempo de conexión que más se repite
        iii. La mediana nos muestra que medida ocupa el lugar central de nuestros datos
    b. Medidas de Posición: Estas sirven para dividir los datos en partes iguales
        i. En este caso por ejemplo podríamos ver el percentil 99 para ver donde cae el 99% de los datos de tiempos entre conexión y así ver la calidad del servicio prestado
    c. Medidas de Dispersión o Variabilidad: Nos dicen que tan homogéneos o estables son los datos
        i. Un desvío estándar o un CV bajo significaría una conexión estable. Pero un desvío estándar o un CV grande indicaría que a veces el tiempo entre conexiones es rápido y otras se cae.
    d. Medidas de Forma
        i. Nos sirven para entender de forma más clara el como se distribuirían los datos, si es una distribución normal, asimétrica o si la distribución de los datos es achatada, normal o elevada. Nos permitiría visualizar gráficamente donde caen o el como están ordenados los datos. En este contexto nos sirve para visualizar mejor la variación de los datos o el como se concentran, ya sea el tamaño de los paquetes, el tiempo de conexión, o la cantidad de paquetes enviados por los usuarios.

2. Es necesario analizar la dispersión y la asimetría de los tiempos de conexión para avalar la estadística ya que la media es muy susceptible a los valores extremos recolectados.

   La dispersión indica la confiabilidad de la media. Una red con media de 200ms y desvío de 5ms es mejor que una con media de 150ms y desvío de 100ms. Una media baja no sirve de nada si la desviación estándar es alta, ya que implica un servicio impredecible.

   La importancia de la Asimetría: Lo ideal en una red es tener muchos usuarios con una buena conexión y muy pocos o ninguno con una conexión lenta o con problemas, esto se podría ver mediante la asimetría, donde en una asimetría positiva (con cola a la derecha) se podría validar como la se cumple esta característica, puesto que caerían la mayoría de los datos en un tiempo de conexión menor, pero habrían algunos con un tiempo de conexión muy grande (conexión más lenta o problemas de red) que estarían desplazando la media con estos valores extremos.
   Si la asimetría fuera negativa, indicaría que la mayoría de los usuarios tiene problemas, lo cual sería crítico.

3. Media = 250 ms; Coeficiente de Asimetria: 2.5
   Un coeficiente de asimetría de 2.5 es positivo y bastante alto (mayor a 0).
   Esto indica una Asimetría Positiva, es decir que la cola de la distribución se alarga hacia la derecha (hacia los valores altos).

Lo cual nos deja los datos de la siguiente manera:
$\bar{X} > Med > Mo$

En este contexto se podría interpretar que la gran mayoría de los usuarios tiene tiempos de conexión bajos (menores a la media de 250ms). Probablemente, la mediana sea mucho menor a 250ms (quizás 50ms o 100ms).

Sin embargo, hay un pequeño grupo de usuarios (la cola derecha) que tiene tiempos de conexión extremadamente altos (quizás 2000ms, 5000ms, etc).

Conclusión: Esos pocos usuarios con lag extremo están "arrastrando" el promedio hacia arriba hasta 250ms. La red funciona bien para la mayoría, pero hay casos puntuales graves que investigar.

4. Las medidas de variabilidad ayudan a complementar el análisis de las estadísticas recolectadas puesto que estas nos muestran la dispersión de los valores observados.

   Si el valor de estas medidas de dispersión o variabilidad es pequeño, los datos estarán estrechamente agrupados alrededor de la MTC, entonces dicha MTC se considera representativa. Caso contrario la MTC no es confiable.

   En este caso por ejemplo, si calculáramos la desviación estándar, se podría establecer un tiempo entre conexiones normal, por ejemplo: Media $\pm$ 2 Desviación Estándar. Y cualquier tiempo de conexión fuera de este rango se tomaría como una anomalía.
   Y el Coeficiente de Variación nos permitiría comparar la estabilidad entre diferentes servidores, independientemente de si uno es rápido y el otro lento. Nos diría cual es más "predecible".

2) Da dos definiciones diferentes de la probabilidad de eventos. **1.p**

Para asignar probabilidad a eventos, podemos proceder de dos maneras:
- Definición Clásica
  La probabilidad de un evento puede definirse como
  n° casos favorables/n° casos posibles

- Definición Frecuentista

$$P(A) = \lim_{n \to \infty} \frac{f_n}{n}$$

f_n cantidad de veces que ocurre el evento A en n ensayos

3) Defina probabilidad condicional y de un ejemplo referido a sistemas de información. **1.p**

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Suponiendo los eventos:
A = El ping de la red hogareña aumenta
B = Se envía un paquete pesado
A intersección B = El ping de la red hogareña aumenta, al mismo tiempo que se envía un paquete pesado

A/B = El ping de la red hogareña aumenta dado que se envió un paquete pesado

4) Defina independencia de eventos y nombrar al menos dos propiedades. **1.p**

Si $P(A \cap B) = P(A) \cdot P(B)$ A y B son indep.

Obs 1 Si A y B son indep $\Rightarrow P(A/B) = P(A)$

2 $P(A/B) = P(A) \iff P(B/A) = P(B)$

Propiedades

Propiedades

1. P(A) y P(A`) no son indep
2. Sean A y B ∈ $\mathcal{A}$ dos eventos excluyen y P(A)>0 P(B)>0, => no son indep
3. Sea A ∈ $\mathcal{A}$ y P(A) = 0 o P(A) = 1, entonces A es independiente de cualquier otro evento
4. Sean A y B dos eventos independientes => A y B`, A` y B`, A` y B también lo son

5) Sea (X, Y) un vector aleatorio con probabilidad de masa:

$$P_{ij} = P\left(X = x_1\ Y = y_j\right)$$

a)¿Qué tipo de variable aleatoria es? justifique su respuesta

b)Determine el valor de verdad de la proposición:

$$P\left(X = x_i, Y = y_j\right) = P_{ij} = p_i \cdot p_j$$

Justifique su respuesta. **1.5p**

a) Es una variable aleatoria bidimensional discreta.
   Porque su rango es un conjunto finito o infinitamente numerable de puntos. Esto lo sabemos porque el enunciado usa una función de probabilidad de masa conjunta, la cual solo está definida para valores puntuales de las variables X, Y.
b) Es falso, puesto que dicha proposición solamente es verdadera si se tratase de una variable aleatoria bidimensional independiente, y el enunciado no menciona que lo sean.

6) En un servidor de atención al cliente, cada llamada que ingresa puede ser un cliente nuevo o de un cliente que ya llamó anteriormente. La probabilidad de que una llamada corresponda a un cliente nuevo es del 40%. Se analizaron 20 llamadas recibidas en un día y se desea estudiar la probabilidad de que cierto números de ellas provengan de clientes nuevos.

a) Definir la variable en estudio, modelo de distribución de la variable y los parámetros. justifique.**1.p**
b) Calcula la esperanza de la variable definida.**0.5p**

Sea X la v.a. que cuenta la cantidad de llamadas correspondientes a un cliente nuevo:

$X \sim B(n, p) \Rightarrow X \sim B(20, 0{,}4)$

$$P(X=x) = \binom{n}{x} p^x \cdot (1-p)^{n-x}$$

$\mu = p \cdot n$ $\quad Var(X) = q \cdot p \cdot n = (1-p) \cdot p \cdot n$

b) $\mu = p \cdot n$
$\mu = 0{,}4 \cdot 20 = 8$

7) Un investigador desea analizar el tiempo promedio de ejecución de un algoritmo en un determinado servidor. Se supone que los tiempos de ejecución siguen una distribución normal con desvío estándar de 8 segundos. Se toman 15 ejecuciones al azar, obteniendo como media muestral 52 segundos. Se quiere determinar si el tiempo promedio de ejecución es significativamente mayor a 50 segundos. El valor de p obtenido en el test de hipótesis es de p = 0.166. Nivel de significancia es de 0.10%.

a)Detalle los pasos de un test de hipótesis en el contexto del problema. **0.75**

b)Utilizando el p-valor, de una conclusión con un Nivel de confianza del 90%. **0.5**

c)Compare la conclusión en el intervalo de confianza del ~~9%~~ 90% para el tiempo promedio de ejecución (47,95 , 56.05). **0.75**.

Datos

$X \sim N(\mu, 8^2)$
$\sigma = 8$ seg
$\mu_0 = 50$ seg

$n = 15$
$\bar{x} = 52$
p-valor $= 0{,}166$
$\alpha = 0{,}1$

a) Test de Hipotesis

1 - Identificar α
$\alpha = 0{,}1$

2 - Identificar parámetro de interés
μ

3 - Plantear parámetro e hipotesis nula
$\mu_0 = 50$ seg
$H_0 : \mu = \mu_0$
$\mu = 50$ seg

4 - Plantear hipótesis alternativa.

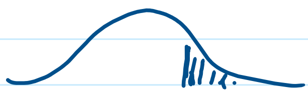
$H_1 : \mu > \mu_0$

$H_1: \mu > \mu_0$
$\quad \mu > 50$ seg

5 - Identificar estadístico de prueba

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim z_\alpha$$

6 - Plantear región de rechazo



$z_\alpha \Rightarrow z_{0,9} = 1,28$

$RR = \{ z / z \geq z_\alpha \}$
$\Rightarrow \{ z / z \geq 1,28 \}$

7 - Calcular cant. necesarias

$$Z = \frac{52 - 50}{8/\sqrt{15}} = 0,9682$$

8 - Se emiten conclusiones

$Z \notin R.R. \Rightarrow$ No se rechaza H_0

No existe evidencia suficiente para rechazar H0 en favor de la alternativa. Es decir que no existe evidencia suficiente para afirmar que el tiempo promedio de ejecución es significativamente mayor a 50 segundos.

b) p-valor $= 0,166$ $\qquad \alpha = 0,1$

p-valor $> \alpha \Rightarrow$ No rechazo H_0
p-valor $\leq \alpha \Rightarrow$ Rechazo H_0

En este caso como p-valor > alpha, no existe evidencia suficiente para rechazar H0 en favor de la alternativa

c) Datos

$\mu \in (47,95 \,,\; 56,05)$ con un N.C. del 90%.

$\mu_0 = 50$ seg $\qquad \bar{x} = 52$ seg

$\Rightarrow \mu_0 \in (47,95 \,,\; 56,05)$ Por lo tanto, no podemos afirmar que la media sea significativamente mayor a 50 segundos

16/10/2025

viernes, 6 de febrero de 2026 16:16

1. En una empresa se registran datos sobre la duración (en minutos) de las llamadas de atención al cliente realizadas por los empleados. El análisis de estos tiempos permite evaluar la eficiencia del servicio y detectar posibles irregularidades en la atención. 1.5p

   (a) Suponga que, al analizar los tiempos de atención, se obtiene:

   Empresa A: Media = 8.5 min, $CV = 20\%$ y Empresa B: Media = 9.0 min, $CV = 35\%$.

   Explique cuál empresa muestra una mayor homogeneidad en los tiempos de atención.

   (b) ¿Cómo ayudarían las medidas de variabilidad a complementar el análisis de los tiempos de atención de las llamadas?

2. Enuncie **detalladamente** la ley de la Probabilidad Total. Ejemplifique en un contexto diferente al de la gasolinera. 1p

3. Sea X una variable aleatoria tal que $E(X)$ está bien definida. Demostrar que, para cualesquiera $a, b \in \mathbb{R}$, se cumple la siguiente propiedad: $E(aX + b) = aE(X) + b$. 1p

4. Sea (X, Y) un vector aleatorio, con probabilidad de masa $p_{ij} = P(X = x_i, Y = y_j)$. 1p

   (a) ¿Qué tipo de variable aleatoria es? Justifique su respuesta.

   (b) Determine el valor de verdad de la proposición $P(X = x_i, Y = y_j) = p_{ij} = p_i \cdot p_j$. Justifique.

5. La función de densidad conjunta de (X, Y) es: 1.5p

$$f(x,y) = \begin{cases} \frac{1}{7}(x+2y), & 2 \le x \le 4,\ 1 \le y \le 3, \\ 0, & \text{caso contrario.} \end{cases}$$

   (a) Verificar que $f(x, y)$ es una densidad válida.

   (b) Calcular las densidades marginales $f_X(x)$ y $f_Y(y)$.

   (c) Verificar si X y Y son independientes.

6. En la guardia de cierto hospital ingresan, en promedio, 6 pacientes por hora. Sea X la variable aleatoria que cuenta la cantidad de pacientes que ingresan al hospital entre las 6 y las 10 de la mañana.

   (a) De los modelos vistos en esta asignatura, ¿cuál es el que mejor describe el comportamiento de la variable aleatoria X? ¿Cuáles son los o el parámetro de este modelo en esta situación en particular? Explique en detalle ese modelo. 1p

   (b) El director del hospital necesita saber cuánto tiempo transcurre entre dos ingresos sucesivos. Si Y es la variable aleatoria que mide el tiempo entre dos ingresos sucesivos, ¿qué modelo (de los vistos en esta materia) sugiere usar para calcular $P(Y < 0.1)$? Explique en detalle ese modelo, especificando cuál sería el parámetro de esa distribución en esta situación en particular. 0.5p

$\lambda = 24$

$\mu = 6$

$\Delta_t = 4$

$r = X$

$\frac{\lambda^r}{r!} \cdot e^{-\lambda}$

$\lambda = \mu \cdot \Delta t$

$X \sim P(\lambda)$

$X \sim P(24)$

$X \sim \mathcal{E}(\alpha)$

$P(X \le x) = 1 - e^{-\alpha \cdot x}$

Relación con Poisson

$1 - e^{-\mu \cdot t}$

$\Rightarrow P(X < 0{,}1) = 1 - e^{-6 \cdot 0{,}1}$

$P(X<0{,}1) = 0{,}4511$

$$f(x,y)=\begin{cases}\frac{1}{7}(x+2y), & 2\le x\le 4,\ 1\le y\le 3,\\ 0, & \text{caso contrario.}\end{cases}$$

$\int_2^4\int_1^3\left(\frac{1}{7}(x+2y)\right)dy\,dx = 1$ • para que sea valida

$\frac{1}{7}\int_2^4\int_1^3 (x+2y)\,dy\,dx$

$\frac{1}{7}\int_2^4\left(\int_1^3 x\,dy + \int_1^3 2y\,dy\right)dx$

$\frac{1}{7}\int_2^4\left(xy\Big|_1^3 + \frac{\cancel{2}y^2}{\cancel{2}}\Big|_1^3\right)dx$

$\frac{1}{7}\int_2^4 (2x + 8)\,dx$

$\frac{1}{7}\left(\int_2^4 2x\,dx + \int_2^4 8\,dx\right)$

$\frac{1}{7}\left(\frac{\cancel{2}x^2}{\cancel{2}}\Big|_2^4 + 8x\Big|_2^4\right)$

$\frac{1}{7}\left(\left[\cancel{16}-4+32-\cancel{16}\right]\right)$

$28/7 = 4$

$4 \neq 1$

<u>Marginales</u>

$f_x(x) = \int_1^3 \frac{1}{7}(x+2y)\,dy$

$\frac{1}{7}\int_1^3 (x+2y)\,dy$

$$\tfrac{1}{7}\left(\int_1^3 x\, dy + \int_1^3 2y\, dy\right)$$

$$\tfrac{1}{7}\left(x.y\Big|_1^3 + \frac{\cancel{2}y^2}{\cancel{2}}\Big|_1^3\right)$$

$$\tfrac{1}{7}\left(3x - 1x + 9 - 1\right)$$

$$\tfrac{1}{7}\left(2x + 8\right) \Rightarrow f_x(x) = \tfrac{1}{7}(2x+8), \quad 2 \le x \le 4$$

$$f_y(y) = \int_2^4 \tfrac{1}{7}(x + 2y)\, dx$$

$$\tfrac{1}{7}\int_2^4 (x + 2y)\, dx$$

$$\tfrac{1}{7}\left(\int_2^4 x\, dx + \int_2^4 2y\, dx\right)$$

$$\tfrac{1}{7}\left(\frac{x^2}{2}\Big|_2^4 + x2y\Big|_2^4\right)$$

$$\tfrac{1}{7}\left(8 - 2 + 8y - 4y\right)$$

$$\tfrac{1}{7}\left(4y + 6\right) \qquad f_y(y) = \tfrac{1}{7}(4y + 6), \quad 1 \le y \le 3$$

$$f_{(x,y)} = f_x(x) \cdot f_y(y) \iff \text{son ind.}$$

• En este caso:

$$f(x,y) = \tfrac{1}{7}(2x + y)$$

$$f_x(x) = \tfrac{1}{7}(2x + 8) \qquad f_x(x) \cdot f_y(y) = \tfrac{1}{49}(2x+8)(4y+6)$$

$$f_y(y) = \tfrac{1}{7}(4y + 6) \qquad \Rightarrow f(x,y) \ne f_x(x)\, f_y(y)$$

$\therefore$ No son indep.

7. En una empresa de logística se desea evaluar si el tiempo promedio de entrega de los pedidos ha mejorado respecto al valor histórico de referencia de 52 minutos. Para ello, se toma una muestra aleatoria de 25 entregas, obteniéndose un promedio de 49.8 minutos y una desviación estándar de 4.5 minutos. Se asume que los tiempos de entrega siguen una distribución aproximadamente normal.

   (a) Explique cómo se construiría un **intervalo de confianza del 95%** para la media poblacional del tiempo de entrega, indicando los pasos y elementos necesarios. 1p

   (b) Plantee el **test de hipótesis** para evaluar si el tiempo promedio de entrega actual es **menor que 52 min**. Utilice un nivel de significancia $\alpha = 0.05$ y detalle claramente cada paso. 1p

   (c) Explique con un ejemplo qué representa el **error de tipo II** en este contexto e interprete sus consecuencias prácticas para la empresa. 0.5p

$T = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$

$t_{(24,\ 1-\alpha)} = t_{tabla}$

IC $95\% = (1-\alpha) \cdot 100$

$0{,}95 = 1-\alpha$

$\alpha = 1 - 0{,}95$

$\alpha = 0{,}05$

Datos

$\alpha = 0{,}05$ N.S.

$\mu_0 = 52$ min

$\bar{x} = 49{,}8$ min

$n = 25$

$S = 4{,}5$ min

. Estadístico a usar: $t_{n-1,\ 1-\alpha/2}$

grados de libertad

$X_i \sim N(\mu, \sigma^2)$

Porque la muestra aleatoria $X_i \sim N(52 \text{ min}, \sigma^2)$ con σ desconocida

$\Rightarrow \bar{x} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$ y $T = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

$$1-\alpha = P\left(-t_{n-1,1-\frac{\alpha}{2}} < T < t_{n-1,1-\frac{\alpha}{2}}\right)$$

$$1-\alpha = P\left(-t_{n-1,1-\frac{\alpha}{2}} < \frac{\bar{x}-\mu}{S/\sqrt{n}} < t_{n-1,1-\frac{\alpha}{2}}\right)$$

$$1-\alpha = P\left(-t \cdot \frac{S}{\sqrt{n}} < \bar{x} - \mu < t \cdot \frac{S}{\sqrt{n}}\right)$$

$$1-\alpha = P\left(-\bar{x} - t \cdot \frac{S}{\sqrt{n}} < -\mu < -\bar{x} + t \cdot \frac{S}{\sqrt{n}}\right)$$

$$1-\alpha = P\left(\bar{x} + t \cdot \frac{S}{\sqrt{n}} > \mu > \bar{x} - t \cdot \frac{S}{\sqrt{n}}\right)$$

$\therefore (\hat{\theta}_i, \hat{\theta}_s) = \theta \pm \mathcal{E}$

$\Rightarrow$ I.C. = Estimador $\pm$ Valor Crítico . Error Estimado

$$\text{I.C.} = \bar{x} \pm t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

y para construir el intervalo:

$$\mu \in \left(\bar{x} - t_{n-1,\,1-\alpha/2} \cdot \frac{S}{\sqrt{n}} \,,\quad \bar{x} + t_{n-1,\,1-\alpha/2} \cdot \frac{S}{\sqrt{n}} \right)$$

con un N.C. del 95%

b) Datos

$X \sim N(52, \sigma^2)$

$\alpha = 0{,}05$ $\quad\quad \bar{X} = 49{,}8 \text{ min}$

$\mu_0 = 52 \text{ min}$ ← param. nulo

μ ← param. de interés

Paso a Paso

1) Definir el N.S. α

$\alpha = 0{,}05$

2) Identificar el parámetro de interes

μ

3) Definir param nulo y proponer hip. nula.

$H_0: \mu = \mu_0$

$\mu = 52 \text{ min}$

4) Plantear hip. alternativa

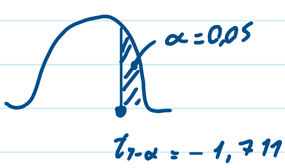
$H_1: \mu < \mu_0$

$\mu < 52 \text{ min}$

5) Identificar Est. de Prueba

$$t_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{\underbrace{n-1}_{24}}$$

6) Establecer R.R.



$R.R. = \{ t \,/\, t \leq -1{,}711 \}$

7 Calcular cantidades necesarias

$$t_0 = \frac{49,8 - 52}{4,5 / \sqrt{25}}$$

$t_0 = -2,4444$

$t \leq -t_{\alpha, n-1}$

$-2,4444 \leq -1,711$ ✓

8 Se toma una decisión (se rechaza o no H_0)

$t_0 \in R.R. \Rightarrow$ Se rechaza H_0 en favor de H_1

La muestra presenta evidencia suficiente como para rechazar la hipótesis nula, con un N.S. del 5%. a favor de la alternativa que indica que el tiempo promedio de entrega actual es menor a 52 min.

c) Error Tipo 1

Ver algo que no hay (F. +)

P(Se rechaza H_0 / H_0 es V)

E.T. 1 = α

$H_0 : \mu : \mu_0$
$H_1 : \mu < \mu_0$ F

Error Tipo 2

No ver algo que SI hay (F. -)

P(No se rechaza H_0 / H_0 es F)

E.T. 2 = β

$H_0 : \mu : \mu_0$
$H_1 : \mu < \mu_0$ V

En este caso el error de Tipo 2 representaría el no reconocer la mejoría en el tiempo de entrega, puesto que este disminuyó en comparación al promedio histórico

Las consecuencias que esto podría traer a la empresa son:

+ no se reconoce la mejoría de x

+ no se premian o mantienen estrategias que lograron x

+ Podrían revertirse cambios que si estaban funcionando

2. Enuncie **detalladamente** la ley de la Probabilidad Total. Ejemplifique en un contexto diferente al de la gasolinera. 1p

Para enunciar la ley de la Prob. Total, primero debemos saber lo que es una part. aleat.

=> Sea $(S, \mathcal{A}, P)$ un espacio de probabilidad y una colección de eventos $A_i \in S$, es una part. aleat. de S si se cumple:

① $A_i \cap A_j = \emptyset, \ \forall i \neq j$

② $\bigcup_{i=1}^{n} A_i = S$

Suponiendo que $\{A_i\}_{i=1}^{n} \subseteq S$ es una partición aleatoria de S y sea B un evento de S, entonces se cumple que:

$$P(B) = \sum_{i=1}^{n} P(B/A_i) \cdot P(A_i)$$

Ley de la Probabilidad Total

Bayes: Sea $\{A_i\}_{i=1}^{n} \subseteq S$ una partición aleatoria de S

y $P(A_i) > 0$ con $i = 1, \dots, n$ $\forall A \in \mathcal{A}$

Y sea $B \subseteq S$, con $P(B) > 0$

Entonces, se cumple lo siguiente:

Def Cond. Def Intersección

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B/A_i) \cdot P(A_i)}{\sum_{j=1}^{n} P(B/A_j) \cdot P(A_j)} \rightarrow \text{P. Priori}$$

P. Posteriori

P. Condicional

Prob Total

Indep. si: $P(A \cap B) = P(A) \cdot P(B)$

Indep. si : $P(A \cap B) = P(A) \cdot P(B)$

18/12/2025

viernes, 6 de febrero de 2026 16:14

Examen de Probabilidad y Estadistica 18/12/2025

1) Se mide el tamaño en MB de la carga de archivos por usuarios.
Media: 58 Mediana: 32 Moda: 12 Rango: 308 Desvio Estandar: 42
a) Que indican las medidas de tendencia central, con estos datos
b) Rango y desvio estándar.
c) Para que sirven las medidas de variabilidad asociadas a las medidas de tendencia central.

2) Defina espacio muestral. De un ejemplo de continuo y discreto.

3) Defina variable aleatoria. De un ejemplo de variable aleatoria bidimensional.

4) Explique la relación entre binomial y poisson.

5) De la formula de intervalo de confianza, sabiendo que tiene distribución normal y varianza conocida.

6) Definir los pasos de test de hipótesis de acuerdo al problema y dar una conclusión. El nivel de significancia era 0.05 y p-valor 0,035

7) Dar la definición de probabilidad condicional, demostrar que P(A/B) + P(A'/B) = 1

8) Dar la definición de eventos independientes. Demostrar que P(A) > 0 P(A') > 0 NO son independientes.

1) Datos

$\bar{x} = 58$ $M_{ed} = 32$ $M_o = 12$ $R = 308$ $\sigma = 42$

a) Con estos datos las medidas de tendencia central indican lo siguiente:
   - El tamaño promedio de los archivos es 58 MB
   - La Mediana indica donde está el centro de la distribución si la misma está ordenada, es decir, que el 50% de los datos es menor o igual a 32 MB y el otro 50% es mayor o igual a 32 MB.
   - La moda nos indica que el tamaño de archivo cargado más observado es 12MB
b) El Rango nos indica que entre el dato más pequeño (el de menor peso en MB) y el más grande (mayor peso en MB), hay una diferencia de 308 MB

   El desvío estándar es de 42 MB, lo que nos indica que nuestra media no es representativa, puesto que los datos están en promedio alejados en 42MB de la media.

c) Las medidas de variabilidad asociadas a las medidas de tendencia central sirven para ver que tan representativas son estas últimas, puesto que estas analizan que tan heterogéneos son los datos. Por ejemplo en este caso, las MTC no son representativas puesto que los datos están muy alejados de la media, calculando el coeficiente de variación tenemos que los datos varían en un 72,41% y cuanto mayor es este más dispersos son los datos.

$$CV = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{42}{58} \cdot 100 = 72,41\%$$

2) Defina espacio muestral. De un ejemplo de continuo y discreto.

   El espacio muestral es el conjunto S donde se encuentran todos los resultados posibles de un experimento aleatorio.

   Este puede ser discreto o continuo:
   - Discreto: Está conformado por elementos finita o infinitamente numerables. Ejemplo: Se lanza un dado S={1, 2, 3, 4, 5, 6}; o S={Cantidad de materias aprobadas}
   - Continuo: Está conformado o contiene al menos un intervalo de números reales. Ejemplo: Se realiza la medición de el voltaje de una batería de un servidor S=[12.4v, 15v]

3) Defina variable aleatoria. De un ejemplo de variable aleatoria bidimensional.

   Una v.a. es toda función X que asigna a cada uno de los elementos de S un valor en R. Esto es $X: S \to R$. Además:

$$\forall x \in \mathbb{R}, \quad \{\omega \in S : X(\omega) \leq x\} \in \mathcal{A}$$

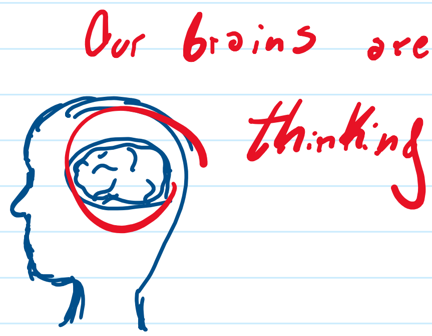
$$F(x) = P(X \leq x) \quad \wedge \quad F(x) : \mathbb{R} \to [0, 1]$$

Sean X la v.a. que mide la cantidad de solicitudes exitosas al servidor, e Y la v.a. que mide la cantidad de solicitudes fallidas al servidor.

Sabiendo que solo hay 2 solicitudes, y que la probabilidad de que la solicitud sea exitosa es 0.8, entonces:

X \ Y	0	1	2
0	0	0	0,04
1	0	0,32	0
2	0,64	0	0

X	Y
0	2
1	1
2	0



$$P(X=2, Y=0) = 0,8 \cdot 0,8 = 0,64$$

$$P(X=1, Y=1) = 2 \cdot 0,8 \cdot 0,2 = 0,32$$

$$P(X=0, Y=2) = 0,2 \cdot 0,2 = 0,04$$

$$1$$

4) Explique la relación entre binomial y poisson.
   La distribución de poisson está modelada sobre eventos muy raros de Binomial, por lo tanto la relación entre estas se dá cuando n tiende a infinito y p tiende a 0.

   Como regla práctica, esta puede aplicarse cuando n>100, p<0.05 y n*p <20

   La relación está dada entonces como una distribución de Poisson de parámetro (n*p):

$$S: \quad X \sim \mathcal{B}(n, p) \quad \wedge \quad n > 100, \quad p < 0,05, \quad n.p < 20$$

$$\Rightarrow X \approx \mathcal{P}(n.p)$$

$$P(X = x) = \frac{(n.p)^x}{x!} e^{-np}$$

5) De la formula de intervalo de confianza, sabiendo que tiene distribución normal y varianza conocida.

   Como tenemos una distribución normal y varianza conocida, podemos estimar la media utilizando z.

$$1-\alpha = \left(-z_{\alpha/2} \leq Z_1 \leq z_{\alpha/2}\right)$$

$$\vdots$$

$$\Rightarrow 1-\alpha = \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

6) Definir los pasos de test de hipótesis de acuerdo al problema y dar una conclusión. El nivel de significancia era 0.05 y p-valor 0,035

Datos

$\alpha = 0,05$ $\qquad$ p-valor $= 0,035$

Al ser p-valor el menor nivel de significancia posible para el cual se debe rechazar H0, podemos a partir de este tomar una decisión.

p-valor $\leq \alpha$ Rechazo H_0

p-valor $> \alpha$ No rechazo H_0

En este caso p-valor <= alpha, entonces hay evidencia suficiente para rechazar H0

De igual forma, la consigna pide los pasos del test de hipotesis, los cuales son:

1. Identificar α
2. Identificar parámetro de interés
3. Definir el parámetro nulo y proponer H_0
4. Proponer H_1
5. Identificar el estadístico de prueba adecuado
6. Establecer la región de rechazo
7. Calcular las cant. necesarias
8. Emitir conclusiones

7) Dar la definición de probabilidad condicional, demostrar que P(A/B) + P(A'/B) = 1

Sea $(S, \mathcal{A}, P)$ un espacio de probabilidad, y dos eventos

A y $B \in \mathcal{A}$, con $P(B) > 0$

$\Rightarrow P(A/B)$ es llamada prob. condicional y se define:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Además, si $P(B) = 0$

$P(A/B) = P(A)$

- Demostrar $P(A/B) + P(A'/B) = 1$

$P(A/B) + P(A'/B) =$ Def. de Prob Cond.

$\frac{P(A \cap B)}{P(B)} + \frac{P(A' \cap B)}{P(B} =$ Denominador común

$\frac{1}{P(B)} \cdot (P(A \cap B) + P(A' \cap B)) =$ Prop. de $\mathcal{A}$

$\frac{1}{P(B)} \cdot (\cancel{P(A \cap B)} + \underbrace{P(B) - \cancel{P(A \cap B)}}) =$ Resolvemos

$$\boxed{\frac{P(B)}{P(B} = 1}$$

- Queda demostrado entonces que $\boxed{P(A/B) + P(A'/B) = 1}$

8) Dar la definición de eventos independientes. Demostrar que P(A) > 0 P(A') > 0 NO son independientes.

Independencia de Eventos

Sea $(S, \mathcal{A}, P)$ un E.P. y dos eventos A y $B \in \mathcal{A}$, estos serán independientes cuando:

$$P(A \cap B) = P(A) \cdot P(B)$$

Obs: $P(A/B) = P(A) \iff P(B/A) = P(B)$

Demostrar que si $P(A) > 0$ y $P(A') > 0$ no son indep.

$$P(A \cap A') = P(A) \cdot P(A')$$

$$P(\varnothing) = P(A) \cdot P(A')$$

$0 \neq P(A) \cdot P(A')$ • Como $P(A) > 0$ y $P(A') > 0$, es imposible que de 0

Como la igualdad no se cumple, no son independientes lunes, 16 de febrero de 2026 18:55

$$\frac{k \cdot n}{NoC} = \frac{\cancel{2} \cdot n}{\cancel{4}} = \frac{1}{2} \cdot n = \frac{n}{2}$$

$$L_{inf} + \frac{I_{ck} - F_{Md-1}}{f_{Med}} \cdot A = 40 + \frac{50 - 45}{20} \cdot 10$$

$$= 40 + \frac{5\cancel{0}}{2\cancel{0}}$$

$$= 40 + 2,5 = 42,5$$

$$CV = \frac{S}{\bar{X}} \cdot 100 \rightarrow \underline{\text{Grupo A}}$$

$$\frac{7}{70} \cdot 100 = 10\%$$

$$\underline{\text{Grupo B}}$$

$$\frac{12}{160} \cdot 100 = 7,5\%$$

En una distribución que se aproxima a la Normal, con $\overline{X} = 100$ y $S = 15$, ¿qué porcentaje aproximado de datos esperarías encontrar entre los valores 70 y 130?

Regla del 68, 95, 99

68, $\bar{X} \pm S$

$95, \; \bar{x} \pm 2s$

$99{,}73 \; \bar{x} \pm 3s$

X: v.a. que cuenta la cant. defectuosos

$$X \sim h(\overset{N}{5}, \overset{n_1}{2}, \overset{n}{2})$$

$$P(X = r) = \frac{\binom{n_1}{r}\binom{N-n_1}{n-r}}{\binom{N}{n}}$$

$$P(X=2) = \frac{\binom{2}{2}\binom{5-2}{2-2}}{\binom{5}{2}} =$$

$$\frac{1 \cdot 1}{10} = 0{,}1$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0{,}2}{0{,}5} = 0{,}4$$

$$P(A'/B) = \frac{1 - P(A \cap B)}{P(B)} = \frac{1-0,2}{0,5} = \frac{0,8}{0,5} =$$

Prob. Total

$$\sum_{j=1}^{n} P(B/A_j) \cdot P(A_j)$$

M_1: 50% $\quad P(M_1) = 0,5 \quad P(F_{M_1}) = 0,25$

M_2: 30% $\quad P(M_2) = 0,3 \quad P(F_{M_2}) = 0,2$

M_3: 20% $\quad P(M_3) = 0,2 \quad P(F_{M_3}) = 0,1$

$$0,25 \cdot 0,5 + 0,3 \cdot 0,2 + 0,2 \cdot 0,1 = P(B)$$
$$= 0,205$$

Bayes

5. En el caso anterior de los equipos de audio, si un cliente trae un equipo a reparar, ¿cuál es la probabilidad de que sea de la marca M_2?

$$\frac{P(B/M_2) \cdot P(M_2)}{P(B)} = \frac{0,3 \cdot 0,2}{0,205} = 0,29268 \approx 0,2927$$

7. Se extrae una moneda de una caja con 2 monedas honestas y 1 moneda con dos caras. Se lanza y sale 'Cara'. ¿Cuál es la probabilidad de que la moneda extraída sea la deshonesta?

$P(H) = 2/3$ $\qquad$ $P(C/H) = (1/2)$

$P(D) = 1/3$ $\qquad$ $P(C/D) = 1$

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B/A_i) \cdot P(A_i)}{\sum_{j=1}^{n} P(B/A_j) \cdot P(A_j)} =$$

$$= \frac{1 \cdot 1/3}{(2/3 \cdot 1/2) + (1/3 \cdot 1)} = \frac{1/3}{2/3} = 3/6 = 0{,}5$$

$P(A_1) = 2/4 \quad \{e_1, e_2\}$ $\qquad$ $P(A_1 \cap A_2) = 1/4$

$P(A_2) = 2/4 \quad \{e_1, e_3\}$

$P(A_3) = 2/4 \quad \{e_1, e_4\}$ $\qquad$ $= P(A_1) \cdot P(A_2) = 2/4 \cdot 2/4 = 4/16 = 1/4$

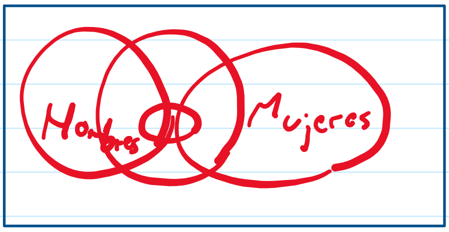
$1/4 \neq 1/8$ no son ind. en su totalidad.

17/2/25

martes, 17 de febrero de 2026 18:57

$P(S/M) \cdot P(M) = 0{,}18$

5. De acuerdo con los registros de un médico clínico, el 60% de sus pacientes son mujeres; el 30% de sus pacientes son mujeres con sobrepeso y el 10% de las pacientes con sobrepeso tienen hipotiroidismo. Se elige al azar un paciente de este médico. Calcular la probabilidad de que sea mujer, tenga sobrepeso y padezca de hipotiroidismo.



$P(M) = 0{,}6$ $\qquad P(S \cap M) = 0{,}3$

$P(H/(S \cap M)) = 0{,}1$

$$P(M \cap S \cap H) = \underline{P(M) \cdot P(S/M)} \cdot P(H/(S \cap M))$$

$$P(S \cap M) \cdot P(H/(S \cap M))$$

$$0{,}3 \cdot 0{,}1 = 0{,}03$$

11. De acuerdo con el registro médico del Ejemplo 4 (pág. 12), el 60% de los pacientes son mujeres, el 30% de las mujeres tienen sobrepeso y el 10% de ellas tiene hipotiroidismo. Si se elige un paciente al azar, ¿cuál es la probabilidad de que sea mujer, tenga sobrepeso y padezca hipotiroidismo?

$P(M) = 0{,}6$ $\qquad P(S/M) = 0{,}3$

$P(H/(S \cap M)) = 0{,}1$

$$\frac{P(S \cap M)}{(M)} = 0{,}3$$

$$P(S \cap M) = 0{,}3 \cdot 0{,}6$$

$$P(S \cap M) = 0{,}18$$

$$P(M \cap S \cap H) = P(M) \cdot P(S/M) \cdot P(H/S \cap M)$$

$$0{,}18 \cdot 0{,}1 = 0{,}018$$

Cond

$$\underline{P(A_i/B)} = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B/A_i) \cdot P(A_i)}{\sum P(B/A_i) \cdot P(A_i)}$$

Priori

$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B/A_i) \cdot P(A_i)}{\sum_{j=0}^{n} (B/A_j) \cdot A_j}$ ← Priori

Posteriori

Se tiene una v.a. discreta X con valores $\{1, 2\}$ y probabilidades $P(X = 1) = 0.3$ y $P(X = 2) = 0.7$. ¿Cuál es su Esperanza $E(X)$?

$E(x) = \sum_{i=0}^{n} x_i \cdot p(x_i)$

6. Si conocemos la Esperanza $E(X) = \mu$ y la Varianza $Var(X) = \sigma^2$, ¿cuál es la varianza de la variable transformada $Y = aX + b$?

A. $a\sigma + b$

B. $a^2\sigma^2$

✓ ¡Exacto!
Por propiedad de la varianza, las constantes sumadas (b) no afectan la dispersión, y las constantes multiplicativas (a) salen al cuadrado.

$P(|X - \mu| \geq t\sigma) \leq 1/t^2$

9. La Desigualdad de Chebyshev establece que $P(|X - \mu| \geq t\sigma) \leq 1/t^2$. Si elegimos $t = 2$, ¿qué porcentaje máximo de datos puede estar fuera de los 2 desvíos estándar?

A. 75%

B. 25%

✓ ¡Exacto!
Para $t = 2$, el límite es $1/2^2 = 1/4 = 0.25$. Esto significa que como máximo el 25% de los datos están alejados más de 2σ de la media.

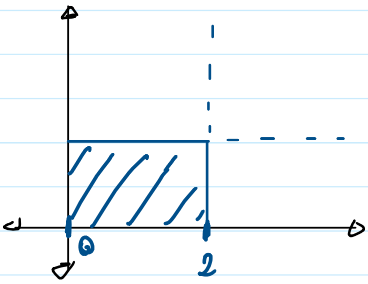
10. Al realizar la normalización o estandarización de una v.a. X mediante la transformación $Z = (X - \mu)/\sigma$, ¿cuáles son la esperanza y varianza de la nueva variable Z?

A. $E(Z) = 0$ y $Var(Z) = 1$

✓ ¡Exacto!

Esta transformación centra la variable en el cero y escala su dispersión para que el desvío estándar sea la unidad.

11. Dada una v.a. continua X con función de densidad $f(x) = kx$ para $0 \leq x \leq 2$ (y cero en otro caso), ¿cuál debe ser el valor de la constante k para que sea una función válida?



D. 1/2

✓ Respuesta correcta

Para que sea una densidad válida, la integral en todo el rango debe ser 1: $\int_0^2 kxdx = [k\frac{x^2}{2}]_0^2 = 2k = 1 \implies k = 1/2$.

$$\boxed{\int_{-\infty}^{\infty} f(x).dx = 1} \Rightarrow \int_0^2 k\,x\,dx = \left[k\,\frac{x^2}{2}\right]_0^2 = k\,\frac{\cancel{4}^{\,2}}{\cancel{2}_{\,1}} = 2k$$

$$\Rightarrow 2k = 1$$
$$k = 1/2$$

$$F(x) \begin{cases} 0 & x < 2 \\ 0,5 & 2 \leq x < 7 \\ 1 & x \geq 7 \end{cases}$$

12. Sea X una v.a. discreta con la siguiente función de distribución acumulada: $F(x) = 0$ si $x < 2$, $F(x) = 0.5$ si $2 \leq x < 7$, y $F(x) = 1$ si $x \geq 7$. ¿Cuál es el valor de $P(X = 2)$?

13. Para una v.a. continua con densidad $f(x) = \frac{e^{-x/1000}}{1000}$ para $x > 0$, ¿cuál es la probabilidad de que la variable tome un valor mayor a 1000?

$$P(x > 1000) = 1 - P(x \leq 1000) = 1 - F(1000)$$

$$= 1 - \int_{0}^{1000} f(t) \, dt$$

$$\frac{e^{-x/1000}}{1000}$$

$$= 1 - \left[-\frac{1}{e} + 1 \right]$$

$$= 1 + \frac{1}{e} - 1$$

$$= 1/e = e^{-1}$$

15. Si una v.a. discreta X tiene valores $\{1, 2, 3\}$ con probabilidades $\{0.2, 0.5, 0.3\}$, ¿cuál es el valor de $E(X^2)$?

$$\underbrace{0,2}_{p} \cdot \underbrace{1}_{X}^{2} + 0,5 \cdot 2^{2} + 0,3 \cdot 3^{2} = 4,9$$

Usando la Desigualdad de Chebyshev, si queremos asegurar que la probabilidad de estar fuera del intervalo $(\mu - t\sigma, \mu + t\sigma)$ sea a lo sumo del 10%, ¿cuál es el valor mínimo de t?

Desigualdad Chebyshev

$$P(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2} \Rightarrow \frac{1}{\sqrt{10}^{2}} = 0,1$$

Estandarizar una V.a.

$$Z = \frac{(X-\mu)}{\sigma}$$

Si una v.a. X tiene media $\mu = 10$ y desvío $\sigma = 2$, y se obtiene un valor estandarizado $Z = 1.5$, ¿cuál era el valor original de X?

$$1,5 = \frac{X - 10}{2} \Rightarrow 3 + 10 = X$$

$$13 = X$$

$$\int_0^1 x^2 \cdot (2 \cdot x) = 2 \int_0^1 x^3$$

$$= 2 \cdot \left[\frac{x^4}{4} \right]_0^1$$

$$= 2 \cdot 1/4 = 1/2$$

¿Cuál es la varianza de $Z = X + Y$ si sabemos que $Var(X) = 5$, $Var(Y) = 5$ y la correlación es $\rho_{XY} = 0.5$?

$$\rho_{xy} = \frac{Cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}} = \frac{Cov(X,Y)}{\sqrt{5} \cdot \sqrt{5}} = 0,5$$

$$Cov(X,Y) = 0,5 \cdot (\cancel{\sqrt{5}})^{\cancel{2}}$$

$$Cov(X,Y) = 2,5$$

$Var(Z) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$

$$1^2 \cdot 5 + 1^2\, 5 + 2 \cdot 1 \cdot 1 \cdot 2,5 = 15 \checkmark$$

35. En la esperanza de una función de dos variables $E[g(X,Y)]$, ¿qué función $g(x,y)$ se utiliza para calcular la Covarianza directamente?

D. $(x-\mu_X)(y-\mu_Y)$

✓ ¡Exacto!

Por definición, la covarianza es la esperanza del producto de los desvíos de cada variable respecto a su propia media.

36. Si $f(x,y)=e^{-(x+y)}$ para $x,y>0$. ¿Son X e Y independientes?

A. No, porque ambas variables están en el exponente

B. Sí, porque se puede escribir como $e^{-x}\cdot e^{-y}$

✓ ¡Exacto!

La densidad conjunta es el producto de dos funciones que dependen solo de una variable cada una, lo que satisface la definición de independencia.

37. Se define la Esperanza Condicional $E(Y/X=x)$. ¿Qué tipo de objeto matemático es el resultado?

C. Una función que depende de x

✓ ¡Exacto!

Al fijar un valor de X, calculamos un promedio de Y. Si cambiamos el valor de x, el promedio puede cambiar, definiendo una función.

D. $kg\cdot cm$

✓ ¡Exacto!

La covarianza es la esperanza del producto de los desvíos, por lo que su unidad es el producto de las unidades originales.

39. En el Ejemplo 6c (pág. 27), se muestra una tabla donde $Cov(X,Y)=0$. ¿Por qué se concluye que NO son independientes?

B. Porque $P(X = x, Y = y) \neq P(X = x)P(Y = y)$ para algún punto

✓ ¡Exacto!

Basta con encontrar una sola celda donde no se cumpla la factorización para descartar la independencia, aunque la relación lineal sea nula.

40. ¿Cuál es el valor de $P(X \leq 2, Y > 10)$ expresado en términos de la Función de Distribución Conjunta $F(x, y)$?

B. $F_X(2) - F(2, 10)$

✓ Respuesta correcta

Representa toda la franja hasta $X = 2$ ($F_X(2)$) menos la parte donde $Y \leq 10$ ($F(2, 10)$), dejando la zona donde $Y > 10$.

Si $X \sim \mathcal{P}(\lambda)$ (Poisson), ¿cuál es la relación entre su Esperanza y su Varianza?

D. $E(X) = Var(X) = \lambda$

✓ ¡Exacto!

Es una propiedad única del modelo de Poisson: la media y la varianza coinciden numéricamente con el parámetro de la tasa.

20/02

viernes, 20 de febrero de 2026 13:15

12. Un científico inocula ratones con un virus ($p = 1/6$) hasta que **tres** contraigan la enfermedad. **¿Cuál es la expresión correcta para calcular** la probabilidad de que necesite exactamente 8 ratones?

$\binom{7}{2}(1/6)^3(5/6)^5$

¡Exacto!

Usamos Binomial Negativa con $r = 3$ y $x = 8$. La fórmula es $\binom{x-1}{r-1}p^r(1-p)^{x-r}$, resultando en $\binom{7}{2}(1/6)^3(5/6)^5$.

Si un cajero recibe en promedio 25 personas por hora, ¿cuál es la probabilidad de que concurran **exactamente 20** personas en un lapso de **2 horas**?

$X \sim P(\lambda)$

$X \sim P(50)$

$\lambda = \mu \cdot \Delta t$

$\lambda = 25 \cdot 2$

$\lambda = 50$ personas

$$P(X=r) = \frac{\lambda^r}{r!} \cdot e^{-\lambda}, \quad r \in \mathbb{N}_0$$

$E(X) = Var(x) = \lambda$

$$P(X=20) = \frac{50^{20}}{20!} \cdot e^{-50}$$

$P(X=20) = 7{,}56 \cdot 10^{-7}$

En un vector aleatorio con distribución Multinomial $(X_1, \ldots, X_k)$, ¿cuál es el valor de la Covarianza entre dos resultados distintos $Cov(X_i, X_j)$?

C. $-np_ip_j$

¡Exacto!

Debido a que el número total de ensayos n es fijo, si un resultado ocurre más veces, los otros deben ocurrir menos, generando una correlación negativa.

Tenés 1000 plantines (400 infectados). Elegís 10 al azar. ¿Por qué es aceptable usar una distribución Binomial para aproximar este problema en lugar de la Hipergeométrica?

Porque el cociente $n/N = 0.01$ es menor a 0.1

¡Exacto!

Cuando la muestra es muy pequeña respecto a la población, el hecho de no reponer los elementos no altera significativamente las probabilidades de cada extracción.

16. Dada una distribución Binomial con $n = 400$ y $p = 0.005$, ¿cuál es el valor de λ para realizar la aproximación por Poisson?

A. 20

B. 2

¡Exacto!

El parámetro de Poisson se obtiene multiplicando $\lambda = n \cdot p = 400 \cdot 0.005 = 2$.

En el modelo Hipergeométrico, ¿qué representa el término $\frac{N-n}{N-1}$ en la fórmula de la varianza?

Factor de corrección por población finita

¡Exacto!

Este término reduce la varianza respecto al modelo binomial debido a que el muestreo sin reposición de una población finita disminuye la incertidumbre.

TCL

sábado, 21 de febrero de 2026 15:47

$$\boxed{\frac{m^4}{S^4} - 3}$$ Kurtosis

$$\boxed{\frac{\bar{X} - Med}{S} \cdot 3}$$ c.A. Pearson

$\bar{X} < Med < Mo$ As Neg

$CV_1 > CV_2 \Rightarrow$