MATD234/2025-A

Probabilidad

Y

ESTADÍSTICA

25/04/2025

Presentación DE LOS Datos

Muestra: Tamaño n
Datos: $x_1, x_2, \dots, x_n$
Presentación: Tabla de frecuencia

Definiciones:

- Frecuencia (absoluta) del dato x_i: $f(x_i)$ es el número de veces que aparece x_i en la muestra.
- Frecuencia acumulada: $F(x_i)$ es el número de datos que son menores o iguales a x_i.
- Frecuencia relativa: $fr(x_i) = f(x_i)/n$
- Frecuencia relativa acumulad: $Fr(x_i) = F(x_i)/n$

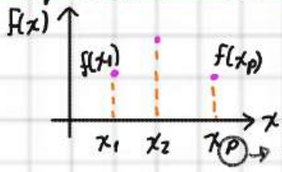
ejemplo: X: Edad (años) de los estudiantes de la EPN (grado)
Tamaño: $n=30$

Datos: 17, 18, 20, 24, 22, 20, 21, 21, 23, 25, 20, 19, 18, 20, 22, 21, 23, 18, 17, 20, 22, 21, 24, 22, 20, 21, 19, 18, 20, 23

Tabla de frecuencias:

x	$f(x)$	$F(x)$	$fr(x)$	$Fr(x)$
17	2	2	0,067	0,067
18	4	6	0,133	0,20
19	2	8	0,067	0,267
20	7	15	0,233	0,50
21	5	20	0,1667	0,667
22	4	24	0,133	0,80
23	3	27	0,10	0,90
24	2	29	0,067	0,967
25	1	30	0,033	1

- Representación gráfica:



se utiliza "p" porque tomamos la cantidad de datos distintos

Esta presentación de datos se denomina "de datos individuales" y se recomienda cuando la muestra es relativamente pequeña.
Cuando la muestra es grande, se puede presentar los datos de forma agrupada. A los datos se los agrupa en intervalos llamados intervalos de clase.

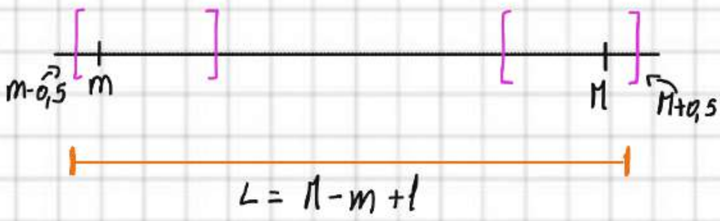
Existen varios métodos para construir intervalos. Una característica de estos es que los intervalos deben ser de igual longitud.

Un método para agrupar datos es el siguiente:

Supuestos:

1.- Suponemos que los datos son enteros, si no lo son, los transformamos en enteros con potencias de 10. (1,63 → 163)
2.- Longitud entera.
3.- El punto medio debe ser entero, pues este será el representante de todos los datos en el intervalo.
4.- Ningún dato debe coincidir con un extremo de intervalo.

Metodología: m: mínimo M: máximo



Para que se cumpla la cuarta consideración, hagamos que los extremos de intervalo tengan 0,5 de longitud decimal.

- **r**: número de intervalos que se va a usar. Se sugiere que se usen entre 7 y 13 intervalos.
- **w**: longitud de los intervalos. $w = \lceil \frac{L}{r} \rceil$ → entero inmediato superior.
- **e**: valor del exceso. $e = wr - L$

Para garantizar que el punto medio sea entero, la longitud debe ser impar.

Entre las alternativas factibles, se puede escoger la que tenga menor exceso, y se lo reparte "equitativamente" entre el primer y el único intervalo.
↳ si el exceso es 2, se reparte 1-1; si el exceso es 3, se reparte de forma 1-2 o 2-1.

ejemplo: X: Nota Prueba de ingreso a la U. (sobre 1000)

muestra: $n=180$: n grande ⇒ datos agrupados

Datos: $x_1, x_2, \dots, x_{180}$

mínimo: $m=638$ máximo: $M=987$

$$L = M - m + 1 = 987 - 638 + 1 = 350$$

- si $r=7$: $w=\lceil \frac{L}{r} \rceil = 50$; $e=0$ ✗ longitud par
- si $r=8$: $w=\lceil \frac{L}{r} \rceil = 44$; $e=2$ ✗ longitud par
- si $r=9$ $w=\lceil \frac{L}{r} \rceil = 39$; $e=1$
- si $r=10$ $w=\lceil \frac{L}{r} \rceil = 35$; $e=0$ → Elegida
- si $r=11$ $w=\lceil \frac{L}{r} \rceil = 32$; $e=2$ ✗ longitud par
- si $r=12$ $w=\lceil \frac{L}{r} \rceil = 30$; $e=10$ ✗ longitud par
- si $r=13$ $w=\lceil \frac{L}{r} \rceil = 27$; $e=1$

Escogemos $r=10$ intervalos con longitud $w=35$

El primer intervalo debe empezar en $m-0,5=637,5$

#	Intervalo	Punto Medio	$f(i)$	$F(i)$	$fr(i)$	$Fr(i)$
1	[637,5 ; 672,5]	655				
2	[672,5 ; 707,5]	690				
3	[707,5 ; 742,5]	725				
4	[742,5 ; 777,5]	760				
5	[777,5 ; 812,5]	795				
6	.	830				
7	.	865				
8	.	900				
9		935				
10		970				

Cuando trabajamos con intervalos, se considera que la variable es continua, es decir, si se puede unir con líneas en el gráfico

- Histograma de frecuencias:

Las medidas estadísticas sirven para describir lo que ocurre en la muestra.

Observación: Si al construir los intervalos, ninguna de las alternativas, se pueden analizar otras opiones (5, 6, 14, 15).

Medidas Estadísticas

- Medidas de tendencia central:

Media: es el promedio aritmético de todos los datos o el promedio ponderado

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{x} = \frac{1}{n}\sum_{j=1}^{r} x_j\, f(x_j)$$

↳ promedio aritmético — promedio ponderado

ejemplo: (edad) $n=30$ Datos: 17, 17, 18, 18, . . ., 24, 24, 25

$$\bar{x} = \frac{17+17+\dots+25}{n}$$
$$= \frac{2(17)+4(18)+\dots+2(24)+25}{n}$$

ejemplo: Nota de aprobación en el saew

Materia	# de créditos	Nota
Fund.	3	—
Cál. Dif.	4	—
Alg. lin.	5	32
Pen. Cient	1	38
Tic's	1	40
Fund.	3	28
Calc. Dif.	4	34

Nota de aprobación:

$$\frac{5(32)+38+40+3(28)+4(34)}{5+1+1+3+4}$$

$\to$ 32, 71

28/04/2025

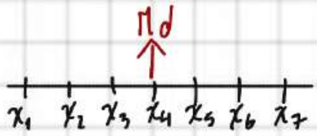
X: Variable   Muestra: Tamaño n Datos: $x_1, x_2, \dots, x_n$

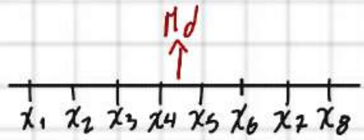
ejemplo: X: edad (años)

$\bar{x} = 20{,}63$ (años)

La media va a ser el representante de todos los datos.

Mediana: es un valor, el cual va a ser mayor o igual que el 50% de los datos. Se representa como Md.



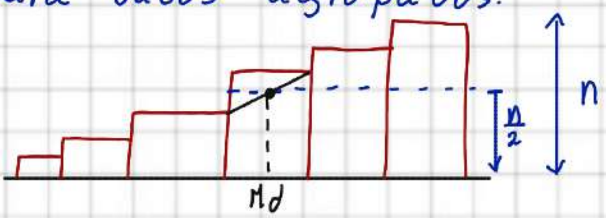


Para datos individuales:

Si n es impar: $Md = \frac{x_{n+1}}{2}$

Si n es par: $Md = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$

Para datos agrupados:



$$Md = L + \frac{\frac{n}{2} - F}{f}\,\omega$$ donde:

L es el límite inferior del intervalo mediana y es aquel cuya frecuencia acumulada es la primera que es mayor o igual a $\frac{n}{2}$.

F es la frecuencia acumulada hasta L.

f es la frecuencia del intervalo

ω es la longitud del intervalo.

ejemplo:

#	intervalo	Punto medio	f(i)	F(i)
1	99,5 - 104,5	102	3	3
2	104,5 - 109,5	107	7	10
3	109,5 - 114,5	112	13	23
4	114,5 - 119,5	117	19	42
5	119,5 - 124,5	122	12	54
6	124,5 - 129,5	127	8	62
7	129,5 - 134,5	132	4	66

$\bar{x} = 117,30$

$Md = ?$

$n = 66 \Rightarrow \frac{n}{2} = 33 \rightarrow$ la mediana está en el cuarto intervalo

$\Rightarrow L = 114,5$

$F = 23$

$f = 19$

$w = 5$

$$Md = 114,5 + \frac{33-23}{19} \cdot 5$$
$$= 117,13$$

Cuando la media no es un buen representante de la muestra, se toma a la mediana. Esto ocurre cuando se tienen datos atípicos o aberrantes

ejemplo: Una pequeña empresa de servicios tiene a sus empleados y sus respectivos sueldos:

Empleado	Sueldo
1 secretaria	550
1 conserje	500
4 ingenieros	1800
6 tecnólogos	1300
5 ayudantes	470
gerente	8000

x: sueldos de los empleados

x	$f(x)$	$F(x)$
470	5	5
500	1	6
550	1	7
1300	6	13
1800	4	17
8000	1	18

$\bar{x} = 1466,67$ $\quad n = 18 \therefore n$ es par

$$Md = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = 1300$$

En este caso, la mediana es un mejor representante de la muestra, pues sabemos que el 50% de los empleados ganan 1300 \$ o menos.

- **Medidas de dispersión**

Son valores que indican la dispersión de los datos respecto de la media.

Desviación Promedio: $D_p = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| = \frac{1}{n}\sum_{j=1}^{p}|x_j - \bar{x}|\, f(x_j)$ → a este valor no se lo ocupa mucho

Desviación Estándar o típica: S es la raíz cuadrada de la varianza: S^2. Donde:

$$S^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{(n-1)}\sum_{j=1}^{p}(x_j - \bar{x})^2 f(x_j)$$

Observación:

$$S^2 = \frac{1}{(n-1)}\left[\sum_{j=1}^{p} x_j^2 f(x_j) - n\bar{x}^2\right]$$

Varianza de la muestra

S → muestra σ → población

Observación: Cuando los datos son de la población, la varianza población es:

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^{p}(x_j - \bar{x})^2 f(x_j)$$

La desviación estándar es una medida de población por lo siguiente:

$\bar{x} \pm S : [\bar{x} - S ; \bar{x} + S]$

en este intervalo estará aproximadamente el 68% de los datos.

$\bar{x} \pm 2S : [\bar{x} - 2S ; \bar{x} + 2S]$

en este intervalo estará aproximadamente el 95% de la muestra.

$\bar{x} \pm 3S : [\bar{x} - 3S ; \bar{x} + 3S]$

en este intervalo estará aproximadamente el 99% de la muestra.

ejemplo:

#	intervalo	Punto medio	f(i)	F(i)
1	99,5-104,5	102	3	3
2	104,5-109,5	107	7	10
3	109,5-114,5	112	13	23
4	114,5-119,5	117	19	42
5	119,5-124,5	122	12	54
6	124,5-129,5	127	8	62
7	129,5-134,5	132	4	66

$$S^2 = \frac{1}{65}[31212 + 80143 + 163072 + 260091 + 178608 + 129032 + 69696 - 908113,14]$$

$$S^2 = \frac{1}{65}[3740,86]$$

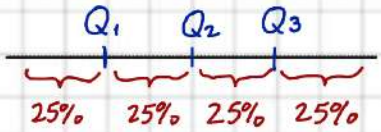
$$S^2 = 57,55$$

- **Medidas de Posición:**

Son valores que dividen a los datos en grupos con igual porcentaje de datos. La **mediana** es también una medida de posición.

Cuartiles: dividen a la muestra en 4 grupos, cada uno con el 25% de los datos. Se los denota como Q_1, Q_2, Q_3

El diagrama de caja sirve para determinar valores atípicos. Es una medida de dispersión



Q_1: el 25% de los datos es menor o igual que el.
Q_2: el 50% de los datos es menor o igual que el.
Q_3: el 75% de los datos es menor o igual que el.

Se utilizan en análisis estadísticos.

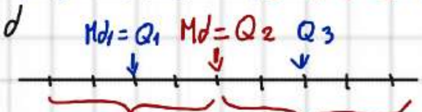
Quintiles: se utilizan mucho para estudios socioeconómicos. Se denotan C_1, C_2, C_3, C_4. → 20%

Deciles: dividen a los datos en 10 grupos. → 10%

Percentiles: P_K: valor K-ésimo tal que el K% de los datos es menor o igual que el.

Observación:
- $Q_1 = P_{25}$, $Q_2 = P_{50}$, $Q_3 = P_{75}$
- $C_1 = P_{20}$, $C_2 = P_{40}$, ...
- $D_1 = P_{10}$, $D_2 = P_{20}$, ...

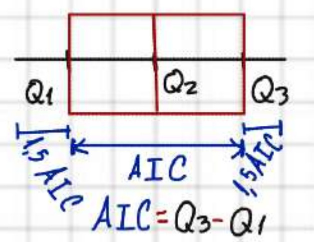
Datos individuales (Cuartiles): Por definición, $Q_2 = Md$



calculamos la mediana = Q_1

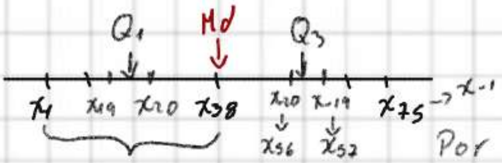
Para calcular Q_3, lo hacemos por simetría respecto a los extremos

Diagrama de caja:



AIC = $Q_3 - Q_1$

ejemplo: $n = 75$, ∴ impar $\Rightarrow Md = x_{\frac{n+1}{2}} = x_{38} = Q_2$



$n_1 = 38$ ∴ par

$Md_1 = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$

$= \frac{x_{19} + x_{20}}{20} = Q_1$

Por simetría: $Q_3 = Md_2 = \frac{x_{-19} + x_{-20}}{2}$

$= \frac{x_{n-19+1} + x_{n-20+1}}{2}$

$= \frac{x_{57} + x_{56}}{2}$

Observación: Para datos agrupados, los percentiles se calculan similar a lo de la mediana, usando interpolación. Obteniendo que para P_K,

$\to$ valor de la variable $\qquad np \to \#$ de datos $\leq P_K$

$$P_K = L + \frac{np - F}{f}\, \omega$$

donde $p = \frac{K}{100}$ y L es el límite inferior del intervalo que contiene al percentil y es aquel cuya frecuencia acumulada (F) es mayor o igual que p.

05/05/2025

ejercicios

Ejercicio del Deber

2. De los datos de una tabla de distribución de frecuencias, con 5 intervalos de clase y ancho de clase común, se observó que: $Q_2 = 24$, $x_1 = 16$, $x_3 = 24$, $n_3 = 2n_1$, $n_4 = n_2$, $n_5 = 2n_2$. ¿Qué porcentaje del total son menores de 30?

- Tabla de frecuencias

#	Intervalo	Punto Medio	$f(i)$	$F(i)$
1	14 - 18	16	n_1	n_1
2	18 - 22	20	n_2	n_1+n_2
3	22 - 26	24	$2n_1$	$3n_1+n_2$
4	26 - 30	28	n_2	$3n_1+2n_2$
5	30 - 34	32	$2n_2$	$3n_1+4n_2$

ω: longitud de intervalo

$x_3 - x_1 = 2\omega$

$24 - 16 = 2\omega$

$\omega = 4$

$Q_2 = P_{50} = 24 \in 3^{\circ}$ intervalo $\qquad L = 22 \qquad F = n_1 + n_2 \qquad f = 2n_1$

$$\Rightarrow 24 = 22 + \frac{np - (n_1 + n_2)}{2n_1} \cdot 4$$

$$2 = \frac{\frac{3n_1 + 4n_2}{2} - (n_1 + n_2)}{n_1} \cdot 2$$

$$1 = \frac{\frac{n_1 + 2n_2}{2}}{n_1}$$

$$2n_1 = n_1 + 2n_2$$

$$n_1 = 2n_2$$

#	Intervalo	Punto Medio	$f(i)$	$F(i)$
1	14 - 18	16	$2n_2$	$2n_2$
2	18 - 22	20	n_2	$3n_2$
3	22 - 26	24	$4n_2$	$7n_2$
4	26 - 30	28	n_2	$8n_2$
5	30 - 34	32	$2n_2$	$10n_2$

$P_{12} = 30 \in 4^{\circ}$ intervalo

$np = 8n_2$

$$p = \frac{8n_2}{n} = \frac{8n_2}{10n_2} = \frac{4}{5} = 80\%$$

20. La producción diaria de leche en cierta zona del país está dada por:

Intervalos de clase (litros de leche)	x_i	Frecuencias (número de días)	$F(i)$
5617.5 - 5692.5	5655	33	33
5692.5 - 5767.5	5730	37	70
5767.5 - 5842.5	5805	31	101
5842.5 - 5917.5	5880	29	130
5917.5 - 5992.5	5955	28	158
5992.5 - 6067.5	6030	25	183
6067.5 - 6142.5	6105	27	210
6142.5 - 6217.5	6180	31	241

20.1. Si por cada litro, la pasteurizadora paga 0,28 dólares, ¿Cuál es la cantidad total que la pasteurizadora paga y cuál es la cantidad promedio que paga diariamente?

20.2. ¿Cuál es la desviación estándar de la producción de leche diaria en la zona?.

20.3. ¿En cuántos días se puede asegurar la producción no ha superado la cantidad de 6000 litros?

20.4. ¿Cuál es la máxima cantidad de litros diarios que se puede asegurar se han obtenido en los 150 días de menor producción?

20.5. ¿En cuántos días se puede asegurar la producción se halla entre 5800 y 6050 litros?

20.6. ¿Cuál es la máxima cantidad de litros diarios que se puede asegurar se han obtenido en los 120 días de menor producción y cuál es la mínima cantidad de litros diarios que se puede asegurar se han obtenido en los 120 días de mejor producción?

1) Pago total $= 0{,}28(33\cdot 5655 + 37\cdot 5730 + 31\cdot 5805 + \dots + 31\cdot 6180)$
$= 398441{,}4$

Pago promedio $= 1653{,}28$

Otra forma: $\bar{x}\cdot 0{,}28 = 5904{,}59 \cdot 0{,}28$
$= 1653{,}28$

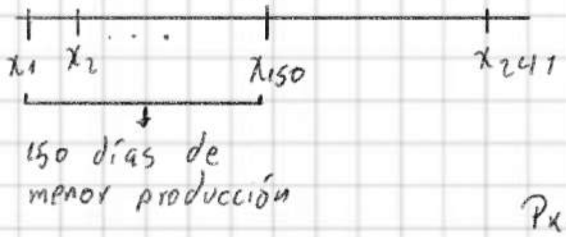
2) $S = 175{,}70$

3) $P_x = 6000$ $\quad np = ?$ $\quad L = 5992{,}5$ $\quad F = 158$ $\quad f = 25$

$$6000 = 5992{,}5 + \frac{np - 158}{25}\cdot 7$$

$np = 160{,}5 \rightarrow$ En 161 días la producción diaria no superó los 6000 litros.

4) max de la producción en estos 150 días. → valor de la variable ⇒ percentil $np = 150$



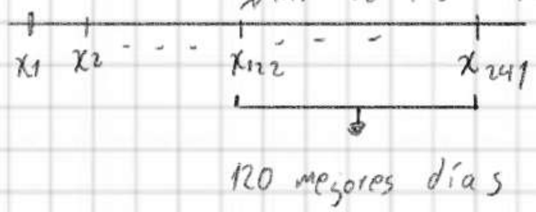
$$P_k = 5917,5 + \frac{150 - 130}{28} \cdot 75$$

$$P_k = 5971,07$$

5) Hallar los percentiles tales que $P_{k_2} - P_{k_1} =$ Valor que necesitamos

6) Primera parte se saca igual que el 20.4

Min de los 120 días → Percentil → $np = 122$



- **Medidas de forma.**

Asimetría: sirve para ver si hay una cola mas larga que otra. Vamos a trabajar con las medidas de Fisher.

Curtosis: sirve para ver la forma de la campana. Vamos a usar la medida de Fisher.

Probabilidad

Vamos a trabajar con experimentos (o pruebas) aleatorios.

Ω: espacio muestral. Conjunto de posibles resultados.

ejemplo:

1.- Lanzamiento de una moneda.
Ω: {cara, sello}

2.- Lanzamiento de un dado.
Ω: $\{1, 2, 4, 5, 6\}$

3.- Lanzamiento de una moneda y un dado.
Ω: $\{(c,1), (c,2), (c,3), (c,4), (c,5), (c,6), (s,1), (s,2), (s,3), (s,4), (s,5), (s,6)\}$

Ω es el equivalente a $\mathcal{U}$ en teoría de conjuntos.

$A \subseteq \Omega$: eventos o sucesos.

$\emptyset \subseteq \Omega$. $\emptyset$ es un evento imposible de que ocurra.

ejemplo: A: dado par. → ej: 3
$A = \{(c,2), (c,4), (c,6), (s,2), (s,4), (s,6)\}$

> Observación:
> Los eventos, al ser subconjuntos, se puede operar con ellos ($\cap, \cup, -, ^c, \dots$) para obtener nuevos eventos.

Si $A \cap B = \emptyset$, se dice que A y B son mutuamente excluyentes.

Si el experimento se realiza n veces, el número de veces que ocurre un evento A, se denomina frecuencia de ocurrencia: n_A.

La frecuencia relativa de ocurrencia del evento: $fr(A) = \frac{n_A}{n}$

ejemplo: lanzamiento de una moneda.

$\Omega = \{C, S\}$

$A = \{C\}$

$n = 7$ Resultados: C, S, S, C, S, S, C $\quad n_A = 3 \Rightarrow fr(A) = \frac{3}{7} = 0,4285$

$n = 50$ Resultados: C, S, S, C, S, S, C, ..., C $\quad n_A = 23 \Rightarrow fr(A) = \frac{23}{50} = 0,46$

$n = 1000$ Resultados: .... $\quad n_A = 502 \Rightarrow fr(A) = 0,502$

$n = 10000$ Resultados: .... $\quad n_A = 4997 \Rightarrow fr(A) = 0,4997$

En general, cuando n es "grande" $(n \to \infty)$, el valor de la frecuencia relativa tiende a un valor fijo: $P(A) = \lim_{n \to \infty} fr(A)$.

En el ejemplo: $P(A) = 0.5$

A este valor se le denomina: Probabilidad de ocurrencia del evento A.

Este punto de vista intuitivo sugiere que la probabilidad proviene de la frecuencia relativa, y esta tiene, entre otras, las siguientes propiedades.

i) $fr(\Omega) = 1$

ii) $0 \leq fr(A) \leq 1$

iii) $A \cap B = \emptyset \Rightarrow fr(A \cup B) = fr(A) + fr(B)$

Por lo tanto, la probabilidad también tiene estas propiedades.

i) $P(\Omega) = 1$

ii) $0 \leq Pr(A) \leq 1$

iii) $A \cap B = \emptyset \Rightarrow Pr(A \cup B) = Pr(A) + Pr(B)$

De manera formal:

$$P: \mathcal{P}(\Omega) \to \mathbb{R}$$
$$A \mapsto Pr(A)$$

Como la teoría de probabilidad subyace en la teoría de conjuntos, las propiedades de conjuntos sirven para determinar propiedades en probabilidades.

Para el cálculo de probabilidades de eventos, vamos a usar el método efectivo:

$P(A) = \frac{|A|}{|\Omega|}$ → cardinalidad de A; → cardinalidad de Ω

$P(A) = \frac{\#\text{casos fav.}}{\#\text{casos tot.}}$

Propiedades de Probabilidades:

1.- $P(\emptyset)=0$ Dem: $\emptyset \cap \Omega = \emptyset \Rightarrow \emptyset$ y Ω son mutuamente excluyentes

$$\Rightarrow P(\emptyset \cup \Omega) = P(\emptyset) + P(\Omega)$$
$$\Rightarrow P(\Omega) = P(\emptyset) + P(\Omega)$$
$$\Rightarrow 0 = P(\emptyset)$$

2.- $P(A^c) = 1 - P(A)$ Dem: $A \cap A^c = \emptyset \Rightarrow P(A \cup A^c) = P(A) + P(A^c)$

$$\Rightarrow P(\Omega) = P(A) + P(A^c)$$
$$\Rightarrow 1 = P(A) + P(A^c)$$
$$\Rightarrow P(A^c) = 1 - P(A^c)$$

3.- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

4.- $P(A) = P(A \cap B) + P(A \cap B^c)$

5.- $A \subseteq B \Rightarrow P(A) \leq P(B)$

6.- $P(A^c \cap B^c) = P[(A \cup B)^c]$

7.- $P(A^c \cup B^c) = P[(A \cap B)^c]$

⋮

ejemplo: En una oferta de ropa, un almacén pone a la venta 24 pantalones buenos, 36 pantalones con falla, 33 camisas buenas y 40 camisas con falla. Se considera los siguientes eventos:

A: pantalón B: bueno

C: camisa D: fallo

Calcular:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{60}{133} + \frac{57}{133} - \frac{24}{133} = \frac{93}{100}$$

$$P(A \cup C) = P(A) + P(C) - P(A \cap C) = \frac{60}{133} + \frac{73}{133} - 0 = 1$$

$P(B \cap C) =$

$P(B \cup C) =$

Tabulando los datos:

	Bueno	Malo	Total
Pantalón	24	36	60
Camisa	33	40	73
Total	57	76	133

$P(A) = \frac{60}{133}$ $P(B) = \frac{57}{133}$

$P(C) = \frac{73}{133}$ $P(D) = \frac{76}{133}$

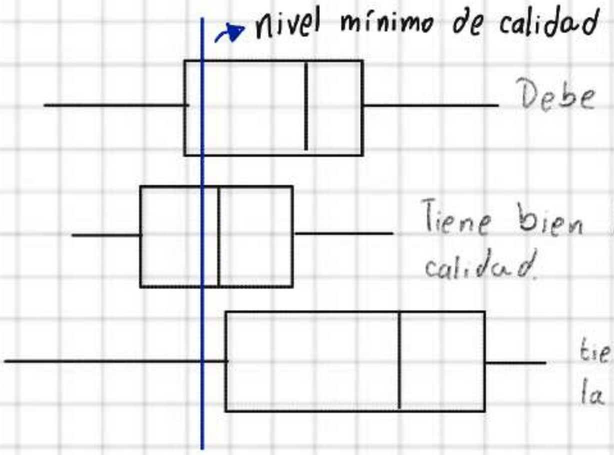
En el curso de probabilidad y estadística hay 46 alumnos, de los cuales 12 son mujeres. Se va a formar una directiva conformada por presidente, secretario y tesorero al azar.

a) ¿Cuántas directivas se pueden formar?
b) ¿Cuál es la probabilidad de que presidente sea mujer?
c) ¿Cuál es la probabilidad de que en la directiva se encuentre al menos una mujer?

$\Omega = \{MMM, MMH, MHM, HMM, MHH, HMH, HHM, HHH\}$

09/05/2025

ejemplo: Se analizan los productos de 3 empresas distintas y se realizan sus respectivos diagramas de caja con respecto al nivel mínimo de calidad.



Debe mejorar variabilidad y calidad

Tiene bien la variabilidad pero debe mejorar la calidad.

tiene bien la calidad pero hay que mejorar la variabilidad.

Métodos de Conteo

Son técnicas que nos permiten medir la cardinalidad de un conjunto. Se basan en 2 principios básicos:

Principio de multiplicación: suponga que se tienen K actividades, cada una de ellas se puede realizar de n_i formas. $(i = 1, 2, \ldots, K)$

Experimento: realizar todas las actividades una a continuación de otra (no importa el orden).

Entonces, el número de formas para hacer el experimento es $n_1 \cdot n_2 \cdot \ldots \cdot n_K$.

ejemplo: Suponga que un jarabe para la tos se realiza en 3 laboratorios, puede ser de tamaño grande o pequeño, y puede tener sabor natural o artificial. ¿De cuántas maneras se puede adquirir un jarabe?

1° actividad: marca: $n_1 = 3$
2° actividad: tamaño: $n_2 = 2$
3° actividad: sabor: $n_3 = 2$

Hay que realizar las 3 actividades

$\Rightarrow$ principio de multiplicación: $\#$ formas $= n_1 \cdot n_2 \cdot n_3 = 12$

ejemplo: Continuación del ejercicio de las directivas.

a) Experimento: nombrar directiva (Presidente, Secretari@, Tesorer@)

Por actividades:

1° act: presid: $n_1 = 46$
2° act: secr: $n_2 = 45$
3° act: tes: $n_3 = 44$

$\rightarrow |\Omega| = \#$ directivas $= n_1 \cdot n_2 \cdot n_3 = 91080$

siempre en probabilidad, trabajar con 4 cifras decimales.

b) A: directivas con presidente mujer.
Por actividades

1° act: pres. mujer : $n_1 = 12$
2° act: secretari@: $n_2 = 45$
3° act: tesorer@ : $n_3 = 44$

$|A| = \#$directivas con presidente mujer $= 12 \cdot 45 \cdot 44 = 23760$

$$\Rightarrow P(A) = \frac{|A|}{|\Omega|} = \frac{23760}{91080} = 0,2608$$

c) B: directivas con al menos una mujer

B^c: directivas sin mujeres.

$$P(B) = 1 - P(B^c)$$
$$= 1 - \frac{|B^c|}{|\Omega|}$$
$$= 1 - \frac{35904}{91080}$$
$$P(B) = 0,6057$$

Principio de Adición: suponga que se tienen K actividades, cada una de ellas se puede realizar de n_i formas ($i: 1, 2, \ldots, K$)

Experimento: Realizar una de las actividades.

Entonces, el número de formas de hacer el experimento es: $n_1 + n_2 + \ldots + n_K$

ejemplo: Suponga que un accidente es observado por 5 niños, 6 mujeres y 4 hombres. ¿De cuántas maneras se puede contar el accidente?

Para enterarse del accidente, es necesario realizar solo una de las actividades.

1° act: niño: $n_1 = 5$
2° act: mujer: $n_2 = 6$
3° act: hombre: $n_3 = 4$

ejemplo: 3 amigos van a realizar una competencia atlética. ¿De cuántas maneras pueden llegar a la meta?

1° actividad: llegan separados : $n_1 = 3 \cdot 2 \cdot 1 = 6$
2° actividad: 1 ganador, 2 juntos : $n_2 = 3$
3° actividad: 2 ganadores, 1 final: $n_3 = 3$
4° actividad: llegan todos juntos: $n_4 = 1$

Dado que el experimento se cumple al realizar solo 1 de las 4 actividades, es un principio de adición. Es decir, pueden ser 13 formas.

Permutaciones: si de un conjunto de n elementos, se toman en orden k de ellos para formar un arreglo, cada uno de estos arreglos se denomina una permutación. El número de posibles permutaciones que pueden formarse está dado por:

Sin repetición: ${}_nP_k = \frac{n!}{(n-k)!}$ Con repetición: ${}_nPR_k = n^k$

Demostración: Por actividades (construcción del arreglo)

	Sin Repetición	Con Repetición
1° actividad: 1° lugar	$n_1 = n$	$n_1 = n$
2° actividad: 2° lugar	$n_2 = n-1$	$n_2 = n$
⋮		⋮
k° actividad: k° lugar	$n_k = n-k+1$	$n_k = n$
Total:	${}_nP_k = n_1 \cdot n_2 \cdot \ldots \cdot n_k$ $= n(n-1)\ldots(n-k+1) \cdot \frac{(n-k)(n-k-1)\ldots 2 \cdot 1}{(n-k)(n-k-1)\ldots 2 \cdot 1}$ $= \frac{n!}{(n-k)!}$	${}_nPR_k = n \cdot n \cdot n \ldots n$ $= n^k$

En el ejemplo de la directiva, cada directiva es un arreglo, por lo tanto, tienen un orden. Es decir

${}_{46}P_3$

ejemplo: Cuántos números de 4 dígitos se pueden formar (0, 1, 2, 3, 4, 5, 6, 7, 8, 9)

Por actividades:

1° Escoger las cifras de los miles: $n_1 = 9$

2° Escoger las 3 cifres restantes: $n_2 = {}_{10}PR_3 = 1000$

Combinaciones: de un conjunto de n elementos se toma un subconjunto de k elementos (no importa el orden); cada uno de estos subconjuntos se denomina una combinación. El número de combinaciones que se puede dar está formado por:

Sin Repetición: $nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$

Con Repetición: $nCR_k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$

Observación: Las combinaciones sin repetición equivalen a formar 2 subconjuntos: el primero con k elementos y el segundo con $n-k$ elementos.
Si en vez de formar 2 subconjuntos, se forman m conjuntos, el primero con n_1 elementos, el segundo con n_2 elementos y así sucesivamente hasta el último con n_m elementos.

$$nC_{n_1, n_2 \ldots n_m} = \frac{n!}{n_1! n_2! \cdots n_m!}$$

ejemplo: En un campeonato de fútbol participan 5 equipos. ¿Cuántos partidos se deben realizar?

Equipos: A, B, C, D, E

Partidos: {A,B} {A,C} {A,D} {A,E}, {B,C}, {B,D}, {B,E}, {C,D}, {C,E}, {D,E}

No hay orden, por lo tanto es una combinación

$$5C_2 = \binom{5}{2}$$

ejemplo: Para un campeonato de basquet se inscriben 9 equipos y se van a formar 2 grupos: uno con 4 equipos y otro con 5 equipos. ¿Cómo se pueden formar los grupos?

Por combinaciones: $\# total = {}_9C_4$ o $\# total = {}_9C_5$ o $\# total = {}_9C_{4,5}$

$$\frac{9!}{4!\,5!} = 126$$

12/05/2025

- En una urna se tienen 8 bolas rojas y 6 blancas. Se van a extraer 3 bolas de manera simultánea. Calcular la probabilidad de:

  A: 3 rojas

  B: 2 rojas, 1 blanca

  C: 1 roja, 2 blancas

A: por actividades.

1° actividad extraer 3 bolas rojas

$|\Omega| = {}_{14}C_3$

- De un naipe se toman 5 cartas al azar.
  a) Cuál es la probabilidad de que salga un As.
  b) Cuál es la probabilidad de que salgan los 4 Ases.
  c) Cuál es la probabilidad de que salgan 4 Ases y una vieja (J, Q, K)
  d) Cuál es la probabilidad de que salga una escalera (9, 10, J, Q, K)
  e) Cuál es la probabilidad de que salgan 3 de un palo, 2 de otro palo
  f) Cuál es la probabilidad de que salga al menos un As

a) $P(\text{un as}) = \dfrac{{}_4C_1 \cdot {}_{48}C_4}{{}_{52}C_5}$

b) $P(\text{4 ases}) = \dfrac{{}_4C_4 \cdot {}_{48}C_1}{{}_{52}C_5}$

c) $P(\text{4 ases, 1 vieja}) = \dfrac{{}_4C_4 \cdot {}_{12}C_1}{{}_{52}C_5}$

d) $P(9, 10, J, Q, K) = \dfrac{{}_4C_1 \cdot {}_4C_1 \cdot {}_4C_1 \cdot {}_4C_1 \cdot {}_4C_1}{{}_{52}C_5}$

e) 1° actividad: escoger 1 palo
2° actividad: escoger 3 cartas del palo
3° actividad: escoger el 2° palo
4° actividad: escoger 2 cartas de 2° palo

$P(3 \in 1^\circ \text{palo y } 2 \in 2^\circ \text{palo}) = \dfrac{{}_4C_1 \cdot {}_{13}C_3 \cdot {}_3C_1 \cdot {}_{13}C_2}{{}_{52}C_5}$

f) $P(\text{al menos un as}) = P(\text{1 as} \cup \text{2 ases} \cup \text{3 ases} \cup \text{4 ases})$

$= P(\text{1 as}) + P(\text{2 ases}) + P(\text{3 ases}) + P(\text{4 ases})$

$= \dfrac{{}_4C_1 \cdot {}_{48}C_4}{{}_{52}C_5} + \dfrac{{}_4C_2 \cdot {}_{48}C_3}{{}_{52}C_5} + \dfrac{{}_4C_3 \cdot {}_{48}C_2}{{}_{52}C_5} + \dfrac{{}_4C_4 \cdot {}_{48}C_1}{{}_{52}C_5}$

$= 0{,}3411$

Otra forma: Por complemento

$P(\text{al menos un as}) = 1 - P(\text{ningún as})$

$= 1 - \dfrac{{}_{48}C_5}{{}_{52}C_5}$

$= 0{,}3411$

- En una caja hay 8 bolas rojas numeradas y 6 bolas blancas numeradas. Se extrae una bola al azar. ¿Cuál es la probabilidad de que si la bola es roja, tenga el número 5?

- **Probabilidad Condicionada:** se utiliza cuando se tienen 2 eventos, uno que ya ocurrió y uno que se espera que ocurra.

Si tomamos

A: evento que ya ocurrió.

B: evento que se espera que ocurra.

La probabilidad de que ocurra B sabiendo que A ocurrió, se dice que es la probabilidad de B condicionada a A se denota por:

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

y a A se le dice la condición.

Observación: la fórmula de la probabilidad condicionada se puede reescribir como:

$$P(A \cap B) = P(A)\,P(B \mid A)$$

Esta fórmula se la utiliza para calcular probabilidades de eventos que ocurren de manera consecutiva.

Observación: cuando la ocurrencia de un evento no depende de la ocurrencia del otro evento, se dice que son independientes. En ese caso, $P(B \mid A) = P(B)$ y $P(A \mid B) = P(A)$

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

Definición: $P(A \cap B) = P(A)\,P(B) \Leftrightarrow A$ y B son independientes.

ejemplo: lanzamiento de un dado y una moneda.

A: lanzamiento del dado
B: lanzamiento de la moneda.

En este caso, los eventos son independientes.

ejemplo: En el experimento de las bolas, ¿cuál es la probabilidad de que salgan en el orden r, b, r?

R_1: la 1º rosa
B_2: la 2º blanca
R_3: la 3º rosa

$$P(R_1 \cap B_2 \cap R_3) = P(R_1) \cdot P(B_2 | R_1) \cdot P(R_3 | R_1 \cap B_2)$$

$$= \frac{8}{14} \cdot \frac{6}{13} \cdot \frac{7}{12}$$

$$= \frac{2}{13} = 0{,}1538$$

16/05/2025

ejemplo: En un colegio hay 15 profesores, de los cuales 8 son mujeres. Se va a nombrar 2 inspetores.

a) Cuál es la probabilidad de que sean del mismo sexo.

b) Cuál es la probabilidad de que sean un hombre y una mujer.

A: Inspectores del mismo sexo.
B: 1 hombre y 1 mujer.
Ω: Conjunto de todas las posibles designaciones.

$P(A) = ?$
$P(B) = ?$

$|\Omega| = 15\,C_2 = \binom{15}{2} = 105$

$|A| = 7C_2 + 8C_2 = 49$

$|B| = |\Omega| - |A|$ por complemento
$= 56$

ejemplo: Se lanza un dado 2 veces. Calcule la probabilidad de obtener al menos un 5.

A_1: 5 en el 1° lanzamiento
A_2: 5 en el 2° lanzamiento
A: al menos un 5 $\Rightarrow A: A_1 \cup A_2$

$$\begin{aligned} P(A) &= P(A_1 \cup A_2) \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \cdot P(A_1) \\ &= \frac{1}{6} + \frac{1}{6} - \left(\frac{1}{6} \cdot \frac{1}{6}\right) \\ &= \frac{11}{36} \,// \end{aligned}$$

ejemplo: En una caja se tienen 8 bolas rojas, 3 blancas y 9 azules. Se van a extraer 3 bolas sin reposición. Hallar la probabilidad de que:

a) Las 3 sean rojas
b) 2 sean rojas y 1 blanca
c) Al menos 1 sea blanca
d) 1 de cada color
e) Se extraigan en el orden rojo, blanco y azul.

a) $P(3 \text{ rojas}) = \dfrac{{}_8C_3}{{}_{20}C_3}$

b) $P(2 \text{ rojas}, 1 \text{ blanca}) = \dfrac{{}_8C_2 \cdot {}_3C_1}{{}_{20}C_3}$

c) $P(\text{al menos } 1 \text{ blanca}) = P(1 \text{ blanca} \cup 2 \text{ blancas} \cup 3 \text{ blancas})$

$= P(1 \text{ blanca}) + P(2 \text{ blancas}) + P(3 \text{ blancas})$

$= \dfrac{{}_3C_1 \cdot {}_{17}C_2}{{}_{20}C_3} + \dfrac{{}_3C_2 \cdot {}_{17}C_1}{{}_{20}C_3} + \dfrac{{}_3C_3 \cdot {}_{17}C_0}{{}_{20}C_3}$

$= 0{,}4035$

Otra forma: $P(\text{al menos } 1 \text{ blanca}) = 1 - P(\text{ninguna blanca})$

$= 1 - \dfrac{{}_{17}C_3}{{}_{20}C_3}$

$= 0{,}4035$

d) $P(1 \text{ cada color}) = \dfrac{{}_8C_1 \cdot {}_3C_1 \cdot {}_9C_1}{{}_{20}C_3} = 0{,}18$

$P(r, b, a) = P(r_1 \cap b_2 \cap a_3)$

$= P(r_1) \cdot P(b_2 | r_1) \cdot P(a_3 | r_1 \cap b_2)$

$= \dfrac{8}{20} \cdot \dfrac{3}{19} \cdot \dfrac{9}{18}$

$= \dfrac{216}{6840}$

$= 0{,}0316$

- Repita el problema para el caso con reposición.

ejemplo: Para una rifa se van a vender 38 boletos, de los cuales 7 son premiados. Se van a tomar aleatoriamente 5 boletos. Cuál es la probabilidad de que haya
a) 1 premiado
b) 3 premiados
c) al menos 1 premiado

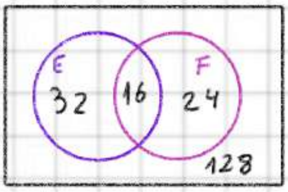
a) $P(1 \text{ premiado}) = \dfrac{{}_7C_1 \cdot {}_{31}C_4}{{}_{38}C_5}$

b) $P(3 \text{ premiados}) = \dfrac{{}_7C_3 \cdot {}_{31}C_2}{{}_{38}C_5}$

c) $P(\text{al menos 1 premiado}) = 1 - P(\text{ningún premiado})$

$= 1 - \dfrac{{}_{31}C_5}{{}_{38}C_5}$

ejemplo: De 200 aspirantes a un cargo, 48 tienen experiencia, 40 tienen formación académica y 32 tienen experiencia pero no tienen formación. Si se va a elegir al aspirante al azar, cuál es la probabilidad de que sea:
a) Con experiencia y formación académica
b) Sin experiencia
c) Con experiencia dado que tiene formación
d) Que sea sin experiencia o con formación



E: tiene experiencia
F: tiene formación
$P(E) = \frac{48}{200}$
$P(F) = \frac{40}{200}$
$P(E \cap F^c) = \frac{32}{200}$

$P(E) = P(E \cap F) + P(E \cap F^c)$

$\frac{48}{200} = P(E \cap F) + \frac{32}{200} \Rightarrow P(E \cap F) = \frac{16}{200}$

a) $P(E \cap F) = \frac{16}{200}$

b) $P(E^c) = 1 - P(E) \Rightarrow P(E^c) = \frac{132}{200}$

c) $P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{16}{200}}{\frac{40}{200}} = \frac{16}{40}$

d) $P(E^c \cup F) = P(E^c) + P(F) - P(F \cap E^c)$

$= 1 - P(E) + P(F) - [P(F) - P(F \cap E)]$

$= 1 - \frac{48}{200} + \frac{16}{200}$

$= \frac{168}{200}$

ejemplo: Se tienen 2 cajas con monedas. En la caja número 1 se tienen 5 monedas de oro y 8 monedas de plata. En la caja 2 hay 7 monedas de oro y 8 de plata. Se lanza una moneda común: si sale cara, se toma la caja 1 se toma una moneda al azar; si sale sello, se realiza con la caja 2. ¿Cuál es la probabilidad de que la moneda sacada (o de la caja 1 o de la caja 2) sea de oro?

I: elegir caja 1
II: elegir caja 2
O: moneda de horo

$O = I \cap O \cup II \cap O$

$P(O) = P(I \cap O \cup II \cap O)$
$= P(I \cap O) + P(II \cap O)$
$= P(I) \cdot P(O|I) + P(II) \cdot P(O|II)$
$= \frac{1}{2} \cdot \frac{5}{13} + \frac{1}{2} \cdot \frac{7}{15} = 0{,}4615$

ejemplo: En una gasolinera hay 3 despachadores: Luis, Francisco y José De datos históricos, se tiene que Luis atiende al 45% de los clientes, Francisco al 30% y José al 25%. Luis limpia el 35% de los parabrisas de los clientes, Francisco el 40% y José el 60%. Si a un cliente le han limpiado el parabrisas, ¿Cuál es la probabilidad de que haya sido atendido por Francisco?

	Atención	Servicio
Luis	45%	35%
Francisco	30%	40%
José	25%	60%

S: dieron servicio de limpieza

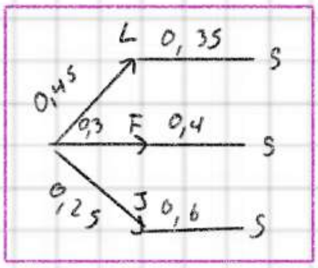
$$P(F \mid S) = \frac{P(F \cap S)}{P(S)}$$

- $S = L \cap S \cup F \cap S \cup J \cap S$

$$\begin{aligned} P(S) &= P(L \cap S) + P(F \cap S) + P(J \cap S) \\ &= P(L) \cdot P(S \mid L) + P(F) P(S \mid F) + P(J) P(S \mid J) \\ &= 0{,}45 \cdot 0{,}35 + 0{,}30 \cdot 0{,}40 + 0{,}25 \cdot 0{,}60 \\ &= 0{,}4275 \end{aligned}$$

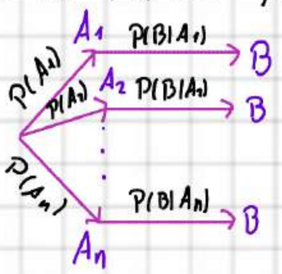
$$P(F \cap S) = P(F) P(S \mid F) = 0{,}30 \cdot 0{,}40 = 0{,}12$$

$$P(F \mid S) = \frac{0{,}12}{0{,}4275} = 0{,}2526$$



Probabilidad total: Sean los eventos excluyentes $A_1, A_2, \ldots, A_n$, uno de los cuales debe ocurrir y sea B un evento que se espera ocurra posteriormente. Entonces:

$$P(B) = \sum_i P(A_i) P(B \mid A_i)$$



- **Teorema de Bayes:** $P(A_i|B) = \dfrac{P(A_i)P(B|A_i)}{P(B)}$

19/05/2025

variables aleatorias

Suponemos que un experimento aleatorio tiene espacio muestral Ω. Una variable aleatoria (v.a.) denotada en $X, Y, W, V, \ldots$ y es una función que a cada elemento de Ω le asocia un número.
En general una variable aleatoria es de la forma:

$$X : \Omega \longrightarrow \mathbb{R}$$
$$\omega \longmapsto X(\omega)$$

ejemplo: Bolas en una urna. 8 rojas, 6 blancas y 4 azules.

Experimento: extraer 3 bolas $\quad \Omega = \{\{R,R,R\}, \{R,R,B\}, \{R,R,A\}, \ldots, \{A,A,A\}\}$

Sea X: # de bolas extraídas

$\{R,R,R\} \longmapsto 3$
$\{R,R,B\} \longmapsto 2$
$\vdots$
$\{A,A,A\} \longmapsto 0$

ejemplo: lanzamiento de 2 dados.

$\Omega = \{(1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (2,6), \ldots, (6,6)\}$

X: suma de los resultados

$(1,1) \longmapsto 2$
$\vdots$
$(6,6) \longmapsto 12$

$\text{Rec } X = \{2, 3, \ldots, 12\}$

> El recorrido o imagen de la variable aleatoria es el conjuntos de posibles valores que puede tomar. Se denota $\text{Rec } X$.

Y: producto de los resultados

$(1,1) \longmapsto 1$
$(6,6) \longmapsto 36$

$\text{Rec } Y = \{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30, 36\}$

Discreto: # finito o infinito numerable de elementos
Continuo: intervalos

ejemplo: Tomar un foco, prenderlo y dejarlo prendido hasta que se queme.

X: Tiempo de vida útil

Si el recorrido es un conjunto discreto, se dice que la variable aleatoria es discreta, cuando el recorrido es continuo, se dice que la variable aleatoria es continua.

- **Variable aleatoria discreta:** Sea X una variable aleatoria discreta con recorrido

$\operatorname{Rec} X = \{x_1, x_2, \ldots, x_n\}$ $\vee$ $\operatorname{Rec} X = \{x_1, x_2, \ldots\}$

El análisis de este tipo de variable aleatoria es muy similar al de datos estadísticos, en dónde se calculaban la frecuencia relativa, la frecuencia acumulada relativa, la media y la varianza

Datos Estadísticos	v.a.(Rec X)
frecuencia relativa	función de probabilidad
frecuencia relativa acumulada	Función de probabilidad acumulada
media (medidas estadísticas)	media (parámetros)
varianza (medidas estadísticas)	varianza (parámetros)

Función de probabilidad: es la que permite calcular la probabilidad de que una variable tome determinado valor. Se denota como $p(x)$ y se define como

$$p(x) = P(X = x) = P(\omega / X(\omega) = x)$$

El valor de esta función de probabilidad se lo debe calcular para todos los elementos del recorrido.

$x = x_i : p(x_i) = P(X = x_i) = P(\omega / X(\omega) = x_i)$ $\quad i = 1, \ldots, n$

si $x \neq x_i : p(x) = 0$

$$p(x) = P(X = x) = \begin{cases} p_1; & x = x_1 \\ p_2; & x = x_2 \\ \vdots & \\ p_n; & x = x_n \\ 0; & \text{otros casos} \end{cases}$$

→ esto normalmente se omite

ejemplo: En una urna hay 6 bolas blancas, 8 rosas y 4 azules. Se van a extraer 3 bolas con reemplazo. Sea la variable aleatoria

X: número de bolas rojas extraídas.

$Rec\, X = \{0,1,2,3\}$ la función de probabilidad se debe calcular para los valores: $x=0$, $x=1$, $x=2$ ∧ $x=3$.

$x=0$:
$$\begin{aligned} p(0) = P(X=0) &= P(\omega / X(\omega) = 0) \\ &= P(R_1^c \cap R_2^c \cap R_3^c) \\ &= P(R_1^c) \cdot P(R_2^c | R_1^c) \cdot P(R_3^c | R_1^c \cap R_2^c) \\ &= \frac{10}{18} \cdot \frac{10}{18} \cdot \frac{10}{18} \\ &= 0{,}17146 \end{aligned}$$

$x=1$:
$$\begin{aligned} p(1) = P(X=1) &= P(\omega / X(\omega) = 1) \\ &= P(R_1 \cap R_2^c \cap R_3^c \cup R_1^c \cap R_2 \cap R_3^c \cup R_1^c \cap R_2^c \cap R_3) \\ &= P(R_1 \cap R_2^c \cap R_3^c) + P(R_1^c \cap R_2 \cap R_3^c) + P(R_1^c \cap R_2^c \cap R_3) \\ &= P(R_1) P(R_2^c | R_1) P(R_3^c | R_1 \cap R_2^c) + \dots \\ &= \frac{8}{18} \cdot \frac{10}{18} \cdot \frac{10}{18} + \dots \\ &= 0{,}41152 \end{aligned}$$

$\vdots$

Otra forma: $x=1$: $p(1) = P(X=1) = 3C_1 \cdot \frac{8}{18} \cdot \frac{10}{18} \cdot \frac{10}{18} = 0{,}41152$

$x=2$: $p(2) = P(X=2) = 3C_2 \cdot \frac{8}{18} \cdot \frac{8}{18} \cdot \frac{10}{18} = 0{,}32921$

$x=3$: $p(3) = P(X=3) = 3C_3 \cdot \frac{8}{18} \cdot \frac{8}{18} \cdot \frac{8}{18} = 0{,}08781$

$$p(x) = \begin{cases} 0{,}17146; & x=0 \\ 0{,}41152; & x=1 \\ 0{,}32921; & x=2 \\ 0{,}08781; & x=3 \end{cases}$$

ejemplo: Experimento sin reemplazo.

X: número de bolas rojas extraídas $Rec\,X = \{0, 1, 2, 3\}$

$x=0$: $p(0) = P(X=0) = P(\omega / X(\omega)=0)$

$$= \frac{8C_0 \cdot 10C_3}{18C_3} = 0{,}1471$$

$x=1$: $p(1) = P(X=1) = P(\omega / X(\omega)=1)$

$$= \frac{8C_1 \cdot 10C_2}{18C_3} = 0{,}4412$$

$x=2$: $p(2) = P(X=2) = P(\omega / X(\omega)=2)$

$$= \frac{8C_2 \cdot 10C_1}{18C_3} = 0{,}3431$$

$x=3$: $p(3) = P(X=2) = P(\omega / X(\omega)=3)$

$$= \frac{8C_3 \cdot 10C_0}{18C_3} = 0{,}0686$$

La función de probabilidad satisface que
$0 \leq p(x) \leq 1$
y así mismo,
$\sum p(x_i) = 1$

$$p(x) = \begin{cases} 0{,}1471 \\ 0{,}4412 \\ 0{,}3431 \\ 0{,}0686 \end{cases}$$

→ Función de distribución

- **Función de distribución Acumulada: $F(x)$** se define por:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

ejemplo: Ejemplo 1:

- $F(1) = P(X \leq 1) = \sum_{x_i \leq 1} p(x_i) = p(0) + p(1)$
- $F(2{,}3) = P(X \leq 2{,}3) = \sum_{x_i \leq 2{,}3} p(x_i) = p(0) + p(1) + p(2)$
- $F(5) = P(X \leq 5) = \sum_{x_i \leq 5} p(x_i) = p(0) + p(1) + p(3) = 1$

Para propósitos prácticos, es conveniente calcular la función de distribución por intervalos, en los intervalos que denominan los elementos del recorrido.

$x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_n$

Intervalos:

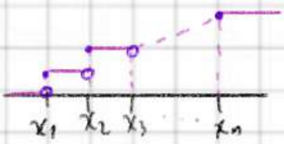
$(-\infty, x_1): F(x)=0$
$[x_1, x_2): F(x)=p_1$
$[x_2, x_3): F(x)=p_1+p_2$
$\vdots$
$[x_n, +\infty): F(x)=1$

de donde:

$$F(x)=P(X \le x) = \begin{cases} 0 & ; x < x_1 \\ p_1 & ; x_1 \le x < x_2 \\ p_1+p_2 & ; x_2 \le x < x \\ \vdots & \\ 1 & ; x_n \le x \end{cases}$$

ejemplo: Ejemplo de las bolas sin reemplazo.

$$p(x) = \begin{cases} 0{,}17146 & ; x=0 \\ 0{,}41152 & ; x=1 \\ 0{,}32921 & ; x=2 \\ 0{,}08781 & ; x=3 \end{cases} \Rightarrow F(x) = \begin{cases} 0 & ; x<0 \\ 0{,}17146 & ; 0 \le x < 1 \\ 0{,}58298 & ; 1 \le x < 2 \\ 0{,}91219 & ; 2 \le x < 3 \\ 1 & ; 3 \le x \end{cases}$$



La función de distribución satisface las siguientes propiedades:

1.- $F(-\infty)=0$
2.- $F(+\infty)=1$
3.- F es no decreciente

Si $F(x)$ es conocido, entonces la función de probabilidad queda determinada por:

$$p(x_i) = F(x_i) - F(x_{i-1})$$

ejemplo: En un negocio se considera la variable aleatoria.

X: Pérdidas o ganancias mensuales.

La correspondiente función de distribución es

$$F(x) = \begin{cases} 0 & ; \quad x < -3 \\ 0{,}05 & ; \quad -3 \leq x < -1 \\ 0{,}20 & ; \quad -1 \leq x < 2 \\ 0{,}45 & ; \quad 2 \leq x < 3 \\ 0{,}70 & ; \quad 3 \leq x < 4.5 \\ 0{,}92 & ; \quad 4.5 \leq x < 5 \\ 1 & ; \quad 5 \leq x \end{cases}$$

1.- ¿Cuáles son los posibles valores de pérdidas o ganancias que puede haber?

2.- ¿Cuál es la función de probabilidad.

Rec X: $\{-3, -1, 2, 3, 4.5, 5\}$

Función de probabilidad

$x = -3 : p(-3) = F(-3) - F(-\infty) = 0{,}05 - 0 = 0{,}05$

$x = -1 : p(-1) = F(-1) - F(-3) = 0{,}20 - 0{,}05 = 0{,}15$

$x = 2 : p(2) = F(2) - F(-1) = 0{,}45 - 0{,}20 = 0{,}25$

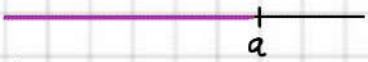
$x = 3 : p(3) = F(3) - F(2) = 0{,}70 - 0{,}45 = 0{,}25$

$x = 4.5 : p(4.5) = F(4.5) - F(3) = 0{,}92 - 0{,}70 = 0{,}22$

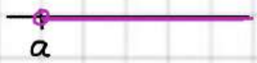
$x = 5 : p(5) = F(5)$

La función de distribución es de utilidad en el cálculo de probabilidades en intervalos, de la siguiente manera:

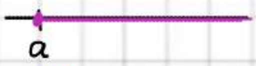
$F(a) = P(X \leq a)$



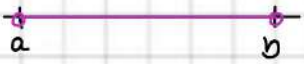
$P(X > a) = 1 - P(X \leq a) = 1 - F(a)$



$P(X \geq a) = 1 - F(a) + p(a)$

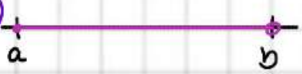


$P(a < X < b) = F(b) - F(a) - p(b)$



$P(a \leq X \leq b) = F(b) - F(a) + p(a)$

$P(a \leq X < b) = F(b) - F(a) - p(b) + p(a)$



$P(a < X \leq b) = F(b) - F(a)$

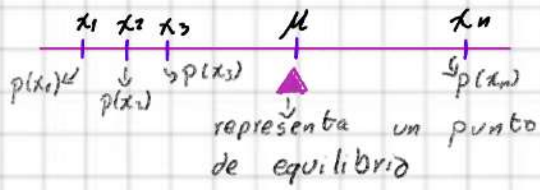
26/05/2025

> Observación: $F(-\infty) = 0$
> $F(+\infty) = 1$
> F es no decreciente

> Observación: $p(x_i) = F(x_i) - F(x_{i-1})$

- **Parámetros:** Son valores que caracterizan a la variable aleatoria

☆ Media / Valor Esperado / Esperanza: $\mu = E[X] = \sum_i x_i \, p(x_i)$



La media corresponde al valor que se esperaría ocurra cuando se va a realizar un experimento.

Corresponde al promedio si el experimento se realiza muchas veces.

☆ Varianza: $\sigma^2 = V[X] = Var[X] = E[(X-\mu)^2] = \sum_i (x_i-\mu)^2 p(x_i)$

Observación:
$\sigma^2 = E[X^2] - (E[X])^2$

Observación: la raíz cuadrada de la varianza se denomina Desviación Estandar o típica.

$\sigma = \sqrt{\sigma^2}$

ejemplo: En una tienda, mensualmente se pueden obtener pérdidas o ganancias. La ocurrencia está dada por la siguiente tabla:

X: Pérdidas/ganancias mensuales (en miles de dólares).

$$p(X) = \begin{cases} 0{,}03 & ; x=-4 \\ 0{,}08 & ; x=-2 \\ 0{,}06 & ; x=-1 \\ 0 & ; x=0 \\ 0{,}10 & ; x=1 \\ 0{,}15 & ; x=2 \\ 0{,}24 & ; x=3 \\ 0{,}30 & ; x=4 \\ 0{,}04 & ; x=5 \end{cases}$$

a) Describa completamente la variable aleatoria.

b) Halle las siguientes probabilidades:

i) Que hayan ganancias
ii) Que no hayan pérdidas
iii) Que sabiendo que va a haber ganancias, sea de al menos 2 mil
iv) Sabiendo que va a haber pérdidas, que no supere los 3 mil
v) Que se encuentre entre 500 y 2 800

Función de Distribución:

$$F(x) = P(X \leq x) = \begin{cases} 0 & ; x < -4 \\ 0{,}03 & ; -4 \leq x < -2 \\ 0{,}11 & ; -2 \leq x < -1 \\ 0{,}17 & ; -1 \leq x < 0 \\ 0{,}17 & ; 0 \leq x < 1 \\ 0{,}27 & ; 1 \leq x < 2 \\ 0{,}42 & ; 2 \leq x < 3 \\ 0{,}66 & ; 3 \leq x \, 4 \\ 0{,}96 & ; 4 \leq x < 5 \\ 1 & ; 5 \leq x \end{cases}$$

Parámetros:

a)

• media: $\mu = E[X] = \sum x_i p(x_i)$

$= (-4)(0,03) + (-2)(0,08) + \dots + (5)(0,04)$

$\mu = E[X] = 2,18$

• varianza: $\sigma^2 = E[x^2] - (E(x))^2$

$= [(-4)^2(0,03) + (-2)^2(0,03)^2 + \dots + (5)^2(0,24)] - 2,18^2$

$\sigma^2 = 4,7676$

• desviación estandar: $\sigma = \sqrt{\sigma^2}$

$\sigma = 2,1834$

i) $P(X > 0) = 1 - P(X \leq 0) = 1 - F(0) = 1 - 0,17 = 0,83$

ii) $P(X \geq 0) = 1 - P(X < 0) = 1 - [F(0) - p(0)] = 1 - [0,17 - 0] = 0,83$

iii) $P(X \geq 2 \mid x > 0) = \dfrac{P(X \geq 2 \cap x > 0)}{P(x > 0)} = \dfrac{P(X \geq 2)}{P(x > 0)}$

$= \dfrac{1 - P(X < 2)}{1 - P(x \leq 0)}$

$= \dfrac{1 - [F(2) - p(2)]}{1 - F(0)}$

$= \dfrac{1 - [0,42 - 0,15]}{1 - 0,17}$

$= \dfrac{0,73}{0,83} = 0,8795$

iv) $P(-3 \leq X < 0 \mid X < 0) = \dfrac{P(-3 \leq X < 0 \cap X < 0)}{P(X < 0)} = \dfrac{P(-3 \leq x < 0)}{P(x < 0)}$

$= \dfrac{F(0) - F(-3) - p(0) - p(-3)}{F(0) - p(0)}$

$=$

v) $P(0,5 \leq X \leq 2,8) = F(2,8) - F(0,5) + p(0,5)$

$= 0,42 - 0,17 + 0$

$= 0,25$

5. En cierto laboratorio se ofrecen 4 tipos de exámenes: el primero dura 20 minutos y es utilizado el 45% de las veces; el segundo dura 30 minutos y es utilizado el 25% de las veces; el tercero dura 40 minutos y es utilizado el 20% de las veces; el cuarto dura 60 minutos y es utilizado el 10% de las veces.
El costo de los exámenes está dado por $C(X) = 10.000 - 3X + 2X^2$ (centavos de dólar), donde X representa el número de minutos empleados en el examen.

a) Halle la probabilidad que por un examen se deba pagar entre 20.000 y 50.000 centavos

b) Halle el costo esperado y la varianza del costo

X: Tiempo de duración de un examen.

$$p(x) = \begin{cases} 0,45 & ; x = 20 \\ 0,25 & ; x = 30 \\ 0,20 & ; x = 40 \\ 0,10 & ; x = 60 \end{cases} \qquad F(x) = \begin{cases} 0 & ; x < 20 \\ 0,45 & ; 20 \le x < 30 \\ 0,70 & ; 30 \le x < 40 \\ 0,90 & ; 40 \le x < 60 \\ 1 & ; 60 \le x \end{cases}$$

media: $\mu = E[x] = \sum x_i \, p(x_i) = 30,5 \ (min)$

varianza: $\sigma^2 = V[x] = E[x^2] - (E[x])^2 = 154,75 \ (min^2)$

desviación estandar: $\sigma = \sqrt{\sigma^2} = 12,4399 \ (min)$

Costo: $C = 10000 - 3X + 2X^2$

- Función de probabilidad por el costo

Si el tiempo es $x = 20$ (min): costo servicio $C = 10740$

y la probabilidad de ocurrencia es 0,45

Así,

$$p(c) \begin{cases} 0,45 & ; C = 10740 \\ 0,25 & ; C = 11710 \\ 0,20 & ; C = 13080 \\ 0,10 & ; C = 17020 \end{cases} \qquad F(c) = \begin{cases} 0 & ; C < 10740 \\ 0,45 & ; 10740 \le C < 11710 \\ 0,70 & ; 11710 \le C < 13080 \\ 0,90 & ; 13080 \le C < 17020 \\ 1 & ; 17020 \le C \end{cases}$$

a) $P(20000 \le C \le 50000) = F(50000) - F(20000) + p(20000)$
$= 1 - 1 + 0 = 0$

b) $\mu_c = E[C] = E[10000 - 3X + 2X^2] = 10000 - 3E[X] + 2E[X^2]$
$= 10000 - 3 \cdot 30,5 + 2(154,75 + 30,5^2]$
$= 11058,5$

Propiedades del valor esperado y de la varianza

- $E[K] = K$
- $E[aX \pm bY] = a\,E[X] \pm b\,E[Y]$ → la esperanza es lineal
- $V[K] = 0$
- $V[aX \pm bY] = a^2 V[X] + b^2 [Y]$ → se da cuando las variables son independientes

ejemplo: Suponga que para una rifa se han impreso 500 boletos, de los cuales 5 son premiados. Se van a extraer aleatoriamente 8 boletos. ¿Cuál es la probabilidad de que entre los 8 haya:

a) 1 premiado?
b) al menos un premiado?
c) a lo mucho 2 premiados?

X: # boletos premiados entre los 8 extraídos $Rec\,X = \{0, 1, 2, 3, 4, 5\}$

Función de probabilidad

$$x=0 : p(0) = P(X=0) = \frac{5C_0 \cdot 495C_8}{500C_8} = 0{,}9222$$

$$x=1 : p(1) = P(X=1) = \frac{5C_1 \cdot 495C_7}{500C_8} = 0{,}0756$$

$$x=2 : p(2) = P(X=2) = \frac{5C_2 \cdot 495C_6}{500C_8} = 0{,}0021$$

$$x=3 : p(3) = P(X=3) = \frac{5C_3 \cdot 495C_5}{500C_8} = 0{,}00002$$

$$x=4 : p(4) = P(X=4) = \frac{5C_4 \cdot 495C_4}{500C_8} = 0{,}0000001$$

$$x=5 : p(5) = P(X=5) = \frac{5C_5 \cdot 495C_3}{500C_8} = 0{,}0000000002$$

$$p(x) = P(X=x) = \begin{cases} 0{,}9222 & ; x=0 \\ 0{,}0755 & ; x=1 \\ 0{,}0021 & ; x=2 \\ 0{,}00002 & ; x=3 \\ 0 & ; x=4 \\ 0 & ; x=5 \end{cases}$$

- $P(X=1) = p(1) = 0{,}0755$
- $P(X \geq 1) = 1 - P(X < 1)$
  $= 1 - [F(1) - p(1)]$

30/05/2025

- **Variables aleatorias continuas:** El recorrido siempre van a ser intervalos. X se dice una variable aleatoria continua si y solo si $P(X=x)=0$

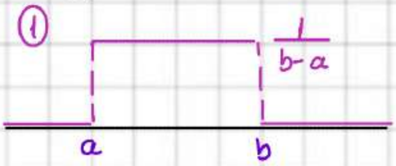
Función de densidad: $f(x)$, cumple que

1.- $f(x) \geq 0$

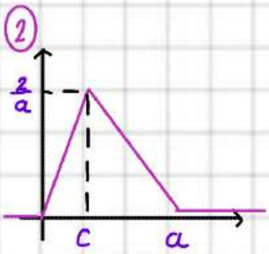
2.- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

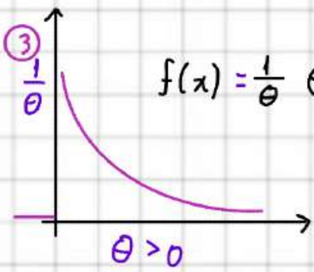
$\boxed{\text{cualquier función, cuya área bajo la curva es 1, es una función de densidad.}}$

ejemplo:

① 

$$f(x)=\begin{cases} \dfrac{1}{b-a} & ; a \leq x \leq b \\ 0 & ; \text{otros casos} \end{cases} \iff f(x)=\frac{1}{b-a} ; a \leq x \leq b$$

② 

③ 

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$$

Mostremos que es una densidad

P.D: $f(x) \geq 0 \rightarrow$ se cumple !!

P.D: $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{+\infty} f(x)\,dx$

$$= \int_{0}^{\infty} \frac{1}{\theta} e^{-\frac{x}{\theta}}\,dx$$

$$= \frac{1}{\theta} \int_{0}^{\infty} e^{-\frac{x}{\theta}}\,dx$$

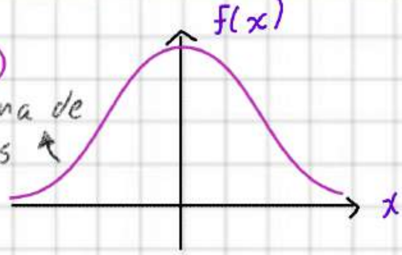
$$= \frac{1}{\theta} \frac{e^{-\frac{x}{\theta}}}{\left(-\frac{1}{\theta}\right)} \Bigg|_{0}^{\infty}$$

$$= -e^{-\frac{x}{\theta}} \Big|_{0}^{\infty}$$

$$= 0 + 1 = 1$$

Traer tabla de derivadas e integrales

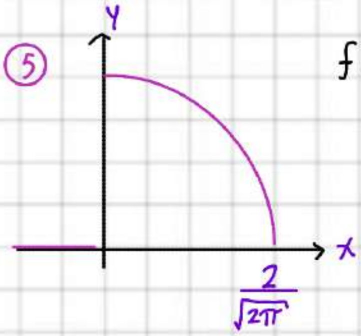
④ campana de Gauss



$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$ → Esta función no tiene primitiva!

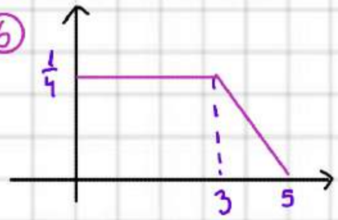
P.D. $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$

⑤



$$f(x) = \sqrt{\frac{4}{\pi} - x^2} \;;\; 0 \le x \le \frac{2}{\sqrt{\pi}}$$

⑥



$$f(x) = \begin{cases} \frac{1}{4} & ; 0 \le x \le 3 \\ -\frac{x}{8} + \frac{5}{8} & ; 3 \le x \le 5 \end{cases}$$

$y = mx + b$

$(3, \frac{1}{4}): \frac{1}{4} = 3m + b$

$(5, 0): \underline{0 = 5m + b}$

$-\frac{1}{4} = 2m$

$m = -\frac{1}{8}$; $b = \frac{5}{8}$

En la desigualdad, la igualdad puede o no estar incluida

Función de Distribución Acumulada: $F(X) = P(X = x) = \int_{-\infty}^{x} f(x)\,dx$

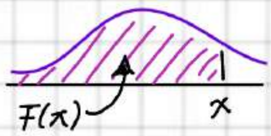
$F(X)$ cumple que:

1.- $F(-\infty) = 0$

2.- $F(+\infty) = 1$

3.- F es no decreciente

4.- F es continua

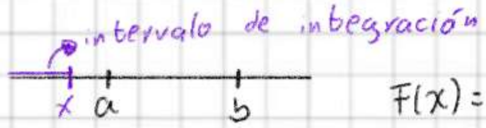


ejemplo:

(1) $f(x)=\dfrac{1}{b-a}$; $a\leq x\leq b$

Dado que f está dada por intervalos, entonces calculamos a F por intervalos.

$x<a$:



$$F(x)=\int_{-\infty}^{x}\cancelto{0}{f(x)\,dx}=0$$

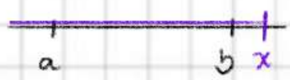
$a\leq x\leq b$:

$$F(x)=\int_{-\infty}^{x}f(x)\,dx=\int_{-\infty}^{a}\cancelto{0}{f(x)\,dx}+\int_{a}^{x}f(x)\,dx$$

$$=\frac{1}{(b-a)}\,x\Big|_{a}^{x}$$

$$=\frac{x-a}{b-a}$$

$b<x$:



$$F(x)=\int_{-\infty}^{x}f(x)\,dx=\int_{-\infty}^{a}\cancelto{0}{f(x)\,dx}+\int_{a}^{b}f(x)\,dx+\int_{b}^{\infty}\cancelto{0}{f(x)\,dx}$$

$$=\int_{a}^{b}f(x)\,dx$$

$$=\frac{1}{(b-a)}\,x\Big|_{a}^{b}=\frac{(b-a)}{(b-a)}=1$$

$$\Rightarrow F(x)=\begin{cases}0 & ; \; x<a\\ \dfrac{x-a}{b-a} & ; \; a\leq x\leq b\\ 1 & ; \; x>b\end{cases}\quad //$$

② $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$; $x \geq 0, \; \theta > 0$

Calculemos $F(x)$ por intervalos

$$x < 0: \; F(x) = \int_{-\infty}^{x} f(x)\,dx = 0$$

$$x \geq 0; \; F(x) = \int_{-\infty}^{x} f(x)\,dx = \cancel{\int_{-\infty}^{0} f(x)\,dx}^{\,0} + \int_{0}^{x} f(x)\,dx$$

$$= \int_{0}^{x} \frac{1}{\theta} e^{-\frac{x}{\theta}}\,dx$$

$$= \frac{1}{\theta} \int_{0}^{x} e^{-\frac{x}{\theta}}\,dx = 1 - e^{-\frac{x}{\theta}}$$

$$\Rightarrow F(x) = \begin{cases} 0 \; ; & x < 0 \\ 1 - e^{-\frac{x}{\theta}} \; ; & x \geq 0 \end{cases}$$

③ $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$; $-\infty < x < \infty$

$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}}\,dx$$

④ $f(x) = \begin{cases} \frac{1}{4} \; ; & 0 \leq x \leq 3 \\ -\frac{x}{8} + \frac{5}{8} \; ; & 3 \leq x \leq 5 \end{cases}$

Por intervalos:

$x < 0: \; F(x) = 0$

$0 \leq x \leq 3: \; F(x) = \frac{x}{4}$

$3 \leq x \leq 5: \; F(x) = -\frac{1}{16}x^2 + \frac{5}{8}x - \frac{9}{16}$

$x > 5: \; F(x) = 1$

$$\Rightarrow F(x) = \begin{cases} 0 & ; \; x < 0 \\ \frac{x}{4} & ; \; 0 \leq x \leq 3 \\ -\frac{1}{16}x^2 + \frac{5}{8}x - \frac{9}{16} & ; \; 3 \leq x \leq 5 \\ 1 & ; \; x < 5 \end{cases}$$

• $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{x} f(x)\,dx$

$$= \frac{1}{4} x \Big|_{0}^{x} = \frac{x}{4}$$

• $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{3} f(x)\,dx + \int_{3}^{x} f(x)\,dx$

$$= 0 + \frac{3}{4} + \left(-\frac{x^2}{16} + \frac{5}{8}x\right)\Big|_{3}^{x}$$

$$= \frac{3}{4} + \left[\left(-\frac{1}{16}x^2 + \frac{5}{8}x\right) - \left(-\frac{9}{16} + \frac{15}{8}\right)\right]$$

$$= -\frac{1}{16}x^2 + \frac{5}{8}x - \frac{9}{16}$$

Observación: Si para una variable aleatoria, es conocida la función de distribución, entonces la función de densidad está dada por:

$$f(x) = F'(x)$$

Observación: Para el cálculo de probabilidades en intervalos,

$$P(X \leq a) = P(X < a) = F(a)$$
$$P(X \geq a) = P(X > a) = 1 - F(a)$$
$$P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) = F(b) - F(a)$$

Para el cálculo de probabilidades se usa únicamenta la función de distribución.

ejemplo: Para el ejemplo anterior (4) calcular:

a) $P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - \frac{1}{4} = \frac{3}{4}$

b) $P(X \leq 2 \mid X > 1) = \dfrac{P(X \leq 2 \cap X > 1)}{P(X > 1)}$

$$= \frac{P(1 < X \leq 2)}{P(X > 1)} = \frac{F(2) - F(1)}{1 - F(1)} = \frac{\frac{2}{4} - \frac{1}{4}}{1 - \frac{1}{4}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

c) $P(1.5 \leq X \leq 3.8) = F(3.8) - F(1.5)$

$$= -\frac{1}{16}(3.8)^2 + \frac{5}{8}(3.8) - \frac{9}{16} - \frac{1}{4}(1.5) = 0,535$$

- **Parámetros:**

☆ Media/Valor Esperado/Esperanza: $\mu = E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$

☆ Varianza: $\sigma^2 = V[X] = Var[X] = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$

Observación: $\sigma^2 = E[X^2] - (E[X])^2$

Estos parámetros tienen las mismas propiedades que en los datos discretos.

02/06/2025

Leyes de Probabilidad

Modelos Teóricos de Distribuciones de Probabilidad

Discretos

Definición: Una prueba o experimento se dice que es de Bernoulli cuando solo se tienen solo 2 posibles resultados. A uno de estos resultados se le denomina éxito y su probabilidad se denota como p, al otro se le llama fracaso y su probabilidad se denota con $q = 1-p$.

ejemplo: Un insecticida tiene una efectividad del 92%. se aplica a un insecto.
Resultado:
- muere el insecto: Fracaso
- sobrevive el insecto: Éxito

$p = 0{,}08 \qquad q = 0{,}92$

- **Ley Binomial:** Una prueba de Bernoulli se la realiza n veces, se considera la variable aleatoria

X: número de éxitos en las n pruebas de Bernoulli. ~ Binomial

Rec X: $\{0, 1, \dots, n\}$

Función de Probabilidad: $P(X=k) = b(k, n, p)$

$x=0$: $P(X=0) = P(F \cap F \cap \dots \cap F) = q^n$

$x=1$: $P(X=1) = P(E \cap F \cap F \cap \dots \cap F) \cup P(F \cap E \cap F \cap \dots \cap F) \cup \dots \cup P(F \cap F \cap \dots \cap F \cap E)$

$= P(E \cap F \cap F \cap \dots \cap F) + \dots + P(F \cap F \cap \dots \cap F \cap E)$

$= p q^{n-1} + \dots + q^{n-1} p = n p q^{n-1}$

Otra forma: $P(X=1) = {}_nC_1 \, p^1 q^{n-1}$

En general: $P(X=k) = b(k, n, p) = \binom{n}{k} p^k q^{n-k} \; ; \; k = 0, 1, \dots, n$

Función de Distribución: $P(X \leq k) = B(k, n, p) = \sum_{i=0}^{k} b(i, n, p)$ Suele estar en tablas

Parámetros:
- $\mu = E[X] = np$
- $\sigma^2 = V[X] = npq$

efectividad del 80%

ejemplo: En el ejemplo del insecticida, se va a aplicar a 20 insectos. Cuál es la probabilidad de:

a) Sobrevivan al menos 3
b) sobrevivan a lo más 5
c) la esperanza

Tomamos la variable aleatoria

X = número de insectos que sobreviven entre los 20 ~ Binomial

función de probabilidad: $P(X=k) = b(k,n,p) = \binom{n}{k} p^k q^{n-k}$
función de distribución: $P(X\leq k) = B(k,n,p) = \sum_{i=0}^{k} b(i,n,p)$
Parámetros: $\mu = np$; $\sigma^2 = npq$

a) $P(X\geq 3) = 1 - P(X\leq 2) = 1 - B(2, 20, 0.20)$

$= 1 - [P(X=0) + P(X=1) + P(X=2)]$

$= 1 - [\binom{20}{0}0.20^0\ 0.80^{20} + \binom{20}{1}0.20^1\ 0.80^{19} + \binom{20}{2}0.20^2\ 0.80^{18}]$

$= 0.7939$

Tablas: $P(X\geq 3) = 1 - B(2, 20, 0.20) = 1 - 0{,}2061 = 0{,}7939$

b) $P(X\leq 5) = B(5, 20, 0.20) = 0{,}8042$

c) $E[X] = \mu - 20(0{,}20) = 4$

Observación: Para que una variable aleatoria siga la ley binomial se deben cumplir las siguientes condiciones

i) Población infinita.
ii) Las pruebas deben ser independientes.
iii) La probabilidad de éxito es constante.

ejemplo: Un examen consta de 15 preguntas de opción múltiple. Cada pregunta tiene 4 opciones de respuesta y solo 1 es verdadera. Si el examen se responde al azar, cuál es la probabilidad de:

a) Acierte a lo menos 5

b) Apruebe el examen → 60%

c) Acierte 6 preguntas

Éxito: aciertos → $p=0.25$

Sea X: # respuestas correctas, en las 15 preguntas ~ Binomial

a) $P(X \leq 5) = B(5, 15, 0.25) = 0{,}8516$

b) $P(X \geq 9) = 1 - P(X \leq 8) = 1 - B(8, 15, 0.25) = 1 - 0{,}9958 = 0{,}0042$

c) $P(X=6) = b(6, 15, 0.25) = \binom{15}{6} 0.25^6 \, 0.75^9 = 0{,}0917$

Tablas: $P(X=6) = B(6, 15, 0.25) - B(5, 15, 0.25) = 0{,}9434 - 0{,}8516 = 0{,}0917$

06/06/2025

- **Ley Hipergeométrica:** Población: N #éxitos en la población: r

X: #éxitos en una muestra de tamaño n

$$P(X=k)=h(k;n;r;N)=\frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}$$

$$P(X\leq k)=H(k,n,r,N)=\sum_{i=0}^{k} h(i,r,n,N)$$

$$\mu=E[X]=np \quad ; \quad \sigma^2=V[X]=npq\left(\frac{N-n}{N-1}\right) \quad \text{con } p=\frac{r}{n}$$

también se le llama Ley de Pascal

- **Ley Binomial Negativa:** Se intercambian los papeles de la ley Binomial.

Binomial	Binomial Negativa
# pruebas: n: fijo	# pruebas: variable
# éxitos: k: variable	# éxitos: fijo

Se considera la variable aleatoria
X: # intentos necesarios para lograr r éxitos. $\text{Rec } X=\{r, r+1, \dots\}$

Función de Probabilidad:

$$P(X=k)=\underbrace{\binom{k-1}{r-1}p^{r-1}q^{(k-1)-(r-1)}}_{\text{en los }(k-1)\text{ primeros intentos logró }(r-1)\text{ éxitos: binomial}}\cdot p$$

p → éxito en el último intento

$$\Rightarrow P(X=k)=\binom{k-1}{r-1}p^r q^{k-r}$$

$$P(X\leq k)=\sum_{i=0}^{k} P(X=i)$$

$$\mu=E[X]=\frac{r}{p} \quad ; \quad \sigma^2=V[X]=\frac{rq}{p^2}$$

ejemplo: Un basquetbolista encesta el 90% de las veces, ¿cuál es la probabilidad de que falle por 4^{ta} vez en el décimo quinto intento?

Éxito: falle en el lanzamiento : $p = 0.10$

número de éxitos : $r = 4$

X : # lanzamientos realizados para fallar 4 veces ~ Binomial Negativa

$$P(X=15) = \binom{4}{3} 0.10^4 \; 0.90^{11} = 0{,}001255$$

$$\mu = E[X] = \frac{r}{p} = \frac{4}{0.10} = 40$$

ejemplo: Cuando se realiza una prueba de laboratorio, la probabilidad de que sea exitosa es del 40%. El costo de los materiales es de \$10, si hay que repetir la prueba, el uso del laboratorio es de \$7.

a) ¿Cuál es la probabilidad de que en el 12^{vo} intento se obtengan 3 resultados éxitosos?

b) ¿Cuántos intentos se espera realizar para obtener los 3 resultados exitosos?

c) ¿Cuál es el costo que se esperaría se deba incurrir para lograr los 3 resultados exitosos?

Éxito: Resultado exitoso : $p = 0.40$

se desean 3 resultados exitosos : $r = 3$

X : # intentos para lograr 3 éxitos ~ Binomial Negativa

a) $P(X=12) = \binom{11}{2} 0.40^3 \; 0.60^9 = 0{,}0354$

b) $E[X] = \mu = \frac{r}{p} = \frac{3}{0.40} = 7.5$ → Este valor no se redondea

c) Costo : $C = (\text{valor cada intento}) \cdot (\#\text{intentos}) + (\text{Valor penalización})$

$= 10X + 7(X-1) = 17X - 7$

$E[C] = \mu_C = E[17X - 7] = 17\,E[X] - 7 = 17 \cdot 7.5 - 7 = 120{,}5$

ejemplo: Una determinada marca de gaseosa le gusta al 75% de personas. a) ¿Cuál es la probabilidad de que entre 10 personas seleccionadas al azar, al menos a 5 de ellas les guste esta marca? b) ¿Cuál es la probabilidad de qué a la quinta persona consultada sea la tercera persona a la que le gusta esta marca? c) Repita el literal si el número de personas a consultar es de 200.

Éxito: gusta marca NN : $p = 0.75$

a) X: # personas que les gusta la marca entre 10 seleccionados.
$\sim$ Binomial

$P(X \geq 5) = 1 - P(X \leq 4) = 1 - B(4; 10; 0.75) = 1 - 0{,}0197$

b) Y: # personas a consultar, por encontrar 3 a quienes les gusta la marca $\sim$ Binomial Negativa

$P(Y = 5) = \binom{4}{2} 0.75^3 \, 0.25^2 = 0{,}1582$

c) Población finita: $N = 200$; $r = 150$

U: # personas que les gusta la marca entre 10 seleccionados (Población finita) $\sim$ Hipergeométrica

$P(X \geq 5) = 1 - P(X \leq 4) = 1 - H(4; 10; 150; 200) \approx 1 - B($

Observación: Cuando en la ley Binomial Negativa se toma $r = 1$ (un éxito) se obtiene la llamada Ley Geométrica.

- Ley Geométrica:

X: # intentos para lograr 1 éxito $\sim$ Geométrica

$$P(X = k) = g(k; p) = p q^{k-1}$$

$$P(X \leq k) = G(k; p) = \sum_{i=0}^{k} g(i, p)$$

$$\mu = E[X] = \frac{1}{p} \; ; \; \sigma^2 = V[X] = \frac{q}{p^2}$$

- Ley de Poisson: Matemáticamente, esta ley se la obtiene como el límite de la ley Binomial cuando la probabilidad de éxito es tan pequeña que se necesitan "muchos" intentos.

☆ Binomial $\xrightarrow[p\approx 0]{n\to\infty}$ Poisson

Poisson: procesos de conteo cuando se conoce el promedio de ocurrencias en una determinada unidad de medida.

ejemplo: A una terminal llegan un promedio de 230 buses cada día.
X: # buses que llegan en un día

ejemplo: A un museo llegan en promedio 80 personas por día.

ejemplo: A un banco llegan en promedio de 20 clientes cada 30 min.

En general, se considera la variable aleatoria
X: # ocurrencias en la unidad de medida sabiendo que el promedio es λ ~ Poisson

Función de Probabilidad:
$$P(X=K)=f(K;\lambda)=\frac{e^{-\lambda}\lambda^{K}}{K!}\quad ; K=0,1,2,\ldots$$

Función de Distribución:
$$P(X\leq K)=F(K;\lambda)=\sum_{i=0}^{K} f(i,\lambda) \to \text{Tablas}$$

ejemplo: A la clase de probabilidad llegan en promedio 4 atrasados en cada clase. Cuál es la probabilidad de que a la próxima clase lleguen:
a) a lo más 5 atrasados.
b) 2 atrasados.

X: # alumnos atrasados a una clase ; $\lambda=4$ ~ Poisson

a) $P(X\leq 5)=P(X=5)+P(X=4)+\ldots+P(X=0)$
$$=\frac{e^{-4}4^{5}}{5!}+\frac{e^{-4}4^{4}}{4!}+\cdots+\frac{e^{-4}4^{0}}{0!}=0{,}7851$$

b) $P(X=2)=f(2;4)=\dfrac{e^{-4}4^{2}}{2!}=0{,}1465$

Con tablas:

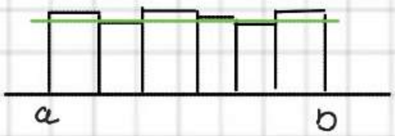
a) $P(X \leq 5) = F(5,4) = 0,7851$

b) $P(X = 2) = F(2,4) - F(1,4) = 0,2381 - 0,0916 = 0,1465$

09/06/2025

Modelos Teóricos Continuos

- **Ley Uniforme:** se utiliza cuando al trabajar con una muestra, el histograma de frecuencias es prácticamente el mismo valor para cada intervalo, de manera que el perfil de este histograma se puede ajustar con una línea horizontal.



v.a. → $X \sim \mathcal{U}[a;b]$

función de densidad: $f(x) = \frac{1}{b-a} \; ; a \leq x \leq b$

Función de distribución: $F(x) = \begin{cases} 0 & ; x < a \\ \frac{x-a}{b-a} & ; a \leq x \leq b \\ 1 & ; b < x \end{cases}$

$\mu = E[X] = \frac{a+b}{2}$

$$\mu = E[X] = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

$$= \int_{-\infty}^{a} x\, f(x) dx + \int_{a}^{b} x\, f(x) dx + \int_{b}^{+\infty} x\, f(x) dx$$

$$= \frac{1}{(b-a)} \frac{x^2}{2} \Big|_a^b$$

$$= \frac{1}{2} \frac{1}{(b-a)} \left[b^2 - a^2 \right]$$

$$= \frac{a+b}{2}$$

$\sigma^2 = V[X] = \frac{(b-a)^2}{12}$

$$\sigma^2 = E[X^2] - (E[X])^2$$

$$= \int_a^b x^2 \cdot \frac{1}{(b-a)} dx - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{1}{(b-a)} \frac{x^3}{3} \Big|_a^b - \frac{(a+b)^2}{4}$$

ejemplo: La hora de llegada a clases sigue una ley uniforme entre las 14:50 y las 15:05. Cuál es la probabilidad de que:

a) llegue a tiempo

b) llegue a tiempo sabiendo que hasta las 14:55 aún no llegaba

x: Hora de llegada $\sim U[2{:}50; 3{:}05]$ $\begin{cases} 2{:}50 \doteq 2.8333 \\ 3{:}05 \doteq 3.0833 \end{cases}$

- Función de distribución

$$F(x) = \begin{cases} 0 & ; \; x < 2.8333 \\ \dfrac{x - 2.8333}{0.25} & ; \; 2.8333 \leq x \leq 3.0833 \\ 1 & ; \; 3.0833 < x \end{cases}$$

$$E[x] = \frac{a+b}{2} = 2.9583 \doteq 2{:}5749 = 2h\ 57'\ 30''$$

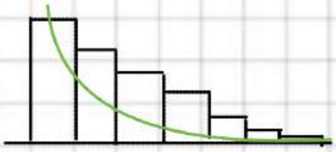
a) $P(x \leq 3{:}00) = P(x \leq 3) = F(3) = \dfrac{3 - 2.8333}{0.25} = 0.6667$

b) $P(x \leq 3{:}00 \mid x > 2{:}55) = P(x \leq 3 \mid x > 2.9167)$

$$= \frac{P(2.9167 < x \leq 3)}{1 - P(x \leq 2.9167)}$$

$$= \frac{F(3) - F(2.9167)}{1 - F(2.9167)}$$

- **Ley Exponencial:** $x \sim \mathcal{E}(\theta)$ Se usa cuando al realizar el histograma, la frecuencia va disminuyendo rápidamente.



F. densidad: $f(x) = \dfrac{1}{\theta} e^{-\frac{x}{\theta}} \; ; \; x \geq 0$

F. distribución: $F(x) = \begin{cases} 0 & ; \; x < 0 \\ 1 - e^{-\frac{x}{\theta}} & ; \; x \geq 0 \end{cases}$

$$\mu = E[x] = \int_{-\infty}^{+\infty} x f(x)\, dx = \int_0^{+\infty} x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = \frac{1}{\theta} \int_0^{+\infty} x e^{-\frac{x}{\theta}} dx = \frac{1}{\theta} \left[-\theta x e^{-\frac{x}{\theta}} + \theta \int e^{-\frac{x}{\theta}} dx \right] \Bigg|_0^{\infty}$$

$$= \frac{1}{\theta} \left[-\theta x e^{-\frac{x}{\theta}} + \theta \left(-\theta e^{-\frac{x}{\theta}} \right) \right] \Bigg|_0^{\infty} = [-x - \theta] e^{-\frac{x}{\theta}} \Bigg|_0^{\infty}$$

$$= \lim_{x \to \infty} (-x - \theta) e^{-\frac{x}{\theta}} + \theta = \lim_{x \to \infty} -\frac{x + \theta}{e^{\frac{x}{\theta}}} + \theta = \theta$$

$\mu = E[x] = \theta$ $\qquad\qquad$ $\sigma^2 = V[x] = \theta^2$

La distribución exponencial se usa para cálculos de tiempo de espera y de tiempo de vida útil.

Observación: La ley exponencial no tiene memoria.

Teorema: Sea $X \sim \mathcal{E}(\theta)$, entonces $P(X \geq s+t \mid X \geq s) = P(X \geq t)$

Demostración:

$$P(X \geq s+t \mid X \geq s) = \frac{P(X \geq s+t \cap X \geq s)}{P(X \geq s)}$$
$$= \frac{P(X \geq s+t)}{1 - P(X < s)}$$
$$= \frac{1 - P(X < s+t)}{1 - P(X < s)}$$
$$= \frac{1 - F(s+t)}{1 - F(s)} = \frac{1 - \left[1 - e^{-\frac{s+t}{\theta}}\right]}{1 - \left[1 - e^{-\frac{s}{\theta}}\right]}$$
$$= \frac{e^{-\frac{s+t}{\theta}}}{e^{-\frac{s}{\theta}}} = e^{-\frac{t}{\theta}}$$
$$= 1 - \left[1 - e^{-\frac{t}{\theta}}\right]$$
$$= 1 - F(t)$$
$$= 1 - P(X < t)$$
$$= P(X \geq t)_{//}$$

ejemplo: La duración del examen de estadística sigue una función exponencial. El 80% de los alumnos entregan antes de los 90 mimutos. Cuál es la probabilidad de que

a) entregue pasado los 80 minutos si a los 60 no entrega

b) entre 4 alumnos elegidos al azar, al menos 2 de ellos entreguen antes de los 80 minutos.

X: tiempo que tarda un alumno en realizar el examen $\sim \mathcal{E}(\theta)$

Sabemos que $P(X < 90) = 0.80$. Dado que $F(x) = 1 - e^{-\frac{x}{\theta}}$

$\Rightarrow F(90) = 0{,}80$

$1 - e^{-\frac{90}{\theta}} = 0{,}80$

$e^{-\frac{90}{\theta}} = 0{,}20$

$-\frac{90}{\theta} = \ln(0.20)$

$\theta = -\frac{90}{\ln(0.20)} \approx 55.92$ (min)

a) $P(x>80 \mid x>60) = P(x>20) = 1-P(x\leq 20)$

$= 1-F(20)$

$= 1-\left[1-e^{-\frac{20}{55.92}}\right] \approx 0.6993$

Definimos una nueva variable:

Y: # alumnos que entregan antes de los 80 minutos $\sim$ Binomial

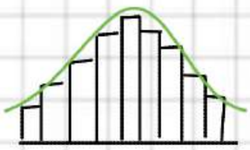
éxito: entrega antes de los 80 minutos ; $p = P(x<80) = 0.7609$

b) $P(Y\geq 2) = 1-P(Y\geq 1) = 1-[P(Y=0)+P(Y=1)]$

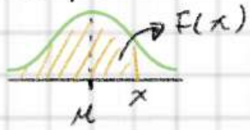
$= 1-\left[\binom{4}{0}0.7609^{0}\,0.2391^{4} + \binom{4}{1}0.7609^{1}\,0.2391^{3}\right]$

$= 0.9551$

- **Ley Normal:** $X \sim N(\mu;\sigma^2)$ Se la utiliza cuando el perfil del histograma se puede ajustar mediante una campana.



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



Distribución: $F(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$ → Esta integral no tiene primitiva, por lo que se la debe evaluar analíticamente usando métodos numéricos.

media: $E[x] = \mu$

varianza: $V[x] = \sigma^2$

ejemplo: La estatura de una mujer adulta está distribuida normalmente con una media de 155 cm y una desviación estándar de 8 cm.

a) Cuál es la prob. de que una mujer seleccionada al azar mida más de 170 cm.

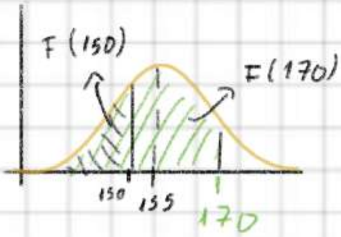
b) Cuál es el porcentaje de mujeres que miden menos de 150 cm.

$\mu = 155$ cm $\qquad \sigma = 8$ cm

Sea X: estatura mujer $\sim N(155; 64)$

Densidad: $f(x) = \frac{1}{\sqrt{16\pi}} e^{-\frac{(x-155)^2}{128}}$

Distribución: $F(x) = \frac{1}{\sqrt{16\pi}} \int_{-\infty}^{x} e^{-\frac{(x-155)^2}{128}} \, dx$



a) $P(X > 170) = 1 - P(X \leq 170)$
$= 1 - F(170)$

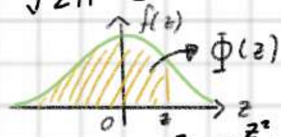
b) $P(X \leq 150) = F(150)$

Observación: en la tabla se observa que el valor acumulado hasta $z=5$ es $0.99999...$, lo que significa que la cola a la derecha de $z=5$ es prácticamente 0.
Para valores de $z<0$ se utiliza la simetría: $\Phi(-a) = 1 - \Phi(a)$.

Caso Particular: $\mu = 0$; $\sigma^2 = 1$
$\Rightarrow$ Ley Normal Estándar
Se usa la letra Z tal que
$Z \sim (0,1)$

distribución

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$; $-\infty < z < \infty$



tablas

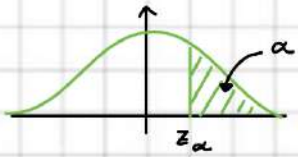
$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} \, dz$

Densidad

16/06/2025

Observación: Cuando se realiza estadística, se utiliza un valor de la ley normal estándar denotado como z_α que representa el valor de la variable Z en el cual el valor de la derecha tiene valor α.

$z_\alpha : P(Z > z_\alpha) = \alpha$



Dado un valor de α, para encontrar el correspondiente z_α, en la tabla se busca el valor complementario.

ejemplo: $\alpha = 2.5\% \Rightarrow \alpha = 0.025 \Rightarrow z_\alpha = z_{0.025} = 1.96$
$\alpha = 5\% \Rightarrow \alpha = 0.05 \Rightarrow z_\alpha = z_{0.05} = 1.645$
$\alpha = 1\% \Rightarrow \alpha = 0.01 \Rightarrow z_\alpha = z_{0.01} = 2.33$
$\alpha = 3\% \Rightarrow \alpha = 0.03 \Rightarrow z_\alpha = z_{0.03} = 1.88$

Teorema de estandarización: sea $X \sim N(\mu; \sigma^2)$, entonces

$$Z = \frac{X-\mu}{\sigma} \sim N(0;1)$$ → Proceso de centrar y reducir

Observación: $F(a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$

ejemplo: X: Estatura mujer $\sim N(155; 8^2)$

a) ¿Qué % mide menos de 150?
b) ¿Qué % mide más de 170?
c) ¿Cuál es la probabilidad de que mida 160?

a) $P(X<150) = P\left(\frac{X-\mu}{\sigma} < \frac{150-\mu}{\sigma}\right)$

$= P\left(Z < \frac{150-155}{\sigma}\right)$

$= P(Z < -0.625)$

$= \Phi(-0.625) = 1 - \Phi(0.625) \simeq 1 - \Phi(0.63) \simeq 1 - 0.7357 \simeq 0.2643$

b) $P(X>170) = 1 - P(X \leq 170)$

$= 1 - F(170)$

$= 1 - \Phi\left(\frac{170-155}{8}\right) = 1 - \Phi(1.88) = 1 - 0.9699 = 0.0301$

c) $P(X=160) = 0$ → la variable es continua, por lo que la probabilidad de que la variable tome un valor exacto es 0.

ejemplo: Suponga que la edad en un determinado grupo racial está normalmente distribuida. El 15% de la población tiene menos de 12 años; el 18% tiene más de 60.
a) ¿Cuál es la edad promedio de la población?
b) ¿Qué porcentaje tiene mas de 50 años?
c) ¿Qué porcentaje tiene entre 30 y 50 años?

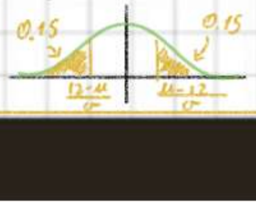
X: Edad $\sim N(\mu; \sigma^2)$ Datos: • $P(X<12) = 0.15$ (i) • $P(X>60) = 0.18$ (ii)

(i) $F(12) = 0.15 \iff \Phi\left(\frac{12-\mu}{\sigma}\right) = 0.15 \Rightarrow \frac{\mu-12}{\sigma} = z_{0.15} = 1.04$

(ii) $P(X \leq 60) = 0.82 \Rightarrow F(60) = 0.82 \Rightarrow \Phi\left(\frac{60-\mu}{\sigma}\right) = 0.82$

$\Rightarrow \frac{60-\mu}{\sigma} = z_{0.18} = 0.92$

Si el valor acumulado es menor al 50%, la variable es negativa.



de donde se tiene

$$\begin{cases} \mu - 12 = 1.04\sigma & (1) \\ 60 - \mu = 0.92\sigma & (2) \end{cases}$$

$(1)+(2) \Rightarrow 48 = 1.96\sigma$

$\boxed{\sigma = 24.49}$

$\boxed{\mu = 37.47}$

Distribución log-normal   Distribución de Weibul
Distribución gamma
Distribución beta

$$P(x>50) = 1 - P(x \leq 50)$$
$$= 1 - F(50)$$
$$= 1 - \Phi\left(\frac{50-37.4}{24.49}\right) = 1 - \Phi(0.51)$$
$$= 1 - 0.6950 = 0.3050$$

$$P(30 \leq x \leq 50) = F(50) - F(30)$$
$$= \Phi\left(\frac{50-37.47}{24.49}\right) - \Phi\left(\frac{30-37.47}{24.49}\right)$$

ejemplo: Sea X una variable normal, ¿cuál es la probabilidad de que la variable esté en:

a) $\mu \pm \sigma$?
b) $\mu \pm 2\sigma$?
c) $\mu \pm 3\sigma$?
d) Repetir el ejercicio para $X \sim \mathcal{E}(\theta)$

a) $\mu \pm \sigma : [\mu-\sigma ; \mu+\sigma]$

$$P(\mu-\sigma \leq x \leq \mu+\sigma) = F(\mu+\sigma) - F(\mu-\sigma)$$
$$= \Phi\left(\frac{\mu+\sigma-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-\sigma-\mu}{\sigma}\right)$$
$$= \Phi(1) - \Phi(-1)$$
$$= \Phi(1) - [1 - \Phi(1)]$$
$$= 2\Phi(1) - 1$$
$$= 2(0.8413) - 1 = 0.6826$$

b) $\mu \pm 2\sigma : [\mu-2\sigma ; \mu+2\sigma]$

$$P(\mu-2\sigma \leq x \leq \mu+2\sigma) = F(\mu+2\sigma) - F(\mu-2\sigma)$$
$$= \Phi\left(\frac{\mu+2\sigma-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-2\sigma-\mu}{\sigma}\right)$$
$$= \Phi(2) - \Phi(-2)$$
$$= 2\Phi(2) - 1$$
$$= 2(0.9772) - 1 = 0.9544$$

d) $x \sim \mathcal{E}(\theta)$; $\mu = \theta$; $\sigma^2 = \theta^2$; $F(x) = 1 - e^{-\frac{x}{\theta}}$; $x > 0$

$\mu \pm \sigma : [\mu-\sigma ; \mu+\sigma]$

$$P(\mu-\sigma \leq x \leq \mu+\sigma) = F(\mu+\sigma) - F(\mu-\sigma)$$
$$= F(\theta+\theta) - F(\theta-\theta)$$
$$= F(2\theta) - F(0)$$
$$= (1 - e^{-2}) - 0 = 1 - e^{-2} \simeq 0.8646$$

$\mu \pm 2\sigma : P(\mu-2\sigma \leq x \leq \mu+2\sigma) = F(\mu+2\sigma) - F(\mu-2\sigma)$

$$= F(\theta+2\theta) - F(\theta-2\theta) = F(3\theta) - \cancel{F(-\theta)}^{\,0}$$
$$= 1 - e^{-3} \simeq 0.9502$$

20/06/2025

Vectores Aleatorios

Cuando sobre un espacio muestral Ω se define 2 o más variables aleatorias, estas se las puede poner en un arreglo para formar un vector denominado vector aleatorio.

ejemplo: Lanzamiento de 2 dados. $\Omega = \{(1,1);(1,2);\ldots;(6,6)\}$

Sean las variables aleatorias

X: suma de los resultados $\quad Rec\,X = \{2,3,4,\ldots,12\}$

Y: producto de los resultados $\quad Rec\,Y = \{1,2,3,4,5,6,8,9,10,12,15,16,18,20,24,25,30,36\}$

Así, con estas variables, considerando los componentes de un vector: $(X;Y)$ se obtiene un vector aleatorio.

ejemplo: En un curso se va a formar una comisión de 2 alumnos, donde existen:

- 15 matemáticos
- 12 físicos
- 14 economistas

Sean las variables aleatorias

X: # matemáticos en la comisión

Y: # físicos en la comisión

Con estas v.a. formamos el vector $(X;Y)$ $\quad Rec(X;Y) = Rec\,X * Rec\,Y$

$Rec(X;Y) = \{(0,0);(0,1);(0,2);(1,0);(1,1);(1,2);(2,0);(2,1);(2,2)\}$

Para este tipo de vectores, el análisis es similar al de una variable aleatoria; se debe determinar la función de probabilidad (Probabilidad Conjunta), la función de distribución (Distribución Conjunta) y los parámetros: valor esperado (Valor esperado del vector) ^ varianza (Matriz de Varianzas-Covarianzas).

Consideramos un vector $(X;Y)$

- Probabilidad Conjunta: $p(x,y) = P(X=x; Y=y)$

debe cumplir:

i) $0 \le p(x;y) \le 1$

ii) $\sum p(x_i;y_j) = 1$

Observación: los valores de la función de probabilidad se suelen presentar en una tabla de doble entrada.

ejemplo: Ejemplo de la comisión.

$(0,0) = p(0,0) = P(X=0, Y=0) = \frac{{}_{14}C_2}{{}_{41}C_2} = 0.1109$

$(0,1) = p(0,1) = P(X=0, Y=1) = \frac{{}_{12}C_1 \cdot {}_{14}C_1}{{}_{41}C_2} = 0.2048$

$(0,2) = p(0,2) = P(X=0, Y=2) = \frac{{}_{12}C_2}{{}_{41}C_2} = 0.0805$

$(1,0) = p(1,0) = P(X=1, Y=0) = \frac{{}_{15}C_1 \cdot {}_{14}C_1}{{}_{41}C_2} = 0.2561$

$(1,1) = p(1,1) = P(X=1, Y=1) = \frac{{}_{15}C_1 \cdot {}_{12}C_1}{{}_{41}C_2} = 0.2195$

$(1,2) = p(1,2) = P(X=1, Y=2) = 0$

$(2,0) = p(2,0) = P(X=2, Y=0) = \frac{{}_{15}C_2}{{}_{41}C_2} = 0.1280$

$(2,1) = p(2,1) = P(X=2, Y=1) = 0$

$(2,2) = p(2,2) = P(X=2, Y=2) = 0$

Tabla de Probabilidades

		x			
		0	1	2	
y	0	0.1109	0.2561	0.1295	0.4965
	1	0.2048	0.2195	0	0.4243
	2	0.0805	0	0	0.0805
		0.3962	0.4756	0.1295	1

$P_y(y)$

$P_x(x)$

- Distribución Acumulada Conjunta: $F(x,y) = P(X \leq x ; Y \leq y)$
$= \sum_{y_i \leq y} \sum_{x_i \leq x} p(x_i ; y_i)$

Cuando se tiene el vector aleatorio, en algunos casos puede ser de interés analizar de manera separada a cada una de sus componentes. Las correspondientes funciones de probabilidad se denominan probabilidades marginales o función de probabilidad marginal.

En el ejemplo:

- $F(1.3 ; 1.8) = P(X \leq 1.3 ; Y \leq 1.8)$
$= p(0,0) + p(0,1) + p(1,0) + p(1,1)$
$= 0.1109 + 0.2048 + 0.2561 + 0.2195$
$= 0.7913$

Probabilidades marginales: se analiza por separado cada componente del vector.

X: $p_x(x) = P(X = x) = \sum_j p(x_i, y_j)$

Y: $p_y(y) = P(Y = y) = \sum_i p(x_i, y_j)$

Para el ejemplo:

X ; $Rec X = \{0, 1, 2\}$

$p_x(0) = P(X=0) = \sum_j (0, y_j)$

$= p(0,0) + p(0,1) + p(0,2)$

En la tabla de la probabilidad conjunta, los valores de las prob. marginales corresponden a las sumas de las filas y columnas.

Observación: Si no consideramos el vector aleatorio, sino únicamente a la variable X,

X: $p_x(0) = P(X=0) = \frac{{}_{15}C_0 \cdot {}_{26}C_2}{{}_{41}C_2}$

ejemplo de la comisión

Si se tiene la prob. conjunta, se puede determinar las prob. marginales, pero el recíproco no siempre es posible.

Solo es posible cuando las v.a. son independientes

Se dice que 2 variables $X \wedge Y$ son independientes cuando el valor de una variable no aporta información sobre el valor que tomará otra variable.

Definición: $X \wedge Y$ se dicen independientes si y solo si $p(x; y) = p_x(x) \cdot p_y(y)$.

Distribución multinomial: Una población se dice que es multinomial cuando cada elemento puede ser clasificado en una de k categorías. Si se toma una muestra de tamaño n y se consideran las variables aleatorias

X_1: # elementos de la muestra en la categoría 1.
X_2: # elementos de la muestra en la categoría 2
$\vdots$
X_k: # elementos de la muestra en la categoría k

se forma el vector aleatorio: $(X_1, X_2, \ldots, X_k)$.

☆ **Probabilidad Conjunta:**

$$p(x_1, x_2, \ldots, x_k) = P(X_1 = x_1;\ X_2 = x_2;\ \ldots;\ X_k = x_k) = \frac{n!}{x_1!\, x_2! \ldots x_k!}\, p_1^{x_1} \cdot p_2^{x_2} \ldots p_k^{x_k}$$

donde: p_i: prob. de estar en la i-ésima categoría.

ejemplo: Suponga que en una institución, el 25% de los profesores están calificados como excelentes, 40% como buenos, 20% como regulares y 15% como pésimos. Se va a tomar una muestra de 70 profesores. ¿Cuál es la probabilidad de que entre los 70 hayan 20 excelentes, 35 buenos, 10 regulares y 5 pésimos?

X_i: # profesores en la categoría i; $i = 1, 2, 3, 4$
Vector: (X_1, X_2, X_3, X_4)

$$p(20, 35, 10, 5) = P(X_1 = 20,\ X_2 = 35,\ X_3 = 10,\ X_4 = 5)$$
$$= \frac{70!}{20!\, 35!\, 10!\, 5!}\, 0.25^{20}\, 0.40^{35}\, 0.20^{10}\, 0.15^{5}$$

Lo anterior es un análisis con variables discretas. Cuando las variables son continuas, el procedimiento es similar.

bidimensional

Vector aleatorio continuo: Densidad Conjunta: $f(x;y)$ satisface que
i) $f(x,y) \geq 0$ ii) $\iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = 1$

Distribución conjunta: $F(x;y) = P(X \leq x;\, Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x;y)\,dy\,dx$

Densidades marginales:
- X: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$
- Y: $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

$X \wedge Y$ se dicen independientes ssi $f(x,y) = f_X(x) \cdot f_Y(y)$

Distribuciones marginales:
- $F_X(x) = \int_{-\infty}^{\infty} F(x,y)\,dy$
- $F_Y(y) = \int_{-\infty}^{\infty} F(x,y)\,dx$

$X \wedge Y$ se dicen independientes ssi $F(x,y) = F_X(x) \cdot F_Y(y)$

23/06/2025

ejemplo: Sea $(X;Y)$ un vector aleatorio de componentes geométricamente distribuidas con igual probabilidad de éxito e independientes. Sea la variable $Z = X + Y$. Encuentre la función de probabilidad de Z.

- X: geom: $P(X=x) = p^1 q^{x-1} = p_X(x)$; $Rec\,X = \{1, 2, \ldots\}$
- Y: geom: $P(Y=y) = p^1 q^{y-1} = p_Y(y)$; $Rec\,Y = \{1, 2, \ldots\}$

Por independencia: $p(x;y) = p_X(x) \cdot p_Y(y) = p^1 q^{x-1}\, p^1 q^{y-1} = p^2 q^{x+y-2}$

Sea $Z = X + Y \Rightarrow Rec\,Z = \{2, 3, \ldots\}$

- Función de probabilidad

$$
\begin{aligned}
P(Z=z) &= P(X+Y=z) \\
&= P(X=x;\, Y=z-x) \\
&= p(x;\, z-x) \\
&= p^2 q^{x+(z-x)-2} \\
&= p^2 q^{z-2} \;;\; z = 2, 3, 4, \ldots
\end{aligned}
$$

Vector aleatorio continuo: Sean X, Y variables aleatorias continuas $\Rightarrow (X, Y)$: vector aleatorio bidimensional continuo.

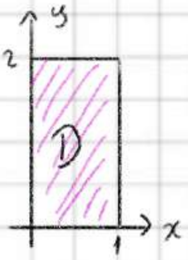
Densidad Conjunta: $f(x;y)$ cumple que: i) $f(x;y) \geq 0$ ii) $\iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = 1$

Distribución conjunta: $F(x;y) = P(X \leq x; Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x;y)\,dx\,dy$

Densidades marginales: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$

ejemplo: Sea la función $f(x;y) = x^2 + \frac{xy}{3}$; $0 \leq x \leq 1$; $0 \leq y \leq 2$

a) Pruebe que es una densidad.
b) Halle las probabilidades marginales.
c) Encuentre la función de distribución conjunta.
d) Determine si son independientes.



i) $(x;y) \in D : f(x,y) \geq 0$

ii) $\iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx$

$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{0} \overset{0}{\cancel{f(x,y)}}\,dy + \int_{0}^{2} f(x,y)\,dy + \int_{2}^{\infty} \overset{0}{\cancel{f(x,y)}}\,dy\right] dx$

$= \int_{0}^{1}\int_{0}^{2} f(x,y)\,dy\,dx$

$= \int_{0}^{1} \left(x^2 y + \frac{xy^2}{6} \right)\Big|_{0}^{2} dx$

$= \int_{0}^{1} \left(2x^2 + \frac{2x}{3} \right) dx = \left(\frac{2}{3}x^3 + \frac{x^2}{3} \right)\Big|_{0}^{1} = \frac{2}{3} + \frac{1}{3} - 0$

$= 1$

Densidades Marginales:

- $f_x(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^2 f(x,y)\,dy = \int_0^2 \left(x^2 + \frac{xy}{3}\right) dy = \left(x^2 y + \frac{xy^2}{6}\right)\Big|_0^2$

$$= 2x^2 + \frac{2}{3}x \; ; \; 0 \le x \le 1$$

- $f_y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^1 f(x,y)\,dx = \int_0^1 \left(x^2 + \frac{xy}{3}\right) dx = \left(\frac{1}{3}x^3 + \frac{x^2 y}{6}\right)\Big|_0^1$

$$= \frac{1}{3} + \frac{y}{6} \; ; \; 0 \le y \le 2$$

Distribución conjunta

$$F(x,y) = P(X \le x ; Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)\,dy\,dx = \int_0^x\int_0^y \left(x^2 + \frac{xy}{3}\right) dy\,dx$$

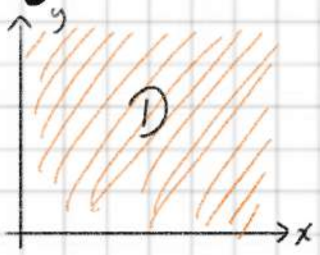
$$= \int_0^x \left(x^2 y + \frac{xy^2}{6}\right)\Big|_0^y dx = \int_0^x \left(x^2 y + \frac{xy^2}{6}\right) dx$$

$$= \left(\frac{x^3 y}{3} + \frac{x^2 y^2}{12}\right)\Big|_0^x$$

$$= \frac{x^3 y}{3} + \frac{x^2 y^2}{12} \; ; \; \begin{array}{l} 0 \le x \le 1 \\ 0 \le y \le 2 \end{array}$$

- No son independientes, pues en este caso $f_x(x) \cdot f_y(y) \neq F(x,y)$

ejemplo: $f(x,y) = \frac{1}{6} e^{-\frac{x}{2} - \frac{y}{3}} \; ; \; x \ge 0 ; \; y \ge 0$



i) la función exponencial nunca es negativa

ii) $\iint_{\mathbb{R}^2} f(x,y)\,dy\,dx = \int_0^{\infty}\left[\int_0^{\infty} \frac{1}{6} e^{-\frac{x}{2} - \frac{y}{3}}\,dy\right] dx$

$$= \int_0^{\infty}\left[\frac{1}{6}\int_0^{\infty} e^{-\frac{x}{2}} e^{-\frac{y}{3}}\,dy\right] dx$$

$$= \int_0^{\infty}\left[\frac{1}{6} e^{-\frac{x}{2}} \int_0^{\infty} e^{-\frac{y}{3}}\,dy\right] dx$$

$$= \frac{1}{6}\int_0^{\infty} e^{-\frac{x}{2}} \left(-3 e^{-\frac{y}{3}}\Big|_0^{\infty}\right) dx$$

$$= \frac{1}{2}\int_0^{\infty} e^{-\frac{x}{2}}\,dx = \frac{1}{2}(-2)\,e^{-\frac{x}{2}}\Big|_0^{\infty} = 1$$

Densidades Marginales:

- $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^{\infty} \frac{1}{6} e^{-\frac{x}{2}-\frac{y}{3}}\,dy$

$$= \frac{1}{6} e^{-\frac{x}{2}} \int_0^{\infty} e^{-\frac{y}{3}}\,dy = \frac{1}{6} e^{-\frac{x}{2}} \left(-3 e^{-\frac{y}{3}}\right)\Big|_0^{\infty} = \frac{1}{2} e^{-\frac{x}{2}} ; x \geq 0$$

- $f_Y(y) = \frac{1}{3} e^{-\frac{y}{3}} , y \geq 0$

$\Rightarrow X \wedge Y$ son independientes, pues:

$$f_X(x) \cdot f_Y(y) = \left(\frac{1}{3} e^{-\frac{y}{3}}\right)\left(\frac{1}{2} e^{-\frac{x}{2}}\right) = \frac{1}{6} e^{-\frac{x}{2}-\frac{y}{3}} = f(x;y)$$

Probabilidades Condicionales

Cuando se tiene de dato que una de las variables ha tomado un valor específico, a esto se lo toma como condición, y se obtiene una probabilidad condicional.

- Para el caso discreto: Si $Y = y_0 \Rightarrow P(X = x \mid Y = y_0) = p(x \mid y_0)$

$$\Rightarrow p(x \mid y_0) = P(X = x \mid Y = y_0) = \frac{P(X = x; Y = y_0)}{P(Y = y_0)} = \frac{p(x; y_0)}{P_Y(y_0)}$$

$$\boxed{p(x \mid y_0) = \frac{p(x; y_0)}{P_Y(y_0)} \quad \wedge \quad p(y \mid x_0) = \frac{p(x_0; y)}{P_X(x_0)}}$$

Observación: $p(x \mid y) = P(X = x \mid Y = y) = \frac{p(x;y)}{P_Y(y)} \wedge p(x \mid y) = P(Y = y \mid X = x) = \frac{p(x;y)}{P_X(x)}$

Observación: En el caso de variables continuas, las funciones de probabilidad condicionada se obtienen de manera similar a las discretas, obteniéndose:

$$f(x \mid y) = \frac{f(x;y)}{f_Y(y)} \quad \wedge \quad f(y \mid x) = \frac{f(x;y)}{f_X(x)}$$

discretas

Observación: $X \wedge Y$ son independientes $\Leftrightarrow p(x \mid y) = p_X(x) \wedge p(y \mid x) = p_Y(y)$

$X \wedge Y$ son independientes $\Leftrightarrow f(x \mid y) = f_X(x) \wedge f(y \mid x) = f_Y(y)$

continuas

ejemplo: $f(x,y) = x^2 + \frac{xy}{3}$; $0 \le x \le 1$

- $f_x(x) = 2x^2 + \frac{2x}{3}$
- $f_y(y) = \frac{1}{3} + \frac{y}{6}$

$$f(x|y) = \frac{f(x,y)}{f_y(y)} = \frac{x^2 + \frac{xy}{3}}{\frac{1}{3} + \frac{y}{6}} \qquad f(y|x)$$

Características de un vector aleatorio

Valor Esperado: $E[(x;y)] = (E[x];E[y])$

> **Observación:** el valor esperado es lineal.
> - $E[aX \pm bY] = aE[X] \pm bE[Y]$
>
> si $X \wedge Y$ son indep.
> $\Rightarrow E[X;Y] = E[X]\,E[Y]$

Matriz de Varianzas-Covarianzas:

Varianza: $V[x] = \sigma_x^2 = E[(X - E[x])^2]$

Covarianza: $cov(x;y) = E[(X - E[x])(Y - E[x])]$

$$cov(x;y) = \begin{cases} \sum_i \sum_j (x_i - \mu_x)(y_j - \mu_y)\, p(x_i, y_j) \ ; & \text{discreto} \\ \iint_{\mathbb{R}^2} (X - E[x])(Y - E[y])\, f(x,y)\, dx\, dy \ ; & \text{continuo} \end{cases}$$

Matriz de Varianzas-Covarianzas:

$$M = \begin{pmatrix} V[x] & cov(x;y) \\ cov(x;y) & V[y] \end{pmatrix}$$

> **Observación:**
> 1. $cov(X;Y) = cov(Y;X)$
> 2. $cov(x;y) = E[XY] - E[X]\,E[Y]$
> 3. si $X \wedge Y$ son independientes $\Rightarrow cov(X,Y) = 0$, pero el recíproco no es cierto.
> 4. Si $X \wedge Y$ son independientes $\Rightarrow V[aX \pm bY] = a^2 V[X] + b^2 V[Y]$
> 5. $V[X+Y] = V[X] + V[Y] + 2\,cov(X,Y)$

> **Observación:** para medir el grado de relación lineal entre 2 variables, se utiliza la correlación. Siempre es un número en el intervalo $[-1,1]$. Cuando valo 0 significa que no tienen ningún tipo de relación lineal.
>
> $$\rho(X;Y) = \frac{cov(x;y)}{\sqrt{V[x]\,V[y]}}$$

Transformación de un Vector Aleatorio

Cuando se tiene una transformación de un vector aleatorio $(x;y)$ se va a transformar en un vector $(U;V)$ mediante ecuaciones que dependan de $x \wedge y$.

$$(x;y) \longrightarrow (U;V)$$ → Estas transformaciones deben ser inversibles

$$\begin{cases} U = U(x;y) \\ V = V(x;y) \end{cases} \Rightarrow \begin{cases} x = x(U;V) \\ y = y(U;V) \end{cases}$$

Dominio $(x;y) = D$ → plano xy

si $f(x,y)$ = densidad conjunta de $(x;y)$

$\Rightarrow g(U;V) = f(x(U;V);y(U;V)) \cdot |J|$ = densidad conjunta de $(U;V)$

donde: $J = \begin{vmatrix} \frac{\partial x}{\partial U} & \frac{\partial x}{\partial V} \\ \frac{\partial y}{\partial U} & \frac{\partial y}{\partial V} \end{vmatrix}$ → Matriz Jacobiana

Jacobiano

El nuevo dominio para el vector (U,V), D^* (→ plano UV), se lo determina al ver en qué se transforman los elementos de la frontera (lo que limita o delimita el dominio original).

27/06/2025

ejemplo: Sean $X \wedge Y$ v.a. independientes y exponencialmente distribuidas con $\mu=1$. Considere el valor aleatorio (x,y).

a) Encuentre la función de densidad conjunta con su dominio

b) Encuentre la función de distribución conjunta.

c) $P(x+y \leq 1)$

d) Considere la transformación $U = x+y \wedge V = \frac{x}{x+y}$ y encuentre su densidad conjunta con su dominio.

$x \sim \mathcal{E}(1)$, $y \sim \mathcal{E}(1)$: v.a. independientes

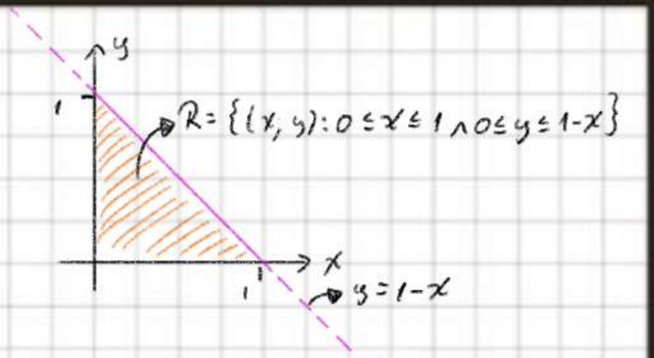
a) $f_x(x) = e^{-x}$; $x \geq 0$ $\qquad f_Y(y) = e^{-y}$; $y \geq 0$

Por independencia: $f(x;y) = f_x(x) \cdot f_y(y) \Rightarrow f(x,y) = e^{-(x+y)}$; $x \geq 0$, $y \geq 0$

b) $F_x(x) = 1 - e^{-x}$; $x \geq 0$ $\qquad F_Y(y) = 1 - e^{-y}$; $y \geq 0$

Por independencia: $F(x,y) = F_x(x) \cdot F_y(y) \Rightarrow F(x,y) = (1-e^{-x})(1-e^{-y})$; $x \geq 0$, $y \geq 0$

c) $P(X+Y \le 1) = P(Y \le 1-X)$
$= P((X;Y) \in R)$
$= \iint_{\mathbb{R}^2} f(x,y)\, dx\, dy$
$= \int_0^1 \int_0^{1-x} f(x,y)\, dy\, dx$



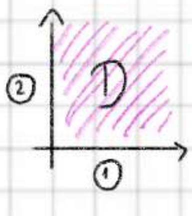
d) $\begin{cases} U = X + Y \\ V = \dfrac{X}{X+Y} \end{cases} \Rightarrow \begin{cases} X = UV \\ Y = U - UV \end{cases}$

Cálculo del Jacobiano: $J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix}$

$= -uv - u(1-v)$
$= -uv - u + uv$
$= -u \Rightarrow |J| = u$

Nuevo dominio (para el vector (U,V)):

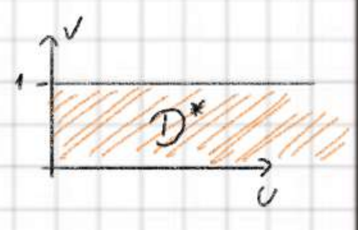
Dominio original: $x \ge 0$, $y \ge 0$ $\Rightarrow U = X+Y \ge 0$



- Frontera: $Y=0$, $x \ge 0$ ① $\wedge$ $x=0$, $y \ge 0$ ②
  - En la transformación:
    $U = X+Y$: ①: $U = X + 0 = X \ge 0$
    ②: $U = 0 + Y = Y \ge 0$

$\Rightarrow 0 \le X \le X + Y$ $\qquad (x+y) = 0 \le \dfrac{x}{x+y} \le 1$

$\Rightarrow U \ge 0;\ 0 \le V \le 1$



$\Rightarrow$ El nuevo dominio (D^*) es $[0;+\infty) * [0;1]$

Densidad conjunto para (U,V): $g(u,v) = f(x(u,v); y(u,v))\,|J|$
$= e^{-uv-(u-uv)} \cdot u = u\, e^{-u}$

Densidades Marginales:

Para V:
$$g_V(v) = \int_{-\infty}^{\infty} g(u,v)\,du = \int_{0}^{\infty} u e^{-u}\,du$$
$$= \left(-u e^{-u} - e^{-u}\right)\Big|_0^{\infty}$$
$$= \lim_{u\to\infty}\left[-\cancel{\frac{u}{e^{u}}}^{\,0} - \cancel{e^{-u}}^{\,0}\right] - (0-1)$$
$$= 1$$

Para U:
$$g_U(u) = \int_{-\infty}^{\infty} g(u,v)\,dv = \int_{0}^{1} u e^{-u}\,dv$$
$$= u v e^{-u}\Big|_0^{1} = u e^{-u}, \; u \geq 0$$

ejemplo: En una empresa los ingresos y los egresos son aleatorios. Los ingresos anuales se pueden modelizar mediante una exponencial de $\theta = 2$ (millones) y los egresos con una exponencial de $\theta = 1$ (millón). Para un análisis financiero, se suele analizar el índice $\frac{\text{ingresos}}{\text{egresos}}$. Describa probabilísticamente este índice.

- ingresos: $X \sim \mathcal{E}(2)$ $\Rightarrow f_X(x) = \frac{1}{2} e^{-\frac{x}{2}}\; ; x \geq 0$
- egresos: $Y \sim \mathcal{E}(1)$ $\Rightarrow f_Y(y) = e^{-y}\; ; y \geq 0$
- índice: $\frac{X}{Y}$: descripción

Si $X \wedge Y$ son independientes $\Rightarrow$ densidad conjunta $= f_X(x) \cdot f_Y(y)$

$$f(x;y) = \frac{1}{2} e^{-\frac{x}{2} - y} \; ; x \geq 0, y \geq 0$$

Dado que el vector V no está especificado, lo podemos tomar libremente.

Sean $U = \frac{X}{Y} \wedge V = Y$

Transformación:
$$\begin{cases} U = \frac{X}{Y} \\ V = Y \end{cases} \Rightarrow \begin{cases} X = UV \\ Y = V \end{cases}$$

Jacobiano:
$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

- nuevo dominio para el vector $(U;V)$: $v \geq 0$, $u \geq 0$

Densidad conjunta: $g(u,v) = f(x(u,v);\, y(u,v)) \cdot |J|$

$$= \frac{1}{2} v\, e^{-\frac{uv}{2} - v}$$

- densidad marginal de U

$$g_U(u) = \int_{-\infty}^{\infty} g(u,v)\, dv = \frac{1}{2} \int_0^{\infty} v\, e^{-\frac{uv}{2} - v}\, dv$$

30/06/2025

Estadística Inferencial

Población: X: variable de estudio **Parámetros:** media: μ; varianza: σ^2 (desconocidos)
Distribución de probabilidad
Muestra: Tamaño n **Datos:** $x_1, x_2, \dots, x_n$

Medidas estadísticas: $\begin{cases} \text{Media muestra}: \bar{x} \\ \text{Varianza muestra}: s^2 \end{cases} \Rightarrow$ Valores de v.a.
- Media muestral: $\bar{X}$
- Varianza muestral: S^2

Las distribuciones de probabilidad de estas medidas estadísticas se denominan distribuciones de muestreo y son las que permiten realizar las inferencias estadísticas: encontrar valores estimados de los parámetros de la población.

Teorema del Límite Central (TLC):

Si de una población infinita se toma una muestra aleatoria de tamaño n, la media de la muestra, $\bar{x}$ es un valor de la variable aleatoria media muestral: $\bar{X}$; para lo cuál se tiene que el valor esperado es $E[\bar{X}] = \mu$ y la varianza es $V[\bar{X}] = \sigma^2/n$. Además, cuando n es suficientemente grande, la media muestral $\bar{X}$ tiene comportamiento asintóticamente normal.

Observación: en la parte práctica, se considera que n es grande si $n \geq 30$.

$n \to \infty$:
$\bar{X} \sim N(\mu; \frac{\sigma^2}{n})$

$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0;1)$

ejemplo: En una población se define una variable aleatoria X con $Rec\, X = \{20, 45, 72\}$
a) Describa la variable aleatoria
b) Suponga que se van a tomar muestras de tamaño 2.
c) Repita el literal anterior con muestras de tamaño 3.

$$p(x) = P(X = x) = \begin{cases} \frac{1}{3}; & x = 20 \\ \frac{1}{3}; & x = 45 \\ \frac{1}{3}; & x = 72 \end{cases}$$

$$\mu = E[X] = \sum_i p(x_i) = \frac{137}{3}$$

$$\sigma^2 = V[X] = E[X^2] - (E[X])^2 = \frac{4058}{9}$$

b) muestra: $n=2$

Posibles muestras	Media de la muestra
(20; 20)	40/2
(20; 45)	65/2
(20; 72)	92/2
(45; 20)	65/2
(45; 45)	90/2
(45; 72)	117/2
(72; 20)	92/2
(72; 45)	117/2
(72; 72)	144/2

- Variable aleatoria $\bar{X}$: media muestral
  $Rec\bar{X} = \{\frac{40}{2}; \frac{65}{2}; \frac{90}{2}; \frac{92}{2}; \frac{117}{2}; \frac{144}{2}\}$
- función de probabilidad:

$$p(\bar{x}) = P(\bar{X}=\bar{x}) = \begin{cases} \frac{1}{9}; & \bar{x} = 40/2 \\ \frac{2}{9} & \bar{x} = 65/2 \\ \frac{1}{9} & \bar{x} = 90/2 \\ \frac{2}{9} & \bar{x} = 92/2 \\ \frac{2}{9} & \bar{x} = 117/2 \\ \frac{1}{9} & \bar{x} = 144/2 \end{cases}$$

- $E[\bar{x}] = \mu_{\bar{x}} = \frac{137}{3}$
- $V[\bar{x}] = \sigma^2_{\bar{x}} = \frac{\sigma^2}{2} = \frac{2029}{2}$

ejemplo: Una máquina envasadora de refrescos envasa frascos con una media de 16 onzas y una varianza de 0.8 onzas^2. Se va a tomar una muestra con tamaño $n=50$ frascos. ¿Cuál es la probabilidad de que la media de esta muestra sea:

a) mayor a 15.9 ?

b) menor a 16.2 ?

X: cantidad de llenado $\mu = 16\ (oz)$; $\sigma^2 = 0.8(oz^2)$ Muestra: $n=50$

Datos: $x_1, x_2, \ldots, x_{50}$ media de la muestra: $\bar{x} = \frac{1}{50}\sum_i x_i$

Por TLC, dado que $n=50$ es "grande"

$$E[\bar{x}] = \mu = 16 \qquad V[\bar{x}] = \frac{\sigma^2}{n} = \frac{0.8}{50}$$

$$\bar{x} \sim N(16; \frac{0.8}{50})$$

$$\Rightarrow P(\bar{x} > 15.8) = 1 - P(\bar{x} \leq 15.8)$$
$$= 1 - F(15.8)$$
$$= 1 - \Phi\left(\frac{15.8 - 16}{\sqrt{0.8}/\sqrt{50}}\right)$$
$$= 1 - \Phi(-1.58)$$
$$= 1 - [1 - \Phi(1.58)]$$
$$= 0.9429$$

$P(\bar{x} < 16.2)$

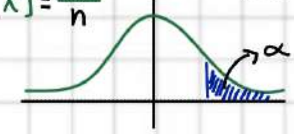
a_1 **Distribución de muestreo para media muestral:** cuando vamos a trabajar con una población infinita, de varianza σ^2 conocida y n grande, por TLC $\bar{x} \sim N(\mu; \sigma^2/n) \Leftrightarrow Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0; 1)$.

$V[\bar{x}] = \frac{\sigma^2}{n}$

a_2 Cuando la población es infinita, con varianza σ^2 desconocida y/o muestra pequeña, entonces

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim \text{t-student con } (n-1) \text{ grados de libertad.}$$

$V[\bar{x}] = \frac{s^2}{n}$



Se requiere que la población sea normal.

La distribución de probabilidad t-estudent tiene comportamiento similar a la normal estándar. Su gráfica tiene forma de campana.

Observación: los grados de libertad se refiere a lo siguiente:
Si se tiene una muestra de tamaño n y se conoce su media, se pueden asignar libremente $n-1$ valores de la variable. El n-ésimo dato se calcula con los datos asignados.

ejemplo: $n = 6$; $\bar{x} = 2$

Se pueden asignar libremente $n-1=5$ valores a la variable:

$10, 12, -7, 25, -12 \quad x_6$

$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5$

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$2 = \frac{1}{6}(10 + 12 - 7 + 25 - 12 + x_6) \Rightarrow x_6 = -16$$

b_1 Cuando se toma una población finita N, varianza σ^2 conocida, n grande, entonces

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \sim N(0;1) \qquad E[\bar{x}] = \mu$$

$$V[\bar{x}] = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$ (factor de corrección

b_2 Si se toma una población finita N, varianza σ^2 desconocida y/o muestra pequeña

$$T = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \sim t\text{-student con } (n-1) \text{ g.l.}$$

varianza de la población

ejemplo: Supongamos que en la politécnica hay 8000 alumnos y se está analizando la repitencia. Se conoce que en promedio la repitencia es 4.1 (# de materias) con $\sigma = 1.3$. Se va a tomar una muestra de 45 alumnos. ¿Cuál es la probabilidad de que la media de esta muestra sea menor a 4?

Población: estudiantes EPN . $N = 8000$ $\rightarrow$ Población finita

Variable de estudio: X : repitencia (# materias)

$\mu = 4.1$; $\sigma = 1.3$

Muestra: $n = 45$

datos: $x_1, x_2, \dots, x_{45}$ media de la muestra: $\bar{x} = \frac{1}{45}(x_1 + x_2 + \dots + x_{45})$

$P(\bar{x} < 4) = ?$

$$\Rightarrow Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \sim N(0,1) \Rightarrow E[\bar{x}] = \mu = 4.1$$

$$V[\bar{x}] = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{1.3^2}{45}\left(\frac{8000-45}{8000-1}\right) = 0.0373$$

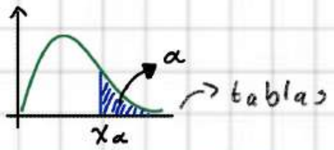
$$\bar{x} \sim N(4.1;\ 0.0373) \Leftrightarrow Z = \frac{\bar{x}-4.1}{\sqrt{0.0373}} \sim N(0;1)$$

$$P(\bar{x} < 4) = F(4) = \Phi\left(\frac{4-4.1}{\sqrt{0.0373}}\right) = \Phi(-0.52) = 1 - \Phi(0.52) = 1 - 0.6985 = 0.3015$$

07/07/2025 ¿Qué es un estimador de máxima verosimilitud?

Distribución de muestreo de $V[x] = S^2$: no está estandarizada, no está transformada

★ **Población normal:** la variable aleatoria $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ sigue una distribución llamada Chi-cuadrada (o Ji-cuadrada) con $(n-1)$ g.l.



Estimador de Parámetros:

Definición: Un estimador de un parámetro θ es una medida estadística que permite estimar el parámetro. $\theta \rightarrow$ parámetro a estimar $\hat{\theta} \rightarrow$ estimador

Se dice que el estimador es insesgado cuando su valor esperado es igual al parámetro. **Insesgado:** $E[\hat{\theta}] = \theta$.

Estimadores Puntuales

- Estimador para μ:
  $\hat{\mu} = \bar{x}$ es insesgado: $E[\hat{\mu}] = E[\bar{x}] \underset{\text{por T.L.C.}}{=} \mu \Rightarrow \mu \approx \bar{x}$
- Estimador para σ^2:
  $\hat{\sigma}^2 = s^2$ es insesgado $\Rightarrow \sigma^2 \approx s^2$

$E = |\bar{x} - \mu|$: error de la estimación.

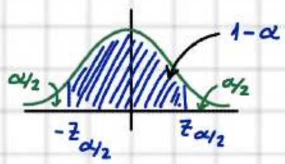
> **Observación:** para tener una idea del valor del error de la estimación se puede usar una estimación por intervalo. Estas estimaciones se denominan intervalos de confianza.

Intervalos de confianza: Un intervalo de confianza para el parámetro θ es un intervalo que contiene al parámetro con una probabilidad denominada $1-\alpha$, llamada valor de confianza.

Intervalo de confianza, con nivel de confianza $(1-\alpha)$ es $[\theta_1 ; \theta_2]$.
Si $P(\theta_1 \leq \theta \leq \theta_2) = 1-\alpha$

Construcción del intervalo de confianza

En el caso a_1 se tiene $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0;1)$



Por construcción: $P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1-\alpha$

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

$P\left(-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{x}-\mu \leq Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$

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

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

$\Rightarrow$ El intervalo de confianza, con nivel de confianza $(1-\alpha)$ para la media μ es:

$$\mu \in \left[\bar{x}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} ; \bar{x}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Error máximo de la estimación: $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

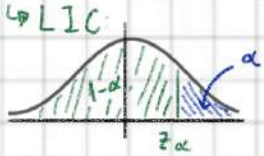
$\Rightarrow E = Z_{\alpha/2} \sqrt{V[\bar{x}]}$ $\Rightarrow \mu \in \bar{x} \pm E$ con $E = Z_{\alpha/2} \sqrt{V[\bar{x}]}$

↓

Intervalo bilateral

Observación: En algunos casos, es necesario trabajar con intervalos de confianza unilaterales, que tienen solo intervalo superior o inferior. Su determinación es similar a lo anterior.

- Límite inferior:
  ↳ LIC:



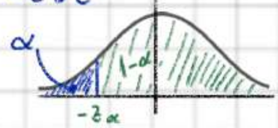
$P(z \leq z_\alpha) = 1-\alpha$

Dado que $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

↳:

$\mu \in [\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}; +\infty[$

- Límite superior:
  ↳ LSC:



$P(z > -z_\alpha) = 1-\alpha$

↓ :

$\mu \in]-\infty; \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}]$

Si no se dice nada del intervalo, se dice que es bilateral.

Observación: Cuando se va a calcular el tamaño de muestra, a priori se asume que va a ser grande (a_1, b_1). Si resulta ser pequeña, se la recalcula con la variable T, tomando el valor calculado como tamaño de muestra aproximado.

ejemplo: Un productor de lámparas quiere promocionar su producto a través de la duración. Para esto, toma una muestra de tamaño $n=100$ y encuentra que la media de la muestra es $\bar{x}$ 583 (h) y desviación estándar $s=95$ (h). Encuentre los intervalos de confianza con 95 % de confianza e interprete.

Para el caso en el que σ^2 sea desconocida, en lugar de la variable Z, trabajamos con la variable T (con población normal)

Observación: cuando no se indica o se especifica el nivel de confianza, por defecto se toma del 95% $(1-\alpha = 0.95)$

Observación: cuando se especifica el error de estimación, se puede hallar el valor mínimo de muestra.

a_1: $E = Z_{\frac{\alpha}{2}} \sqrt{V[\bar{x}]}$

$E = Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}$

$E = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$n = \left(Z_{\frac{\alpha}{2}} \frac{\sigma}{E}\right)^2$

a_2: $E = t_{\frac{\alpha}{2}} \sqrt{V[\bar{x}]}$

$E = t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

$n = \left(t_{\frac{\alpha}{2}} \frac{s}{E}\right)^2$

b_1: $E = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

$n = \left(Z_{\frac{\alpha}{2}} \frac{\sigma}{E}\right)^2 \left(\frac{N-n}{N-1}\right)$

$n(N-1) = \left(Z_{\frac{\alpha}{2}} \frac{\sigma}{E}\right)^2 (N-n)$

$n(N-1) = N\left(Z_{\frac{\alpha}{2}} \frac{\sigma}{E}\right)^2 - n\left(Z_{\frac{\alpha}{2}} \frac{\sigma}{E}\right)^2$

$n\left[N-1+\left(Z_{\frac{\alpha}{2}} \frac{\sigma}{E}\right)^2\right] = N\left(Z_{\frac{\alpha}{2}} \frac{\sigma}{E}\right)^2$

$n = \frac{N\left(Z_{\frac{\alpha}{2}} \frac{\sigma}{E}\right)^2}{\left[N-1+\left(Z_{\frac{\alpha}{2}} \frac{\sigma}{E}\right)^2\right]}$

b_2: $E = t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N}}$

Dado que la población es infinita y σ^2 desconocida $\Rightarrow$ 0.2

Intervalo bilateral: $\mu \in \bar{x} \pm t_{\frac{\alpha}{2}} \sqrt{V[x]}$

Nivel de confianza: $95\% \Rightarrow 1-\alpha = 0.95$

$\Rightarrow \alpha = 0.05 \ , \ \frac{\alpha}{2} = 0.025$

$$t_{\frac{\alpha}{2}} = t_{0.025;99} = 1.9842$$
$$t_{\alpha} = t_{0.05;99} = 1.6604$$

$$V[x] = \frac{s^2}{n} = \frac{95^2}{100} \Rightarrow \sqrt{V[x]} = 9.5$$

- Intervalo bilateral: $\mu \in 583 \pm 1.9842 \cdot 9.5$

$$\mu \in [564.15 ; 601.85]$$

- Intervalos unilaterales:

LIC: $\mu \in [\bar{x} - t_{\alpha}\sqrt{V[x]} \ ; +\infty[$
$\mu \in [583 - 1.6604 \cdot 9.5 \ ; +\infty[$
$\mu \in [567.22 ; +\infty[$

LSC: $\mu \in]-\infty ; \bar{x} + t_{\alpha}\sqrt{V[x]}]$
$\mu \in [0 ; 583 + 1.6604 \cdot 9.5]$
$\mu \in [0 ; 598.77]$

11/07/2025

ejemplo: En una población, se conoce que la desviación estándar es $\sigma = 45$. Para estimar la media, se ha tomado una muestra de tamaño $n = 32$, en la cual, se tiene que $\bar{x} = 187$; $s = 36$. Hallar el intervalo de confianza para un valor de confianza del 98%. Asuma que la población es de tamaño $N = 1865$.

Reemplazamos:

$$V[x] = \frac{45^2}{32}\left(\frac{1865-32}{1865-1}\right) = 62.23 \qquad \sqrt{V[x]} = 7.89$$

- Nivel de confianza: $98\% \Rightarrow 1-\alpha = 0.98$

$\frac{\alpha}{2} = 0.01 \Rightarrow z_{\alpha/2} = z_{0.01} = 2.33$

$$\Rightarrow E = 2.33 \cdot 7.89 = 18.38$$

$$\Rightarrow \mu \in 187 \pm 18.38$$

$$\mu \in [268.62 ; 305.38]$$

Si se desea que el error sea a lo más 22, ¿Con qué tamaño de muestra se debe trabajar?

$E = 22$; $n = ?$

Dado que es población finita: N $\qquad \sigma^2$: conocido (b1)

$$n = \frac{N z^2 \sigma^2}{(N-1)E^2 + z^2\sigma^2} = \frac{1865 \cdot 2.33^2 \cdot 45^2}{1864 \cdot 22^2 + 2.33^2 \cdot 45^2} = 22.45$$

$$\Rightarrow \boxed{n = 23}$$

Como la muestra es pequeña, recalculamos con la variable t (b2)

$$n = \frac{N t^2 \underbrace{s^2}_{\approx \sigma^2}}{N E^2 + t^2 \underbrace{s^2}_{\approx \sigma^2}} \approx \frac{N t^2 \sigma^2}{N E^2 + t^2 \sigma^2}$$

- si suponemos que $n = 23 \Rightarrow t_{\alpha/2} = t_{0.01, 22} = 2.5083$ no hay coincidencia

$$\Rightarrow n = \frac{1865 \cdot 2.5083^2 \cdot 45^2}{1865 \cdot 22^2 + 2.5083^2 \cdot 45^2} = 25.9 \Rightarrow n = 26$$

- si suponemos que $n = 26 \Rightarrow t = \frac{\alpha}{2} = t_{0.01, 25} = 2.4851$

$$\Rightarrow n = \ldots = 25.48 \Rightarrow n = 26$$

Estimador de la varianza poblacional

Para estimar la varianza poblacional, se tiene que $\hat{\sigma}^2 = s^2 \Rightarrow \sigma^2 \approx s^2$

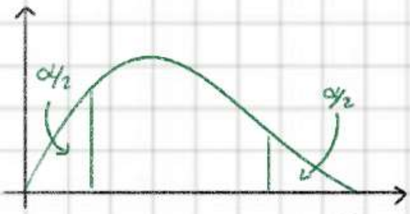
Para estimar la varianza σ^2 (intervalo de confianza) se tiene que la variable $\chi^2 = \frac{(n-1)s^2}{\sigma^2} \sim$ chi cuadrado con $(n-1)$ g.l. Siguiendo un procedimiento similar al de la media, se encuentra que el intervalo de confianza, con nivel de confianza $\sigma^2 \in \left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}} ; \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right]$ (finita y infinita)

la población debe ser normal. Si no es normal, los cálculos que se lleguen a hacer no tienen sentido.

Intervalos unilaterales:

- LIC: $\sigma^2 \in \left[\frac{(n-1)s^2}{\chi^2_{\alpha}} ; +\infty\right]$
- LSC: $\sigma^2 \in \left[0 ; \frac{(n-1)s^2}{\chi^2_{1-\alpha}}\right]$

Demostración:



$$P\left[\chi^2_{1-\frac{\alpha}{2}} \leq \chi^2 \leq \chi^2_{\frac{\alpha}{2}}\right] = 1-\alpha$$

$$P\left[\chi^2_{1-\frac{\alpha}{2}} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2}}\right]$$

$$P\left[\frac{1}{\chi^2_{1-\frac{\alpha}{2}}} \geq \frac{\sigma^2}{(n-1)s^2} \geq \frac{1}{\chi^2_{\alpha/2}}\right] = 1-\alpha$$

$$P\left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}} \geq \sigma^2 \geq \frac{(n-1)s^2}{\chi^2_{\alpha/2}}\right] = 1-\alpha$$

$$P\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right] = 1-\alpha$$

ejemplo: Con el ejemplo anterior, asumir que no se conoce la varianza.

muestra: $n = 32$; $\bar{x} = 187$; $s^2 = 36^2$

Estimación puntual: $\hat{\sigma}^2 = s^2 = 36^2 \Rightarrow \sigma^2 \approx 36^2$

Intervalo de confianza: nivel de confianza de 98%

$1-\alpha = 0.98 = \frac{\alpha}{2} = 0.01$; $1-\frac{\alpha}{2} = 0.99$

- $\chi^2_{\alpha/2} = \chi^2_{0.01;31}$

$\frac{\alpha}{2} = 0.01 \rightarrow$ está en la tabla
g.l $= 31 \rightarrow$ no está en la tabla

está entre 30 g.l $\wedge$ 40 g.l

$\Rightarrow \chi^2_{0.01;30} = 50.892 \wedge \chi^2_{0.01;40} = 63.691$

En la recta de interpolación: $\chi^2_{0.01} = m(g.l) + b$

g.l. $= 30$; $50.892 = 30m + b$
g.l. $= 40$; $63.691 = 40m + b$

$\Rightarrow m = 1.28$; $b = 12.492$

$\Rightarrow \chi^2_{0.01;31} = 1.28(31) + 12.492 = 54.172$

> Observación: cuando el valor buscado no está en la tabla, en la tabla z se toma el valor más próximo. En otros casos (t-student, chi cuadrado, ...) se usa interpolación lineal. Para ello se buscan 2 valores entre los que esté el valor buscado. Se determina la recta que pasa por los dos puntos

- $\chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0.99;31}$

$\chi^2_{0.99;30} = 14.953$ $\qquad\qquad$ $\chi^2_{0.99;40} = 22.164$

$g.l. = 30 \Rightarrow 14.953 = 30m + b$
$g.l. = 40 \Rightarrow 22.164 = 40m + b$

$\Rightarrow m =$

Estimación de la proporción

En una población binomial (cada elemento es considerado como éxito o fracaso). Si la proporción o probabilidad de éxito (p) es desconocida, se la estima a través de una muestra de tamaño n "grande" ($n \geq 100$). Si x es el número de éxitos en la muestra, la proporción de éxitos en la muestra es $\hat{p} = \frac{x}{n}$. Este valor es un estimador insesgado de p. Para determinar un intervalo de confianza para p tomamos a la distribución normal como una aproximación de la distribución binomial.

Binomial $\sim$ **Normal**

Binomial:

$Rec\ X = \{0, 1, \ldots, n\}$

$P(X=k) = \binom{n}{k} p^k q^{n-k}$

$P(X \leq k) = \sum_{i=0}^{k} P(X=i)$

$\mu = np \ ; \ \sigma^2 = npq$

Normal:

$Rec\ X = \mathbb{R}$

Probabilidad: intervalos

Para la aproximación, se toma la llamada corrección por contuinidad:

$$\text{Punto: } x \overset{\text{binomial}}{=} k \longrightarrow \text{Intervalo: } \underbrace{\left[k - \tfrac{1}{2} ; k + \tfrac{1}{2}\right]}_{\text{normal}}$$

ejemplo: moneda trucada: cara (éxito) $\to p = 0.60$; sello (fracaso) $\to q = 0.40$
se lanza 10 veces.

X: # veces que sale cara en los 10 lanzamientos

$$P(X \geq 4) = P(X=4) + P(X=5) + \dots + P(X=10)$$
$$= 1 - P(X \leq 3) = 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)]$$
$$= 1 - B(3; 10; 0.60)$$
$$= 1 - 0.0548 = 0.9452$$

Con aproximación: $\mu = np = 6$; $\sigma^2 = npq = 2.4$

Por la corrección por continuidad:

$$\{4, 4, 6, \dots, 10\} \to [3.5 ; 10.5]$$

$$P(X \geq 4) \approx P(3.5 \leq X \leq 10.5) = F(10.5) - F(3.5)$$
$$= \Phi\left(\frac{10.5-6}{\sqrt{2.4}}\right) - \Phi\left(\frac{3.5-6}{\sqrt{2.4}}\right)$$
$$= \Phi(2.90) - \Phi(-1.61)$$
$$= \Phi(2.90) - 1 + \Phi(1.61) = 0.9444$$

Siguiendo un procedimiento similar al caso de la media, se encuentra que el intervalo de confianza para la proporción con nivel de confianza $1-\alpha$ es:

- Población infinita: $p \in \hat{p} \pm E$; $E = Z_{\frac{\alpha}{2}} \sqrt{V[\hat{p}]}$; $V[\hat{p}] = \frac{\hat{p}\hat{q}}{n}$
- Población finita: N: $V[\hat{p}] = \frac{\hat{p}\hat{q}}{(n-1)} \left(\frac{N-n}{N}\right)$

ejemplo: En una elección en la que participan 8000 estudiantes, hay 2 candidatos para dirigir Fepon, candidato A $\wedge$ candidato B. Para analizar los porcentajes de aceptación de los candidatos, se ha tomado una muestra de 180 alumnos. 94 votos son para el candidato A. Aproxime la proporción para este candidato en la población.

$N = 8000$ $\quad$ $n = 180$ $\quad$ $x = 94$ $\quad \Rightarrow \hat{p} = \frac{x}{n} = \frac{94}{180} = 0.5222$

Intervalo de confianza: $p \in \hat{p} \pm E$

$V[\hat{p}] = 1.36 \cdot 10^{-3}$ $\quad$ $\sqrt{V[\hat{p}]} = 0.037$

$Z_{\frac{\alpha}{2}} = 1.96 : E = 1.96 \cdot 0.037 = 0.07252$

$\Rightarrow p \in [0.5222 \pm 0.07252]$

$\Rightarrow p \in [0.4497 ; 0.59474]$

Para calcular el tamaño de la muestra cuando se ha especificado el error, se tiene:

- Población infinita:
  $E = Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \Rightarrow n = \left(\frac{Z_{\alpha/2} \sqrt{\hat{p}\hat{q}}}{E}\right)^2$
- Población finita:
  $E = Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}\left(\frac{N-n}{N}\right)} \Rightarrow n =$

Pero dado que aparecen $\hat{p} \wedge \hat{q}$, se toman estos valores de una muestra previa. Si esto no se dispone, se toma $\hat{p} = 0.5$

14/07/2025

Pruebas Hipótesis

Cuando se realiza un estudio estadístico, a priori se pueden realizar conjeturas sobre el valor de los parámetros.

ejemplo: X: edad de un estudiante politécnico.
Conjetura: la edad promedio es 23 años.

Para validar estas conjeturas, se toma una muestra. Si el valor propuesto no está en el intervalo de confianza, se rechazaría la propuesta, pero si el valor propuesto pertenece al intervalo de confianza, no se lo rechaza y queda como un valor posible. El no rechazar, no significa aceptar. Queda como un valor posible.

$\mu_0 = 23$ muestra $\Rightarrow$ intervalo de confianza:

$\mu \in [21.8 ; 22.7]$

$\mu \in [22.4 ; 23.1]$

A la conjetura o valor propuesto se le denomina Hipótesis nula y se la denota como H_0: conjetura o suposición sobre el valor del parámetro.

En el caso que se deba rechazar la hipótesis nula, se debe aceptar un valor alterno, denominado Hipótesis alterna, que se la denota como H_a v H_1: usualmente es el complemento de Ho.

Para tomar la decisión de rechazar o no rechazar la hipótesis nula, se lo hace con el correspondiente intervalo de confianza, pero se puede trabajar con un estadístico de prueba , deducido a partir del intervalo de confianza, el cuál se lo compara con un valor llamado crítico, el cuál establece el rechazo, denominado criterio o región de rechazo, en base al cuál se toma la decisión.

Observación: las pruebas de hipótesis están diseñadas para rechazar la hipótesis nula y aceptar la hipótesis alterna (se debe plantear la hipótesis alterna).
Por otro lado, la hipótesis nula se la rechaza cuando no está en el intervalo, la probabilidad de que esto ocurra es α y se le denomina nivel de significancia.

Prueba de hipótesis para la media μ

Para el caso a_1:

$H_0: \mu = \mu_0$ $\qquad$ $H_a: \mu \neq \mu_0$

Estadístico de prueba: $Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ Criterio de rechazo: $|Z_0| > Z_{\alpha/2}$

Dada una muestra de tamaño n (grande) con medidas estadísticas: x^2, s^2.

- Intervalo de confianza con nivel de confianza $(1-\alpha)$ es $\mu \in \bar{x} \pm E$, donde $E = Z_{\alpha/2} \sqrt{V[\bar{x}]}$

$$V[\bar{x}] = \frac{\sigma^2}{n}$$

$$\mu \in \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma^2}{n} = \left[\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

Esta prueba se denomina prueba a dos colas.

si el valor propuesto μ_0 no está en el intervalo: rechazar H_0
esto ocurre si $|\bar{x}-\mu_0| > E$

$\Rightarrow |\bar{x}-\mu_0| > Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$\Rightarrow \left| \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \right| > Z_{\alpha/2}$

Sea $Z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$: Estadístico de prueba

$\Rightarrow$ Se rechaza Ho si $|Z_0| > Z_{\alpha/2}$: Criterio de rechazo

valor crítico

Ha:	$\mu \neq \mu_0$	$\mu < \mu_0$	$\mu > \mu_0$
Rechazo:	$\lvert Z_0 \rvert > Z_{\alpha/2}$	$Z_0 < -Z_\alpha$	$Z_0 > Z_\alpha$

Observación: cuando se trabajan con intervalos unilaterales, se obtienen las pruebas de hipótesis a una cola. El procedimiento es similar al anterior, y se puede resumir en la siguiente tabla:

$\mu = \mu_0$	$\mu = \mu_0$
Ho: $\mu \leq \mu_0$	Ha: $\mu \geq \mu_0$
Ho: $\mu > \mu_0$	Ha: $\mu < \mu_0$

ejemplo: X: Estatura de los hombres estudiantes politécnicos)
(# estudiantes politécnicos = 7000)

Conjetura: estatura promedio de al menos 170 cm.

Plantee una prueba de hipótesis para validar la conjetura.

$H_0: \mu \geq 170$ $\quad$ $H_a: \mu < 170$

Estadístico de prueba: $t_0 = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}}}$

Criterio de rechazo: $t_0 < -t_\alpha$

Población: $N = 7000$ $\quad$ muestra: $n = 28$

Datos: 171, 175, 160, 160, 165, 170, 176, 168, 177, 179, 174, 172, 177, 163, 179, 174, 168, 169, 170, 178, 160, 173, 180, 168, 188, 178, 168, 170

Medidas estadísticas: $\bar{x} = 171.73$; $s = 6.98$

$$\Rightarrow t_o = \frac{171.73 - 170}{\frac{6.98}{\sqrt{28}}\sqrt{\frac{7000-28}{7000}}} = 1.31 \qquad \Rightarrow -t_\alpha = -t_{0.05;27} = -1.7033$$

Decisión: no se cumple el criterio de rechazo $\Rightarrow$ no se rechaza H_0
$\Rightarrow$ es posible que la estatura promedio sea de al menos 170 cm.

> **Observación:** en problemas prácticos, la hipótesis nula se la plantea acorde a los resultados obtenidos en la muestra. Así, por ejemplo, en el problema anterior, $\bar{x} = 171.73$.
>
> $H_a: \mu > 170$ ó $\mu \neq 170$
>
> Con esta información, replanteemos la hipótesis alterna:
>
> $H_o: \mu = 170$ $\quad H_a: \mu > 170$ $\quad$ Est. Prueba: $t_o = 1.31$
>
> Crit. rechazo: $t_o > t_\alpha$; $t_\alpha = t_{0.05;27} = 1.7033$
>
> Decisión: no se cumple el criterio, por lo tanto, no se rechaza Ho. Es posible que la estatura promedio sea de 170 cm.

ejemplo: X: tiempo de viaje del domicilio a la universidad
Conjetura: tiempo promedio es de 60 min.

$H_o: \mu = 60$ $\qquad$ $H_a: \mu < 60$

Población: $N = 7000$ $\qquad$ muestra: $n = 17$; 105, 30, 5, 120, 20, 120, 20, 20, 90, 30, 30, 40, 20, 90, 60, 120, 30.

$\bar{x} = 55.88$ $\qquad$ $s = 41.65$

Estadístico de prueba: $t_o = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}}} = -0.408$

Criterio de rechazo: $t_o < -t_\alpha$; $-t_\alpha = -t_{0.05;16} = -1.7459$

Decisión: no se cumple el criterio, por lo tanto, no se rechaza Ho.
Es posible que el tiempo promedio de viaje sea de 60 min.

Calcule el intervalo de confianza con el límite inferior.

LIC: $\mu \in \left[\bar{x} - t_\alpha \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}} ; +\infty\right[$

$$\mu \in \left[55.88 - 1.7459 \frac{41.65}{\sqrt{17}}\sqrt{\frac{7000-17}{7000}} ; +\infty\right[\Rightarrow \mu \in [38.27; +\infty[$$

21/07/2025

Prueba de hipótesis para la varianza σ^2

$H_0: \sigma^2 = \sigma_0^2$ $\qquad H_a: \sigma^2 \neq \sigma_0^2 \qquad \sigma^2 < \sigma_0^2 \qquad \sigma^2 > \sigma_0^2$

Estadístico de Prueba: $\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$ Rechazo: $\chi_0^2 > \chi^2_{\alpha/2} \vee \chi_0^2 < \chi^2_{\alpha/2}$ $\qquad \chi_0^2 < \chi^2_{1-\alpha} \qquad \chi_0^2 > \chi^2_{\alpha}$

ejemplo: Se está realizando un estudio socioeconómico en una determinada población, para el cuál, los datos de la muestra se ven agrupados en la siguiente tabla:

#	Ingresos	Personas	x_i
1	470.5-575.5	4	523
2	575.5-680.5	8	628
3	680.5-785.5	12	733
4	785.5-890.5	17	838
5	890.5-995.5	13	943
6	995.5-1100.5	6	1048
7	1100.5-1205.5	3	1153
		63	

Dado que el perfil del histograma se puede ajustar con una campana, se podría asumir que los datos vienen de una población normal.

Por otro lado, dado que
$R = 1205.5 - 470.5 = 735 \approx 6\sigma$
$\Rightarrow \sigma = 123$

Plantee pruebas de hipótesis para las conjeturas dadas.

En la muestra: $\bar{x} = 828 \qquad s = 159.70 \Rightarrow s^2 = 25504.09$

Prueba de hipótesis:

Media: $H_0: \mu = 850 \qquad H_a: \mu < 850$

Estadístico de Prueba: $t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = -1.09$

Criterio de rechazo: $t_0 < -t_\alpha$
$\hookrightarrow -t_\alpha = -t_{0.05;\,62} = -1.6698$

Decisión: No se cumple el criterio $\Rightarrow$ No se rechaza H_0.
Es posible que el sueldo promedio de la población sea de $850.

Varianza: $H_0: \sigma^2 = 15129 \qquad H_a: \sigma^2 > 15129$

Estadístico de Prueba: $\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} = 104.52$

Criterio de rechazo: $\chi_0^2 > \chi^2_\alpha$
$\hookrightarrow \chi^2_\alpha = \chi^2_{0.05;\,62} \approx 79.082$

Decisión: Si se cumple el criterio, por lo tanto, si se rechaza H_0.
Es decir, la varianza poblacional es mayor a $15129.

Prueba de hipótesis para la proporción p

En el caso de una población binomial, se puede suponer un valor de la proporción de éxitos. Dependiendo de lo que se observe en una muestra grande ($n>100$), se plantea la hipótesis alterna. Para rechazar o no rechazar la hipótesis nula, se debería determinar el intervalo de confianza para la proporción (calculado con la distribución normal como aprox de la binomial) y siguiendo un procedimiento similar al de la media, es suficiente trabajar con un estadístico de prueba.

$H_0: p=p_0$ $\quad$ $H_a: p\neq p_0 \quad p<p_0 \quad p>p_0$

Estadístico de prueba: $Z_0=\dfrac{\hat{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}}$ $\quad$ Rechazo: $|Z_0|>Z_{\alpha/2} \quad Z_0<-Z_\alpha \quad Z_0>Z_\alpha$

→ Población infinita. Si es finita, debe ir el factor de corrección.

ejemplo: En una determinada comunidad, se supone que hay igual número de hombres y mujeres. Para validar esta suposición, se ha tomado una muestra de 130 personas, entre las cuales 80 son mujeres. Realice una prueba de hipótesis para validar la suposición.

Éxito: la persona es mujer

$n=130$; $x=80$; $\hat{p}=\frac{80}{130}=0.6153$

Prueba de Hipótesis:

$H_0: p=0.5$ $\quad$ $H_a: p>0.5$

Estadístico de prueba: $Z_0=\dfrac{0.6153-0.5}{\sqrt{\frac{0.5\cdot 0.5}{130}}}=2.63$

Rechazo: $Z_0>Z_\alpha$

↳ $Z_\alpha=Z_{0.05}=1.645$

Decisión: el criterio se cumple, por lo tanto se rechaza H_0.

Consulto: pruebas de hipótesis para pruebas apareadas

Comparación DE Muestras

En ciertas ocasiones, es necesario realizar comparación de 2 o más poblaciones a través de sus parámetros. Para ello, se toma muestras en cada población y se determina el intervalo de confianza para la relación que pueden tener 2 parámetros.

ejemplo: 1 Se desea comparar la repitencia entre 2 facultades.
2 Comparar el sueldo los ing. matemáticos y de los físicos.
3 Analizar la efectividad de una medicina para una dolencia.

Este tipo de comparaciones se lo hace a través de las medias.

① X: Repetencia   Facultad A: Población 1: μ_1
Facultad B: Población 2: μ_2

Congetura: $\mu_1 = \mu_2 \Leftrightarrow \mu_1 - \mu_2 = 0$: Hipótesis nula.

② X: ingreso mensual   Población 1: ing. mat. : μ_1
Población 2: físicos : μ_2

Congetura: físicos ganan \$50 más que los ing. mat.
$\mu_2 = \mu_1 + 50 \Leftrightarrow \mu_2 - \mu_1 = 50$

③ X: efectividad de la medicina   muestras apareadas

Población 1: enfermos antes de suministrar medicina: μ_1
Población 2: enfermos después de suministrar medicina: μ_2

Congetura:

Para el caso de muestras independientes que sean o no de la misma población, se desea tomar una congetura de la forma

$$H_0: a_1\mu_1 + a_2 b_2 = b$$

Para encontrar el intervalo de confianza y el correspondiente estadístico de prueba, se tiene Estimador de $a_1\mu_1 + a_2\mu_2$ es $\widehat{a_1\mu_1 + a_2\mu_2} = a_1\bar{x}_1 + a_2\bar{x}_2$ (valor esperado).

Para la varianza se tiene que

$$V[a_1\bar{x}_1 + a_2\bar{x}_2] = a_1^2 V[\bar{x}_1] + a_2^2 V[\bar{x}_2] = a_1^2 \frac{\sigma_1^2}{n_1} + a_2^2 \frac{\sigma_2^2}{n_2}$$

Y el intervalo de confianzo $(a_1\mu_1 + a_2\mu_2) \in [(a_1\bar{x}_1 + a_2\bar{x}_2) \pm E]$; donde $E = z_{\alpha/2}\sqrt{a_1^2\frac{\sigma_1^2}{n_1} + a_2^2\frac{\sigma_2^2}{n_2}}$

Estadístico de Prueba: $z_0 = \dfrac{(a_1\bar{x}_1 + a_2\bar{x}_2) - b}{\sqrt{a_1^2\frac{\sigma_1^2}{n_1} + a_2^2\frac{\sigma_2^2}{n_2}}}$ → Válido para poblaciones infinitas, muestras grandes y varianzas conocidas

valor propuesto

Cuando se trabaja con la variable T, se suman los g.l. de ambas muestras

ejemplo: Se desea comparar el peso de las tortugas al nacer en cautiverio y libres. Para ello, de las que están en cautiverio se ha tomado una muestra $n_1 = 40$ tortugas, con $\bar{x}_1 = 18.5$ (g) ∧ $s_1 = 4.8$ (g). De las tortugas libres, se ha tomado $n_2 = 35$; $\bar{x}_2 = 16.3$ (g) ∧ $s_2 = 5.4$ (g).

Conjetura: tortugas en cautiverio nacen pesando en promedio 2 g más que las libres.

$H_0: \mu_1 = \mu_2 + 2 \iff H_0: \mu_1 - \mu_2 = 2$ $\qquad H_a: \mu_1 - \mu_2 \neq 2$

Intervalo de confianza: $(\mu_1 - \mu_2) \in [(\bar{x}_1 - \bar{x}_2) + E]$

$$E = t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \rightarrow t_{0.025;73} = 1.9930$$

$$E = 1.9930\sqrt{\frac{4.8^2}{40} + \frac{5.4^2}{35}} = 2.37$$

$(\mu_1 - \mu_2) \in [(18.5 - 16.3) \pm 2.37]$

$(\mu_1 - \mu_2) \in [-0.17; 4.57]$

Dado que el valor propuesto $\mu_1 - \mu_2 = 2$ está en el intervalo, no se rechaza Ho.

25/07/2025

Comparación con muestras pequeñas

Las poblaciones deben ser normales, de igual varianza ($\sigma_1^2 = \sigma_2^2 = \sigma^2$). La varianza es desconocida, por lo tanto hay que estimarlo. La estimación de σ^2 se lo toma con un promedio ponderado de las varianzas muestrales, el peso de la ponderación se lo toma a los grados de libertad (suma de los g.l. de ambas muestras):

$$\sigma^2 = S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$$

y el intervalo de confianza: $(a_1\mu_1 + a_2\mu_2) \in [(a_1\bar{x}_1 + a_2\bar{x}_2) \pm E]$. Donde:

$$E = t_{\alpha/2}\sqrt{a_1^2\frac{s_1^2}{n_1} + a_2^2\frac{s_2^2}{n_2}} = t_{\alpha/2}\,\sigma\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2}} = t_{\alpha/2}\,S_p\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2}}$$

Estadístico de prueba: $t_0 = \dfrac{(a_1\bar{x}_1 + a_2\bar{x}_2) - b}{S_p\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2}}}$

> Observación: dado que en el análisis anterior, las varianzas poblacionales deben ser iguales, primero hay que realizar una prueba de hipótesis para ver si es factible que las varianzas sean iguales.

Prueba de hipótesis para igualdad de varianzas

Suponga que en las muestras, las varianzas son $\underbrace{S_m^2}_{\text{varianza menor}} \wedge \underbrace{S_M^2}_{\text{varianza mayor}}$.

$H_0: \sigma_M^2 = \sigma_m^2 \iff H_0: \dfrac{\sigma_M^2}{\sigma_m^2} = 1$

$H_a: \sigma_M^2 > \sigma_m^2 \qquad H_a: \dfrac{\sigma_M^2}{\sigma_m^2} > 1$

Estadístico de Prueba: $F_0 = \dfrac{S_M^2}{S_m^2}$ distribución de Fisher

Criterio de Rechazo: $F_0 > F_\alpha$

ejemplo: Se desea comparar el peso al nacer de 2 tipos de ranas, para lo cual, se han tomado muestras en cada especie, cuyos datos se muestran a continuación:

Especie 1: $n_1 = 15$: 7, 7.5, 8.1, 6.8, 6.9, 7.2, 7.5, 6.6. 7.1, 6.9, 6.8, 7.4, 7.6, 8.1, 7.1

Especie 2: $n_2 = 12$: 6.9, 7.3, 7.5, 7.1, 6.7, 8, 7.7, 7.4, 8.1, 8.4, 9, 7.8

¿Es posible aceptar que el peso promedio de las 2 especies al nacer es el mismo?

$H_0: \mu_1 = \mu_2$

- $n_1 = 15$
- $n_2 = 12$
- $\bar{x}_1 = 7.24$
- $\bar{x}_2 = 7.6583$
- $S_1^2 = 0.2069$
- $S_2^2 = 0.4281$

Dado que las muestras son pequeñas, para comparar las medias, las varianzas poblacionales deben ser iguales.

- Prueba de hipótesis para igualdad de varianzas. $\alpha = 0.05$

$$H_0: \sigma_2^2 = \sigma_1^2 \quad \Longleftrightarrow \quad H_0: \frac{\sigma_2^2}{\sigma_1^2} = 1$$

$$H_a: \sigma_2^2 > \sigma_1^2 \qquad\qquad H_a: \frac{\sigma_2^2}{\sigma_1^2} > 1$$

Estadístico de prueba: $F_0 = \frac{S_2^2}{S_1^2} = \frac{0.4281}{0.2069} = 2.069$

Criterio de rechazo: $F_0 > F_\alpha \mapsto F_\alpha = F_{0.05;\,11;\,14} = 2.565$

g.l. denominador

g.l. numerador

Decisión: No se cumple el criterio de rechazo, por lo tanto no se rechaza Ho. Es posible que las varianzas poblacionales sean iguales.

- Supongamos que las varianzas poblacionales son iguales: $\sigma_1^2 = \sigma_2^2 = \sigma^2 \mapsto$ desconocido

$$\hat{\sigma}^2 = S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-2)} = 0.3042 \qquad S_p = 0.5516$$

- Prueba de hipótesis para comparar medias:

$$H_0: \mu_1 = \mu_2 \quad \Longleftrightarrow \quad H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 \neq \mu_2 \qquad\qquad H_a: \mu_1 - \mu_2 \neq 0$$

Intervalo de confianza:

$(\mu_1 - \mu_2) \in [(\bar{x}_1 - \bar{x}_2) \pm E]$

$E = t_{\alpha/2}\, S_p \sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2}}$ $\quad a_1 = 1$; $a_2 = -1$

$t_{0.025;\,25}$

$E = 0.44$

$\Rightarrow (\mu_1 - \mu_2) \in [-0.42 \pm 0.44]$

$\Rightarrow (\mu_1 - \mu_2) \in [-0.86 \,;\, 0.02]$

Decisión: dado que el valor propuesto $\mu_1 - \mu_2 = 0$ si está en el intervalo de confianza, no se puede rechazar Ho. Es decir, es posible que los pesos promedio sean iguales.//

Error tipo 1: rechazar Ho cuando es verdadera. $P(1)=\alpha$
Error tipo 2: aceptar Ho cuando es falsa. $P(2)=\beta$) se puede determinar el n mín a utilizar.

Observación: Observando lo que ocurre en las muestras, la media de la segunda especie es ligeramente mayor a la de la primera. Otra opción para plantear la hipótesis es

$$H_0: \mu_1 = \mu_2 \iff H_0: \mu_1 - \mu_2 = 0$$
$$H_a: \mu_1 < \mu_2 \qquad H_a: \mu_1 - \mu_2 < 0$$

y el intervalo de confianza unilateral

$$(\mu_1-\mu_2) \in \left[(\bar{x}_1-\bar{x}_2) - t_\alpha S_p\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\ ;\ \infty\right) \to t_\alpha = t_{0.05;25}$$

$$\Rightarrow (\mu_1-\mu_2) \in [-0.7849\ ;\ +\infty) \ni 0$$

Así, no se rechaza Ho.

Por otro lado, si se trabaja con el estadístico de prueba:

Estadístico de prueba: $t_0 = \dfrac{(\bar{x}_1-\bar{x}_2)-0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = -1.966$

Criterio de rechazo: $|t_0| > t_{\alpha/2} \to t_{\alpha/2} = t_{0.05;25} = 2.0595$

Decisión: No se cumple el criterio de rechazo, no se rechaza Ho.

Para el caso de comparar poblaciones binomiales, se sigue un procedimiento similar. $\to$ Consultar

28/07/2025

Pruebas de hipótesis para Datos de Frecuencias

En estas pruebas no se trabaja directamente con los parámetros de una población, sino con los datos obtenidos en una muestra. La conjetura que se realiza sobre la población o sus parámetros se los considera como si fueran ciertos y se calculan los valores que se esperaría obtener bajo la hipótesis planteada. Si en efecto, estas suposiciones son ciertas, los valores observados en la muestra y los valores esperados no tendrían diferencias significativas.

ejemplo: Suponga que a los profesores se les calificaría dentro de las siguientes categorías y se asume el porcentaje de profesores en cada categoría. Se toma una muestra de tamaño $n=50$ tal que:

	Suposición	Muestra	Esperado
Excelente	10%	8	5
Muy bueno	15%	12	7.5
Bueno	30%	18	15
Regular	30%	10	15
Pésimo	15%	2	7.5

A primera vista, se observa que existe una diferencia grande entre la suposición de la muestra

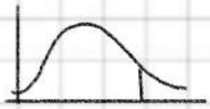
Para validar la suposición planteada, se deben comparar las frecuencias observadas en la muestra o_i con las frecuencias esperadas suponiendo que las conjeturas son ciertas e_i. Si en efecto, la conjetura fuera cierta, no debería haber diferencias significativamente grandes. Por el contrario, si la conjetura no fuera cierta, habrán diferencias significativamente grandes entre la muestra y lo esperado.

Para medir las diferencias entre lo esperado y la muestra se usa el estadístico:

$$\chi_0^2 = \sum_i \frac{(o_i - e_i)^2}{e_i}$$

Si las diferencias son grandes, el estadístico tomará valores grandes, es decir, estará en la cola.

Criterio de rechazo: $\chi_0^2 > \chi_\alpha^2$



Observación: En estas pruebas, Ha siempre es la más general.

ejemplo: Suponga que en un peaje hay 6 carriles. Se conjetura que los conductores tienen determinada preferencia por el carril a utilizar, y la conjetura es la siguiente. Igualmente, en una muestra de $n=500$ vehículos se contabilizó lo siguiente.

Observación: Los g. l. de la variable chi cuadrado es el número de celdas de comparación menos 1.

Carril	Conjetura	Muestra
1	30%	160
2	30%	135
3	15%	63
4	15%	68
5	7%	38
6	3%	36

Observación: el valor esperado se calcula como $e_i = n p_i$ con p_i la probabilidad

Observación: se acostumbra usar una sola tabla para colocar los datos observados y esperados.
↳ se pone entre ( )

Prueba de hipótesis para la pobl. multinomial.

$H_0: p_1 = 0.30;\ p_2 = 0.30;\ p_3 = 0.15;\ p_4 = 0.15;\ p_5 = 0.07;\ p_6 = 0.03$

Suponiendo que Ho fuese cierta, en el 1° carril se espera que de los 500 vehículos pasen el 30%: $500 \cdot 0.30 = 150$

Tabla de comparación:

#carrill	Frecuencia
1	160 (150)
2	135 (150)
3	63 (75)
4	68 (75)
5	38 (35)
6	36 (15)

Estadístico de prueba:
$\chi_0^2 = 34.397$

Criterio de rechazo:
$\chi_0^2 > \chi_\alpha^2 \mapsto \chi_\alpha^2 = \chi^2_{0.05;5} = 11.070$

Decisión: Sí se rechaza H_0 y se acepta la alterna.

Prueba de hipótesis para determinar independencia de v.a

ejemplo: En una población hay personas adultas, donde se definen las variables:

X: Estado civil

Y: Nivel de escolaridad

¿El estado civil es dependiente de la escolaridad?

Para probar si 2 variables son independientes, se toma una muestra y se colocan los resultados en una tabla (matriz) llamada tabla de contingencia, donde una variable va a estar dada en filas y la otra en columnas.

		X_2					
		C_1	C_2	C_3	...	C_n	Total
X_1	K_1	x_{11}	x_{12}	x_{13}		x_{1n}	$x_{1\cdot}$
	K_2	x_{21}	x_{22}	x_{23}		x_{2n}	$x_{2\cdot}$
	K_3	x_{31}	x_{32}	x_{33}		x_{3n}	$x_{3\cdot}$
	$\vdots$						
	K_m	x_{m1}	x_{m2}	x_{m3}		x_{mn}	$x_{m\cdot}$
Total		$x_{\cdot 1}$	$x_{\cdot 2}$	$x_{\cdot 3}$		$x_{\cdot n}$	N

Para calcular los valores esperados se requiere la probabilidad o porcentaje de estar en la celda P_{ij}

$$P_{ij} = (x \in K_i \cap x \in C_j)$$

Suponiendo que las variables son independientes,

$$P_{ij} = P(x \in K_i)\,P(x \in C_j) = p_i \cdot p_j = \frac{x_{i\cdot}}{N} \cdot \frac{x_{\cdot j}}{N}$$

Dado que estos valores no son conocidos, hay que estimarlos.

Valores esperados: $e_{ij} = N \cdot p_{ij} = N \frac{x_{i\cdot}}{N} \cdot \frac{x_{\cdot j}}{N} = \frac{x_{i\cdot} \cdot x_{\cdot j}}{N} = \frac{\text{Total de Fila} \cdot \text{Total de Columna}}{N}$

Estadístico de prueba: $X_0^2 = \sum_{i,j} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$ Criterio de rechazo: $X_0^2 > X_\alpha^2$

g.l. = (#filas − 1)·(#columnas − 1)

ejemplo: X_1: Materia que dicta un profesor (básica, intermedia, profesionalizante y complementaria)

X_2: Nivel de comprensión del alumno (muy bueno, bueno, regular, pésimo).

Se ha tomado una muestra, cuyos datos están en la siguiente tabla:

		X_1				
		Básica	Intermedia	Profes.	Complem.	
X_2	M.B.	40	15	10	30	75
	B.	30	15	5	35	85
	R.	40	5	2	8	55
	P.	15	5	0	2	22
		105	40	17	75	237

- Calculamos las frecuencias esperadas:

$e_{11} = 33.22$	$e_{12} = 12.65$	$e_{13} = 5.37$	$e_{14} = 23.73$
$e_{21} = 37.65$	$e_{22} = 14.34$	$e_{32} = 6.09$	$e_{42} = 26.89$
$e_{31} = 24.36$	$e_{32} = 9.28$	$e_{33} = 3.94$	$e_{43} = 17.40$
$e_{41} = 9.74$	$e_{42} = 3.71$	$e_{34} = 1.57$	$e_{44} = 6.96$

Observación: Cuando la suposición es sobre el tipo de distribución de probabilidad de la variable, a estas pruebas se las denomina prueba de bondad de ajuste. El procedimiento es similar y se debe considerar lo siguiente:

1.- Si se supone que es una variable discreta (binomial, poisson, etc.), la comparación entre los valores esperados en la muestra se los debe hacer en todos los elementos del recorrido de la v.a.

2.- Si la variable es continua, la comparación se la debe hacer por intervalos, que cubran todo el recorrido de la v.a.

3.- Se recomienda que para la comparación, en la tabla de valores observados, cada celda tenga frecuencia de al menos 4. Si esto no ocurre, se reagrupan celdas.

ejemplo: El tiempo de espera para ser atendido en una ventanilla, se supone que sigue una distribución exponencial. Para ello, se ha tomado una muestra, cuyos datos están dados en la siguiente tabla:

Tiempo de espera (minutos)	#personas
0-5	3
5-10	12
10-15	18
15-20	14
20-25	8
25-30	5
30-35	2

Observación: en estas pruebas de bondad de ajuste, el estadístico de prueba χ_0^2 tiene $K-r-1$ g.l. donde K es el número de celdas de comparación y r es el número de parámetros necesarios para calcular los valores esperados.

X_0: Tiempo de espera (min) $\qquad H_0: X \sim \mathcal{E}(\theta)$

Si Ho fuese cierta:

Función de distribución: $F(x) = 1 - e^{-\frac{x}{\theta}}$

Dado que θ es desconocido, hay que estimarlo:

$\theta = E[x] = \mu \wedge$ se tiene que $\hat{\mu} = \bar{x}$

$\hat{\theta} = \bar{x} = 15.32 \Rightarrow F(x) = 1 - e^{-\frac{x}{15.32}}$

Para que en cada celda de comparación la frecuencia sea de al menos 4, realizamos un agrupamiento entre la primera y la última celda.

Tiempo Espera	#personas
0-10	15
10-15	18
15-20	14
20-25	8
25-35	7
35-40	0
40-45	0
45-50	0
⋮	⋮

Tiempo Espera	#personas
0-10	15
10-15	18
15-20	14
20-25	8
25→	7

Calculamos las frecuencias esperadas: primero calculamos las probabilidades de los intervalos.

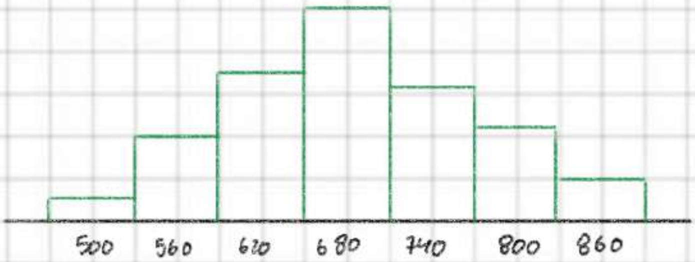
- $P(0 \leq x \leq 10) = F(10) - F(0) = (1 - e^{-\frac{10}{15.32}}) - 0 = 0.4793 \Rightarrow e_1 = n \cdot p_1 = 62 \cdot 0.4793 = 29.72$
- $P(10 \leq x \leq 15) = F(15) - F(10) = 0.1449 \Rightarrow e_2 = 8.9838$
- $P(15 \leq x \leq 20) = F(20) - F(15) = 0.1046 \Rightarrow e_3 =$
- $P(20 \leq x \leq 25) = F(25) - F(20) = 0.0755 \Rightarrow e_4 =$
- $P(25 \leq x) = 1 - F(25) = 0.1956 \Rightarrow e_5 =$

01/08/2025

ejemplo: Se está analizando el ingreso mensual en una determinada comunidad. Para ello, se ha tomado una muestra, cuyos datos agrupados son los siguientes:

#	ingresos	# personas	Puntos Medios
1	470-530	3	500
2	530-590	10	560
3	590-650	18	620
4	650-710	25	680
5	710-770	16	740
6	770-830	15	800
7	830-890	5	860
		89	

Realice el histograma de frecuencias.



X: ingreso mensual $\qquad$ $H_0: X \sim N(\mu; \sigma^2)$

Para esta conjetura, realizamos una prueba de bondad de ajuste: comparar las frecuencias observadas en la muestra con las frecuencias q se esperaría obtener si Ho fuese cierta, para ello necesitamos calcular los valores de probabilidad y como suponemos que es normal, se necesita el valor de la media: μ y de la varianza: σ^2. Como son desconocidos, estimamos con los de la muestra:

$$\hat{\mu} = \bar{x} = 685.39 \qquad \hat{\sigma}^2 = s^2 = 88.92^2 = 7906.77$$

Para la comparación de frecuencias en cada celda de comparación, la frecuencia debe ser ≥ 4. Como en la primera celda, la frecuencia es 3, reagrupamos celdas:

- Tabla de comparación:

Ingreso	frecuencia	
← 590	13	(12.66)
590-650	18	(18)
650-710	25	(23.65)
710-770	16	(19.46)
770-830	12	(10.64)
830->	5	(4.60)

- Estadístico de prueba:

$$\chi_0^2 = \sum \frac{(o_i - e_i)^2}{e_i} = 0.97$$

- Criterio de rechazo: $\chi_0^2 > \chi_\alpha^2$

$$\chi_\alpha^2 = \chi_\alpha^2 = \chi^2_{\alpha;\,k-r-1}$$
$$= \chi^2_{\alpha;\,6-2-1}$$
$$= \chi^2_{\alpha;0.05;3} = 7.815$$

- Decisión: No se cumple el criterio, no se rechaza Ho.

- Probabilidad:

$p_1: P(\chi < 590) = F(590)$
$= \Phi\left(\frac{590-\mu}{\sigma}\right)$
$= \Phi(-1.07)$
$= 0.1423 \Rightarrow 12.66$

$p_2: P(590 < \chi < 650) = F(650) - F(590)$
$= \Phi(-0.40) - \Phi(-1.07)$
$= 0.2023 \Rightarrow 18$

$p_3: P(650 < \chi < 710) = F(710) - F(650)$
$= \Phi(0.28) - F(-0.40)$
$= 0.2657 \Rightarrow 23.65$

$p_4: P(710 < \chi < 770) = F(770) - F(710)$
$= \Phi(0.95) - \Phi(0.28)$
$= 0.2186 \Rightarrow 19.46$

$p_5: P(770 < \chi < 830) = F(830) - F(770)$
$= \Phi(1.63) - \Phi(0.95)$
$= 0.1195 \Rightarrow 10.64$

$p_6: P(830 < \chi) = 0.0517 \Rightarrow 4.60$

ejemplo: Se está analizando los accidentes laborales que ocurren cada semana en una empresa. Para ello, se ha tomado una muestra, cuyos datos se muestran en la siguiente tabla:

#accidentes	#semanas
0	4
1	6
2	8
3	5
4	4
5	4
6	3
7	1
8	1
	36

Variable de estudio:

χ: # accidentes por semana

$H_0: \chi \sim$ Poisson: $P(\chi = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ con $\lambda = E[x] = \mu$

λ desconocido $\Rightarrow \hat{\lambda} = \hat{\mu} = \bar{x} = 2.94$

- Tabla de comparación:

#accidentes	frecuencias
0	4 (1.90)
1	6 (5.59)
2	8 (8.22)
3	5 (8.06)
4	4 (5.92)
5	4 (3.48)
6 o más	5 (2.83)

- Probabilidades:

$p_1: P(X=0) = \dfrac{e^{-2.94}(2.94)^0}{0!} = 0.05 \Rightarrow 1.90$

$p_2: P(X=1) = 0.16 \Rightarrow 5.59$

$p_3: P(X=2) = 0.23 \Rightarrow 8.23$

$p_4: P(X=3) = 0.22 \Rightarrow 8.06$

$p_5: P(X=4) = 0.16 \Rightarrow 5.92$

$p_6: P(X=5) = 0.09 \Rightarrow 3.48$

$p_7: P(X \geq 6) = 1 - P(X \leq 5)$

$= 1 - [P(X=1) + P(X=2) + \ldots + P(X=5)]$

$\Rightarrow 2.83$

- Estadístico de prueba: $\chi_0^2 = \sum \dfrac{(o_i - e_i)^2}{e_i} = 5.88$

- Criterio de rechazo: $\chi_0^2 > \chi_\alpha^2$ $\rightarrow$ $\chi_\alpha^2 = \chi^2_{\alpha;k-r-1}$

$= \chi^2_{0.05;5} = 11.070$

- Decisión: No se cumple el criterio de rechazo, no se rechaza Ho. Es posible que la cantidad de accidentes siga una distribución de Poisson.//