Media aritmetica

$$\bar{X} = \frac{\sum x_i \cdot f_i}{n}$$

Media armónica

$$\bar{X}_{ar} = \frac{n}{\sum 1/x_i}$$

Promedios velocidades, tiempos, rendimient.

Media geométrica

$$\bar{X}_{ge} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots}$$

Porcentajes, tasas!

Coeficiente de variación (CV)

$$CV = \frac{S}{|\bar{X}|}$$

¿cual es el más homogeneo? % de error

Rango

$$R = max(x_i) - min(x_i)$$

Mediana

- Par $\Longrightarrow \frac{a+b}{2}$
- Impar $\longrightarrow$ El del medio
- agrupados $\longrightarrow Me = L_{i-1} + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot C$

L_{i-1} $\frac{N}{2} \geq F_i$ "el primero que pase"

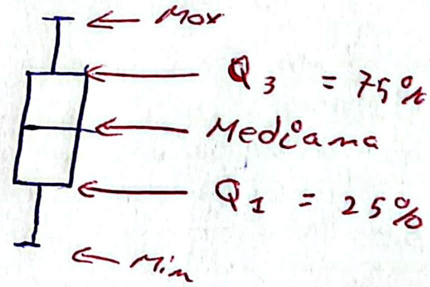
Moda

- agrupados $\longrightarrow M_0 = L_{i-1} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot C$

L_{i-1} el f_i que mas se repite

$$\Delta_1 = f_i - f_{i-1}$$
$$\Delta_2 = f_i - f_{i+1}$$

Caja y bigotes



Varianza

$$S^2 = \frac{\sum x_i^2 \cdot f_i}{n} - \bar{x}^2$$

Desviación típica

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

tallo y hojas

1	2 7
2	1 4 7
3	0

12, 17, 24, 21, 27, 30

Cuartiles y percentiles

$25 \Rightarrow \frac{25}{100}$

- cuartil $\longrightarrow C_\alpha = L_{i-1} + \frac{\alpha \cdot N - F_{i-1}}{f_i} \cdot C$

EJM

$P_{75} \Rightarrow \frac{75}{100} \cdot N \Rightarrow F_i > \frac{75}{100} \cdot N$

El primero que pase eso

Covarianza (S_{xy})

$$S_{xy} = \frac{\sum X_i \cdot Y_i \cdot f_i}{n} - \bar{X} \cdot \bar{Y}$$

>0 Directo

<0 Inverso

$=0$ Incorrelado

Coeficiente de correlación

$$r = \frac{S_{xy}}{S_x S_y} \quad [-1, 1]$$

Desviación típica, no varianza

$Y = a + bx$ Recta de regresión

$$\Downarrow$$

$$Y = \bar{Y} + \frac{S_{xy}}{S_x^2}(x - \bar{x})$$

(Recta de regresión de Y sobre X)

$\overline{A \cup B} = \bar{A} \cap \bar{B}$

$\overline{A \cap B} = \bar{A} \cup \bar{B}$

$P(B - A) = P(B) - P(A)$

$P(A \cap \bar{B}) = P(A - B) = P(A) - P(A \cap B)$

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

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

Se elige aleatoriamente una mujer ¿P. de que sea fumadora?

$$P(F|M) = \frac{P(F \cap M)}{P(M)}$$

Variables discretas:

Discretos → nº de hojas

Continuos → tiempo transcurrido

- función de probabilidad

$P(3) = 1/6$

$P[X > x] = 1 - P[X \leq x]$

1 → 1/6
2 → 1/6
3 → 1/6
⋮
Dado

- función de distribución

	1	2	3	4	5	6
$P[x=x]$	1/6	1/6	1/6	1/6	1/6	1/6
$F(x)$	1/6	2/6	3/6	4/6	5/6	6/6

$F(\infty) = 1$

$F(-\infty) = 0$

Variables continuas

- **Función de densidad**

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \qquad P(a \le x \le b) = \int_a^b f(x)\,dx \qquad P(X \le a) = \int_{-\infty}^{a} f(t)\,dt$$

- **Función de distribución**

$$F(x) = \int_{-\infty}^{x} f(t)\,dt \qquad F(-\infty) = 0 \qquad F(\infty) = 1$$

Pasar de función de densidad a función de distribución

$$f(x)\begin{cases} x & 0 \le x \le 1 \\ 2-x & 1 \le x \le 2 \\ 0 & \text{Resto} \end{cases} \rightarrow \text{Intervalos} \rightarrow \begin{cases} x < 0 \\ 0 \le x \le 1 \\ 1 < x < 2 \\ x > 2 \end{cases}$$

1.º tramo $(x < 0)$

$$\int_{-\infty}^{x} 0\,dt = 0$$

2.º tramo $(0 \le x \le 1)$

$$\int_{-\infty}^{0} 0\,dt + \int_0^x \cancel{t}\,d\cancel{x} = \cancel{0} = 0 + \left[\frac{t^2}{2}\right]_0^x \; \frac{x^2}{2} - \frac{0^2}{2} = \frac{x^2}{2}$$

3.º tramo $(1 < x < 2)$

$$\int_{-\infty}^{0} 0\,dt + \int_0^1 t\,dt + \int_1^x 2-x = 0 + \left[\frac{t^2}{2}\right]_0^1 + \left[2t - \frac{t^2}{2}\right]_1^x =$$

4 tramo

$$\int \ldots = 1$$

$$F(x)\begin{cases} 0 & x < 0 \\ \dfrac{x^2}{2} & 0 < x < 1 \\ 2 - \dfrac{x^2}{2} - 1 & 1 < x < 2 \\ 1 & x > 2 \end{cases}$$

Esperanza

- **Discretos**

$$E(x) = \sum_{i=1}^{n} x_i \cdot P(x_i)$$

- **Continuos**

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

Varianza

- **Discretos**

$$\sigma_x^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 \cdot P(x_i) = \sum(x^2 P(x)) - E[x]^2$$

- **Continuos**

$$\sigma_x^2 = E[x^2] - E[x]^2$$

Ej. V. Discretos

x puede ser 30, 40, 50, 60 y su probabilidad = 0'4, 0'2, 0'1, 0'3

$$E(x) = 30 \cdot 0'4 + 40 \cdot 0'2 + 50 \cdot 0'1 + 60 \cdot 0'3 = 12 + 8 + 5 + 18 = 43$$

$$\sigma_x^2 = \sum(x^2 \cdot P(x)) - E(x)^2 = (30^2 \cdot 0'4 + 40^2 \cdot 0'2 + 50^2 \cdot 0'1 + 60^2 \cdot 0'3) - 43^2 = 161$$

Ej. Continuos

$$f(x) = \begin{cases} \frac{x}{12} & \text{si } 1 < x < 5 \\ 0 & \text{Resto} \end{cases}$$

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(t)\, dt = \int_1^5 \frac{x^2}{12}\, dx = \frac{1}{36}\left[x^3\right]_1^5 = 31/9$$

$$\sigma_x^2 = E(x^2) - E(x)^2 = \quad - (31/9)^2 =$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 \cdot f(t)\, dt =$$

Distribuciones discretos

- **Binomial**

$\sum \binom{n}{x} p^x \cdot q^{n-x}$ $\quad E(x) = n \cdot p$ $\quad \sigma_x^2 = n \cdot p \cdot q$ $\quad B(n, p)$

- **Poisson**

$P(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$ $\quad E(x) = \lambda$ $\quad \sigma_x^2 = \lambda$

$(\lambda, \sqrt{\lambda}) \longrightarrow N(0,1)$

Si $\lambda > 16$

Ejy

$\lambda = 8$ $\quad P(x>12) = 1 - P(x \leq 12) = 1 - \sum_{i=0}^{12} \frac{12^x}{x!} \cdot e^{-12} = 0'0\ 638$

Distribuciones continuas

- **Continuas (uniforme)**

$f(x) = \begin{cases} 0 & \text{Resto} \\ \frac{1}{b-a} & a < x < b \end{cases}$ $\quad F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$

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

- **Normal**

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

$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$

$\boxed{\begin{array}{l} * \\ N(\mu, \sigma) \\ z = \frac{x-\mu}{\sigma} \end{array}}$

- $P(z \geq a) = 1 - P(z < a)$
- $P(z \leq -a) = 1 - P(z < a)$
- $P(z \geq -a) = P(z < a)$
- $P(|z| < a) = 2 \cdot P(z < a) - 1$

$P(a < z < b) = P(z < b) - P(z < a)$

Factor de corrección

Si $>$ $\Rightarrow +0'5$

Si $\geq$ $\Rightarrow -0'5$

Ejm

• Distribución normal con media 7000 y desviación típica 600

¿ $P(6280 \leq x \leq 7120)$?

$x \equiv$ vida del semiconductor $\rightarrow N(7000, 600)$

$$Z = \frac{x - 7000}{600} \rightarrow \left.\begin{matrix} \frac{6280 - 7000}{600} \\ \frac{7120 - 7000}{600} \end{matrix}\right\} \rightarrow P(-1'2 \leq Z \leq 0'2) = \ldots = 0'4Q2$$

Teorema Central del Límite (TCL)

Si μ y σ son conocidos y $n > 30$

$\Downarrow$

$$* N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Longrightarrow Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \longrightarrow N(0, 1)$$

• Si población infinita
• " tiene sustitución
• " $N \geq 20$

$$\Bigg\} \; N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$\Downarrow$ Si no

$$N\left(\mu, \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}\right)$$

t-student

cuasivarianza muestral: $\hat{S}$

varianza muestral = S

- t-student con $n-1$ grados de libertad si $n < 30$
- si $n > 30 \Rightarrow T \to N(0,1)$

$$T = \frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} = \frac{\bar{X}-\mu}{S/\sqrt{n-1}}$$

$$B(N, p) \sim\to N(N\cdot p, \sqrt{N\cdot p\cdot q})$$

$$\sim\to N(\lambda, \sqrt{\lambda})$$

Ejy

$$P(-8 \leq \bar{x}-\mu \leq 8) = P\left(\frac{-8}{10'85/\sqrt{16}} \leq \bar{x}-\mu \leq \frac{8}{10'85/\sqrt{16}}\right):$$

$$T = \frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} = \frac{\bar{X}-\mu}{10'85/\sqrt{16}} \to t_{15}$$

- De proporciones

$$\hat{p} \sim N\left(p, \sqrt{\frac{p\cdot q}{n}}\right)$$

$\boxed{1 \text{ cosa} \quad \text{u otra}}$ no hay + opciones

1 moneda 100 veces

¿+ 55 caras?

$p = 0'5$, $q = 0'5$, $n = 100$

$$N\left(0'5, \sqrt{\frac{0'5\cdot 0'5}{100}}\right)$$

$$P(x > 0'55) = P\left(z > \frac{0'55-0'5}{0'05}\right) = P(z > 1) = \ldots = 0'1587$$

$$N(0'5, 0'05) \sim\to N(0,1)$$

- Diferencia de medias

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

• Chi-cuadrado

$\chi^2_{(n-1)}$ chi-cuadrado con n-1 grados de libertad

$$\frac{n\, S_n^2}{\sigma^2}$$

muestra de 25, varianza de 6, tenga una varianza muestral > 9'45

$$\frac{n\, S_n^2}{\sigma^2} = \frac{25 \cdot 9'45}{6} = 39'4$$

$$P(S_n^2 > 9'45) = 1 - P(S_n^2 < 9'45) = 1 - P(\chi^2_{n-1} < 39'4) = 1 - 0'975 = 0'025$$

Intervalos de confianza

$$\left(\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} , \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

• Una distribución normal con desviación típica = 20.
Para una muestra de 250 se obtiene una media de 90 ¿IC. al 90%?

$$\alpha = 100 - 90 = 10 \Rightarrow \alpha/2 = 5 = 5\% = \underbrace{0'05}$$

$$\Downarrow$$

$Z_{\alpha/2}$ lo que deje a la IS: 95% = 0'95

$$\Downarrow$$

$$\left(90 - \left(1'645 \cdot \frac{20}{\sqrt{250}}\right) ; 90 + \left(\quad \right)\right) \leftarrow Z_{\alpha/2} = 1'645$$

$$(90 - 2'08 ; 90 + 2'08) = IC_{90\%} = (87'92 , 92'08)$$

• Si σ es desconocida $\Rightarrow$ se sustituye por cuasidesviación típica S_n

$$\left(\bar{x} - Z_{\alpha/2} \frac{S_n}{\sqrt{n-1}} ; \bar{x} + Z_{\alpha/2} \frac{S_n}{\sqrt{n-1}}\right)$$

• Si σ desconocida y $n < 30$

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

I.C. de Binomial

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\cdot q}{n}} \; ; \; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\cdot q}{n}}\right)$$

- muestra de 150, selecciono 50

$$\hat{p} = \frac{50}{150} = 0'33$$

I.C. Chi-cuadrado

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

- 15 vendedores con media 5 y varianza 2 ¿IC al 90%?

$\alpha/2 = 0'05 \Rightarrow 1-\alpha/2 = 0'95$, 14 grados libertad

$\chi^2_{\alpha/2\,(n-1)} = 6'57$ $\qquad$ $\chi^2_{1-\alpha/2\,(n-1)} = 23'68$

$$\left(\frac{14\cdot 2'14}{23'68} \leq 2 \leq \frac{14\cdot 2'14}{6'57}\right) = [1'27,\ 4'57]$$

$$s^2 = \left(\frac{n}{n-1}\right)\cdot S_n^2 = \frac{15}{14}\cdot 2 = 2'14$$

Intervalo de confianza para la varianza al 90% $\Rightarrow$ chi-cuadrado

ejm: Población → todos los de UAM
Muestra → Estudiantes

Variable → conjunto de valores que puede tomar una característica

tipos
- → cualitativas
  - Nominales (Estado civil, colores) [No se puede ordenar]
  - Ordinales (suspenso, Notable, sobresaliente, Rico, pobre)
- → cuantitativas
  - Discretas (Nº de hijos, nº de casos) [Nºs enteros]
  - ~~~~ Continuas (peso) [se puede medir]

$\boxed{\forall = \text{cualquiera, Para todo}}$

N = Nº total de la muestra

ejm N = 150

cuantas familias tienen hasta 4 hijo ⇒ 121

Valores que toma la variable	Frecuencia absoluta	Frecuencia Relativa (absoluta / N)	Frecuencia acumulada %
0	26	0'173 x100	26
1	42	0'28 x100	26+42=68
2	32	0'213 x100	68+32=100
3	21	0'14 x100	100+21=121
4	14	0'093 x100	121+14=135
5	11	0'073 x100	135+11=146
6	4	0'027 x100	146+4=150
	150	100	100

clases = 6

Rango = Max − Min

Amplitud = Rango / clases = 5

[90−95), [95−100)

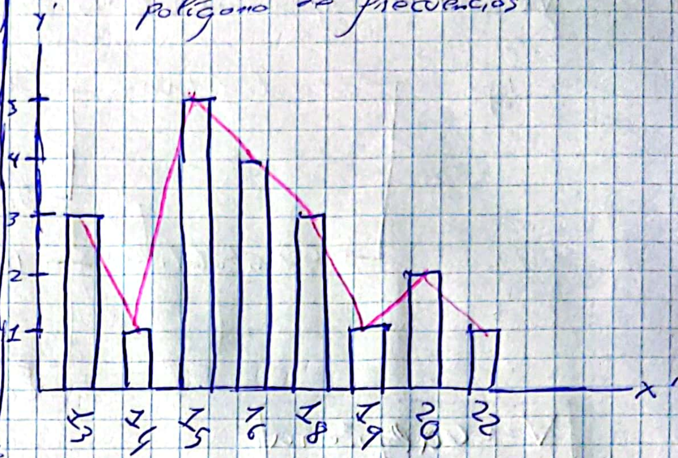
MC = Marca de clase = ~~~~ pillar el centro 92'5 $= 90 + \left(\frac{\text{Amplitud}}{2}\right)$ (Min)

Histograma.

Tema 1 Distribuciones, frecuencias y representaciones

	f. Absoluta	f. Relativa	f. acumulada
13	3	15	15
14	1	5	20
15	5	25	45
16	4	20	65
18	3	15	80
19	1	5	85
20	2	10	95
22	1	5	100
	20	100	

polígono de frecuencias



N = nº de muestras

Media = $\bar{x} = \dfrac{\text{Suma de los datos}}{N}$

S^2 = varianza = $\dfrac{(x_1-\bar{x})^2+(x_2-\bar{x})^2 \cdots + (x_n-\bar{x})^2}{N}$

S = desviación típica = $|\sqrt{S^2}|$

Rango = $V_{máx} - V_{mín}$

Amplitud = $\dfrac{\text{Rango}}{\text{clases}}$

Coeficiente de variación = CV = $\dfrac{S}{|\bar{x}|}$ = (%)

Moda = Mo = Nº que más se repite

Mediana = Me → Impar (N) ⇒ Se pilla el Nº del medio
→ Par (N) ⇒ Si pillan los 2 valores centrales y se dividen entre 2.

Media aritmética = $\bar{X} = \dfrac{(x_1 \cdot f_1)+(x_2 \cdot f_2)+(x_3 \cdot f_3)+(x_n \cdot f_n)}{N}$

Marca de clase = para hayar la media aritmética en intervalos, cogemos el punto medio del intervalo (40-60 ⇒ 50)

- **Media aritmética**

$$\overline{x} = \frac{x_1 + x_2 + \dots + x_n}{N} = \frac{\sum_i x_i}{N}$$

- **Media aritmética a partir de datos agrupados**

$$\overline{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots x_n \cdot f_n}{f_1 + f_2 + \dots f_n} = \frac{\sum x_i \cdot f_i}{N}$$

- Para intervalos, se utiliza Mc (Marca de clase) que es el valor intermedio

- **Media aritmética ponderada**

P = Peso del dato

$$\overline{x} = \frac{x_1 P_1 + x_2 P_2 \dots + x_n P_n}{P_1 + P_2 \dots + P_N} = \frac{\sum_i x_i P_i}{\sum_i P_i}$$

- **Media armónica**: Inversa de la media aritmética

$$\overline{x}_a = \frac{n}{1/x_1 + 1/x_2 \dots + 1/x_n} = \frac{n}{\sum_i 1/x_i}$$

para promediar velocidades, tiempos, rendimiento, ...

- Si uno de los valores es 0 ó próximo, no se puede calcular

- **Media geométrica**

$$\overline{x}_g = \sqrt[n]{x_1 \cdot x_2 \dots x_n}$$

Para promediar porcentajes, tasas, Nº índices, ...

- **Mediana**

Me → Impar ⇒ El valor del centro
→ Par ⇒ La suma de los 2 valores /2

- Primero se tiene que ordenar de Menor a Mayor

Para valores continuos o agrupados

$$Me = x_{med} = L_i + (L_i - L_{i-1}) \cdot \frac{\frac{N}{2} - F_{i-1}}{f_i}$$

L_i: límite inferior de la clase / f_i: frecuencia absoluta de la clase media

F_{i-1}: Frecuencia acumulada hasta la clase media

$L_i - L_{i-1}$: Anchura de la clase (c)

EJM	f_i	F_i
[60, 63)	5	5
[63, 66)	18	23
[66, 69)	42	65
[69, 72)	27	92
[72, 75)	8	100
	100	

clase Media [66, 69)

L_i 66 $\qquad \frac{N}{2} = 50$

F_{i-1} 23

N 100

C 3

f_i 42

$$M_e = x_{med} = 66 + \frac{50 - 23}{42} 3 = 67'93$$

- Moda: Valor que más se repite de la variable

 - Para datos agrupados

$$M_o = L_{i-1} + \frac{\Delta_1}{\Delta_1 + \Delta_2} C$$

Δ_1 Diferencia de frecuencias absolutas de la clase modal y la premodal

Δ_2 Diferencia de frecuencia absolutas de la clase modal y postmodal

* con los datos de antes

EJM.

L_{i-1} 66

$\Delta_1 = 42 - 18 = 24$

$\Delta_2 = 42 - 27 = 15$

C 3

$$M_o = 66 + \frac{24}{24 + 15} 3 = 67'85$$

$$\boxed{\text{42 porque f. acumulada es el 1º que pasa del 50\%}}$$

¿L_{i-1} en la mediana?

tenemos $N = 100 \Rightarrow \frac{N}{2} = 50 \Rightarrow$ El primer $F_i \geq 50$ es nuestro L_{i-1}

- **Varianza**

- Para conjunto de datos

$$s^2 = \frac{\sum x_i^2}{N} - \bar{x}^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

- Para datos agrupados

$$s^2 = \frac{\sum f_i x_i^2}{N} - \bar{x}^2$$

- Para frecuencias relativas

$$s^2 = \sum n_i \cdot x_i^2 - \bar{x}^2 = \sum n_i (x_i - \bar{x})^2$$

Ej:

x_i	f_i	n_i = frecuencia relativa
0	40	0'44
1	26	0'29
2	14	0'16
3	6	0'07
4	3	0'03
5	0	0'00
6	1	0'01
	90	1'00

$$n_i (x_i - 1)^2$$

$$0'03 (4 - 1)^2 = 0'27$$

⋮

Varianza = 1'4

- **Desviación típica**

- Para conjunto de datos

$$s = \sqrt{\frac{\sum x_i^2}{N} - \bar{x}^2}$$

- Para datos agrupados

$$s = \sqrt{\frac{\sum f_i x_i^2}{N} - \bar{x}^2}$$

- Para frecuencias relativas

$$s = \sqrt{\sum n_i \cdot x_i^2 - \bar{x}^2}$$

EJ 11 ¿Media, varianza y desviación típica?

600
470
170
430
300

Media $= (600 + 470 + 170 + 430 + 300) / 5 = \underline{394} = \bar{X}$

Varianza $= s^2 = \dfrac{(600-394)^2 + \ldots\ldots}{5} = 21\,704$

desviación típica $= s = \sqrt{s^2} = 147'32$

- Coeficiente de variación

$$CV = \frac{s}{|\bar{x}|}$$

¿cuál es el más homogéneo? % de error

- MEDA Mediana de las Desviaciones Absolutas

$$MEDA = \text{mediana}\,[x_i - \text{Mediana}]$$

- Rango

$$R = \max(x_i) - \min(x_i)$$

- Medidas de posición
  - Cuantiles Cuando se divide a la serie en 4 partes
  - Percentiles Cuando se divide a la serie en 100 partes

  - de 1 a 100

    $N=80$, $25\% \Rightarrow (80 \cdot 25)/100 \Rightarrow$ Dato que está en la posición 20

  - Decil 1 a 10

    $N=80 \Rightarrow$, 4 decil $\Rightarrow (80 \cdot 4)/10$

  - Cuantil 1 a 4

    $N=80$, cuantil 3 $\Rightarrow (80 \cdot 3)/4$

	M. clase	f_i	F_i
40 - 50	45	5	5
50 - 60	55	10	15
60 - 70	65	21	36
70 - 80	75	11	47
80 - 90	85	5	52
90 - 100	95	3	55
100 - 130	115	3	58

$58 = N$

percentil de 75 → 75/100

cuartil 3 → 3/4

$P_{75} =$

$\frac{75}{100} \cdot N = 43'5$

$F_i > 43'5 \Rightarrow F_i = 47$

$f_i = 11$

$Mc = 75$

$F_{i-1} = 36$

$L_{i-1} = 70$

$L_i = 80$

$$P_{75} = C_{0'75} = L_{i-1} + \frac{0'75 \cdot 58 - F_{i-1}}{f_i}(L_i - L_{i-1}) =$$

$$= 70 + \frac{0'75 \cdot 58 - 36}{11}(80 - 70) = 76'8$$

Cuartil

$$C_\alpha = L_{i-1} + (L_i - L_{i-1}) \cdot \frac{\alpha \cdot N - F_{i-1}}{f_i}$$

Cuartil 2/4 y percentil 50/100

$N = 58$

$\frac{50}{100} \cdot 58 = 29$

$F_i > 29 \Rightarrow F_i = 36$

$f_i = 21$

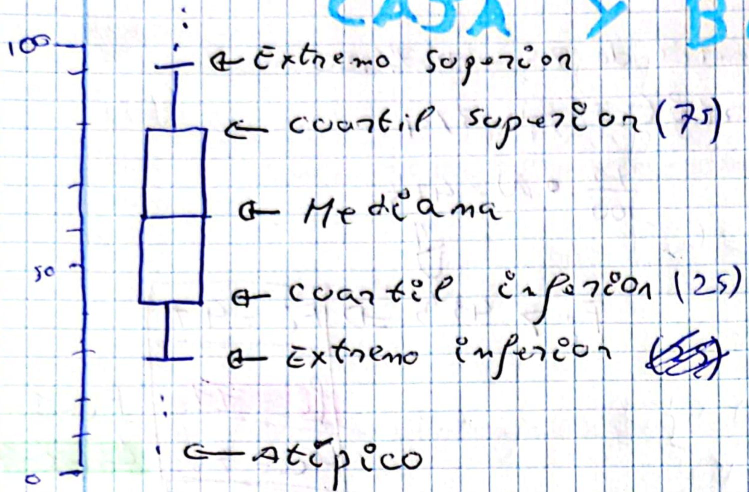
$L_{i-1} = 60$

$L_i = 70$

$$C_{0'5} = 60 + (70 - 60) \cdot \frac{0'5 \cdot 58 - 15}{21} =$$

$$= 66'6$$

CAJA Y BIGOTES



$\boxed{y = a + bx}$

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

$S_y^2 = b^2 S_x^2$

$S_y = |b| S_x$

EJM

La media de $x = 8$ y la varianza $= 4$

¿Que transformación? para tener Media $y = 42$ y desviación típica $= 10$

Media → $\bar{y} = a + b\bar{x} \Rightarrow 42 = a + b\cdot 8$

d. típica → $S_y = b\cdot S_x \Rightarrow 10 = b\cdot 2 \Rightarrow \underline{b = 5}$

$\Downarrow$

$a = 42 - (5\cdot 8) = \underline{2} \Rightarrow y = 2 + 5x$

covarianza sin agrupar

$$S_{xy} = \frac{\sum x_i \cdot y_i}{N} - \bar{x} \cdot \bar{y}$$

con agrupados

$$S_{xy} = \frac{\sum f_i \, x_i \cdot y_i}{N} - \bar{x} \cdot \bar{y}$$

>0 Directa 

<0 Inversa 

$=0$ Incorreladas 

coeficiente de correlación

$$r = \frac{S_{xy}}{S_x S_y} \quad [\text{Entre } -1 \text{ y } 1]$$

S_{xy} covarianza

S_x, S_y } Desviación estandar

$$y = a + bx$$

$$b = \frac{S_{xy}}{S_x^2} \qquad a = \bar{y} - \underbrace{\frac{S_{xy}}{S_x^2}}_{b} \cdot \bar{x}$$

$$y = \bar{y} + \frac{S_{xy}}{S_x^2}(x - \bar{x})$$

---

x = Nº de accidentes

y = nº de vehiculos a + de 120 km/h

x =	5	7	2	1	9
y =	15	18	10	8	20

¿Recta regresión y coeficiente de correlación?

$$y = \bar{y} + \frac{S_{xy}}{S_x^2}(x - \bar{x})$$

$$\bar{x} = \frac{5+7+2+1+9}{5} = 4'8 \qquad S_x^2 = \frac{\sum x_i^2}{N} - \bar{x}^2 = 8'96$$

$$\bar{y} = \frac{15+18+10+8+20}{5} = 14'2 \qquad S_y^2 = \frac{\sum y_i^2}{N} - \bar{y}^2 = 20'96$$

$$S_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \cdot \bar{y} = 13'64$$

$$y = 14'2 + \frac{13'64}{8'96}(x - 4'8) \Rightarrow y = 6'89 - 1'52x$$

• coeficiente de correlación

$$r = \frac{13'64}{\sqrt{8'96}\,\sqrt{20'96}} = 0'995$$

• $x = 6 \Rightarrow$ ¿y?

$$y = 6'89 - 1'52(6) \Rightarrow y = 16'01$$

Tema -3-

$$P(A) = \frac{\text{casos favorables (cf)}}{\text{casos posibles (cp)}} \qquad P(A) = \frac{N_A}{N}$$

Probabilidad relativa $0 \leq P(x) \leq 1$

~~Suceso/Evento elemental~~ :

$E_2 = \{x / x = 1,2,3,4,5,6\}$ $\quad E_1\{x / 0 \leq x \leq 100\}$

$$P(A_2) = \frac{P(A_2 \cap B_1) + (A_2 \cap B_2)}{N}$$

sabiendo que ha ocurrido B, ...

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

se elige aleatoriamente una mujer (B_2), ¿proba. de fumar (A_1)?

$$P(A_1/B_2) = \frac{P(A_1 \cap B_2)}{P(B_2)}$$

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

$$P(A \cap B \cap B) = P(A \cap B) \cdot P\left(\frac{C}{A \cap B}\right)$$

A times (20) $\quad$ Ambos (1) = $P(A \cap B)$

B New (16)

$$\cancel{P(A)} \; P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0'01}{0'20} = \boxed{0'05}$$

A_1 Azul

A_2 verde

A_3 ~~Rojo~~ verde

$$P(A_1 \cap A_2 \cap A_3) = \frac{2}{10} \times \frac{5}{9} \times \frac{4}{8} = \frac{1}{18}$$

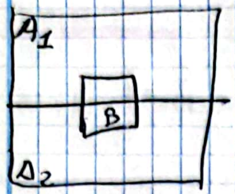
$$P(A_3 / A_1 \cap A_2) = \frac{4}{8}$$

¡¡ Si son independientes !!

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

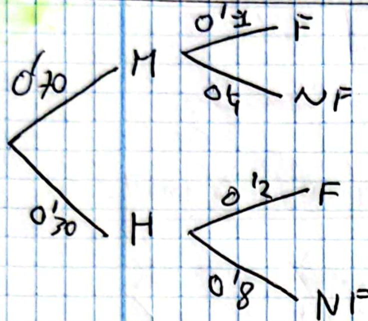
$$P(A \cap B) = P(A) \cdot P(B)$$

teorema de pro. total



$$P(B) = P(B\cap A_1) + P(B\cap A_2)$$

$$P(B) = P(A_1)\cdot P(B/A_1) + P(A_2)\cdot P(B/A_2) \qquad \Big|\ \sum^{i} P(A_i)\cdot P(B/A_i)$$



$$P(F) = P(M\cap F) + P(H\cap F) =$$
$$= P(M)\cdot P(F/M) + P(H)\cdot P(F/H) =$$
$$= 0'7\cdot 0'1 + 0'3\cdot 0'2 = \underline{13\%}$$

teorema de Bayes

¿un individuo al azar, es fumador, pro. H?

$$P(H/F) = \frac{P(H\cap F)}{P(F)} = \frac{P(H)\cdot P(F/H)}{P(F)} = \frac{0'3\cdot 0'2}{0'13} = \underline{0'46}$$

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

$$P(F) = P(M \cap F) + P(H \cap F) =$$
$$= P(M) \cdot P(F/M) + P(H) \cdot P(F/H)$$

$$P(H/F) = \frac{P(H \cap F)}{P(F)} = \frac{P(H) \cdot P(F/H)}{}$$

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

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

$$P(B - A) = P(B) - P(A)$$

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

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

---

se elige aleatoriamente una ~~fum~~ mujer ¿ P. de que sea fumadora?

$$P(F|M) = \frac{P(F \cap M)}{P(M)}$$

EJERCICIOS - Tema 2 -

1)

x	5	7	10	13	15
y	2	3	4	5	6

$N = 5$

$\bar{x} = 10$ $\quad S_x^2 = \frac{\sum x^2}{N} - \bar{x}^2 = 13'6$ $\quad S_{xy} = \frac{\sum x \cdot y}{N} - \bar{x} \cdot \bar{y} = \frac{226}{5} - \bar{x} \cdot \bar{y} = 5'2$

$\bar{y} = 4$

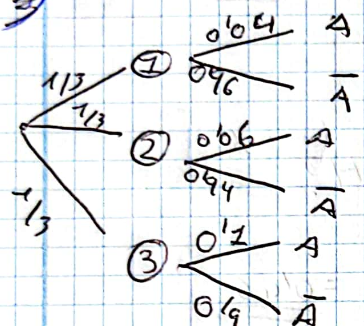
$S_y^2 = \frac{\sum y^2}{N} - \bar{y}^2 = 2$ $\quad y = \bar{y} + \frac{S_{xy}}{S_x^2}(x - \bar{x})$

$\boxed{y = -5'61 + 0'38x}$

$r = \frac{S_{xy}}{S_x \, S_y} = \boxed{0'997}$

EJERCICIOS - Tema 3 -

3)



a) $\cancel{P(A1) = \frac{1/3 \cdot 0'04 + 0'06 + 0'1}{3} = 0'06}$

P(A)

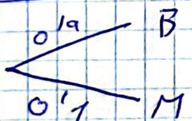
a) $P(1\bar{A}) = \left(\frac{1}{3} \cdot 0'04\right) + \left(\frac{1}{3} \cdot 0'94\right) + \left(\frac{1}{3} \cdot 0'9\right) +$

$+ \left(\frac{1}{3} \cdot 0'96\right) + \left(\frac{1}{3} \cdot 0'06\right) + \left(\frac{1}{3} \cdot 0'9\right) +$

$+ \left(\frac{1}{3} \cdot 0'96\right) + \left(\frac{1}{3} \cdot 0'94\right) + \frac{1}{3} \cdot 0'1) = \frac{\sum}{3}$

---

1 pedido



Tema -4- Inferencia estadística

- No se conocen exactos los valores de las variables

Variables Aleatorias

* Discretas: El conjunto de posibles valores es numerable
  - Nº de páginas de un libro
* Continuas: El conjunto de posibles valores NO es numerable
  - tiempo de respuesta de 1 servicio (1 día y 2 horas)
  - Cantidad de agua consumida en un mes

Función de probabilidad ó **Función de masa**. (Variable discreta)

X es una variable con posibles valores $\{x_1, x_2, \dots\}$.

La función de probabilidad es el conjunto de probabilidades con las que X toma cada uno de esos valores.

EJM X = resultado de lanzar un dado

X	1	2	3	4	5	6
$P[X = x]$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- **Propiedades**

- $0 \leq P[X = x] \leq 1$
- $\sum_i P[X = x_i] = 1$
- $P[X \leq x] = \sum_{i, x_i < x} P[X = x_i]$
- $P[X > x] = 1 - P[X \leq x]$

Función de distribución (variables discretas)

Especifica para cada valor del espacio muestral, la probabilidad de aparición de un valor menor o igual

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

EJM X = Resultado de lanzar un dado

X	1	2	3	4	5	6
$P[X = x]$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$F(x)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	$\frac{6}{6}$

- **Propiedades**
- $F(-\infty) = 0$
- $F(\infty) = 1$
- si $x \leq y \Rightarrow F(x) \leq F(y)$

Para sustituir la función de probabilidad en variables continuas

⇓

Usamos la Función de densidad

- Especifica la probabilidad de aparición de cada valor del espacio muestral

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

- Positiva en todos sus puntos

La función distribución para variables continuas se define como:

En el caso discreto, por la probabilidad de que X tome un valor menor o igual a x especificado

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)\,dt$$

- propiedades

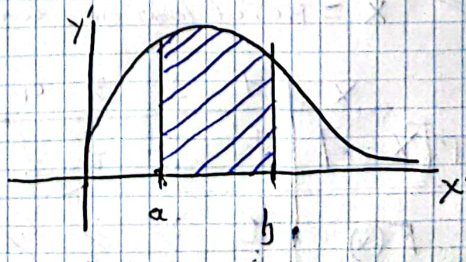
Son funciones de tipo suave, NO Escalones

- $F(-\infty) = 0$
- $F(\infty) = 1$
- $F(x)$ es continua
- si $x \leq y \Rightarrow F(x) \leq F(y)$

- Propiedades función de densidad continua

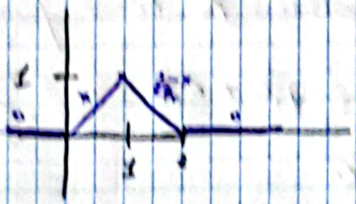
- $f(x) \geq 0$
- $P(a \leq x \leq b) = \int_a^b f(x)\,dx$
- $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$



EJ 14 Dada la siguiente función de densidad, obtener la función de distribución

$$f(x) = \begin{cases} x & 0 \le x \le 1 \\ 2-x & 1 \le x \le 2 \\ 0 & \text{otros valores} \end{cases}$$



$$F(x) \begin{cases} 0 & x < 0 \\ & 0 \le x \le 1 \\ & 1 \le x \le 2 \\ & x > 2 \end{cases}$$

1º tramo $(x < 0)$

$$\int_{-\infty}^{x} 0 \, dt = 0$$

2º tramo $(0 \le x \le 1)$

$$\int_{-\infty}^{0} 0 \, dt + \int_{0}^{x} t \, dt = \frac{x^2}{2}$$

3º tramo $(1 \le x \le 2)$

$$\int_{-\infty}^{0} 0 \, dt + \int_{0}^{1} t \, dt + \int_{1}^{x} (2-t) \, dt = 0 + \left[\frac{t^2}{2}\right]_0^1 + \left[2t - \frac{t^2}{2}\right]_1^x = \boxed{\frac{1}{2}} + \boxed{2x - \frac{x^2}{2}} - \boxed{2 + \frac{1}{2}} =$$

$$= 2x - \frac{x^2}{2} - 1$$

4º tramo $(x > 2)$

$$\int_{-\infty}^{0} 0 \, dt + \int_{0}^{1} t \, dt + \int_{1}^{2} (2-t) \, dt + \int_{2}^{\infty} 0 \, dt = 1$$

Su función de distribución:

$$F(x) \begin{cases} 0 & x < 0 \\ \frac{x^2}{2} & 0 \le x \le 1 \\ +2 - \frac{x^2}{2} - 1 & 1 \le x \le 2 \\ 1 & x > 2 \end{cases}$$

Ejs.

Una variable aleatoria X tiene función de densidad

$$f(x) = \begin{cases} 12x^2(1-x) & 0 < x < 1 \\ 0 & \text{Resto} \end{cases}$$

$$P(X \leq 0'5) = \int_{-\infty}^{0'5} f(t)\,dt = \int_{0}^{0'5} 12t^2(1-t)\,dt = 0'3125$$

$$P(0'2 \leq x \leq 0'5) = \int_{0'2}^{0'5} 12t^2(1-t)\,dt = 0'2853$$

$$P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0 & x \leq 0 \\ 12\left(\frac{x^3}{3} - \frac{x^4}{4}\right) & 0 < x \leq 1 \\ 1 & x > 1 \end{cases}$$

Esperanza (ó media)

- Discretas $= \mu_x = E(x) = \sum_{i=1}^{n} X_i \cdot p(x_i)$

- continuas $= \mu_x = E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$

Varianza

- Discretas: $\sigma_x^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 \cdot p(x_i)$

- continuas $= \sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x) \cdot dx = E[x^2] - E[X]^2$

Propiedades Esperanza

$E[ax] = a \cdot E[x]$

$E[X+Y] = E[x] + E[Y]$

$E[ax + bY] = aE[x] + b\,E[Y]$

Propiedades varianza

$Var[x] = 0 \iff X$ es contínua

a es constante $\Rightarrow Var[ax] = a^2 \cdot Var[x]$

a y b " " $\Rightarrow Var[ax+b] = a^2 \cdot Var[x]$

EJM (Variable discreta)

Calcula el beneficio esperado apostando 100 € en la ruleta:

a/ un número cualquiera.

• La probabilidad de acierto es de $1/37$

$\Downarrow$

si apostamos 100 € $\Rightarrow$ La variable aleatoria X : beneficio obtenido

* $X = -100$ si no acierto

$\Downarrow$

$P(X = -100) = \frac{36}{37}$

* $X = 3500$ si acierto

$\Downarrow$

$P(X = 3500) = \frac{1}{37}$

Por tanto:

$$E(x) = -100\left(\frac{36}{37}\right) + 3500\left(\frac{1}{37}\right) = -2'7 €$$

b/ Rojo frente a negro

si sale rojo $\Rightarrow$ +100

si negro ó 0 $\Rightarrow$ −100

$$E(x) = -100\left(\frac{19}{37}\right) + 100\left(\frac{18}{37}\right) = -2'7 €$$

EJM (Va. discreta)

Una variable aleatoria X puede tomar los valores 30, 40, 50, 60 con probabilidades 0'4, 0'2, 0'1, 0'3.

Representa la función de probabilidad y de distribución

X	$P(X=x)$	$x \cdot P(X=x)$	$x^2 \cdot P(X=x)$
30	0'4	12	360
40	0'2	8	320
50	0'1	5	250
60	0'3	18	1080
	1	43	2010

$$E(x) = \sum_{i=1}^{k} P(x=x_i) = 12+8+5+18 = 43$$

$$V(X) = \sum x^2 P(X=x_i) - E(x)^2 = 210 - 43^2 = 161$$

$$\sigma = \sqrt{161} = 12'69$$

Ej: (v.a. continua)

La altura de un arbol sigue una V.a. con la siguiente función de densidad

$$f(x) = \begin{cases} \frac{x}{12} & \text{si } 1 < x < 5 \\ 0 & \text{El resto} \end{cases}$$

¿Esperanza y Varianza?

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_{1}^{5} \frac{x^2}{12}\, dx = \frac{1}{36}\left[x^3\right]_1^5 = 3'444 = \frac{31}{9}$$

$$\underbrace{E[x^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{1}{48}\left[x^4\right]_1^5 = 13}$$

$$Var(x) = E[x^2] - E[x]^2 = 13 - \left(\frac{31}{9}\right)^2 = 1'1358$$

Momentos de una V.A: Son los valores esperados de ciertas funciones de las variables.

- no centrados
- centrados en media
- absolutos

¿Como se si $f(x)$ es una variable de densidad?

- Si es positiva para todos sus puntos
- compruebo la integral $-\infty$, a ∞ es igual a 1

Momentos no centrados

Si X es una variable aleatoria y $r \in \mathbb{N}$, se define el momento no centrado de orden r como:

$$m_r = \mu_r = E[X^r]$$

siempre que exista la esperanza

- Si es discreta $\Rightarrow \mu_r = E[x^r] = \sum x^r p(x)$
- Si es continua $\Rightarrow \mu_r = E(x^r) = \int_{-\infty}^{\infty} x^r f(x)\, dx$

Momentos centrados en la media

Si X es v.Aleatoria tal que $\exists\, E[X]$ y $r \in \mathbb{N} \Rightarrow$ se define el momento centrado de orden r como:

$$\mu_r = E[(X - E[X])^r]$$

siempre que exista dicha esperanza

- Si X es discreta $\Rightarrow \mu_r = \sum_x (x-\mu)^r p(x)$
- Si X es continua $\Rightarrow \mu_r = \int_{-\infty}^{\infty} (x-\mu)^r p(x)$

El momento centrado de orden 2 = Varianza

Momentos absolutos

Si X es una variable aleatoria y $\alpha \in \mathbb{R}^+$, se define el momento absoluto de orden α de X como:

$$E[X^\alpha]$$

siempre que exista la esperanza

- Si X es discreta $\Rightarrow E(X^\alpha) = \sum_x |x|^\alpha p(X = x)$
- Si X es continua $\Rightarrow E(X^\alpha) = \int_{-\infty}^{\infty} |x|^\alpha f(x)\, dx$

Discretas	Continuas
• Distribución binomial	• Distribución uniforme (continua)
• " de Poisson	• " Exponencial
	• " Normal

Distribuciones Discretas

• **Bernoulli de parámetro p**

• Solo admite 2 resultados

⇓

X solo admite 1 y 0

$$\mu_x = p \qquad \sigma_x^2 = p \cdot q$$

Ejm

10% desempleados

$x = 1$ (Desempleado) $\qquad P(x=1) = 0'1$

$x = 0$ (NO desemp.) $\qquad P(x=0) = 0'9$

• **Binomial**

• Supongamos un resultado elemental K, de éxitos en N pruebas

La probabilidad será $= p^K (1-p)^{n-K}$

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \sum_{I=1}^{N} \binom{N}{x_i} p^{x_i} (1-p)^{n-x_i} & 0 \leq X \leq N \\ 1 & x \geq N \end{cases}$$

$$E(x) = n \cdot p \qquad VAR = pN \cdot q$$

Ejm

Nº de caras al lanzar 20 veces una moneda.

Nº de aprobados si se presentan 80 alumnos.

Nº de familias con solo 1 hijo en una población de 120 familias

Binomial $\xrightarrow{a}$ Normal

$$B(n, p) \longrightarrow N(n \cdot p, \sqrt{n \cdot p \cdot q})$$

Propiedades

1. La distribución binomial se puede obtener como suma de N variables aleatorias independientes Bernoulli con el mismo parámetro "p".

2.- $X \to B(N, P)$
$Y \to B(M, P)$

$X + Y \to (N+M, P)$

EJM

La probabilidad de éxito de una vacuna es 0'72, se administra a 15 pacien[tes]

a) Ninguno sufra la enfermedad
b) todos sufran la enfermedad
c) 2 contraigan la enfermedad

$B(15, 0'72)$

a) $P(X=15) = \binom{15}{15} \cdot 0'72^{15} \cdot 0'28^{0} = 0'724\%$

b) $P(X=0) = \binom{15}{0} \cdot 0'72^{0} \cdot 0'28^{15} = 5'097 \cdot 10^{-7}\%$

c) $P(X=13) = \binom{15}{13} \cdot 0'72^{13} \cdot 0'28^{2} = 11'50\%$

EJM

La probabilidad de que el carburador de un coche salga defectuoso es 4 por 100

a) El Nº de defectuosos en un lote de 1000
b) varianza

a) $E(x) = N \cdot P = 1000 \cdot 0'04 = 40$

b) $vari = N \cdot P \cdot q = 1000 \cdot 0'04 \cdot 0'96 = 38'4$

• POISSON

X representa el Nº de eventos $\boxed{\text{independientes}}$ que ocurren a velocidad constante en en un intervalo de tiempo.

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$E(x) = \lambda \qquad VAR(x) = \lambda$$

- Se utiliza cuando los procesos son impredecibles o de ocurrencia aleatoria

---

EJM

Si las llamadas que recibe una centralita son sucesos independientes (El que llame 1 persona llame no afecta a que llame otra), y llaman con una media de 5 por min.

X = Nº de llamadas en 1 min $\quad X \rightarrow$ poisson $\lambda = 5$

En 1 hora $\quad Y \rightarrow$ poisson $\lambda^* = 5 \times 60$

EJM

Una centralita 480 llamadas por hora y la centralita tiene la capacidad de 12 por min.

¿P de que en 1 min no sea posible atender la llamada?

$$P(x>12) = 1 - P(x \leq 12) = 1 - \sum_{x=0}^{12} \frac{8^x}{x!} e^{-8} = 0'0638$$

$$x = \text{"Nº de llamadas por min"} = \frac{480}{60} = 8$$

EJM

Un servidor recibe una media de 7 accesos al minuto, se quiere calcular prob. de que reciba más de 10 en 1 min

X = Nº de accesos en un minuto será una POISSON de media 7: $X \rightarrow P(7)$

$$P(x>10) = 1 - P(x \leq 10) = 1 - \sum_{0}^{10} \frac{7^x}{x!} \cdot e^{-7} = 0'099$$

Ejm

A una gasolinera llegan, de media, 3 coches por minuto

a) ¿Cual es la función de probabilidad?

b) ¿Pro. de que lleguen 2 en 1 min?

c) ¿Pro. " " " 12 en 5 min?

a) La función de probabilidad es una POISSON con $\lambda = 3$

$$P(x) = \frac{3^x}{x!} e^{-3}$$

b) $$P(2) = \frac{3^2}{2!} \cdot e^{-3} = 22'4\%$$

c) $\lambda = 3 * 5 = 15$

$$P(12) = \frac{15^{12}}{12!} e^{-15} = 8'28\%$$

Distribuciones Continuas

- **Distribución Continuas (uniforme)**

  - Cuando una variable aleatoria toma valores en un intervalo finito $-\infty < a < b < +\infty$

p.d.
$$f(x) \begin{cases} 0 & \text{Resto} \\ \frac{1}{b-a} & a \leq x \leq b \end{cases}$$

f.d
$$F(x) \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

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

Ejm Se selecciona un Nº al azar del intervalo $[2,6]$ y definimos una variable aleatoria X = "Nº seleccionado"

¿$P(x \leq 5)$?

$$X \sim U(2,6); \; P(x \leq 5) = \int_2^5 f(x)\, dx = \int_2^5 \frac{1}{6-2} dx = \left[\frac{x}{4}\right]_2^5 = \frac{3}{4} = 0'75$$

o

$$P(x \leq 5) = F(5) = \frac{x-2}{6-2} = \frac{5-2}{6-2} = \frac{3}{4} = 0'75$$

$$E(x) = \frac{a+b}{2} = 4$$

Distribuciones Continuas (Exponencial)

- Suelen ser los sucesos que miden tiempo que transcurre entre 2 sucesos.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{Resto} \end{cases} \qquad F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & \text{Resto} \end{cases}$$

$$E(x) = \frac{1}{\lambda} \qquad \sigma_x^2 = \frac{1}{\lambda^2}$$

Distribuciones continuas (Normal, Gauss)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$E(x) = \mu \qquad VAR(x) = \sigma^2$$

tipificación:

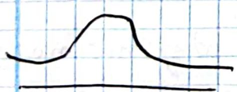
La variable $X \to N(\mu, \sigma)$ se transforma mediante la siguiente expresión:

$$\boxed{Z = \frac{X-\mu}{\sigma}}$$

$$f(Z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right)$$

$$F(x) = F_Z\left(\frac{x-\mu}{\sigma}\right)$$

- $F(-x) = 1 - F(x)$



- $P[Z \geq a] = 1 - P[Z \leq a]$
- $P[Z \geq -a] = P[Z \leq a]$
- $P[Z \leq -a] = 1 - P[Z \leq a]$
- $P[|Z| \leq a] = 2P[Z \leq a] - 1$

- $P(Z > 1'24) = 1 - P(Z \leq 1'24)$
- $P(Z \leq -0'72) = P(Z \geq 0'72) = 1 - P(Z < 0'72)$
- $P(0'5 \leq Z \leq 1'76) = P(Z \leq 1'76) - P(Z \leq 0'5)$

Ejm

a) $P(Z \leq 2'17)$

$P(Z \leq 2'17) = 0'9850$

b) $P(Z \geq 1'32)$

$P(Z \geq 1'32) = 1 - P(Z \leq 1'32) = 1 - 0'9066 = 0'0934$

c) $P(1'52 < Z < 2'03)$

$P(1'52 < Z \leq 2'03) = P(Z \leq 2'03) - P(Z \leq 1'52) = 0'9788 - 0'9357 = 0'0431$

d) $P(Z \geq -1'32)$

$P(Z \geq -1'32) = P(Z \leq 1'32) = 0'9066$

Ejm La vida de un semiconductor se distribuye normalmente con media 7000 horas y desviación típica 600 horas

¿$P(6280 \leq X \leq 7120)$?

$X \equiv$ vida del semiconductor (en horas) $\sim N(7000, 600)$

$$Z = \frac{X - 7000}{600} \sim N(0,1), \quad \frac{6280 - 7000}{600} = -1'2, \quad \frac{7120 - 7000}{600} = 0'2$$

$$P(6280 \leq X \leq 7120) = P(-1'2 \leq Z \leq 0'2) =$$

$$= P(Z \leq 0'2) - P(Z \leq -1'2) = P(Z \leq 0'2) - (1 - P(Z < 1'2)) =$$

$$= 0'5793 - 0'1151 = \underline{\underline{0'4642}}$$

Ejm Se sabe que, debido a los procesos de llenado, el contenido de una lata de bebida de 33 cl y una desviación típica de 2 cl.

a) La función de densidad

La función de densidad es una Normal (33, 2)

$$f(x = x_i) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{1}{8}(x_i - 33)^2\right)$$

b) contenido superior a 35 cl

$$P(X \geq 35) = P\left(Z \geq \frac{35 - 33}{2}\right) = 1 - P(Z \leq 1) = 15'87\%$$

c) Si un pack consta de 6 latas ¿Proba. inferior a 192 cl?

sea $Y = 6$ latas $\Rightarrow$ se distribuye $N(6*33, \sqrt{6}*2) = N(198, 4'899)$

$$P(Y < 192) = P\left(Z < \frac{192 - 198}{4'899}\right) = P(Z < -1'22) = 0'112$$

Factor de corrección

- si es $>$ ⇒ sumo 0'5
- si es $\geq$ ⇒ Resto 0'5

Tema -5-

<Conceptos>

- Inferencia: Extraer juicios o conclusiones a partir de ciertos supuestos, sean estos generales o particulares.

- Población: Es el conjunto total de datos que existen sobre un variable.

- Parámetro: Cantidad numérica calculada sobre una población.

- Muestra estadística: Una muestra es una parte de la población de datos.

- Escoger una muestra adecuada corre a cargo de las distintas técnicas de muestreo.

- Estadístico: cantidad numérica calculada sobre una muestra

- Estimador: Si un estadístico se usa para aproximar un parámetro, también se suele llamar estimador.

- Estadísticos muestrales: Cualquier cantidad obtenida de una muestra con el propósito de obtener un parámetro poblacional se llama: estadística muestral.

- Error muestral: Cualquier medida conlleva algún error

Se ha obtenido una muestra aleatoria de tamaño 25 de una población con media $\mu = 15$, si la media de la muestra es $\bar{X} = 12$, entonces a la diferencia observada.

$$\varepsilon = \bar{x} - \mu = 12 - 15 = \underline{\underline{-3}}$$

Se toman muestras de tamaño 2 de una población consistente en tres valores 2, 4 y 6, para simular una población "grande" de manera que el muestreo pueda realizarse un gran número de veces, supondremos realizarlo con reemplazo.

La media poblacional es: $\mu = \frac{2+4+6}{3} = 4$

Muestras ordenadas	$\bar{x}$	ERROR MUESTRAL $\varepsilon = \bar{X} - \mu$
(2, 2)	2	2-4 = -2
(2, 4)	3	3-4 = -1
(2, 6)	4	4-4 = 0
(4, 2)	3	3-4 = -1
(4, 4)	4	4-4 = 0
(4, 6)	5	5-4 1
(6, 2)	4	4-4 = 0
(6, 4)	5	5-4 = 1
(6, 6)	6	6-4 = 2

La media de las medias muestrales es:

$$\bar{x} = \frac{2+3+4+3+4+5+4+5+6}{9} = 4$$

La suma de los errores muestrales es:

$$\varepsilon = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_9 = 0$$

Métodos básicos de inferencia estadística:

- Estimación: Proceso mediante el que establecemos que valor debe tener un parámetro según deducciones que realizamos.
- contraste de hipótesis: Su objetivo es comprobar si una estimación se adapta a los valores poblacionales.

<Métodos de muestreo>

- Muestreo aleatorio simple: Cada elemento de la población tiene la misma probabilidad de ser elegido para la muestra.

- Muestreo aleatorio sistemático: Se eligen los elementos de la población con una periodicidad determinada.

- Muestreo aleatorio estratificado: Los elementos de la población se dividen en estratos o clases.

- Muestreo aleatorio por conglomerados: Los elementos de la población se encuentran en conglomerados que se suponen representativos de la población.

En todos los casos se puede dar:
- Muestreo con remplazamiento
- Muestreo sin remplazamiento

- Simple:
  - todos tienen la misma probabilidad de ser elegidos
  - la selección se realiza en una sola etapa
  - tiene un coste muy alto su aplicación

- Sistemático:
  - Parecido al anterior pero más cómoda la elección
  - Para obtener una muestra "n" ~~de~~ en una población "N", se ordenan y numeran los elementos de la población.
  - Ojo: Si la muestra contiene periodicidades, obtendremos una muestra sesgada.

- Estratificado:
  - Se aplica cuando sabemos que existen ciertos factores que pueden influir en el estudio y queremos asegurarnos de tener cierta cantidad mínima de individuos
    - Hombres y Mujeres
    - Jóvenes, adultos y ancianos
  - Se realiza entonces una m.a.s.

- Conglomerado:
  - se aplica cuando no se dispone de una lista de todos los individuos, pero sabemos que se encuentran agrupados naturalmente en grupos.
  - se eligen varios grupos al azar, y entonces podemos estudiar todos los individuos de los grupos
  - Cuidado extrapolar resultados

<< Estimación puntual >>

- Cuando se usa un solo valor extraído de la muestra para estimar el parámetro desconocido de la población.

El valor usado se le llama estimador.

Media muestral

$$\bar{x} = \frac{x_1 + x_2 + \dots}{N}$$

Varianza muestral

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

Cuasi-varianza muestral

$$\hat{S}^2 = \frac{\sum (x_i - \bar{x})^2}{N-1}$$

- Propiedades de los estimadores -

- Insesgadez: Un estimador es insesgado cuando la esperanza matemática del este es igual al parámeto que se desea estimar.
- Eficiencia: Un estimador es más eficiente cuando su varianza es reducida.
- Consistencia: Un estimador consistente es aquel que a medida que la muestra crece, se aproxima cada vez más al valor del parámetro real.

Para la Media muestral

$$E(\bar{x}) = \mu$$

$$VAR(\bar{x}) = \frac{\sigma^2}{N}$$

Para la Varianza muestral

$$E(s^2) = \frac{n-1}{n}\sigma^2$$

Para la Cuasivarianza muestral

$$E(S_{n-1}^2) = \sigma^2$$

<Distribución muestral>

para la media muestral:

$E(\bar{x}) = \mu \rightarrow$ (Media)

$VAR(\bar{x}) = \dfrac{\sigma^2}{n}$

EJM

• Se han seleccionado muestras aleatorias de tamaño 20 en una población grande.
Se calcula la media muestral "s" para cada muestra.
La colección de todas estas medias muestrales recibe el nombre de "distribución muestral de desviaciones típicas".

EJM

• Se eligen muestras de tamaño $n=2$, con reemplazo de la población de valores 0, 2, 4, 6.

a/ Media y desviación estandar poblacional?

$$\mu = \frac{0+2+4+6}{4} = 3$$

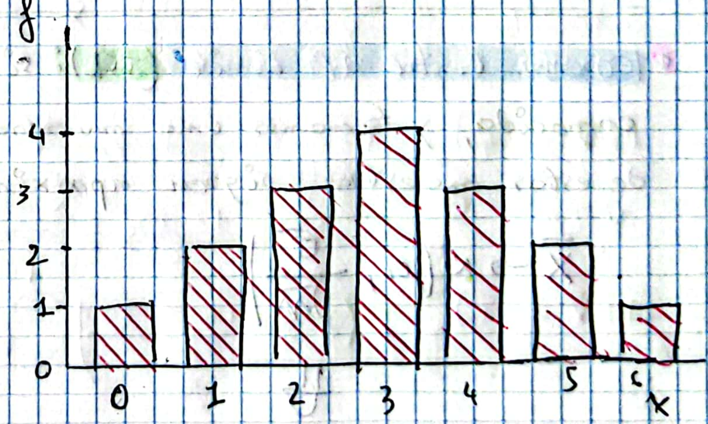
$$\sigma = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{56}{4} - 3^2} = 2'2361$$

b/ Media y desviación típica de la distribución muestral de medias, gráficos.

Muestra	$\bar{x}$
(0,0)	0
(0,2)	1
(0,4)	2
(0,6)	3
(2,0)	1
⋮	
(6,6)	6

$\Rightarrow$

$\bar{x}$	f
0	1
1	2
2	3
3	4
4	3
5	2
6	1
	16



Media muestral

$$\mu_{\bar{x}} = \frac{\sum (f \cdot \bar{x})}{\sum f} = \frac{0\cdot 1 + 1\cdot 2 \ldots}{16} = \boxed{3}$$

b) Distribución muestral de frecuencias de la desviación típica, desviación típica de la distribución muestral

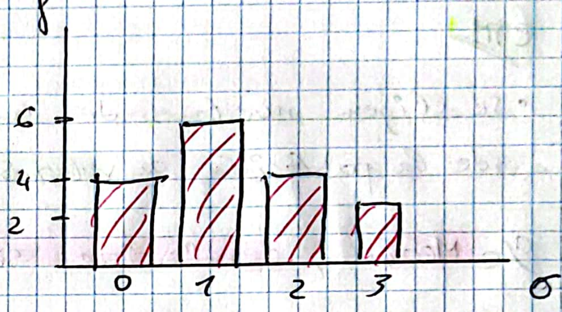
Muestra	x_1	x_2	$\bar{x}$	s^2	$s = \sigma$
(0,0)	0	0	0	0	0
(0,2)	0	2	1	1	1
(0,4)	0	4	2	4	2
(0,6)	0	6	3	9	3
(2,0)	2	0	1	1	1
⋮	⋮	⋮	⋮	⋮	⋮
(6,6)	6	6	6	0	0

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$s^2 = \frac{(0-3)^2 + (6-3)^2}{2} = \frac{9+9}{2} = 9$$

$\bar{x}^2 = 184$

$s = \sigma$	f
0	4
1	6
2	4
3	2
	16



$$s = \sqrt{\frac{\sum (\bar{x}_i - \mu_x)^2}{\sum f}} = \sqrt{\frac{\sum \bar{x}_i^2}{\sum f} - \mu_x^2} = \sqrt{\frac{184}{16} - 3^2} = 1'5811$$

Media muestral.
pag ante

• Teorema Central del Límite (TCL): si una población tiene μ y σ conocida, y tomamos una muestra n ($n > 30$), las medias de estas muestras siguen aproximadamente la distribución:

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\Downarrow$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$$

• Factor de corrección para población finita:

En la práctica, a la hora de seleccionar se hace SIN reemplazo;
En estos casos es necesario emplear un "Factor de corrección" para la desviación estandar, llamado "coeficiente corrector de poblaciones finitas" o "coeficiente de exhaustividad".

$$\sqrt{\frac{N-n}{N-1}}$$

→ Esta corrección implica el siguiente cambio:

$$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

n: tamaño de la muestra.

N: tamaño población.

por tanto:

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)$$

COMIENZO — Distribución muestral de medias

¿Es una población infinita? — sí → $N\left(\mu_x, \frac{\sigma}{\sqrt{n}}\right)$

NO ↓

Tiene sustitución — sí → $N\left(\mu_x, \frac{\sigma}{\sqrt{n}}\right)$

NO ↓

Es $N \geq 20n$ — sí → $N\left(\mu_x, \frac{\sigma}{\sqrt{n}}\right)$

↓

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)$$

Aproximar poisson → Normal

$N(\lambda, \sqrt{\lambda})$

• si $\lambda > 16$

EJM

• Para la distribución muestral de medias del ejercicio anterior

a) Error muestral de cada media

Muestra	$\bar{X}$	$\varepsilon = \bar{X} - \mu$
(0,0)	0	$0-3 = -3$
(0,2)	1	$1-3 = -2$
(0,4)	2	$2-3 = -1$
(0,6)	3	$3-3 = 0$
(6,6)	6	$6-3 = 3$

$\Sigma \varepsilon_i^2 = 40$

b) Media y desviación típica de los errores muestrales

$$\mu_\varepsilon = \frac{-3 + -2 + -1 + 0 + \dots + 3}{16} = 0$$

$$S_\varepsilon = \sigma_\varepsilon = \sqrt{\frac{\Sigma \varepsilon_i^2}{N} - \bar{x}^2} = \sqrt{\frac{40}{16} - 0^2} = 1'5811$$

EJM

• Distribución muestral de la media: Población infinita varianza conocida

Sea X población con distribución $N(\mu = 90, \sigma = 20)$

a) si se obtiene una m.a.s. de tamaño 16 ¿Probabi. de que muestra X sea > 92?

$$\left.\begin{array}{l} E(x) = \mu = 90 \\ VAR = \dfrac{\sigma^2}{n} = \dfrac{20^2}{16} \Rightarrow \sigma_{\bar{x}} = \dfrac{\sqrt{400}}{16} = 5 \end{array}\right\} \bar{X} \sim N(90, 5)$$

$$P(\bar{X} \geq 92) = P\left(Z \geq \frac{92-90}{5}\right) = P\left(Z \geq \frac{2}{5}\right) = \boxed{0'3446}$$

b) Determinar el tamaño muestral para que la probabilidad de que la media muestral sea menor o igual que 98 sea 0'99

$$0'99 = P(\bar{X} \leq 98) = P\left(N\left(90, \frac{20}{\sqrt{n}}\right) \leq 98\right) = P\left(Z \leq \frac{98-90}{\frac{20}{\sqrt{n}}}\right) =$$

$$P\left(Z \leq \sqrt{n} \cdot \frac{2}{5}\right) \Rightarrow \sqrt{n} \cdot \frac{2}{5} = 2'33 \Rightarrow n = \left(\frac{2'33 \cdot 5}{2}\right)^2 = n = 33'9 \Rightarrow \boxed{n \geq 34}$$

Distribución muestral de media Poblaciones finitas

Varianza conocida

La estatura de 1000 estudiantes sigue una dist. normal media = 174'5 cm y $\sigma = 6'9$ cm.

~~muestra de~~ 200 muestras de tamaño 25 SIN Reemplazo

a) Nº de medias muestrales entre 172'5 y 175'8

$N = 1000$, $n = 25$ $\mu = 174'5$ $\sigma = 6'9$

$$\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{6'9}{\sqrt{25}}\sqrt{\frac{1000-25}{1000-1}} = 1'36$$

$$Z = \frac{172'5 - 174'5}{1'36} = -1'47$$

$$Z = \frac{175'8 - 174'5}{1'36} = 0'96$$

$$P(-1'47 \leq Z \leq 0'96) = 0'761$$

Como hay 200 muestras $\Rightarrow 200 \cdot 0'761 = 152$ medias

b) Nº de medias muestrales que caen por debajo de 172

$$P\left(Z \leq \frac{172 - 174'5}{1'36}\right) = P(Z \leq -1'83) = 0'0336$$

$\downarrow$

x 200 muestras $\Rightarrow$ 7 medias

* Poblaciones Infinitas // Varianza desconocida

Parámetros muestrales

Media muestral $\bar{x} = \dfrac{x_1 + x_2 \dots + x_n}{n}$

Varianza muestral $\sigma^2 = \dfrac{\Sigma (x_i - \bar{x}_0)^2}{n}$

Cuasi-varianza muestral: $S^2 = \dfrac{\Sigma (x_i - \bar{x})^2}{n-1}$

- desconocemos σ^2 pero si tipificamos

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

- si desconocemos σ, se puede sustituir por una t-student con n-1 grado libertad

$$T = \frac{\bar{x} - \mu}{\hat{S} / \sqrt{n}} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n-1}}}$$ t-student

- se utiliza si tamaño muestra $n \leq 30$

- si $n > 30 \Rightarrow T \to N(0,1)$

EJM

Supongamos X = "Nº de microorganismos de partículas", esta normalmente distribuida.

16 mediciones se obtiene Cuasidesviación típica 10'8585

- obtener probabilidad de que media muestral no difiera de la media poblacional + de 8 unidades.

$\hat{S} = 10'8585$ $\quad n = 16$

$$T = \frac{\bar{x} - \mu}{\frac{10'8585}{\sqrt{16}}} = t_{16-1} = t_{15}$$

$$P(-8 \le \bar{X} - \mu \le 8) = P\left(\frac{-8}{\frac{10'8585}{\sqrt{16}}} \le \bar{x} - \mu \le \frac{8}{\frac{10'8585}{\sqrt{16}}}\right) =$$

$$= P(-2'947 \le t_{15} \le 2'947) = 1 - 2P(t_{15} \le 2'947) =$$

$= 1 - 2 * (0'995) = 0'99$ es la probabilidad de que la media poblacional difiera en + de 8 unidades.

EJM • Se realizan 36 mediciones, se obtiene cuasi desviación típica 12. Obtener probabilidad de que muestra no difiera en 5

$$T = \frac{\bar{X} - \mu}{\hat{S}/\sqrt{n}} = \frac{\bar{X} - \mu}{\frac{12}{\sqrt{36}}} = \frac{\bar{X} - \mu}{2} \rightarrow t_{35} \cong N(0,1)$$

$$P(-5 \le \bar{X} - \mu \le 5) = P\left(\frac{-5}{2} \le \frac{\bar{X} - 5}{2} \le \frac{5}{2}\right) = P(-2'5 \le Z \le 2'5) = 0'98758$$

Distribución muestral de proporciones

Si tiene población infinita, distribuida como una binomial, para cada muestra se determina el estadístico proporción de éxito $\hat{p}$.

$$\mu_p = P$$

$$\sigma_p = \sqrt{\frac{P \cdot (1-P)}{n}} \quad (1-P) \rightarrow = q$$

• A medida que $n \geq 30$, se aproxima a una normal

$$\hat{P} \sim N\left(P, \sqrt{\frac{Pq}{n}}\right)$$

EJM • Si tiramos una moneda 100 veces
¿Probabilidad de obtener +55 caras?

$p = 0'5 \quad q = 0'5 \quad n = 100$

$$\mu_p = 0'5 \qquad \sigma_p = \sqrt{\frac{0'5 \cdot 0'5}{100}} = 0'05 \Rightarrow N(0'5, 0'05)$$

$$P(X \geq 0'55) = P\left(Z > \frac{0'55 - 0'5}{0'05}\right) = P(Z > 1) = 1 - P(Z \le 1) = \boxed{0'1587}$$

Distribución de diferencia de medias

- Si $\overline{X}_1$ y $\overline{X}_2$ son independientes.

- $\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$

- $\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

$$\overline{X}_1 - \overline{X}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

EJ/

2 tipos de gasolina

1º desviación estandar 1'23 Km/L y n = 35

2º desviación estandar 1'37 Km/L y n = 42

a) ¿Cuál es la probabilidad de que la primera gasolina de un rendimiento promedio mayor de 0'45 Km/L que la segunda?

$$P(|\overline{X}_1 - \overline{X}_2| > 0'45) = P(Z > 1'52) = \boxed{0'06426}$$

$$Z = \frac{0'45 - (0-0)}{\sqrt{\frac{1'23^2}{35} + \frac{1'37^2}{42}}} = 1'52$$

b) ¿Cuál es la probabilidad de que la diferencia en rendimientos promedio se encuentre entre 0'65 y 0'83 Km/L a favor de la 1?

$$P(0'65 \leq |\overline{X}_1 - \overline{X}_2| \leq 0'83) = P(2'19 \leq Z \leq 2'80) = \boxed{0'0117}$$

$$Z = \frac{0'65 - 0}{\sqrt{\frac{1'23^2}{35} + \frac{1'37^2}{42}}} = 2'19$$

$$Z = \frac{0'83 - 0}{\sqrt{\frac{1'23^2}{35} + \frac{1'37^2}{42}}} = 2'80$$

La duración media de marca A es 18, B es 16, sus desviaciones típicas son 3 y 5. Se toman 75 de A y 50 de B. ¿probabilidad duración media de A ≥ 1 que B?

$\mu_1 = 18$ $\mu_2 = 16$ $\sigma_1 = 3$ $\sigma_2 = 5$ $n_1 = 75$ $n_2 = 50$

n_1 y $n_2 > 30$

⇓

$$\bar{X}_1 - \bar{X}_2 \sim \left(18-16, \sqrt{\frac{3^2}{75} + \frac{5^2}{50}} \right) = N(2, 0'787)$$

⇓

$$P(\bar{X}_1 - \bar{X}_2 > 1) = P\left(Z > \frac{1-2}{0'787}\right) = \boxed{0'8980}$$

Cuasivarianza muestral

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 \ldots}{n-1}$$

$$s'^2 = \frac{n}{n-1} \qquad E(s'^2) = \sigma^2$$

• Encontrar la probabilidad de que una muestra aleatoria de 25 observaciones, de una población normal con varianza igual a 6, tenga una varianza muestral > 9'45

$\frac{n S_n^2}{\sigma^2}$ se distribuye como una $\underbrace{\chi^2_{(n-1)}}_{\text{chi-cuadrado con}}$ n-1 grados de libertad

⇓

24 grados de libertad → $\frac{25 \cdot 9'45}{6} = 39'4$

$$P(S_n^2 > 9'45) = 1 - P(S_n^2 < 9'45) = 1 - P(\chi^2_{(n-1)} < 39'4) = 1 - 0'975 = \boxed{0'025}$$

14 • Una población se compone de 5 números: 2, 3, 6, 8, 11, muestras de tamaño 2 con reemplazo.

a) Media y desviación típica

$$\mu = \frac{2+3+6+8+11}{5} = 6$$

$$\sigma^2 = \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5} = 10'8$$

$$\sigma = \sqrt{10'8} = 3'2863$$

b) Media y desviación de la distri. muestral de medias

hay 25 muestras de tamaño 2

$$\overline{\mu}_x = \frac{150}{25} = 6 = \mu \qquad \sum \quad (2,2) \rightarrow \frac{2+2}{2}, \quad (2,3) \rightarrow \frac{2+3}{2} \ldots$$

$$S_x^2 = \text{Varianza} = \frac{\sum (x_i - \overline{\mu}_x)^2}{m} = \frac{135}{25} = 5'4$$

$$S_x = \sqrt{5'4} = 2'32$$

c) varianzas muestrales de cada una de las 25

$$S_x^2 = \frac{\sum (x_i - \overline{X})^2}{n}$$

muestra $\boxed{(11, 2)}$

media $\boxed{(6'5)}$

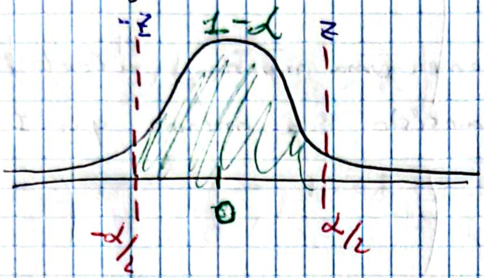
$$S_x^2 = \frac{(11-6'5)^2 + (2-6'5)^2}{2} = 20'25$$

$$\mu_S^2 = \frac{\text{suma de todas las varianzas}}{25} = \boxed{5'4}$$

<Estimación por intervalos de confianza>

- Estimación mediante intervalos de confianza: Determinar un posible rango de valores o intervalo, en el que el valor de un parámetro de la población se encuentra dentro de unos límites.

- Nivel de confianza: "Probabilidad" de que el intervalo calculado contenga al verdadero valor del parámetro.



$P(Z \geq Z_{\alpha/2}) = \alpha/2$

$P(Z \leq -Z_{\alpha/2}) = \alpha/2 \Rightarrow P(Z \leq Z_{\alpha/2}) = 1-\alpha/2$

$N(\mu, \sigma)$ a un nivel confianza $(1-\alpha)$:

$$\left(\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

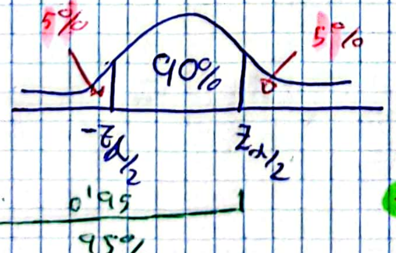
Ej: El tiempo diario que jovenes estudiantes dedican a actividades deportivas, se puede aproximar por una variable aleatoria con distribución normal de $\sigma = 20$ min.

a) Para una muestra aleatoria simple de 250 estudiantes se ha obtenido un tiempo medio de 90 min.
Calcula el intervalo de confianza al 90%

$N(\mu, 20)$

$N = 250$ Preguntan I.C al 90% $\Rightarrow \alpha = 10\%$

$\bar{X} = 90$



- Busco Z que me deje a la izq. el 95%

- Busco en tabla normal -

$Z_{\alpha/2} = 1'645$

Sabiendo $Z_{\alpha/2} = 1'645$, calculo el error $\quad Error = Z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$

$$Error = 1'645 \frac{20}{\sqrt{250}} = \underline{2'08}$$

I.C.: $\quad IC = (\bar{X} - Error, \bar{X} + Error)$

$$IC = (90 - 2'08, 90 + 2'08) \Rightarrow \boxed{IC_{90\%} = (87'92, 92'08)}$$

b) ¿Que tamaño mínimo debe tener una muestra aleatoria simple para que el error máximo cometido sea menor que 1 min al 90%?

Error

$$\Downarrow$$

$$1'645 \cdot \frac{20}{\sqrt{N}} \leq 1 \quad \Rightarrow N \geq 1082'41$$

- Si σ es desconocida:

$$\left(\bar{X} - Z_{\alpha/2} \frac{S_n}{\sqrt{n-1}} ; \bar{X} + Z_{\alpha/2} \frac{S_n}{\sqrt{n-1}}\right)$$

ojo o bien $n \geq 30$, o la distribución de partida es normal

- Intervalo media poblacional, σ desconocida y $n \leq 30$

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \rightarrow t_{n-1}$$

$$\Downarrow$$

$$\left(\bar{X} - t_{\alpha/2 ; n-1} \frac{S}{\sqrt{n}} ; \bar{X} + t_{\alpha/2 ;} \frac{S}{\sqrt{n}}\right)$$

Ej. Se hicieron 10 mediciones sobre la resistencia de cierto tipo de alambre que dieron valores:

$\bar{X} = 10'48$ Ohms y $S = 1'36$ Ohms

¿Intervalo de confianza al 90%?

$\alpha = 1 - 0'90 = 0'10 \Rightarrow \alpha/2 = 0'05$

$t_{0'05,9} = 1'833$

$$\left(\bar{X} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}\right)$$

$$\left(10'48 - 1'833 \frac{1'36}{\sqrt{10}}, 10'48 + 1'833 \frac{1'36}{\sqrt{10}}\right)$$

$$\boxed{IC = [9'69, 11'27]}$$

I. C. para una binomial

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\,q}{n}}} \Rightarrow IC = \left(\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-p)}{n}}, \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-p)}{n}}\right)$$

Ej. Se ha obtenido una muestra de 150 vendedores, de los seleccionados, 50 no han llegado al mín establecido

- IC para los que No lo han conseguido al 80%

$1 - \alpha = 80\% \Rightarrow \alpha = 20\%$

La proporción de la muestra $= 50/150 = 0'33$

10 − 10
80
90

$\leq 90 \Rightarrow$ cuantiles están entre $-1'28$ y $1'28$

tabla

$$IC = \left(0'33 - 1'28\sqrt{\frac{0'33(1-0'33)}{150}}; 0'33 + 1'28\sqrt{\frac{0'33(1-0'33)}{150}}\right)$$

IC $[0'28, 0'38]$

* I.C. para la varianza σ^2 de una población con distribución Normal

$$IC\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)} ; \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}(n-1)}\right)$$

Ej. Muestra de 15 vendedores, la media y varianza (miles €) son 5 y 2.

¿I.C. al 90%?

$\alpha/2 = 0'05 \Rightarrow 1-\alpha/2 = 0'95$; 14 grados de libertad

$\underbrace{\chi^2_{\alpha/2}(n-1) = 6'57}_{\text{tabla}}$ $\underbrace{\chi^2_{1-\alpha/2}(n-1) = 23'68}_{\text{tabla}}$

Chi-cuadrado

$$IC = \left(\frac{14 \cdot 2'14}{23'68} \le 2 \le \frac{14 \cdot 2'14}{6'57}\right) \Rightarrow [1'27 ; 4'57]$$

como nos dan la varianza muestral

$$S^2 \left(\frac{n}{n-1}\right) s_n^2 = \frac{15}{14} * 2 = 2'14$$