

Probabilidad y Estadística

<https://github.com/AdriCri22/Probabilidad-Estadistica-PE-FIB>

Bloque 1 – Cálculo de probabilidades

Conjuntos/Diagramas de Venn

Árboles

$P(A)$ → Significa probabilidad de que pase A

Propiedades:

- $0 \leq P(A) \leq 1$ → Si una probabilidad supera estos valores has hecho algo mal.
- $P(A_0 \cup A_1 \cup \dots \cup A_n) = P(A_0) + P(A_1) + \dots + P(A_n)$ si $A_i \cap A_j = \emptyset$ per $i \neq j$
- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ → Unión
- $P(A \cap B) = P(A) \cdot P(B)$ → Intersección
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$ → Probabilidad condicionada $P(\text{dato incierto}|\text{dato conocido})$

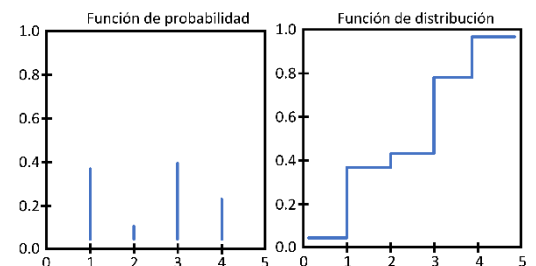
Independencia de VA → Hay 2 maneras de comprobarlo:

- Si $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$, es decir si $P(A \cap B) = P(A) \cdot P(B)$
- Si $P(B|A) = P(B|\neg A) = P(B)$, es decir si todas las filas/columnas de la tabla de probabilidades condicionadas coinciden para A y $\neg A$.

Bloque 2 – Variable Aleatoria

Variable Aleatoria (VA) → Variable que recoge todos los posibles sucesos

- Discreta (VAD) → Cuando las variables aleatorias son números naturales
 - Función de probabilidad (p_x) → Se define a partir de cada uno de los posibles valores de k, son probabilidades puntuales.
 - Función de distribución (F_x) → Se define la probabilidad acumulada, es ir sumando las probabilidades puntuales $F_{X(k)} = P(X \leq k) = \sum_{j \leq k} p_x(j)$.
- Continua (VAC) → Cuando las variables aleatorias son decimales
 - Función de densidad (f_x) → de una VAC es la función que recubre el área dónde está definida la variable cumpliendo: $\int_{-\infty}^{+\infty} f_X(x)dx = 1$
 - Función de distribución (F_X) → de una VAC define la probabilidad acumulada: $F_X(k) = P(X \leq k) = \int_{-\infty}^k f_X(x)dx$



Nota: que las probabilidades resulten en decimales es independiente a las VA.

Un **cuantil** (α) → $0 \leq \alpha \leq 1$ → X_α (cuantil α de X) si cumple $F_X(X_\alpha) = \alpha$ → es el problema inverso al cálculo de probabilidades acumuladas.

En VAC cuando pregunta por el mínimo → $F(x) = 1 - \alpha$ → Porque nos preguntan por el área superior
por el máximo → $F(x) = \alpha$

Esperanza (media)

$$\text{VAD} \rightarrow E(X) = \sum(k \cdot p(k))$$

$$E(X^2) = \sum(k^2 \cdot p(k))$$

$$\text{VAC} \rightarrow E(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 \cdot f_X(x) dx$$

Variación es el error al cuadrado (para poder representar el error por encima y debajo de esta media) de la predicción sobre una media de todas las posibilidades.

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

$$\text{VAD} \rightarrow V(X) = \sum[(k - E(X))^2 \cdot p(k)]$$

$$\text{VAC} \rightarrow V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 \cdot f_X(x) dx$$

Desviación tipo/estándar indica el error medio de predicción $\rightarrow \sigma = \sqrt{V(X)}$

Pares de Variables

Probabilidad conjunta $\rightarrow P_{X,Y}(x, y) = P(X \cap Y)$

Probabilidad condicionada $\rightarrow P_{X,Y}(X) = P_{X,Y}(x)/P_Y(y)$

$$V(X \pm Y) = V(X) + V(Y) \pm 2 \cdot \text{Cov}(x, y)$$

Función de probabilidad conjunta

	X_1	...	X_n
Y_1			
...			
Y_n			

Covarianza \rightarrow Indica el grado de dependencia hay entre variables

$$\rightarrow \text{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

$E(X \cdot Y) \rightarrow$ Se hace la tabla de probabilidades de las 2 variables aleatorias y

$$\sum X_i \cdot Y_i \cdot P(X_i \cap Y_i)$$

$$\text{VAD} \rightarrow \text{Cov}(X, Y) = \sum_{\forall x} \sum_{\forall y} (x - E(X)) \cdot (y - E(Y)) \cdot p_{XY}(x, y)$$

Correlación $\rightarrow \text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}, \quad -1 \leq \text{corr}(X, Y) \leq 1$

- Cuanto más cerca del 1 la relación es mayor
- Cuanto más cerca del -1 la relación es inversa
- Cuando es 0 no existe relación entre ambas variables

Integrales

$$\int 0 dx = C$$

$$\int k dx = kx + C$$

$$\int x^n dx = \frac{x^{n+1}}{n+1} + C, \quad n \neq -1$$

$$\int \frac{1}{x} dx = \ln x + C$$

$$\int e^x dx = e^x + C$$

Integral definida

$$\int_a^b f(x) dx = [F(x)]_a^b = F(b) - F(a)$$

Bloque 3 – Modelos de variable aleatoria

Modelos VAD:

- Modelo Bernoulli $X \sim \text{Bern}(p) \rightarrow$ Sirve cuando se hace 1 solo experimento para 2 posibles resultados: 0 (“no éxito”) y 1 (“éxito”). **(Valores enteros + 2 resultados 0 o 1)**

Función de probabilidad:
$$\begin{aligned} \text{Si } k = 0 &\rightarrow P(k) = q = 1 - p \\ \text{Si } k = 1 &\rightarrow P(k) = p \end{aligned}$$

Función de distribución:
$$F_x(k) = \sum_{i \leq k} P_x(i)$$

$$\begin{aligned} p &= \text{Probabilidad de 1 (“éxito”)} \\ q &= 1 - p = \text{Probabilidad de 0 (“no éxito”)} \end{aligned}$$

$$E(X) = p$$

$$V(x) = p \cdot q$$

- Modelo Binomial $X \sim B(n, p) \rightarrow$ Número de éxitos en la repetición de n pruebas Bernoulli independientes con probabilidad constante p . **(Valores enteros + prob.)**

Función de probabilidad:
$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

$$\begin{aligned} p &= \text{Probabilidad de 1 (“éxito”)} \\ q &= 1 - p = \text{Probabilidad de 0 (“no éxito”)} \end{aligned}$$

Función de distribución:
$$F_x(k) = \sum_{i \leq k} P_x(i) \quad (\text{tablas})$$

$$E(x) = n \cdot p$$

$$n = \text{Número de pruebas}$$

$$V(X) = n \cdot p \cdot q$$

Nota: $\binom{n}{k} = \frac{n!}{k!(n-k)!} \rightarrow$ Calculadora: n (SHIFT) nCr k

- Modelo Geométrico $X \sim \text{Geom}(p) \rightarrow$ Número de intentos de un experimento de Bernoulli hasta observar el primer “éxito”.

Función de probabilidad:
$$P(X = k) = p \cdot q^{k-1}$$

$$\begin{aligned} p &= \text{Probabilidad de 1 (“éxito”)} \\ q &= 1 - p = \text{Probabilidad de 0 (“no éxito”)} \end{aligned}$$

Función de distribución:
$$F_x(k) = 1 - q^k$$

$$E(x) = 1/p$$

$$k = \text{Número de intentos}$$

$$V(X) = q/p^2$$

- Modelo Binomial negativo $X \sim BN(r, p) \rightarrow$ Número de repeticiones de un experimento de Bernoulli hasta observar cierto número de “éxitos”.

Función de probabilidad:
$$P(X = k) = \binom{k-1}{r-1} \cdot p^r \cdot q^{k-r}$$

$$\begin{aligned} p &= \text{Probabilidad de 1 (“éxito”)} \\ q &= 1 - p = \text{Probabilidad de 0 (“no éxito”)} \end{aligned}$$

Función de distribución:
$$F_x(k) = \sum_{i \leq k} P_x(i)$$

$$k = \text{Número de repeticiones}$$

$$E(x) = 1/p$$

$$r = \text{Número de 1 (“éxitos”)}$$

$$V(X) = q/p^2$$

- Modelo Poisson $X \sim P(\lambda) \rightarrow$ Número de ocurrencias favorables en un determinado intervalo o espacio. **(Valores enteros + tiempo)**

Función de probabilidad:
$$P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

$$k = \text{Número de repeticiones}$$

Función de distribución: $F_x(k) = \sum_{i \leq k} P_x(i)$ (tablas) $\lambda =$ Tasa media de ocurrencias por unidad considerada

$$E(x) = \lambda$$

$$V(X) = \lambda$$

Para calcular el **cuantil** $P(X \leq x_\alpha) = \alpha$

Se calcula la probabilidad acumulada de α por arriba y por abajo, es decir para las dos VA más próximas a α .

1. Si nos piden que el valor sea por debajo, pero con un mínimo \rightarrow VA de arriba de α
2. Si nos piden que el valor sea por encima, pero con un mínimo \rightarrow VA de debajo de α
3. Si nos piden que el valor sea por debajo, pero máximo \rightarrow VA de abajo de α
4. Si nos piden que el valor sea por arriba, pero máximo \rightarrow VA de arriba de α

¿Por qué? Cuantil de un mínimo $\rightarrow F(x) = 1 - \alpha$, cuantil de un máximo $F(x) = \alpha$

Arriba/Abajo dependiendo de si queremos que superar o no el α .

Modelos VAC:

- Modelo Exponencial $X \sim \text{Exp}(\lambda) \rightarrow$ Distribución del tiempo entre llegadas (ocurrencias) en un proceso de Poisson.

Función de densidad: $f_x(x) = \lambda \cdot e^{-\lambda x}$

Probabilidad de pasar x tiempo sin que pase algo

Función de distribución: $F_x(x) = 1 - e^{-\lambda x}$

$$E(x) = 1/\lambda$$

$\lambda =$ Tasa media de ocurrencias por unidad considerada

$$V(X) = 1/\lambda^2$$

- Modelo Uniforme $X \sim U(a, b) \rightarrow$ Probabilidad de pertenecer a un intervalo concreto en un rango, depende de la longitud del intervalo. (**Rangos**)

Función de densidad: $f_x(x) = \frac{1}{b-a}$ para $a < x < b$

Función de distribución: $F_x(x) = \frac{x-a}{b-a}$ para $a < x < b$ $a =$ Valor mínimo del rango de X
 $b =$ Valor máximo del rango de X

$$E(x) = (b+a)/2$$

$$V(X) = (b-a)^2/12$$

- Modelo Normal $X \sim N(\mu, \sigma) \rightarrow$

Función de densidad: $f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$\mu =$ Esperanza

Función de distribución: ?

$\sigma =$ desviación estándar

$$E(x) = \mu$$

$$V(X) = \sigma^2$$

- Estandarizar (preguntan probabilidad dándonos los parámetros) $P\left(Z < \frac{k-\mu}{\sigma}\right) = \%$

$\rightarrow \frac{k-\mu}{\sigma} = \text{prob. tabla}$ (si $\frac{k-\mu}{\sigma}$ es negativo $\rightarrow \frac{k-\mu}{\sigma} = 1 - \text{prob. tabla}$).

- Cuando te dan un % y 2 de tres variables (k, μ, σ) y te piden que averigües la restante

$$P\left(Z < \frac{k - \mu}{\sigma}\right) = \% \text{ que te dicen} \rightarrow \frac{k - \mu}{\sigma} = \text{prob que corresponde en la tabla}$$

TCL (Teorema Central de Límite):

- Distribución de la variable suma $S_n = \sum X_{i_{n \text{ gran}}} \rightarrow N(n\mu, \sigma\sqrt{n}) \rightarrow \frac{S_n - n\mu}{\sigma\sqrt{n}} \rightarrow N(0,1)$
- Mitjana $\bar{X}_n = \frac{\sum X_i}{n} \rightarrow N(\mu, \sigma\sqrt{n}) \rightarrow \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \rightarrow N(0,1)$
- Se puede pasar de distribución binomial a normal si $n > 30 \rightarrow \mu = E(X)$ y $\sigma = \sqrt{V(X)}$
Se hace cuando k es muy grande
- Se puede pasar de distribución de Poisson a normal $\rightarrow \mu = \lambda$ y $\sigma = \sqrt{\lambda}$
Se hace cuando k es muy grande
- Cuando piden un intervalo con probabilidad de $x\%$ \rightarrow calcular $c = x + (1 - x)/2$
Intervalo máximo $\rightarrow P(Z < M) = P\left(Z < \frac{M - \mu}{\sigma}\right) = c \rightarrow M = \mu + \sigma \cdot c$
Intervalo mínimo $\rightarrow m = \mu - \sigma \cdot c$

Observaciones de VAC:

- Por definición $P(X = x) = 0$, por lo que $f_x(x)$ no es una probabilidad
- $P(x < a) = P(x \leq a)$ $a \in \mathbb{N}$
- Las funciones de distribución calculan $P(x < a)$ para calcular $P(x > a) = 1 - P(x < a)$
- $P(x < -a) = 1 - P(x < a)$
- $P(x > -a) = P(x < a)$

Nota: cuando nos dan los datos con ciertas unidades y nos dicen calcular con otras pasar los parámetros.

Tablas

distribución **NORMAL ESTANDARIZADA**

z	0.00	0.01	0.02	0.03
0.0	0.5000	0.5040	0.5080	0.5120
0.1	0.5398	0.5438	0.5478	0.5517
0.2	0.5793	0.5832	0.5871	0.5910
0.3	0.6179	0.6217	0.6255	0.6293
0.4	0.6554	0.6591	0.6628	0.6664

$$P(Z < 0.21) = 0.5832$$

$$P(Z \leq ?) = 0.59 \rightarrow X = 0.22$$

Función de distribución **BINOMIAL**

N	X	0,05	0,1	0,15
2	0	0,9025	0,8100	0,7225
	1	0,9975	0,9900	0,9775
3	0	0,8574	0,7290	0,6141
	1	0,9928	0,9720	0,9393
	2	0,9999	0,9990	0,9966

$$X \sim B(n = 3, p = 0.15)$$

$$P(X \leq 2) = 0.9966$$

$$P(X = 2) = P(X \leq 2) - P(X < 1) = 0.9966 - 0.9393 = 0.0573$$

$$P(X > 2) = 1 - P(X \leq 2) = 1 - 0.9966 = 0.0034$$

Si hay que redondear se redondea al más próximo

Bloque 4 – Inferencia estadística

Conceptos básicos:

- **Parámetro:** parte de la **población** que queremos hacer la estimación.
- **Muestra:** es una fracción del parámetro que se usa para obtener datos, ya que analizar a toda la parte de la población que queremos analizar se haría muy difícil.
- **Estadístico:** Cualquier indicador que se obtenga a partir de los datos de la muestra.
- **Estimador:** estadístico de una muestra que se usa para obtener el valor de un parámetro de la población.

Estimación puntual de μ es la media muestral $\rightarrow \bar{x} = \frac{\sum x_i}{n}$

Error tipo o error estándar es la variabilidad (valores muy dispersos) del estimador $\rightarrow se = \sigma/\sqrt{n}$

Cuando la σ es desconocida $\rightarrow \widehat{se} = 1/\sqrt{n}$

Parámetro (θ) (población)	Estimador ($\bar{\theta}$) (Muestra)
μ (esperanza, media poblacional)	\bar{x} (media muestral)
σ^2 (variancia poblacional)	s^2 (variancia muestral)
σ (desviación tipo poblacional)	s (desviación tipo muestral)
π (probabilidad)	p (proporción)

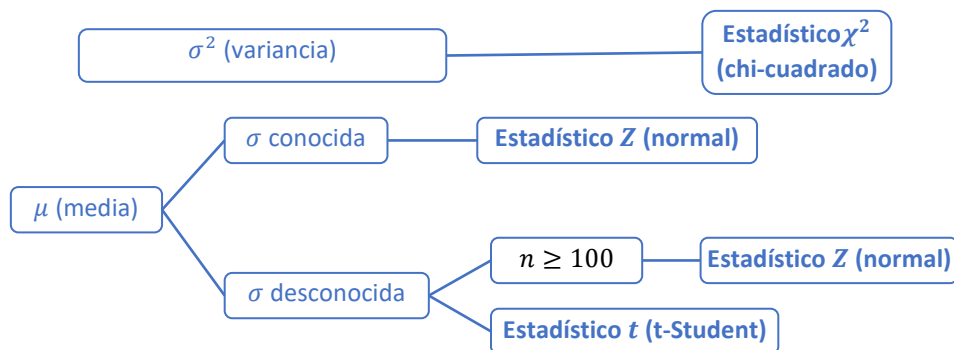
Propiedades de los estimadores:

- No tener sesgo (biaix en catalán) \rightarrow Cuando la diferencia entre la esperanza y la estima es nula.
- Ser eficiente \rightarrow Cuando no tiene sesgo y la variancia es menor.

Intervalos de confianza (estimar un parámetro)

Mecánica

- 1- Definir el estadístico usado
- 2- Especificar distribución



- 3- Indicar las premisas necesarias para decir que sigue la distribución
- 4- Delimitar el nivel de confianza (usualmente $1 - \alpha = 95\% \rightarrow \alpha = 5\%$)
Dónde $\alpha = 1 - \text{porcentaje que nos dicen}$
- 5- Calcular el intervalo \rightarrow Usando la distribución especificada
- 6- Interpretar el resultado \rightarrow El tanto % de las veces VA estará en el intervalo dado

El **error tipo** es igual al denominador del estadístico.

- **Error tipo de la media:** es la desviación “habitual” de la media muestral \bar{x} respecto a la media de la población μ . Se calcula S^2/\sqrt{n} .

- **Error tipo de la proporción:** es la desviación “habitual” de la proporción de la muestra p respecto a la proporción real π . Se calcula $\sqrt{\frac{\pi \cdot (\pi - 1)}{n}}$

Nota: $\pi = p$, pero si no hay diferencia $\rightarrow \pi = p = 0,5$.

Pruebas de Hipótesis (Refutar un parámetro)

- 1- Escoger una variable según los objetivos del estudio (La variable que queremos demostrar)
- 2- Escoger un diseño y un estadístico

π (probabilidad)

Estadístico Z Binomial

- 3- Definir la hipótesis nula $H_0: \mu = \text{media}$ o $\pi = \text{proporción}$ y la hipótesis alternativa H_1
 - $H_1: \mu \neq \text{media}$ o $\pi \neq \%$ bilateral
 - $H_1: \mu \neq \text{media}$ o $\pi < \%$ unilateral
- 4- Especificar la distribución del estadístico si H_0 fuera cierto (y sus premisas)
- 5- Contrastar H_0 Dos alternativas para hacerlo
 - a. Si $|z| > z_{1-\alpha}$ (unilateral) o $|z| > z_{1-\alpha/2}$ (bilateral) $\rightarrow H_0$ Se rechaza (es poco fiable)
 - b. Calcular el valor P \rightarrow Si $P < \alpha \rightarrow H_0$ Se rechaza (es poco fiable)
- 6- Añadir la estimación para el intervalo $IC(1 - \alpha)$

Intervalos de confianza

`binom.test()` # Binomial

Estadístico t-Student

`t.test(datos1, datos2, var.equal=TRUE, conf.level=prob)$conf.int` # Independientes

`t.test(Diferencia*2, var.equal=TRUE, conf.level=prob)$conf.int` # Apareadas

`var.test()` # Fisher

`prop.test()` # Proporciones

`chisq.test()` # Proporciones Pearson

`mean()` # Calcula la media

`sd()` # Calcula la desviación

`var()` # Calcula la variancia

`pt, pf, pnorm` # Para calcular el p-valor

`qt, qf, qnorm` # Para calcular el punto crítico

Bloque 5 – Diseño de experimentos

Tipos de recogida de datos:

- Muestras **independientes**: cada caso se mide de manera independiente (los datos pertenecen a una variable o a otra, no a ambas).
- Muestras **apareadas**: cada caso da lugar a dos medidas, pares de medidas (los datos pueden pertenecer a más de 1 variable a la vez).

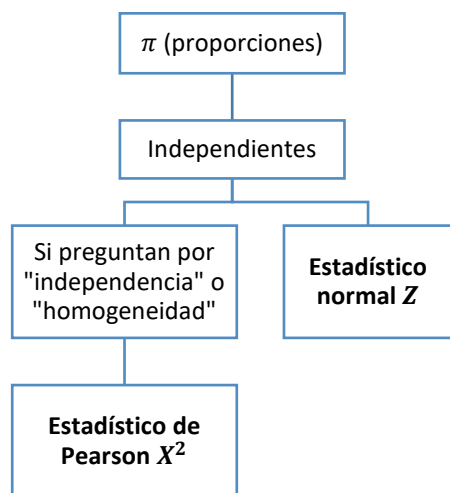
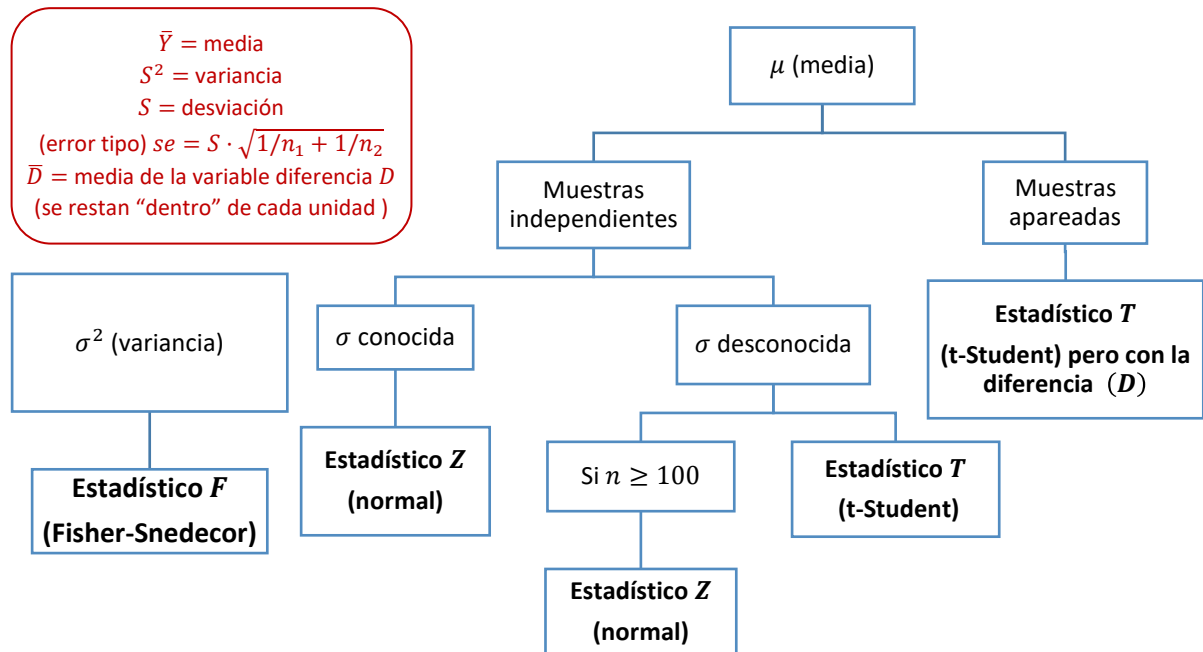
Guía:

1. Definir las **variables**.
2. Decidir el **estadístico**.
3. Definir la **hipótesis**:
 - i. La hipótesis nula **siempre es la misma**: $H_0: \mu_A = \mu_B$ (μ o σ^2)
 - ii. La hipótesis alternativa depende de si queremos que sea
 - i. Bilateral (ambos extremos): $H_1: \mu_A \neq \mu_B$ (μ o σ^2)
 - ii. Unilateral (un extremo): $H_1: \mu_A < \mu_B$ o $H_0: \mu_A > \mu_B$ (μ o σ^2)
4. **Distribución del estadístico** que sigue H_0 y sus premisas.
5. **Calcular** el estadístico
6. Calcular el **P-valor** (tener en cuenta la hipótesis alternativa) o el **punto crítico**:
 - i. Si es bilateral \rightarrow calcular el p-valor de uno de los extremos y multiplicar por 2.
 - ii. Si es unilateral \rightarrow calcular el p-valor de un extremo.

Si es el **punto crítico**:

- $|t| > t_{1-\alpha}$ (unilateral) \rightarrow Rechazar H_0 .
- $|t| > t_{1-\alpha/2}$ (bilateral) \rightarrow Rechazar H_0 .

7. **Conclusión:** mirar que criterio de decisión tiene el estadístico para rechazar la H_0 .



$$p = \frac{\text{número de aciertos}}{\text{número total}}$$

$$f_{ij} = \text{frecuencia de casos observados en la fila "i", columna "j"}$$

$$e_{ij} = \frac{e_i \cdot e_j}{e_N} = \frac{(\text{total de la fila i}) \cdot (\text{total de la columna j})}{\text{total de los totales}}$$

f_{ij}	C	D	Total
A	f_{11}	f_{12}	Total fila 1
B	f_{21}	f_{22}	Total fila 2
	Total col. 1	Total col. 2	Total de los totales

e_{ij}	C	D	Total
A	e_{11}	e_{12}	Total fila 1
B	e_{21}	e_{22}	Total fila 2
	Total col. 1	Total col. 2	Total de los totales

$(f_{ij} - e_{ij})^2 / e_{ij}$	C	D	Total
A	X_{11}	X_{12}	
B	X_{21}	X_{22}	
			$\sum X^2$

Bloque 6 – Previsión

Estudios observacionales: se trata de estudios dónde vemos lo que sucede y sirven para predecir, anticipar, prever...

Estudios experimentales: se trata de estudios en los que podemos interactuar, por lo que podemos intervenir y cambiar el futuro.

Modelo: $Y_i = \beta_0 + \beta_1 \cdot X_i + \varepsilon_i$ Parte determinista de Y
Parte aleatoria de Y

Parámetros

β_0 = Constante al origen

Y_i = Valor de la variable respuesta Y en el caso i -ésimo

β_1 = Pendiente de la recta

X_i = Valor que toma la condición X en el caso i -ésimo

ε_i = Error aleatorio o distancia de la recta del caso i -ésimo / error de predicción

$S^2 = \sigma^2$ = Variancia residual o variancia de los $\varepsilon_i \rightarrow$ cuanto mayor sea σ los valores estarán más dispersos y por lo tanto mayor variabilidad habrá.

S = Desviación típica del término aleatorio del modelo

$\hat{\beta}_0 = b_0$ = Estimación del término independiente

$\hat{\beta}_1 = b_1$ = Estimación término lineal / pendiente estimada

} Son los estimadores de β_0 y β_1

\bar{Y} = Media de la variable respuesta

S_{XY} = Covariancia muestral

$r = r_{XY}$ = Correlación muestral / coeficiente de Pearson

$r_{XY}^2 = R^2$ = coeficiente de determinación

$0 \leq R^2 \leq 1 \rightarrow \begin{cases} \text{más cerca del 1} \rightarrow \text{más capacidad predictiva y poca variabilidad} \\ \text{más cerca del 0} \rightarrow \text{menos capacidad predictiva y mucha variabilidad} \end{cases}$

S_Y = Desviación tipo de la variable respuesta / error de estimación del término independiente

S_X = Desviación tipo de la condición / error de estimación del término lineal

Interpretación de los parámetros

Los **parámetros** de la recta han de ser interpretados de acuerdo con sus unidades

La **pendiente** se interpreta:

- Experimentos: La respuesta Y tendrá un cambio esperado de β_1 (unidades de Y) por cada incremento de 1 unidad de la causa X .
- Previsión: una variación de 1 unidad en la variable X se asocia a una variación de β_1 unidades en la variable Y .

La **variancia residual** se interpreta:

- Experimentos: variabilidad de la variable Y .
- Previsión: error de predicción de la variable Y , conociendo el valor de X . (son las fluctuaciones de nuestra previsión, es decir cuanto por encima y por debajo de nuestra valor nos podríamos equivocar).

La **constante** se puede interpretar como el valor que toma la respuesta en ausencia de la variable predictora.

Si el error estándar es demasiado grande \rightarrow Para **disminuir** S_{b1} hay que aumentar n (número de observaciones).

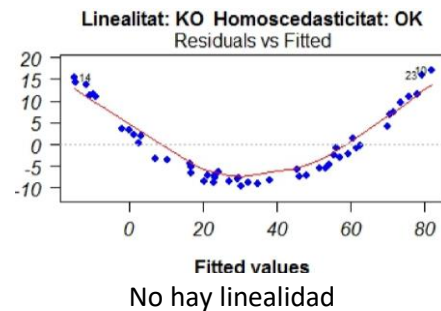
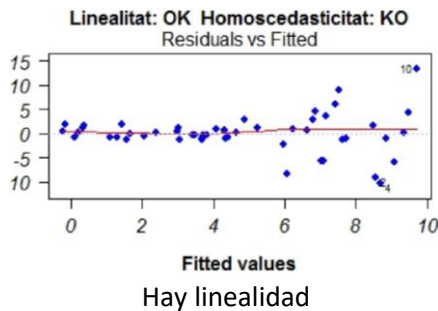
Inferencia se puede hacer la prueba de hipótesis para β_0 , β_1 y σ^2 (seguir el procedimiento habitual, tomando la distribución t-Student como estadístico).

La inferencia con intervalo de confianza: $IC(\mu_1 - \mu_2, 95\%) = (\bar{y}_1 - \bar{y}_2) \pm t_{n_1+n_2-2, 1-\alpha/2} \cdot \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$, el resultado representaría que por ejemplo la muestra 2 tarde de media entre X e Y segundos menos con un intervalo de confianza del 95%.

Para **validar el modelo lineal** hay que mirar las premisas:

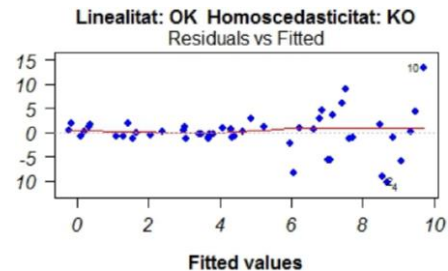
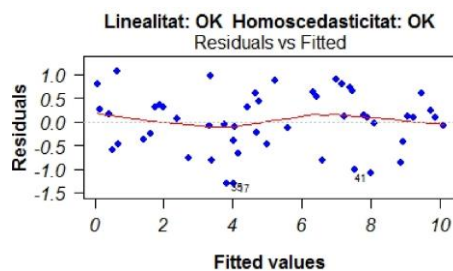
- En la parte determinista:

- **Linealidad:** en el rango que se nos da que se mantenga igual sobre el eje de la y

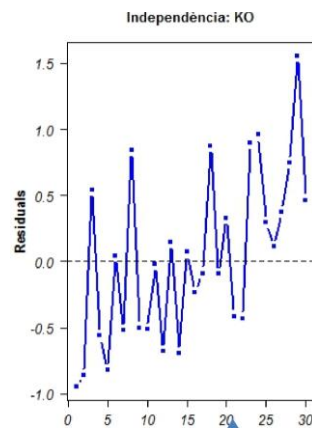
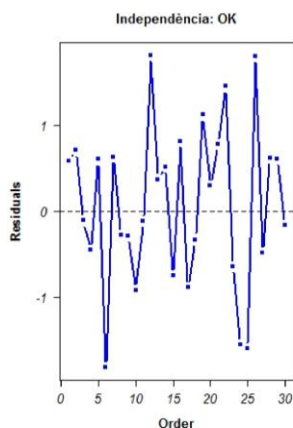


- En la parte aleatoria:

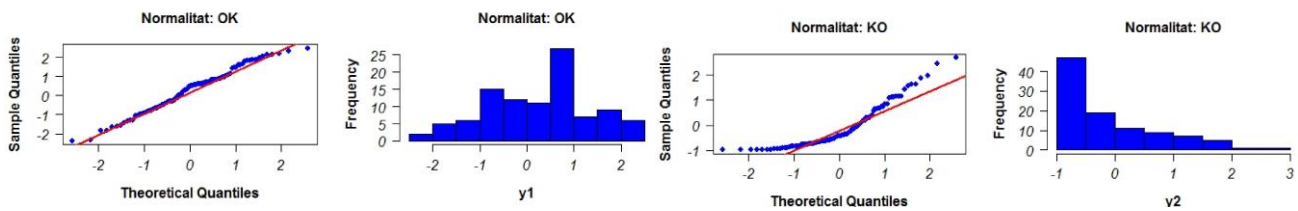
- **Homoscedasticidad:** misma σ^2 para cualquier caso/dato, cuando miramos el gráfico vemos que todos los datos tienen el mismo error.



- **Independencia:** Detecta si existe o no dependencia entre los datos.



- **Normalidad:** en el caso de un gráfico que se mantengan los datos sobre la recta qqnorm y en el caso de un histograma que forme una campana.



```
datos <- read.table("clipboard",header=TRUE) # Leemos los valores copiados
# La función lm se usa para ajustar un modelo lineal, sea un modelo de regresión lineal,
# de análisis de varianza o de análisis de covarianza
mod.lm <- lm(var_respuesta ~ condicion, datos)
# Obtenemos una salida más detallada e informativa
summary(mod.lm)
# Permite modificar distintos parámetros de la ventana gráfica
par(cex.lab = 1.2, cex.axis = 1.2, las = 1, font.lab = 2, font.axis = 3)
# Dibuja el gráfico
plot(var_respuesta ~ condicion, datos, pch = 19, col = 4, cex = 1.2)
# Permite sobreponer una recta de regresión a un gráfico de dispersión
abline(mod.lm, col = 2, lwd = 3)

## Ejemplo para validación lineal
par(mfrow=c(2,2))
plot(lm(var_respuesta ~ condicion),c(2,1)) # QQ-Norm i Standard Residuals vs. Fitted
hist(rstandard(lm(var_respuesta ~ condicion))) # Histograma dels residuos estandarizados
plot(1:10,rstandard(lm(var_respuesta ~ condicion)),type="l") # Ordren de los residuos
```

`summary(mod.lm)` - Ejemplo e interpretación de los datos:

Call:

`lm(formula = Preu ~ Capacitat, data = datos)`

Residuals:

Min	1Q	Median	3Q	Max
-102.32	-20.89	12.48	36.99	89.21

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	$b_0 = 386.3889$	$S_{b_0} = 69.6313$	$t_{b_0} = 5.549$	$p - valor_{b_0} = 4.40e-05$ ***
Capacitat	$b_1 = 2.4133$	$S_{b_1} = 0.4097$	$t_{b_1} = 5.891$	$p - valor_{b_1} = 2.28e-05$ ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: $S = 57.94$ on *grados de libertad* = 16 degrees of freedom

Multiple R-squared: $R^2 = 0.6844$, Adjusted R-squared: 0.6647

F-statistic: 34.7 on 1 and 16 DF, p-value: 2.278e-05

$p - valor_{b_x} < 0.05 \rightarrow$ Indica que es un valor significativo, que es importante para nuestra muestra, cuanto más cercano a 0 mejor.

$$t_{b_0} = b_0 / S_{b_0}$$

$$t_{b_1} = b_1 / S_{b_1}$$