ESTADÍSTICA

¿Qué es?

Es relacionada con el METODO CIENTÍFICO en la colección y análisis de datos,

OBJ?

Deducir o inferir conclusiones y tomar desiciones ante cond de incertidumbre

TIPOS: (I) Inferencial
(II) Descriptiva.

(I) Busca lograr generalizaciones y toma de desiciones
∴ A partir de los datos de la MUESTRA concluir sobre una POBLACION.

(II) Abarca la obtención, organización, resumen y presentación de los datos en forma convenientemente útil y de fácil comunicación, además de calcular medidas con dicha info.

DESCRIPTIVA	INFERENCIAL
OBTENCIÓN, PRESENTACIÓN Y DESCRIPCIÓN DE DATOS NUMÉRICOS.	TECNICAS PARA TOMAR DECISIONES CON BASE EN EL ANALISIS DE LOS DATOS OBTENIDOS.

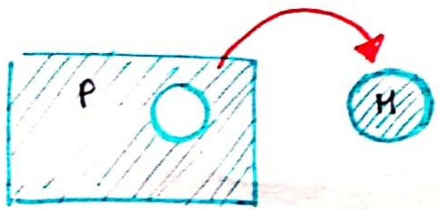
CONCEPTO: Población - Muestra

POBLACIÓN:

Conj de elementos con una o más características en común. Puede estar formado por elementos animados o no, o por las mediciones que sobre estos elementos se realicen.

MUESTRA:

Es una parte o subconj de la población previamente especificada. Sus componentes deben poseer las características comunes que los de la población de donde ella proviene.



CONCEPTO: Parametro - Estimador

PARAMETRO: Valor númerico que describe las caract. de una POBLACIÓN.

Ej) El promedio de materias aprobadas de todos los alumnos de 2do Año es de tres.
(El parametro vale 3)

ESTADÍSTICO: Valor numerico que describe una caract. de la MUESTRA.

Ej) El promedio de materias aprobadas de un grupo de 25 alumnos de 2do Año. es de tres.
(El estadístico/Estimador es 3).

Muestra ← MUESTREO → Clasificación:

Muestra

Es un subconjunto de datos pertenecientes a una población.

→ REQUISITOS: Para ser Representativa.

CUALITATIVAMENTE

- **Homogenea:** Estar compuesta por elementos ∈ al Universo previamente definido.
- **Adecuada:** Incluye todas las variables esenciales de los elementos que ∃ en el universo.
- **No Viciada:** Presenta variaciones de determinadas variables en más o menos la misma frecuencia que en el universo.

CUANTITATIVAM.

Debe ser lo suficientemente grande!

QUÉ ES UN EXPERIMENTO?

Es una acción mediante la cuál se obtiene un resultado y que implica la observación de este.

→ **Aleatorio:** El resultado no se puede predecir con exactitud.

→ **Determinista:** El resultado se puede predecir con exactitud.

MUESTREO

DEF: Es la técnica empleada para recoger una muestra a partir de una población o un subgrupo de la población.

OBJ:
- Estimar parametros
- probar hipotesis
- Inferir sobre la población

según la INFO de la MUESTRA

Clasificación:

- Por la forma de considerar un Evento:
  - SIN reposición
  - con reposición
- Por la forma de tomar la muestra:
  - OPINATICO (Influye la opinion propia)
  - SIN NORMA (Sin usar un metodo)
  - ALEATORIO (Misma posibilidad de elegir los elementos)
- Por el nº de Muestras:
  - Simple (1M)
  - Múltiple (+ 1M)

Representación de datos cuantitativos

- V. A. discreta
  - Diagrama de Bastones — Ej) Frecuencia abs ó relativas.
  - Diagrama de Escalones — Ej) Frecuencia abs acumulada.
- V. A. Continua
  - Histograma — Conjunto de rectáng con bases en el eje x, centr en las marcas de clase, long iguales a los tamaños de los int y alturas iguales a los frecuencia de cada int.
  - Polígono - Ojiva — línea de trazos que se obtiene uniendo las frecuencias correspondientes a los ptos medios de los int de clase del histograma.

Representación de datos cualitativos

- Gráf. de Barras → En el eje horiz se representan las categorías de la variable y sobre el eje vertical las frecuencias respectivas de cada categoría.

  Barras comparativas

  Barras apiladas

- Diagrama de Torta. (Pie Chart) → Consta de un círculo que se divide en tantos sectores circulares como categorías tenga la variable, cuyas áreas son proporcionales a las frecuencias correspondientes.

MEDIDAS DE RESUMEN

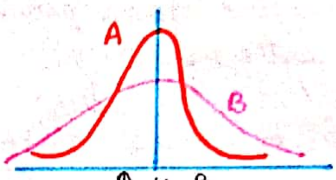
Sintetizan con un valor características de la distribución de la variable en el conj de datos observados.
Su OBJ es ser REPRESENTATIVAS

• POSICIÓN

Identifican la ubicación de la distribución de la variable a lo largo de una escala. Caracterizan al conjunto de valores de la variable mediante un valor gralmente de tendencia central.

• DISPERSIÓN

Aportan info sobre el grado de de variabilidad de una variable. a partir de un valor numerico.



A y B
Presentan misma altura "media".
B se encuentra más disperso.

C y D
Presentan igual dispersión.
Se ubican de distintas maneras.

MEDIDA DE TENDENCIA CENTRAL

Son valores que intentan representar al conj de datos y tienden a ubicarse hacia el centro de la distribución.

- MODO:
Es el valor del carácter que se representa con mayor frecuencia en la muestra o población. Presenta el mayor nº de obs.

• V.a. Discreta: Valor con mayor frecuencia

Ej)

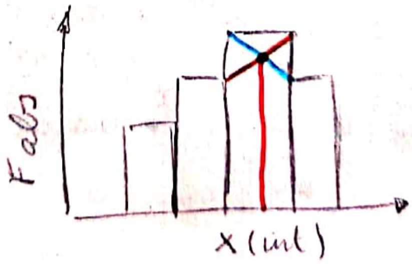
X	F
0	2
1	4
2	3

Modo = 1

• V.a. continua: Se busca el int modal (aquel con mayor frecuencia)

FORMULA → $Mo = Li + \left[\frac{F_{Mo} - F_{Mo-1}}{(F_{Mo} - F_{Mo-1}) + (F_{Mo} - F_{Mo+1})} \right] \cdot C$ ← Ampl del Int

Gráficamente:
- Trazar AB
- Trazar CD
- Trazar ⊥ a la inserción de AB y CD.
- • MODO



Escaneado con CamScanner

- **MEDIANA**: Es el valor central de la distribución, es decir, el valor dentro del rango de la variable / el 50% obs es inferior o igual a dicho valor y el 50% de las obs es superior o igual al mismo.

↳ Esc ord
↳ Datos asimétricos

↳ Depende de la posición que ocupan los datos en la ordenación y no del valor numérico de los mismos.

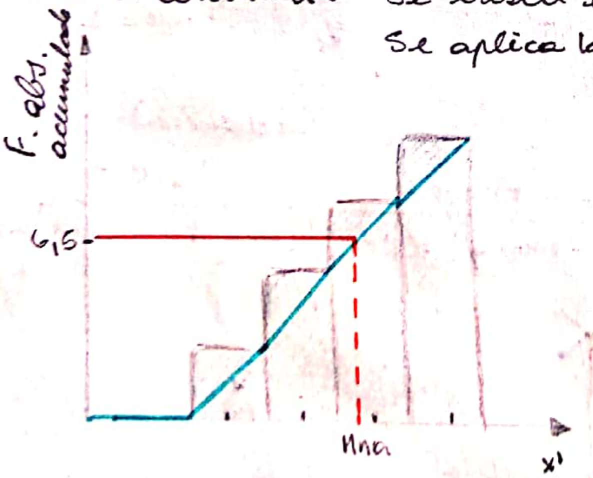
VENTAJAS	DESVENTAJAS
• Fácil calcular / interpretación • Es única (A partir de la convención) • No está influenciada por los valores extremos.	• No tiene en cuenta los valores de la variable, sino solo su posición.

- Para V. discreta: se ordenan los datos.

POCOS DATOS IMPAR (Valor central)	POCOS DATOS PAR (Promedio de valores centrales)	MUCHOS VALORES
$Mna = \frac{n+1}{2}$	$Mna = \frac{n/2 + (n/2+1)}{2}$	$Mna = \frac{n+1}{2}$
El resultado indica la posición del valor de la Mna.	El resultado indica el valor de la Mna entre los dos valores.	En la tabla de frecuencia se busca el valor de Fi que supera al resultado obtenido. La obs de ese dato es la Mna.

- Para V. continua: Se busca el int mediana $Mna^{\circ} = \frac{n+1}{2}$

Se aplica la FORMULA: $Mna = L_i + \left[\frac{\left(\frac{n+1}{2}\right) - F_{i-1}}{(F_i - F_{i-1})}\right] \cdot c$



- Trazar ojiva
- Trazar horizontal desde Mna° hasta tocar la ojiva
- Trazar ⊥ → Mna.

Ej) $n = 12 \rightarrow Mna^{\circ} = \frac{12+1}{2} = 6,5$

- MEDIA ARITMETICA:

Es el valor que tomaría la variable si estaría repartida entre todos los ind que se forman la muestra. Es única y tiene las mismas unidades que la variable.

↳ se calcula sumando todos los valores de las obs y dividiendo por el nº de datos considerados

↳ . Sí la muestra tiene tamaño n ⟶ ESTIMADOR $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$

. Sí la población tiene tamaño N ⟶ PARAMETRO $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$

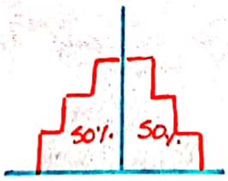
. Si se tiene datos agrupados en int ⟶ Media Aritmetica Ponderada.

$$\bar{X} = \frac{\sum^{n} x_i' . f_i}{\sum_{i=1}^{n} f_i}$$

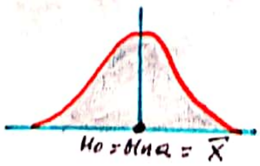
VENTAJAS	DESVENTAJAS	PROPIEDADES
. Fácil de calcular . Única c/grupo de datos . Fácil Interpretación . Usa todos los valores para su cálculo.	. Se ve muy influencia por valores extremos. Pierde representatividad . No puede calcularse en caso de int abierto	I. La suma de los desvíos entre el valor de y la media arit es cero. Datos NO Agrupados ⟶ $\sum^{n} (x_i - \bar{x}) = 0$ Datos agrupados ⟶ $\sum (x_i' - \bar{x}) f_i = 0$ II. $\sum_{i=1}^{n} (x_i - \bar{x})^2 \leq \sum_{i=1}^{n} (x_i - a)^2 \quad \forall\, a \in R$ III. $x = \frac{\bar{x}_1 . n_1 + \bar{x}_2 . n_2 + \dots}{n_1 + n_2 + \dots} = \bar{X} = \frac{\sum x_i}{n}$ IV. $\bar{X} = \frac{\sum K}{n} = \frac{nK}{K} = K$ con $x_i = K \quad \forall\, i = 1 \dots n$. V. Sí $y_i = K . x_i \longrightarrow \bar{y} = K.\bar{x}$ VI. Sí $y_i = x_i \pm K \longrightarrow \bar{y} = \bar{x} \pm K$

OBS! Asimetría

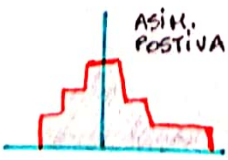
• Una distribución es simetrica si sus mitades son imagen especular.

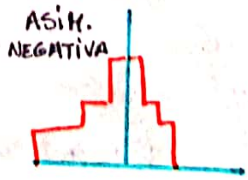


• En las dist simetricas MEDIA y MEDIANA coinciden ⟷ la MODA igual coincide en el máx central.

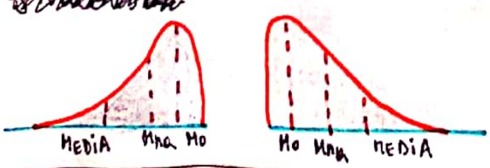


• La asimetría es positiva o negativa en función de que lado se encuentra la cola de la distribución





• La media tiende a desplazarse hacia los valores extremos (cola)



RANGO (R): Es la diferencia entre los valores máx y mín del conj de datos

$R = X_{máx} - X_{mín}$

↳ Es una medida pobre.

↳ Es sensible a valores extremos.

↳ se emplea en control de calidad con muestras pequeñas.

VARIANZA (S^2 o σ^2): Mide el grado de dispersión de los datos con referencia a la media aritmetica. Su unidad es el cuadrado de la variable. Es el promedio de los cuadrados de los desvíos de cada obs respecto a la media aritmetica:

↳ V. Poblacional: $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$

↳ V. Muestral: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

DESVENTAJAS:
- Es sensible a valores extremos (alejados de la media)
- Su unidad es al cuadrado

Propiedades:

I. Es una cantidad NO negativa Var(x) = 0, ya que, se trata de una Σ de nº positivos xq son cuadrados. $S^2_x \geq 0$

II. La varianza de una cte es cero, debido a que las x_i son iguales. $S^2_x = 0$

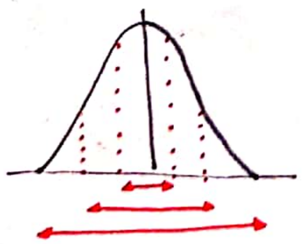
III. Si los valores de la variable se multiplican o dividen por una cte cambia la dispersión de los datos. Si $y_i = K \cdot x_i \longrightarrow V(Y) = V(K \cdot X)$

IV. Si una variable se le suma una cte la varianza NO cambia.

↳ El efecto de sumar o restar a cada valor de la variable una cte es que toda la distribución hacia la D(+) o I(−) una dist igual a la cte pero sin cambiar la dispersión. $Var(X+C) = Var(X)$

DESVIACIÓN ESTÁNDAR [S − σ]: Es la raíz cuadrada de la var, el cual permite resolver la interpretación de los resultados en terminos del problema.

$S = \sqrt{S^2}$



$\mu \pm \sigma = 68\%$

$\mu \pm 2\sigma = 95\%$

$\mu \pm 3\sigma = 99,7\%$

COEF. DE VARIACIÓN (CV): Cuando se desea comparar variabilidad de dos conj distintos, no siempre se puede emplear la desv. estandar

↳ El coef de var expresa la S como porcentaje de la media arit y es adimensional.

$C.V(\%) = \frac{S(x)}{X} \cdot 100$

PROBABILIDADES

Algunas definiciones:

- Es que tan posible es que ocurra un evento determinado. Mide el grado de certidumbre de que un suceso pueda ocurrir. Suele expresarse de 0 a 1

• EXPERIMENTO:
Acción, proceso u operación en el que se obtienen resultados definidos y que conllevan a la observación de estos resultados. Si es aleatorio su resultado es al azar.

• RESULTADO:
Valor que se obtiene de una sola repetición del experimento. (ensayo)

• ENSAYO:
Acto que lleva a un resultado determinado, de entre los posibles resultados distintos del exp.

ESPACIO MUESTRAL
Conj formado por todos los resultados posibles de un exp aleatorio. Se denomina "S"

PTO MUESTRAL:
uno de los resultados posibles del exp aleat.
$s \in S$

según el nº de elementos

- Discreto finito: nº finito de elementos Ej) lanzar un dado
- Discreto Infinito: nº inf numerable de elementos Ej) lanzar un dado hasta que salga 6
- Continuo: nº inf no numerable de elementos Ej) lanzar una flecha al blanco y medir la dist entre la flecha y el centro del blanco

Principios

- MULTIPLICACIÓN:
Si los conj $A_1, A_2, \dots A_K$ contienen respectivamente $n_1, n_2 \dots n_K$ elementos
$n_1 \cdot n_2 \dots n_K$ maneras de combinar un elemento de A_1, uno de A_2 y uno de A_K

- ADICIÓN:
Si los conj $A_1, A_2 \dots A_K$ contienen respect. $n_1, n_2 \dots n_K$, existen $n_1 + n_2 \dots + n_K$ maneras en que se puede dar el 1ro o el 2do o el K-esimo conj.

Uso de Esquemas

- **ARBOL:**
Secuencia lineal de pasos.
Puede tener ramas truncas.

- URNA:
Se trata de una extracción de elementos, pero se confecciona un arboriograma.
Pueden ser $\underset{\text{S. Ind}}{\text{CON}}$ ó $\underset{\text{S. Dep}}{\text{SIN reposición}}$.

- Suceso - Clase Exhaustiva:

↳ A_i es un suceso A respecto a un esp muestral S asociado a un exp E, es simplemente un conj de resultados posibles del mencionado experimento.

Ej) Arrojar un dado y observar su cara superior.

→ A = SUCESO PAR = $\{2,4,6\}$

→ B = SUCESO IMPAR = $\{1,3,5\}$

$\therefore A$ y $B \subset S$

↳ clase exhaustiva:

Si se tiene los sucesos $A_1, A_2, \ldots A_k$ estos forman una clase exhaustiva de "S" si:

I) $A_j \cap A_i = \emptyset$ para cualquier $j \neq i$; si los k sucesos son mutuamente exclusivos (No suceden simultaneamente)

II) $\bigcup_{i=1}^{k} A_i = S$, los k sucesos "completan" el espacio muestral "S".

- TEORIAS DE PROBABILIDAD: HERRAMIENTAS PARA USAR!

↳ Frecuencia relativa $\tilde{F}(A)$ → Es el nº de veces que se da el resultado A en relación al nº de veces que se repite el exp → $\tilde{F}(A) = F(A)/n$

Propiedades que presenta:

- $0 \leq \tilde{F}(A) \leq 1$
- $\tilde{F}(A) = 1$ sí y solo sí A ocurre cada vez (siempre) en las n repeticiones.
- $\tilde{F}(A) = 0$ si A NO OCURRE nunca en las n repeticiones.
- Si A y B son dos sucesos mutuamente excluyentes → $\tilde{F}(A \cup B) = \tilde{F}(A) + \tilde{F}(B)$
- Si $\tilde{F}(A)$ converge en un valor fijo se denomina $P(A)$, cuando $n \to \infty$.

↳ Teoría Clásica

La prob de un evento A es $P(A) = \dfrac{\text{nº de casos favorables de A}}{\text{nº de casos totales}}$

con resultados equiprobables y prob entre 0 y 1.

Todos los resultados = posib de ocurrir.

↳ Teoría Axiomatica

La prob del suceso A cumple:

La prob del suceso A (P(A)) es el nº real que satisface →

I) $0 \leq P(A) \leq 1$

II) $P(S) = 1$

III) Si $A_1, A_2, A_3 \ldots$ son sucesos que se excluyen mutuamente

$P\left(\bigcup_{i=1}^{k} A_j\right) = P(A_1) + P(A_2) + P(A_3) + \ldots$

Definiciones:

• EVENTO MUTUAMENTE EXCLUYENTE:
La ocurrencia de uno impide la ocurrencia del otro en un mismo exp básico.
A y B son mutuamente excluyente
$\longleftrightarrow$ $A \cap B = \emptyset$.

• PROB CONDICIONAL:
Dados dos sucesos A y B, la prob de que ocurra A habiendo ocurrido B se escribe $P(A/B)$
∴ A está condicionado por la aparición previa de B.

• EVENTOS PROB. INDEPENDIENTES:
La ocurrencia de uno NO modifica la prob de ocurrencia del otro en más de un ensayo o exp básico.
A y B son ind si $P(A) = P(A/B)$
$P(B) = P(B/A)$

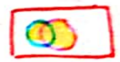
- EVENTOS COMPUESTOS -

Se combinan 2 ó más eventos simples usando los terminos { Y / O / DADO

Tipos:

I. $P(\overline{A}) = P(A_c)$ Prob de evento complementario → Relación: $P(\overline{A}) = 1 - P(A)$

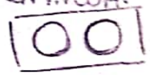
II. $P(A \cup B) =$ Prob de que ocurra A o B



III. $P(A \cap B) =$ Prob de que ocurra A Y B

IV. $P(A/B) =$ Prob de que ocurra A sí ocurrió B.

OBS! NO EXISTE INTERSECCIÓN SI LOS SUCESOS SON INCOMP



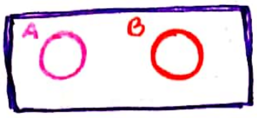
$P(A \cap B) = 0$

• EVENTO COMPLEMENTARIO:
Es el conj de todos los ptos muestrales que no pertenecen al conjunto.
Dado A, es $\overline{A}$ su complemento

Ejemplo:
Exp: lanzar un dado muestral donde $S = \{1,2,3,4,5,6\}$
Eventos posibles:
$A = \{$Sacar un 1$\}$ Evento Simple
$B = \{$Sacar un nº par$\} = \{2,4,6\}$ Evento Comp.

• OBSERVACIONES: (ya que $P(A \cap B) = 0$)

↳ Evento excluyentes:



$A \cap B = 0 \longrightarrow P(A \cup B) = P(A) + P(B)$

Sea A y B dos sucesos cualquiera, se dice que son mutuamente excluyente si ambos sucesos NO pueden ocurrir al mismo tiempo.

↳ Evento NO excluyente:

$A \cap B \neq 0 \longrightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Sea A y B dos sucesos cualquiera, se dice que son mutuamente excluyente si ambos sucesos pueden ocurrir al mismo tiempo.

Ejemplos:

(I) Calcular la prob de sacar un nº par o impar al lanzar un dado

$S = \{1, 2, 3, 4, 5, 6\}$

A = sale nº par → $P(A) = \frac{3}{6}$ B = Sale un nº impar → $P(B) = \frac{3}{6}$

$P(A \cup B) = P(A) + P(B) = \frac{3}{6} + \frac{3}{6} = 1$

(II) En una cafetería el 70% de los clientes toma café, el 40% lee el periodico mientras espera ser atendido y solo al 25% de los clientes les gusta leer el periodico mientras toma su cafe.

¿ Prob de que TOME CAFE ó LEA EL PERIODICO ?

$P(T) = 0{,}7$ $P(L) = 0{,}4$ $P(T \cap L) = 0{,}25$

$P(T \cup L) = P(T) + P(L) - P(T \cap L)$
$= 0{,}7 + 0{,}4 - 0{,}25$
$= 0{,}85$

$P(\bar{T} \cap B) = P(\bar{T}) \cdot P(B/\bar{T})$
$= P(\bar{T}) \cdot P($

↳ Eventos Independientes:

Dos ó más eventos son ind cuando la ocurrencia de uno no afecta la prob de ocurrencia del otro / $P(A/B) = P(A)$ ó $P(B/A) = P(B)$

↳ FORMULA $P(A/B) = P(A)$ Prob condicional

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

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

▸ Eventos dependientes:

Dos o más eventos son dep cuando la ocurrencia de uno afecta la prob de ocurrencia del otro /

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

$P(A/B) \neq P(A)$

$P(A \cap B) - P(A/B)$

Escaneado con CamScanner

Experimento Bernoulli

Modelos de Prob:

- Binomial
- Hipergeom.
- Poisson

• Es un exp aleatorio que tiene dos resultados posibles (ÉXITO ó FRACASO) y son mutuamente excluyentes.

• Prob éxito = p y Prob de fracaso $q = 1 - p$ ya que, son sucesos complementarios.

↳ Esperanza de x: $E(x) = p$.

↳ Varianza de x: $Var(x) = p \cdot q = p \cdot (1-p)$.

En general: A partir de un exp

RESULTADO EXP	x	$P(x)$	$x \cdot P(x)$	$x^2 \cdot P(x)$
D	1	p	$1 \cdot p = p$	$1^2 \cdot p$.
ND	0	$1-p$	$0 \cdot (1-p) = 0$	$0^2 \, (1-p)$
T		1	p	p

Esperanza: $E[x] = \sum_{i=1}^{2} x_i \, P(x_i) = 1 \cdot p + 0\,(1-p) = p$

Varianza: $Var(x) = E(x^2) - E(x)^2 = p - p^2 = p(1-p)$

• BINOMIAL

. Si al exp Bernoulli se lo repite "n" veces, se tiene resultados posibles mutuam excluyentes e ind. ∴ las prob se mantienen ctes.

. Es una distrib de prob discreta que nos dice el porcentaje en que es probable obtener un resultado entre dos posibles al realizar un nº n de pruebas. La prob NO puede ser − ni mayor a 1.

CARACTERISTICA:
- Solo 2 posibles resultados
- Resultados ind
- Prob de éxito cte en todas las repeticiones
- El exp se repite n veces igual condiciones
- Cuando hay extracción de elementos se deben reponer.

• HIPERGEOMETRICA

CARACTERISTICAS:
- El resultado puede ser EXITO ó FRACASO.
- Se realizan n ensayos o repeticiones donde n es finito.
- Las repeticiones son SIN REPOSICIÓN.
  ∴ No son sucesos ind, ya que, cambia la composición de la población y por ende la prob. de éxito.
- La población esta formado por N elementos con K exitos y $N-K$ fracasos.

• Sea x el nº de EXITOS en la muestra n se debe cumplir que $n \leq N$, es decir, que se extra una muestra de tamaño n de una población de N elementos sin reposición.

- $x: \{0, 1, 2 \dots n\}$ - OBS! : calcula la prob de exitos (x) en la muestra con n elementos.

OBS!

$N \begin{cases} K & \text{(cant de exito de la pob)} \\ N-K & \text{(cant de fracaso de la pob)} \end{cases}$

$n \begin{cases} x & \text{(cant de exito en la muestra)} \\ n-x & \text{(cant de fracaso de la muestra)} \end{cases}$

TERMINOS DE LA FORMULA:

$\binom{N}{n} \longrightarrow$ Total de casos / n elementos de la población.

$\binom{K}{x} \longrightarrow$ Total casos favorables / cuanta de las muestras de n elementos tiene x exitos y $(n-x)$ fracasos. Donde los x se extraen

$\binom{N-K}{n-x} \longrightarrow$ de los K exitos de la población

$\binom{K}{x}\binom{N-K}{n-x} \longrightarrow$ x exitos y fracasos restantes.

ESPERANZA: $E(Hip) = n(K/N)$ con n = cant de elementos de la muestra
K/N = Prob de éxito.

PEX

VARIANZA: $Var(Hip) = n \underbrace{\frac{K}{N}}_{PEX} \cdot \underbrace{\frac{(N-K)}{N}}_{PFr} \cdot \underbrace{\frac{(N-n)}{(N-1)}}_{\text{Factor de corrección}}$ con $\frac{(N-K)}{N}$ = Prob de fracaso

Poisson

- Cada una de las v. a representa el n° total de ocurrencia de un FENOMENO en un continuo.

- Expresa la prob de un n° K de ocurrencia producidas en un continuo fijo (ej. El tiempo), si estos eventos ocurren con una frecuencia media conocida y son ind del continuo.

- La dist de probabilidad de Poisson es una dist de prob de una V.d que ~~nos~~ propoiciona la prob de que ocurra un determinado suceso un n° de veces K en un int de tiempo, espacio, vol, etc.

Donde el parametro λ es el n° medio de veces que ocurre el suceso en un int continuo que puede ser: tiempo, espacio, vol, etc.

$\hookrightarrow$ Su función de prob es $P(X=K) = \dfrac{e^{-\lambda} \cdot \lambda^{K}}{K!}$ para $X = 0, 1, 2 \ldots$

e: log. natural.

$\hookrightarrow$ COND: I. $P(X_i) \geq 0 \quad \forall i = 1 \ldots n$
II. $\Sigma P(X_i) = 1$

I) $\lambda > 0$ y $e^{-\lambda} > 0$ y $X! > 0 \longrightarrow \therefore \dfrac{e^{-\lambda} \cdot \lambda^{X}}{X!} > 0$

promedio de aciertos q ocurre en un int cont.

II) serie de Madaurin de la función exponencial donde: $P_x(x) = e^{-\lambda} e^{\lambda} = e^{0} = 1.$

Escaneado con CamScanner

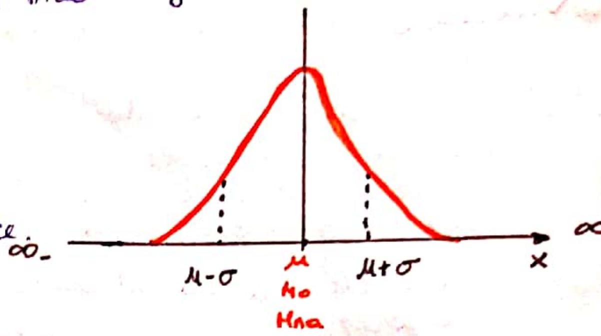
Distribuciones Discretas

Bernoulli	Binomial	Hipergeométrica	Poisson
P(A) = π P($\overline{A}$) = 1 - π	x: número de éxitos p: probabilidad de éxito q: probabilidad de fracaso n: número de repeticiones del experimento básico.	x: número de éxitos k: Número de éxitos en la población N: cantidad de elementos de la población n: número de repeticiones del experimento básico	x: número de sucesos que ocurren en un continuo. λ: promedio de aciertos que ocurren en un intervalo continuo
$P_x(x) \begin{cases} \pi \text{ si } x\text{: éxito } (x=1) \\ (1-\pi) \text{ si } x\text{: fracaso } (x=0) \end{cases}$	Bin (x=.. p=.,n=..)= $\binom{n}{x} p^x . q^{n-x}$ para x = 0,1,.....n $\binom{n}{x} = \frac{n!}{x!(n-x)!}$	Hip(x=..,n=..,k=..,N=..) = $\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ x .... n x como suma = K	$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ Para x = 0,1,2,.....
Sólo dos resultados posibles. Mutuamente excluyentes.	- El experimento básico se repite un número finito de veces **n**. - Existen dos resultados posibles; Éxito o Fracaso (mutuamente excluyentes) - Las respuestas son independientes. - Las probabilidades se mantienen constantes. • con reposición	- El experimento básico se repite un número finito de veces **n**. - Existen dos resultados posibles; Éxito o Fracaso (mutuamente excluyentes) - Las respuestas son **dependientes.** - Las probabilidades **no** se mantienen constantes. • Sin reposición	-Es posible dividir al intervalo continuo considerado en subintervalos. -La probabilidad de un acierto permanece constante a lo largo de los intervalos. -La probabilidad de más de dos aciertos en un subintervalo es suficientemente pequeña como para ignorarla. -Los aciertos son independientes ya que los intervalos no se superponen. - La probabilidad de que un acierto sencillo ocurra en un intervalo muy corto es proporcional a la longitud del intervalo.
E(x) = π V(x) = π (1-π)	E(Xb) =n.p V(Xb)=n.p.(1-p)	E(Xh) =n (K/N) $V(Xh) = n\frac{K}{N} \cdot \frac{(N-K)}{N} \frac{(N-n)}{(N-1)}$	E(Xp) = λ V(Xb) = λ

- Distribución Normal:

Es la distribución continua que más se utiliza en estadística; es un modelo que aproxima el valor de una variable aleatoria a una situación ideal dependiendo de la media y desviación estándar.

→ PROPIEDADES:



- Es simetrica respecto a la media μ.
- Tiene un máx en la media μ.
- Crece hasta la media μ y luego decrece.
- En los ptos $\mu-\sigma$ y $\mu+\sigma$ presenta ptos de inflexión.
- El eje de abscisas es una asintota de la curva.
- Tiene forma de campana.
- Simetrica respecto a la media μ, mediana Mna y moda Mo.

→ Se caracteriza por dos parametros: la media μ y la varianza σ^2.

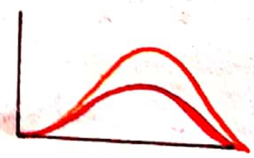
NOTACIÓN $X \sim N(\mu, \sigma^2)$ donde $E[X] = \mu$ y $V[X] = \sigma^2$

→ Su función es: $F(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$; $x \in R$

→ Interpretación geométrica:

. La media (μ) → Factor de traslación

. La Varianza (σ^2) → Factor de escala



OBS!!

Para conocer la función de dist F(x) se debe integrar la función de densidad de prob, implica un calculo no optimo.

$\therefore$ F(x) esta tabulada.

OBS! σ determina el grado de dispersamiento de la curva

$\sigma\uparrow$ más se dispersan los datos entorno a la media y la curva será más plana.

- Cálculos de probabilidades asociados a una curva normal específica:

→ Si X es una variable aleatoria que sigue una dist normal con media μ y desviación típica σ, se puede transformar en una variable aleatoria Z que sigue una distribución normal con $\mu = 0$ y $\sigma = 1$ definida "Dist normal estándar".

Donde $Z = \dfrac{X - \mu}{\sigma}$ es una traslación y cambio de escala de la variable original.

(*) Nos permite así comparar entre dos valores de dos distribuciones normales diferentes, para saber cual de los dos ~~extremos~~ es más extremo.

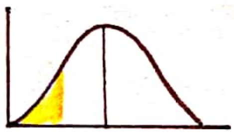
- TRABAJO CON DISTRIBUCIONES TABULADAS:

– distribución Normal:

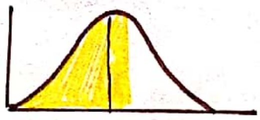
Z	valor de Z
valor de Z	probabilidades

Z	valor de Z
valor de Z	probabilidades

para $Z < 0$



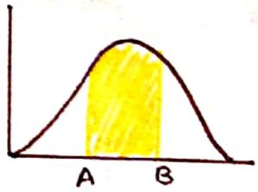
para $Z > 0$



La prob se acumulan de izq a dch.

OBS! PROBABILIDAD ENTRE DOS VALORES = PROB. DEL VALOR MAYOR − PROB. DEL VALOR MENOR

$P(A \leq Z \leq B) = P(Z \leq B) - P(Z \leq A)$



– distribución Chi-cuadrado de Pearson con n grados de libertad:

. Se define como $\chi_n^2 = x_1^2 + x_2^2 + \ldots + x_n^2$ donde $x_1; x_2 \ldots x_n$ son variables ind y $x_i \sim N(0;1)$ para $i = 1, 2, \ldots n$

. Se obtiene al sumar n v.a que se distribuyen como normales.

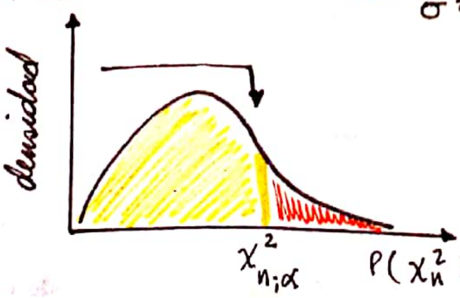
. La función de densidad es asimétrica positiva.

. Si aumenta n → la función de densidad es más simétrica. (tiende a la normal)

. Presenta el parámetro grados de libertad:

. Con $\chi_{n-1}^2 = \dfrac{(n-1)\, s^2}{\sigma^2}$

{ · Buscar gl
{ · Buscar valor de la variable



$P(\chi_n^2 \geq \chi_{n;\alpha}^2) = \alpha$

gl \ P	PROBABILIDAD
GRADO DE LIBERTAD	Valores de la variable

$\mu = n$

$\sigma^2 = 2n$

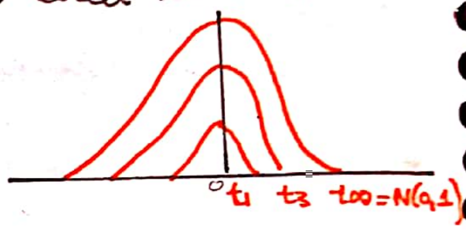
- distribuciones t de student con n grados de libertad

Sea "z" una v.a que sigue una distribución $N(0,1)$
Sea "χ_n^2" una v.a que sigue una distribución chi-cuadrado con n grados de libertad. Ambas variables independientes forman la nueva variable

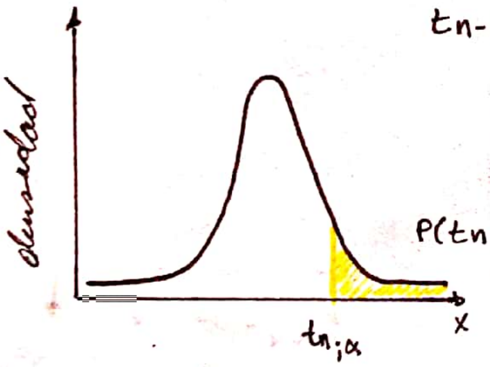
$$t_n = \frac{z}{\sqrt{\chi_n^2/n}}$$

la cual sigue una distribución t-student con n grados de libertad

- cuando los grados de libertad aumentan, más cerca esta a $N(0,1)$
- simetrica respecto a cero.
- los valores lejanos a cero son anómalos.



$$t_{n-1} = \frac{x-\mu}{s}$$



$$P(t_n \geq t_{n;\alpha}) = \alpha$$

gl \ P	Probabilidad
grado de libertad	variable

$\mu = E(x) = 0$

$\sigma^2 = \frac{n}{n-2}$

OBS! . A medida que aumenta el tamaño de la muestra la distribución de probabilidad del promedio muestral se hace cada vez más acampanada, concentrandose alrededor del promedio de la población original.

- A medida que $n \uparrow \rightarrow \sigma^2 \downarrow$ y $\mu = cte.$

$$\begin{cases} - \mu = E(\bar{x}) \quad \forall n \\ - \sigma^2_{\bar{x}} = \dfrac{\sigma^2_x}{n} \end{cases}$$

A partir de esta idea tenemos: Distribución muestral de $\bar{X}_n$

- Sea X una v.a / $E(X) = \mu$ y $V(X) = \sigma^2$ y una muestra aleatoria $n \longrightarrow$

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

$$V(\bar{x}) = \frac{\sigma^2}{n}$$

$$\text{Si } x \sim N(\mu; \sigma^2) \longrightarrow \bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$$

* Si X se distribuye normalmente $\rightarrow$ la media aritmetica $\bar{x}$ también.
* La varianza de la media muestral < varianza original.

- TEOREMA CENTRAL DEL LIMITE:

  "Dada una población cualquiera con μ y σ^2 la distribución muestral de $\bar{x}$, calculada a partir de las muestras aleatorias de tamaño n con reemplazo de ésta, estará distribuida en forma aprox normal con $\mu_{\bar{x}} = \mu$ y $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$ donde la aprox es mejor cuando n es mayor"

tamaño n es bueno si $n \geq 30$.

$n \downarrow \sigma^2 \uparrow$

Intervalo de confianza:

- Es un rango de valores, derivados de los estadísticos de la muestra, que posiblemente incluya el valor de un parámetro desconocido.

- CONFIANZA: La confianza es la probabilidad de contener con un intervalo un parámetro.

	Parámetro Θ		Estimador $\hat{\Theta}$
μ	La media poblacional se define como el promedio de todos los elementos de la población. Es la media de la característica de un grupo.	$\bar{x}$	La media muestral es la media aritmética de los valores muestrales aleatorios extraídos de la población.
σ^2	La varianza poblacional es una medida de la dispersión de los datos de la población. Promedio de la distancia desde cada pto de datos en una población a la media, al cuadrado e indica como se distribuyen los ptos de datos en la población.	S^2	Varianza de la muestra de una población.
π	Fracción de individuos en una población con una determinada característica	p	

- Propiedades de los estimadores:

- Se dice estimador INSESGADO si $E[\hat{\theta}] = \Theta$
- El estimador $\hat{\Theta}_1$ es más eficiente que $\hat{\theta}_2$ si $Var(\hat{\Theta}_1) < Var(\hat{\Theta}_2)$
- Un estimador es consistente si $\lim_{n \to \infty} \hat{\theta} = \Theta$
- Un estimador SUFICIENTE para un parámetro es aquel capaz de recoger o resumir toda la info que la muestra de una V.a X contiene.

I.C PARA μ CON VARIANZA CONOCIDA

Si $(x_1, x_2, x_3 \dots x_n)$ es una muestra aleatoria de tamaño n de una v.a X donde $X \sim N(\mu ; \sigma^2)$ con σ^2 conocida, un int confianza para μ de nivel $1-\alpha$ es $\left[\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$

- Se define estimador puntual $\longrightarrow \mu = \bar{x}$
- A partir del estimador definido para el parametro se construye el estadistico $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Donde Z es el "pivote" que contiene el verdadero parámetro μ y que bajo las cond dadas $Z \sim N(0,1)$.

PARAMETRO A ESTIMAR:
μ con σ^2 conocida.
VARIABLE A USAR:
$Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
· INT CONFIANZA:
$\bar{x} \pm Z_{(1-\frac{\alpha}{2})} \cdot \dfrac{\sigma}{\sqrt{n}}$

DEDUCCIÓN:

$P(LI \leq \mu \leq LS) = 1 - \alpha$

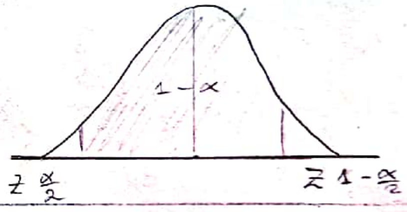
$P(Z_{0\frac{\alpha}{2}} \leq Z \leq Z_{1-\frac{\alpha}{2}}) = 1 - \alpha$ reemplazo por el estadistico que C parametro

$P\left(Z_{0\frac{\alpha}{2}} \leq \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$

despejo el parametro μ

$P\left(\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$

OBS! Por ser simetrica se emplea $Z_{1-\frac{\alpha}{2}}$



I.C PARA μ CON VARIANZA DESCONOCIDA

Si $(x_1 ; x_2 ; \dots x_n)$ es una muestra aleatoria de tamaño n de una v.a X donde $X \sim N(\mu ; \sigma^2)$ con σ^2 desconocida, un int de confianza para μ de nivel $1-\alpha$ es $\left[\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} ; \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \right]$

- Como se desea hallar μ usamos como estimador $\bar{x}$, pero se debe buscar otro estadistico, ya que, σ^2 es desconocida $\therefore$ se propone como pivote una VA parecida a Z reemplazando σ por un estimador adecuado. En este caso se estima σ con S y se propone como pivote $T = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$ donde T sigue la distribución de Student con parametro n-1.

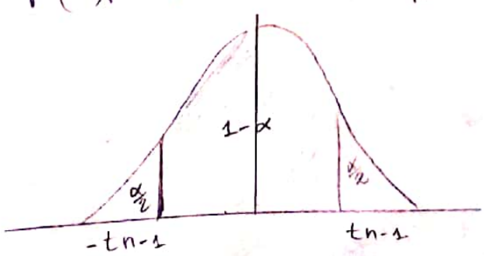
DEDUCCIÓN

$P(LI \leq \mu \leq LS) = 1 - \alpha$

$P(t_{n-1 ; \frac{\alpha}{2}} \leq t \leq t_{n-1 ; 1-\frac{\alpha}{2}}) = 1 - \alpha$

se reemplaza t y despeja μ. con $t = \dfrac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$

$P\left(\bar{x} - t_{n-1 ; 1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1 ; 1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\right) = 1 - \alpha$



PARAMETRO A ESTIMAR
μ con σ^2 desconocida
VARIABLE A USAR
$t_{n-1} = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$
· IC: $\bar{x} \pm t_{(n-1 ; 1-\frac{\alpha}{2})} \cdot \dfrac{S}{\sqrt{n}}$

INTERVALO DE CONFIAZA PARA LA PROPORCION

Si $\hat{p}$ es la proporción de observaciones de una muestra aleatoria de tamaño n que verifican una propiedad de interés, entonces un int de confianza para la proporción p de la población que cumple dicha propiedad de nivel aprox $1-\alpha$ es
$$\left[\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \; ; \; \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

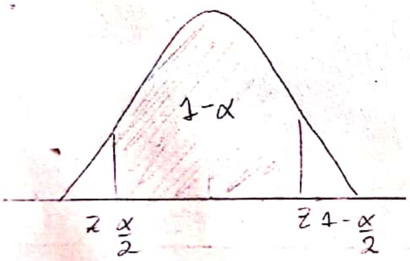
- Para realizar un int para la proporción π se utiliza el estimador puntual $\hat{p}$ y la distribución normal.
- Se define el estadístico a usar $Z = \dfrac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$.

DEDUCCIÓN.

$$P(z_{\frac{\alpha}{2}} \leq z \leq z_{1-\frac{\alpha}{2}}) = 1-\alpha$$

$$P\left(p - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$$

Se reemplaza Z por su estadístico y despeja π.



PARAMETRO A ESTIMAR
π
VARIABLE A UTILIZAR
$Z = \dfrac{p-\pi}{\sqrt{\frac{p(1-p)}{n}}}$
I.C.,
$p \pm z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{p(1-p)}{n}}$

INTERVALO DE CONFIANZA PARA LA VARIANZA

Si $(x_1 ; x_2 ; \ldots x_n)$ es una muestra aleatoria de una v.a X, donde $X \sim N(\mu ; \sigma^2)$ un intervalo de confianza para σ^2 de nivel $1-\alpha$ es
$$\left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \; ; \; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2} ; n-1}} \right]$$

- Se define es estimador puntual de σ^2 a $S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$
- A partir del estimador se construye el estadístico $\chi = \dfrac{(n-1)S^2}{\sigma^2}$
- El estadístico contiene al parametro desconocido a estimar σ^2 y tiene una distribución Chi cuadrado con n-1 grados de libertad

DEDUCCIÓN

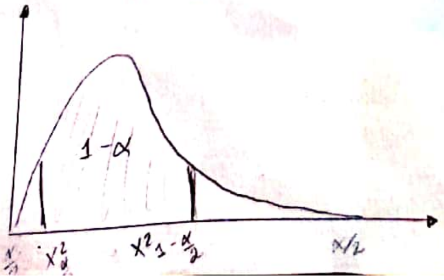
$$P(\chi^2_{\frac{\alpha}{2}} < \chi^2 < \chi^2_{1-\frac{\alpha}{2}}) = 1-\alpha$$

$$P\left(\chi^2_{n-1 ; \frac{\alpha}{2}} < \frac{(n-1) \cdot S^2}{\sigma^2} < \chi^2_{n-1 ; 1-\frac{\alpha}{2}}\right) = 1-\alpha$$

se reemplaza por el estadístico

$$P\left(\frac{(n-1)S^2}{\chi^2_{n-1 ; 1-\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1 ; \frac{\alpha}{2}}}\right) = 1-\alpha$$

se acomoda y usan propiedades de desigualdades



PARAMETRO A ESTIMAR σ^2
VARIABLE A USAR
$\chi^2_{n-1} = \dfrac{(n-1)S^2}{\sigma^2}$
$LI = \dfrac{(n-1)S^2}{\chi^2_{(n-1)(1-\frac{\alpha}{2})}}$ $LS = \dfrac{(n-1)S^2}{\chi^2_{(n-1)(\frac{\alpha}{2})}}$

PRUEBA DE HIPOTESIS / DOCIMAS DE HIPOTESIS

HIPOTESIS ESTADISTICA

Es una afirmación de que algo es verdadero.

· Se realizan sobre:
- Parametros de una Población
- Distribución de probabilidad de una variable aleatoria.

PASOS:

INTERES 1° → H_0 → H_1 2°

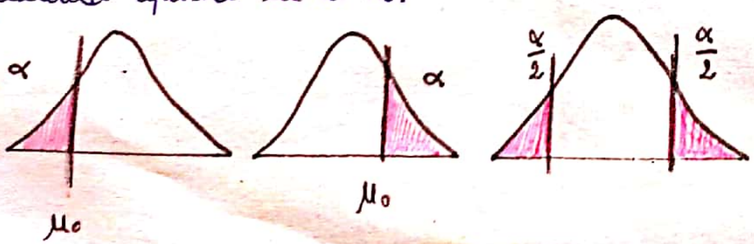
Hipótesis NULA: Afirmación sobre un parametro poblacional que tiene un valor específico. Pto inicial de la investigación.

Hipótesis ALTERNATIVA: Complemento de la hip nula.

OBS!!
Se trabaja suponiendo que la primera es una afirmación verdadera → La decisión se toma según la nula (se rechaza o NO la HIP NULA)

Tipos $\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$ $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$

La igualdad aparece en la H_0.



Para llevar a cabo la experiencia sobre el comportamiento de algun índice o la exigencia del cumplimiento de una Norma, lleva a realizar proposiciones sobre el valor de algún parámetro estadístico. Proc. para juzgar si una propiedad que se supone en la población es compatible con lo obs de la muestra

Las proposiciones se deben contrastar con la realidad (mediante el muestreo de datos) para tomar una decisión entre aceptar o rechazar la proposición.

Regla que determina si se puede aceptar o rechazar una afirmación acerca de una población dep de la evidencia de la muestra.

{ Las proposiciones se conocen como → HIPOTESIS
El procedimiento para decidir → PRUEBA DE HIPOTESIS

⊛ TIPO DE ERRORES:

No se puede eliminar la posibilidad de errores, se considera la prob de que suceda:

α: Prob de cometer ET1
Prob de Rechazar Ho siendo V.

β: Prob de cometer ET2
Prob de No Rechazar Ho siendo F.

DECISIÓN	Ho V	Ho F
No Rechaza Ho	Decisión correcta	ET2
Rechaza Ho	ET1	Decisión correcta

PASOS PARA REALIZAR UNA DOCIMA DE HIPOTESIS

I. Definir HIPOTESIS según el parametro de interes

II. Definir un nivel de significación α

III. Seleccionar estadístico o variable pivotal

IV. Dibujar estadístico o v. pivotal
Marcar la región crítica
Definir ~~formula~~ regla de decisión.

V. Reunir datos muestrales y calcular el valor del estadístico o variable pivotal

VI. Decisión y conclusión en terminos del problema

P.H
⊛ Parametricas:
Conj fijo de parametros que se utilizan para determinar un modelo poblacional
→ Se conoce la dist poblacional o se puede estimar (Gaussiana)

PRUEBA DE HIPOTESIS NO PARAMETRICAS

- Prueba de hipotesis apicables a mediciones nominal u ordinal.
- Estadistica No parametrica: Estudia las pruebas y modelos estadisticos, cuya distribución no puede ser definida a PRIORI, los datos observados determinan.
- Se trabaja con la prueba de Chi cuadrado (x^2):
  Al usar una distribución X^2 se puede determinar si las diferencias entre los conjuntos de frecuencia son
  → significativas
  → demasiados grandes para representar la muestra

x^2 características:

- x^2 es siempre ≥ 0 xq es una $\Sigma\square$
- varía de 0 a ∞.
- familia de distribuciones que dependen de 1 parametro. (gl)

Prueba de hip No parametricas más usadas:

- BONDAD DE AJUSTE → Dist determinada? 1 MUESTRA
- HOMOGENEIDAD → Son Homog? + 1 MUESTRA
- INDEPENDENCIA → Son ind? 1 MUESTRA

- **BONDAD DE AJUSTE:**

Se parte de una muestra aleatoria de tamaño n, proveniente de una población cuya distribución de probabilidad es desconocida. {Binomial, Poisson, Hiperg...}

La H_0 indica que los datos se ajustan a una distribución determinada

- Las obs n se acomodan en una tabla de frecuencias ~~esperada E_i~~. con K int de clases. (O_i)
- De la dist de prob se calcula la frecuencia esperada (E_i)
- Estadístico: $\chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}$

con dist χ^2 y $K-1-r$ gl. donde r = nº de parámetros de la pob a esti...

- n gl ↑ → la distribución se aproxima a la normal.
- La hip se rechaza si: $\boxed{\chi^2 > \chi^2_{1-\alpha;K-1-r}}$
- Si se tiene frecuencias esperadas pequeñas (valores menores de 5) → combinar int de clases adyacentes

- **PRUEBA DE INDEPENDENCIA:**

Se utiliza para determinar si la pertenencia de una v a categorías es diferente como función de la pertenencia a la categoría de una 2da variable. (si están o no relacionadas)

- Se utiliza una <u>tabla de contigencia</u>, permite poner a prueba si dos criterios de clasificación de una misma muestra son IND o no.
- Se calcula las frecuencias esperadas / $\sum$ frecuencia esperadas $= \sum$ frecuencias observadas
- Estadístico: $\chi^2 = \sum_{i=1}^{f} \cdot \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

r = nº categorías de la variable fila

c = nº de categorías de la variable columna

O_{ij} = nº obs

E_{ij} = nº esperado

- Grados de libertad:

$g.l = (f-1).(c-1)$

PRUEBA DE HOMOGENEIDAD:

- Se comparan dos ó más poblaciones independientes.
- En cada población se toma una muestra y los individuos de la misma se clasifican según la variable Y que puede tomar m valores.

Una de las dos variables involucradas es controlada por el investigador (modelo predeterminado)

- Sirve para contrastar dos hipótesis:

H_0: Todas las poblaciones tienen la misma distribución para la variable y.

H_1: La distribución de la variable Y en algunas poblaciones es diferente.

Se desea comprobar que los datos de las distintas muestras son homogeneos.

- Estadístico: $\chi^2 = \sum^{F} \sum^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
- grados de libertad: $gl = (f-1)(c-1)$
- OBS! El valor esperado se calcula de forma diferente que en Indep. ya que, esta fijado de ante mano el total de filas o columnas.

extensión

P. IND	P. HOMO
• Es una prueba de Chi □ de dos ó + variables. El objetivo es determinar si la 1ra V esta relacionada o No con la 2da V.	• La prueba se aplica a UNA sola V. categorica de dos o diferentes pob. • Se utiliza para determinar si los recuentos de frecuencia se distribuyen de forma identica en las diferentes pob.
1 MUESTRA	+ de una muestra de diferentes poblaciones
2 ó + VARIABLES	1 VARIABLE

REGRESIÓN Y CORRELACIÓN

DESER.

Regresión: v fija, v aleatoria

Correlación: Dos v. aleatorias

Análisis de Regresión y correlación requiere de 2 o + variables cuantitativas en forma simultánea.

Regresión

- Estudia la relación funcional de una o más variables respecto de otra. (v. cuantitativas)
- "Simple" xq interviene una sola variable independiente.
- OBJ!

  Hallar una función o modelo matemático para predecir y estimar el valor de una variable a partir de valores de otra.

Y: v. dependiente / respuesta. Es la variable que se desea predecir o estimar y.

X: v. independiente / predictora. Es la variable que provee las bases para estimar.

- Los valores de la v. ind X son fijos. Es decir, que son PRESELECCIONADOS por el investigador. Se mide sin error o se desprecia el error.
- ∀ valor de "x" ∃ una subpoblación de valores de "y" normalmente distribuida.
- Las medias de las subpoblaciones se sitúan sobre la línea recta del modelo teórico.

Correlación

- Estudia la magnitud o grado de asociación entre las variables.

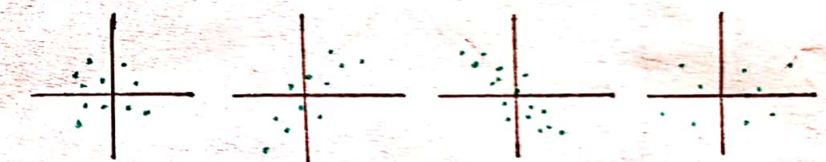
- X e Y tienen una distribución de probabilidad conjunta normal bivariada:
  - para ∀ "x" ∃ una subpoblación de valores de "y" que sigue una distribución normal.
  - para ∀ "y" ∃ una subpoblación de valores de "x" que sigue una dist normal.

variación en cov

$$\rho = \frac{cov(xy)}{\sqrt{Var(x)Var(y)}}$$

variación de x, variación de y

- coeficiente de correlación
  - → poblacional ρ
  - → Muestral r



r cerca de 0 | r>0 | r<0 | No hay relación lineal

Ptos dispersos, x e y son Ind.

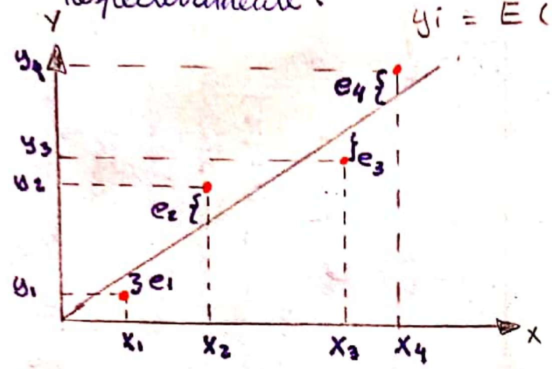
- Propiedades de ρ:
  - → Es adimensional
  - → Toma valores entre [-1; 1]
  - → Si no hay correlaciones lineales → $\rho = 0$
  - → Relación lineal perfecta entre dos variables → $\rho = 1$, $\rho = -1$
  - → Cuanto más cerca está ρ de 1 o -1 mejor será el grado de relación lineal.

Más sobre regresión . . . .

- **MODELO TEÓRICO:** $\mu_{y/x} = \alpha + \beta x$

- Implica que todas las medidas de las subpoblaciones de "y" están sobre la misma recta.
- α y β son los coeficientes de regresión de la población y geométricamente representan la oo y la pendiente de la recta, respectivamente.

$$y_i = E(y/x) = \alpha + \beta x + \underbrace{e_i}_{error}$$



- **MODELO ESTADÍSTICO:** $\mu_{y/x} = \alpha + \beta x + e$

- Los valores de y son estadísticamente independientes
- Se tiene en cuenta el error "e" $\quad e = y - (\alpha + \beta x)$
- Los errores están normalmente distribuidos $S^2 = S^2$ común de las subpoblaciones de valores "y".

α: ordenada al origen
Indica el valor medio poblacional de la variable respuesta Y cuando $X = 0$.
OBS! La interpretación No tiene sentido si x no puede valer cero.

β: Es la pendiente de la línea de regresión lineal.
Inica el cambio del valor medio poblacional de la v. respuesta Y cuando x se incrementa en una unidad.

- **MODELO ESTIMADO** $\hat{y} = a + bx$

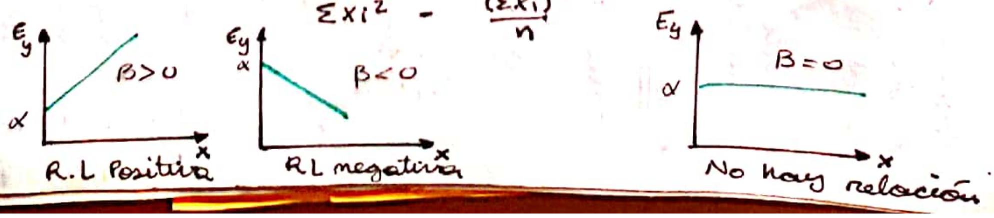
a es un estimador de α $\longrightarrow$ oo: Indica el valor promedio estimado de la variable respuesta cuando $x = 0$

b es un estimador de β $\longrightarrow$ pendiente: Indica el cambio promedio estimado en la v respuesta cuando la v predictora aumenta en una unidad adicional.

- MÉTODO DE LOS MÍNIMOS CUADRADOS:
Consiste en minimizar la dist vertical entre los ptos y la recta.

$$a = \bar{y_i} - b\,\bar{x_i} \qquad b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

Casos posibles:



Inferencia

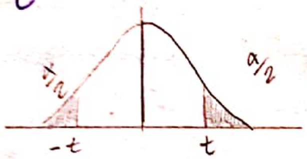
correlación y Regresión lineal

correlación

.PRUEBA:

- Hipotesis para el coef de correlación:

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$$



- α
- V. pivotal $t_{n-2} = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

OBS!
Al trabajar con dos v aleatorias es posible calcular el coeficiente de correlación r → Indica la asociación entre las v Y y X.

- Si $\rho = 0$ NO EXISTE ASOCIACIÓN ENTRE LAS VARIABLES
- Si $\rho \neq 0$ EXISTE ASOCIACION ENTRE LAS VARIABLES

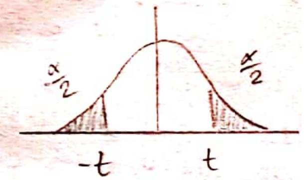
Regresión

Se utilizan los estimadores

- $a \cong N(E(a)) = \alpha$
- $b \cong N(E(b)) = \beta$
- $\hat{y} \cong N(E(\hat{y})) = \alpha + \beta x$
- Error estándar de la estimación: Mide la dispersión o alejamiento promedio de los ptos con respecto a la recta estimada. Se ó $S_{Y/X}$
- Desviación típica de los coeficientes de Regresión → Sb, Sa

OBS!
La hip más importante en regresión es la de la pendiente β, ya que, indica si hay una relación funcional entre las variables.

- PRUEBA DE HIP:
- Hip del coef de regresión: $\begin{cases} H_0: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$
- α
- V. pivotal $t = \dfrac{b - \beta}{Sb} \approx t_{n-2}$



- Sí $\beta = 0$ ∄ relación funcional poblacional de las variables
- Sí $\beta \neq 0$ ∃ relacion funcional poblacional de las variables

OBS!
Si se desea se puede buscar un int confianza que C β.
/ $IC_{1-\alpha\%}: b \pm (t_{n-2;\,1-\frac{\alpha}{2}}) \cdot Sb)$

Analisis de la Varianza

¿Cuál es su enfoque?

El análisis de la varianza se basa en la partición de sumas de cuadrados y grados de lib. asociados con la variable respuesta Y.

- Variación de Y_i ⟶ desviaciones: $(Y_i - \bar{y}_i$
- Variación Total ⟶ $SC_{TOT} = \Sigma (y_i - \bar{y}_i)^2$

Relación:

$$\underset{T}{(Y_i - \bar{Y})} = \underset{R}{(\hat{Y}_i - \bar{Y})} + \underset{E}{(Y_i - \hat{Y}_i)}$$

T: desviación total

R: desv. del valor ajustado por la regresion con respecto a la media general

E: desv. de la observacion con respecto a la línea de regresión

↓

(*) Para que los desvios no se anulen se elevan al cuadrado y se trabaja algebraicamente...

↓

$$\underbrace{\Sigma (y_i - \bar{Y})^2}_{SC_{TOT}} = \underbrace{\Sigma (Y_i - \hat{Y})^2}_{SC\ er} + \underbrace{\Sigma (\hat{Y}_i - \bar{Y})^2}_{SC\ reg}$$

SC_T: Suma total de cuadrados

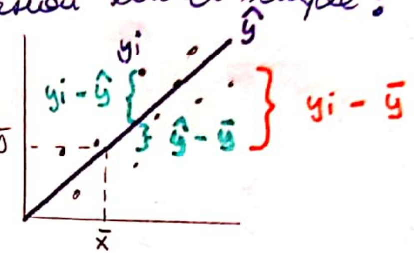
SC_r: Suma de cuadrados de la regresión

SC_e: Suma de cuadrados del error.

⟶ Si se divide por los g.l respectivos: $(n-1), (1), (n-2)$ a cada termino se obtienen los CUADRADOS MEDIOS DEL A. VARIAZA con dist Chi cuadrado.

↓

En regresion lineal simple:



(*) Cuadrados medios:

CM es el cuadrado medio del error o cuadrado medio residual. Es un estimador insesgado de σ^2.

$$CM_e = \frac{SC_e}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

OBS! Otras fórmulas:

- $SC_{reg} = b^2 \Sigma (x_i - \bar{x})^2$
- $SC_{er} = \Sigma (y_i - \bar{Y})^2 - b^2 \Sigma (x_i - \bar{x})^2$

Tabla de Analisis de la Varianza

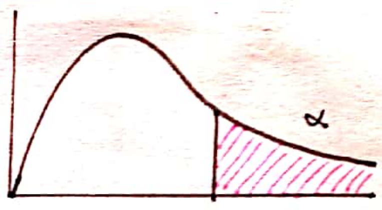
Se utiliza para confirmar o comprobar las hipotesis ...

I. $\begin{cases} H_0 = \beta = 0 & \nexists \text{ relacion funcion poblacional} \\ H_1 = \beta \neq 0 & \exists \text{ relación funcional poblacional} \end{cases}$

II. $\alpha = \ldots$

III. $F_{1;n-2} \cong \frac{CM\,regresión}{CM\,error}$

IV. Dibujo de la variable:
- Asimétrica positiva
- Cociente de varianzas
- Cociente de Chi-Cuadrado



$F_{(1;n-2)}$

Regla de decisión:
- Rechazo Ho si $F_{cal} \geq F_{tabla}$
- No rechazo Ho si $F_{cal} < F_{tabla}$

Fuentes de Variación	g.l	Suma de Cuadrados	Cuadrados medios	E(CM)	F
Debido a la regresión	1	$Sc_{reg} = b^2 \Sigma(x_i - \bar{x})^2$	$CM_{reg} = \frac{Sc_{reg}}{1}$	$\sigma^2_e + \beta^2 \Sigma(x - \bar{x})^2$	$\frac{CM_{reg}}{CM_e}$
Debido al error	$n-2$ → se estiman 2 parametros	Sc_{er}	$CM_{er} = Sc_{er}/n-2$	σ^2_e	"
Total	$n-1$	$Sc_{TOT} = \Sigma(y_i - \bar{y}_i)^2$	X	X	"

No se usa en la pract

Tabla F

La distribución F es una distribución continua que relaciona dos v. aleatorias ind con distrib. Chi Cuadrado, divididas por sus g.l.

Es descripta por los gl del num y denominador

- n_1 = gl numerador (V_1)
- n_2 = gl denominador (V_2)

$F = CM_{reg} / CM_e$

USO:
→ Buscar tabla según el α deseado!

	n_1 gl
n_2 gl	valor de F

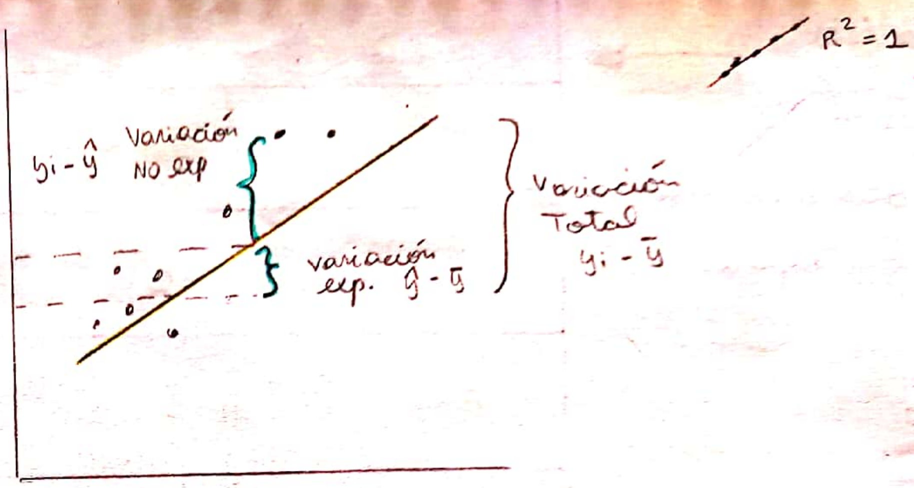
OBS!
Se llega a las mismas conclusiones usando dif distribució

Coeficiente de Determinación:

R^2 mide el porcentaje o parte de la variabilidad explicada por la regresión.

R^2 = Razón de la variación explicada a la variación total.

- $R^2 = 0$ entonces la variación total es toda NO explicada.
- $R^2 = 1$ entonces la variación total es toda explicada.
- R^2 varía de 0 a 1.
  No puede ser < 0.



OBS! Cuanto R^2 esté más cerca a 1 mayor se ajusta el modelo a la variable que estamos intentando explicar.

Def: El coef. de determinación ~~del~~ ~~ejemplo~~ que se ~~viene~~ desarrollado se define como la proporción de la varianza total de la variable explicada por la regresión.

OBS! R^2 indica el grado de ajuste de la recta de regresión a los valores de la muestra. Entre más se aproxima a 1 mejor será el ajuste.

ANOVA

"Análisis de la Varianza en diseño de experimentos"

Definiciones

• UNIDAD EXPERIMENTAL:
Parte más pequeña de material experimental a la que se le aplica el tratamiento y en donde se mide o registra la variable que se investiga.
Ej) experimento forestal → U.E = Árbol

• TRATAMIENTO:
Conjunto de procedimientos que se aplican a una U.E en un diseño elegido y cuyos efectos son registrados (respuesta).

• OBSERVACIÓN: Valor que asume la V. respuesta en una determinada realización.

• FACTOR:
Variable ind controlada por el experimentador. Se estudia su efecto sobre la V. dep. o respuesta
Símbolo: A, B, C...

• NIVEL DEL FACTOR:
Es cada una de las categorías, valores o formas específicas del Factor. Símbolo: a_n, b_n, c_n

Ej) • Factor A: Tipo de riego
Niveles: goteo (a_0); Aspersión (a_1); Secano (a_2)

DISEÑO ESTADÍSTICO EXPERIMENTAL (DEE)

Proceso de planificación del experimento mediante el cual se recolectan datos que al ser analizados por métodos estadísticos nos conducen a conclusiones válidas.

→ OBJ!

- Proporcionar métodos para obtener la MAYOR CANTIDAD DE INFO en forma objetiva, confiable y al mínimo costo.
- Controlar la variabilidad que esta presente en todo experimento y que afecta a los resultados,

→ UNIFACTORIAL: Un solo factor con sus respectivos niveles. (cada nivel del factor es un tratamiento). ← se usa este

→ MULTIFACTORIAL: Dos o más factores. Un tratamiento es la combinación de un nivel del factor.

ERROR EXPERIMENTAL

Que refleja?

- Variación del material exp
- errores de medición, observación, experimentación...

ANOVA 18–18:30
DOC 18:30 19:15
REG 19:30 20:30

Principios básicos del diseño experimental

✱ REPETICIÓN:
Es la reproducción del experimento básico, es decir, la reproducción de la asignación de un tratamiento a una U.E.
Son observaciones de un mismo tratamiento en diferentes U.E.

OBJ
- → Estimar el error exp.
- → Aumenta alcance de la inferencia del experimento.

✱ ALEATORIZACIÓN:
Consiste en la asignación al AZAR de los tratamientos a las U.E para que no se presente sesgo.

OBJ { Hace válidos los procesos de inferencia asegurando conclusiones confiables.

✱ CONTROL LOCAL:
Se basa en controlar el error experimental tomando medidas para hacer el DEE más eficiente reduciendo el error exp.

como evitar? →
- Incremento del nº de repeticiones en el exp.
- uso de un DEE apropiado
- uso de material Homogéneo
- Perfeccionamiento de técnicas en el exp.

DDE: Existen diferentes tipos de diseños experimentales, durante la práctica se trabaja con DCA.

DISEÑO COMPLETAMENTE ALEATORIZADO: Los tratamientos en estudios se distribuyen al azar en forma irrestricta sobre todas las U.E, siendo el nº de repeticiones por tratamiento igual o diferente

Ventajas:
- Fácil de planear y analizar
- Flexible en el empleo del nº de rep y tratamiento
- Permite trabajar con el máx nº de g.l para la Σ de errores

Desventajas - Uso de material experimental no homogéneo

Consta de ... $\boxed{\text{GRUPOS}} \xrightarrow[\text{a c/u}]{\text{se aplican}} \boxed{\text{TRATAMIENTO}} \xrightarrow[\text{por repetición}]{\text{se obtienen}} \boxed{\text{OBSERVACIONES}}$

MODELO QUE UTILIZA: $Y_{ij} = \mu_i + \varepsilon_{ij}$ con $i = 1 \dots t$; $j = 1 \dots r$

- Y_{ij} = es el valor observado de la j-ésima OBSERVACIÓN del i-ésimo TRATAMIENTO.
- μ_i = Valor medio poblacional de i-ésimo TRATAMIENTO.
- ε_{ij} = Es el error aleatorio de la j-e OBSERVACIÓN del i-e TRATAMIENTO.
- τ_i = Es el efecto medio poblacional del i-ésimo TRAT.
- μ = Valor medio poblacional gral.

Si se reemplaza $\mu_i = \mu + \tau_i$

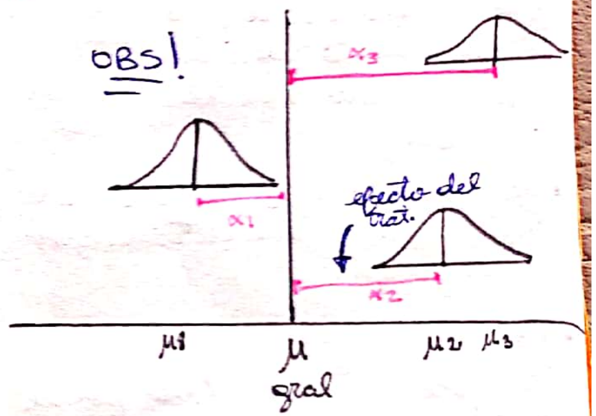
$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad \text{con } i = 1 \dots t \text{ ; } j = 1 \dots r$$

SUPUESTOS DEL MODELO:

- Y_{ij} sigue una dist Normal para cada $T_i \rightarrow y_{ij} \sim N(\mu_i ; \sigma_i^2)$
- Y_{ij} son v. ind $\therefore$ $cov(Y_{ij} ; Y_{kl}) = 0$
- La varianza de los tratamientos son iguales $\sigma_1^2 = \sigma_2^2 = \sigma_t^2 = \sigma^2$

$\left\{ \begin{array}{l} \cdot\ \varepsilon_{ij} \sim N(0 ; \sigma^2) \\ \cdot\ E(Y_{ij}) = \mu_i \\ \cdot\ Var(Y_{ij}) = \sigma^2 \end{array} \right.$

OBS!



El efecto del tratamiento corre o distancia la distribución de la μ gral.

$\hookrightarrow$ si hay efecto $\rightarrow$ distrib separada de μ

$\hookrightarrow$ si NO hay efecto $\rightarrow$ distrib sobre μ (Ej) $\mu_1 = \mu$

Partición de la suma de cuadrados:

$$\underbrace{\sum_i \sum_j (y_{ij} - \bar{\bar{y}})^2}_{SC_{TOT}} = \underbrace{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2}_{SC_{ee}} + \underbrace{\sum_i \sum_j (\bar{y}_i - \bar{y})^2}_{SC_{TRAT}}$$

$$SC_{TOT} = SC_{ee} + SC_{TRAT}$$

$$\underbrace{r \cdot t}_{n} - 1 \qquad t(r-1) \qquad (t-1)$$

Cuadrados medios:

$$CM_{error} = \frac{SC_{error}}{t(r-1)} = \hat{\sigma}^2 \qquad CM_{TRAT} = \frac{SC_{TRAT}}{t-1}$$

Los CM son varianzas muestrales estimadores de varianzas poblacionales (parametros).

El cociente entre estos sigue una distrib F.

$$F = \frac{CM_{trat}}{CM_{ee}} \approx F(t-1;\ t(r-1))$$

TABLA ANOVA:

Fuent variación	suma de cuadrados	g.l	Cuadrados medios	F
Entre Tratamient	SC_{TRAT}	$t-1$	$\frac{SC_{trat}}{t-1}$ (I)	$\frac{I}{II}$
Dentro tratamiento (EE)	SC_{ERROR}	$t(r-1)$	$\frac{SC_{error}}{t(r-1)}$ (II)	
Total	SC_{TOTAL}	$tr-1$	–	

- HIPOTESIS EN TERMINOS DE LOS EFECTOS:

$$\begin{cases} H_0: \tau_i = 0 \\ H_1: \tau \neq 0 \end{cases}$$

. Todos los trat tienen el mismo efecto sobre la v. rpta.
. No todos " " " ""

- DOCIMAS EN TERMINOS DE LOS PROMEDIOS:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_t \\ H_1: \mu_i \neq \mu \end{cases}$$