Cálculo de probabilidad.

1) Experimentos y fenomenos aleatorios. son aquellos que al ser repetidos en condiciones uniformes presentan resultados variables de manera que no puede predecirse con exactitud el resultado de una repetición individual.

2) Se llama espacio muestral al conjunto formado por todos los posibles resultados imaginables (o puntos muestra) de un experimento aleatorio.

3) Se denomina suceso o evento a todo subconjunto del espacio muestral.

4) Denominamos suceso contrario a todos los puntos muestras que no pertenecen al mismo, se simboliza con la misma letra pero raya sobre él.

Ejemplo "experimento aleatorio"

elegir dos comprimidos al azar de una producción y verificar si la cantidad de droga es la especifica o no

$S = \{EE; EN; NE; NN\}$.

5) Sucesos → seguro = suceso tal que al realizar el experimento siempre debe presentarse.
→ imposible = suceso tal que al realizar el experimento jamás ~~debe~~ puede presentarse.

6) Operaciones entre sucesos.

- Unión = de los k sucesos será un suceso formado por todos los puntos muestras que pertenecen al menos a uno.
- Intersección = de los k sucesos será un suceso formado por todos los puntos muestras que pertenecen a todos simultáneamente.
- sucesos mutuamente excluyentes. cuando no pueden ocurrir simultáneamente. (su intersección es nula).

Ejemplo= elegir um comprimido de uma producción y determimar la camtidad de droga

$S = \{x / 0 \leq x \leq a\}$

$C = \{x / 50 \leq x \leq a\}$

$F = \{x / 10 \leq x \leq 80\}$

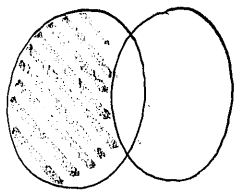
$G = \{x / 10 \leq x \leq 40\}$

$C \cup F = \cancel{\{x / 10 \leq x \leq 50\} \cup \{x / 80 \leq x \leq a\}}$ $\{x / 10 \leq x \leq a\}$

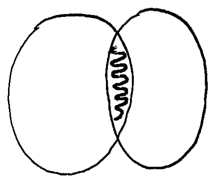
$C \cap F = \{x / 50 \leq x \leq a\} \rightarrow \{x / 50 \leq x \leq 80\}$

$C \cup \bar{C} = \{x / 0 \leq x \leq a\} = S$

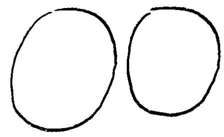
$\underbrace{C \cap G}_{\text{sucesos mutuamente excluyentes}} = \varnothing$



umiom de dos sucesos



imtersección de dos sucesos



sucesos mutuamemte excluyemtes

*) Probabilidad= medida de la posibilidad que ocurra o presemte determimado suceso em um experimemto aleatorio.

* propiedades de probabilidad.

$P(A) \geq 0.$

$P(S) = 1.$

$$\boxed{\text{Entonces}, \ 0 \leq P(A) \leq 1.}$$

(*) La probabilidad de um suceso seguro es igual a 1, ahora que la probabilidad de um suceso sea 1 mo mos garamtiza que el mismo sea seguro.

(*) La probabilidad de um suceso imposible es igual a 0, ahora que la probabilidad de um suceso sea 0 mo mos garamtiza que el mismo sea imposible.

Definiciones de probabilidad → clásica 1)
↓
Frecuencial 2)

1) **Clásica**

$$P(A) = \frac{\text{n° de elementos del espacio muestral favorables a "A"}}{\text{n° total de elementos en el espacio muestral}}$$

*** requerimientos**; espacio muestral "S" finito. (variable cuantitativa discreta).
y todos los elementos sean equiprobables.

2) **Frecuencial** = sea E un experimento aleatorio, A un suceso asociado con él y n el n° de repeticiones de E.

- Frecuencia absoluta $f(A)$, es igual al n° de veces que se presentó el suceso "A" en el experimento aleatorio
- Frecuencia relativa $h(A)$, es igual al cociente entre el n° de veces que se repitió el suceso A en el experimento (frecuencia absoluta) y "n" n° total de repeticiones de E.

Ejemplo = elegir un individuo y determinar su grupo sanguíneo.

S = {grupo A; grupo B; grupo AB; grupo O}.

suceso A = que el individuo sea de g. A.

n = 10 veces $f(A) = p$ "veces que se presentó el suceso "A"".

- ¿Cómo cambia la frecuencia relativa si se modifica la cantidad de veces que repetimos el experimento? } Observamos que $h(A)$ se aproxima a un valor al aumentar el n° de repeticiones.

* Concepto frecuencial o a posteriori define la probabilidad de un suceso A, $P(A)$, es el nº que tienden las frecuencias relativas cuando el número de repeticiones tiende a infinito.

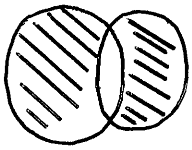
$$\boxed{P(A) = \lim_{n \to \infty} h(A)}$$

* la probabilidad de un suceso es entonces una idealización de la frecuencia relativa

* la frecuencia relativa es una estimación de la probabilidad.

* <u>Regla de la suma</u>

* nos permite calcular la probabilidad de la unión de dos sucesos.



definimos $\mathcal{E}$ experimento aleatorio, A, B y C sucesos asociados a él

√ para dos sucesos =

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

* demostración

√ para tres sucesos =

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

√ caso particular = si A y B son mutuamente excluyentes $(A \cap B) = \emptyset$

$$P(A \cap B) = P(A) + P(B).$$

deducimos que =

$$P(A) = 1 - P(\bar{A}).$$

demostración =

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = P(S) = 1$$

$$\Rightarrow P(A) + P(\bar{A}) = 1$$

$$P(A) = 1 - P(\bar{A}).$$

* demostración regla de la suma para dos sucesos.

$$f(a \cup b) = f(a \cup b) + \underbrace{f(\bar{a} \cup b)} + f(a \cup \bar{b})$$

$$= \cancel{f(a \cup b)} + f(a) - \cancel{f(a \cap b)} + f(b) - f(a \cap b)$$

$$= f(a) + f(b) - f(a \cap b).$$

$$\frac{f(a \cup b)}{n} = \frac{f(a) + f(b) - f(a \cap b)}{n}$$

$$h(a \cup b) = h(a) + h(b) - h(a \cap b) \Rightarrow \boxed{P(a \cup b) = P(a) + P(b) - P(a \cap b)}$$

si A y B son mutuamente excluyentes

$$P(A \cap B) = 0$$

* **Probabilidad condicional** = la probabilidad de un suceso A condicionado a la ocurrencia de un suceso "b" se define como =

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \; ; \; P(B) \neq 0.$$ demostración.

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

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

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

* sucesos independientes = dos sucesos son independientes si la probabilidad de uno no cambia habiéndose presentado o no el otro suceso.

$$P(A/B) = P(A/\bar{B}) = P(A)$$
$$P(B/A) = P(B/\bar{A}) = P(B)$$

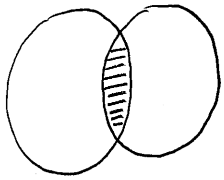
las probabilidades son iguales habiéndose presentado o no el otro suceso.

$$\Downarrow$$

$$P(A \cap B) = P(A/B) . P(B)$$ *1

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

Regla del producto = nos permite calcular la intersección de dos o más sucesos



recordando =

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

$$P(B) \neq 0$$

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

$$P(A) \neq 0$$

* sucesos independientes 1

"Estadísticas pensadas como variables aleatorias".

Ejemplo = un experimento consiste en elegir $n = 5$ alícuotas de agua de río y medir la cc. de arsénico

$x_1; x_2, \ldots, x_5$ } medición cc. de Arsénico en las cinco alícuotas.

- Al realizar la descripción del conjunto de datos, se obtiene un valor numérico de cada una de las estadísticas.

$\bar{X}$ y S^2

S^2: varianza muestral

$\bar{X}$: media aritmética

Si repetiríamos el mismo experimento obtendríamos diferentes resultados de la media y el desvío estándar ya que las mediciones en la cc. de As variarían por estar sujetas a los errores aleatorios.

Las estadísticas varían al variar la muestra $\Longrightarrow$ son variables aleatorias pensadas en el muestreo reiterado de n valores de la variable "x".

¿Cómo definimos una variable aleatoria desde el punto de vista matemático? $\longrightarrow$ dada una variable aleatoria x con distribución de probabilidad $f(x)$, consideremos que el experimento consiste en elegir al azar "n" valores de x.

El espacio muestral correspondiente es:

$$S = \{(x_1, x_2, \ldots, x_n) / x_i \in \mathbb{R} \; i = 1, \ldots, n\}$$

son tomadas de una muestra aleatoria.

Una estadística definida sobre este espacio muestral es una variable aleatoria ya que asigna a cada punto muestral un número real.

¿Cómo definimos muestra aleatoria? al realizar el experimento obtendremos un conjunto de "n" valores.

Sin embargo antes de realizar el experimento hay una incertidumbre sobre cada x_i, es decir, la i-ésima observación. A cada una de estas se las considera como una variable aleatoria, con igual distribución que la variable x.

El conjunto de variables aleatorias independientes $(x_1, x_2, \ldots, x_n)$ c/u con la misma distribución que la variable aleatoria x original y mismos parámetros, recibe el nombre de muestra aleatoria.

Entonces una estadística definida sobre una muestra aleatoria es una variable aleatoria.

Ya que las estadísticas son variables aleatorias, interesa conocer su distribución de probabilidad que se denota "distribución en el muestreo"

distribución muestreo = distribución de probabilidad de una estadística

1) distribución en el muestreo de la media aritmética

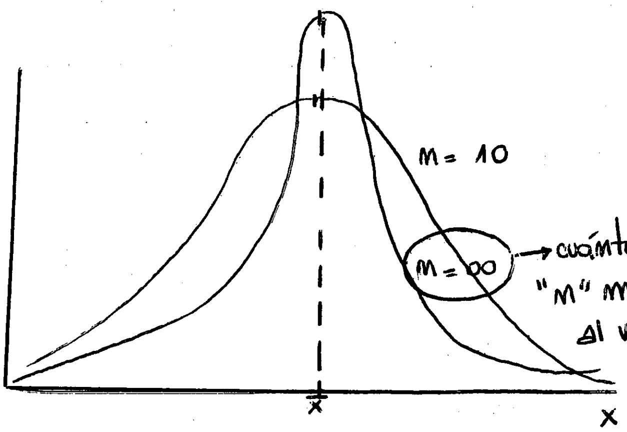
X = variable aleatoria cuyo parámetros son u_x y σ_x^2

definimos $(x_1, x_2, \ldots, x_n)$ como una muestra aleatoria de la variable x

$$\bar{X} = \frac{1}{n} \cdot (x_1, x_2, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$E(\bar{x}) = u_x \qquad V(\bar{x}) = \frac{\sigma_x^2}{n} \quad ó \quad V(\bar{x}) = \frac{\sigma_x}{\sqrt{n}}$$

Si la variable aleatoria x es normal (se prueba con gráfico de probabilidad normal), entonces, la media aritmética tiene distribución normal.



→ cuánto más grande es el "n" mas nos acercamos al verdadero valor promedio.

Teorema del límite central.

La variable aleatoria media aritmética de una muestra aleatoria de extensión "n", suficientemente grande, proveniente de una población con promedio u_x y desvío estándar σ_x se distribuye normalmente con promedio u_x y desvío $\frac{\sigma_x}{\sqrt{n}}$

$$\bar{x} \longrightarrow N\left(u_x \,;\, \frac{\sigma_x}{\sqrt{n}}\right).$$

$$E(S_x^2) = E\left(\frac{\chi^2(n-1)\cdot\sigma_x^2}{n-1}\right) = \frac{\sigma_x^2}{n-1}\cdot E(\chi^2(n-1)) = \frac{\sigma_x^2}{\cancel{n-1}}\cdot \cancel{n-1} = \sigma_x^2$$

$$\boxed{E(S_x^2) = \sigma_x^2.}$$

multiplicas y div.

$$V(S_x^2) = \sigma_x^2(S_x^2) = \sigma_x^2\left(\frac{\chi_{n-1}\cdot\sigma_x^2}{n-1}\right) = \left(\frac{\sigma_x^2}{n-1}\right)^2 \cdot \underbrace{\sigma_x^2(\chi^2(n-1))}_{V(\chi^2(n-1))}$$

$$= \frac{\sigma_x^4}{(n-1)^2}\cdot 2(n-1) = \frac{2\sigma_x^4}{(n-1)}$$

$$\boxed{V(S_x^2) = \frac{2\sigma_x^4}{n-1}}$$

→ con variancia desconocida

$$t_K = \frac{Z^*}{\sqrt{\frac{\chi^2_K}{K}}} = \frac{\frac{\bar{X}-\mu_x}{\sigma_x/\sqrt{n}}}{\sqrt{\frac{\frac{\cancel{(n-1)}\cdot S^2_x}{\sigma_x^2}}{\cancel{(n-1)}}}} = \frac{\frac{\bar{X}-\mu_x}{\cancel{\sigma_x}/\sqrt{n}}}{\frac{S_x}{\cancel{\sigma_x}}} = \frac{\bar{X}-\mu_x}{S_x/\sqrt{n}} = t_{n-1}$$

→ cociente de variancias. → consultar.

$X_1 \to N(\mu_1; \sigma_1^2)$ $\quad X_2 \to N(\mu_2; \sigma_2^2)$

si $\sigma_1^2 = \sigma_2^2 \Longrightarrow$

$$\frac{S_1^2}{S_2^2} \overset{si}{=} \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\frac{\cancel{(n_1-1)}\cdot S_1^2}{\cancel{(n_1-1)}\cdot\sigma_1^2}}{\frac{(n_2-1)\cdot S_2^2}{\cancel{(n_2-1)}}} = \frac{\chi^2_{n_1-1}/n_1-1}{\chi^2_{n_2-1}/n_2-1} = F_{n_1-1;\, n_2-1}$$

* cuánto mayor diferencia exista entre la distribución de x y la distribución normal, mayor n se requiere para que la media aritmética se distribuya normal. En general p' un $n > 30$ la aproximación normal es válida.

2) distribución en el muestreo de la frecuencia relativa. ("caso particular de la media aritmetica").

experimento aleatorio

$E ; \quad S = \{A ; \bar{A}\} \quad X(A) = 1 \quad X(\bar{A}) = 0$

$P(A) = P(X=1) = p \qquad P(\bar{A}) = P(X=0) = 1 - p = q$

$$E(X) = \mu_X = \sum_{j=1}^{2} x_j \cdot p(x_j) = p$$

$$V(X) = \sigma_X^2 = \sum_{j=1}^{2} (x_j - \bar{X})^2 \cdot p(x_j) = p \cdot q$$

$$\bar{X} = \frac{0+1+0+0+0+1+\ldots+0}{n} = \frac{f(A)}{n} = h(A)$$ $\Rightarrow$ frecuencia relativa es una media aritmetica.

$$h(A) \xrightarrow{n > 30} N\left(E[h(A)] = p ; \underbrace{V[h(A)]}_{\sigma^2} = \frac{p \cdot q}{n}\right).$$

x' T.L.C

condición porque no tiene distribución normal.

si $\sigma = \frac{p \cdot q}{\sqrt{n}}$

3) Distribución de la variancia muestral =

Sí $x \longrightarrow N(u_x ; \sigma^2_x)$

Se $(x_1 ; x_2 ; \ldots ; x_n)$ muestra aleatoria de la variable x.

$$S^2_x = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$ → consultar demostración

$$S^2_x \longrightarrow \frac{\chi^2_{n-1} \cdot \sigma^2_x}{n-1}.$$

→ variancia muestral

$x_1 \to N \to Z_1$

$x_2 \to N \to Z_2$

$x_3 \to N \to Z_3$

$$\sum_{i=1}^{n} Z_i^2 = \chi_n^2 = \sum_{i=1}^{n} \left(\frac{x-u}{\sigma}\right)^2 \overset{\uparrow \text{ sumo y resto } \bar{x}}{=} \underbrace{\sum_{i=1}^{n} \left(\frac{x-u+\bar{x}-\bar{x}}{\sigma}\right)^2}$$

$$= \sum_{i=1}^{n} \left(\frac{x_i-\bar{x}}{\sigma}\right)^2 + \underbrace{\sum_{i=1}^{n} \left(\frac{\bar{x}-u}{\sigma}\right)^2} + 2\left(\frac{\bar{x}-u}{\sigma}\right) \cdot \underbrace{\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)}_{=0}$$

↓ binomio cuadrado

$$n \cdot \left(\frac{\bar{x}-u}{\sigma}\right)^2 = \underbrace{\left(\frac{\bar{x}-u}{\sigma/\sqrt{n}}\right)^2}_{Z^2} = \chi^2_1$$

$$= \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{\sigma^2} + \chi^2_1 \Longrightarrow \chi^2_n - \chi^2_1 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{\sigma^2}$$

$$\chi^2_{n-1} = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{\sigma^2}$$

$$\chi^2(n-1) = \frac{(n-1)}{(n-1)} \cdot \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{\sigma^2} = \frac{(n-1) \cdot S^2_x}{\sigma^2} \Longrightarrow$$

$$S^2_x = \frac{\chi^2(n-1) \cdot \sigma^2}{(n-1)} > 0.$$

Distribución de probabilidad = unidad n° 3.

1) Variable aleatoria =

Sea E un experimento aleatorio y S su espacio muestral.

Llamamos variable aleatoria X a una función que asigna a cada uno de los elementos

$s \in S$ un n° real

espacio muestral constituido por "s" elementos.

$$\boxed{X : S \longrightarrow \mathbb{R}}$$

variable aleatoria

Ejemplo = elegir dos comprimidos de la población y verificar si la cantidad de droga es la especificada o no

E = especificada

N = no especificada.

$$S = \{EE;\ EN;\ NE;\ NN\}$$

X = n° de comprimidos en donde la droga posee la cantidad especificada

→ consultar el otro ejemplo

$$X(EE) = 2 \qquad X(EN) = 1 \qquad X(NN) = 0$$

función que le otorga a cada punto muestral del "S" un único n° real.

Variables aleatorias discretas y continuas.

- discreta = puede asumir un n° finito de valores
- continua = puede asumir un n° infinito de valores.

(*) ejemplo = Elegir un comprimido de la producción y verificar que la cantidad de droga es la especifica

√ $S = \{E;\ N\}$

$X(E) = 1 \quad X(N) = 0.$

(*) ejemplo = elegir un comprimido de la producción y determinar si la cantidad de droga es la especificada.

√ $S = \{w \,/\, 0 \leq w \leq a\}$ $\qquad X(w) = w \quad \underbrace{x \in [0;a]}_{\text{campo de variación}}$

√ p/ dos comprimidos

$S = \{w_1; w_2 \,/\, 0 \leq w \leq a\} \quad X(w_1; w_2) = \dfrac{w_1 + w_2}{2} \qquad x \in [0;a]$

- **Función de probabilidad.**

x : variable aleatoria discreta

* variables aleatorias continuas $\nearrow$ discreta $\longrightarrow$ Función probabilidad; $\searrow$ continua $\longrightarrow$ Función densidad de probabilidad

$x = x_1, x_2, x_3, \ldots, x_j, \ldots, x_m, \ldots$

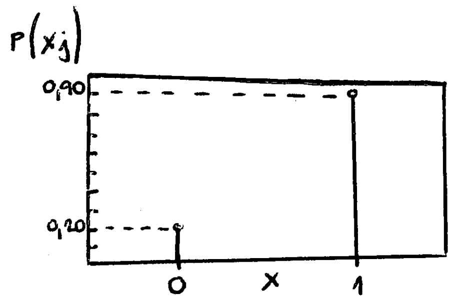
A cada valor posible x_j se le asocia un nº $p(x_j) = P(x = x_j)$, siendo $P(x = x_j)$ la probabilidad de x_j.

Estos nº $p(x_j)$ deben satisfacer :

$$p(x_j) \geq 0 \text{ para todo } j \quad y \quad \sum_{j=1}^{n(\infty)} p(x_j) = 1 \xrightarrow[\text{justificado}]{} P(S) = 1.$$

Ejemplo = elegir 1 comprimido de una producción y verificar si la cantidad de droga es la especificada o no.

$$\left.\begin{array}{l} P(x=0) = 0{,}10 \\ P(x=1) = 0{,}90 \end{array}\right\} \text{ cumpliendo propiedades}$$



- **Función densidad de probabilidad.**

x = variable aleatoria continua

Llamaremos función de densidad de probabilidad de x a una función $f(x)$ que verifica que =

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

y que satisface las siguientes propiedades =

a) $f(x) \geq 0 \ \forall x$.

b) $\int_{-\infty}^{+\infty} f(x) \cdot dx = 1$

Población vs. Muestra

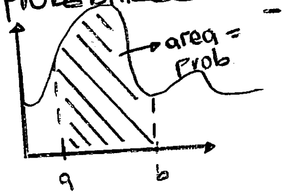
Población

- conjunto total de unidades en lo que se desea realizar una inferencia.
- conjunto de infinitos resultados que podríamos obtener al efectuar en forma hipotética infinitas mediciones de una variable.

⇓

- distribución probabilidad

- variable aleatoria discreta = función de probabilidad → gráfico de bastones
- variable aleatoria continua = función de densidad de probabilidad → área debajo de la curva



⇓

Parámetros = medidas que caracterizan una población

- $E(x) = \mu$ ⟶ medida de posición
- $V(x) = \sigma_x^2$ ⟶ medida de dispersión

* Propiedades de la esperanza y la varianza. (continua).

Muestra

- subconjunto de la población que es seleccionada aleatoriamente con el objetivo de que sea representativa de la misma y permita efectuar inferencias sobre ella.

⇓

- distribuciones de frecuencias

- variables aleatorias discretas → gráfico de bastones
- variables " continuas → histograma de frecuencias.

⇓

Estadísticas = medidas que caracterizan una muestra.

- $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ ⟶ de posición
- $S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ ⟶ de dispersión

Variable aleatoria → continua →

$$E(x) = \sum_{j=1}^{n(\infty)} x_j \cdot p(x_j) = u$$

$$V(x) = \sum_{j=1}^{n(\infty)} (x_j - u)^2 \cdot p(x_j) = \sigma_x^2$$

Variable aleatoria → discreta →

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

$$V(x) = \int_{-\infty}^{+\infty} (x-u)^2 \cdot f(x) \cdot dx = \sigma_x^2$$

* sea x una variable aleatoria (continua o discreta) y K constante.

Propiedades de la variancia.

- $V(K) = 0$
- $V(K \cdot x) = K^2 \cdot V(x)$
- $V(K + x) = V(x)$; $V(K) = 0$

Propiedades de la esperanza.

$E(K) = K$

$E(K \cdot x) = K \cdot E(x)$

$E(K + x) = K + E(x)$

sean $x_1, x_2, x_3, \dots, x_j, \dots, x_n$

$$E(x_1 + x_2 + x_3 + \dots + x_j + \dots; x_n) = E(x_1) + E(x_2) + E(x_3) \dots + E(x_n).$$

sean $x_1, x_2, x_3, \dots, x_j, \dots, x_n$ independientes

$$V(x_1 + x_2 + x_3 + \dots + x_n + \dots + x_j) = V(x_1) + V(x_2) + V(x_3) + \dots + V(x_n)$$

Distribución de probabilidad.
1) distribución de Poisson
2) distribución normal
3) distribución Chi-Cuadrado
4) distribución T-Student.
5) distribución F-Snedecor.

1) distribución de Poisson. = dada una gran cantidad de elementos distribuidos aleatoriamente y uniformemente en un gran volumen, superficie, espacio o período de tiempo, sea:

→ se aplica cuando tenemos un gran nº de elemen

• E = elegir una pequeña porción y contar nº elementos.

$S = \{0, 1, 2, \ldots\}$

x = nº elementos de la pequeño porción

$X(s) = s$ ⇝ función identidad

$X = 0, 1, 2, \ldots, x_j, \ldots$

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

$E(x) = V(x) = \lambda.$

↳ no se puede aplicar poissons si no se cumple que $E(x) = V(x)$ ó equivalentemente $\mu_x = \sigma_x^2$

Ejemplo x = nº árboles enfermos por hectarea en un bosque de pinos.

λ = 10 árboles enfermos / hectarea → dato específico

2) distribución normal =

x = variable aleatoria continua $(-\infty < x < +\infty)$

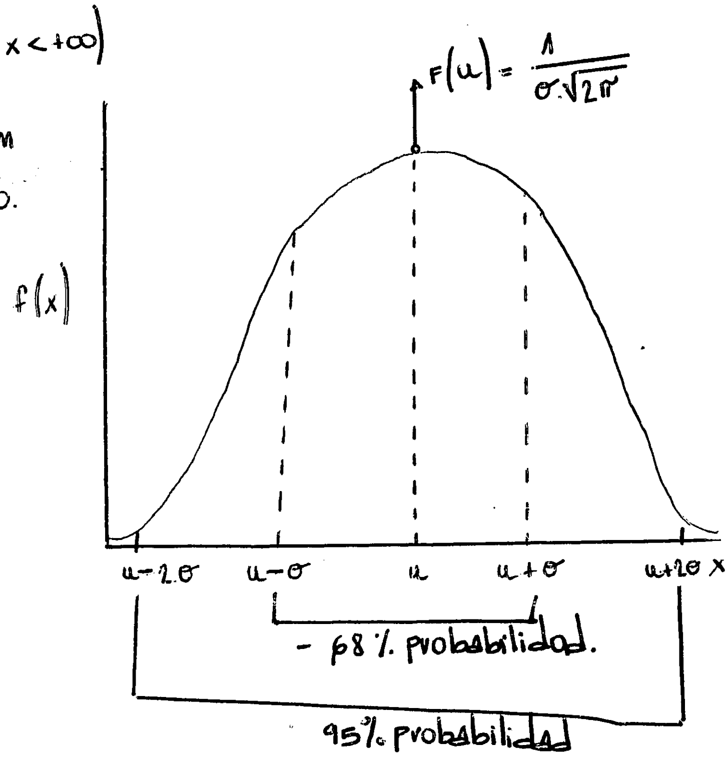
• $P(x \le \mu) = P(x > \mu) = 0{,}50$ simetrica con respecto al promedio.

• parametros μ y σ

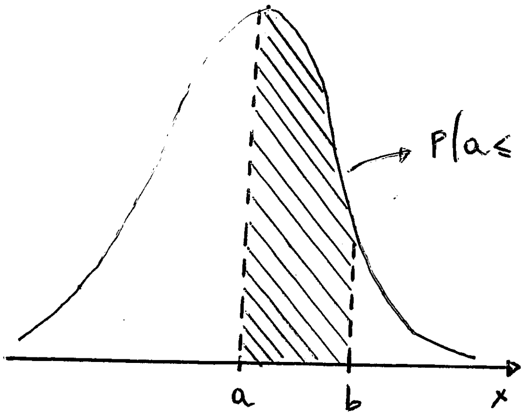
• forma campanular

- puntos de inflexión $\sigma \pm \mu$.

- asintotica con respecto al eje x tendiendo $\pm\infty$.



entre $\mu \pm 3\sigma$ se acumula 99,7 % probabilidad.



$$P(a \leq x \leq b) = \int_a^b f(x)\, dx = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-u}{\sigma}\right)^2} \cdot dx$$

↓

no tiene primitiva

entonces se recurre

variable normal estándar

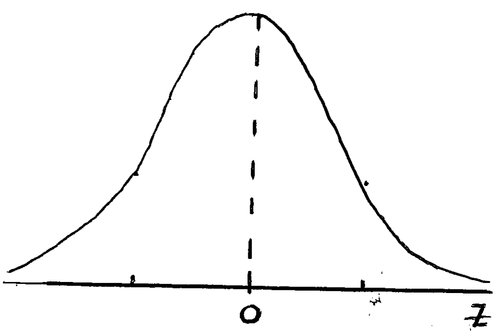
proceso de estandarización → (se transforma a la variable normal x en una variable normal estandarizada z).

$$z \longrightarrow N(0;1) \xrightarrow[\text{si}]{} x \longrightarrow N(u_x; \sigma_x^2).$$

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot\left(\frac{x-u}{\sigma}\right)^2}$$

$$= \underbrace{\varphi(z)}$$

áreas debajo de la curva de densidades de probabilidad están tabuladas.



* estandarización

$$x \longrightarrow N(u;\sigma) \Longrightarrow z = \frac{x-u}{\sigma} \longrightarrow \underbrace{N(0;1)}_{*}$$

$$P(x \leq a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot\left(\frac{x-u}{\sigma}\right)^2} \cdot dx$$

$$= \int_{-\infty}^{\frac{a-u}{\sigma}} \frac{1}{\sigma\cdot\sqrt{2\pi}} \cdot e^{-1/2\cdot(z)^2} \cdot dz$$

$$= P\left(z \leq \frac{a-u}{\sigma}\right)$$

$$E(z) = E\left(\frac{x-u_x}{\sigma_x}\right) = \frac{1}{\sigma_x} \cdot E(x-u_x) = \frac{1}{\sigma_x} \cdot (\overbrace{E(x)}^{u_x} - u_x) = 0$$

$$\sigma_x^2(z) = \sigma^2\left(\frac{x-u_x}{\sigma_x}\right) = \sigma^2\Big[\underbrace{\frac{1}{\sigma_x}}_{K} \cdot \underbrace{(x-u_x)}_{X}\Big] = \left(\frac{1}{\sigma_x}\right)^2 \cdot \sigma^2 \cdot (x-u_x) = \frac{1}{\sigma_x^2} \cdot \sigma_x^2 = 1$$

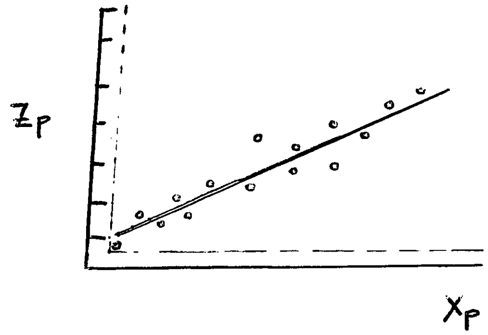
x' prop. de variancia.

* Prueba grafica de normalidad.

$$X \longrightarrow N(u_x ; \sigma_x) \qquad Z = \frac{X - u_x}{\sigma_x}$$

$\Longrightarrow$ deben verificarse una relación lineal entre los percentiles Z_P y X_P.

?



* Distribución chi-Cuadrado =

$Z_1, Z_2, Z_3, \ldots, Z_K$; "K" variables aleatorias normales estandarizadas independientes

$\longrightarrow$ grados de libertad.

$$Z_1^2 + Z_2^2 + Z_3^2 + \ldots + Z_4^2 = \sum_{i=1}^{K} Z_i^2 = \chi^2_K$$

$$E(\chi^2_K) = K$$

$$V(\chi^2_K) = 2 . K$$

* A medida que $K \uparrow$ la distribución se hace más simetrica y más parecida a la normal

⊛ $\chi^2_K \geqslant 0$ siempre

* distribución asimetrica a la derecha.

4) distribución T-studem

Seam z y v dos variables independientes tal que =

Z tiene distribución mormal estandarizada = $Z \longrightarrow N(0;1)$. $-\infty < Z < +\infty$.

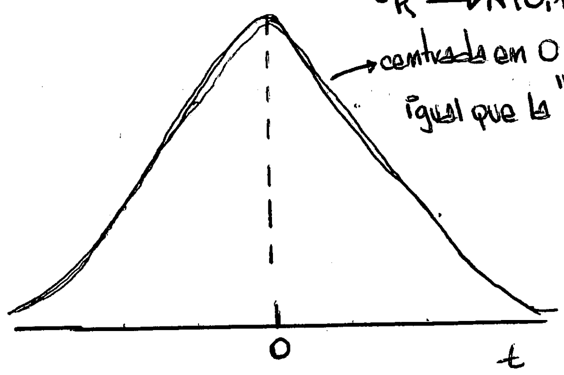
V " " Chi-cuadrado = $V \longrightarrow \chi^2_K$ $V > 0$.

Entonces, la variable aleatoria "T" tiene distribución T-student =

$$t = \frac{Z}{+\sqrt{\frac{V}{K}}} \longrightarrow t_K \qquad -\infty < t < +\infty$$

*forma campanular, alcanza el máximo en 0.
*cuando K↑ la distribución se cc. más en 0
*cuando $K \to \infty$ el gráfico $t_K \longrightarrow N(0,1)$. tiende a la mormal
→centrada en 0 igual que la "Z"

$E(t) = 0$.



5) distribución F-Smedecor =

χ^2_K
χ^2_M > independientes.

$$\underbrace{F_{K.M}}_{*} = \frac{\chi^2_K / K}{\chi^2_M / M}$$

mumerador
demominador

*tiene 2 grados de libertad debido a las dos χ^2

*el grafico parte de (0;0) asímetrico x' derecha.

Error → son aquellas causas que hacen que una medición difiera de su valor real y en que cantidad difiere el valor medido del valor verdadero.

*errores groseros = grandes y conducen a resultados discordantes (potenciales outliers). son faciles de detectar.

*errores sistemáticos = misma magnitud y dirección que las mediciones efectuadas. Se presentan de forma constante en mediciones repetidas bajo condiciones uniformes. Tendencia a sobrevalorar o subvalorar el valor verdadero.

*errores aleatorios = causado por condiciones no controlables que son parte inevitable de toda medición. Su efecto es mínimo. son variables aleatorias.

1) Estadística descriptiva

Objetivo = resumir información contenida en un conjunto de datos.

$\Big\downarrow$ mediante

- gráficos
- tablas
- "estadísticas" = características numéricas.

Definiciones

1) unidad = es el ítem sujeto a observación

ejemplo = alícuota de una solución.

2) variable = característica que puede variar de observación en observación.

ejemplo = concentración, el pH, edad, sexo, peso

3) Conjunto de datos = todas las observaciones registradas de una o más variables conforman un c. de datos.

Variables →

cuantitativas tomam valores numéricos para los cuales tiene sentido trabajar con características numéricas. → *

cualitativas = también llamadas categóricas, no son medibles numéricamentes; cuando la variable resulta de este tipo cada unidad observada resulta incluida en una de dos o más categorías exhaustivas y mutuamente excluyente.

exhaustivas = total

* →

a) discreta = la variable solo toma valores aislados

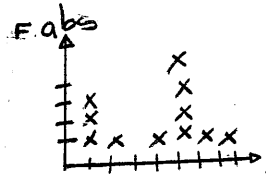
b) continua = la variable toma cualquier n° de un interva-

lo.

amalizamos gráficamente ← * Descripción conjunto de datos, (depende el tipo variable).

Cuantitativas ↔ Cualitativa

a) diagrama de puntos.= permite en forma rápida visualizar el conjunto de datos distribuido en una escala real.



- Valor de la variable eje de las abcisas o eje "x"
(si un valor se repite se marcaría alineadamente en forma vertical).
- Frecuencias absolutas en el eje "y" o eje de ordenadas.

b) diagrama de tallo y hoja = apropiado cuando el nº dato es relativamente pequeño.

Construcción → cada observación está separada o representada en una hoja.

→ generalmente la hoja constituye el último digito de la observación, el tallo el resto.

→ los tallos se ordenan de menor a mayor. en forma vertical

⇝ luego las hojas se unen al tallo alejandose de forma creciente. Siempre aclaramos unidad de la hoja

distribución altura de la planta

20	8
21	245
22	156789
23	∘∘∘
24	∘∘∘
25	
26	

x = Altura de la planta (m)

unidad hoja 0,01 m.

Ejemplo = 20|8 = 2,08 m.

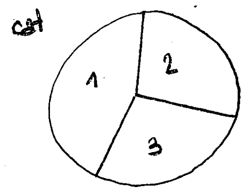
a) gráfico sectores = diagramamos un circulo que representa el 100 % unidades.
El mismo se divide en tanto sectores como categorías existan.
El área de cada sector es proporcional al porcentaje de unidades que pertenecen a esa categoría.

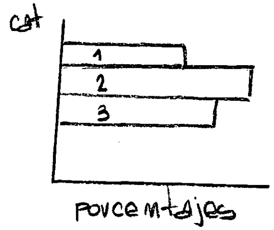
b) gráfico barras = se representa una barra para cada categoría. La longitud de la barra es proporcional al porcentaje de unidades de esa categoría. el ancho es igual para todas.

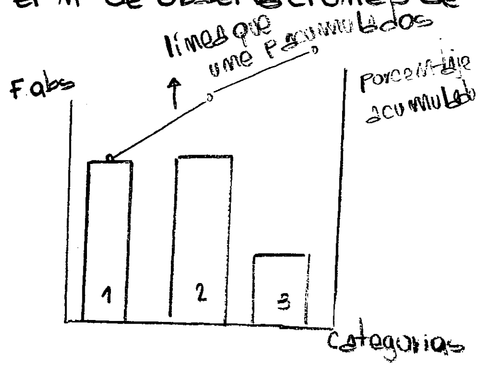
c) diagrama de Pareto = es un gráfico de barras especial

→ suele agregarse una Fila con los porcentajes acumulados

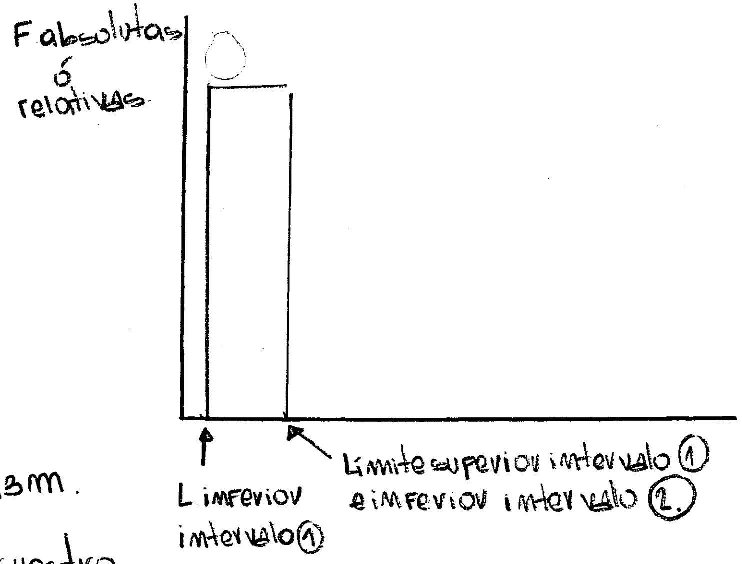
La frecuencia de cada categoría se representa en un eje vertical y las categorías de forma horizontal (en orden descendente de acuerdo con el nº de observaciones de c/u de ellas.)







e) Histograma de frecuencias: caso en donde la variable toma muchos valores diferentes, ocurre generalmente con observaciones de variables continuas.



* se construye tabla de frecuencias.

1) se determina intervalo que incluya todos los valores observados.

Ejemplo: altura de la planta 2,89 m − 2,08 m = 0,81 m tomamos un intervalo de 0,88 m para comenzar en 2,05 m y finalizar en 2,93 m.

2) dividimos el intervalo en subintervalos de igual amplitud, en nuestro caso n = 8. Aplitud de 0,11 m cada subintervalo.

→ observación el nº de intervalos no debe ser ni tan grande, ni tan pequeño con el fin de no perder el objetivo de agrupación de datos ni que se pierda demasiada información original

3) definimos frecuencia absoluta $\to (F_j)$ correspondiente a un intervalo genérico "j" ($j = 1, 2 \ldots n$) al nº de valores de la variable que pertenecen al mismo; sin incluir extremo superior pero si inferior o el caso contrario.

4) También podemos calcular frecuencia relativa (h_j) que se define como el cociente entre F_j y el nº total de datos.

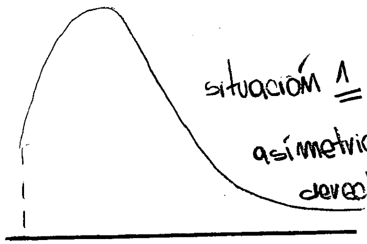
La frecuencia absoluta acumulada $\to F_j$ es igual al nº de observaciones menores o igual al límite superior de ese intervalo.

5) Frecuencia relativa acumulada (H_j) se define como el cociente entre la F_j y el nº total de datos.

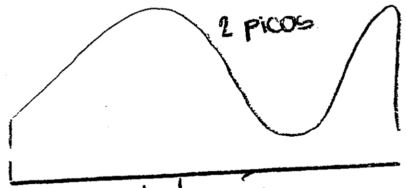
- Sobre cada subintervalo se grafica un rectángulo cuya área representa la Frecuencia (relativa o absoluta) del mismo.
Si los intervalos tienen igual amplitud, entonces el segmento es unitario.

• El gráfico de tallo y hoja nos permite ver características de la distribución. } pueden comparar distribuciones relacionadas.

Distribución
- → simétrica = si al trazar una línea justo sobre el valor medio obtenemos $\underline{\underline{\text{casi}}}$ la imagen especular de la otra parte ↳ ≅.
- → asimétrica = rotamos el diagrama 90°



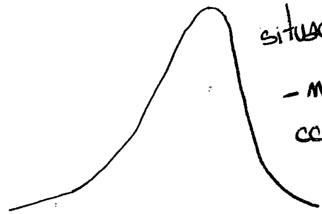
situación 1: asimétrico x' derecha



situación 3: 2 picos

distribución bimodal

* A partir del análisis del diagrama de tallo y hoja → también gráfico de bastones. y el de puntos podemos ubicar aproximadamente el valor central, hablar sobre simetría, nº de picos.
- permiten detección de "outliers" → observaciones lejos del conjunto de datos.



situación 4:
- menor variabilidad
- datos valor central

d) gráfico o diagrama de bastones = muchas veces utilizamos una tabla de frecuencias p' organizar datos.

Cuando el nº de observación es grande pero hay pocos valores diferentes (generalmente variables discretas), es conveniente organizar datos en tabla
- En el eje de las abcisas se representan valores observados de la variable y en el eje de las ordenadas las correspondientes frecuencias relativas y/o absolutas.

(n) observaciones	(b_j) F absoluta	(h_j) F relativa	(F_j) F absoluta acumulada	(H_j) F relativa acumulada
	nº de veces que se repiten los datos	cociente entre frec. absoluta y nº total de datos	el nº total de observaciones menores o iguales a x_j	cociente entre el nº total de datos y frecuencia absoluta acumulada

(3) 3 plantas tienen 6 hojas o menos.

* Diagramas de dispersióm ← se aplica para regresióm y correlacióm

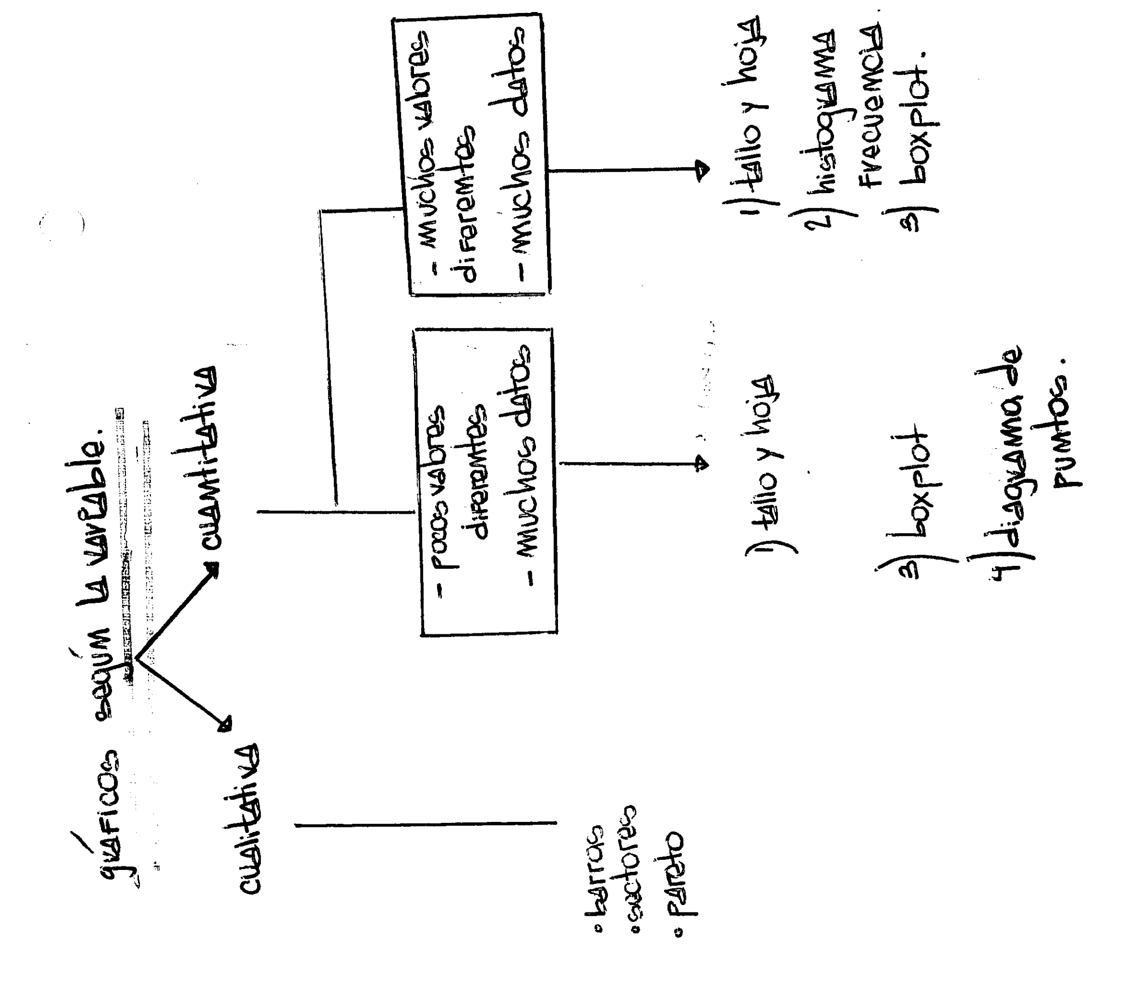
- muestra la relación entre dos variables cuamtitativas

Los valores de uma variable se situam em el eje de las abscisas y el otro em el eje de las ordenadas. El gráfico revela la naturaleza de las variables, analizando:

1) Forma = cuando los puntos del diagrama de dispersióm se situám aproximadamente a lo largo de uma línea recta decimos que la relación es lineal. Otras formas pueden ser agrupaciones, relaciones curvilíneas u otras formas. no claras ni definidas.

2) Dirección = si la relación entre las variables es clara y tiene uma dirección visible decimos que existe uma dirección positiva cuando a valores mayores de uma variable le corresponde mayores de la otra.

3) Intensidad = la fuerza de la relación estará dada por la proximidad de los puntos a la curva que se supone que describe el comportamiento de uma en función de la otra.



Estadísticas.- principales medidas características del conjunto.

dispersión
(se refieren a la variabilidad)

* las medidas de posición son útiles pero resumen sólo parte de la información contenida en el conjunto de datos. Podemos tener dos conjuntos con medidas de posición aproximadamente igual pero difieren variabilidad de valores observados. (en cuánto se alejan del valor central).

1) Rango= es la diferencia entre el valor máximo de la variable y el menor valor observado.

* el rango sólo tiene en cuenta variabilidad valores extremos y no nos da idea cc. de los datos alrededor valor central.
Es útil cuando hay pocos datos. ⟶ nos da idea de la variación

2)_ Varianza= cuando se usa la media aritmética como medida de posición, debe definirse una característica que sirva como medida de variabilidad, en promedio, de los datos respecto de la media. Por lo tanto nos basaremos en el siguiente desvío= $(x_i - \overline{x})$. Pero la suma de esos desvíos es nula, puesto que =

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right) = \sum_{i=1}^{n} x_i - \overline{x}.n = \overset{\text{consultar}}{n.\overline{x} - n.\overline{x}} = 0$$

Posición
(dan idea de localización central).

1) media aritmética= es la suma de los valores observados dividido nº total de datos.
° es la abcisa del centro de gravedad de la distribución de las frecuencias.
$(x_1, x_2, x_3, \ldots, x_i, \ldots, x_n)$ son los n valores observados de la variable x :

$$\overline{x} = \frac{1}{n}.\left(x_1 + x_2 + x_3 + \ldots + x_n\right) = \frac{1}{n}.\sum_{i=1}^{n} x_i$$

(*) la media aritmética depende de todas las observaciones por lo que la presencia de un valor extremo influye sensiblemente en ella. En este caso la media no es buena medida de posición central

2) Fractilas o cuantilos= la fractila de orden "r" es aquel valor tal que r % está comprendido $(0 \le r \le 100)$ de las observaciones son menores o iguales que él.

Mediana= es aquel valor de la variable que se encuentra en el valor central del conjunto de datos. La mitad de las observaciones son menores o iguales que él y la otra mitad mayor o igual.

* consultar deciles y percentilos.

- La mediana se calcula ordenando el conjunto de datos de menor a mayor.
Si el conjunto de es par la mediana se excluye.
Si el " " datos es impar la mediana se consideran.

De aquí se deduce que si se desea definir una medida de variabilidad promedio usaremos una medida de dispersión característica que se basa en la suma de los cuadrados de esos desvíos

$$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

* A mayor variabilidad, mayor S^2

* Si el conjunto de las observaciones provienen de mediciones reiteradas de una misma magnitud, la variancia nos refleja la presición de la medida.

$\uparrow S^2 \downarrow$ - presición

* Si las observaciones provienen de unidades diferentes, la variancia es medida de la homogeneidad del conjunto de datos.

$\downarrow S^2 \uparrow$ + homogeneidad.

3) desviación estándar = se define está medida como la raíz cuadrada positiva de la variancia

$$S = +\sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

se expresa con unidades de la variable y podemos interpretarla como la distancia promedio de las observaciones con respecto a la media.

4) Coeficiente de variación = es una medida adimensional que indica que proporción representa la desviación estándar con respecto a la media

$$C.V = S / |\bar{x}|$$

3) Modo = es el valor de la variable con mayor frecuencia.

- el modo es la única medida de posición que puede usarse para datos provenientes de una variable cualitativa.

5) Rango intercuartil = es la diferencia entre el tercer cuartil y el primero.

$Q_3 - Q_1$.

* medida de dispersión que no se ve afectada por valores extremos.

* Cuando usamos la mediana (Q_2) como medida de posición central, el rango intercuartil es la medida de dispersión adecuada para acompañarla.

Mide la dispersión del 50% datos centrales. Si su valor pequeño nos dice que el 50% datos centrales tienen poca variabilidad.

Boxplot: obtenemos un conjunto de cinco n° que brindan un buen resumen de nuestros datos. Cabe aclarar que en cuanto a la simetria de la distribución resulta más preciso el histograma o diagrama de tallo y hoja. Este gráfico brinda una medida central, la mediana y una idea de la dispersión a través del rango y el rango intercuartil.

* modificando este gráfico podemos detectar potenciales outliers (observaciones que no son típicas del conjunto). Se consideran potenciales outliers:

$$Q_1 - 1,5 . rI$$

$$Q_3 + 1,5 . rI .$$

Control estadístico de calidad.

- Control de calidad → proceso analítico de medición / → proceso de producción

* variación = es una característica observada o medida se debe a una gran cantidad o variedad de causas

Causas
→ asignables = cosas identificables externas que alteran mis resultados
→ no asignables ó aleatorias = todas aquellas cosas no identificables y que no pueden ser controladas.

* Tiempo control
→ 1) "poner el equipo bajo control" = quitar causas asignables.
→ 2) "control periódico"

¿Cuando un proceso está bajo control estadístico? → cuando sólo están presentes causas aleatorias.

El comportamiento colectivo de las variaciones será predecible y es posible fijar límites de control dentro de los cuales se espera que caigan la mayoría de observaciones que realicemos sobre la muestra control.

* muestra control = material de referencia estándar en el cual el nivel del analito se encuentra certificado por una autoridad reguladora.

Los controles se realizan mediante "cartas de control" → ventajas = * previenen ajustes innecesarios del proceso.
* proveen información diagnóstica

- cartas de control de variables
- gráfico de shewart

- cartas de control de atributos.

"Si el proceso está bajo control estadístico las determinaciones de la concentración del analito efectuadas sobre la muestra control tendrán distribución normal con promedio μ y σ_x, llamadas especificaciones de calidad del proceso.

Control de Mediciones → estabilidad del valor medio (μ) (1)

Control de Mediciones → variabilidad del proceso (2)

(1) estabilidad

- determinaciones individuales realizadas sobre la muestra control periodicamente.
  ($n = 1$ por día)

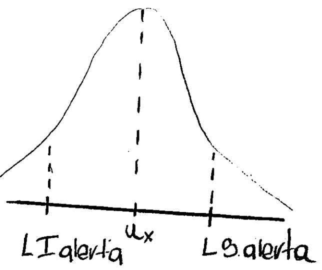
- conjunto de n mediciones realizadas sobre la muestra control periodicamente.
  ($n > 1 \Longrightarrow$ trabajamos con la media aritmetica).

periodicamente. → ojo q' en límites de control periodicos no se estandariza en base a $\bar{x}$ sino x.

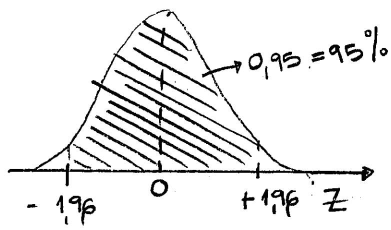
N Límites de alerta $= P\left(\text{LA inferior} < x < \text{L.A. superior}\right)$

$$= P\left(\boxed{\frac{\text{L.A. inferior} - \mu_x}{\sigma_x}} < \frac{x - \mu_x}{\sigma_x} < \boxed{\frac{\text{L.A. superior} - \mu_x}{\sigma_x}}\right)$$

estandarizamos



$$= P\left(Z_{0,025} < Z < Z_{0,975}\right) = 0,95$$



$$= P\left(\underline{-1,96} < Z < \underline{1,96}\right) =$$

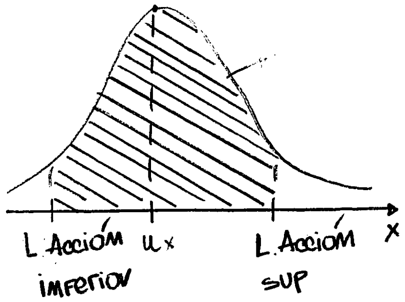
$$\boxed{* \ \frac{\text{L.A. inf} - \mu_x}{\sigma_x} = -1,96} \Longrightarrow \text{L.A. inferior} = -1,96 \cdot \sigma_x + \mu_x$$

$$\boxed{* \ \frac{\text{L.A. superior} - \mu_x}{\sigma_x} = 1,96} \Longrightarrow \text{L.A. superior} = 1,96 \cdot \sigma_x + \mu_x$$

→ Límites de acción. $= P(L.\text{inferior Acción} < x < L.\text{superior Acción})$

estandarización

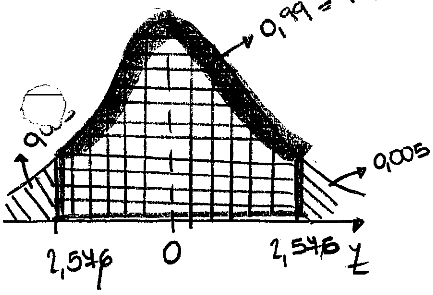
$$= P\left(\frac{L.\text{inferior acción} - u_x}{\sigma_x} < \frac{x - u_x}{\sigma_x} < \frac{L.\text{superior acción} - u_x}{\sigma_x}\right) =$$



$$= P\left(Z_{0,005} < Z < Z_{0,995}\right) = 0,99$$

$$= P\left(-2,546 < Z < 2,545\right) = 0,99$$

$$\Longrightarrow \frac{L.\text{inferior acción} - u_x}{\sigma_x} = -2,546$$



$$L.\text{inferior acción} - u_x = -2,546 \cdot \sigma_x$$

$$\boxed{L.\text{inferior acción} = -2,546 \cdot \sigma_x + u_x}$$

$$\Longrightarrow \frac{L.\text{superior acción} - u_x}{\sigma_x} = 2,546$$

$$\boxed{L.\text{superior acción} = 2,546 \cdot \sigma_x + u_x}$$

conjunto de "n" mediciones realizadas sobre la muestra control periodicamente.

Límites de alerta. $= P(L.\text{alerta inferior} < \bar{x} < L.\text{alerta superior})$

$$= P\left(\frac{L.\text{alerta inferior} - u_x}{\sigma_x/\sqrt{m}} < \frac{\bar{x} - u_x}{\sigma_x/\sqrt{m}} < \frac{L.\text{alerta inferior} - u_x}{\sigma_x/\sqrt{m}}\right)$$

↓

gráficos son los mismos sólo que en vez de x se grafica en función $\bar{X}$.

$$= P\left(Z_{0,025} < Z < Z_{0,975}\right) = 0,95$$

$$= P\left(-1,96 < Z < 1,96\right) = 0,95$$

$$* \frac{L.\text{alerta inferior} - u_x}{\sigma_x/\sqrt{m}} = -1,96 \Longrightarrow L.\text{alerta inferior} = -1,96 \cdot \frac{\sigma_x}{\sqrt{m}} + u_x.$$

$*\ \dfrac{L.\text{alerta superior} - u_x}{\frac{\sigma_x}{\sqrt{n}}} = 1{,}96 \Longrightarrow L.\text{alerta superior} = 1{,}96 \cdot \dfrac{\sigma_x}{\sqrt{n}} + u_x$

Límites de acción $= P\left(\text{Límite inferior de acción} < \bar{x} < \text{Límite superior de acción}\right)$

$$\underset{\text{estandarizo}}{=} P\left(\boxed{\frac{L.i.\text{ Acción} - u_x}{\sigma_x/\sqrt{n}}} < \frac{\bar{x} - u_x}{\sigma_x/\sqrt{n}} < \frac{L.S.\text{ Acción} - u_x}{\sigma_x/\sqrt{n}}\right)$$

* los gráficos son equivalentes a los anteriores solo que se grafica en función de la $\bar{x}$

$$= P\left(Z_{0{,}005} < \bar{x} < Z_{0{,}995}\right) = 0{,}99$$

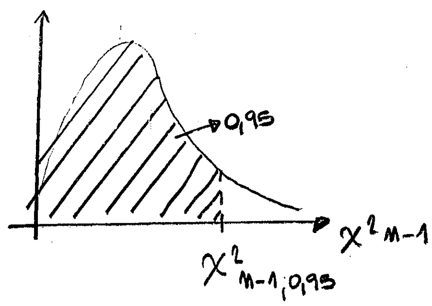
$$= P\left(\underline{-2{,}576} < \bar{x} < 2{,}576\right) = 0{,}99$$

$*\ \dfrac{L.I.\text{Acción} - u_x}{\sigma_x/\sqrt{n}} = -2{,}576 \Longrightarrow L.I.\text{ Acción} = -2{,}576 \cdot \dfrac{\sigma_x}{\sqrt{n}} + u_x$

$*\ \dfrac{L.S.\text{ Acción} - u_x}{\sigma_x/\sqrt{n}} = 2{,}576 \Longrightarrow L.S.\text{ Acción} = \dfrac{2{,}576 \cdot \sigma_x}{\sqrt{n}} + u_x$

2) Límites de control de variabilidad. $\longrightarrow$ analizamos sólo los límites superiores

¿por qué? $\uparrow S_x^2 \downarrow$ precisión



L. Alerta superior $\quad P\left(S_x^2 < L.\text{alerta}\right) = 0{,}95$

$$\underset{\text{estandarizo}}{=} P\left(\frac{S_x^2 \cdot (n-1)}{\sigma_x^2} < \frac{L.\text{alerta} \cdot (n-1)}{\sigma_x^2}\right) = 0{,}95.$$

$$P\left(\chi^2(n-1) < \frac{L.\text{alerta} \cdot (n-1)}{\sigma_x^2}\right) = 0{,}95$$

$$P\left(\chi^2_{(n-1)} < \underline{\chi^2(n-1);0{,}95}\right) = 0{,}95$$

$*\ \dfrac{\text{Límite alerta superior} \cdot (n-1)}{\sigma_x^2} = \chi^2(n-1);0{,}95 \Longrightarrow L.A.\text{superior} = \dfrac{\chi^2(n-1);0{,}95 \cdot \sigma_x^2}{n-1}$

L. Accion superior $= P\left(S_x^2 < L.A.\text{superior} \right)$

$$= \underset{\uparrow \atop \text{estandarizo}}{P} \left(\frac{S_x^2 \cdot \sigma^2_x}{n-1} < \frac{L.A.\text{superior} \cdot \sigma^2_x}{n-1} \right) = 0{,}99$$

$$= P\left(\chi^2_{(n-1)} < \chi^2_{(n-1);\underset{\downarrow}{0{,}99}} \right) =$$

prob que acumula.

$$* \frac{L.A.\text{superior} \cdot \sigma^2_x}{n-1} = \chi^2_{(n-1);0{,}99} \Longrightarrow \underline{L.\text{Alerta superior}} = \frac{\chi^2_{(n-1);0{,}99} \cdot (n-1)}{\sigma^2_x}$$

- Criterios para determinar que un método esta fuera de control estadístico

Expresión general p/ Límites de Control

$\circledast\ \mu_x \pm K \cdot \sigma_x$

ó

$\mu_x \pm K \cdot \frac{\sigma_x}{\sqrt{n}}$ $\circledast$

- (68%) → K = 1
- L. alerta (95%) → K = 2
- L. Acción (99,73%) → K = 3.

* Falta de estabilidad = uno o más puntos fuera de los límites de acción.

- presencia de patrones, que denote falta de aleatoridad
  - → patrón estratificado = poca variabilidad
  - → " ciclico
  - → cambio de promedio
  - → tendencia.

- dos o tres puntos consecutivos entre los límites de alerta y acción
- cuatro o cinco puntos consecutivos entre los límites de K=1

* Falta de precisión = aumento de la variancia.

- se analiza x' medio de la variancia de un conjunto de mediciones realizadas sobre la muestra control.

↓ solo se analiza si $n > 1$

Cuando $S_x^2 \longrightarrow \frac{\chi^2_{(n-1)} \cdot \sigma^2_x}{n-1}$ es grande perdemos precisión

Limites

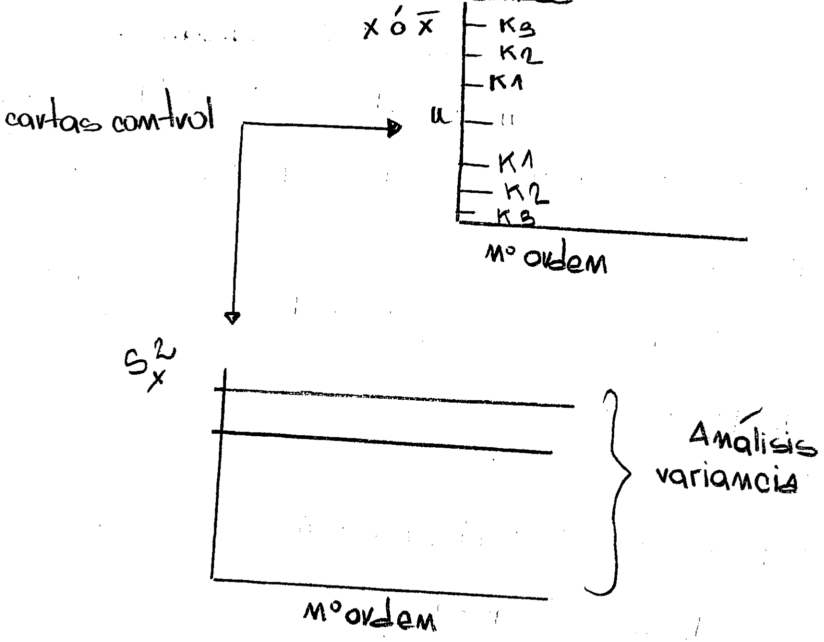
- Alerta
- Acción

$$P(s_x^2 < L.\text{Alerta superior}) = 0{,}95 \qquad P(s_x^2 < L.\text{Acción superior}) = 0{,}99$$

$$L.\text{Alerta superior} = \frac{\chi^2_{(m-1);\,0{,}95} \cdot \sigma^2}{(m-1)} \qquad L.\text{Acción sup} = \frac{\chi^2_{m-1;\,0{,}99} \cdot \sigma^2}{(m-1)}$$

* bajo control → ninguna varianza fuera de los límites de acción
- no debe aumentar la varianza (patrones ascendentes fuera de control, se pierde la presición).

Analisis Estabilidad



ejercicios unidad n°4.

1) a) 0,61 es el valor de una ~~estádistica~~ parámetro. Representa a la población.

b) $n = 200$ semillas de girasol

130 de ellas germinan.

A: "semillas germinadas"

$$h(A) = \frac{130 \text{ semillas}}{200 \text{ semillas}}$$

$h(A) = 0,65$ → este valor representa una estadística porque es sobre una muestra aleatoria.

No necesariamente obtendríamos el mismo valor de $h(A)$ ya que se calcula sobre una variable aleatoria y el nº de semillas germinadas seguramente diriera en las diferentes muestras de ella.

2)_ Cuando hacemos referencia a la "distribución del muestreo" nos referimos a la distribución de la probabilidad de cada una de las estádisticas, ya que definimos a estas como variables aleatorias.

3) x = diámetro del capítulo de las plantas de girasol (cm).

$$X \longrightarrow N(u_x = 20cm ; \sigma_x = 3,2cm).$$

$$\bar{X} \longrightarrow N\Big(\overbrace{u_x}^{E(\bar{X})} ; \overbrace{\frac{\sigma_x}{n}}^{V(\bar{X})}\Big)$$

consultar como lo pienso

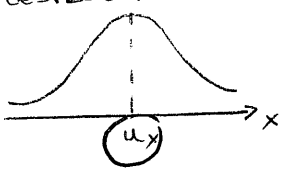
muestra $n = 3$

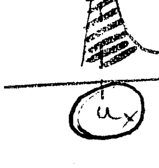
$$V(\bar{X}) = \frac{20cm}{\sqrt{3}} = 11,55 \text{ cm} \qquad V(\bar{X}) = \frac{20cm}{\sqrt{30}} = 3,651 \text{ cm}$$

Los valores se encuentran centrados siempre en los mismos, pero cuánto mayor es el tamaño de la muestra los valores se centran más alrededor de éste porque disminuye el desviación estándar.

→ índice numérico de la dispersión de los datos.

↓ desviación estandar ; ↓ dispersión de los datos





centrados siempre en el mismo valor.

4) x = diámetro del tallo plantas girasol. *consultar

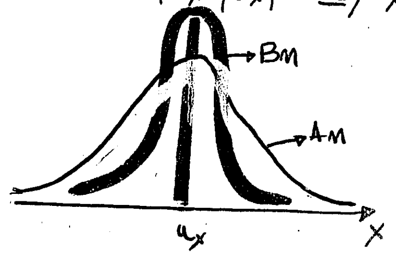
$x \longrightarrow N(u_x ; \sigma_x)$

$\bar{x} \longrightarrow N\left(u_x ; \frac{\sigma_x}{\sqrt{M}}\right)$ → desviación estandar depende tamaño de la muestra

6). x : contenido de As.

$\sigma_x = 2,5 \cdot 10^{-3}$ mg/L. $A_M = 18$ mediciones $B_M = 48$ mediciones.

Si $X \longrightarrow N(u_x ; \sigma_x) \Longrightarrow \bar{X} \longrightarrow N\left(u_x ; \frac{\sigma_x}{\sqrt{M}}\right)$



b) el analista b tiene más chances porque al aumentar el m de la muestra reducís el desvío estándar concentrando los valores cerca de la $\bar{x}$.

7). x : concentración de benceno en agua de desecho

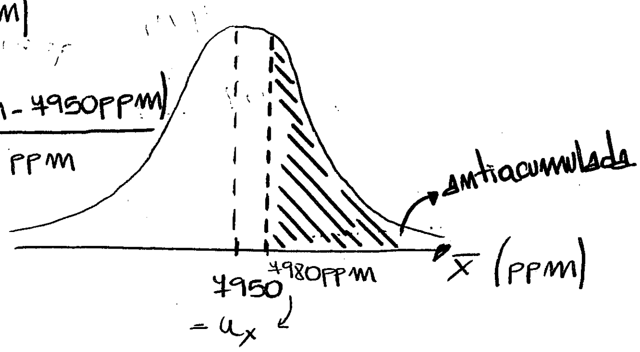
$X \longrightarrow N(u_x = 4950 ; \sigma_x = 100)$ ppm.

a) $M = 40$ $\bar{X} = 4980$ ppm $\Longrightarrow P(\bar{X} > 4980\,ppm)$

$\bar{X} \longrightarrow N\left(u_x = 4950\,ppm ; \frac{100\,ppm}{\sqrt{40}} = \sigma_{\bar{x}}\right)$ 15,81

$P(\bar{x} > 4980\,ppm) = 1 - P(\bar{x} < 4980\,ppm)$

$= 1 - P\left(\frac{\bar{x} - u_x}{\sigma_{\bar{x}}} < \frac{4980\,ppm - 4950\,ppm}{ppm}\right)$



$= 1 - P(z < 1,898)$

$= 1 - P(z < \sim 1,9)$

$= 1 - 0,975$

$\boxed{= 0,025}$

9) $P(A) = 0{,}08$ $\quad P(\bar{A}) =$ A: comprimidos que no cumplen lo especificado.

$P(\bar{A}) = 1 -$

$P(\bar{A}) = 0{,}92$

a) consultar procedimiento.

$h(A) \longrightarrow N\left(E[h(A)] = p = P(A) \,;\, V[h(A)] = \frac{p\,q}{n}\right)$ $\quad q = P(\bar{A})$

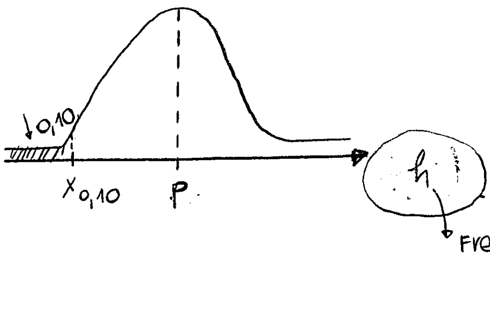
$n = 150$. $\quad$ Si $\longrightarrow n \geq 30$

$P(h \leq 0{,}10) =$

$P\left(\frac{h-u}{\sigma} \leq \frac{0{,}10 - u}{\sigma}\right)$ $\quad u = p \quad \sigma = \frac{p\,q}{n}$

$P\left(Z \leq \frac{0{,}10 - 0{,}08}{0{,}022}\right)$

$P(\underbrace{Z \leq 0{,}91}_{\text{busco valores Z en tabla}}) = 0{,}8159$



b) $n = 10 \longrightarrow$ distribución no es normal

$P(h \leq 0{,}10)$ consultarlo.

10 posibilidades

$P(\bar{C}) = 0{,}08 \Longrightarrow P(C) = 0{,}92$

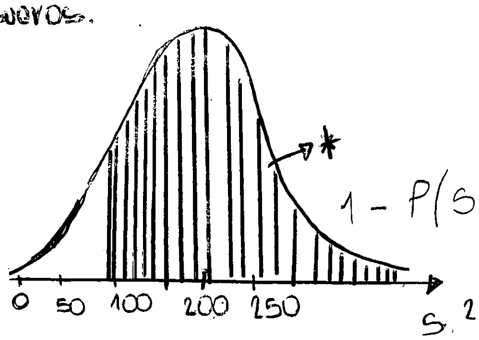
ejercicio nº 10.

x = concentración colesterol en una ~~suero~~ mezcla de sueros.

$\sigma_x = 10$ mg/dL

a) $x \longrightarrow N(u_x \,;\, \sigma_x)$

$S_x^2 \longrightarrow \frac{\chi^2_{(n-1)} \cdot \sigma_x^2}{(n-1)}$; $n = 10$.



b) $S^2_x = 93{,}48$ mg/dL

$\frac{\chi^2(9) \cdot 10^2}{9}$ $\quad$ x' que tengo el desvío

$P(S_x^2 > 93{,}48 \text{ mg/dL}) = ?$ *

$P\left(\frac{\chi^2(9) \cdot 10^2}{9} > 93{,}48 \text{ mg/dL}\right) =$

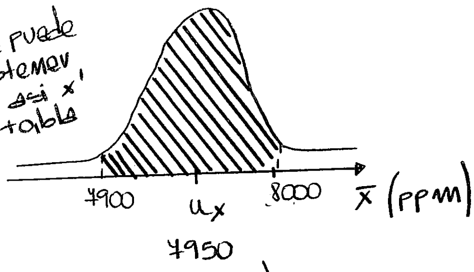
$P\left(\chi^2(9) > \frac{93{,}48 \text{ mg/dL} \cdot 9}{10^2}\right)$

b)

$$P\left(4900 < \bar{X} < 8000\right)$$

no se puede obtener así x' tabla

$$P\left(\frac{4900-4950}{15{,}81} < \frac{\bar{X}-u_x}{\sigma_x} < \frac{8000-4950}{15{,}81}\right) =$$



$$= P\left(\frac{\bar{X}-u_x}{\sigma_x} < \frac{8000-4950}{15{,}81}\right) - P\left(\frac{\bar{X}-u_x}{\sigma_x} < \frac{4900-4950}{15{,}81}\right)$$

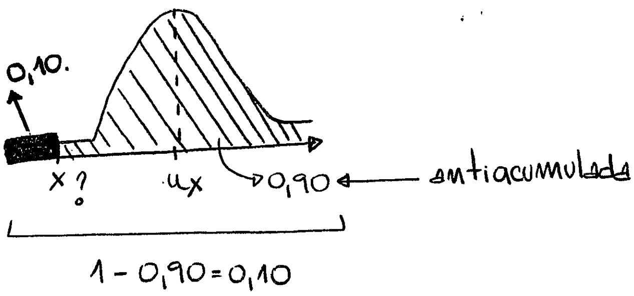
$$= P\left(Z < 3{,}16\right) - P\left(Z < -3{,}16\right)$$

$$= 0{,}9992 - 0{,}0008$$

$$\boxed{= 0{,}9984}$$

c) $n = 40$ $\quad P\left(\bar{X} > x\right) = 0{,}90$

$$P\left(\frac{\bar{X}-u_x}{\sigma_x} > x\right) = 0{,}90$$



$$P\left(\frac{\bar{X}-u_x}{\sigma_x} < \frac{x-u_x}{\sigma_x}\right) = 0{,}10$$

$$P\left(Z_{0,10} < \frac{x - 4980\,ppm}{15{,}81\,ppm}\right)$$

$$Z_{0,10} \cdot 15{,}81\,ppm + 4950\,ppm = x$$

$$-1{,}28 \cdot 15{,}81\,ppm + 4950\,ppm = x$$

$$-20{,}2368\,'' + 4950\,ppm = x$$

$$\boxed{4929{,}764\ ppm = x}$$

Tiempo de control → 1° "poner el equipo bajo control"
→ 2° "controlar periódicamente"

(*) el comportamiento colectivo de las variaciones será predecible y será posible Fijar límites dentro de los cuales se esperan que caigan la mayoría de observaciones que se realizan sobre la muestra control.

↓

material de referencia estándar en el cual el nivel del analito se encuentra certificado por una entidad reguladora.

→ seguimiento del proceso bajo control estadístico. } objetivo = detección causas asignables que generan que el proceso salga del estado de control.

utilizamos = cartas de control.

- previenen ajustes innecesarios de un proceso.
- proveen información diagnóstica

cartas de control de variables

gráfico de Shewart.

carta control atributos } no la vemos.

cuando el proceso se encuentra bajo control estadístico es posible fijar límites dentro de los cuales se esperan que caigan la mayoría de observaciones efectuadas sobre la muestra control si el proceso sigue bajo control.

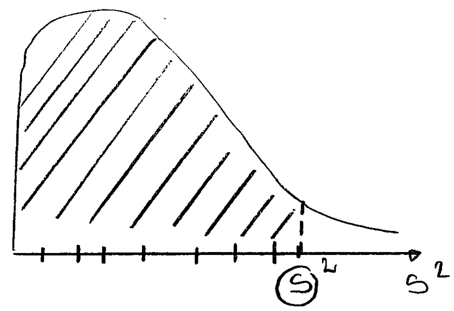
↓ sin causas asignables.

si el proceso esta bajo control las determinaciones de la concentración del analito efectuadas sobre muestra control tendrán distribución normal con promedio μ y desviación estándar σ, llamadas especificaciones de calidad del proceso.

$$= 1 - P\left(\chi^2_{(9)} < \frac{93{,}78\ mg/dL \cdot 9}{10^2}\right)$$

$$= 1 - P\left(\chi^2_{(9)} < 8{,}44\right)$$

$$= 1 - 0{,}50 \Longrightarrow \cong 0{,}50.$$



c) $P\left(S_x^2 < S^2\right) = 0{,}90$

$$P\left(\frac{\chi^2_9 \cdot 10^2}{9} < S^2\right) = 0{,}90$$

tabla

$$P\left(\frac{X^2_{9;0{,}90} \cdot 10^2}{9} < S^2\right) \Longrightarrow \frac{X^2_{9;0{,}90}\ 10^2}{9} = S^2$$

$$\frac{14{,}7 \cdot 100}{9} = S^2$$

$$\boxed{163{,}\overset{\frown}{3} = S^2} \rightarrow$$ se la puede denotar de cualquier forma

- control calidad.

control calidad → proceso analítico de medición / proceso de producción] aplicaciones.

lo que nos interesa ↓

variabilidad → característica observada o medida

¿de qué?

causas

- asignables = causas externas que modifican mis resultados
  ejemplo = operario que no realice su trabajo de forma correcta.
- no asignables = todas aquellas causas que no puedo identificar ni controlar.

"un proceso (analítico de medición o producción) está bajo control estadístico sólo cuando hay causas no asignables". (*)

¿Qué controlamos de un proceso? → estabilidad del valor medio (1)
→ variabilidad del proceso. (2)

(1) control de la estabilidad del valor medio.

↓ conjunto de "n" mediciones realizadas sobre muestra control periódicamente

trabajamos $\bar{X}$

→ determinación individuales realizadas sobre la muestra periódicamente.

trabajamos con "n"

Límites → alerta = límite que esperamos que caigan 95 de 100 mediciones si el proceso sigue bajo control.

→ acción = límite en el que esperamos que caigan 99 de 100 mediciones si la muestra sigue bajo control.

Límites → alerta = " ; → acción = " } igual definiciones que antes

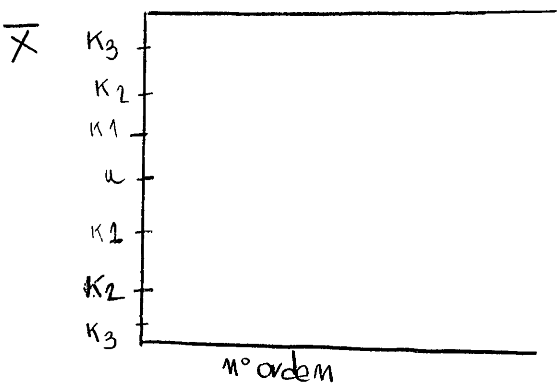
* fórmulas para controles del límite en mediciones individuales

* deducciones fórmulas ≠

Expresión general

$$\mu \pm K \cdot \frac{\sigma_x}{\sqrt{n}}$$

- K = 1 (68%)
- K = 2 (95%) Límites alerta.
- K = 3 (99,73%) Límite acción



- $\underbrace{\text{Li alerta}}_{\text{inferior}} = \mu - 1{,}96 \cdot \sigma$ → proviene del Z que acumula 0,95.

=

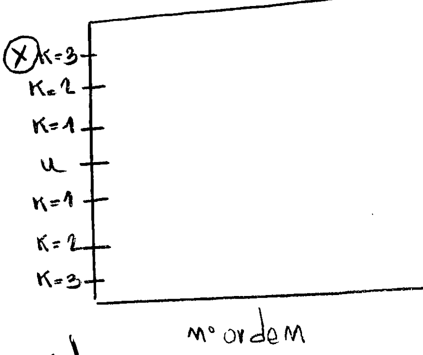
- $\underbrace{\text{Li alerta}}_{\text{superior}} = u + 1{,}96 \cdot \sigma$

$\underbrace{\text{Li acción}}_{\text{inferior}} = u - 2{,}546 \cdot \sigma$ → proviene del Z que acumula 0,99

=

$\underbrace{\text{Li acción}}_{\text{superior}} = u + \underline{2{,}546} \cdot \sigma$

construcción gráfico



Expresión general

$$\mu \pm K \cdot \sigma_x$$

K = 1 (68%)

K = 2 (95%) Límite alerta

K = 3 (99%) Límite acción

- Pérdida de control (reglas)

- uno o más puntos fuera de los límites de acción (dos de tres puntos consecutivos dentro límite alerta y acción).
- presencia de patrones que detome falta de aleatoridad.
- $\underbrace{\text{cuatro de cinco puntos}}_{\text{concecutivos}}$ fuera de los límites k=1

② Control variabilidad de un proceso. (precisión)

* deducción para el final

$$L.alerta = \frac{\chi^2_{(n-1);0,95} \cdot \sigma^2}{(n-1)}$$

$$L.acción = \frac{\chi^2_{(n-1);0,99} \cdot \sigma^2}{n-1}$$

* sólo ponemos límites superiores porque mayor variancia ↓ precisión !! ← lo que nos interesa

ejercicio 11

x: nivel colesterol sérico (mg/dL)

$\mu_x = 210$ mg/dL
$\sigma_x = 3$ mg/dL
$n = 3$

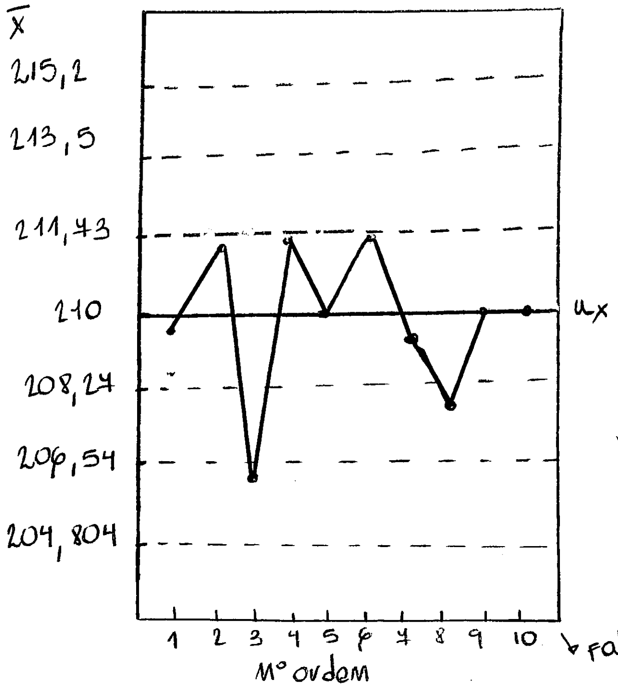
} nos da como dato $\bar{x}$ ⟶ conjunto de "n" mediciones determinadas sobre la muestra

- Limites alerta y acción → variación (precición); → estabilidad

- Limite Alerta ↗ $210 \text{ mg/dL} \pm 2 \cdot \frac{3 \text{ mg/dL}}{\sqrt{3}}$ → superior = 213,5 mg/dL; → inferior = 206,54 mg/dL

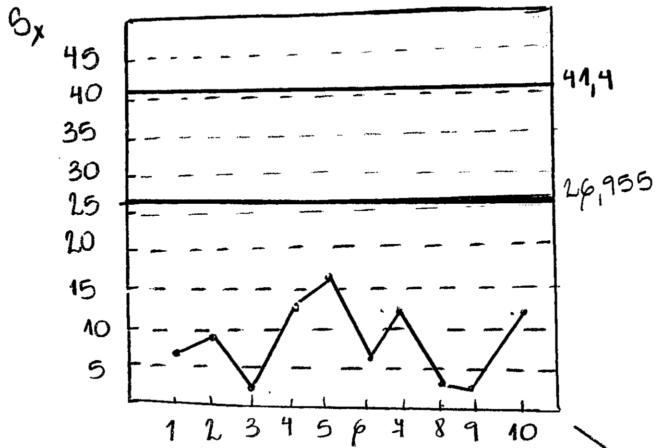
- Limites acción ⟶ $210 \text{ mg/dL} \pm 3 \cdot \frac{3 \text{ mg/dL}}{\sqrt{3}}$ → superior = 215,2 mg/dL; → inferior = 204,804 mg/dL

- Limite 68% ⟶ $210 \text{ mg/dL} \pm 1 \cdot \frac{3 \text{ mg/dL}}{\sqrt{3}}$ → superior = 211,73 mg/dL; → inferior = 208,27 mg/dL



consultar porque para mí los datos están muy dispuestos entre los valores medios y eso implica que se reune más 68% datos $\Longrightarrow$ pérdida de control

faltavom nº ordemes!



faltavom nº de ordemes

* Limite superior alerta $= \dfrac{\chi^2(2);0{,}95 \cdot (\cancel{2}\,3\,mg/dL)^2}{2}$

$$= \frac{5{,}99 \cdot (3\,mg/dL)^2}{2} = \cancel{11{,}98\,mg/dL}\ 26{,}955$$

* Limite superior de acción $= \dfrac{\chi^2(2);0{,}99 \cdot (\cancel{2}\,3\,mg/dL)^2}{2}$

$$= \frac{9{,}21 \cdot (\cancel{2}\,3\,mg/dL)^2}{2} = \cancel{18{,}42}\ 41{,}4\ mg/dL$$

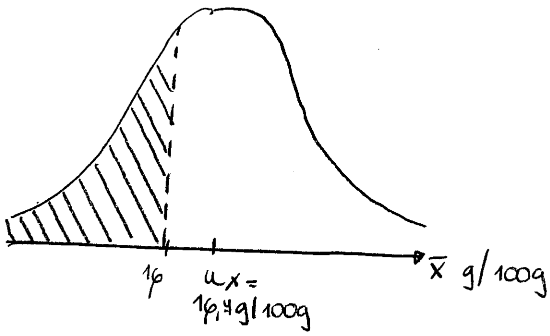
Ejercicios propuestos

13) - x: contenido de fibra total en lentejas (g/g)

$$x \longrightarrow N\left(u_x = 16{,}7\ g/100g \ ;\ \sigma_x = 2{,}2 g/100g\right)$$

a) $n = 12$ $\qquad \bar{x} \longrightarrow N\left(u_x \ ; \dfrac{\sigma_x}{\sqrt{n}}\right)$

$P(\bar{x} < 16{,}00 g/100g) = \ ?$



$$P\left(\frac{\bar{x} - u_x}{\frac{\sigma_x}{\sqrt{n}}} < \frac{16 g/100g - 16{,}7 g/100g}{\frac{2{,}2 g/100g}{\sqrt{12}}}\right) =$$

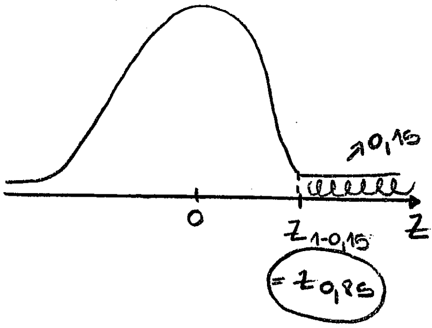
$$P(Z < -0{,}32) = \quad 0{,}1357$$

b) $n = 12$

$$\bar{x} \longrightarrow N\left(u_x \ ; \frac{\sigma_x}{\sqrt{n}}\right).$$

$$P(\bar{x} > \bar{x}') = 0{,}15$$

$$P\left(\frac{\bar{x} - u_x}{\sigma/\sqrt{n}} > \frac{\bar{x}' - u_x}{\sigma_x/\sqrt{n}}\right) = 0{,}15$$



$$1 - P\left(Z_{0,85} < \frac{\bar{x}' - 16{,}7 g/100g}{\frac{2{,}2 g/100g}{\sqrt{12}}}\right) =$$

$$Z_{0,85} = \frac{\bar{x}' - 16{,}7 g/100g}{\frac{2{,}2 g/100g}{\sqrt{12}}}$$

$$1{,}04 \cdot \frac{2{,}2 g/100g}{\sqrt{12}} = \bar{x}' - 16{,}7 g/100g$$

$$0{,}6605\ g/100g + 16{,}7 g/100g = \bar{x}'$$

$$17{,}3605 g/100g = \bar{x}'$$

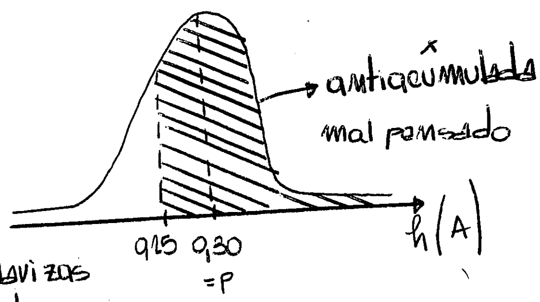
14) A: peces contaminados en un lago

$P(A) = 0{,}30 = 30\%$ $\quad n = 200 \xrightarrow[x' T.C.L]{} h(A) \longrightarrow N\left(E[h(A)] = 0.30 \,;\, V[h(A)] = \frac{0{,}30 \cdot \cancel{0{,}96}^{\,0{,}70}}{200}\right)$

$h(A) \rightarrow N\left(E[h(A)] = 0{,}30 \,;\, V[h(A)] = 0{,}00105\right)$

$P\left(h(A) \overset{\text{a lo sumo}}{\leq} 0{,}25\right) =$ $\quad$ 50 peces contaminados $\Longrightarrow h(50) = 50/200 = 0{,}25$

$P\left(\frac{h(A) - p}{\sqrt{p \cdot q / n}} \leq \frac{0{,}25 - 0{,}30}{\sqrt{0{,}00105}}\right) =$



$1 - P\left(Z \leq \frac{0{,}25 - 0{,}30}{\sqrt{0{,}00105}}\right) =$

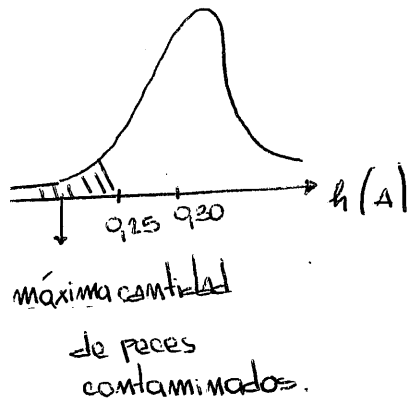
ojo que cuando estandarizamos es el desvío estandar no la variancia

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

$P(Z < -1{,}54) =$

$\underline{\underline{0{,}0618}}$

- tener en cuenta que el término "a lo sumo" representa un máximo.

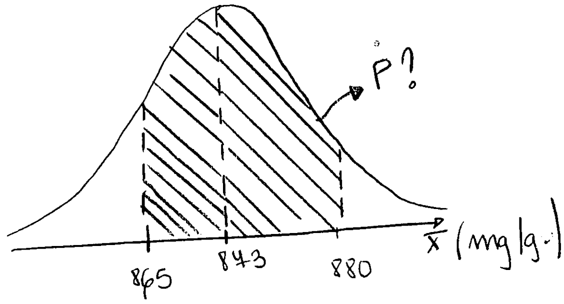


15) - x: contenido de potasio en garbanzos (mg/gr)

$(\mu_x = 843 \text{ mg/gr} \,;\, \sigma_x = 21 \text{ mg/g})$

a) $n = 45$ $\quad P(865 \text{ mg/g} < \bar{x} < 880 \text{ mg/g}) = ?$

$\bar{x} \xrightarrow[x' T.C.L]{} N\left(\mu_x = 843 \text{ mg/g} \,;\, \sigma_x = \frac{21 \text{ mg/g}}{\sqrt{45}}\right)$



$P(865 \text{ mg/g} < \bar{x} < 880 \text{ mg/g}) =$

$P(\bar{x} < 880 \text{ mg/g}) - P(\bar{x} < 865 \text{ mg/g}) =$

$P\left(\frac{\bar{x} - \mu_x}{\sigma_x/\sqrt{n}} < \frac{880 \text{ mg/g} - 843 \text{ mg/g}}{3{,}13 \text{ mg/g}}\right) - P\left(\frac{\bar{x} - \mu_x}{\sigma_x/\sqrt{n}} < \frac{865 \text{ mg/g} - 843 \text{ mg/g}}{3{,}13 \text{ mg/g}}\right) =$

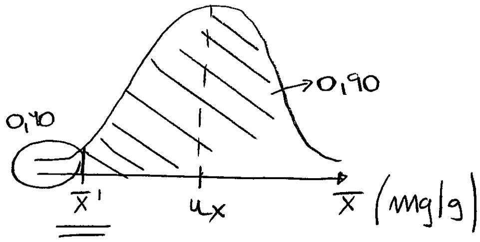
$P(Z < 2{,}24) - P(Z < -2{,}56) = 0{,}9845 - 0{,}0052$

$= 0{,}9823$ ✓

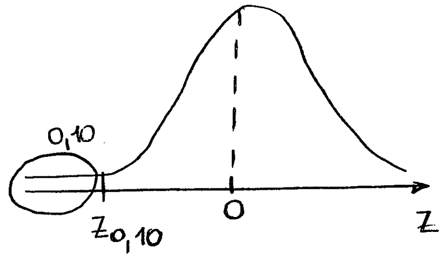
b).- $n = 45$

$\bar{x}' = ?$ $\qquad P(\bar{x} > x') = 0{,}90$

$$P\left(\frac{\bar{x} - u_x}{\sigma_x/\sqrt{n}} > \frac{\bar{x}' - u_x}{\sigma_x/\sqrt{n}}\right) = 0{,}90$$



$$P\left(Z_{0,10} < \frac{\bar{x}' - u_x}{\sigma_x/\sqrt{n}}\right) = 0{,}10$$



$$Z_{0,10} = \frac{\bar{x}' - u_x}{\sigma_x/\sqrt{n}}$$

$$-1{,}28 = \frac{\bar{x}' - 873\ mg/g}{3{,}13\ mg/g}$$

$$-1{,}28 \cdot 3{,}13\ mg/g = \bar{x}' - 873\ mg/g$$

$$-4{,}0064\ mg/g = \bar{x}' - 873\ mg/g$$

$$\underbrace{868{,}9936}_{\cong 869}\ \frac{mg}{g} = \bar{x}$$

Limites = $u_x \pm \sigma_x . K$

datos variable → x = promedio } estabilidad

datos variable → $\bar{x}$ } estabilidad y presición s_x^2 → Limites $u_x \pm \frac{\sigma_x . K}{\sqrt{n}}$

*consultar si se usan los valores calculados sin la raíz cuadrada.

16).- ($\mu = 15{,}0\ mg/dL$; $\sigma_x = 0{,}5\ mg/dL$) $\qquad n = 20$.

x = contenido nitrogeno ureico.

Limites alerta = $\left(u_x \pm K . \frac{\sigma_x}{\sqrt{n}}\right)$; $K = 2$ → superior = 15,224 mg/dL ; inferior = 14,78 mg/dL.

Limites 68% = $\left(u_x \pm K . \frac{\sigma_x}{\sqrt{n}}\right)$; $K = 1$ → superior = 15,112 mg/dL ; inferior = 14,88 mg/dL

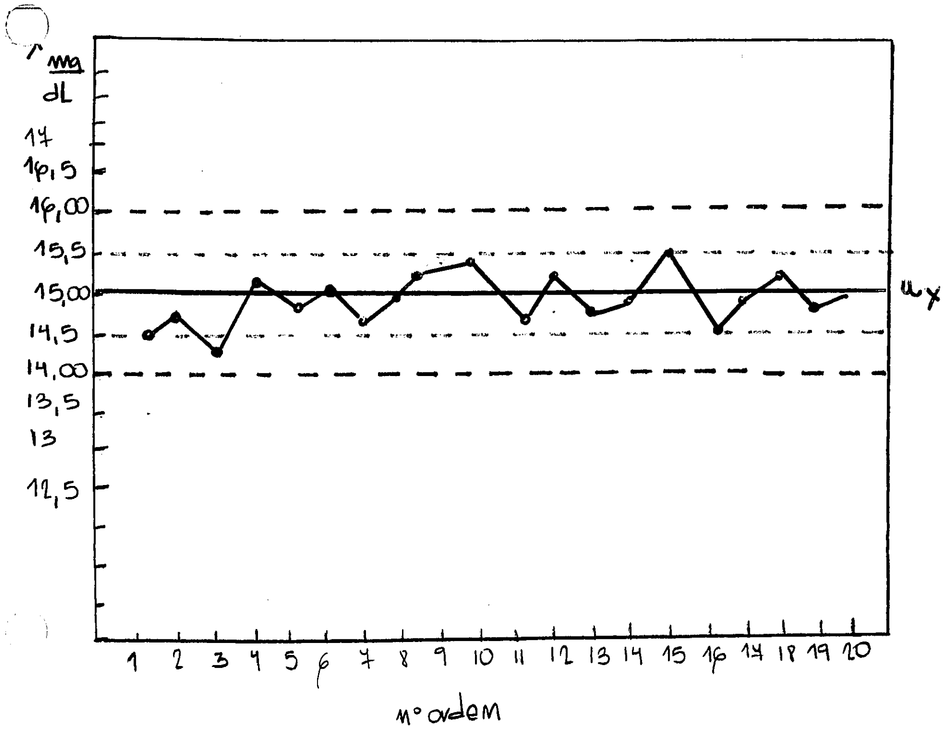
Limites acción = $\left(u_x \pm K . \frac{\sigma_x}{\sqrt{n}}\right)$; $K = 3$ → superior = 15,33 mg/dL ; inferior = 14,665 mg/dL

Limites alerta → superior = 15 mg/dL + 2.0,5 mg/dL = 16 mg/dL
→ inferior = 15 mg/dL - 2.0,5 mg/dL = 14 mg/dL

Limites 68% → superior = 15 mg/dL + 1.0,5 mg/dL = 15,5 mg/dL
→ inferior = 15 mg/dL - 1.0,5 mg/dL = 14,5 mg/dL

Limite alerta → superior = 15 mg/dL + 3.0,5 mg/dL = 16,5 mg/dL
→ inferior = 15 mg/dL - 3,0,5 mg/dL = 13,5 mg/dL

u_x = 15 mg/dL.



100% datos ~ 20.
68% datos ~ ≅13

conclusión = el proceso de determinación de nitrógeno ureico está fuera de control. Los datos no se encuentran fuera de los limites de acción pero en base a la distribución de los mismos durante 20 días notamos la falta de aleatoridad del proceso

17) X : medición pH compuesto químico.

$X \longrightarrow N(u_x ; \sigma_x = 0,1)$.

$\bar{X} \longrightarrow N\left(u_x ; \frac{\sigma_x}{\sqrt{n}}\right)$ → el verdadero valor pH

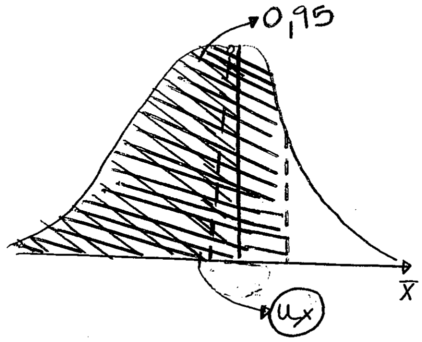
n = ?

considerando que $\boxed{\bar{X}' - u_x} = 0,02$

$P(\bar{X} < \bar{X}' + 0,02)$

$P\left(\frac{\bar{X} - u_x}{\sigma_x/\sqrt{n}} < \frac{0,02}{0,1/\sqrt{n}}\right) =$

$$P\left(Z < \frac{0{,}02}{0{,}1/\sqrt{n}}\right) = 0{,}95$$



$$Z_{0{,}95} = \frac{0{,}02}{0{,}1/\sqrt{n}}$$

x'enunciado

$\bar{x}$ '- $u_x = 0{,}02$

$$1{,}65 \cdot \frac{0{,}1}{\sqrt{n}} = 0{,}02$$

$$1{,}65 \cdot 0{,}1 = \sqrt{n} \cdot 0{,}02$$

$$\frac{1{,}65 \cdot 0{,}1}{0{,}02} = \sqrt{n}$$

$$8{,}25 = \sqrt{n}$$

$68 \text{ ó } 69 \cong n$ → consultar.

18) - x: contenido de aflatoxinastotales en cereales

método - elisa.

$(u_x = 5ppm;\ \sigma_x = 0{,}4ppm)$

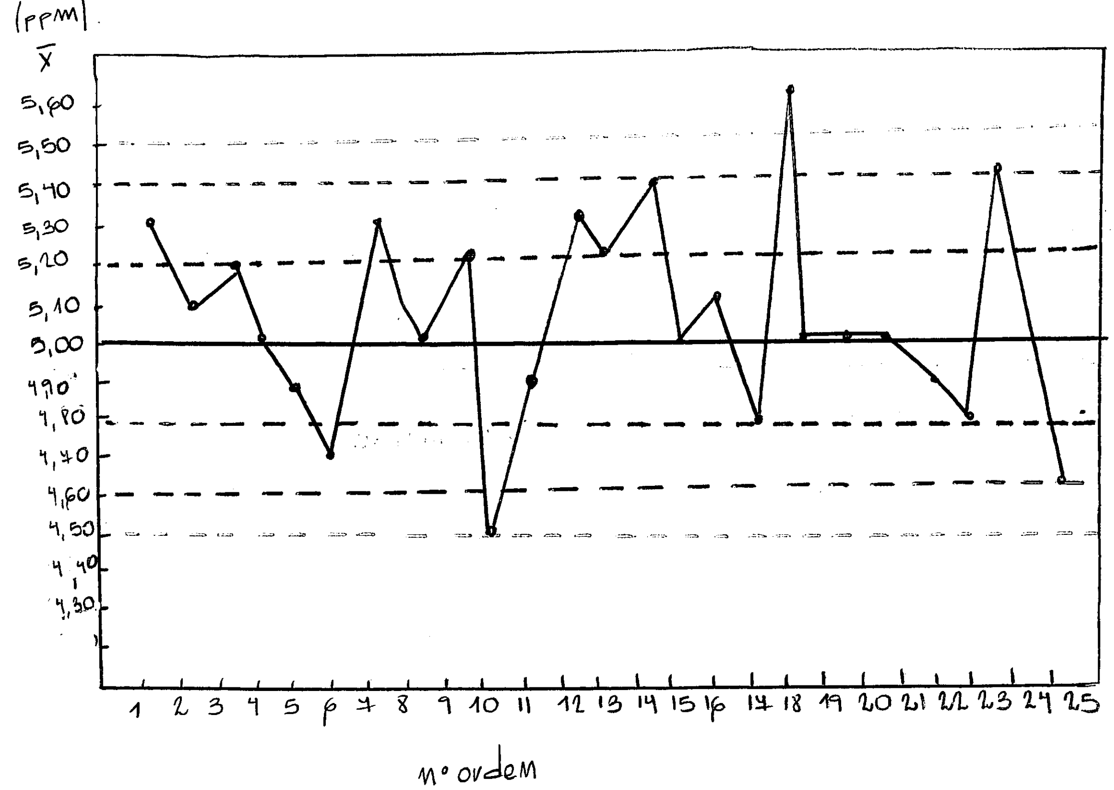
$n = 5$ → conjunto de "n" mediciones diarias, trabajamos con $\bar{x}$.

Limite acción $= u_x \pm 3 \cdot \frac{\sigma_x}{\sqrt{5}} =$ → superior $= 5{,}5$ ppm. → inferior $= 4{,}5$ ppm.

Limites alerta $= u_x \pm 2 \cdot \frac{\sigma_x}{\sqrt{n}} =$ → superior $= 5{,}4$ ppm → inferior $= 4{,}6$ ppm

Limites 68% $= u_x \pm \frac{1 \cdot \sigma_x}{\sqrt{5}} =$ → superior $= 5{,}20$ppm. → inferior $= 4{,}80$ppm.

Análisis estabilidad

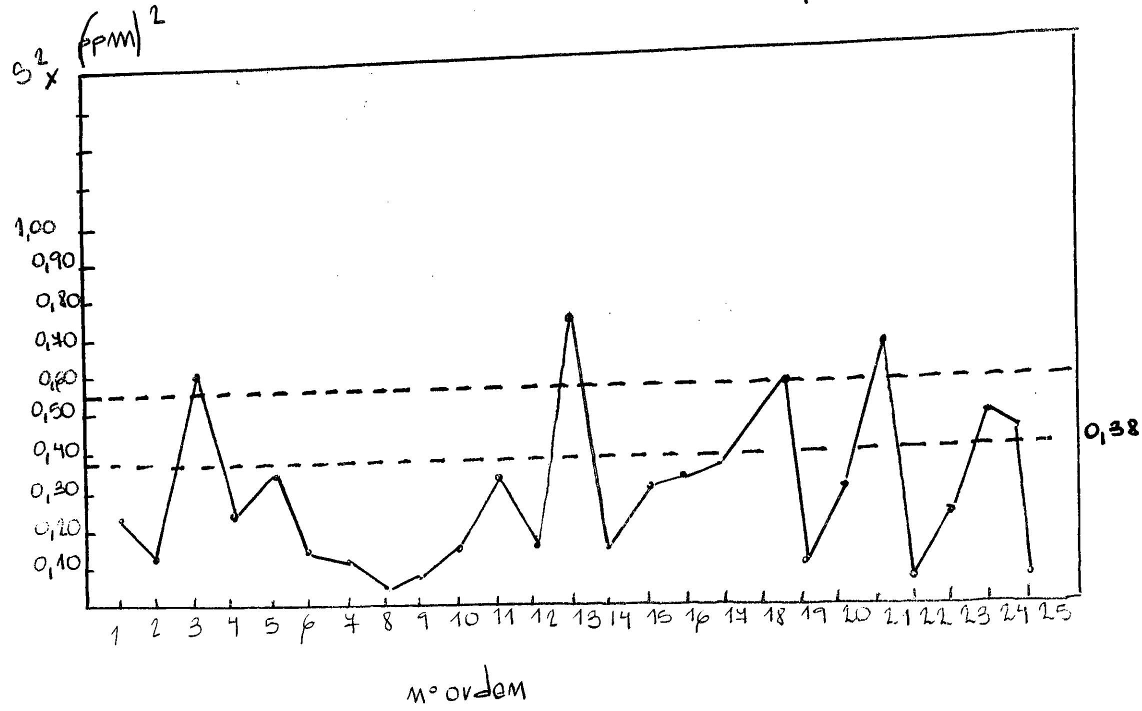


conclusión = en cuanto a la estabilidad del método es coherente, presenta buena aleatoridad, el día 10 y 18 se encontró que el promedio de concentración de aflatoxinas en cereales alcanzó límites de acción pero los días posteriores al suceso el proceso se normalizó

Analisis presición

Límite superior alerta $\longrightarrow \dfrac{\chi^2_{n-1;\,0{,}95} \cdot \sigma_x^2}{n-1} = \dfrac{9{,}49 \cdot (0{,}4)^2}{4} = 0{,}38\ ppm^2$

" " acción $\longrightarrow \dfrac{\chi^2_{n-1;\,0{,}99} \cdot \sigma_x^2}{n-1} = \dfrac{13{,}3\,(0{,}4)^2}{4} = 0{,}53\ ppm^2$



conclusión = el método elisa tiene una presición normal, no esta fuera de control, en base a los ~~#~~ resultados de la variancia pudimos observar que el método sigue siendo preciso aunque los días 3, 13, 21 haya crecido la misma velozmente, porque los días posteriores al suceso se normalizó

20)- x: altura varones (cm).

$x\,(u_x = 140cm;\ \sigma_x = 7\ cm)$.

$n = 140$.

$\bar{x} \xrightarrow[x' T.L.C.]{} \left(u_x = 140cm;\ \dfrac{\sigma_x}{\sqrt{n}} = \dfrac{7cm}{\sqrt{140}}\right)$

$\left(u_x = 140;\ 0{,}592cm\right)$.

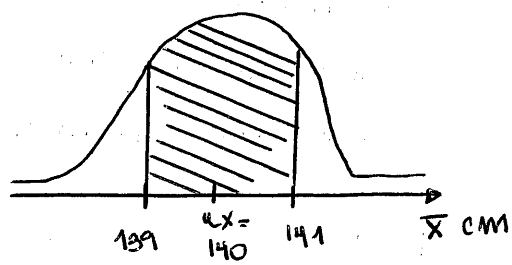
$P\left(139cm < \bar{x} < 141cm\right) =$

$P\left(\bar{x} < 141cm\right) - P\left(\bar{x} < 139cm\right) =$

$P\left(\dfrac{\bar{x}-u_x}{\sigma_x/\sqrt{n}} < \dfrac{141-140}{0{,}592}\right) - P\left(\dfrac{\bar{x}-u_x}{\sigma_x/\sqrt{n}} < \dfrac{139-140}{0{,}592}\right) =$

$P\left(Z < 1{,}69\right) - P\left(Z < -1{,}69\right) =$

$0{,}9545 - 0{,}0455 = 0{,}909$.



21) x: medidas celulares (um).

* errores aleatorios

$x \longrightarrow N\left(u_x = 0\,um;\ \sigma_x^2 = 4um^2\right)$.

$n = 16$

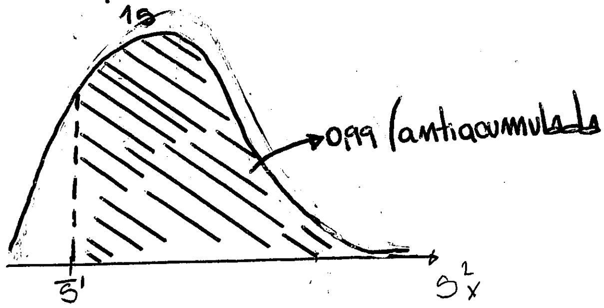
$S_x^2 \longrightarrow N\left(u_x = 0\,um;\ \underbrace{\sigma_x^2}\ = 4um^2\right)$

↳ en la fórmula no se eleva al cuadrado

$P\left(S_x^2 > s'\right) = 0{,}99 \Rightarrow P\left(S_x^2 < s'\right) = 0{,}01$

$S_x^2 \longrightarrow \dfrac{\chi^2_{0{,}99;\,15} \cdot 4um^2}{15}$

$P\left(S_x^2 < \dfrac{\chi_{0{,}01;\,15} \cdot 4um^2}{15}\right) =$



$S_x^2 = \dfrac{\chi_{0{,}01;\,15}) \cdot 4um^2}{15}$

$\dfrac{5{,}23 \cdot 4um^2}{15} = 1{,}39\ um^2$.

Unidad n° 5 - estimación de parámetros.

x = variable aleatoria cuya distribución depende de un parámetro desconocido σ

"característica poblacional"

$x_1; x_2; x_3; x_4 \ldots x_n$ muestra aleatoria de la variable x.

Definimos estimador como "nueva variable aleatoria, función de las n variables que componen la muestra".

le asigna un valor real a c/u de los vectores de la variable

$g(x_1, x_2, x_3, x_4, \ldots, x_n)$.

Ejemplo = si el parámetro poblacional desconocido es "μ_x", entonces un buen estimador del promedio poblacional es la media aritmetica

$$g(x_1; x_2; x_3; x_4 \ldots x_n) = \frac{x_1 + x_2 + x_3 + x_4 + \ldots x_n}{n} = \bar{x}$$

* pero ¿por qué hablamos de buenos o malos estimadores?

→ Propiedades de un buen estimador

$E(\hat{\sigma}) = \sigma$

esperanza del estimador

* $E(\bar{x}) = \mu_x$ * $E(S_x^2) = \sigma_x^2$ * $E(\hat{p}) = P$.

consistente = (si el estimador es insesgado)

$V(\hat{\sigma}) \xrightarrow[n \to \infty]{} 0.$

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

$V(S_x^2) = \frac{2 \cdot \sigma_x^4}{n-1}$ → si $n \to \infty$, entonces $V \to 0$

$V(\hat{p}) = p \cdot q / n$

nos seguimos preguntando... ¿qué es una estimación?

* una estimación se define como el valor observado del estimador luego de analizar el muestreo.

¿qué formas de estimación hay?

puntual =
1- definimos el parámetro a estimar
2- elegir un buen estimador
③ obtener la muestra, a partir de la cual se obtiene el valor observado del estimador, osea, estimación

por intervalo de confianza
encontrar dos números "a" y "b" en donde $a<b$, a partir de la muestra dada, tal que el intervalo de confianza $(a;b)$ cubre el valor del parámetro con una confianza de $(1-\alpha)$.

cuánto mayor confianza hay, mejor
=
☺

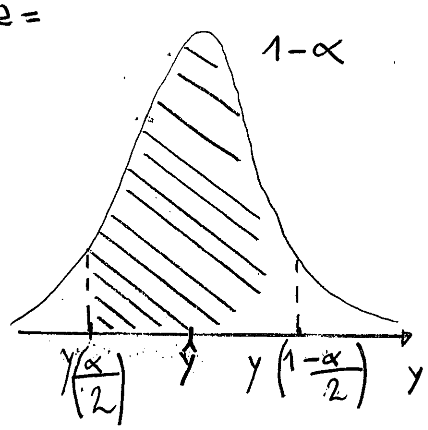
ventajas =
margen de error
semiamplitud del intervalo.

medida objetiva del grado de confianza $(1-\alpha)$.

- Construcción de un intervalo de confianza.

* definir el parámetro a estimar.
* seleccionamos buen estimador.
* fijamos el grado de confianza $(1-\alpha)$.
* definimos una variable "y" que contenga al parámetro a estimar, al estimador y cuya distribución de la probabilidad sea conocida. (si en la expresión interviene otro parámetro el mismo debe conocerse).
* Sobre el eje de variación "y" se eligen dos valores tal que =

$$P\left(y_{\frac{\alpha}{2}} < y < y_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$



* despejamos el valor del parámetro
* obtener muestra del cual se obtiene "a" y "b"

- Siempre concluimos que con un grado de confianza de $(1-\alpha)$, el intervalo $(a;b)$ cubre el valor del parámetro verdadero.

Deducciones de los intervalos de confianza.

- para el promedio de una población normal con varianza conocida.

$- X \longrightarrow \underbrace{N(\overset{?}{u_x}; \sigma_x^2)}$, $\xrightarrow{}$ podemos estandarizar p' una z $\quad \dfrac{\bar{X} - u_x}{\sigma_x / \sqrt{n}} = Z \longrightarrow N(0; 1)$.

lo que quiero estimar

$- \theta = u_x = E(\bar{x})$.

$- \hat{\theta} = \bar{X} = \dfrac{1}{n} \cdot \sum_{i=1}^{n} x_i$

- grado de confianza $1-\alpha$

demostramos =

$$P\left(Z_{\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}} \right) = 1-\alpha$$

x' simetría de la distribución =

$$P\left(-Z_{1-\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}} \right) = 1-\alpha$$

puedo presentarlo como su forma estandarizada.

despejo el parámetro

$$P\left(-Z_{1-\frac{\alpha}{2}} < \underbrace{\frac{\bar{x} - u_x}{\sigma_x / \sqrt{n}}} < Z_{1-\frac{\alpha}{2}} \right) = 1-\alpha$$

despejas p' ambos lados de la igualdad

$$P\left(-\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} < -u_x < -\bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} \right) = 1-\alpha$$

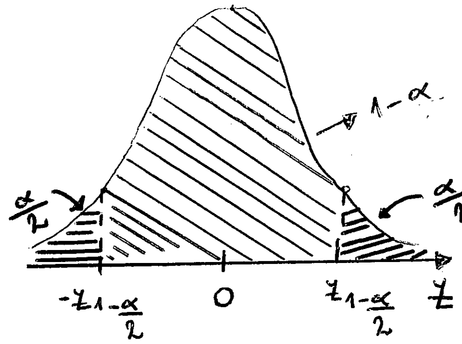
$$P\left(\bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} > \underbrace{u_x} > \bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} \right) = 1-\alpha$$

invierto de lugar (sig paso) x' una cuestión de comodidad visual

$$P\left(\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} < u_x < \bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} \right) = 1-\alpha$$

tengo dos valores más o menos $\bar{X}$ que se calcula mediante la muestra

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} = (a;b) \quad a<b;$$



- promedio de una población normal con varianza desconocida.

* $X \longrightarrow N(u_x; ?)$

$\underbrace{u_x}_{\text{lo q' quiero estimar}}$

* $\theta = u_x = E(\bar{X})$
* $\hat{\theta} = \bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$
* g. confianza $= 1-\alpha$

deducimos

$$t_{n-1} = \frac{\bar{X} - u_x}{s_x/\sqrt{n}}$$

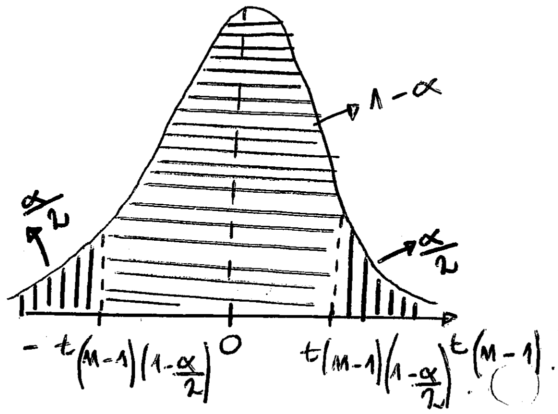
deducción distribución de probabilidad p/ estadísticas capítulo 4

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

por simetría de la distribución

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

↓ conozco su distribución de probabilidad



$$P\left(-t_{(n-1);\left(1-\frac{\alpha}{2}\right)} < \underbrace{\frac{\bar{X} - u_x}{s_x/\sqrt{n}}}_{\text{despejamos p' ambos lados de la igualdad}} < t_{(n-1);\left(1-\frac{\alpha}{2}\right)}\right) = 1-\alpha$$

$$P\left(-\bar{X} - t_{(n-1);1-\frac{\alpha}{2}} \cdot \frac{s_x}{\sqrt{n}} < -u_x < -\bar{X} + t_{(n-1);\left(1-\frac{\alpha}{2}\right)} \cdot \frac{s_x}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\bar{X} + t_{(n-1);1-\frac{\alpha}{2}} \cdot \frac{s_x}{\sqrt{n}} > u_x > \bar{X} - t_{(n-1);\left(1-\frac{\alpha}{2}\right)} \cdot \frac{s_x}{\sqrt{n}}\right) = 1-\alpha$$

Unidad nº 5 = "estimación de parámetros".

x: variable aleatoria cuya distribución depende de un parámetro desconocido "θ". ← característica poblacional

$x_1, x_2, x_3, \ldots, x_i, \ldots, x_n$ una muestra aleatoria de la variable "x".

Estimador = nueva variable aleatoria, función de las "n" variables que componen la muestra aleatoria

$g(x_1; x_2; x_3; \ldots; x_i; \ldots; x_n)$ → Nueva variable, le asigna nº real a cada vector de la variable.

$\neq$

Ejemplo: $\underset{\text{parámetro}}{\theta} = \mu_x \Longrightarrow \underset{\text{estimador}}{\hat{\theta}} = \bar{x}$

$$g(x_1; x_2; x_3; \ldots; x_i; \ldots; x_n) = \frac{x_1 + x_2 + x_3; \ldots; x_i; \ldots; x_n}{n}$$

Estimación = valor observado del estimador luego de realizar el muestreo.

Ejemplo: $\theta = \mu_x$ = cc. promedio del analito en individuos sanos.

valores observados de la cc. en 40 individuos sanos (mg/dL).

1,9; ...; 2,5; ...; 2,1

estimación del promedio poblacional "$\bar{x}$"

$\bar{x} = 2{,}3$ mg/dL.

Propiedades de un buen estimador.

* no viciado o insesgado $E(\hat{\theta}) = \theta$

$E(\bar{x}) = \mu_x$; $E(\underset{\text{varianza muestral}}{s_x^2}) = \sigma_x^2$ → varianza poblacional ; $E[h(A)] = p(A)$.

$$V(\hat{\theta}) \xrightarrow[n \to \infty]{} 0$$

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

si $n \to \infty$, $V(\bar{x})$ tiende a 0.

$$V[h(A)] = \frac{p \cdot q}{n}$$

cuando $n \to \infty$ la varianza tiende a 0.

$$V(s_x^2) = \frac{2 \cdot \sigma_x^4}{n-1}$$

cuando $n \to \infty$ la varianza tiende a 0.

Formas de estimar $\longrightarrow$ puntuales =

1. definir el parámetro puntual
2. seleccionar buen estimador
3. Obtener muestra, a partir de la cual se obtiene valor observado del estimador, o sea, estimación

Formas de estimar $\longrightarrow$ por intervalo de confianza =

cuánto ↑ confianza, mejor.

Encontrar dos nº a y b, $(a<b)$ a partir de la muestra dada, tal que el intervalo $(a;b)$ cubre el valor del parámetro con una confianza $(1-\alpha)$.

ventajas:

- medida objetiva del grado de confianza $1-\alpha$. ♡
- margen de error = semiamplitud intervalo.

* construcción intervalo de confianza

1. definir parámetro a observar
2. seleccionar buen estimador
3. Fijar el grado de confianza $\longrightarrow$ "α"
4. definir variable "y" que contenga el parámetro a estimar, al estimador y cuya distribución de probabilidad sea conocida. (si en la expresión interviene otro parámetro también debe conocerse.

$$P\left(\frac{S^2_x \cdot (n-1)}{\chi^2_{(n-1);\frac{\alpha}{2}}} > \sigma_x^2 > \frac{S^2_x \cdot (n-1)}{\chi_{(n-1);1-\frac{\alpha}{2}}}\right) = 1-\alpha$$

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

Expresión general p' intervalos $= \left(\underbrace{\frac{S^2_x\,(n-1)}{\chi^2_{(n-1)\left(1-\frac{\alpha}{2}\right)}}}_{} \; ; \; \underbrace{\frac{S^2_x\,(n-1)}{\chi^2_{(n-1)\left(\frac{\alpha}{2}\right)}}}_{}\right)$

más grande que $\Rightarrow \underbrace{a}_{} < b$

x' que el numerador es mayor

* <u>Estimación</u> por intervalo de confianza de proporción <u>p' muestras grandes</u>. $(n > 30)$.

- $\theta = P(A) = E[h(A)]$
- $\hat{\theta} = h(A)$
- $n > 30$ x'. TLC se considera que se distribuye $\cong$ normal
- g. confianza $= 1-\alpha$

$$\frac{h-p}{\sqrt{\frac{p \cdot p}{n}}} = Z \longrightarrow N(0;1).$$

$\downarrow$ aproximo $p \cong h$;

$\Rightarrow q = 1-p \cong 1-h \Rightarrow Z \cong \dfrac{h-p}{\sqrt{\frac{h.(1-h)}{n}}}$

$$P\left(Z_{\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

x' simetría de la Z

$$P\left(-Z_{1-\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$\downarrow \cong$

despejo en función del parámetro

$$P\left(-Z_{1-\frac{\alpha}{2}} < \frac{h-p}{\sqrt{\frac{h.(1-h)}{n}}} < Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\bar{x} - t_{(n-1),\left(1-\frac{\alpha}{2}\right)} \cdot \frac{S_x}{\sqrt{n}} < \mu_x < \bar{x} + t_{(n-1),\left(1-\frac{\alpha}{2}\right)} \cdot \frac{S_x}{\sqrt{n}}\right) = 1-\alpha$$

Expresión general = $\boxed{\bar{x} \pm t_{(n-1);\left(1-\frac{\alpha}{2}\right)} \cdot \frac{S_x}{\sqrt{n}}}$

* Estimación por intervalo para la varianza de una población normal

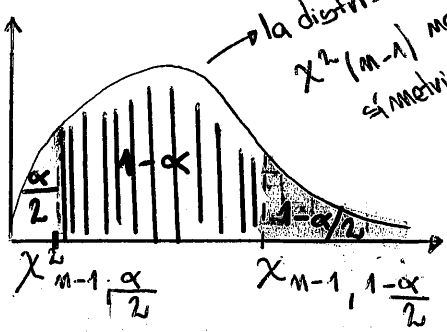
$$X \longrightarrow N(\mu_x ; \underbrace{\sigma_x^2}_{\text{parámetro a estimar}})$$

$$\theta = \sigma_x^2 = E(S_x^2)$$

$$\hat{\theta} = S_x^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

g. confianza = $1-\alpha$

$$\chi^2_{(n-1)} = \frac{S_x^2 \cdot (n-1)}{\sigma_x^2}$$



la distribución $\chi^2(n-1)$ no es simétrica

deducimos =

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

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

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

objetivo es despejar el parámetro

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

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

(-1)

invierto signo de la desigualdad ← Al multiplicar por (-1) invierto numerador y denominador

$$P\left(-Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h.(1-h)}{n}} < h-p < Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h.(1-h)}{n}}\right) = 1-\alpha$$

$$P\left(-h - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h.(1-h)}{n}} < \ominus p < -h + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h.(1-h)}{n}}\right) = 1-\alpha$$

$$P\left(h + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h.(1-h)}{n}} > p > h - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h.(1-h)}{n}}\right) = 1-\alpha$$

$$P\left(h - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h(1-h)}{n}} < p < h + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h.(1-h)}{n}}\right) = 1-\alpha$$

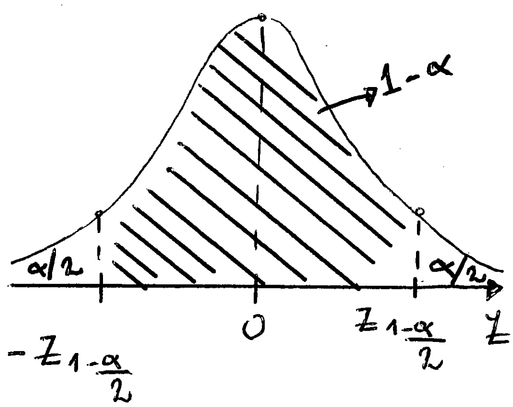
Expresión general = $\boxed{h \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h(1-h)}{n}}}$

por aproximación
$(h \cong p)$ puedo obtener en un intervalo
un n° < 0 ó un n° > 1, por ejemplo

$\left.\begin{array}{l} -0{,}1 \cong 0 \\ 1{,}01 \cong 1 \end{array}\right\}$ siempre aproximo

recordar qué la probabilidad de un
suceso siempre

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



-Cálculo de extensión de una muestra

- para el cálculo del promedio con grado de confianza y presición determinada.

mayor g. confianza, mayor "n" para "d" fijos.

$$n = \left(\frac{Z_{1-\alpha/2} \cdot \sigma_x^2}{d}\right)^2 \longrightarrow (*)$$

mayor precisión de la estimación, mayor n. (p i α fijo).

se mide x i semiamplitud del intervalo (d)

↑precisión ⟹ ↓d = ↑n

*cálculo de la estimación de una proporción con grado de confianza y $\underbrace{\text{semiamplitud}}_{=\text{precisión}}$ dada.

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} \cdot p' \cdot q'}{d^2} \qquad \begin{cases} p' = h \\ q' = 1-h \end{cases} \quad \text{despejas} = \quad d = Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h.(1-h)}{n}}$$

$$\frac{d}{Z_{1-\frac{\alpha}{2}}} = \sqrt{\frac{p' \cdot q'}{n}}$$

$$\frac{d^2}{Z^2_{1-\frac{\alpha}{2}}} = \frac{p' q'}{n}$$

(*)

$$\cancel{Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} = d}$$

$$\cancel{Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{n} = d^2}$$

¡mal despejado!

≈.

$$\cancel{\frac{Z_{1-\alpha/2} \cdot \sigma}{d^{2}} = n}$$

consultar

$$\boxed{n = \frac{p' q' . Z^2(1-\alpha/2)}{d^2}}$$

*calidad ⇝ 1) ↑ confianza $= (1-\alpha)\uparrow$

2) menor semiamplitud intervalo

↑precisión; ↓d ; ↑n.

$$Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} = d$$

$$Z_{1-\frac{\alpha}{2}} \cdot \sigma_x = d.\sqrt{n}$$

$$Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{d} = \sqrt{n} \Longrightarrow \left(\frac{Z_{1-\alpha/2} \cdot \sigma_x}{d.}\right)^2 = n.$$

5. Sobre el eje de variación "y" se eligen dos valores tal que

$$P\left(y_{\alpha/2} < y < y_{\frac{1-\alpha}{2}}\right) = 1-\alpha$$

→ final demostración de los ≠ intervalos.

6. Despejar valor del parámetro

7. Obtener muestra de la cual se calcula el intervalo $(a;b)$.

→ se concluye que con un grado de confianza "$1-\alpha$"; el intervalo $(a;b)$ cubre el valor del parámetro.

Cálculo de extensión muestra.

⇒ p/ estimación de un promedio con g. confianza y presición determinada

$$n = \left(\frac{z_{1-\alpha/2} \cdot \sigma_x}{d}\right)^2$$

→ a mayor heterogeneidad, mayor "n" (p/ α y d fijos).

mayor g. confianza, mayor "n" para "d" fijos.

mayor precisión de la estimación, mayor n (p/ α fijos)

precision medida por semiamplitud del intervalo (d).

p/ la estimación de una proporción con grado de confianza "α" y precisión determinada

$$n = \frac{z^2_{1-\alpha/2} \cdot p' \cdot q'}{d^2}$$

Ejercicio N° 3

$\theta = \mu_x \implies \hat{\theta} = \bar{X}$

μ_x: rendimiento promedio de nueva variedad de maíz en Santa Fe sur.

a). Tomaría una muestra aleatoria de tamaño $n = \left(\frac{Z_{1-\alpha/2} \cdot \sigma}{d} \right)^2$, definiendo previamente el α y el intervalo de confianza con el cual se desea trabajar.

b). valores obtenidos de rendimiento (qq/ha).

88 85 74 69 93 70 87 65 89 79 93 84

Estimaciones puntuales.=

1. definimos el parametro en estudio

$\hat{\theta} = \mu_x$: rendimiento promedio nueva variedad de maíz en Santa Fe sur.

2. seleccionamos buen estimador

$\hat{\theta} = \bar{X}$ → $E(\bar{X}) = \mu_x$; no sesgado
→ $V(\bar{X}) = \frac{\sigma_x^2}{n}$; consistente.

3. Obtenemos la muestra (rendimiento de 12 parcelas (qq/ha)).

$\bar{X} = 81,33$ (qq/ha).

Conclusión = en base a la muestra de tamaño 12 obtenida se concluye que el valor promedio de rendimiento para la nueva variedad de maíz en el sur de Santa Fe es de aproximadamente 81,33 (qq/ha).

Estimación por intervalo de confianza

mivel confianza = $1-\alpha$ $\qquad \alpha = 5\% = 0{,}05$:

mivel confianza = 0,95 = 95%.

$$\bar{x} \pm t_{\left(n-1;\ \cancel{\frac{1-\alpha}{2}}\right)}^{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

$$81{,}33\,(qq/ha) \pm t\left(11;\ \overset{0{,}975}{\cancel{0{,}475}}\right) \cdot \frac{9{,}698}{\sqrt{12}}$$

$$a = 81{,}33\,(qq/ha) - \overset{2{,}20}{\cancel{0{,}129}} \cdot \frac{9{,}698}{\sqrt{12}} \longleftrightarrow b = 81{,}33\,(qq/ha) + \overset{2{,}20}{\cancel{0{,}129}} \cdot \frac{9{,}698}{\sqrt{12}}$$

a = ~~80,969 (qq/ha)~~ 75,171 (qq/ha) $\qquad$ b = ~~81,691 (qq/ha)~~ 87,4891 (qq/ha).

Conclusión = con un nivel de confianza del 95% se concluye que el verdadero valor promedio de rendimiento de la nueva variedad de maíz en el sur de Sta. Fe se encuentra comprendido en el intervalo de ~~(80,969 ; 81,691)~~ (qq/ha). 75,171 ; 87,4891 (qq/ha).

Ejercicio N° 4.

n = 20 (n° de alícuotas medidas).

x = concentración de sodio de una solución. (mg/L).

a) completamos tabla 5.1 (pág 81). tener cuidado interpretación de la pregunta

$\underbrace{g.c.}$ = 90% $\qquad \sigma_x = 0{,}2$ mg/L

$\underline{1-\alpha} = 0{,}9$

(estoy averiguando sobre 1 alícuota particular).

Muestra 4 = $\longrightarrow \bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Longrightarrow 1{,}863\ mg/L \pm 1{,}645 \cdot \frac{0{,}2\ mg/L}{\sqrt{20}}$?

$(1{,}753\ ;\ 1{,}973)\ mg/L$.

Muestra 15 = $\longrightarrow \bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \sigma/\sqrt{n} \Longrightarrow 1{,}909\ mg/L \pm 1{,}645 \cdot \frac{0{,}2\ mg/L}{\sqrt{9}}$

$(1,499\ ;\ 2,019)\ mg/L$

c). $u_x = 2\ mg/L$

Fig. 5a).

p.p intervalos que lo cubren el verdadero valor de la cc. de sodio en la solución es de

$\frac{18}{20}\ \left(\frac{\cancel{16}}{\cancel{18}}\right) =$ [illegible].99

$$100\% \text{ ——— } \overset{20}{\cancel{18}} \text{ alícuotas}$$

$$\boxed{90\% \text{ ——— } \overset{18}{\cancel{16}} \text{ alícuotas}}$$

en el gráfico se muestra que $\overset{18}{\cancel{16}}$ intervalos de confianza contienen el verdadero valor promedio de sodio en la solución.

La alícuota n° 4 y n° 2, en el intervalo de confianza establecido no contienen el verdadero valor promedio de sodio de la solución.

d) alícuota "4" $\xrightarrow{\sigma \text{ desconocido}}$ $\bar{x} \pm t_{8\ ;\ 0,95} \cdot \frac{0,248\ mg/L}{\sqrt{9}} = 1,863\ \frac{mg}{L} \pm 1,86 \cdot \frac{0,248\ mg/L}{\sqrt{9}}$

$(1,709\ ;\ 2,016)\ mg/L$.

alícuota "15" ⟶ $\bar{x} \pm t_{8\ ;\ 0,95} \cdot \frac{0,139\ mg/L}{\sqrt{9}} = 1,909\ mg/L \pm 1,86 \cdot \frac{0,139\ mg/L}{\sqrt{9}}$

$(1,823\ ;\ 1,995)\ mg/L$

Ejercicio 5.

X = valor promedio As. (mg/L).

$\sigma_x = 0{,}04$ mg/L.

a) $\alpha = 0{,}02$ → representa d.

$n = ?$

M.C = 99 %

$1-\alpha = 0{,}99$

$\alpha = 0{,}01$

$Z_{0{,}995}$.

$$n = \left(\frac{Z_{1-0{,}01/2} \cdot 0{,}04 \text{ mg/L}}{0{,}02}\right)^2$$

$$n = \left(\frac{Z_{0{,}995} \cdot 0{,}04 \text{ mg/L}}{0{,}02}\right)^2 =$$

$$n = \left(\frac{2{,}5760 \cdot 0{,}04 \text{ mg/L}}{0{,}02}\right)^2$$

$$\boxed{n = 27}$$

b). $\bar{X} = 0{,}059$ mg/L $\quad n = 30$.

$$\bar{X} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} \Longrightarrow 0{,}059 \frac{\text{mg}}{\text{L}} \pm Z_{0{,}995} \cdot \frac{0{,}04 \text{ mg/L}}{\sqrt{30}}$$

$$(0{,}040 ; 0{,}079) \text{ mg/L}$$

Ejercicio n° 6.

X = capacidad antioxidante de frambuesas (ug Fe/g peso seco).

$\bar{X} = $ se estima q' el verdadero promedio de FRAP es 15,51 ug Fe/g

a) $\bar{X} = 15{,}51$ (ug Fe/g peso seco)

* se estima q' el verdadero promedio de FRAP es 15,51 ug Fe/g y la verdadera variancia 8,11 ug Fe/g.

$\underline{\underline{S}} = 2{,}85$ (ug Fe/g peso seco) ⟶ $S_x^2 = 8{,}11$ ug Fe/g peso seco.

↓ desvío

b). $X \longrightarrow N(\underbrace{u_x}_{\text{estimaremos esto x' intervalo confianza}} ; \sigma_x^2)$

M.C = 99 %

$$\bar{X} \pm T_{13;0{,}995} \cdot \frac{S}{\sqrt{n}} \Longrightarrow 15{,}51 \text{ ug Fe/g seco} \pm 3{,}01 \cdot \frac{2{,}85 \text{ ug Fe/g}}{\sqrt{14}}$$

$$(13{,}21 ; 17{,}8) \text{ ug Fe/g}.$$

Conclución = con un nivel de confianza del 99% se concluye que el verdadero promedio de FRAP en frambuesas salvajes se encuentra comprendido en $(13,21 ; 17,8)$ ug Fe/g

c) - $M.C = 95\%$ $\qquad \sigma_x^2 \Longrightarrow \hat{S}_x^2$ } buen estimador

$1-\alpha = 0,95.$

$\alpha = 0,05.$

$$\left(\frac{(n-1) \cdot S_x^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} ; \frac{(n-1) \cdot S_x^2}{\chi^2_{n-1;\frac{\alpha}{2}}} \right)$$

$$\left(\frac{13 \cdot 8,11 \ ug\,Fe/g^2}{\chi^2_{13;0,975}} ; \frac{13 \cdot 8,11 \ ug\,Fe/g^2}{\chi^2_{13;0,025}} \right)$$

$$\left(\frac{13 \cdot 8,11 \ ug\,Fe/g^2}{24,7} ; \frac{13 \cdot 8,11 \ ug\,Fe/g^2}{5,01} \right)$$

$$(4,27 ; 21,043) \ ug\,Fe/g^2$$

Conclusión = con un nivel de confianza del 95% se concluye que el verdadero valor de variancia poblacional de el contenido de Frap en frambuesas salvajes está comprendido entre $(4,27 ; 21,043)$ ug Fe/g^2.

d) $15,51 \ ug\,Fe/g \pm \underbrace{3,01 \cdot \frac{2,85 \ ug/Fe}{\sqrt{14}}}_{d}$ $\qquad M C = 99\%$

$$d = \frac{L_{sup} - L_{inf}}{2}$$

$$d = \frac{17,8 \ ug\,Fe/g - 13,22 \ ug\,Fe/g}{2}$$

$\underline{\underline{d = 1,145}} \Longrightarrow$ $\qquad n = \left(\frac{Z_{1-\frac{0,01}{2}} \cdot \sigma_x'}{d} \right)^2$

$\sigma_x \simeq S_x = 2,8499$ $\qquad n = \left(\frac{Z_{0,995} \cdot 2,8499}{1,145} \right)^2$

$n \simeq 41$ se necesitan 41 frambuesas aumentar la precisión en la estimación del verdadero valor promedio de Frap en frambuesas salvajes

Ejercicio Nº 9

$g.c = 95\%$

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

$\alpha = 0{,}05.$

$\theta = \sigma_x^2 \Longrightarrow \hat{\theta} = S_x^2$ varianza muestral

determino presicion método

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

$S_x^2 = 6{,}74 \ mg/100g^2$

$n = 10$

$$\left(\frac{(9) \cdot 6{,}74\,(mg/100g)^2}{\chi^2_{10\,;\,0{,}975}} \;;\; \frac{(9) \cdot 6{,}74\,(mg/100g)^2}{\chi^2_{9\,;\,0{,}025}} \right)$$

$$\left(\frac{9 \cdot 6{,}74}{19} \;;\; \frac{9 \cdot 6{,}74}{2{,}70} \right) (mg/100g)^2$$

$$(3{,}21 \;;\; 22{,}6)\,(mg/100g)^2$$

Propuestos.

ejercicio 10.

$\theta =$ pp. aves contaminadas en un parque natural

$n.c = 0{,}90$

$\alpha = 0{,}10$

$\theta = P(A)$

$\hat{\theta} = h(A)$

$I.C = (0{,}14 ; 0{,}26).$

$0{,}14 = h - z_{0{,}95} \cdot \sqrt{\dfrac{h \cdot (h-1)}{95}}$

* para averiguar el valor de h

$= \dfrac{0{,}14 + 0{,}26}{2} = 0{,}20.$

$d = \dfrac{0{,}26 - 0{,}14}{2} = 0{,}06$

consultar

$$n = \left(\frac{z_{1-\alpha/2} \sqrt{p \cdot q}}{d \to Ls - Li} \right)^2 \to p = q = 0{,}5$$

$$n = \left(\frac{z_{0{,}95} \cdot 0{,}71}{0{,}12} \right)^2$$

$$n = \left(\frac{1{,}645 \cdot 0{,}71}{0{,}12} \right)^2$$

$n = 95$

b) aumentando "n" para que d sea más pequeño, o disminuyendo el nivel confianza ejercicio N°4. =

X = pp. personas hábitos alimenticios inadecuados.

P(A) = personas con hábitos inadecuado de alimentación

$P(\bar{A})$ = personas con hábitos adecuados de alimentación

$\theta = P(A)$ $\hat{\theta} = h(A) \longrightarrow E(h[A]) = P(A)$ insesgado.

$V(h[A]) \xrightarrow[n \to \infty]{} 0$ consistente.

$\dfrac{p \cdot q'}{\sqrt{n}}$ tiende a 0. ($n \to \infty$)

b) n =

N.C = 95%

$\alpha = 4,00\% = 0,04$.

$P(A) = \dfrac{12}{40} = 0,3$

$P(\bar{A}) = 0,7$.

$$n = \left(\frac{Z_{1-\frac{0,05}{2}} \cdot \sqrt{0,3 \cdot 0,7}}{0,04} \right)^2$$

$$n = \left(\frac{1,96 \cdot 0,46}{0,04} \right)^2$$

$\boxed{n = 504}$

c) n = 500

P(A) = 124/500.

P(A) = 0,248.

$$h \pm Z_{1-\frac{0,05}{2}} \cdot \sqrt{\frac{h \cdot (1-h)}{n}}$$

$$0,248 \pm 1,96 \cdot \sqrt{\frac{0,248 \cdot (1-0,248)}{500}}$$

$(0,21 ; 0,29)$.

* conclusión = con un grado de confianza del 95% se concluye que el valor verdadero de pp de personas con malos hábitos alimenticios

ejercicio 11.

6

x: ingesta diaria de vitamina D en escolares.

$\cancel{\bar{x} \longrightarrow N\left(4{,}5\,ug;\ \frac{2{,}02\,ug}{\sqrt{32}}\right)}$ $\quad x \longrightarrow N(\mu x = ?\ ;\ \sigma x = ?)$.

$n = 32$ $\quad \bar{x} = 4{,}5\,ug \quad S_x = 2{,}02\,ug$

a) g.c = 95% $\quad \alpha = 1 - 0{,}95$
$\alpha = 0{,}05$.

$\longrightarrow$ desvio poblacional incognita

$$4{,}5\,ug \pm \underset{\cancel{Z_{1-\frac{0{,}05}{2}}}}{T_{n-1;\,1-\frac{\alpha}{2}}} \cdot \frac{2{,}02\,ug}{\sqrt{32}} \Longrightarrow 4{,}5\,ug \pm \underset{\cancel{Z_{0{,}975}}}{T_{31;\,1-0{,}05/2}} \cdot \frac{2{,}02\,ug}{\sqrt{32}} = \cancel{(3{,}80;\ 5{,}19)\,ug}\ (3{,}77;\ 5{,}23)\,ug.$$

$T_{31;\,1-0{,}05/2} \cong T_{30;\,1-\frac{0{,}05}{2}}$

$4{,}5\,ug \pm 2{,}04 \cdot 2{,}02\,ug/\sqrt{32} =$

conclusión: con un nivel de confianza de 95% se estima que el valor promedio real de ingesta diaria de vitamina D esta comprendido entre ~~(3,80; 5,19) ug.~~ $(3{,}77;\ 5{,}23)\,ug$.

b) $d = 0{,}3\,ug$

$$n = \left(\frac{Z_{1-\alpha/2} \cdot \sigma_x}{d}\right)^2$$

$S_x \cong 2{,}02\,ug$ por ser un buen estimador de σ_x.

$$n = \left(\frac{Z_{0{,}975} \cdot 2{,}02\,ug}{0{,}3\,ug}\right)^2$$

$$n = \left(\frac{1{,}96 \cdot 2{,}02\,ug}{0{,}3\,ug}\right)^2 = 175.$$

Ejercicio 15.

$\boxed{p' = q' = 0{,}5.}$

a) $N.C = 95\% = 0{,}95$

$P(A)$ = prob. de pacientes afectados por error en medicación.

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

$\alpha = 0{,}05.$

A = pacientes afectados por error en medicación.

$d = 0{,}3$

$$n = \left(\frac{Z_{1-0{,}05/2} \cdot \sqrt{p \cdot q'}}{0{,}03} \right)^2$$

$$n = \left(\frac{1{,}96 \cdot 0{,}5}{0{,}03} \right)^2$$

$\boxed{n = 1068}$

b) $n = 1068$ pacientes

nº pacientes enfermos = 188

$h = 188 / 1068$

$h = 0{,}176$

$$h \pm Z_{1-0{,}05/2} \cdot \sqrt{\frac{h(1-h)}{1068}} \Longrightarrow$$

$$= 0{,}176 \pm 1{,}96 \cdot \sqrt{\frac{0{,}176(1-0{,}176)}{1068}}$$

$$= (0{,}1532 ; 0{,}1988)$$

conclusión = con un nivel de confianza del 95% se concluye que el verdadero porcentajes de pacientes afectados por error en la medicación está comprendido entre $(0{,}1532 ; 0{,}1988)$.

$= (15{,}32\% ; 19{,}88\%)$.

Ejercicio 14 $\quad x$ = cc. de HDL colesterol (mg/dL)

$\theta = \sigma_x^2 \longrightarrow N(\mu_x ; \sigma_x^2)$

$\hat{\theta} = s_x^2 \longrightarrow \dfrac{\chi^2(n-1) \cdot \sigma_x^2}{(n-1)}$

a) estimación puntual.

$s_x^2 = 4{,}42 \ mg/dL^2$

Se estima que el verdadero valor de la variancia en el método de determinación de cc. de HDL es de $4{,}42 (mg/dL)^2$.

b) estimación por intervalo de confianza.

$n.c = 99\%$

$1-\alpha = 0{,}99$

$\alpha = 0{,}01$

$$\left(\frac{(n-1) \cdot s_x^2}{\chi_{(n-1);1-\frac{\alpha}{2}}} ; \frac{(n-1) \cdot s_x^2}{\chi_{(n-1);\frac{\alpha}{2}}} \right)$$

$$\left(\frac{(9) \cdot 4{,}42 \ mg/dL^2}{\chi_{9;0{,}995}} ; \frac{(9) \cdot 4{,}42 \ mg/dL^2}{\chi_{9;0{,}005}} \right)$$

$$\left(\frac{(9) \cdot 4{,}42 \ mg/dL^2}{23{,}6} ; \frac{(9) \cdot 4{,}42 \ mg/dL^2}{1{,}73} \right)$$

$$(1{,}8 \ ; \ 24{,}55) \ mg/dL^2$$

Ejercicio 15.

x = contenido promedio componente activo en un comprimido.

$x \longrightarrow N(\mu_x ; \sigma_x = 0{,}0068 \ g)$.

$\theta = \mu_x \implies \hat{\theta} = \bar{x}$

a) $d = 0{,}005 \quad n.c = 99\%$

$n = ? \quad 1-\alpha = 0{,}99$

$\alpha = 0{,}01$

$$n = \left(\frac{Z_{1-\frac{0{,}01}{2}} \cdot 0{,}0068 \ g}{0{,}005} \right)^2$$

$$n = \left(\frac{Z_{0{,}995} \cdot 0{,}0068 \ g}{0{,}005} \right)^2$$

$$n = \left(\frac{2{,}5760 \cdot 0{,}0068 \ g}{0{,}005} \right)^2$$

$n = 13.$

b) n.c = 99% $\sigma_x = 0{,}00689$

$1-\alpha = 0{,}99$

$\alpha = 0{,}01$

$$\bar{X} \pm Z_{0{,}995} \cdot \frac{\sigma_x}{\sqrt{n}} =$$

$$0{,}84\,g \pm 2{,}5460\ \frac{0{,}00689}{\sqrt{12}} = \left(0{,}835\,g;\ 0{,}845\,g\right)$$

Con un nivel de confianza del 99% se estima que el verdadero valor promedio de componente activo en los comprimidos se encuentra comprendido entre (0,835; 0,845) g.

Ejercicio 16

$\theta = P(A) \Rightarrow \hat{\theta} = h(A)$

A = prevalencia enfermedad celíaca niños de 3-16 años.

n = 550 niños

$h(A) = \frac{\cancel{7}}{550} = 0{,}0127.$

nivel c. = 99%

$\alpha = 0{,}01$

$$0{,}0127 \pm Z_{0{,}995} \cdot \sqrt{\frac{0{,}012 \cdot (1-0{,}012)}{550}} = \left(0{,}04\%\ ;\ 2{,}5\%\right)$$

$= 0{,}0127 \pm 2{,}5460 \cdot 4{,}77 \times 10^{-3}$

$= 0{,}04\%$;

= con n. confianza del 99% el verdadero porcentaje de niños que padecen celíaquia en edades de 3-16 años cae en (0,04 %; 2,5%).

Unidad nº 6 = ensayo de hipótesis en base a una muestra.

- nos basamos en problemas de "desición" podemos formular "ensayos de hipotesis"

¿Qué es un ensayo de hipótesis? = es un procedimiento estadistico usado para tomar una desición respecto del valor de un parámetro poblacional, en base a la información de una muestra poblacional

Planteamos dos **hipótesis**

- H_0 = hipótesis nula → afirmación sobre una característica poblacional
- H_1 = hipótesis alternativa → afirmación contradictoria a H_0.

siempre involucran parámetros

clasificamos

- simples = establecen un único valor real cómo valor del parámetro
- compuestas = caso contrario (contienen más de una posibilidad).

procedimiento estadistico

Ensayos

- unilateral = se plantea H_1 para un sólo lado (una zona de rechazo).
- bilateral = se plantea H_1 para dos lados. (dos zonas de rechazo)

errores asociados

decisión	situación real: Ho cierta	situación real: Ho falsa
rechazar Ho	error tipo I	potencia
no rechazo Ho		error tipo II.

desición correcta

$P(E_I) = P(\text{rechazar } H_0 / H_0 \text{ cierta}) = \alpha$ → nivel de significación del ensayo.

$P(E_{II}) = P(\text{no rechazar } H_0 / H_0 \text{ falsa}) = \beta$

Potencia $= P(\text{rechazar } H_0 / H_0 \text{ falsa}) = 1 - \beta$.

sucesos complementarios

Regla de decisión

- La decisión se tomará en base a una muestra, con la que construiremos "estadística base del ensayo".

- el procedimiento del ensayo de hipótesis establece una "zona de rechazo de Ho" y una "zona de no rechazo Ho", sobre el eje de variación de la estadística base.

Criterio para establecer zonas. • Fijamos la $P(E_I)$ en un valor fijo $\underline{\alpha}$ y se elige aquella zona de rechazo que haga mínima $P(E_{II})$, es decir maximice la potencia.

α: nivel de significación ensayo

no podemos minimizar dos errores al mismo tiempo

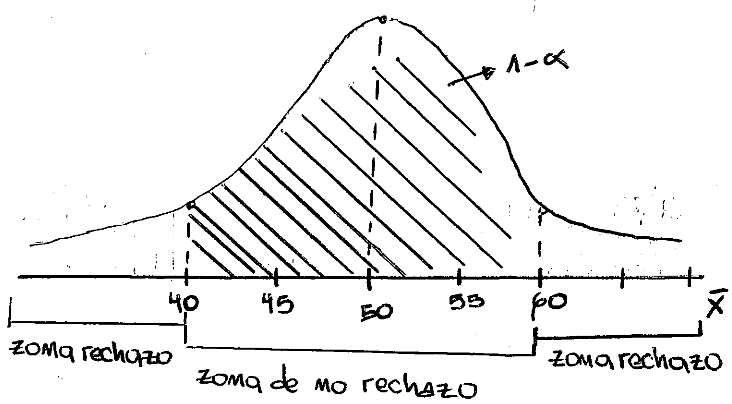
* zona de rechazo = conjunto de valores de la estadística base tal que si el valor observado para la muestra tomada pertenece a dicha región se decide rechazar Ho.

conjunto de valores de la estadística base tal que si el valor observado para la muestra tomada cae en dicha región no rechazo Ho.

Ejemplo:

$H_0)\ u = 50\,mg$ $\qquad H_1)\ u \neq 50\,mg$

* procedimiento: el laboratorio sacará una muestra de 200 comprimidos del fármaco y medirá la concentración de droga en cada uno.



Ensayo de hipótesis para el promedio de una población normal
variancia conocida.

-ensayo unilateral por derecha

$H_0)\ u = u_0$ $\qquad H_1)\ u = u_1 ;\ \ u_1 > u_0$

*estadística básica = $\bar{x}$

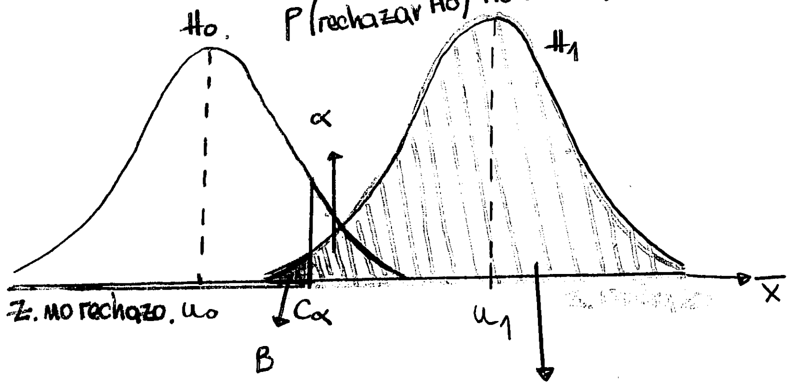
$C\alpha$ = valor crítico por encima del cual rechazo Ho.

$P(E_1) = P(\text{rechazar Ho}/\text{Ho cierta}) = \alpha$

$P(\bar{x} > C\alpha / u = u_0)$.

$= P\left(\frac{\bar{x} - u_0}{\sigma_x/\sqrt{n}} > \frac{C\alpha - u_0}{\sigma_x/\sqrt{n}}\right)$

$= P(Z > Z_{1-\alpha})$



$Z_{1-\alpha} = \frac{C\alpha - u_0}{\sigma_x/\sqrt{n}}$

$u_0 + Z_{1-\alpha} \cdot \frac{\sigma_x}{\sqrt{n}} = C\alpha$

Potencia = P(rechazar Ho/Ho Falsa).

$1 - B$

$P(E_{II}) = P(\text{no rechazar Ho}/\text{Ho Falsa})$.

*rechazo Ho cuando $\bar{x} > C\alpha$

-ensayo unilateral por izquierda.

$H_0)\ u = u_0$ $\qquad H_1)\ u = u_1 ;\ \ u_1 < u_0$.

* estadística base = $\bar{x}$

$P(E_1) = \alpha = P(\text{rechazar Ho}/\text{Ho cierta})$

$P(\bar{x} < C\alpha / u = u_0) = \alpha$.

$P\left(\frac{\bar{x} - u_0}{\sigma_x/\sqrt{n}} < \frac{C\alpha - u_0}{\sigma_x/\sqrt{n}}\right) = \alpha$

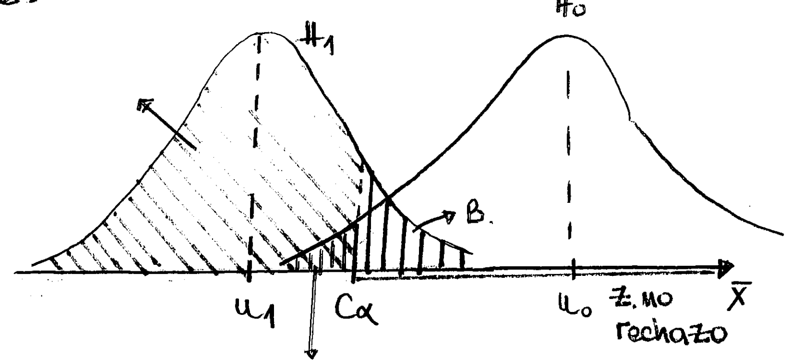
$P(Z < -Z_{1-\alpha}) = \alpha$

x'simetría de la Z.

$-Z_{1-\alpha} = \frac{C\alpha - u_0}{\sigma_x/\sqrt{n}}$

$u_0 - Z_{1-\alpha} \cdot \frac{\sigma_x}{\sqrt{n}} = C\alpha$.

Potencia = P(rechazar Ho / Ho falsa) = 1 - B



$P(E_1) = P(\text{rechazar Ho} / \text{Ho cierta}) = \alpha$

$P(E_{II}) = P\left(\text{no rechazar Ho} / \text{Ho falsa}\right) = B$

* otras formas de toma decisión

* rechazo Ho si $Z_{obs} < -Z_{1-\alpha}$

$$= \frac{\bar{X} - u_0}{\sigma_x/\sqrt{n}} < \frac{C_\alpha - u_0}{\sigma_x/\sqrt{n}}$$

* rechazo si $P_{asociada} < \alpha$

$$\left.\begin{aligned} P_{asociada} &= P(\bar{X} < \bar{X}_{observado}) \\ &= P(Z < Z_{observado}) \end{aligned}\right\}$$ comparas $P(E_1)$ si es mas pequeño rechazas

* para ensayo bilateral se demuestran las dos

* Ensayo de hipotesis para promedio de población para varianza desconocida.

- ensayo unilateral por derecha.

$Ho)\ u = u_0 \qquad H_1)\ u = u_1\ ;\ u_1 > u_0$

estadistica base = $\bar{X}$.

$P(E_1) = P(\text{rechazar Ho} / \text{Ho cierta}) = \alpha$

$$P(\bar{X} > C_\alpha / u = u_0) = \alpha$$

$$P\left(\frac{\bar{X} - u_0}{\sigma_x/\sqrt{n}} > \frac{C_\alpha - u_0}{\sigma_x/\sqrt{n}}\right) = \alpha$$

$$P\left(t_{n-1} > T_{n-1;\,1-\alpha}\right) = \alpha$$

$$\frac{C_\alpha - u_0}{S_x/\sqrt{n}} = t_{n-1;\,1-\alpha}$$

$$C_\alpha = u_0 + t(n-1;\,1-\alpha) \cdot \frac{S_x}{\sqrt{n}}$$

* Otras formas de rechazo

- rechazo Ho si = $t_{obs} = \dfrac{\bar{x} - u_0}{s_x/\sqrt{n}} > t_{crítica} = \dfrac{C\alpha - u_0}{s_x/\sqrt{n}}$

- ensayo unilateral por izquierda

$H_0)\ u = u_0 \qquad H_1)\ u = u_1\ ;\ u_1 < u_0$

Estadística base = $\bar{x}$

$P(E_1) = P(\text{rechazar } H_0 / H_0 \text{ cierta}) = \alpha$

$= P(\bar{x} < C\alpha \ / \mu = u_0) = \alpha$

$P\left(\dfrac{\bar{x} - u_0}{s_x/\sqrt{n}} < \dfrac{C\alpha - u_0}{s_x/\sqrt{n}}\right) = \alpha \qquad \dfrac{C\alpha - u_0}{s_x/\sqrt{n}} < -t(n-1; 1-\alpha)$

$C\alpha = u_0 - t(n-1; 1-\alpha) \cdot \dfrac{s_x}{\sqrt{n}}$

$P(t(n-1) < -t(n-1; 1-\alpha)) = \alpha$

por simetría de la "t".

"cálculo de extensión para muestra de un ensayo" → ¿Qué tan grande debe ser "n"?

$H_0)\ u = u_0 \qquad H_1)\ u = u_1 \quad (u_1 < u_0) \implies$ rechazo Ho si $\bar{x} < C\alpha$

$P(E_{II}) = \beta$

$P(E_{II}) = P(\text{no rechazar } H_0 / H_0 \text{ falsa}) = \beta$

valor crítico deducido anteriormente

$C\alpha = u_0 - z_{1-\alpha} \cdot \dfrac{\sigma_x}{\sqrt{n}}$

$= P(\bar{x} > \underset{*}{C\alpha} / u = u_1) = \beta$

H_1 cierta

$= P\left(\dfrac{\bar{x} - u_1}{\sigma_x/\sqrt{n}} > \left(u_0 - z_{1-\alpha} \cdot \dfrac{\sigma_x}{\sqrt{n}} - u_1\right) \Big/ \sigma_x/\sqrt{n}\right) = \beta.$

$= P\left(Z > \underbrace{-z_{1-\alpha} + \dfrac{u_0 - u_1}{\sigma_x/\sqrt{n}}}_{z_{1-\beta}}\right) = \beta$

$$Z_{1-\beta} = \frac{u_0 - u_1}{\sigma_x / \sqrt{n}} - Z_{1-\alpha}$$

lo pienso multiplicando al término

$$Z_{1-\beta} = \frac{u_0 - u_1}{\sigma_x} \cdot \sqrt{n} - Z_{1-\alpha}$$

$$Z_{1-\beta} + Z_{1-\alpha} = \frac{u_0 - u_1}{\sigma_x} \cdot \sqrt{n}$$

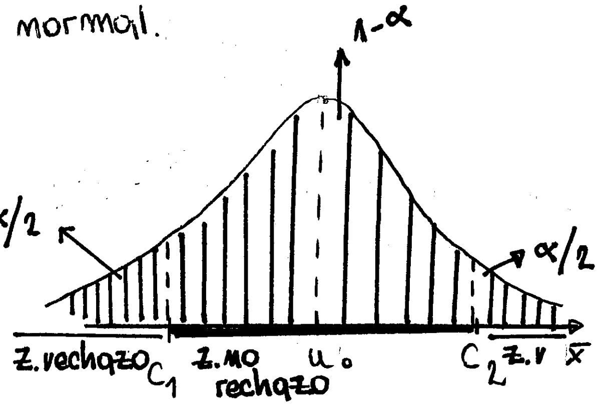
$$\frac{Z_{1-\beta} + Z_{1-\alpha}}{\frac{u_0 - u_1}{\sigma_x}} = \sqrt{n}$$

$$\left[\frac{(Z_{1-\beta} + Z_{1-\alpha}) \cdot \sigma_x}{u_0 - u_1}\right]^2 = \cancel{\sqrt{}}\,n$$

* a menor α y/o β ↑ mayor "n"
* a mayor heterogeneidad en la población ↑ mayor "n"
* a menor diferencia entre los promedios ↑ mayor "n"

* ensayos bilaterales para promedio población normal.

$H_0)\ u = u_0 \qquad H_1)\ u \neq u_0 \qquad P(E_1) = \alpha$



estadística base $= \bar{x}$

* rechazo H_0 sí $\underbrace{\bar{x} < C\alpha_1 \text{ ó } \bar{x} > C\alpha_2}_{\text{dos zonas de rechazo.}}$

√ con variancia conocida =

$$C_1 = u_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}} \qquad C_2 = u_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_x}{\sqrt{n}}$$

– con variancia desconocida =

$$C_1 = u_0 - t_{(n-1;\, 1-\frac{\alpha}{2})} \cdot \frac{\sigma_x}{\sqrt{n}} \qquad C_2 = u_0 + t_{(n-1;\, 1-\frac{\alpha}{2})} \cdot \frac{S_x}{\sqrt{n}}$$

- Ensayo de hipótesis para la variancia de una poblacional normal.

$H_0)\ \sigma_x^2 = \sigma_0^2$ $\quad H_1)\ \sigma_x^2 = \sigma_1^2\ ;\ \sigma_1^2 > \sigma_0^2 \longrightarrow$ ensayo unilateral por derecha.

estadístico base $= S_x^2$

$$P(E_1) = P(\text{rechazar } H_0 / H_0 \text{ cierta}) = \alpha$$

$$= P\left(\sigma_1^2 > \sigma_0^2 / \sigma_x^2 = \sigma_0^2\right) = \alpha$$

* para variancia se estandariza como $\underline{\underline{\chi^2_{n-1}}}$

$$P\left(S_x^2 > C\alpha / \sigma_x^2 = \sigma_0^2\right) = \alpha$$

$$P\left(\frac{S_x^2 \cdot (n-1)}{\sigma_0^2} > \frac{C\alpha \cdot (n-1)}{\sigma_0^2}\right) = \alpha$$

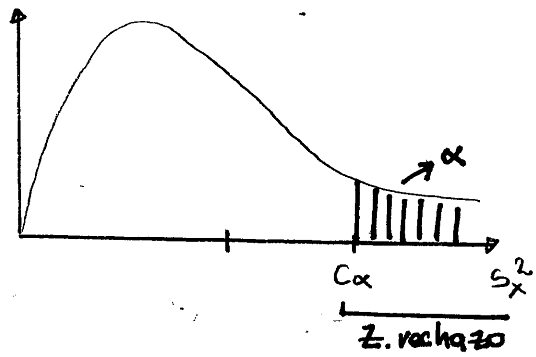
ojo!

$$\cancel{P\left(t(n-1) > t(n-1;\ 1-\alpha)\right) = \alpha} \quad P\left(\chi^2(n-1) > \chi^2(n-1;\ 1-\alpha)\right) = \alpha$$

$$\chi^2(n-1;\ 1-\alpha) = \frac{C\alpha \cdot (n-1)}{\sigma_0^2}$$

$\longrightarrow$ la estandarización es en base a H_0!, entonces usas σ_0^2

$$\frac{\chi^2(n-1;\ 1-\alpha) \cdot \sigma_0^2}{(n-1)} = C\alpha$$



Otras formas de rechazo =

- rechazo H_0 si =

$\underline{S^2 \text{ observada}} > C\alpha$

$$\frac{S_x^2 \cdot (n-1)}{\sigma_0^2} > \frac{\chi^2(n-1,\ 1-\alpha) \cdot \sigma_0^2}{(n-1)}$$

- rechazo H_0 si =

$P_{asociada} < \alpha \Longrightarrow$

$$P_{asociada} = P\left(S_x^2 > S^2_{obs}\right) = P\left(\chi^2_{(n-1)} > \chi^2_{(n-1;\, 1-\alpha)}\right)$$

$\longrightarrow$ siempre nos interesa analizar si la variancia es mayor porque afecta presición

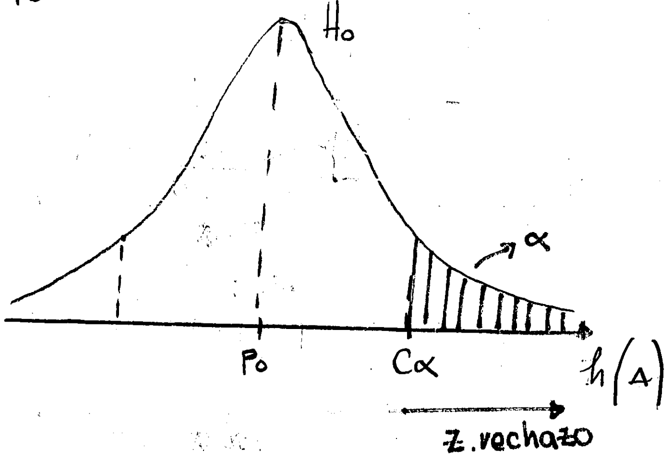
- Ensayo de hipotesis para la <u>probabilidad</u> de un suceso.

(válido p' muestras grandes).

*ensayo unilateral por derecha

$p = P(A)$

$H_0)\ p = p_0$ $\qquad H_1)\ p = p_1 \qquad p_1 > p_0$



Estadistica base = $h(A)$.

$$P(E_1) = P\left(\text{rechazar } H_0 / H_0 \text{ cierta}\right) = \alpha$$

$$P\left(p_1(A) > p_0(A) / p(A) = p_0\right) = \alpha$$

$$P\left(h_1(A) > C\alpha / p(A) = p_0\right) = \alpha$$

$$P\left(\frac{h_1(A) - p_0(A)}{\sqrt{\frac{p_0 q_0}{n}}} > \frac{C\alpha - p_0(A)}{\sqrt{\frac{p_0 . q_0}{n}}}\right) = \alpha$$

$$P\left(Z > Z_{1-\alpha}\right) = \alpha$$

$$Z_{1-\alpha} = \frac{C\alpha - p_0(A)}{\sqrt{\frac{p_0 . q_0}{n}}}$$

$$\boxed{p_0(A) + Z_{1-\alpha} . \sqrt{\frac{p_0 . q_0}{n}} = C\alpha}$$

– Ensayo unilateral por izquierda

$H_0)\ p(A) = p_0 \qquad H_1)\ p(A) = p_1(A)\ ;\ p_1(A) < p_0(A)$

√ Estadística base $= h(A)$.

√ $P(E_1) = P\left(\text{rechazar } H_0 / H_0 \text{ cierta}\right) = \alpha$

$$P\left(p_1(A) < p_0(A) / p(A) = p_0\right) = \alpha$$

$$= P\left(h_1(A) < C\alpha / p(A) = p_0\right) = \alpha$$

$$= P\left(\frac{h_1(A) - p_0}{\sqrt{\frac{p_0 . q_0}{n}}} < \frac{C\alpha - p_0}{\sqrt{\frac{p_0 . q_0}{n}}}\right) = \alpha$$

$$= P\left(Z < \underbrace{-Z_{1-\alpha}}_{= Z\alpha}\right) = \alpha$$

$$-Z_{1-\alpha} = \frac{C\alpha - p_0}{\sqrt{\frac{p_0 . q_0}{n}}}$$

$$p_0 - \frac{Z_{1-\alpha}}{\sqrt{\frac{p_0 . q_0}{n}}} = C\alpha$$

Otras formas de rechazo:

* rechazo H_0 si:

$$Z_{obs} < \underbrace{Z_{crítica}}_{Z\alpha = -Z_{1-\alpha}} = h(A)_{obs} < C\alpha$$

* rechazo H_0 si

$$P_{asociada} < \alpha \ ;\ P_{asociada} = P\left(h(A) < h(A)_{obs}\right)$$
$$= P\left(Z < Z_{obs}\right).$$

Ensayos bilaterales

$H_0)\ p(A) = p_0 \qquad H_1)\ p(A) \neq p_0$

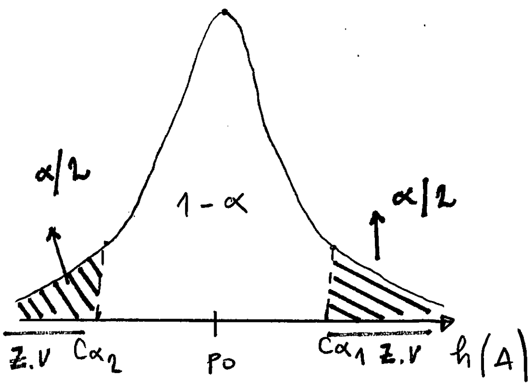
Estadística base $= h(A)$

$P(E_1) = \alpha$

rechazo H_0 si $=$ $h(A) > C\alpha_1$ ó $h(A) < C\alpha_2$

$$C\alpha_1 = p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}$$

$$C\alpha_2 = p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}$$



Capítulo 6 = "ensayo de hipotesis en base a una muestra"

"problemas de desición" $\xrightarrow[\text{Formulamos}]{}$ "ensayo de hipotesis."

definimos = procedimiento estadistico usado para tomar decisiones respecto del valor de un parámetro poblacional, estimado en base a la información de una muestra de la población.

1) Planteamos dos hipótesis

→ H₀) es la afirmación de una característica poblacional que al principio se supone cierta.

→ H₁) es la hipótesis que contradice H₀.

Clasificación

Hipotesis.
- simples = establecen un sólo punto (único valor real) como valor del parámetro.
- compuestas = contienen más de un único valor real.

Ensayos hipotesis
- unilateral
  - izquierda
  - derecha
- bilateral.

2) Toma de desición
- rechazar Ho
- rechazar H1.

Error tipo I = rechazar Ho, siendo Ho cierta ⟶ $P(E_I) = \alpha$

↓ nivel significación ensayo.

Error tipo II = no rechazar Ho, siendo Ho falsa ⟶ $P(E_{II}) = \beta$.

*potencia = (rechazar Ho / Ho falsa) = $1 - \beta$.

↳ probabilidad de elegir bien

• Regla decisión =

- La decisión se toma en base a la información de la muestra, con la cual se construye la estadística base del ensayo.

- el procedimiento de ensayo de hipótesis genera una región crítica o zona de rechazo de Ho y una región de no rechazo, sobre el eje de variación estadística base.

¿cómo establecemos las zonas?

Fijamos P(Ei) en un valor fijo "α", llamado nivel significación y se elige aquella zona de rechazo que haga mínima P(Eii). (maximice potencia).

- Región crítica = conjunto de valores de la estadística base tal que si el valor observado (zona rechazo) p/ la muestra tomada pertenece a dicha región, se decide rechazar Ho.

Ensayo hipótesis →
- 1) establecer las hipótesis
- 2) fijar nivel significación "α" = P(E1).
- 3) Obtener muestra y calcular el valor observado de la estadística base
- 4) Toma decisión
  - → rechazo Ho; Cα observado pertenece región crítica
  - → no rechazo Ho; Cα observado pertenece región no rechazo :)

E.H
- p/ → promedio población normal → varianza conocida / varianza desconocida
  - * cálculo ext. muestra
- p/ → promedio población normal (bilateral)
- p/ → probabilidad suceso
- p/ → varianza población normal.

e) otra forma de expresar la regla decisión es rechazar H_0 cuando =

$$t_{observada} > t_{(n-1);(1-\alpha)}$$

despejas de la regla decisión

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{(n-1);(1-\alpha)}$$

$$t_{observada} = \frac{0{,}0651\ mg/L - 0{,}05\ mg/L}{0{,}021\ mg/L / \sqrt{10}} = 2{,}27$$

$$t_{(9);(0{,}95)} = 1{,}83$$

$$2{,}27 > 1{,}83 \implies \text{se rechaza } H_0$$

f).- $\underbrace{P_{asociada}} < \alpha$

$$P(\bar{x} > \text{observado}) < \alpha$$

estandarizas pluma T

$$\frac{\bar{x} - \mu_x}{s/\sqrt{n}} > \frac{\bar{x} - \mu_x}{s/\sqrt{n}} =$$

antiacumulada

$$T_{(9)} > T_{observada} = 1 - T_{observada}$$

$$T_{(9)} > \frac{0{,}0651\ mg/L - 0{,}05\ mg/L}{0{,}021\ mg/L/\sqrt{10}} = 2{,}27$$

$$T_{(9)} > 2{,}27 = 1 - 0{,}975 = 0{,}025$$

$$\underbrace{0{,}025}_{\text{producto de estandarizar una "T"}} < \underbrace{0{,}05}_{\alpha} \implies \text{se rechaza } H_0.$$

h) se rechaza H_0:

$$\underbrace{P_{asociada}}_{\text{calculada apartado f}} < \alpha$$

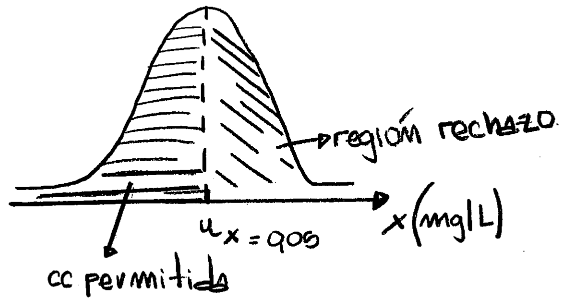
$$0{,}025 > 0{,}01 \implies \text{no se rechaza } H_0.$$

Ejercicio n° 3.

* ¿instalar un equipo tratamiento x' osmosis inversa?, si la cc de As es mayor

x = cc. As en el agua de la provincia de Santa Fe. (mg/L).

a) Ho) $u_{(x)} \leq 0{,}05$ mg/L H1) $u_x > 0{,}05$ mg/L → ensayo de hip. p' promedios (bilateral)



b) tomamos una muestra de agua a lo largo de 10 días y obtenemos la cc. As (mg/L) =

0,042 0,062 0,067 0,069 0,08 0,036 0,086

0,081 0,076 0,022

c) $\bar{x} = 0{,}0651$ mg/L. $S_x = 0{,}021$ mg/L

σ_x → desconocida $\alpha = 0{,}05$

se rechaza Ho si:

$$\bar{x} > u_0 + T_{(n-1);(1-\alpha)} \frac{S_x}{\sqrt{n}}$$

$$\bar{x} = 0{,}0651 \frac{mg}{L}$$

$$C_\alpha = 0{,}05 \, mg/L + T_{(9;\,0{,}95)} \frac{0{,}021 \, mg/L}{\sqrt{10}}$$

$$C_\alpha = 0{,}05 \, mg/L + 1{,}83 \cdot \frac{0{,}021 \, mg/L}{\sqrt{10}}$$

$$C_\alpha = 0{,}0622 \, mg/L$$

$0{,}0651$ mg/L $> 0{,}0622$ mg/L ⟹ rechazamos Ho.

Conclusión = con un nivel de significación del 5% concluimos que el verdadero valor promedio de cc. de arsénico en agua de la provincia de Santa Fe supera 0,05 mg/L y por lo tanto sería conveniente que las autoridades colocaran un equipo p' el tratamiento del agua por osmosis inversa.

Ejercicio N° 4.

Fin trabajo = decidir el cambio de proveedor

X : pureza plaguicida (%) del nuevo proveedor.

$X \longrightarrow N(\mu_x ; \sigma_x = 0{,}1\%)$

$H_0)\ \mu_x \leq 99{,}2\%$ $\qquad H_1)\ \mu_x > 99{,}2\%$

b)- rechazar Ho, con <u>Ho cierta</u> $E(I) = 0{,}05$.

Aceptar Ho ; cuando <u>Ho Falsa</u> $E(II) = 0{,}01$

$\mu_x - \mu = 0{,}15$

$-\mu = 0{,}15 - 0{,}992$

$\boxed{\mu = 0{,}842}$

$$M = \left(\frac{(Z_{1-0{,}05} + Z_{1-0{,}01}) \cdot 0{,}001}{1{,}5 \times 10^{-3}} \right)^2$$

$$M = \left(\frac{(1{,}645 + 2{,}326) \cdot 0{,}001}{1{,}5 \times 10^{-3}} \right)^2$$

$\boxed{M = 7{,}01 \cong 8}$

c) $\bar{X} = 99{,}297\%$.

Otras formas =

se rechaza Ho si

se rechaza Ho si =

$$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma_x}{\sqrt{M}}$$

$\bar{X} = 99{,}297\%$

$$C\alpha = 0{,}992 + Z_{0{,}95} \cdot \frac{1 \times 10^{-3}}{\sqrt{10}}$$

$$C\alpha = 0{,}992 + 1{,}645 \cdot (1 \times 10^{-3}) \cdot \sqrt{10}$$

$$C\alpha = 0{,}9972 \times 100 = 99{,}72\%$$

$99{,}297\% > 99{,}25\% \Longrightarrow$ rechazamos Ho conclusión = el plaguicida propuesto por proveedor "b" tiene mayor pureza y por lo tanto, es conveniente que para la calibración se proceda con el cambio del mismo.

d) Potencia = $1 - \beta$

$$1 - 0{,}01 = 0{,}99.$$

En este caso la potencia es igual a 0,99.

e)- N.C = 95%

conocido

$\sigma = 0{,}1\%$ $\quad \bar{x} = 99{,}294\%$

$n = 10.$

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

$\alpha = 0{,}05$

$$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 99{,}294\% \pm Z_{1-\frac{0{,}05}{2}} \cdot \frac{0{,}1\%}{\sqrt{10}}$$

$$= 99{,}294\% \pm \underbrace{Z_{0{,}975}} \frac{0{,}1\%}{\sqrt{10}}$$

$$= 99{,}294\% \pm 1{,}96 \cdot \frac{0{,}1\%}{\sqrt{10}}$$

$$\boxed{= (99{,}23 \,;\, 99{,}36)\,\%}$$

Ejercicio 5.

x: cc. de hepcidina en pacientes talasemicos. (mg/mL)

$$x \longrightarrow N\left(\mu_x = 21\ mg/mL \,;\, \sigma_x^2\right)$$

a) La variable se distribuye normalmente ya que lo analizamos por la tendencia lineal de los datos en el gráfico de probabilidad normal.

$\alpha = 5\% = 0{,}05.$

$\bar{x} = 13{,}2492\ mg/mL.$

$S_x = 5{,}1233\ mg/mL$

$H_0)\ \mu = 21\ mg/mL$ $\qquad H_1)\ \mu < 21\ mg/mL$

rechazamos Ho si:

$$\bar{x} < \mu_0 - t_{(n-1);(1-\alpha)} \cdot \frac{s}{\sqrt{n}}$$

$\bar{X} = 13,28\ mg/mL$

$$C\alpha = 21\frac{mg}{mL} - t(11;0,95) \cdot \frac{5,12\ mg/mL}{\sqrt{12}}$$

$$C\alpha = 21\frac{mg}{mL} - 1,80 \cdot 1,48\ mg/mL$$

$$C\alpha = 18,336\ mg/mL$$

$$13,28\frac{mg}{mL} < 18,336\ mg/mL \Longrightarrow \text{rechazamos } H_0$$

Conclusión: con un nivel de significación del 5% se estima que el valor promedio de cc. de hepcidina en pacientes portadores talasemicos es menor a 21 mg/mL.

Otras formas de tomar una decisión:

* rechazamos Ho si:

$$T_{observada} < T_{(n-1);(1-\alpha)}$$

$$\frac{\bar{X} - \mu_0}{s/\sqrt{n}} < T_{(n-1);(1-\alpha)} \longrightarrow -5,22 < 1,80 \Longrightarrow \text{rechazamos } H_0$$

$$T_{observada} = \frac{13,28\ mg/mL - 21\ mg/mL}{5,12\ mg/mL/\sqrt{12}} \qquad T_{(11);(0,95)} = 1,80$$

$$= -5,22$$

$P(\text{asociada}) = \cancel{P(\bar{x} < \alpha)}$

rechazamos Ho si = $P(\bar{x} < \bar{x}_{obs}) < \alpha$

$* \; P(\bar{x} < \bar{x}_{observado}) = \dfrac{\bar{x} - u_0}{S/\sqrt{n}} < \dfrac{\bar{x} - u_0}{S/\sqrt{n}}$

$$= T_{11} < \frac{13{,}28\,mg/mL - 21\,mg/mL}{\dfrac{5{,}12\,mg/mL}{\sqrt{12}}}$$

$$= T_{11} < -5{,}22$$

$$\cong 0$$

$p \cong 0 \rightarrow$ tabla de la "T" reducida

$\cong 0 < 0{,}05 \Longrightarrow$ rechazamos Ho

b)- $P\left(\text{rechazar Ho} / u = 14\ \dfrac{mg}{mL}\right) = 1 - \beta \xrightarrow{} $ potencial !! ☺ $\qquad \sigma = 5\,mg/dL$

$P\left(\underbrace{\bar{x} < u_0 - Z_{1-\alpha} \cdot \dfrac{\sigma_x}{\sqrt{n}}}_{a} / u = 14\right) = 1 - \beta \rightarrow$ consultar.

$a = 21\,\dfrac{mg}{dL} - Z_{0{,}95} \cdot \dfrac{5\,mg/dL}{\sqrt{12}}$

$a = 18{,}63\ mg/dL$

$$P\left(\frac{\bar{x} - u_0}{\sigma/\sqrt{n}} < \frac{u_0 - Z_{1-\alpha} \cdot \dfrac{\sigma_x}{\sqrt{n}} - u}{\sigma_x \sqrt{n}}\right) = 1 - \beta$$

$$P\left(Z < \frac{u_0 - u}{\sigma_x \sqrt{n}} - Z_{0{,}95}\right) = 1 - \beta$$

$$P\left(Z < \frac{21\,mg/dL - 14\,mg/dL}{5\,mg/dL/\sqrt{12}} - 1{,}645\right) = 1 - \beta$$

$$P(Z < 1{,}13) = 1 - \beta$$

$$0{,}8708 = 1 - \beta$$

Ejercicio n°6 x = contenido proteínas totales leche

$$X \longrightarrow N(\mu_x, \sigma_x)$$

$\bar{X} = 2{,}895 g/100g$ $S_x = 0{,}215 g/100g$

$n = 30$ $P(E_1) = \alpha = 0{,}05$

$H_0)\ \mu_x \geq 2{,}9 g/100g$ $H_1)\ \mu_x < 2{,}9 g/100g$

se rechaza H_0 sí =

$$\bar{X} < \mu_0 - T(n-1); (1-\alpha) \cdot \frac{S}{\sqrt{n}}$$

$$\bar{X} = 2{,}895 g/100g$$

$$C\alpha = 2{,}9 g/100g - T_{29}; 0{,}95 \cdot \frac{0{,}215 g/100g}{\sqrt{30}}$$

$$C\alpha = 2{,}9 g/100g - \frac{1{,}70 \cdot 0{,}215 g/100g}{\sqrt{30}}$$

$$C\alpha = 2{,}83 g/100g$$

$2{,}895 g/100g > 2{,}83 g/100g \Longrightarrow$ no rechazamos H_0

* otras formas

$T_{observada} < T_{crítica}$

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} < -T(n-1); (1-\alpha) \Longrightarrow -0{,}13 > -1{,}70 \Longrightarrow \text{no rechazamos } H_0$$

$$T_{observada} = \frac{2{,}895 g/100g - 2{,}9 g/100g}{0{,}215 g/100g / \sqrt{30}}$$

$$T_{observada} = -0{,}13$$

$$T_{crítica} = -T(29); (0{,}95) = -1{,}70$$

$P_{asociada} = P(\bar{X} < \bar{X}_{observada}) < \alpha$

$* \alpha = 0{,}05$

$* P\left(\frac{\bar{X} - u_0}{S/\sqrt{n}} < \frac{\bar{X}' - u_0}{S/\sqrt{n}}\right)$

$* P\left(T < \frac{2{,}895g/100g - 2{,}9g/100g}{0{,}215g/100g / \sqrt{30}}\right) =$

$* P\left(T_{(29)} < \overset{\text{siempre positivo}}{(+)} 0{,}13\right) = 1 - 0{,}55$

$= 0{,}45$

$0{,}45 > 0{,}55 \Longrightarrow$ no rechazamos Ho

Conclusion: con 0,05 grado de significación se concluye que el contenido verdadero promedio de proteínas en la leche cumple con lo requerido del código Alimenticio Argentino.

b) $P(E_{II}) = (\text{aceptar Ho} / \text{Ho Falsa}) = B$ $\quad \sigma_X = 0{,}2g/100g$

$P\left(\bar{X} > u_0 - Z(1-\alpha) \cdot \frac{\sigma}{\sqrt{n}} \,/\, u = 2{,}8g/100g\right) = B$

$P\left(\frac{\bar{X} - u_0}{\sigma_X} > \frac{\left(u_0 + Z(1-\alpha) \cdot \frac{\sigma}{\sqrt{n}}\right) - 2{,}8g/100g}{0{,}2g/100g}\right) = B$

$P\left(\frac{\bar{X} - u_0}{\sigma_X} > \frac{2{,}9g/dL - 2{,}8g/100g}{\frac{0{,}2g/100g}{\sqrt{30}}} - Z_{0{,}95}\right) = B$

$P(Z \overset{\text{antiacumulada}}{>} 2{,}74 - 1{,}645) = B$

$1 - P(Z < 1{,}094) = B$

¿cómo graficaría esto?

$1 - 0{,}8621 = B$

$\boxed{0{,}1379 = B}$

$$h < p_0 - Z_{1-\alpha}\sqrt{\frac{p_0 \cdot q_0}{n}} =$$

$h = 0{,}0523$

$$C\alpha = p_0 - Z_{(1-\alpha)}\sqrt{\frac{p_0\, q_0}{n}} =$$

$$C\alpha = 0{,}1 - Z_{0{,}95} \cdot \sqrt{\frac{0{,}1 \cdot 0{,}9}{440}} = 0{,}1 - 1{,}645 \sqrt{\frac{0{,}1 \cdot 0{,}9}{440}} = 0{,}099$$

$0{,}0523 < 0{,}099 \Longrightarrow$ rechazamos h_0.

* ( )ras Formas de calcular $\longrightarrow$ $Z_{observado} < Z_{critico}$

rechazamos Ho si $\dfrac{h - p_0}{\sqrt{\frac{p_0 q_0}{n}}} < -Z_{0{,}95} =$

$$* \; Z_{obs} = \frac{0{,}0523 - 0{,}1}{\sqrt{\frac{0{,}1 \cdot 0{,}9}{440}}} = -3{,}34$$

$* \; -Z_{0{,}95} = -1{,}645$

$-3{,}34 < -1{,}645 \Longrightarrow$ rechazamos Ho.

$P_{asociada} < \alpha$

$$P(h < h') < \alpha$$

$$P\Bigg(\underbrace{\frac{h - p_0}{\sqrt{\frac{p_0 q_0}{n}}}}_{Z} < \frac{h' - p_0}{\sqrt{\frac{p_0 q_0}{n}}}\Bigg) < \alpha = 0{,}05 \Longrightarrow 0{,}0004 < 0{,}05$$ rechazamos Ho.

más chico

$$P_{asoc} = P\left(Z < \frac{0{,}0523 - 0{,}1}{\sqrt{\frac{0{,}1 \cdot 0{,}9}{440}}}\right)$$

$$P_{asociada} = P(Z < -3{,}34) = 0{,}0004$$

c) $P(E_{II}) = (\text{aceptar } H_0 / H_0 \text{ falsa}) = \beta$

$$P\left(\bar{X} > u_0 - z(1-\alpha) \cdot \frac{\sigma_X}{\sqrt{n}} \Big/ u = 2,8g/100g\right) = \beta$$

$$P\left(\frac{\bar{X} - u_0}{\sigma_X} > \frac{u_0 - z(1-\alpha)\frac{\sigma_X}{\sqrt{n}} - u}{\sigma_X/\sqrt{n}}\right) = \beta$$

antiacumulada

$$P\left(z > \frac{u_0 - u}{\sigma_X/\sqrt{n}} - z(1-\alpha)\right) = \beta$$

$$1 - P\left(z < \frac{u_0 - u}{\sigma_X/\sqrt{n}} - z(1-\alpha)\right) = \beta$$

$$1 - P\left(z < \frac{2,9g/100 - 2,8g/100g}{\frac{0,2g/100g}{\sqrt{20}}} - 1,645\right) = \beta$$

$$1 - P(z < 0,59) = \beta$$

$$1 - 0,7224 = \beta$$

$$0,2776 = \beta$$

Ejercicio 4.

A: efectos secundarios en enfermos de artritis por el uso de analgésicos.

$P(A) = ?$ $\quad h = \frac{23 \text{ e.sec.}}{440 \text{ pac.}}$ $\quad P(E_I) = 0,05.$

$h = 0,0523.$

a) $H_0)\ P(A) = 0,1 \qquad H_1)\ P(A) < 0,1$

Se rechaza H_0 si:

$$C\alpha = \chi^2(9; 0{,}99) \cdot \frac{100\ mg/dL^2}{9} =$$

$$C\alpha = 21{,}7 \cdot \frac{100\ mg/dL^2}{9}$$

$$C\alpha = 241\ mg/dL^2$$

$$93{,}79\ mg/dL < 241\ mg/dL^2 \Longrightarrow \text{no rechazo } H_0.$$

* otras formas de desición

a) $\chi^2_{observada} > \chi^2_{critica}$

$$\frac{\underset{\sim}{S_X^2}}{\sigma_X^2/(n-1)} > \chi^2(n-1)(1-\alpha)$$

$$\frac{93{,}79\ mg/dL^2}{100\ mg/dL^2/9} > \chi^2(9)(0{,}99)$$

$$\chi^2_{obs} = 8{,}44$$

$$\chi^2_{crítica} = 21{,}7$$

$$8{,}44 > 21{,}7 \Longrightarrow \text{no rechazamos } H_0$$

b) $P_{asociada} < \alpha = 0{,}01$

$$P(S_X^2 > S^2_{X'}) < 0{,}01 = \alpha \longrightarrow \cong 0{,}50 > 0{,}01 \Longrightarrow \text{no rechazamos } H_0$$

$$P\left(\chi_9 > \frac{S^2_{X'} \cdot (n-1)}{100}\right) =$$

$$P\left(\chi_9 > \frac{93{,}79\ mg/dL^2 \cdot 9}{100\ mg/dL^2}\right) = 1 - P(\chi_9 < 8{,}4411) \cong 1 - 0{,}50 \cong 0{,}50$$

Conclusión = con un nivel de significación 5% se concluye que el porcentaje de enfermos de artritis que toman analgésicos y contusen efectos secundarios es menor al 10%.

b) estimación por intervalo de confianza

n. confianza = 95%

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

$\alpha = 0{,}05.$

$$h \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{h(1-h)}{n}} =$$

$$0{,}0523 \pm Z_{0{,}975} \sqrt{\frac{0{,}0523 \cdot (0{,}9477)}{440}} = (0{,}0315 \ ; \ 0{,}073)$$

$$(3{,}15 \ ; \ 7{,}31)\ \%$$

Conclusión = con un nivel de confianza del 95% se estima que el verdadero porcentaje de enfermos de artritis que toman analgésicos y contusen efecto secundarios esta comprendido entre $(3{,}15 ; 7{,}31)\ \%$.

Ejercicio 8. pág 91.

$P(E_1) = 0{,}01$

$P(\text{rechazar } H_0 / H_0 \text{cierta}) = \alpha = 0{,}01.$

datos muestrales = 148, 180, 185, 182, 180, 177, 171, 157, 194, 173 mg/dL

X = cc. colesterol total mg/dL

precisión método → variancia

$\sigma_X = 10$ mg/dL =

$\sigma^2_X = 100$ mg/dL

a)

$H_0)\ \sigma^2_X = 100 \frac{mg}{dL}$ $H_1)\ \sigma^2_X > 100 \frac{mg}{dL}$

Rechazo H_0 sí =

$$S_x^2 > \chi^2_{(n-1);\,1-\alpha} \cdot \frac{\sigma_x^2}{n-1}$$

$S_x^2 = 93{,}79\ mg/dL^2$

Conclusión = con un nivel de significación del 1% ~~concluimos~~ estimamos que el verdadero valor de la variancia poblacional es 100 mg/dL. No ha aumentado y por lo tanto no hubo disminución en la precisión del método.

Ejercicio 9.

x: contenido de creatinina medidos (mg/dL).

$n = 30$ $\quad u_x = 0{,}8$ mg/dL $\quad$ errores aleatorios $\longrightarrow N(0; \sigma^2)$

$\bar{X} \longrightarrow N\left(u = 0{,}8 \frac{mg}{dL}; \sigma^2\right)$

desconocida

$\bar{X} = 0{,}81$ mg/dL $\quad S_x = 0{,}025$ mg/dL

$\alpha = 0{,}01$.

$H_0)\ u = 0{,}8$ mg/dL $\quad H_1)\ u \neq 0{,}8$ mg/dL

ensayo bilateral.

con que una se cumpla ya rechazamos Ho.

$$\bar{X} < \mu_0 - t(n-1),\left(1-\frac{\alpha}{2}\right) \cdot \frac{S_x}{\sqrt{n}} \quad ó \quad \bar{X} > u_0 + t(n-1)\left(1-\frac{\alpha}{2}\right) \cdot \frac{S_x}{\sqrt{n}}$$

$$\bar{X} = 0{,}81 \frac{mg}{dL}$$

$$C\alpha = u_0 \pm t(n-1)\left(1-\frac{\alpha}{2}\right) \cdot \frac{S_x}{\sqrt{n}}$$

$$C\alpha = 0{,}8 \frac{mg}{dL} \pm t(29; 0{,}995) \cdot \frac{0{,}025 \text{ mg/dL}}{\sqrt{30}}$$

$$C\alpha = 0{,}8 \frac{mg}{dL} \pm 2{,}76 \cdot \frac{0{,}025 \text{ mg/dL}}{\sqrt{30}}$$

$$C\alpha = 0{,}79 \frac{mg}{dL} \quad ó \quad C\alpha = 0{,}813 \text{ mg/dL}$$

$$0{,}81 \frac{mg}{dL} > 0{,}79 \frac{mg}{dL} \Longrightarrow \text{no rechazo } H_0$$

$$0{,}81 \frac{mg}{dL} \cong 0{,}81 \frac{mg}{dL} \Longrightarrow \text{no rechazo } H_0$$

* Rechazamos H_0 si

T observada < T crítica ó T observada > T crítica

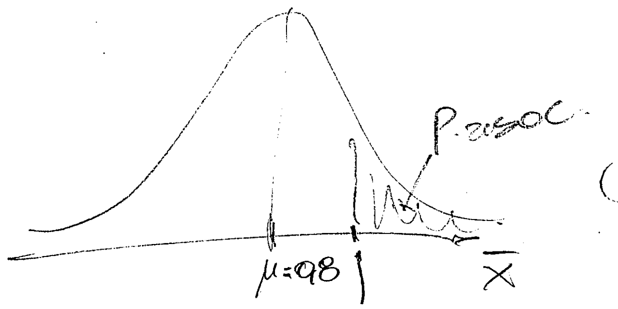
$$\frac{\bar{x} - u_0}{s_x / \sqrt{n}} < -t(n-1)\left(1-\frac{\alpha}{2}\right) \qquad \frac{\bar{x} - u_0}{s_x \sqrt{n}} > T(n-1)(1-\alpha/2)$$

$$T_{obs} = \frac{0{,}81\ mg/dL - 98\ mg/dL}{0{,}025\ mg/d / \sqrt{30}}$$

$T_{obs} = 2{,}19$ T crítica $= \pm 2{,}76$

$2{,}19 > -2{,}76 \Rightarrow$ no rechaza H_0

$2{,}19 < 2{,}76 \Rightarrow$ no rechazo H_0



$\bar{x}_{obs} = 0{,}81$

* $2\, P_{asociada} < \alpha$

$2 P(\bar{x} < \bar{x}') > \alpha$

$$2 \cdot P\left(\frac{\bar{X} - u_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{0{,}81\ mg/dL - 0{,}8\ mg/dL}{0{,}025\ mg/d / \sqrt{30}}\right)$$

$P(\bar{x} > 0{,}81)$

$P\left(\frac{\bar{x} - \mu}{s/\sqrt{n}} > 2{,}19\right)$

$P(t_{29} > 2{,}19)$

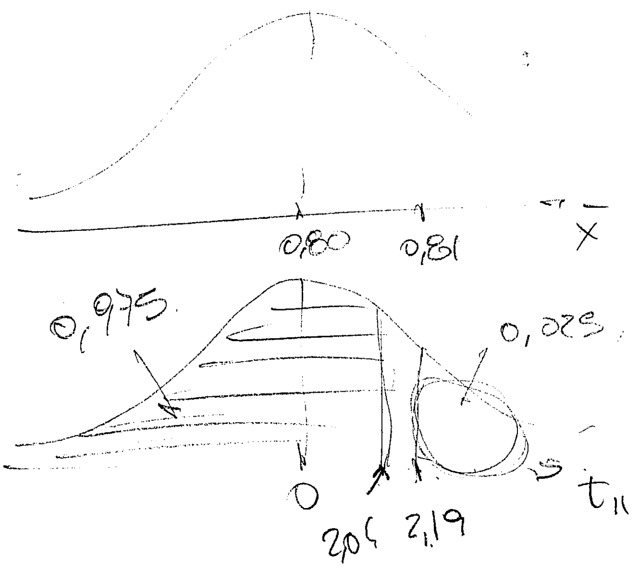
$= 2 * 0{,}025$

$$2 \cdot (t_{29} > 2{,}19) = 2\,[1 - P(t_{29} < 2{,}19)] =$$

$$\cong 2 \cdot [1 - P(t_{29} < 2{,}19)] =$$

$$2 \cdot [1 - 0{,}99] = 0{,}02$$

$0{,}02 < 0{,}05$



$2 * 0{,}025 = 0{,}05$

b) $C.V = \frac{S_x}{\bar{x}} = 0,05$

↑ Formula c.variación (no tenemos ensayo p' coeficiente variación) =

$S_x = 0,05 \cdot 0,81 \; mg/dL$

$S_x^2 = 1,64 \times 10^{-3} \; mg/dL^2 \Leftarrow S_x = 0,0405 \; mg/dL$

$P(E_1) = \alpha = 0,05$

$C.V = \frac{\sigma_x}{\mu_0} = 0,05$

$\sigma_x = 0,05 \cdot 0,8 \; \frac{mg}{dL}$

$\sigma_x = 0,04 \Longrightarrow \sigma_x^2 = 0,0016 \; \frac{mg}{dL}^2$

$H_0) \; \sigma^2_x \leq 0,0016 \; \frac{mg}{dL}^2$

$H_1) \; \sigma^2_x > 0,0016 \; \frac{mg}{dL}^2$

Se rechaza Ho sí =

$$S_x^2 > \chi_{(n-1);(1-\alpha)} \cdot \frac{\sigma_x^2}{n-1}$$

$$S_x^2 = 0,00164 \; \frac{mg}{dL}^2$$

$$C\alpha = \chi_{29;0,95} \cdot \frac{0,0016 \; \frac{mg}{dL}^2}{29}$$

$$C\alpha = 42,6 \cdot \frac{0,0016 \; mg/dL^2}{29}$$

$$C\alpha = 2,35 \times 10^{-3} \; mg/dL^2$$

$0,00164 \; \frac{mg}{dL}^2 < 2,35 \times 10^{-3} \; \frac{mg}{dL} \Longrightarrow$ no rechazamos Ho.

Ejercicio 10

A : prevalencia de bocio en escolares

$h(A) : \dfrac{16 \text{ niños} (+)}{455 \text{ niños}}$ $\quad n: 455$ niños

$p_0 = 0{,}05$

$q_0 = 1 - 0{,}05 = 0{,}95$

$h(A) = \cancel{0{,}04}\ 0{,}03516$

$H_0)\ P(A) \leq 0{,}05 \qquad H_1)\ P(A) > 0{,}05$

Rechazo Ho si:

$$h > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0\, q_0}{n}}$$

$h = \cancel{0{,}04}\ 0{,}03516$

$$C\alpha = Z(1 - 0{,}05) \cdot \sqrt{\frac{0{,}05 \cdot 0{,}95}{455}}$$

$$C\alpha = 1{,}645 \cdot 0{,}01022$$

$$C\alpha = 0{,}017$$

$\cancel{0{,}04} > 0{,}017 \Longrightarrow$ rechazo Ho.

$0{,}03516 > 0{,}017$

Conclusión = con un nivel de significación del 5% se concluye que existe deficiencia del consumo de Yodo en la población ya que la verdadera probabilidad de prevalencia de Iodo en escolares es mayor al 5%.

* se rechaza Ho si =

Zobservada > Zcrítica

$$Z_{obs} = \cancel{\frac{h - p_0}{\sqrt{\frac{p_0\, q_0}{n}}}} \qquad Z_{crítica} = Z_{1-\alpha} = 1{,}645$$

consultarlo.

$$= \cancel{-0{,}98} \cdot \frac{p_0 - h}{\sqrt{\frac{p_0 \cdot q_0}{n}}} = \frac{0{,}05 - 0{,}04}{\sqrt{\frac{0{,}05 \cdot 0{,}95}{455}}} = \frac{0{,}01}{0{,}45} = \cancel{0{,}978}\ 1{,}45$$

Ejercicio 10. ⟶ lo rehice

$h(a) = \dfrac{16 \text{ niños } (+)}{455 \text{ niños.}}$ $p_0 = 0{,}05 \; ; \; q_0 = 0{,}95$

$h(a) = 0{,}035165.$

$H_0)\; P(A) \leqslant 0{,}05$ $H_1)\; P(A) > 0{,}05.$

Rechazo H_0 si:

$$h > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}$$

$h = 0{,}03516$

$C\alpha = 0{,}05 + 1{,}645 \cdot \sqrt{\dfrac{0{,}05 \cdot 0{,}95}{455}}$

$C\alpha = 0{,}067$

$0{,}03516 < 0{,}067 \Longrightarrow$ no rechazamos H_0

Conclusión = con un nivel de significación del 5% concluimos que existe deficiencia de Yodo en la alimentación ya que el verdadera probabilidad de prevalencia de Yodo en escolares es mayor a 0,05.

otras formas de decisión =

* $Z_{observada} > Z_{critica}$

$$Z_{observada} = \frac{-(p_0 - h)}{\sqrt{\frac{p_0 \cdot q_0}{n}}}$$

$Z_{critica} = Z_{0,95} = 1{,}645$

$$= \frac{-(0{,}05 - 0{,}035165)}{\sqrt{\frac{0{,}5 \cdot 0{,}95}{455}}}$$

$-1{,}452 < 1{,}645 \Longrightarrow$ no rechazamos H_0.

$Z_{obs} = -1{,}452$

$P_{asociada} < \alpha$

$P(h > h') = P \qquad \alpha = 0{,}05$

$$P\left(\left(\frac{p_0 - h}{\sqrt{\frac{p_0 \cdot q_0}{n}}}\right) > -\left(\frac{p_0 - h'}{\sqrt{\frac{p_0 \cdot q_0}{n}}}\right)\right) = P$$

$0{,}9265 > 0{,}05 \Rightarrow$ no rechazamos H_0.

$P(Z > -1{,}45) = P$

$1 - P(Z < -1{,}45) = P$

$P \cong 1 - 0{,}0735$

$P \cong 0{,}9265$.

Ejercicio 15

A: úlceras gastroduodenales como efecto secundario del tratamiento con AAS.

$n = 133$ pacientes
13 contrajeron úlceras g. (+)

$h(A) = 13 / \cancel{133}\ 122$
$= \cancel{0{,}098}$
$= 0{,}10656$

$H_0)\ P(A) = 0{,}04 \qquad H_1)\ P(A) \neq 0{,}04$

se rechaza H_0 si =

$$h < p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}} \quad \text{ó} \quad h > p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}$$

$h = 0{,}10656$

$C_\alpha = 0{,}10656 \pm Z_{1-\left(\frac{0{,}05}{2}\right)} \cdot \sqrt{\frac{0{,}04 \cdot 0{,}93}{122}} = 0{,}0613!$

$= 0{,}152$

$0{,}10656 > 0{,}0613 \Rightarrow$ no rechazamos H_0.

$0{,}10656 < 0{,}152 \Rightarrow$ no rechazamos H_0.

Conclusión = con un nivel de significación del 5 % concluimos que el 7 % de pacientes tratados con aspirina (AAS) en baja dosis presenta úlceras gastroduo intestinales.

$P_{asociada} = 2P(\hat{p} > \hat{p}')$

$$* \; 2 \cdot P\left(\frac{p_0 - \hat{p}}{\sqrt{p_0 q_0 / n}} > \left(\frac{p_0 - \hat{p}'}{\sqrt{p_0 q_0}} \right) \right)$$

$$2\left[1 - P(Z < 1,6)\right] = 2 \cdot \left[1 - 0,9452\right]$$
$$= 2 \cdot 0,0548$$
$$= 0,1096$$

$p \cong 0,1096 > 0,05 \Longrightarrow$ no rechazamos Ho.

Ejercicio 16

método Elisa (enzimoinmunoensayo p/ analisis cuantitativo de trazas de gluten en productos panadería, etc)

x : mediciones de gliadina p/ determinar el contenido de gluten.

$\sigma_x = 1,2 ppm \Rightarrow \sigma^2_x = 1,44 ppm^2$

$n = 30$. $\quad \bar{X} = 30,1 ppm \quad S_x = 1,5 ppm \Longrightarrow 2,25 ppm^2$

Ho) $\sigma^2_x = 1,44 ppm$ $\qquad$ H1) $\sigma^2_x > 1,44 ppm^2$

se rechaza Ho sí =

$$S_x^2 > \chi_{(n-1);(1-\alpha)} \cdot \frac{\sigma_x^2}{(n-1)}$$ poblacional

consultar ¿?

$S_x^2 = 2,25 ppm^2$

$C\alpha = \chi_{29;0,95} \cdot \frac{\cancel{2,25 ppm^2}\; 1,44 ppm^2}{29}$

$C\alpha = 42,6 \cdot 0,08 ppm^2$

$C\alpha = 3,31 ppm^2$

$2,25 ppm^2 < 3,31 ppm^2 \Longrightarrow$ no rechazo Ho.

$2,25 ppm^2 > 2,12 ppm^2$

⇓

rechazo Ho.

rechazamos Ho si:

$P_{asociada} < \alpha$

$$P(s_x^2 > s_x^{2'}) =$$

$$P\left(\frac{\chi^2 (n-1)}{\sigma_x^2} > \underbrace{\frac{2,25 \cdot 29}{1,44}}\right) =$$

$$P(\chi^2_{29} > 45,31) =$$

$$1 - P(\chi^2_{29} < 45,31) =$$

$$1 - 0,975 = 0,03$$

$0,03 < 0,05 \Rightarrow$ rechazamos Ho.

rechazamos Ho si:

$$\chi^2_{obs} > \chi^2_{crítica}$$

$$\chi^2_{obs} = \underbrace{\frac{s_x^2 \cdot (n-1)}{\sigma_x^2}} = \underbrace{\frac{2,25 ppm^2 \cdot 29}{1,44 ppm^2}} = 45,31$$

$$\chi^2_{obs} = \chi_{29;0,95} = 42,6$$

$45,31 > 42,6 \Rightarrow$ rechazo Ho

Regla de decisión

- la desición se tomará en base a la información suministrada por la muestra, con la cual construimos una estadística base del ensayo.

- el procedimiento de ensayo de hipótesis establece una región crítica o zona de rechazo de Ho y una región de no rechazo, sobre el eje de variación de la estadística base.

Unidad n° 4 = diseño de experimentos y ensayo de hipotesis en base a dos muestras.

Diseño experimentos = planificación de experimentos para comparar efectos de varias condiciones experimentales (concentración, pH, etc.) sobre una o más variable de interés.

Ejemplo = efecto de tres técnicas de interés para la extracción de Arn.

↑ condición experimental

variable = cc. ARN para las diferentes técnicas

Preguntas para el ejemplo → ¿son las tres técnicas las únicas de interes? (condición experimental)

¿A través de qué variable puedo medir el efecto de las técnicas de extracción?

¿Cuál o cuales de otras condiciones pueden afectar la respuesta?

¿Cuantas alícuotas ($n = ?$) deberán tomarse para ser tratadas con cada una de las técnicas? dependerá si las unidades son homogeneas

¿Cómo asignaremos → aleatoriamente o no? una tecnica a cada conjunto de alícuotas?

¿Que método estadistico se aplicará para el análisis de los datos? → ¿Qué ensayo de hipotesis utilizaría?

Definiciones

1) Factores o criterios = características que diferencian las poblaciones entre sí. Ejemplo = tecnicas

2) niveles o categorias = modalidades del factor en estudio.
Ejemplo = técnica 1, técnica 2, técnica 3

3) Tratamiento = nivel del factor (en caso de un factor en estudio) o combinación niveles (más de dos factores en estudio).
Ejemplo = tratamiento = nivel del factor ⟹ técnica

4) unidad experimental = mínima porción del material experimental que se aplica un tratamiento y luego se mide la variable.

Diseños que nos interesan en este curso. (básicos)

Tipos

→ diseño completamente aleatorizado ①

→ diseño en bloques completos aleatorizados. ②

① *unidades experimentales homogéneas

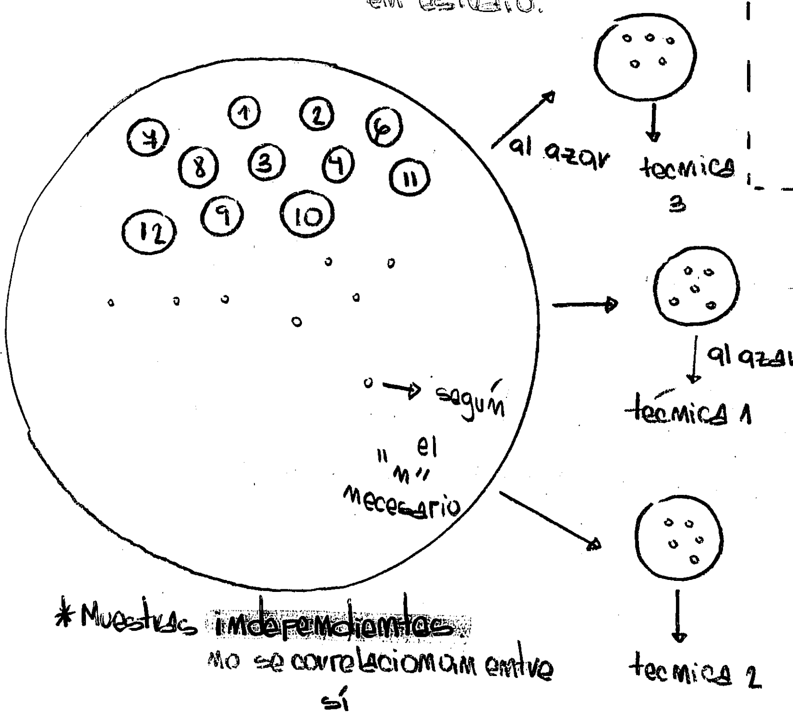
al aplicarle el mismo nivel del factor en estudio a c/u de ellas, reaccionan de la misma forma

diseño $\boxed{\text{completamente}}$ aleatorizado.

¿por qué hablamos de completamente?

→ tanto las subpoblaciones como el nivel del factor son designadas al azar.

Habrá tantas subpoblaciones ó subconjuntos como niveles del factor en estudio.



*Muestras independientes
no se correlacionan entre sí

② *unidades experimentales heterogéneas

las unidades difieren entre sí por diferentes condiciones experimentales y aplicamos el mismo nivel del factor y dan ≠ resultados de la variable "x"

diseño en bloques completos aleatorizados

¿por qué completos?

- se habla que c/u de los bloques se constituyen por unidades homogéneas, pero los tratamientos a c/u de ellas son asignados al azar

c/u de las unidades experimentales que constituyen el bloque

bloque 1	16	13	11
bloque 2	5	6	7
bloque 3	1	2	3
	t_1	t_2	t_3

} heterogéneos entre bloques

Los bloques se constituyen con tantas unidades experimentales como tantos niveles del factor haya

* Muestras dependientes
se correlacionan entre sí

- Método estadístico para tomar decisiones ⟶ ensayo de hipótesis en base a dos muestras

?
La forma en que obtienes las muestras determina

tipo muestras (+)

Luego analizamos el tipo de distribución y en base a eso **Elección del test**

analizaremos

población ⟶ distribución normal / distribución libre.

(+) tipo de unidades ⟶ homogeneas / heterogeneas

- ¿Qué estudiamos? (en el próximo estudiamos mas de dos niveles) ↓ en nuestro curso!
el efecto de la variación de dos niveles de un factor en estudio sobre una variable aleatoria

- planteamos las hipotesis en base a parámetros de dos poblaciones que se diferencian entre sí por los niveles del factor en estudio.

¿Qué puede interesar comparar?

- promedio ⟶ $\mu_1 = \mu_2$ (H_0
- variancia ⟶ $\sigma_1^2 = \sigma_2^2$ (H_0
- probabilidad ⟶ $P_1(A) = P_2(A)$ (H_0.

La desición de rechazar H_0 o no, se tomará en base a las **dos muestras**.

? cómo obtuve mi muestra

⟶ diseño experimental completamente aleatorizado.

⟶ los niveles del factor están asignados naturalmente a cada unidad experimental (sin diseño previo).

Independientes

Ejemplo = Se desea comparar el nivel de colesterol entre hombres y mujeres

Factor = sexo, niveles = F; M

ya están asignados naturalmente a cada población

- diseño experimental en bloques completos aleatorizados

- A cada unidad experimental se le aplican los dos niveles del factor.

1) - Muestras independientes → distribución normal

comparación de promedios de dos poblaciones normales.

Sean x_1 y x_2 variables aleatorias con distribución normal

si es variancia divido por n, sino divido por $\sqrt{n}$ cuando es desvío

- $x_1 \rightarrow N(u_1; \sigma_1^2)$
- $x_2 \rightarrow N(u_1; \sigma_2^2)$.

$$\Longrightarrow \quad \bar{x}_1 \rightarrow N\left(u_1; \frac{\sigma_1^2}{n_1}\right) \qquad \bar{x}_2 \rightarrow N\left(u_2; \frac{\sigma_2^2}{n_2}\right).$$

- Variancias conocidas !!

deseamos probar que =

$$H_0)\ u_1 = u_2 \qquad\qquad H_1)\ u_1 > u_2$$

$$u_1 - u_2 = 0 \qquad\qquad u_1 - u_2 > 0$$

} ensayo unilateral por derecha.

Estadística base $= \bar{x}_1 - \bar{x}_2$ } pero qué distribución tiene la estadística base? !! O

Entonces por el teorema

$$\underbrace{\bar{x}_1 - \bar{x}_2}_{\cancel{w}\ d} \rightarrow N\left(u_1 - u_2; \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}\right)$$

$$P(E_1)\ P(\text{rechazar } H_0 / H_0 \text{ cierta}) = \alpha$$

$$P(E_1) = P\left(u_1 - u_2 > 0 / u_1 - u_2 = 0\right) = \alpha$$

$$= P\left(\underbrace{\bar{x}_1 - \bar{x}_2}_{\cancel{w}\ d} > c\alpha \,/\, \underbrace{u_1 - u_2}_{u_w} = 0\right) = \alpha$$

valor poblacional

$$= P\left(\frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > \frac{c\alpha - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) = \alpha$$

es la raíz cuadrada x' que estandarizas en función del desvío

Teorema de combinación lineal de variables normales

- sean $y_1; \ldots; y_n$; n variables normales independientes con promedios $u_1, \ldots, u_n$ y variancias $\sigma_1^2; \ldots; \sigma_n^2$ -

- definimos nueva variable w, combinación lineal de la y_j variables

$$w = \sum_{j=1}^{n} y_j \cdot a_j$$

Entonces, la nueva variable w, también tiene distribución normal

$$w \rightarrow N\left(\sum_{j=1}^{n} a_j \cdot u_j;\ \sum_{i=1}^{n} a_j \cdot \sigma_j^2\right).$$

$$= P\left(Z > \underbrace{Z_{1-\alpha}}\right) = \alpha$$

$$Z_{1-\alpha} = \frac{C\alpha - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\boxed{Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = C\alpha}$$

Importante

$$\underbrace{\bar{X}_1 - \bar{X}_2}_{} \longrightarrow N\left(u_1 - u_2 ; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

esto no se denota como variable diferencia ¡ojo!

deseamos probar que =

$H_0)$ $u_1 = u_2$

$u_1 - u_2 = 0$

$H_1)$ $u_1 < u_2$ ensayo unilateral por izquierda

$u_1 - u_2 < 0.$

Estadística base $= \bar{x}_1 - \bar{x}_2$

por. T.C.L variables normales definimos

$$\bar{d} \longrightarrow N\left(u_1 - u_2 ; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$P(E_1) = P(\text{rechazar } H_0 / H_0 \text{cierta}) = \alpha$

$$= P\left(u_1 - u_2 < 0 / u_1 - u_2 = 0\right) = \alpha$$

$$= P\left(\bar{x}_1 - \bar{x}_2 < C\alpha / \underbrace{u_1 - u_2 = 0}\right) = \alpha$$

producto de la nueva variable definida

$$= P\left(\frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < \frac{C\alpha - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) = \alpha$$

$$= P\left(Z < \underbrace{Z_\alpha}_{-Z_{1-\alpha}}\right) = \alpha$$

$$-Z_{1-\alpha} = \frac{C\alpha - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\boxed{-Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = C\alpha}$$

* Variancias desconocidas.

1° paso $\longrightarrow$ Test de igualdad de variancias.

rechazo Ho si $F_0 = \dfrac{s_1^2}{s_2^2}$; (siendo $s_1^2 > s_2^2$)

$$F_0 > C\alpha$$

Estadística básica del test $= \dfrac{s_1^2}{s_2^2}$

$$P(E_1) = P(\text{rechazar Ho}/\text{Ho cierta}) = \alpha$$

$$P\left(F_0 > C\alpha / \sigma_1^2 = \sigma_2^2\right) = \alpha$$

$$P\left(F_{n_1-1;\, n_2-1} > \underbrace{C\alpha}_{F_{n_1-1;\, n_2-1;\, 1-\alpha}}\right) = \alpha$$

$$\Longrightarrow F_{n_1-1;\, n_2-1} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \quad \text{si } \sigma_1^2 = \sigma_2^2 \Longrightarrow F = \frac{s_1^2}{s_2^2} \longrightarrow F_{n_1-1;\, n_2-1}.$$

si rechazo Ho $\Longrightarrow \sigma_1^2 \neq \sigma_2^2$ (2)

si no rechazo Ho $\Longrightarrow \sigma_2^2 = \sigma_1^2$ (1)

(1) Comparación de promedio para varianzas supuestamente iguales. $\sigma_1^2 \cong \sigma_2^2$

$$\bar{x}_1 - \bar{x}_2 = \bar{d} \longrightarrow N\left(\mu_1-\mu_2 ; \sqrt{\frac{\sigma_1^2}{m_1}+\frac{\sigma_2^2}{m_2}}\right).$$

$\downarrow$ como $\sigma_1^2 \cong \sigma_2^2$

$$\sigma \cdot \sqrt{\frac{1}{m_1}+\frac{1}{m_2}} = Sa \cdot \sqrt{\frac{1}{m_1}+\frac{1}{m_2}}$$

$$Ho) \; \mu_1 = \mu_2 \qquad \underbrace{\mu_1-\mu_2}_{d} = 0$$

$$H_1) \; \mu_1 > \mu_2 \qquad \underbrace{\mu_1-\mu_2}_{d} > 0$$

} ensayo unilateral por derecha.

Estadística base $= \bar{x}_1 - \bar{x}_2 = \bar{d}$ MO.

$$P(E_1) = P\left(\text{rechazar Ho} / \text{Ho cierta}\right) = \alpha$$

$$P\left(\mu_1-\mu_2 > 0 / \mu_1-\mu_2 = 0\right) = \alpha$$

$$P\left(\bar{x}_1-\bar{x}_2 > C\alpha / \mu_1-\mu_2 = 0\right) = \alpha$$

$$P\left(\frac{\bar{d}-0}{Sa \cdot \sqrt{\frac{1}{m_1}+\frac{1}{m_2}}} > \underbrace{\frac{C\alpha - 0}{Sa\sqrt{\frac{1}{m_1}+\frac{1}{m_2}}}}\right) = \alpha$$

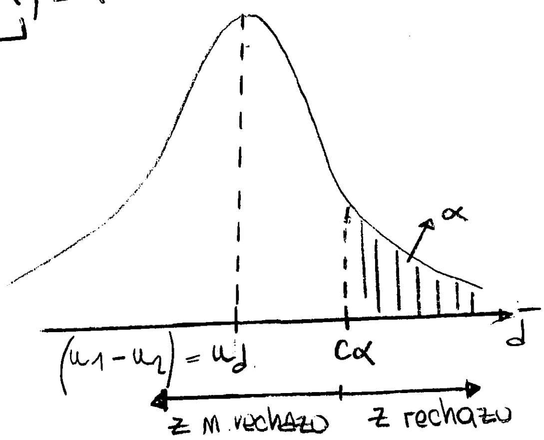
$$P\left(t_{m_1+m_2-1} > \underbrace{t_{m_1+m_2 \overset{-2}{-1};\, 1-\alpha}}\right) = \alpha$$

$$t_{m_1+m_2\overset{-2}{-1};\,1-\alpha} = \frac{C\alpha - 0}{Sa\sqrt{\frac{1}{m_1}+\frac{1}{m_2}}}$$

$$\boxed{t_{m_1+m_2\overset{-2}{-1};\,1-\alpha} \cdot Sa\sqrt{\frac{1}{m_1}+\frac{1}{m_2}} = C\alpha}$$

ojo que es

$t_{m_1+m_2-2}$ ✓



$H_0)\ u_1 = u_2$ $\qquad$ $H_1)\ u_1 < u_2$

$\underbrace{u_1 - u_2}_{d} = 0$ $\qquad$ $\underbrace{u_1 - u_2}_{d} < 0$

Estadistico base $= \underbrace{\bar{X}_1 - \bar{X}_2}_{\bar{d}}$

$$\bar{d} \longrightarrow N\left(u_1 - u_2 ; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_2}}\right)$$

$\downarrow$ siendo $\sigma_1^2 \simeq \sigma_2^2$

$$\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = Sa \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$P(E_1) = P\left(\text{rechazar } H_0 / H_0 \text{ cierta}\right) = \alpha$$

$$P\left(u_1 - u_2 < 0 \,/\, u_1 - u_2 = 0\right) = \alpha$$

$$P\left(\underbrace{\bar{X}_1 - \bar{X}_2}_{\bar{d}} < C\alpha \,/\, u_1 - u_2 = 0\right) = \alpha$$

$$P\left(\frac{\bar{d} - 0}{Sa \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < \underbrace{\frac{C\alpha - 0}{Sa \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}\right) = \alpha$$

$$P\left(t_{n_1 + n_2 - 2} < \underbrace{t_{n_1 + n_2 - 2 ; \alpha}}_{= -t_{n_1 + n_2 - 2 ; 1-\alpha}}\right) = \alpha$$

$$-t(n_1 + n_2 - 2 ; 1 - \alpha) = \frac{C\alpha - 0}{Sa \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\boxed{-t(n_1 + n_2 - 1 ; 1 - \alpha) \cdot Sa \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = C\alpha}$$

(2) comparación de promedios para variancias supuestamente $\neq$.

$$\sigma_1^2 \neq \sigma_2^2$$

$H_0)\ u_1 = u_2$

$\underbrace{u_1 - u_2 = 0}_{d}$

$H_1)\ u_1 > u_2$

$u_1 - u_2 > 0$ } ensayo unilateral por derecha

$\underbrace{d \longrightarrow N\left(u_1 - u_2;\ \sigma_1^2 + \sigma_2^2\right).}_{\text{consultar.}}$

Estadística base del test $= \bar{d} = \bar{x}_1 - \bar{x}_2$

$$P(E_1) = P(\text{rechazar } H_0 / H_0 \text{ cierta}) = \alpha$$

$$P(u_1 - u_2 > 0 / u_1 - u_2 = 0) = \alpha$$

$$P(\underbrace{\bar{x}_1 - \bar{x}_2}_{\bar{d}} > C\alpha / u_1 - u_2 = 0) = \alpha$$

$$P\left(\frac{\bar{d} - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} > \underbrace{\frac{C\alpha - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}}\right) = \alpha$$

* estudiar formula F = y Sa (tabla de fórmulas)

$$P(t_F > \underbrace{t_{F;1-\alpha}}) = \alpha$$

$$t_{F;1-\alpha} = \frac{C\alpha - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t_{F;1-\alpha} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = C\alpha.$$

- Muestras independientes / comparación de dos probabilidades.
(válido para muestras $n \geq 30$).

$H_0)\ p_1 = p_2$

$p_1 - p_2 = 0$

$H_1)\ p_1 > p_2$

$p_1 - p_2 > 0.$ } ensayo unilateral por derecha

Estadística base del test = $h_1 - h_2$

$$h_1 \xrightarrow[n \geq 30]{} N\left(p_1 ; \frac{p_1 q_1}{n_1}\right) \text{ varianza}$$

$$h_2 \xrightarrow[n \geq 30]{} N\left(p_2 ; \frac{p_2 \cdot q_2}{n_2}\right)$$

$$h_1 - h_2 \xrightarrow[n \geq 30]{} N\left(\underbrace{p_1 - p_2}_{=0} ; \underbrace{p \cdot q}_{\text{pienso que Ho cierta}} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

$$N\left(p_1 - p_2 ; \frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}\right)$$

$p_1 = p_2 \to$ lo saco como factor!

$$P(E_1) = P(\text{rechazar Ho} / \text{Ho cierta}) = \alpha$$

$$P(p_1 - p_2 > 0 / p_1 - p_2 = 0) = \alpha$$

$$P(h_1 - h_2 > c_\alpha / p_1 - p_2 = 0) = \alpha$$

$$P\left(\frac{h_1 - h_2 - 0}{\sqrt{p \cdot q \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > \frac{c_\alpha - 0}{\sqrt{p \cdot q \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right) = \alpha$$

$$P(Z > Z_{1-\alpha}) = \alpha$$

$$Z_{1-\alpha} = \frac{c_\alpha - 0}{\sqrt{p \cdot q \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\boxed{Z_{1-\alpha} \cdot \sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = c_\alpha}$$

- Muestras dependientes $\longrightarrow$ distribución normal.

Ensayo hipotesis para comparación de promedios.
muestras dependientes.

. Sean $(x_{11}; x_{21})$, $(x_{12}; x_{22})$ pares de observaciones que alcanzan el último par como $(x_{1m}; x_{2m})$ conjunto de m pares de observaciones.

par	tratamiento 1	tratamiento 2	diferencia
1	x_{11}	x_{21}	$x_{11} - x_{21} = d_1$
2	x_{12}	x_{22}	$x_{12} - x_{22} = d_2$
...	$x_{1...}$	$x_{2...}$	$x_{1...} - x_{2...} = d_{...}$
m	x_{1m}	x_{2m}	$x_{1m} - x_{2m} = d_m$

nueva variable d

- suponemos que "d" $\longrightarrow N(u_d; \sigma_d^2)$
- entonces $\bar{d} \longrightarrow N\left(u_d; \frac{\sigma_d}{\sqrt{m}}\right)$.

$H_0)\ u_d = 0$

$H_1)\ u_d > 0$ } Ensayo unilateral por derecha.

- Estadística base = $\bar{d}$

$P(E_1) = P(\text{rechazar } H_0 / H_0 \text{ cierta}) = \alpha$

$= P(u_d > 0 / u_d = 0) = \alpha$

$= P(\bar{d} > C\alpha / u_d = 0) = \alpha$

$= P\left(\frac{\bar{d} - 0}{\cancel{\sigma_d}/\sqrt{m}} > \frac{C\alpha - 0}{\cancel{\sigma_d}/\sqrt{m}}\right) = \alpha$ $\longrightarrow$ no tenemos dato poblacional

(sd sobre σ_d tachado)

~~$P(Z > Z_{1-\alpha}) = \alpha$~~

$P(t_{m-1} > t_{m-1;1-\alpha}) = \alpha$

$t_{m-1;1-\alpha}$

$\cancel{Z_{1-\alpha}} = \frac{C\alpha - 0}{\cancel{\sigma_d}/\sqrt{m}}$ (sd)

$$\boxed{Z_{1-\alpha} \cdot \frac{\sigma_d}{\sqrt{m}} = C\alpha}$$

$t_{m-1;1-\alpha} \cdot \frac{s_d}{\sqrt{m}} = C\alpha$

Ensayo para comparación de probabilidad de dos sucesos. Muestras independientes

$H_0)\ p_1 = p_2$ $\qquad$ $H_1)\ p_1 \neq p_2$

$p_1 - p_2 = 0$ $\qquad$ $p_1 - p_2 \neq 0$

consultar si esta forma es válida.

$h_1 \xrightarrow[n>30]{} N\left(p_1 ; \frac{p_1 \cdot q_1}{n_1}\right)$

Estadística base p' test = $h_1 - h_2$.

$h_2 \xrightarrow[n>30]{} N\left(p_2 ; \frac{p_2 \cdot q_2}{n_2}\right)$.

$P(E_1) = P\left(\text{rechazar } H_0 / H_0 \text{ cierta}\right) = \alpha$

$\Rightarrow h_1 - h_2 \longrightarrow N\left(\underbrace{p_1 - p_2 = 0 ; \sqrt{p \cdot q \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}_{}\right)$

H_0 cierta. se deduce de acá

$P(E_1) = P\left(p_1 - p_2 \neq 0 / p_1 - p_2 = 0\right) = \alpha$

$N\left(p_1 - p_2 = d ; \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right)$

$P(E_1) = P\left(h_1 - h_2 < C_1 / p_1 - p_2 = 0\right) = \frac{\alpha}{2}$ ó $P(E_1) = P\left(h_1 - h_2 > C_2 / p_1 - p_2 = 0\right) = \frac{\alpha}{2}$

$$P\left(\frac{h_1 - h_2 - 0}{\sqrt{p \cdot q \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} < \frac{C_1 - 0}{\sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right) = \frac{\alpha}{2}$$

$$P\left(\frac{h_1 - h_2 - 0}{\sqrt{p \cdot q \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > \frac{C_2 - 0}{\sqrt{p \cdot q \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right) = \frac{\alpha}{2}$$

$$P\left(Z < Z_{\frac{\alpha}{2}} = -Z_{1-\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

$$P\left(Z > Z_{1-\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

$$-Z_{1-\frac{\alpha}{2}} = \frac{C_1 - 0}{\sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$Z_{1-\frac{\alpha}{2}} = \frac{C_2 - 0}{\sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = C_1$$

$$0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{p \cdot q \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = C_2$$

tratamiento 2

tratamiento 1	A	$\bar{A}$	T
A		$f(A;\bar{A})$	
$\bar{A}$	$f(\bar{A};A)$		
T'			m'

$$m' = f(\bar{A};A) + f(A;\bar{A}).$$

defino $p'_1(A) = P(\overset{\text{tratamiento 1}}{A};\bar{A})$ = probabilidad del suceso A en el tratamiento 1 en la subpoblación que manifestaron respuestas distintas en cada tratamiento. → si bien hay diferentes respuestas son dependientes porque a cada unidad se aplica ambos tratamientos.

- planteamos hipotesis =

$H_0)\ p'_1(A) = p'_2(A)$ $\qquad$ $H_1)\ p'_1(A) < p'_2(A)$; ensayo unilateral por izquierda.

equivalentemente =

$H_0)\ p_1'(A) = 0{,}5$ $\qquad$ $H_1)\ p'_1(A) < 0{,}5$

Estadistica base del test = $h' = \dfrac{f(A;\bar{A})}{m'}$

si $m' > 30$ =

$$h' \longrightarrow N\left(p'_1 ; \frac{p'_1 \cdot (1-p'_1)}{m'}\right)$$

$$h' \xrightarrow[\text{cierta}]{H_0} N\left(0{,}5 ; \frac{0{,}5 \times 0{,}5}{m'}\right)$$

$$P(E_1) = P(\text{rechazar } H_0 / H_0 \text{ cierta}) = \alpha$$

$$P(p'_1(A) < 0{,}5 / p'_1(A) = 0{,}5) = \alpha$$

$$P(h' < C\alpha / p'_1 = 0{,}5) = \alpha$$

$$P\left(\frac{h' - 0{,}50}{\sqrt{\frac{0{,}50 \times 0{,}50}{m'}}} < \frac{C\alpha - 0{,}5}{\sqrt{\frac{0{,}5 \times 0{,}5}{m'}}}\right) = \alpha$$

$$P(Z < \underbrace{Z_\alpha}_{-Z_{1-\alpha}}) = \alpha \Longrightarrow -Z_{1-\alpha} = \frac{C\alpha - 0{,}5}{\sqrt{\frac{0{,}5 \times 0{,}5}{m'}}}$$

$$\boxed{0{,}5 - Z_{1-\alpha} \cdot \sqrt{\frac{0{,}5 \cdot 0{,}5}{m'}} = C\alpha.}$$

Otras formas de rechazo =

* rechazo Ho sí =

$$t_{observada} = \frac{\bar{d} - 0}{S_d/\sqrt{n}} > t_{crítica} = t_{(n-1\,;\,1-\alpha)}.$$

* rechazo Ho sí $P_{asociada} < \alpha$

$$P_{asociada} = P\left(t_{n-1} > t_{n-1;\,1-\alpha}\right) = P\left(\bar{d} > \bar{d}'_{observado}\right)$$

Luego comparamos con nivel de significación del ensayo.

- Muestras dependientes / comparación de probabilidades.

Ensayo de hipótesis

- una misma unidad experimental se somete a dos distintos tratamientos y la variable es cualitativa.

unidad	tratamiento 1	tratamiento 2
1	A	A
2	$\bar{A}$	A
ooo		
ooo		
ooo		
ooo		
n − 1	A	$\bar{A}$
n	$\bar{A}$	$\bar{A}$

tratamiento 2

tratamiento 1	A	$\bar{A}$	+
A	$\cancel{f(A;A)}$	$f(A;\bar{A})$	
$\bar{A}$	$f(\bar{A};A)$	$\cancel{f(\bar{A};\bar{A})}$	
+			n

• no me estaría brindando datos para mi interés rechazo H_0 sí =

$$Z_{observada} = \frac{h' - 0{,}5}{\sqrt{\frac{0{,}5 \cdot 0{,}5}{n'}}} < -Z_{1-\alpha}$$

rechazo H_0 sí

$$P_{asociada} = \underbrace{P(h < h')}_{\text{equivalente}} < \alpha$$

$P(Z < -Z_{1-\alpha}) = ?$ → comparo con el α.

Ahora bien... cuando la distribución es libre llamamos "libre distribución" → no paramétrico.

Las hipótesis no se definen sobre parámetros

- el único supuesto es que sea continua la variable.

- Tests a libre distribución =
  - Test Wilcoxon (1)
  - Test Mann Whitney (2)

(1) Test wilcoxon = prueba de rangos de signo para muestras de observaciones pareadas.

↓

muestras dependientes.

Ejemplo = comparar contaminación de CO_2 en ≠ horarios (8am y 10am) en diferentes esquinas de la ciudad.

H_0) las distribuciones para X_1 y X_2 tienen la misma localización

H_1) las dos distribuciones tienen diferentes localizaciones (ensayo bilateral). o dos colas

H_1) la distribución de X_1 está desplazada hacia la derecha de X_2 (ensayo unilateral o una cola)

↓ (también puede ser al revés).

¿Cómo procederíamos?

1° Cálculo todas las variables "d" y realizo gráfico de probabilidad normal, si $d \not\to N(u_d; \sigma_d^2)$

$\Downarrow$

$$\boxed{d_j = X_{aj} - X_{bj}}$$

* aplico Test Wilcoxon

- eliminamos diferencias nulas.
- se asigna el rango a cada uno de los valores absolutos de las diferencias (si dos o más valores resultan iguales se asigna a c/u de ellos el promedio de los rangos que se hubieran asignado a c/u y se continúa enumerando con el siguiente nº.

Calculamos
→ suma de rangos para las diferencias negativas ($T-$)
→ " " " para las diferencias positivas ($T+$).

- en una prueba de 2 colas ⟶ rechazo Ho si: $\underbrace{T(\text{min} - \text{ó} +)}_{\text{tomo como referencia } T \text{ menor.}} \leq T_0$

T_0 ⟶ valor crítico (tabla) 9 $\xrightarrow{7}$ niveles de significancia para prueba bilateral.

- en prueba de una cola (b desplazado hacia la derecha de a) ⟶ rechazo
$a - b < 0$
Ho si = $T+ \leq T_0$

- en prueba de una cola (a desplazado hacia la derecha de b) ⟶ rechazo Ho si
$a - b > 0$

$T- \leq T_0$.

② Test Mann - whittney para "muestras independientes"

se aplica sí $X_1 \not\rightarrow N$, $X_2 \not\rightarrow N$.

¿Cómo procedemos? definimos estadísticas.

U_a = se obtiene contando valores de la muestra A que preceden cada valor de la muestra b. U_a es la suma de todas esas enumeraciones.

U_b = se obtiene contando el nº de valores de la muestra "b" que preceden la muestra a. U_b es la suma de todas esas enumeraciones.

$$\boxed{U_a + U_b = n_1 \times n_2}$$

U_0, es el valor crítico ← tabla 8.

- valor que acumula $1-\alpha$ prob. (1 cola)
- valor que acumula $1-\frac{\alpha}{2}$ prob. (2 colas).

- Para test de una cola (b desplazado a la derecha de a), se rechaza Ho sí = $U_b \leq U_0$.

- Para test de una cola (a desplazado a la derecha de b), rechazo Ho sí = $U_a \leq U_0$.

- Para test de dos colas, rechazo Ho cuando U (el menor de U_a o U_b) $\leq U_0$

Observación = si se obtienen observaciones con el mismo valor en ≠ muestras el cálculo de las estadísticas U_a y U_b no pueden efectuarse de esta forma, debemos aplicar otro procedimiento.

Unidad nº 4. Objetivo = comparar 2 o más poblaciones. (2 o más muestras).

→ planificación experimento p/ comparar efectos de varias condiciones experimentales sobre una o más variables de interés en ciertos tipos de unidades.

Ejemplos =

1) El efecto de dos antibióticos, sobre infección bacteriana
producirán cambios experimentales sobre la variable x
variable x.

Definiciones

* Factores o criterios = características que distinguen o diferencian las poblaciones entre sí.

En el ejemplo = antibiotico.

* Nivel = modalidad factor en estudio.

1 Factor en estudio

= tratamiento

En el ejemplo, nivel o tratamiento = antibiótico a, antibiótico b.

* unidad experimentales = mínima porción de material experimental al que asignamos un tratamiento y luego medimos el valor de la variable.

En el ejemplo, unidad experimental = paciente.

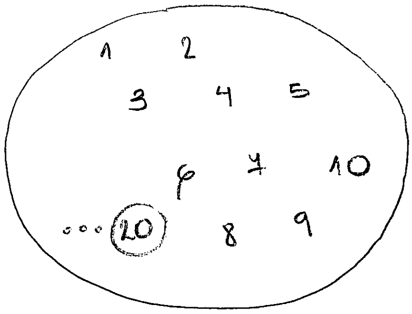
Tipos de experimentos
→ 1 diseño completamente aleatorizados.
→ 2 diseño en bloques completamente aleatorizados

1) unidades experimentales homogéneas → si se aplica el mismo nivel del factor en estudio reaccionan igual salvo por situaciones logicas del azar

- al azar le asigno aleatoriamente nivel del factor, y habrá tantos grupos como niveles del factor haya; esos grupos también son al azar.
→ muestras independientes.

* x'eso completamente aleatorizado.

2) comparar unidades experimentales heterogeneas. → unidades experimentales difieren en ciertas condiciones experimentales y al aplicarle el mismo nivel del factor en estudio dan resultados ≠ variable "x".



	M1	M2	M3	M4
bloque 1	16	13	10	5
bloque 2	18	7	12	15
bloque 3	3	14	6	19
bloque 4	17	1	4	11
bloque 5	2	9	20	8

} entre bloques son heterogeneos entre sí.

→ cada bloque es homogeneo entre sí y habrá tantas unidades que lo compongan como niveles de factor en estudio.

→ muestras dependientes.

Una vez que definimos el tipo experimento, ensayamos la hipotésis.

* se estudia el efecto que produce la variación de dos niveles de un factor en estudio.

antibiótico $\begin{cases} \nearrow a \\ \searrow b \end{cases}$

* planteamos hipotesis respecto parámetros de dos poblaciones que se diferencian en los niveles del factor.

Por ejemplo:

$H_0)\ \mu_1 = \mu_2$ → comparamos el promedio de dos poblaciones

$H_0)\ \sigma_1^2 = \sigma_2^2$ → " precisión de dos poblaciones

$H_0)\ P_1(A) = P_2(A)$ → comparamos probabilidad de dos poblaciones.

* La desición de rechazar o no rechazar H_0 se tomará en base a dos muestras. → la técnica estádistica depende del tipo de muestra.

* Muestras independientes. → comparación promedios p' dos poblaciones normales.

(variancia poblacional conocida).

→ comp. de probabilidad (muestras grandes).

↓

comparación promedios p' poblaciones normal (variancia poblacional desconocida)

* planteamos → corroboramos si pueden suponerse = ó ≠.

$H_0)\ \sigma_1^2 = \sigma_2^2 \qquad H_1)\ \sigma_1^2 > \sigma_2^2$

$P(E_1) = \alpha$ 1 es el numerador, 2 el denominador 2   siendo $S_{x_1}^2 > S_{x_2}^2$ ⊛ considerar los dos ($n_i = 1, 2$)

rechazo $H_0)$ si $F_{obs} = \dfrac{S_1}{S_2^2} > F_{(n_i - 1);(1-\alpha)}$

O sí: $P_{asoc} < \alpha$

⊛ continua atrás.

Sean x_1 y x_2 variables con distribución normal, la primera con parámetros $(\mu_1; \sigma_1)$ y la segunda $(\mu_2; \sigma_2)$.

Esto implica que:

$\bar{X}_1 \to N\left(\mu_1; \dfrac{\sigma_1^2}{\sqrt{n_1}}\right) \qquad \bar{X}_2 \to N\left(\mu_2; \dfrac{\sigma_2^2}{\sqrt{n_2}}\right)$

deseamos probar que =

$H_0)\ \mu_1 = \mu_2 \qquad H_1)\ \mu_1 > \mu_2$

es decir

$\mu_1 - \mu_2 = 0 \qquad \mu_1 - \mu_2 > 0$

Estadistica base test = $\bar{X}_1 - \bar{X}_2$.

↓

pero cómo se distribuye esta estadística?

x! Teorema combinación Lineal de variables normales:

"sea $y_1, \ldots, y_n$ variables normales independientes con promedio $\mu_1, \ldots; \mu_n$ y variancia $\sigma_1^2, \ldots, \sigma_n^2$.

se define nueva variable w_j combinación lineal de variable y_j =

$$W = \sum_{j=1}^{n} a_j \cdot y_j$$

entonces la nueva variable también se distribuye normalmente"

$$W \to N\left(\sum_{j=1}^{n} a_j \mu_j; \sum_{j=1}^{n} a_j^2 \sigma_j^2\right)$$

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

se rechaza H_0 sí =

$\bar{X}_1 - \bar{X}_2 >$ [redacted] ; $P(e_1) = \alpha$

↘ variancia pob. conocida

$$c\alpha = 0 + Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

equivalentemente =

$$Z_{obs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > \underbrace{Z_{1-\alpha}}_{Z\ \text{crítico}\ \alpha}$$

o sí $P_{asociada} < \alpha$

$P(\bar{X}_1 - \bar{X}_2 > (\bar{X}_1 - \bar{X}_2)')$

(*) puede estimarse una variancia en común o no.

si las variancias son iguales (test previo no rechazo Ho)

si las variancias son diferentes (test previo rechazo Ho) (*)₁

se estima variancia en común = $S_M^2 = \frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{n_1+n_2-2}$

y p'ensayar prueba $H_0)\ \mu_1-\mu_2=0$ vs $H_1)\ \mu_1>\mu_2$

$P(E_1)=\alpha$, se rechaza Ho sí =

$$\bar{X}_1-\bar{X}_2 > C\alpha = T(n_1+n_2-2);(1-\alpha)\cdot S_a\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}.$$

* equivalentemente

$P_{asoc} < \alpha$ ó $T_{obs} = \frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} > t(n_1+n_2-2);1-\alpha$

(*)₁ no estimamos variancia en común

$H_0)\ \mu_1-\mu_2=0$ $H_1)\ \mu_1-\mu_2>0$

$$\bar{X}_1-\bar{X}_2 > C\alpha = t_f;(1-\alpha)\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

* equivalentemente = $t_{obs} = \frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = t_f;(1-\alpha)$

formulita

ó $P_{asociada} < \alpha$

* comparación de dos probabilidades para muestras grandes.

$H_0)\ p_1(A) = p_2(A)$ $\qquad H_1)\ p_1(A) > p_2(A)$

$p_1(A) - p_2(A) = 0$ $\qquad p_1(A) - p_2(A) > 0.$

* estadística basica test = $h_1 - h_2$

Si n_1 y n_2 son lo suficientemente grandes, (> 30), se distribuye normalmente

entonces =

$$h_1 \longrightarrow N\left(p_1 ; \frac{p_1(1-p_1)}{n_1}\right) \qquad h_2 \longrightarrow N\left(p_2 ; \frac{p_2(1-p_2)}{n_2}\right).$$

entonces distribución estadística base =

$$h_1 - h_2 \longrightarrow N\left(p_1 - p_2 = 0 ; p(1-p) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

estimamos p con h = $h = \frac{f_1 + f_2}{n_1 + n_2}$

Rechazo H_0 sí =

$$h_1 - h_2 > C_\alpha = 0 + z_{1-\alpha} \cdot \sqrt{h(1-h)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Muestras dependientes:

par tratamiento

par	Tratami. 1	Tratami. 2	diferencia
1	X_{11}	X_{12}	$d_1 = X_{11} - X_{12}$
2	X_{21}	X_{22}	$d_2 = X_{21} - X_{22}$
3			
...			
m			

nueva variable "d" =

$d_m = X_{1.m} - X_{2.m}$.

distribución nueva variable

$$d \longrightarrow N(\mu d; \sigma_d^2) \Longrightarrow \bar{d} \longrightarrow N\left(\mu d; \frac{\sigma d}{\sqrt{m}}\right)$$

comparación dos promedios =

$H_0)\ \mu d = 0 \qquad H_1)\ \mu d > 0$

Estadística base: $\bar{d}$ $\quad P(E_1) = \alpha$

rechazo H_0 si:

$$\bar{d} > C\alpha = 0 + t_{(m-1);1-\alpha} \frac{sd}{\sqrt{m}}$$

equivalentemente

$$T_{observada} = \frac{\bar{d}}{sd/\sqrt{m}} > \underbrace{t_{(m-1);(1-\alpha)}}_{\text{Teórica}}$$

o sí = $P_{asociada} < \alpha$.

$$P(\bar{d} > \bar{d}'_{observado}) = P\left(\frac{\bar{d} - \mu d}{\sigma d/\sqrt{m}} > \underbrace{\frac{\bar{d}' - \mu d}{\sigma d/\sqrt{m}}}_{n^o}\right)$$

T

$1 - P(T < n^o) = P_{asoc}$.

comparación de dos probabilidades.

unidad	Trat 1	Trat 2
1	A	$\bar{A}$
2	$\bar{A}$	A
...	...	...
m	$\bar{A}$	A

Trat 2

Trat 1	A	$\bar{A}$
A	$f(A;A)$	$f(A;\bar{A})$
$\bar{A}$	$f(A;\bar{A})$	$f(\bar{A};\bar{A})$

sólo ellos me dan info acerca de dos tratamientos

Tratamiento 2

Trat 1	A	$\bar{A}$	Total
A		$A\bar{A}$	
$\bar{A}$	$\bar{A}A$		
Total			n'

$n' = f(\bar{A}, A) + f(\bar{A}, A)$

si $p_1(A) = p_2(A)$ las unidades q' tuvieron respuestas $\neq$ se deben a su comportamiento al azar esperándose 50% haya presentado suceso A tratamiento 1 y $\bar{A}$ del tratamiento 2 $(A; \bar{A})$, mientras el otro 50% al revés $(\bar{A}; A)$.

planteamos hipótesis

$H_0)\ p_1(A) = p_2(A)$ vs $H_1)\ p_1(A) > p_2(A)$

estadística básica test =

$$h' = \frac{f(A; \bar{A})}{n'}$$

si $n > 30 \Longrightarrow h'$ tiene distribución normal

$$h' \longrightarrow N\left(p_1'; \frac{p_1'(1-p_1')}{n'}\right)$$

$$h' \xrightarrow[\text{cierto}]{H_0} N\left(0{,}5; \frac{0{,}5 \cdot 0{,}5}{n'}\right)$$

Luego con $P(E_1) = \alpha$, rechazo H_0 si:

$$h' > C\alpha = 0{,}50 + z_{1-\alpha}\sqrt{\frac{0{,}50 \cdot 0{,}50}{n'}}$$

equivalentemente

$$z_{obs} = \frac{h' - 0{,}5}{\sqrt{\frac{0{,}5 \cdot 0{,}5}{n'}}} > \underbrace{z_{1-\alpha}}_{z \text{ crítico}}$$

notar que $1 - p_1' = P(\bar{A}; A) = p_2' \Longrightarrow H_0)\ p_2'(A) = 0{,}5$

o $H_1)\ p'(A) < 0{,}5$.

Ejercicio 5 =

x = porcentaje de recuperación de cada comprimido (%)

n = 20 comprimidos

- Factor = temperatura
- niveles = temperatura "a" y temperatura "b"
- unidad experimental = comprimido.
- modelo experimental = completamente aleatorizado. ⟹ muestras independientes

Temperatura a	98,8	100,6	101,1	100,1	100	98,6	98,8	101,8	100,3	98,6
Temperatura b	96,6	96,1	98,2	97,2	95,6	97,6	96,9	93,8	96,4	97,3

$\bar{X}a = 99,87\,\%$ $\quad S_{xa} = 1,1324\,\%$ $\quad S^2_{xa} = 1,28\,\%^2$

$\bar{X}b = 96,57\,\%$ $\quad S_{xb} = 1,23\,\%$ $\quad S^2_{xb} = 1,51\,\%^2$

* variancia poblacional desconocida

$H_0)\ \sigma^2_{x_a} = \sigma^2_{x_b}$ $\qquad H_1)\ \sigma^2_{x_b} > \sigma^2_{x_a}$ $\qquad P(E_1) = 0,05$

siendo $S^2_{x_b} > S^2_{x_a}$

rechazamos H_0 si =

$$F_0 = \frac{S^2_{xb}}{S^2_{xa}} > F(\underline{n_2}-1, \underline{n_1}-1)\,;\,(1-\alpha)$$

siempre 1° el de mayor variancia muestral.

$$F_{obs} = \frac{1,51\,\%^2}{1,28\,\%^2} = 1,18$$

$$F(9;9)\,(0,95) = 3,18$$

$1,18 < 3,18 \implies$ no rechazo H_0

Conclusión = con un nivel de significación 5% se concluye que

$\sigma^2_{x_a} = \sigma^2_{x_b}$

* estimamos variancia común =

$$s_M^2 = \frac{s_{xa}^2 \cdot (n_1 - 1) + s_x^2 b\,(n_2 - 1)}{n_1 + n_2 - 2}$$

$$s_M^2 = \frac{1{,}28\%^2 \cdot 9 + 1{,}51\%^2 \cdot 9}{18}$$

$$s_M^2 = 1{,}395\%^2$$

$H_0)$ $\mu a = \mu b$
$\mu a - \mu b = 0$

$H_1)$ $\mu a - \mu b > 0$ → ensayo unilateral a la derecha.

rechazo Ho si = $(\overline{X}a - \overline{X}b) > \delta + t_{(n_1+n_2-2);(1-\alpha)} \cdot Sa \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

$\sqrt{s_M^2} = 1{,}18\,\%$

$\overline{X}a - \overline{X}b = 99{,}84\% - 96{,}54\%$
$= 3{,}3\,\%$

$C\alpha = 0 + t_{18;0{,}95} \cdot \cancel{1{,}395\%}\ 1{,}18\% \sqrt{\frac{1}{10} + \frac{1}{10}}$

$C\alpha = 1{,}73 \cdot \cancel{1{,}395\%}\ 1{,}18\% \cdot \sqrt{\frac{1}{10} + \frac{1}{10}}$

$C\alpha = 0{,}913$

$3{,}3\% > 0{,}91\,\%$ ⟹ rechazo Ho.

rechazo Ho si:
* v. observado est > Tcrítica

$$\frac{\overline{X}a - \overline{X}b}{Sa\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{18;0{,}95}$$

Pasociada < $\alpha = 0{,}05$

$P\left(\overline{X}_1 - \overline{X}_2 > (\overline{X}_1 - \overline{X}_2)'\right) =$

↳ consultar !! ◡

v. observado = 6,25 ⟹ 6,25 > 1,395

rechazo Ho.

Tcrítica = 1,395

$$P\left(\frac{(\overline{X_1}-\overline{X_2})-0}{S_a\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > \frac{3,3\,\%}{1,18\%\sqrt{\frac{1}{10}+\frac{1}{10}}}\right) =$$

$$P\left(T_9 > 6,25\right) =$$

$$1 - \overset{\text{consultar}}{P\left(T_9 < 6,25\right)} \cong 0$$

$\cong 0 < 0,05 \Longrightarrow$ rechazamos Ho.

Conclusión = con un nivel significación 5% se concluye que a mayor temperatura hay mayor degradación del componente activo.

<u>Ejercicio 6</u>

x = concentración paracetamol (% p p)

Factor = técnica

nivel o tratamiento = técnica "a" y técnica "b".

unidad experimental = lotes

- modelo experimental = modelo en bloque completamente aleatorizados $\Longrightarrow$ muestras dependientes

$P(E_1) = 0,05$

$d \longrightarrow N\left(\mu_d ; \sigma_d^2\right)$

d (% p/p)

1	1,48
2	0,66
3	0,24
4	0,21
5	-0,1
6	-0,61
7	-0,1
8	0,09
9	-0,07
10	-0,21

$\bar{d} = 0,159\ \%pp \longrightarrow N\left(\mu_d ; \frac{\sigma_d}{\sqrt{n}}\right)$

$S_d = 0,57\%\ p/p \quad S^2_d = 0,32\,(\%\,p/p)^2$

$H_0) \; u_d = 0$ $\qquad$ $H_1) \; u_d \neq 0$

ensayo bilateral.

rechazo H_0 si =

$$\bar{d} < 0 - t_{n-1,\,1-\frac{\alpha}{2}} \cdot \frac{s_d}{\sqrt{n}} \quad ó \quad \bar{d} > 0 + t_{n-1;\,1-\frac{\alpha}{2}} \cdot \frac{s_d}{\sqrt{n}}$$

$\bar{d} = 0{,}159 \; \% \; p/p$

$$C\alpha = \pm \, t_{9;\,0{,}975} \cdot \frac{0{,}54 \; \% \; p/p}{\sqrt{10}}$$

$$C\alpha = \pm \, 2{,}26 \cdot \frac{0{,}54 \; \% \; p/p}{\sqrt{10}}$$

$$C\alpha = \pm \, 0{,}41 \; \% \; p/p$$

$0{,}159 \; \% \; p/p > -0{,}41 \; \% \; p/p \Longrightarrow$ no rechazo H_0

no

$0{,}159 \; \% \; p/p < 0{,}41 \; \% \; p/p \Longrightarrow$ Rechazamos H_0.

Rechazamos H_0 si =

$$T_{obs} < -t\left(n-1;\, 1-\frac{\alpha}{2}\right) \quad ó \quad T_{obs} > t\left(n-1;\, 1-\frac{\alpha}{2}\right)$$

$$\frac{\bar{d} - 0}{s_d/\sqrt{n}} < -t_{9;\,0{,}975} \quad ó \quad \frac{\bar{d} - 0}{s_d/\sqrt{n}} > t_{9;\,0{,}975}$$

$$\frac{0{,}159 \; \% \; p/p}{0{,}54 \; \% \; p/p \,/\, \sqrt{10}} = T_{obs}.$$

$0{,}88211 = T_{obs}$

$0{,}88211 > -2{,}26 \Longrightarrow$ no rechazo H_0

$0{,}88211 < 2{,}26 \Longrightarrow$ no rechazamos H_0.

Rechazamos Ho si =

$\underbrace{u_d = 0}$ planteado en mi hipotesis

$\bar{d}$ esta a la derecha de u !!

$2\left(P_{asociada}\right) < \alpha$

$2\left(P_{asociada} = p = P\left(\bar{d} > \bar{d}'\right)\right)$

$$2\left[P\left(\frac{\bar{d}-0}{Sd/\sqrt{n}} > \frac{\bar{d}'-0}{Sd/\sqrt{n}}\right)\right]$$

$$2\left[P\left(T_9 > \frac{0{,}159 \% \, p/p}{0{,}57 \% \, p/p/\sqrt{10}}\right)\right] =$$

$$2\left[1 - P\left(T_9 < 0{,}8821\right)\right] =$$

$$2\left[1 - 0{,}80\right] = 0{,}40$$

$p = 0{,}40 > 0{,}05 \Longrightarrow$ no rechazamos Ho.

Conclusión = con un nivel de significación del 5% se concluye que el valor promedio del contenido de paracetamol es el mismo p/ ambas técnicas

Ejercicio 4

X = rendemiento promedio de soja kg/ha

Factor = variedad

nivel = ciclo intermedio, ciclo largo

unidad experimental = lotes de plantas

modelo experimental = completamente aleatorizado $\Rightarrow$ muestras son independientes

tes u_i = rendimiento promedio de la variedad ciclo intermedio.

u_l = rendimiento promedio de la variedad ciclo largo

u_l = rendimiento promedio de planta de soja c. intermedio

$\sigma^2_{x_i}$ = variancia poblacional del rendimiento de planta de soja c largo

$\sigma^2_{x_l}$ = " " " " " " "

$H_0)\ \sigma^2_{x_i} = \sigma^2_{x_f}$ $\qquad H_1)\ \sigma^2_{x_i} > \sigma^2_{x_f}$ $\qquad P(E_1) = 0{,}05$

siendo $S^2_{x_i} > S^2_{x_f}$

$$F_0 = \frac{S^2_{x_i}}{S^2_{x_f}} > F_{(n_i-1);(n_f-1);1-\alpha}$$

* varianza poblacional desconocida

$$F_{obs} = \frac{(201{,}5\ kg/ha)^2}{(174{,}3\ kg/ha)^2}$$

$F_{obs} = 1{,}34$ $\qquad F(14;18;0{,}95) \cong F(15;20;0{,}95) \cong 2{,}20$

$1{,}34 < 2{,}20 \implies$ no rechazo H_0

Conclusión = con un nivel significación 5% se concluye que $\sigma^2_{x_i} = \sigma^2_{x_f}$.

Se estima variancia en común:

$$S^2_n = \frac{(n_i-1) \cdot S^2_{x_i} + (n_f-1)\, S^2_{x_f}}{n_i + n_f - 2}$$

$$S^2_n = \frac{14 \cdot (201{,}5\ kg/ha)^2 + 18 \cdot (174{,}3\ kg/ha)}{15 + 19 - 2}$$

$$S^2_n = 34852{,}51\ kg/ha^2$$

$$\implies S_a = \sqrt{34852{,}51\ kg/ha^2}$$

$$S_a = 186{,}7\ kg/ha$$

$H_0)\ \mu_i = \mu_f$ $\qquad H_1)\ \mu_i > \mu_f$

$\mu_i - \mu_f = 0$ $\qquad \mu_i - \mu_f > 0$

se rechaza H_0 si:

$$(\bar{X}_i - \bar{X}_f) > 0 + t_{n_i+n_f-2\,;\,(1-\alpha)} \; S_a \sqrt{\frac{1}{n_i} + \frac{1}{n_f}}$$

$$\bar{X}_i - \bar{X}_f = 3001{,}3 \text{ kg/ha} - 2741{,}8 \text{ kg/ha} = 259{,}5 \text{ kg/ha}$$

$$C\alpha = t_{32\,;\,0{,}95} \cdot 186{,}7 \text{ kg/ha} \cdot \sqrt{\frac{1}{15} + \frac{1}{19}}$$

$$C\alpha \cong t_{40\,;\,0{,}95} \cdot 186{,}7 \text{ kg/ha} \cdot \sqrt{\frac{1}{15} + \frac{1}{19}}$$

$$C\alpha \cong 1{,}63 \cdot 186{,}7 \text{ kg/ha} \sqrt{\frac{1}{15} + \frac{1}{19}}$$

$$C\alpha \cong 105{,}11 \text{ kg/ha}$$

$259{,}5$ kg/ha $> 105{,}11$ kg/ha $\Rightarrow$ rechazamos Ho.

* rechazo Ho si =

$T_{obs} > T_{crítica}$

$$\frac{(\bar{X}_i - \bar{X}_f) - 0}{S_a\sqrt{\frac{1}{n_i} + \frac{1}{n_f}}} = T_{obs}$$

$T_{crítica} = t_{32\,;\,0{,}95} \cong 1{,}63$

$T_{obs} = 4{,}024 > T_{crítica} = 1{,}63 \Rightarrow$ rechazo Ho

$$\frac{259{,}5 \text{ kg/ha}}{186{,}7 \text{ kg/ha} \cdot \sqrt{\frac{1}{15} + \frac{1}{19}}} = T_{obs}$$

$$4{,}024 = T_{obs}$$

* rechazo Ho si =

$P_{asociada} < \alpha = 0{,}05$

$$P\left(\bar{X}_i - \bar{X}_f > (\bar{X}_i - \bar{X}_f)'\right) = P\left(\frac{(\bar{X}_i - \bar{X}_f) - 0}{S_a\sqrt{\frac{1}{n_i} + \frac{1}{n_f}}} > \frac{(\bar{X}_i - \bar{X}_f)' - 0}{S_a \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_f}}}\right) =$$

$$P(t_{32} > 4{,}024) \cong 1 - P(t_{40} < 4{,}024) \cong 0{,}0003$$

$P < \alpha \Rightarrow$ rechazo Ho.

b) Estimación x' intervalo de confianza

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

$\alpha=0{,}05$

$\theta=\mu_1-\mu_2$ $\quad 1-\alpha=0{,}95$

$\hat{\theta}=\overline{X}_1-\overline{X}_2$

$$(\overline{x}_i-\overline{x}_p) \pm t_{(n_i+n_p-2);\,1-\frac{\alpha}{2}} \cdot \frac{S_a}{\sqrt{\frac{1}{n_i}+\frac{1}{n_p}}}$$

$$259{,}5\,kg/ha \pm t_{32;\,0{,}975} \cdot \frac{186{,}7\,kg/ha}{\sqrt{\frac{1}{15}+\frac{1}{19}}} =$$ → consultar

$$259{,}5\,kg/ha \pm 2{,}02 \cdot \frac{186{,}7\,kg/ha}{\sqrt{\frac{1}{15}+\frac{1}{19}}} =$$

$$259{,}5\,\frac{kg}{ha} \pm 2{,}02 \cdot 186{,}7\,kg/ha \cdot \sqrt{\frac{1}{15}+\frac{1}{19}}$$

Problema n° 8.

X = nivel promedio de IL-2 (U/L) ✓

Factor = donación de sangre

niveles = antes donación y después de la donación

u. experimental = paciente

m. experimental = en bloques completamente aleatorizados.

d U/L
0,07
−2,19
−1,39
2,32
−3,05
−2,62
0,76
0,29
−0,47
−4,5
−2,43
2,78

$d \rightarrow N(\mu_d;\, \sigma_d^2)$

$\bar{d}=-0{,}87\,(U/L)$; $S_d=2{,}18\,(U/L)$

Figura 7-1 = según el diagrama de puntos ... consultarlo!

b)- $P(E_1) = 0{,}05$

$H_0)\ u_d = 0$ $\qquad$ $H_1)\ u_d \neq 0$ → no especifica si disminuye o aumenta

→ ensayo bilateral

$$\bar{d} < 0 - t_{(n-1),\,1-\frac{\alpha}{2}} \cdot \frac{S_d}{\sqrt{n}} \quad ó \quad \bar{d} > 0 + t_{(n-1),\,1-\frac{\alpha}{2}} \cdot \frac{S_d}{\sqrt{n}}$$

$\bar{d} = -0{,}84\ (U/L)$

$$C\alpha = \pm\, t_{11\,;\,0{,}975} \cdot \frac{2{,}18\ U/L}{\sqrt{12}}$$

$$C\alpha = \pm\ 2{,}20 \cdot \frac{2{,}18\ U/L}{\sqrt{12}}$$

$$C\alpha = \pm\ 1{,}4$$

$\bar{d} = -0{,}84\ U/L > -1{,}4\ U/L \Longrightarrow$ no rechazo Ho

$\bar{d} = -0{,}84\ U/L < 1{,}4\ U/L \Longrightarrow$ no rechazo Ho.

rechazo Ho si =

* $T_{observada} < -t_{(n-1)\,;\,1-\frac{\alpha}{2}}$ ó $T_{observada} > T_{(n-1)\,;\,1-\frac{\alpha}{2}}$

$$T_{obs} = \frac{-0{,}84\ U/L \cdot \sqrt{12}}{2{,}18\ U/L} = -1{,}3820$$

$T_{crítica} = \pm 2{,}20$

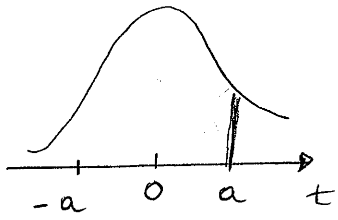
$T_{obs} = -1{,}3820 > -2{,}20 \Longrightarrow$ no rechazo Ho

$T_{obs} = -1{,}3820 < 2{,}20 \Longrightarrow$ no rechazo Ho rechazo Ho si =

$2 \, P_{asociada} < \alpha = 0,05$

$$2\left[P(\bar{d} < \bar{d}')\right] =$$

$$2.\left[P\left(\frac{\bar{d} - 0}{Sd/\sqrt{n}} < \frac{\bar{d}' - 0}{Sd/\sqrt{n}}\right)\right] =$$



$$2\left[P\left(T_{11} < \frac{-0,87\,UI/L}{2,18\,UI/L/\sqrt{12}}\right)\right] =$$

$$2.\left[P\left(T_{11} < -1,38\right)\right] =$$

$$2.\left[1 - P\left(T_{11} < 1,38\right)\right] =$$

$$2.\left[1 - 0,90\right] = 0,2$$

$0,2 > 0,05 \Rightarrow$ no rechazo Ho.

conclusión = con un nivel de significación 5% se concluye que el nivel promedio de IL-2 se modifica por el efecto donación de sangre.

*consultar x' los "supuestos"

Ejercicio nº 9

A = enfermar de VIH.

✓ Factor = vacunación
✓ Nivel = vacuna o placebo
✓ unidad exp = pacientes
✓ modelo exp completamente aleatorizado ⟹ muestras independientes

$$\begin{cases} p_1 = \text{probabilidad enfermar estando vacunado.} \\ p_2 = \text{probabilidad enfermar estando en placebo.} \end{cases}$$

$n_1 ; n_2$ muestras > 30 $\longrightarrow p = \hat{h} = h_1 - h_2$

$$h \longrightarrow N\left(p_1 - p_2 = 0 \; ; \; p \cdot (1-p) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

estimamos p con h; pp. de la muestra total.

$$h = \frac{F_1 + F_2}{n_1 + n_2} = \frac{n^{\circ} \text{ enfermos VIH}}{\text{total}} = \frac{24 + 21}{1503} = 0{,}03$$

$h_1 = 24/741 = 0{,}03239$ $\qquad h_2 = 21/762 = 0{,}02756$

~~$H_1)\ p_1 < p_2$~~

~~$H_0)\ p_1 = p_2$~~

$p_1 - p_2 = 0$

$$(h_1 - h_2) < -z_{1-\alpha} \sqrt{h(1-h) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$h' = (0{,}03239 - 0{,}02756) = 4{,}83 \times 10^{-3}$$

$$C\alpha = -z_{0{,}95} \cdot \sqrt{\frac{45}{1503} \cdot \left(1 - \frac{45}{1503}\right) \cdot \left(\frac{1}{741} + \frac{1}{762}\right)} =$$

$$C\alpha = -1{,}645 \cdot 8{,}792633 \times 10^{-3}$$

$$C\alpha = -0{,}0145$$

$h' = 5{,}63 \times 10^{-3} > -0{,}0145 \Longrightarrow$ no rechazo Ho

Rechazo Ho sí = Zobs < Zcrítica

$$Z_{obs} = \frac{(h_1 - h_2) - 0}{\sqrt{h \cdot (1-h) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$Z_{crítica} = Z_{0,95} = 1,645$$

$$Z_{obs} \quad \frac{4,83 \times 10^{-3}}{\sqrt{\frac{45}{1503} \cdot \left(1 - \frac{45}{1503}\right) \cdot \left(\frac{1}{741} + \frac{1}{762}\right)}}$$

$$Z_{obs} = 0,54$$

$$0,54 > -1,645 \Rightarrow \text{no rechazo Ho.}$$

Rechazo Ho sí =
Pasociada < $\alpha = 0,05$

$$P\left(h_1 - h_2 < (h_1 - h_2)'\right) =$$

$$P\left(\frac{(h_1 - h_2) - 0}{\sqrt{h \cdot (1-h) \cdot \left(\frac{1}{n_1} + \frac{1}{n_1}\right)}} < \frac{(h_1 - h_2)' - 0}{\sqrt{h \cdot (1-h) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right) =$$

$$P(Z < 0,54) = 0,11$$

$$1 \cong 0,11 > 0,05 \Rightarrow \text{no rechazo Ho.}$$

Conclusión = con un nivel de significación del 5% se concluye que la vacuna no es eficaz contra la prevención del VIH, ya que la probabilidad de enfermar estando vacunado o no, son estimadamente iguales.

Ejercicio 10

Lámpara Led \ Lámpara Hg	+	-	total
+		15	15
-	17		
total	17		32

$32 \longrightarrow n'$

A = detectar anticuerpos por tecnica de inmunof.

- Factor = lámpara
- nivel = Led ; Hg
- unidad experimental = muestra sangre
- modelo experimental = en bloques completamente aleatorizados $\Longrightarrow$ muestras dependientes.
- $n > 30 \Longrightarrow h \xrightarrow[\text{Ho cierta}]{} N\left(0{,}50 \,;\, \frac{0{,}50 \times 0{,}50}{n'}\right)$.

$$Ho)\ P_{Hg}(+) = 0{,}5$$

$$H_1)\ P_{Hg}(+) \neq 0{,}5 \qquad P(E1) = 0{,}05 = 5\%$$

ensayo bilateral.

$$\underbrace{h'}_{\frac{F(A,\bar{A})}{n'}} < 0{,}5 - Z_{1-\frac{\alpha}{2}} \cdot \frac{0{,}5}{\sqrt{n'}} \quad \text{ó sí} \quad h' > 0{,}5 + Z_{1-\frac{\alpha}{2}} \cdot \frac{0{,}5}{\sqrt{n'}}$$

$$h' = \frac{17}{32} = 0{,}53125$$

$$C\alpha = 0{,}5 \pm Z_{0{,}975} \cdot \frac{0{,}5}{\sqrt{32}} = 0{,}5 \pm 1{,}96 \cdot \frac{0{,}5}{\sqrt{32}}$$

$$= 0{,}327$$

$$= 0{,}673$$

$h' = 0{,}53125 > 0{,}327 = C\alpha \Longrightarrow$ no rechazo Ho

$h' = 0{,}53125 < 0{,}673 = C\alpha \Longrightarrow$ no rechazo Ho

* Rechazo Ho si =

$Z_{obs} < Z_{crítico}$ ó $Z_{obs} > Z_{crítico}$

$\frac{h' - 0{,}5}{0{,}5 \cdot \sqrt{32}} = Z_{ob}$ $\quad Z_{crítico} = -Z_{0{,}975}$

$0{,}011048 = Z_{obs}$ $\quad Z_{crítico} = -1{,}96 \longrightarrow Z_{obs} = 0{,}011048 > -1{,}96 = Z_{crit} \Longrightarrow$ no rechazo Ho

* Rechazamos Ho sí =

$2 P_{asociada} < \alpha = 0{,}05$

$2 P(h' < (h')') =$

$$2P\left(\frac{h - 0{,}5}{0{,}5/\sqrt{32}} < \frac{-h' + 0{,}5}{0{,}5/\sqrt{32}} \right) =$$

$$2P(Z < -0{,}35) = 0{,}3632 \cdot 2 = 0{,}7264$$

$P_{asociada}\ 0{,}7264 > 0{,}05 = \alpha \Longrightarrow$ no rechazo Ho

Conclusión = con nivel de significación del 5 % se concluye que la probabilidad de detectar anticuerpos por técnica de inmuno fluorecencia indirecta es la misma con lámparas Led y de Mercurio usadas en el microscopio.

Ejercicio 11.

a) X = concentración de hemoglobina (g/dL)

- Factor = tranfusión globulos rojos
- niveles = globulos rojos conservados menos de 8 días (I)
  globulos rojos " por más tiempo (II)
- unidad experimental = paciente
- modelo experimental completamente aleatorizado ⟹ muestras ind.

trabajo $P(E1) = 0{,}05$

* variancia poblacional desconocida.

$H_0)\ \sigma_1^2 = \sigma_2^2$

$H_1)\ \sigma_1^2 > \sigma_2^2$

siendo $S^2_{x1} = 1{,}64\ (g/dL)^2 > S^2_{x2} = 1{,}02\ (g/dL)^2$

rechazo Ho sí =

$$F_{obs} = \frac{S^2_{x1}}{S^2_{x2}} > F_{n_1-1;\, n_2-1;\, 1-\alpha}$$

$$F_{obs} = \frac{1{,}64\,(g/dL)^2}{1{,}02\,(g/dL)^2} = 1{,}61$$

$F_{obs} = 1{,}61 > F_{crítica} = 1 \Longrightarrow$ rechazo Ho

$$\sigma_1^2 \neq \sigma_2^2$$

$$F_{crítica} = F_{(1210);(1218);0{,}95}$$

$F_{crítica} = 1$.

$H_1)\ u_1 \neq u_2$

$u_1 - u_2 \neq 0 \longrightarrow$ ensayo bilateral.

$H_0)\ u_1 = u_2$

$u_1 - u_2 = 0$

Estadístico base $= \bar{X}_1 - \bar{X}_2$

rechazo Ho si =

$$(\bar{X}_1 - \bar{X}_2) < 0 - t_{f;\frac{1-\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \text{ ó si } (\bar{X}_1 - X_2) > t_{f;\frac{1-\alpha}{2}} \cdot \sqrt{\frac{Sn_1}{n_1} + \frac{Sn_2}{n_2}}$$

$$\bar{X}_1 - \bar{X}_2 = 7{,}69\,(g/dL) - 7{,}76\,(g/dL) = -0{,}07\,g/dL$$

$$C\alpha = \mp t_{f;\frac{1-\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$C\alpha = \pm t_{\infty;0{,}975} \cdot \sqrt{\frac{1{,}64\,g/dL^2}{1211} + \frac{1{,}02\,g/dL^2}{1219}}$$

$$C\alpha = \pm 1{,}96 \cdot \sqrt{2{,}191 \times 10^{-3}\,g/dL^2}$$

$$C\alpha = \pm 0{,}092\,g/dL^2$$

$\bar{X}_1 - \bar{X}_2 = -0{,}07\,g/dL > -0{,}092\,g/dL = C\alpha$

$\Longrightarrow$ no rechazo Ho

$\bar{X}_1 - \bar{X}_2 = -0{,}07\,g/dL < 0{,}092\,g/dL$

$\Longrightarrow$ no rechazo Ho

$$f = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{(n_1-1)} + \frac{(S_2^2/n_2)^2}{(n_2-1)}}$$

$$f = \frac{(1{,}64\,g/dL^2/1211) + (1{,}02\,g/dL^2/1219)^2}{\frac{(1{,}64\,g/dL^2/1211)^2}{(1211-1)} + \frac{(1{,}02\,g/dL^2/1219)^2}{(1219-1)}}$$

$$f = \frac{\left(2{,}191 \times 10^{-3}\,(g/dL)^2\right)^2}{1{,}52 \times 10^{-9}\,(g/dL)^4 + 5{,}75 \times 10^{-10}\,(g/dL^4)}$$

$$f = \frac{4{,}8 \times 10^{-6}\,\cancel{g/dL^4}}{2{,}095 \times 10^{-9}\,\cancel{g/dL^4}}$$

$$f = 2291 = \infty$$

rechazo Ho sí =

Tobs < Tcrítica ó Tobs > Tcrítica

$$\frac{(\bar{x}_1-\bar{x}_2)-0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} < -t_{f;\,1-\frac{\alpha}{2}}$$

$$Tobs = \frac{-0{,}07\ g/dL}{\sqrt{2{,}191\times10^{-3}\ g/dL^2}}$$

$Tobs = -1{,}496$

$-1{,}496 > -1{,}96 \Longrightarrow$ no rechazo Ho

$-1{,}496 < 1{,}69 \Longrightarrow$ no rechazo Ho

$Tcrítica = \pm t_{f,\,0{,}945}$

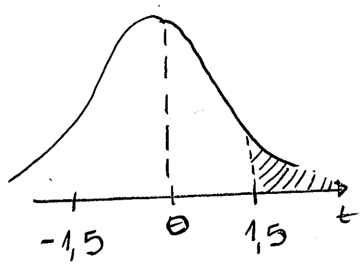
$Tcrítica = \pm 1{,}69$

rechazo Ho sí =

Pasociada $< \alpha$

$$p = 2P\left(\bar{x}_1-\bar{x}_2 < (\bar{x}_1-\bar{x}_2)'\right) =$$

$$2.P\left[\frac{(\bar{x}_1-\bar{x}_2)-0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} < \frac{(\bar{x}_1-\bar{x}_2)'-0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}\right]$$



$P(T_\infty < -1{,}496) = P(T_\infty > 1{,}5)$

$= 1 - P(T_\infty < 1{,}5)$

$$2.P\left(T_\infty < -1{,}496\right) =$$

$$2.\left[1-P(T_\infty < 1{,}5)\right] = 2.\left[1-0{,}95\right]$$

$$= 2.\left[0{,}25\right]$$

$$= 0{,}1$$

$Pas = 0{,}1 > \alpha \Longrightarrow$ no rechazo Ho

Conclusión con un nivel de significación del 5 % se concluye que el grupo I y el grupo II no difieren respecto a su nivel de hemoglobina basal.

b)

A = muerte a los 90 días.

$P(A)_I$ = probabilidad de muerte a los 90 días en pacientes del grupo I.

$P(A)_{II}$ = " " " " " " " en pacientes del grupo II.

Trabajamos $P(E_1) = 0{,}05 = 5\%$

H_0) $P(A)_I = P(A)_{II}$

$P(A)_I - P(A)_{II} = 0$

H_1) $P(A)_I < P(A)_{II}$ → ensayo unilateral x' izquierda

$P(A)_I - P(A)_{II} < 0$

Estadístico base = $h_1 - h_2$

$h_1 = \frac{430}{1211} = 0{,}36$

$h_2 = \frac{448}{1219} = 0{,}37$

$h = \frac{430+448}{1211+1219} =$

$h = 0{,}36$

Rechazo Ho si =

$$(h_1 - h_2) < -Z_{1-\alpha} \sqrt{h.(1-h).\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} =$$

$h_1 - h_2 = -0{,}0124.$

$$C\alpha = -Z_{0,95} \cdot \sqrt{0{,}36 \cdot (1-0{,}36) \cdot \left(\frac{1}{1211}+\frac{1}{1219}\right)} =$$

$$C\alpha = -1{,}645 \cdot \sqrt{3{,}793 \times 10^{-4}} = -0{,}03204$$

$-0{,}0124 > -0{,}03204 \Longrightarrow$ no rechazamos Ho

Rechazo ho si =

Zobs < Zcrítico

$$\frac{(h_1 - h_2) - 0}{\sqrt{h.(1-h)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = Zobs$$

Zcrítico = $-Z_{1-\alpha}$

Zcrítico = $-1{,}645$

$-0{,}637 > -1{,}645 \Longrightarrow$ no rechazo Ho

$$\frac{-0{,}0124}{\sqrt{0{,}36\,(1-0{,}36).\left(\frac{1}{1211}+\frac{1}{1219}\right)}} = Zobs$$

→ Zobs = $-0{,}637$

rechazo Ho sí =
$P_{asociada} < \alpha$

$$P_{asociada} = P\left[(h_1 - h_2) < (h_1 - h_2)'\right]$$

$$= P\left[\frac{(h_1-h_2)-0}{\sqrt{h(1-h)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} < \underbrace{\frac{(h_1-h_2)'-0}{\sqrt{h(1-h)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}}_{Z_{obs}}\right]$$

$$= P(Z < -0{,}634) = 0{,}2611$$

$P_{asoc.} = 0{,}2611 > 0{,}05 = \alpha \Longrightarrow$ no rechazo Ho.

Conclusión = con un nivel de significación del 5% se concluye que la probabilidad de muerte a los 90 días es igual para el g. I y el g II.

Ejercicio 15 (propuestos).

A = muerte intrahospitalaria.

P(A) = probabilidad de muerte intrahosp.

h_{Mg} = F. relativa del suceso muerte intrahospitalaria en pacientes tratados con Mg

h_p = " " " " " tratados con placebo.

* Factor = tratamiento
* niveles = Mg; placebo
* unidad exp = pacientes con IAM
* modelo experimental completamente aleatorizado

$H_1)$ $P(A)_{Mg} < P(A)_P$ $\quad P(E_1) = 0{,}05$.

$P(A)_{Mg} - P(A)_P < 0$ → ensayo unilateral x' izq.

$H_0)$ $P(A)_{Mg} = P(A)_P$

$P(A)_{Mg} - P(A)_P = 0$

$h = \frac{4+17}{96+98} =$ → Frecuencia relativa p/ suceso muerte intrahospitalaria

* estadística base = $h_{Mg} - h_p$

rechazamos Ho si =

$$h_{Mg} - h_p < -Z_{1-\alpha} \cdot \sqrt{h \cdot (1-h) \cdot \left(\frac{1}{n_{Mg}} + \frac{1}{n_P}\right)} =$$

$$\frac{4}{96} - \frac{17}{98} < -Z_{0,95}\sqrt{0{,}10825(1-0{,}10825) \cdot \left(\frac{1}{96}+\frac{1}{98}\right)} =$$

$C\alpha = -1{,}645 \cdot 0{,}045 = -0{,}0734$

$h_{Mg} - h_P = -0,132$

$h_{Mg} - h_P = -0,132 < -0,0734 = C_\alpha \Longrightarrow$ rechazo H_0 !!

*rechazo H_0 sí =

$Z_{obs} < Z_{crítico}$

$Z_{crítico} = -Z_{0,95}$

$= -1,645$

$$\frac{(h_{Mg} - h_P) - 0}{\sqrt{h(1-h) \cdot \left(\frac{1}{n_{Mg}} + \frac{1}{n_P}\right)}} = Z_{obs}$$

$-2,95 < -1,645 \Longrightarrow$ rechazo H_0.

$$\frac{-0,132}{0,045} = Z_{obs}$$

$-2,95 = Z_{obs}$

*rechazo H_0 si =

$$P_{asociada} = P\left[(h_{mg} - h_P) < (h_{mg} - h_P)'\right] =$$

$$P\left[\frac{(h_{mg} - h_P) - 0}{\sqrt{h \cdot (1-h) \cdot \left(\frac{1}{n_{Mg}} + \frac{1}{n_P}\right)}} < \underbrace{\frac{(h_{mg} - h_P) - 0}{\sqrt{h(1-h)\left(\frac{1}{n_{Mg}} + \frac{1}{n_P}\right)}}}_{Z_{obs}}\right]$$

$$P\left[Z < -2,95\right] = 0,0015$$

$0,0016 < 0,05 \Longrightarrow$ rechazo H_0.

Conclusión = con un nivel de significación 5% concluimos que la probabilidad de muerte intrahospitalaria es menor cuando los pacientes son tratados con Mg

Problema nº 16

$CV = \frac{S_x}{\bar{X}} = \frac{\sigma_x}{u_x}$

X = trazas o co. Hg en agua (ug/L).

$X \longrightarrow N(u_x; \sigma_x)$

Factor = técnica

Niveles = EAA (a) ; EM (b)

unidad exp = alícuota de agua

modelo exp = ~~en bloques comp. aleatorizados~~ comp. aleatorizado

Muestras ~~dependientes~~ independientes.

$CV_a = \frac{S_a}{0{,}051\ ug/L} = 0{,}015$

$S_a = 0{,}015 \cdot 0{,}051\ ug/L$

$S_a = 7{,}65 \times 10^{-4}\ ug/L$

$H_1)\ \sigma_b^2 > \sigma_a^2$ siendo $S_b^2 > S_a^2$

$CV_b = \frac{S_b}{0{,}066\ ug/L} = 0{,}031$

$H_0)\ \sigma_a^2 = \sigma_b^2$

$S_b = 0{,}031 \cdot 0{,}066\ ug/L$

$S_b = 2{,}04 \times 10^{-3}\ ug/L$

rechazo Ho si = $F_{obs} > F(n_a - 1; n_b - 1; 1-\alpha)$

$F_{obs} = \frac{4{,}3 \times 10^{-6}\ ug/L^2}{5{,}9 \times 10^{-7}\ ug/L^2}$

$* S_b^2 = 4{,}3 \times 10^{-6}\ ug/L^2$

$* S_a^2 = 5{,}9 \times 10^{-7}\ ug/L^2$

$F_{obs} = 7{,}28$

$F_{crítica} = F(8; 9; 995) = 3{,}23$

$F_{obs} = 7{,}28 > 3{,}23 = F_{crítica} \Longrightarrow$ rechazo Ho.

* Con un nivel de significación del 5 % se concluye que las dos técnicas probadas no miden con igual presición el contenido de mercurio en agua (ug/L).

Problema nº 25

H = hipoglucemia → P(H) = probabilidad hipoglucemia; → $P(\bar{H})$ = probabilidad de no tener hipoglucemia

$P(H; \bar{H}) = P_2$

$P(\bar{H}; H) = P_1$

$H_0)\ \cancel{P(H)_{art} = P(H)_{control}}$

$P_1 = P_2$

$H_1)\ P(H)_{art} < P(H)_{cont}$

$H_0)\ \underbrace{P(H)_{art} = 0{,}5}$

$P_1 = 0{,}5$

$H_1)\ \cancel{P(H)_{art} < 0{,}5}$

$P_1 < 0{,}5$.

	artificial: Hipoglucemia H	artificial: no hipoglucemia $\bar{H}$
control: Hipoglucemia H	X	$H; \bar{H}$
control: $\bar{H}$ no hipoglucemia	$\boxed{\bar{H}; H}$	~~X~~

n' = 31

→ ~~mei nt.~~ Probabilidad hipoglucemia art ↓

$f(\bar{H}; H) = f(H; \bar{H}) \Longrightarrow P(\bar{H}$ en control y si H con páncreas artificial$) = 0{,}5$

sucesos complementarios

$$h' < 0{,}5 - Z_{1-\alpha} \cdot \frac{0{,}5}{\sqrt{n'}}$$

$$h' = \frac{9}{31} = 0{,}3003$$

$h' = 0{,}3003 < 0{,}355 \Longrightarrow$ rechazo H_0

$$C\alpha = 0{,}5 - Z_{0{,}95} \cdot \frac{0{,}5}{\sqrt{32}}$$

$$C\alpha = 0{,}5 - 1{,}645 \cdot \frac{0{,}5}{\sqrt{32}}$$

$$C\alpha = 0{,}355$$

Conclusión = con un nivel de significación del 5% concluimos que la probabilidad de tener hipoglucemia reduce con el pancreas artificial.

Únicamente para el d.b.c.a ∞ podemos evaluar si el efecto bloque produce variación, es decir, si existe diferencia significativa en ellos.

$H_0)\ B_1 = B_2 = B_3 = \dots = B_J = 0$ $\qquad$ $H_1)$ al menos un $B_j \neq 0\ ;\ j = 1, \dots, J.$

rechazo H_0 si =

$$F_{obs} = \frac{C.M\ bloque}{C.M.\ error} \geq F_{J-1\ ;\ (I-1).(J-1)\ ;\ 1-\alpha}$$

Diseño experimental completamente aleatorizado

Fuentes de variación	Suma de cuadrados	grados de libertad	Cuadrados medios	Fobs.	
Tratamiento	$\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{I\circ} - \bar{y}_{\circ\circ})^2$	$(I-1)$	$S.C.Trat/(I-1)$	$\frac{C.M.trat}{C.M.error} = F_{obs}^{*}$	
Error	$\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \bar{y}_{I\circ})^2$	$(I.J-I)$	$S.C.error/(I.J-I)$		
Total	$\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \bar{y}_{\circ\circ})^2$	$(I.J-1)$			

¡No olvidar estudiar demostración supuestos!

$P(F > F_{obs}) < \alpha$

Diseño experimental en bloques completos aleatorizados ∇

Fuentes de variación	Suma de cuadrados	grados de libertad	cuadrados medios	Fobs	P.
Tratamiento	$\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{I\circ} - \bar{y}_{\circ\circ})^2$	$(I-1)$	$S.C.trat/I-1$	$\frac{C.M.\ tratamiento}{C.M.\ error}$	
bloque	$\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{\circ j} - \bar{y}_{\circ\circ})^2$	$(J-1)$	$S.C\ bloque/J-1$	$\frac{C.M\ bloque}{C.M.error}$	
error	$\sum_{i=1}^{I}\sum_{j=1}^{J}(e_{ij})^2$	$(I-1)(J-1)$	$S.C.error/(I-1).(J-1)$		
Total	$\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \bar{y}_{\circ\circ})^2$	$I.J-1$			

– Si Ho es cierta y no hay efectos del factor en estudio, ambos cuadrados medios tienen igual esperanza

$$E\left(C.M._{error}\right) = \sigma^2 = E\left(C.M.\ tratamiento\right) = \sigma^2 + \underbrace{\frac{J}{I-1} \cdot \sum_{i=1}^{I} \alpha_i}_{=0}$$

– Si [redacted] entonces, $\underline{E\left(C.M\ tratamiento\right)} > E\left(C.M.\ error\right)$.

- $$E\left(C.M.\ tratamiento\right) = \sigma^2 + \underbrace{\frac{J}{I-1} \cdot \sum_{i=1}^{I} \alpha_i}_{\neq 0}$$

– Para realizar prueba de hipótesis se comparan ambos $\underbrace{\text{C. Medios}}$ mediante un cociente, por su distribución, este cociente resulta una Fobs=

recordamos se distribuyen como $\dfrac{\chi^2_K \cdot \sigma^2}{K}$

$$F_{obs} = \frac{C.M.\ trat}{C.M.\ error} = \frac{\dfrac{\chi^2_{I-1} \cdot \sigma^2}{I-1}}{\dfrac{\chi^2_{I.J-1} \cdot \sigma^2}{I.J-1}}$$

entonces $\underline{\text{rechazo Ho sí}}$ =

$$\boxed{F_{obs.} = \frac{C.M.\ trat}{C.M.\ error} > \frac{F_{I-1;\, \overbrace{I.J-I}^{*};\, 1-\alpha}}{F_{crítica.}}}$$ d.c.a

$$F_{obs} = \frac{C.M.\ trat}{C.M.\ error} = \frac{\dfrac{\chi^2_{I-1} \cdot \sigma_x^2}{(I-1)}}{\dfrac{\chi^2_{(I-1).(J-1)} \cdot \sigma_x^2}{(I-1)(J-1)}} \longrightarrow F_{(I-1);\,(I-1).(J-1)}.$$

Atenti que los grados del error son distintos

rechazo Ho sí –

$$\boxed{F_{obs} > F_{(I-1);\, \underbrace{(I-1).(J-1)}_{*};\, 1-\alpha}}$$ d.b.c.a.

d o A.

variabilidad debida al Factor en estudio. = se analiza si las observaciones varían debido a qué fueron sometidas a diferentes niveles del factor en estudio.

$$- S.C.\,trat = \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$$

- g. lib = $I - 1$

- cuánto mayor es el efecto del factor, la diferencia entre la media debido al tratamiento y la media general aumenta ⟹ aumenta S.C. tratamiento = hay mayor variabilidad debido al factor

⇝ estás sumas de cuadrado tienen distribución χ^2_K

• **variabilidad debido a los errores aleatorios.**

es la variabilidad inherente a los procesos biológicos o naturales estudiados, es la que produce que unidades que perciben el mismo tratamiento brinden respuestas diferentes.

$$S.C.\,error = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_{i\bullet})^2$$

- g. Libertad = $I . J - I$.

---

si dividimos estás sumas de cuadrados por sus grados de libertad obtenemos "C. Medios" ⟶ estimadores de la varianza poblacional.

$$\text{- C. Medio del tratamiento} = \frac{S.C.\,Tratamiento}{g.\,libertad} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2}{I - 1}$$

si no rechazo Ho $\alpha_i = 0$. $E(C.M.t) = \sigma^2$

$$E\left(\begin{matrix}\text{C. Medio}\\ \text{tratamiento}\end{matrix}\right) = \sigma^2 + \frac{J}{I-1} \cdot \sum_{i=1}^{I} \alpha_i^2$$

* en este diseño experimental los bloques constituyen una nueva fuente de variación

S.C.Total = S.C. tratamiento + S.C. error + $\underbrace{\text{S.C. bloque}}$

nuevo factor que incrementa variación

* S.C. Totales = $\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij}-\bar{y}_{\circ\circ})^2$ g. Lib. = $I.J-1$

* S.C. bloque = $\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{\circ j}-\bar{y}_{\circ\circ})^2$ g. Lib. = $J-1$

* S.C. tratamiento = $\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{i\circ}-\bar{y}_{\circ\circ})^2$ g. Lib. = $I-1$

* S.C. error = $\sum_{i=1}^{I}\sum_{j=1}^{J}(e_{ij})^2$ g. Lib. = $(I-1).(J-1)$

3) * supuestos del modelo

$$\boxed{e_{ij} \longrightarrow NI(0; \sigma^2)}$$

* tienen distribución normal
* tienen esperanza 0 y varianza cte.
* son independientes.

cómo consecuencia =

$$y_{ij} \longrightarrow N(u_i; \sigma^2).$$

4) Planteó de hipótesis

¿Qué deseamos estudiar? si existen diferencias estadísticamente significativas entre los valores medio de la variable estudiada en cada uno de los niveles del factor.

$\boxed{H_0)\ u_1 = u_2 = \dots = u_I}$ (el promedio de Y es el mismo para cualquier nivel).

$H_1)$ al menos un u_i difiere del resto. $i = 1, 2, \dots, I$.

5) Fundamentos técnica Anova -

- Para determinar si al menos uno de los promedios difieren del resto, se analiza la variabilidad existente de la variable "y", identificando posibles causas de variación de la misma. ⟶ Fuentes de variación

1) variabilidad total de las observaciones = suma de cuadrados totales.

$$\underline{S.C.T} = \sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij}-\bar{y}_{\circ\circ})^2$$

si dividimos x' I.J (nºobs) menos uno.

$$S.C.T \longrightarrow \chi^2_{I.J-1}.\sigma^2$$

La variabilidad total de las observaciones proviene de dos fuentes = ⟶ para diseño completamente aleatorizado.

1) suma cuadrados debido al tratamiento.

2) suma cuadrados debido al error aleatorio.

diseño en bloques completos aleatorizados

*Tenemos un factor con I niveles. Se forman "J" bloques cada uno de ellos I unidades homogéneas. A cada unidad del bloque se le asigna al azar un nivel del factor.

"Y_{ij}"

Niveles del factor

	Bloques 1	2	...	...	J	
1	y_{11}	y_{12}	...	...	y_{1J}	$\rightarrow \bar{y}_{I\circ}$ } media debido al trat.
2	y_{21}	y_{2}				
...	...					
...	...					
I	y_{I1}					

↓ Medias en cada bloque $\bar{y}_{\circ J}$

- Media debido al tratamiento $= \frac{1}{J} \cdot \sum_{j=1}^{J} x_{ij}$ → ojo que la variable es "y".

$\bar{y}_{i\circ}$

- Media debido al bloque $= \frac{1}{I} \cdot \sum_{i=1}^{I} x_{ij}$

$\bar{y}_{\circ J}$

- Media general $= \frac{1}{I.J} \cdot \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}$

$\bar{y}_{\circ\circ}$

2) Ecuación de las observaciones

$$\boxed{y_{ij} = \alpha_i + B_j + e_{ij} + u_i}$$

"y_{ij}" = j-ésima observación para el i-ésimo nivel del factor (correspondiente al j-esimo bloque).

α_i = efecto de i-esimo nivel del factor.

B_j = efecto del j-ésimo bloque.

e_{ij} = error aleatorio debido a la observación del j-ésimo bloque y el i-ésimo nivel del factor.

u_i = valor medio de la variable "y" bajo estudio.

diseño completamente aleatorizados.

* Tenemos un factor con I niveles y realizamos "J" observaciones en cada nivel

balanceado.

- simbolizamos la variable de interés con la letra "y"
- c/u observaciones "Y_{ij}"

1) esquema de las observaciones:

varía "j"

$\bar{Y}_{I\circ}$ media correspondiente al I nivel

Niveles Factor

1	y_{11}	y_{12}	...	y_{1j}	...	y_{1J}	$\rightarrow \bar{y}_{1\circ}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2J}	$\rightarrow \bar{y}_{2\circ}$
...	$y_{...1}$	$y_{...2}$	...	$y_{...j}$	...	$y_{...J}$	$\rightarrow \bar{y}_{...\circ}$
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{iJ}	$\rightarrow \bar{y}_{i\circ}$
...	"	"	"	"	"	"	
I	y_{I1}	y_{I2}	...	y_{Ij}	...	y_{IJ}	$\rightarrow \bar{y}_{I\circ}$

$\bar{Y}_{i\circ}$ Media del nivel $= \frac{1}{\cancel{I}} \cdot \sum_{j=1}^{J} y_{ij}$

$\frac{1}{J}$ } sumas y divides x el nº de obs.

Media general $= \underbrace{\frac{1}{I.J}}_{\text{nº total de obs.}} \sum_{j=1}^{J} \sum_{i=1}^{I} y_j$

$\bar{y}_{\circ\circ}$

2) Ecuación observaciones

$$\boxed{y_{ij} = \alpha_i + u + e_{ij}}$$

"y_{ij}" = j-ésima observación debido al i-ésimo nivel del factor en estudio".

u = promedio o valor medio de la variable "y".

e_{ij} = error aleatorio debido a la j-ésima observación del i-ésimo nivel del factor en estudio.

α_i = efecto del i-ésimo nivel del factor.

Anova = análisis de la variancia. unidad n° 8. Ejemplo = pág 116 - ejercicio 4 -

↳ inferencia basada en más de dos muestras.

definición = conjunto de técnicas que nos permite evaluar el efecto del cambio de los niveles del factor o factores en estudio sobre el nivel medio de una variable.

- Recordamos =

1) Factor = condición experimental cuyo efecto sobre una variable de interés es estudiado.

2) niveles = cada una de clases y/o categorías del factor en estudio.

3) tratamiento = combinación de cada uno de los factores con sus respectivos niveles. (todas las combinaciones posibles).

4) unidad experimental = mínima porción de material a la que se le aplica un tratamiento.

Existen diferentes tipos de diseños experimentales para evaluar efectos de los FACTORES a través de anova.

- Tipos de diseños experimentales =
  - cantidad de factores =
    - unifactorial
    - bifactorial
    - multifactorial
  - naturaleza de los factores.
    - o efectos fijos = niveles considerados son los únicos que se desean estudiar.
    - o efectos aleatorios = niveles estudiados son elegidos al azar entre una población de posibles niveles.
  - según naturaleza de unidades experimentales.
    - unidades heterogéneas
      - * no todas las unidades son similares entre sí, esto genera que se comporten de forma distinta aunque se les diera el mismo tratamiento.
      - * para este tipo de unidades acudimos a un diseño en bloques completos aleatorizados

        ↓ dando lugar a =

        Muestras dependientes.
    - unidades homogéneas = * si se les aplicara a todas las unidades el mismo tratamiento se obtendrían resultados similares que solo varían aleatoriamente.
      - * para este tipo de unidades elegimos un diseño completamente aleatorizado

        ↓ dando lugar a =

        Muestras independientes
      - Caso particular = sin diseño previo (los niveles del factor separan naturalmente las unidades. Ejemplo = región a donde pertenece).

durante este curso aplicamos caso unifactorial a efectos fijos.

Umidad nº 8. Amova.

- Conjunto de técnicas para amalizar el efecto que produce el cambio de miveles de umo o más factores sobre el valor medio de la variable.

↓ mosotros mos emfocamos em 1 solo factor pero más de um mivel.

Ejemplo =
Factor "amtibiotico"
mivel del factor = $\underbrace{\text{"a", "b"}}_{\text{si fueram sólo dos estaríamos em base a dos muestras.}}$, "c"

Tipo de diseños experimentales
- según camtidad factores =
  - • umifactorial
  - • bifactorial
  - • trifactorial = multifactorial
- según maturaleza de factores
  - → diseño a efectos fijos → los miveles comsiderados som los úmicos que posee el factor o úmicos que deseamos estudiar
  - → diseño efectos aleatorizados. → los miveles estudiados fuerom elegidos al azar emtre población de posibles miveles
- según maturaleza de umidades experimentales =
  - homogémeas
    - • si se aplicara a todas las umidades experimentales el mismo tratamiento c/u, obtendríamos resultados similares que sólo varíam aleatoriamente.
      - → caso particular = sim diseño previo, umidades separadas maturalmente. Ejemplo = raza.
      - – Modelo completamente aleatorizado ⟹ muestras independientes.
  - heterogémeas
    - • si se aplicaría el mismo tratamiento obtendríamos ≠ resultados por mo ser similares emtre sí. → empleamos diseño em bloques completos aleatorizados.
      - se formam bloques homogemeos (umidades similares emtre sí y demtro de cada bloque se asigma um tratamiento a cada umidad)
      - – caso particular = el bloque esta formado x' uma sola umidad que recibe todos los tratamientos.

1) Esquematización observaciones. P' ~~modelo completamente aleatorizado.~~

- Factor ⟶ "I" niveles ; "J" observaciones
- La variable de interés lleva la letra "y". (c/u observaciones "y_{ij}").

nivel del Factor	1	y_{11}	y_{12}	$y_{1\circ\circ}$	y_{1j}	y_{1J}	⟶ $\bar{y}_{1\circ}$ ⟶ variando en "J"
	2	y_{21}	y_{22}	$y_{2\circ\circ}$	y_{2j}	y_{2J}	⟶ $\bar{y}_{2\circ}$
	∘∘∘	∘∘∘	∘∘∘	∘∘∘	∘∘∘	∘∘∘	
	i	y_{i1}	y_{i2}	∘∘∘	y_{ij}	y_{iJ}	⟶ $\bar{y}_{i\circ}$
	∘∘∘	∘∘∘	∘∘∘	∘∘∘	∘∘∘	∘∘∘	
	∘∘∘	∘∘∘	∘∘∘	∘∘∘	∘∘∘	∘∘∘	
	∘∘∘	∘∘∘	∘∘∘	∘∘∘	∘∘∘	∘∘∘	
	I	y_{I1}	y_{I2}	$y_{I\circ\circ}$	y_{Ij}	∘∘ y_{IJ}	$\bar{y}_{I\circ}$

Media cada nivel $\bar{y}_{I\circ}$ $= \frac{1}{J} \circ \underbrace{\sum_{j=1}^{J} y_{ij}}_{\text{varia en "J"}}$

Media general $\bar{y}_{\circ\circ}$ $= \frac{1}{IJ} \sum_{J=1}^{J} \sum_{i=1}^{I} y_{Ij}$

IJ ⟶ nº total observaciones!

Ecuación modelo "diseño completamente aleatorizado" ⟶ $y_{ij} = \mu + \alpha_i + e_{ij}$

- y_{ij} ⟶ jésima observación bajo el nivel "I" factor.
- μ ⟶ Promedio general variable y
- α_i ⟶ efecto de iesimo nivel Factor
- e_{ij} ⟶ error aleatorio de la j-ésima observación bajo el i-ésimo nivel del factor.

$\boxed{\text{supuestos del modelo.}}$ = este modelo considera a los errores aleatorios ε_{ij} como variables aleatorias.

$e_{ij} \longrightarrow N(0; \sigma^2)$

- tienen distribución normal.
- $E(e_{ij}) = 0$ y variancia cte $= \sigma^2$
- son independientes.

Planteó de Hipotesis.

$H_0)$ $u_1 = u_2 = \circ\circ\circ = u_I$

(el promedio y: igual para cualquier nivel).

$H_1)$ al menos un u_i difiere del resto

$i = 1, \circ\circ\circ, I$

(no todos los niveles tienen el mismo promedio).

Fundamento de la técnica = para determinar si al menos 1 promedio difieren, se analiza la variabilidad existente en la variable "y".

1) <u>suma de cuadrados totales</u> = variabilidad total presentes en las observaciones.

$$\underline{SC\ total} = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\underbrace{y_{ij}}_{\text{c/una de las observaciones}} - \underbrace{\bar{y}_{\circ\circ}}_{\text{media general}}\right)^2$$

• Si a la suma de cuadrados totales se lo divide por $(I.J-1)$ siendo I.J el nº total de observaciones obtenemos la variancia de todos los datos.

$$\text{variancia total} = \frac{SC\ totales}{(I.J-1)}$$

proviene → Factor en estudio a)

" → errores aleatorios b)

a) <u>variabilidad debida al factor en estudio</u> = las observaciones pueden variar por haber sido sometidas a diferentes niveles del factor

$$\underline{SC\ tratam} = \sum_{j=1}^{J}\sum_{i=1}^{I}\left(\bar{y}_{\circ i} - \bar{y}_{\circ\circ}\right)^2 = J.\sum_{i=1}^{I}\left(\bar{y}_{\circ i} - \bar{y}_{\circ\circ}\right)$$

↑ efecto factor = $\boxed{\bar{y}_{\circ i} \neq}$ medias de c/u niveles diferentes entre sí.

- Hay diferencia respecto al promedio general; la suma es mayor, hay más variabilidad debida al factor

b) <u>variabilidad debida a los errores aleatorios</u> = es la variabilidad inherente a procesos biológicos o naturales estudiados y es lo que produce que las unidades que recibieron el mismo tratamiento brinden respuestas diferentes.

$$\underline{S.C.\ error} = \sum_{j=1}^{J}\sum_{i=1}^{I}\left(y_{ij} - \bar{y}_{i\circ}\right)^2$$

$\underbrace{\text{S.C.totales}}$ = S.C. tratamiento + S.C. error.

sí ↓

dividimos c/u de ellas x' $(I.J-1)$ grados de libertad obtenemos <u>cuadrados medios</u>. ⟶ estimador varianza poblacional.

$* \; SC \longrightarrow \underbrace{\chi^2_{(I.J-1)} \cdot \sigma^2}$ distribución como una chi-cuadrado.

- $C.M.error = \dfrac{S.C.error}{(I.J-1)}$ $\quad E(C.M\,error) = \sigma^2 \quad$ $C.M\,error \longrightarrow \dfrac{\chi^2_{(I.J-1)} \cdot \sigma^2}{(I.J-1)}$

- $C.M.\,tratamiento = \dfrac{S.C.\,tratamiento}{(I.J-1)}$ $\quad E(C.M\,trat) = \sigma^2 + \dfrac{J}{I-1} \sum_{i=1}^{I} \alpha_i^2$

si Ho es cierta = (no hay efecto del factor).

$\alpha_i = 0 \;\forall i \Longrightarrow E(CM_{trat}) = \sigma^2 \quad C.M\,trat \longrightarrow \dfrac{\chi^2_{(I-1)} \cdot \sigma^2}{I-1}.$

p' probar Ho es falsa y hay efectos del factor

<u>$E(CM\,trat) > E(C.M\,error)$</u>

↘ se analiza x' un cociente

$$F_{obs} = \frac{CM\,trat}{C.M.\,error} \longrightarrow F_{I-1;\,I.J-1}.$$

si Ho falsa el valor del cociente es alto

rechazo Ho si =

$$F_{obs} > F_{(I-1);(I.J-1);(1-\alpha)}.$$

- Todo el análisis lo resumimos en el cuadro anova.

Fuente variación	S. Cuadrados	g. L	Cuadrado M.	Fobs	P. asociada
tratamiento	$J.\sum_{i=1}^{I}(\bar{Y}_{\circ i}-\bar{Y}_{\circ\circ})^2$	I-1	$\frac{\text{S. Cuadrados}}{\text{g. L}}$	$\text{Fobs} = \frac{\text{C.M. trat}}{\text{C.M. error}}$	P(Fobs > Fcrítica)
Error	$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{\circ i})^2$	I.J-I	$\frac{\text{S.C. error}}{\text{g. L}}$		
Total	$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{\circ\circ})$	I.J-1			

- "Comparación múltiples" ⟶ si rechazamos Ho, nos genera dudas

¿Cuál o cuáles tratamientos ≠ resto?

- analizamos tabla donde figuran los tratamientos ordenados según sus medias. (de menor a mayor).

- la tabla tiene en su última columna asteriscos o puntos. Los asteriscos alineados indican que los tratamientos no difieren significativamente en sus promedios.

- si dos tratamientos no tienen "*" o "." alineados sus promedios difieren significativamente.

Corroborar supuestos en los "residuos"
- → varianza constante ⇝ gráfico de residuos vs. nivel del factor
- → Independencia = se garantiza en el planeamiento y diseño del experimento.
- → distribución normal = gráfico de prob. normal.

$e_{ij} = \hat{e}_{ij} = Y_{ij} - \hat{\alpha}_i - \hat{\mu}$

$\hat{\mu} = \bar{Y}_{\circ\circ}$

$\hat{\alpha}_i = \bar{Y}_{i\circ} - \bar{Y}_{\circ\circ}$

$e_{ij} = Y_{ij} - (\bar{Y}_{i\circ} - \cancel{\bar{Y}_{\circ\circ}}) - \cancel{\bar{Y}_{\circ\circ}}$

$\boxed{e_{ij} = Y_{ij} - \bar{Y}_{i\circ}}$

2) Esquematización observaciones p' modelo diseño bloques completamente aleatorizados.

M. Factor	bloques 1	2	...	3	...	j	J
1	Y_{11}	Y_{12}	...	Y_{13}	...	Y_{1j}	Y_{1J}
2	Y_{21}	Y_{22}	...	Y_{23}	...	Y_{2j}	Y_{2J}
3	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
I	Y_{I1}	Y_{I2}	...	Y_{I3}	...	Y_{Ij}	Y_{IJ}

media c/ nivel

$$\bar{Y}_{i\circ} = \frac{1}{J} \cdot \sum_{j=1}^{J} Y_{ij}$$

media c/ bloque.

$$\bar{Y}_{\circ j} = \frac{1}{I} \sum_{j=1}^{J} Y_{ij}$$

$$\left.\bar{Y}_{\circ\circ} = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij} \right\} \text{media general.}$$

Ecuación modelo = $Y_{ij} + \alpha_i + \beta_j + \varepsilon_{ij}$

Y_{ij}: jesima observación bajo el iésimo factor.

α_i: efecto del I-ésimo nivel del factor

β_j: efecto del J-ésimo bloque

ε_{ij}: error aleatorio de la observación del i-ésimo nivel del factor en el jésimo bloque.

supuestos

$$\varepsilon_{ij} \longrightarrow N(0; \sigma^2)$$

σ^2: constante.

Fuente variación:
- Tratamiento
- bloque
- error aleatorio.

S.C.Total = S.C. Trat + S.C. bloque + S.C. error

$$S.C. total = \sum_{i=1}^{I} \sum_{j=1}^{J} \left(y_{ij} - \bar{y}_{\cdot\cdot} \right)^2$$

$$S.C. Tratamiento = J \cdot \sum_{i=1}^{I} \left(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot} \right)^2$$

$$S.C. bloque = I \sum_{j=1}^{J} \left(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot} \right)^2$$

$$S.C. error = \sum_{i=1}^{I} \sum_{j=1}^{J} e_{ij} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot} \right)^2$$

- Cuadro anova =

F. variación	S. Cuadrados	g. Libertad	C. Medios	F	Prob. aso.
Tratamiento	$J \cdot \sum_{i=1}^{I} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$I - 1$	S.C trat / g.L	$F_{obs} = \frac{S.C. trat}{S.C. Error}$	$P(F_{obs} > F_{critica})$
bloques	$I \sum_{j=1}^{J} (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$	$J - 1$	S.C. bloques / g.L		
Error	S.C. error		S.C error / g.L		
Total	S.C. total	$I \cdot J - 1$			

$(I-1) \cdot (J-1)$

$$F_{obs} = \frac{S.C. tratamiento}{S.C. Error} \longrightarrow F_{(I-1);(J-1)\cdot(I-1)}$$

Regla decisión =

rechazo Ho si =

$$F_{obs} > F_{crítica} = F_{(I-1);(I-1)\cdot(J-1);1-\alpha}$$

c) 1.- Independencia de las observaciones = Fue considerado en el planteó y diseño experimental.

2.- supuesto de normalidad.

$$\varepsilon_{ij} \longrightarrow N(0, \sigma^2)$$

3.- Variancia constante de los errores aleatorios.

d) Ho.- $u_a = u_b = u_c = u_d$ H1) al menos un u_i, siendo $i = a, b, c$ ó d sea distinto.

e) S.C. totales = $\sum_{j=1}^{J} \sum_{i=a}^{I} (y_{ij} - \bar{y}_{\circ\circ})^2$

$\longrightarrow$ S.C. Tratamiento = $J. \sum_{i=a}^{I} (y_{i\circ} - \bar{y}_{\circ\circ})^2$

$\longrightarrow$ S.C. errores = $\sum_{i=a}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_{i\circ})^2$

F) est. básica = $\dfrac{\text{C M tratamiento}}{\text{C M error}}$ = Fobs $\xrightarrow[\text{Ho cierta}]{}$ $F(I-1); (I.J-1)$

h) Zona de rechazo Ho = "unilateral a la derecha"

i) Probabilidad asociada se define como la probabilidad de observar un valor de F mayor o igual que el observado cuando Ho cierta $\longleftarrow$ siempre trabajas con esto.

ojo q' en modelo d.b.c.a g.L error = $I.J - I$ ~~I.J-1~~

j) completamos cuadro anova

F. variación	S. Cuadrados	g. L	C. Medios	F observada	P. asociada
-Tratamiento ~~Zona~~	0,002127 mg/m³	4 - 1 = 3	$7,09 \times 10^{-4}$ mg/m³	50,19	$P(F > 50,19)$ $\longrightarrow$ consultar
-error	0,00022.6 mg/m³	20-4=16 ~~20-1~~ ~~19~~	$1,4125 \times 10^{-5}$ mg/m³		
-Total	0,002353 mg/m³	19			

Ejercicio n° 2

y = cc de gases carbonicos (mg/m³).

Factores = zonas

nivel factor = tratamiento = zona a, zona b, zona c, zona d.

unidad exp = alicuota aire

" exp = homogeneas ⟹ muestras independientes

- modelo exp = completamente aleatorizado.
  - efectos fijos
  - unifactorial

a).- $y_{ij} = u + \alpha_i + \varepsilon_{ij}$

u = promedio general de cc de gases carbónicos (mg/m³)

α_i = el efecto de la i-ésima zona $I = a, b, c, d$.

ε_{ij} = error aleatorio de la j-ésima alícuota de aire medida en la i-esima zona

$I = a, b, c, d$ $J = 1, 2, 3, 4, 5$

b) $y_{ij} = u + \alpha_i + \varepsilon_{ij}$

$-\varepsilon_{ij} = u + \alpha_i + y_{ij}$

$\varepsilon_{ij} = y_{ij} - u - \alpha_i \longrightarrow \hat{\varepsilon}_{ij} = e_{ij} = y_{ij} - \hat{u} - \hat{\alpha}_i$

$\hat{u} = \bar{Y}_{\cdot\cdot}$

$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$

$e_{ij} = y_{ij} - \cancel{\bar{Y}_{\cdot\cdot}} - (\bar{y}_{i} - \cancel{\bar{Y}_{\cdot\cdot}})$

$\boxed{e_{ij} = y_{iJ} - \bar{y}_{i\cdot}}$ deducción formula general.

$e_{b3} = y_{b3} - \bar{y}_{3\cdot}$

$e_{b3} = 0{,}031\ mg/m^3 - 0{,}0328\ mg/m^3$

$e_{b3} = -1{,}8 \times 10^{-3}\ mg/m^3$

c) 1) supuesto independencia de las observaciones

2) supuesto distribución normal errores aleatorios

3) supuesto variancia constante.

* en base a los gráficos concluimos que los residuos se distribuyen normalmente ya que tienen un comportamiento lineal en el gráfico de probabilidad normal. En el gráfico de residuos vs. nivel factor observamos que los datos se distribuyen uniformemente por lo que concluimos que la variancia es constante.

d) Ho) $u_1 = u_2 = u_3$

H1) al menos un u_i $(I = 1,2,3)$ es $\neq$ del resto.

$$S.C\ Totales = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_{\cdot\cdot})^2$$

$$S.C.\ Tratamiento = J \sum_{i=1}^{I} (y_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

$$S.C.\ bloques = I \sum_{j=1}^{J} (y_{\cdot j} - \bar{y}_{\cdot\cdot})^2$$

F) Estadística básica

$$\frac{C\,M\ Tratamiento}{C.M.\ Error} \xrightarrow[\text{Ho cierta}]{} F_{(I-1);\ (I-1).(J-1)}$$

g)

F.V	S.C.	g.l	C.M	F	P
solución	926,77	2	463,385	44,89	>0,005
día	1118,94	5	223,78	21,67	>0,005
error	103,23	10	10,323		
Total	2148,94	17			

S.C.T = 2148,94 = 926,77 + 1118,94 + S.C error

S.C Error = 103,23.

K) Rechazo Ho si

$F_{obs} > F_{critica}$ Trabajamos $P(E_1) = 0{,}05$.

$F_{obs} = 50{,}19$

$F_{critica} = F(3;16;0{,}95) \cong 3{,}29$

$50{,}19 > 3{,}29 \Rightarrow$ rechazo Ho.

Conclusión = con un nivel de significación del 5% se concluye que al menos un valor medio de la cc de gases carbónicos (mg/m³) difiere del resto evaluado en las diferentes zonas.

l) Test comparaciones multiples

nivel	n_i	Media			
c	5	0,0194	*		
b	5	0,0328		*	
d	5	0,0422			*
a	5	0,0462			*

↓ balanceado.

concluimos = d y a presentan un mayor promedio de cc gases carbonicos (mg/m³) mas grande que b y c, están más contaminadas esas zonas.
Los promedios de c y b difieren significativamente del resto.

Ejercicio 4. =

a) Y = nº colonias bacterianas en contenedores de leche.
Factor - solución lavado.
nivel = solución 1, solución 2, solución 3.
modelo exp = d.b.c.a $\Rightarrow$ muestras dependientes

$$Y_{ij} = \mu + \alpha_i + B_j + \varepsilon_{ij}$$

* μ = crecimiento bacteriano promedio de leche
* α_i = efecto de la i-ésima solución de lavado
* B_j = efecto del j-ésimo día
* ε_{ij} = error aleatorio del j-ésimo dia debido a la solución i-ésima.

Rechazo Ho si =

$P(\varepsilon_1) = 0{,}01 = 1\%$

$F_{obs} = \dfrac{C.M\ trat}{C.M\ error} > F_{critica} = F(I-1);(I-1).(J-1);(1-\alpha)$

$F_{obs} = 44{,}89$

$F_{critica} = F_{2;10;0{,}99} = 7{,}56$

$F_{obs} > F_{critica} \Longrightarrow$ rechazo Ho

Conclusión = con un nivel de significación del 1% concluyo que al menos un valor promedio del crecimiento bacteriano en contenedores de leche varía en presencia de algunas de las soluciones

$P_{asociada} = P\left(F_{(I-1);(I-1).(J-1)} > F_{obs}\right) =$

$1 - P\left(F_{2;10} < 44{,}89\right) \cong 1 - 0{,}995 \cong 0{,}005$

$\Longrightarrow P_{asociada} \cong 0{,}005 < \alpha = 0{,}01 \Rightarrow$ rechazo Ho.

$P_{asociada} = P\left(F_{(J-1);(I-1).(J-1)} > F_{obs}\right) =$

$1 - P\left(F_{5;10} < 21{,}67\right) \cong 1 - 0{,}995 \cong 0{,}005$

* Conclusión = con un nivel de significación 5% podemos concluir en base al método de scheffe que la solución 3 es la más efectiva p/ retardar el crecimiento bacteriano en contenedores de leche (presenta un menor valor medio del nº colonias bacterianas) se puede concluir que u_3 difiere significativamente de u_1 y u_2.

Análisis de correlación = unidad n° 9

Objetivo: estudiar el comportamiento de 2 variables aleatorias. (x e y)

1) estimar el comportamiento promedio de y en función de x.
2) estimar el comportamiento promedio de x en función de y.

"cambio de roles"

3) obtener una medida del grado de asociación entre ambas variables.

Esquema general de las observaciones.

X	Y
x_1	y_1
x_2	y_2
...	...
x_i	y_i
x_I	y_I

$(x_i ; y_i)$ pares observados de valores correspondiente a la i-ésima unidad.

$i = 1, 2, 3, \ldots, I$ → ya no tenemos en cuenta a valores observados de y_i correspondiente a la i-ésima unidad de la variable x.

Ampliamos el concepto de observación

- deseamos medir la fuerza con que se relacionan x e y.

¿Qué herramientas tenemos?

$$\underbrace{Cov_{xy}}_{\text{covarianza}} = \frac{1}{I-1} \cdot \sum_{i=1}^{I} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

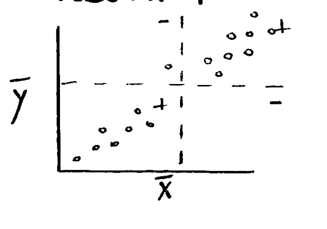
→ ¿nos permite obtener una magnitud identificable de la fuerza de asociación entre las dos variables?

- no, debido a que su rango de variación no está acotado.
- al cambiar arbitrariamente las unidades de medida puede hacerse grande o muy pequeño.

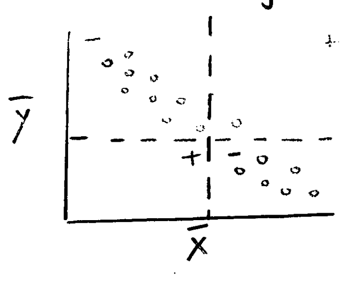
⟹ ¿qué nos brinda su información?

el signo de la covarianza nos permite establecer el tipo de relación

relación positiva.



relación negativa.



Además la covariancia fue un pie para pensar en cómo podríamos dimensionar esta fuerza. $\xrightarrow{\text{surgió}}$ Coeficiente de correlación lineal de Pearson.

$$\boxed{r = \frac{Cov\, x/y}{S_x \cdot S_y}}$$

propiedades.

- su valor no depende de qué variable se considera como independiente y cómo dependiente. ⟵ no sería útil para el concepto de correlación

* es independiente de las unidades en que se miden las variables. "$-1 \leq r \leq 1$"

midiendo intensidad:

si $|r| < 0{,}5$, la relación lineal entre variables es débil

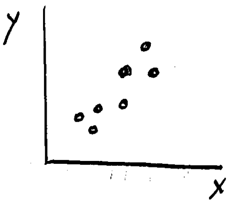
si $|r| \geq 0{,}8$ la relación lineal entre las variables es intensa.

si $0{,}5 \leq |r| \leq 0{,}8$ la relación lineal entre las variables es moderada.

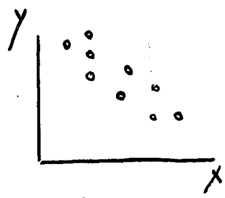
* el signo de "r" depende de la covarianza lo cual nos determina que los valores observados $(x_i; y_i)$ están sobre una recta de pendiente positiva o negativa

* r^2 = coeficiente de determinación

Ejemplos =

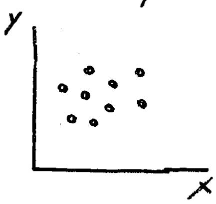


$r \cong 1+$

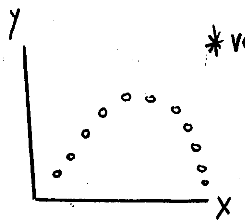


$r \cong 1-$

$r \cong 0$ → no existe asociación



→ * relación no lineal



• Coeficiente de correlación poblacional $\longrightarrow$ parámetro

$$\rho = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$

si x e y son independientes

pero $\rho = 0 \not\Rightarrow$ x e y sean independientes sino que habla de ausencia de relación lineal entre las variables. $(\rho = 0 \Rightarrow B = 0)$.

- $\rho = +1$ ó $-1 \Longleftrightarrow u_{x/y} = \alpha + Bx$ $\underbrace{(B \neq 0)}$

hay relación lineal entre las variables debido a que existe un cambio promedio de la variable dependiente cuando la variable independiente aumenta en 1 unidad.

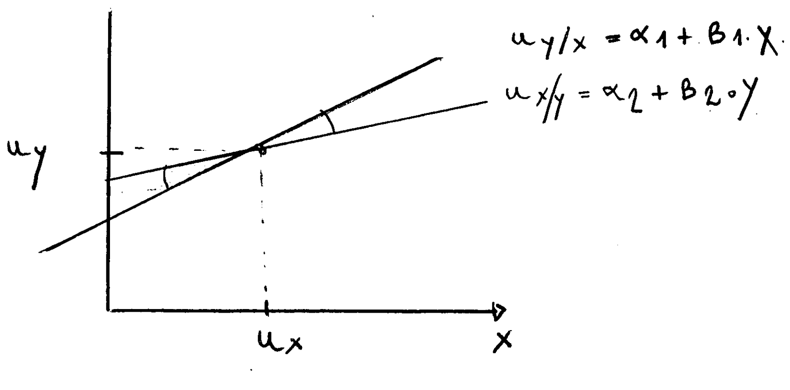
- Podemos estimar las rectas de regresión $\longrightarrow$
  - $u_{x/y} = \alpha_1 + B_1 \cdot y$
  - $u_{y/x} = \alpha_2 + B_2 \cdot x$.

$\downarrow$

- se tienen que intersectar $(\bar{x}; \bar{y})$.

- si $r = 0$ las rectas son paralelas a los ejes y su intersección es perpendicular. (no hay asociación).

- si $r = 1$ las dos rectas coinciden.



- si $\hat{\alpha} > 90° \Rightarrow$ no hay correlación $(r \cong 0)$

- $\alpha \cong 0 \longrightarrow$ correlación perfecta $(1 \cong r$ ó $-1 \cong r)$

- Las funciones de regresión poblacional siempre se intersectan en $(u_x; u_y)$

- Las funciones de regresión muestral siempre se intersectan en $(\bar{x}; \bar{y})$.

-Otra forma de calcular los coeficientes de correlación =

muestral = $r = \pm\sqrt{b_1 \cdot b_2}$ poblacional = $\rho = \pm\sqrt{B_1 \cdot B_2}$

- pendiente $\longrightarrow$ cuando b_1 y b_2 ó B_1 y B_2 son + $\Longrightarrow$ r ó ρ son +.