I T

RESUMEN ESTADISTICA
~ FINAL ~

Descriptiva (Tema I)

ESTADISTICA

Inferencial: se ocupa de las tecnicas para tomar decisiones basandose en el analisis de los datos obtenidos.

Descriptiva: incluye obtención, presentacion y descripcion de datos numericos.

PARAMETRO ↔ Población: conjunto cuyos elementos tienen una o más caract. en común.

ESTIMADOR ↔ Muestra: subconjunto de la población. Sus componentes deben tener caract. comunes iguales q' en la pobl.

MUESTREO

- Población es infinita
- " es homogenea ← cuando
- El proceso es destructivo.

es →

- ALEATORIO: todos los elementos de la población tienen la misma posibilidad de integrar la muestra
- OPINATICO: si el investigador se deja dominar por su opinion
- SIN NORMA: sin usar reglas fijas, sin usar metodo determ.

FENOMENO ALEATORIO → es un fenomeno empirico que se caracteriza por ser observado bajo ciertas condiciones. No siempre se obtiene el mismo resultado sino que los diferentes resultados ocurren con regularidad estadistica.

ligados.

EVENTO ALEATORIO

Tiene la propiedad de que la frecuencia relativa se acerca a un valor limite cdo el num. de observaciones tiende a infinito.

PROBAB. DEL EVENTO ALEATORIO

Es el valor limite de la frecuencia relativa.

EXPERIMENTO → Acción mediante la cual se obtiene un resultado e implica la observación de ese.

ALEATORIO: Aquel en el que el resultado no se puede predecir con exactitud.

DETERMINISTA: Aquel en el que el resultado se puede predecir con total exactitud.

VARIABLES

DISCRETA: la variable solo puede tomar ciertos valores en su amplitud de variación.

CONTINUA: la variable puede tomar cq' valor comprendido en la amplitud de variación.

ALEATORIA: Es la cualidad, medida o conteo resultado de un experimento aleatorio.

RESULTADOS

- Constantes → No numericas, de atributos o cualitativas.
  - Nominales
  - Ordinales
- Variables → Numericas o cuantitativas
  - Discretas → Cardinales
  - Continuas.
    - Intervalos
    - Cocientes o razones

Histograma: Representación gráfica de una distribución de frecuencias de una variable cuantitativa continua.

EL AREA DE CADA RECTANGULO REPRESENTA A LA FRECUENCIA ABSOLUTA.

—o—

MEDIDAS DE POSICION.

Media Aritmetica: es el valor que tomaria la variable si estuviese uniformemente repartida entre todos los individuos de la muestra.

Mediana: es el valor central de la distribución, a ambos lados del cual se reparten por mitades las observ. (se deben ordenar los valores)

* Nro. de Orden → da el valor de la posición donde está la mediana.

Modo: es el valor del caracter q' se presenta con mayor frecuencia, aquel que le corresponde el mayor número de observaciones.

Cuartilos: al igual q' la mediana dividen a las observaciones en cuartos. (25%, 50% y 75%).

Semi-Rango: es el valor "central" de los valores distintos de la variable, entre el valor minimo y el valor max.

Prop. de la Media Aritmetica.

1. $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$

2. $\sum_{i=1}^{n} (x_i - \bar{x})^2 \leq \sum_{i=1}^{n} (x_i - A)^2$ → valor cq'.

3. $\bar{x}_{grande} = \bar{\bar{x}}$

4. $\bar{x}_{\text{¢}} = \text{¢}$

5. Si $y_i = x_i + c$ $\forall i$ desde 1 hasta n;
   $\Rightarrow M(y) = M(x \pm c) = \bar{x} \pm c$

6. Si $y_i = x_i . c$ $\forall i$ desde 1 hasta n,
   $\Rightarrow M(y) = M(x . c) = \bar{x} . c$

7. Si $y_i = x_{1i} \pm x_{2i} \Rightarrow M(y) = M(x_1) \pm M(x_2)$
   es decir: $\boxed{\bar{y} = \bar{x}_1 + \bar{x}_2}$

MEDIDAS DE DISPERSIÓN

Rango: Diferencia entre la mayor y la menor de las observaciones. No proporciona una medida de la variabilidad con respecto a alguna medida de posición.

Varianza: media de los cuadrados de los desvíos con respecto a la media aritmetica de la variable. cdo se quiere estimar la varianza de la población a partir de una muestra se trabaja con n-1 g.l., así la estimación es insesgada.

Desvío Estandar: Raíz cuadrada de la varianza. Se dice que los valores se desvían un determinado valor respecto a la media aritmetica. Se puede interpretar en terminos del problema.

Coeficiente de Variación: mide el porcentaje de la variabilidad relativa al promedio. Es independiente de las unidades de medida, es útil para comparar variabilidades, en las que se ha trabajado con ≠ unidades de medida.

Prop. de la Varianza.

① $Var(x) > 0$

② $Var(c) = 0$

③ $Var(x+c) = Var(x) \Leftrightarrow Var(x+c) = Var(x) + Var(c) = Var(x) + 0 = Var(x)$.

④ $Var(x \cdot c) = c^2 \cdot Var(x)$.

→ MEDIA ARITMETICA:

- VENTAJAS:
  - Emplea en su calculo toda la información disponible
  - Se expresa en las mismas unidades q' la variable en estudio.
  - Es el centro de gravedad de toda la distribución, representando a todos los valores observados.
  - Es un valor unico.
  - Se trata de un concepto familiar para todas las personas.
  - Es útil para llevar a cabo procedimientos estadisticos como la comparación de medias de varios conjuntos de datos.

- DESVENTAJAS:
  - Se ve adversamente afectada por valores extremos, perdiendo representatividad.
  - si el conjunto de datos es muy grande puede ser tedioso su calculo manual.
  - No se puede calcular para datos cualitativos.
  - No se pueden calcular para datos que tengan clases de extremo abierto, tanto superior como inferior.

→ MEDIANA

- VENTAJAS:
  - Fácil de calcular si el número de observaciones no es muy grande
  - No se ve influenciada por valores extremos, ya que solo influyen valores centrales.
  - Fácil de entender
  - Se puede calcular para cq' tipo de datos cuantitativos, incluso los datos con clase de extremo abierto.
  - Es la medida de tendencia central más representativa en el caso de variables q' solo admiten la escala ordinal
- DESVENTAJAS:
  - No utiliza en su cálculo toda la información disponible
  - No pondera cada valor por el nº de veces q' se ha repetido
  - Hay q' ordenar los datos antes de determinarla.

→ MODO.

- VENTAJAS:
  - No requiere cálculos
  - Puede usarse tanto para datos cualitativos como cuantitativos
  - Fácil de interpretar
  - No se ve influenciada x valores extremos
  - Se puede calcular en clase de extremo abierto.
- DESVENTAJAS:
  - Para conjuntos pequeños de datos su valor no tiene casi utilidad, si es que existe solo tiene significado en el caso de una gran cantidad de datos.
  - No utiliza toda la información disponible
  - No siempre existe, si los datos no se repiten.
  - En ocasiones, el azar hace que una sola observación no sea representativa del valor más frecuente del conjunto de datos.
  - Difícil de interpretar si los datos tienen 3 o más modas.

Probabilidad (Tema II)

PROBABILIDAD → es una frecuencia relativa, se calcula haciendo el cociente entre el nº de veces que puede esperarse que ocurra el suceso y el nº de resultados posibles.

$0 \leq P(A) \leq 1$ y $\Sigma P(A) = 1$

Experimento: acción, proceso u operación en el que se obtienen resultados bien definidos y q' conllevan a su observación.

Resultado: lo que se obtiene de un solo ensayo del experimento, es decir de una sola repetición del mismo.

Ensayo: acto que lleva a un resultado determinado, de entre los posibles resultados ≠ del experimento.

Espacio Muestral: conjunto formado por todos los resultados posibles de un experimento aleatorio. "**S**"

Punto Muestral: punto del espacio muestral, uno de los resultados posibles del experimento aleatorio. "**s**"

PRINCIPIO DE LA MULTIPLICACIÓN → si los conjuntos $A_1, A_2, \ldots A_k$ contienen respectivamente a $n_1, n_2 \ldots n_k$ elementos. $\exists\ n_1 \cdot n_2 \cdot \ldots \cdot n_k$ maneras de combinar un elemento de A_1, uno de A_2, uno de A_k.

PRINCIPIO DE LA ADICIÓN → Si los conjuntos $A_1, A_2 \ldots A_k$ contienen respectivamente $n_1, n_2 \ldots n_k$ elementos, existen $n_1 + n_2 \ldots + n_k$ maneras en que se puede dar el primero o el segundo, o el k-ésimo conjunto.

Suceso → Un suceso A respecto a un espacio muestral S asociado al experimento E, es simplemente un conjunto de resultados posibles del experimento.

Por ser $A \subset S$, puede estar formado por ningún, uno, dos elementos de S. o hasta coincidir con S.

- Si S es infinito no numerable, aparecen dificultades teóricas y no se tienen en cuenta los A.
- S mismo es un A y $\emptyset$ tmb.
- Si S está compuesto por n elementos, hay 2^n sucesos contenidos en él.

- Clase exhaustiva: Si se tienen los sucesos $A_1, A_2, \ldots A_k$ estos forman una clase exhaustiva de S si cumplen con:

① → $A_j \cap A_i = \emptyset$ para cq' $j \neq i$ (si los k sucesos son mutuamente exclusivos)

② → $\bigcup_{i=1}^{k} A_i = S$ (los k sucesos completan el espacio S).

E_1 = arrojar un dado y observar la cara superior del mismo.

- Suceso imposible → $M = \{$obtener nº 7$\}$ no puede ocurrir nunca.
- Suceso cierto → $N = \{$obtener algún nº entre 1 y 6$\}$

$N = \{x / x \in \mathbb{N} / 1 \leq x \leq 6\}$.

Frecuencia Relativa → $h_i = \frac{f_i}{n}$ ... recordando.

↳ Dado un resultado A, su frecuencia relativa es el nº de veces que se da A en relación con el nº de veces que se da el experimento.

Propiedades:

$\tilde{f}(A) = f(A)/n$ como frec. relat. del suceso A.

↳ No es una f. Pero con el tiempo ∃ una estabilidad.

① $0 \leq \tilde{f}(A) \leq 1$

② $\tilde{f}(A) = 1 \iff$ A ocurre siempre en las n repeticiones.

③ $\tilde{f}(A) = 0 \iff$ A no ocurre nunca.

④ Si A y B son mutua// excluyentes $\Rightarrow \tilde{f}(A \cup B) = \tilde{f}(A) + \tilde{f}(B)$

⑤ $f(A)$ converge a un valor fijo $p(A)$, cuando $n \to \infty$. $\to p = \lim_{n \to \infty} \frac{f_A}{n}$.

Teoría Axiomática de la Probabilidad. → Sea E un experimento aleatorio y S el espacio muestral asociado con ese experimento aleatorio y sea P una función que asigna un nº $\mathbb{R}$ a cada suceso $A \subset S$.

↳ Probabilidad del suceso A $[P(A)]$ nº $\mathbb{R}$ que cumple con:

① $0 \leq P(A) \leq 1$

② $P(S) = 1$

③ Si A_1, A_2, A_3 ... son mutua// excluyentes de par en par, $\Rightarrow P(\bigcup_{i=1} A_i) = P(A_1) + P(A_2) + P(A_3) + \ldots$

Definición axiomática de la probabilidad:

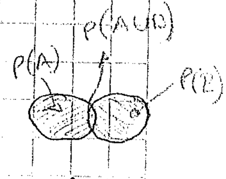
$$\boxed{P(A \cup B) = P(A) + P(B) \text{ si A y B son mutua// excluyentes}}$$

Propiedades de $P(A)$

① Si $\emptyset$ es el vacío $\Rightarrow P(\emptyset) = 0$

② Si A y A' son sucesos complementarios $\Rightarrow P(A) = 1 - P(A')$

③ Sean A y B dos sucesos cq! $\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cup B)$
Si son mutua// excluyentes es $P(A \cup B) = P(A) + P(B)$



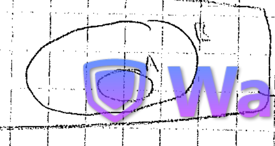
④ Si A, B y C son tres sucesos cq! $\Rightarrow P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cup B) - P(A \cup C) - P(B \cup C) + P(A \cup B \cup C)$.

⑤ Sean A y B dos sucesos tales que $A \subset B \Rightarrow P(A) \leq P(B)$.

↳ Demo: $P(A) \leq P(B)$; si $A \subset B$.

$B = A \cup (B \cap A)$

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



Por el P Axioma $P(B \cap A) \geq 0 \to P(A) \geq P(B) \to P(A) \leq P(B)$

Eventos Compuestos.

1. Probabilidad de evento complementario.

$P(\bar{A}) = P(A_c)$ ⇒ conjunto de todos los puntos muestrales que no pertenecen al conjunto

2. Probabilidad de que ocurra A ó B. (Regla de la Suma)

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ⇒ sean A y B dos eventos en S.

ó

$P(A \cup B) = P(A) + P(B)$ ⇒ Si A y B son mutuall excluyentes.

3. Probabilidad de que ocurran A y B. (Regla del Producto)

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

* Eventos mutuamente excluyentes → la ocurrencia de uno impide la ocurrencia del otro en un mismo ensayo. $A \cap B = \emptyset$.

Si A y B son independientes. → $P(A \cap B) = P(A) \cdot P(B)$.

4. Probabilidad de que ocurra A si ya ha ocurrido B.

* Eventos Probabilísticall independiente: la ocurrencia de uno no modifica la probabilidad de ocurrencia del otro en mas de un ensayo.

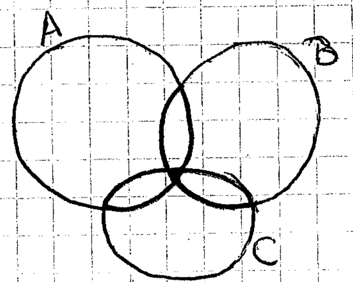
A y B son independientes si $P(A) = P(A/B)$

En general, dados A y B.

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

(*) Generalización de la regla de la suma: Sean A, B y C sucesos no excluyentes

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



Teorema de Bayes → $P(B_i/A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i) \cdot P(A/B_i)}{\sum P(B_i) \cdot P(A/B_i)}$

- Supongamos que se tiene una clase exhaustiva de "n" sucesos. $B_i = \{B_1, B_2, \ldots B_n\}$, que por eso mismo constituye una partición de S.

* Porque los conjuntos B_i forman una partición si se da que:
① $B_i \cap B_j \neq \emptyset$; $\forall i \neq j$.
② $\cup B_i = 1$.
③ $P(B_i) > 0$; $\forall i$.

- Por tratarse de una clase exhaustiva los sucesos B_i son mutuamente excluyentes y completan el espacio muestral.

$$S = B_1 \cup B_2 \cup \ldots \cup B_n.$$

- Supongamos, que además, se tiene un subconjunto $A \subset S$.

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)$$

Por 3º axioma:

$$P(A) = P(A \cap B_1) \cup P(A \cap B_2) \cup \ldots \cup P(A \cap B_n)$$

- Aplicando el teorema del producto es:

$$P(A \cap B_1) = P(B_1) \cdot P(A/B_1)$$

y entonces

$$P(A) = P(B_1) \cdot P(A/B_1) + P(B_2) \cdot P(A/B_2) + \ldots + P(B_n) \cdot P(A/B_n)$$

$$P(A) = \sum P(B_i) \cdot P(A/B_i)$$

- Por otro lado, se quiere calcular la probabilidad de que, habiéndose obtenido A, este provenga de un B_i.

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

y como $P(B_i \cap A) = P(B_i) \cdot P(A/B_i)$.

- Se tiene finalmente:

$$P(B_i/A) = \frac{P(B_i) \cdot P(A/B_i)}{\sum P(B_i) \cdot P(A/B_i)}.$$

Funcion de Probabilidad (Tema III)

* Es un grafico, una tabla, una fórmula o algún otro sistema que nos permite especificar todos los valores posibles de una variable aleatoria junto con sus probabilidades respectivas.

* Especifica las probabilidades de que una variable sea igual a un valor determinado.

Para Variable Discreta (FUNCION DE CUANTIA)

Para una f. de cuantia se debe cumplir:

① $p(x_i) \geq 0 \quad \forall i = 1, \dots n.$

② $\sum p(x_i) = 1$

Y se la define como:

$$p(x) \begin{cases} p(x) & \forall i = 1, \dots n \\ 0 & \text{en otro caso.} \end{cases}$$

Para Variable Continua: (FUNCION DE DENSIDAD)

Para una f. de densidad se debe cumplir:

① $f(x_i) \geq 0 \quad \forall x_i \in (-\infty ; +\infty).$

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

③ Si $A: \{x / a \leq x \leq b\}$ entonces:

$$f(A) = \int_a^b f(x) \cdot dx.$$

Y se la define como

$$F(x) = \begin{cases} f(x_i) & \forall i = 1 \dots n \\ 0 & \text{para otro valor.} \end{cases}$$

Función de Distribución Acumulativa

Para V. Discreta:

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

a) $0 \leq F(x) \leq 1 \quad ; \forall x.$

b) $F(-\infty) = 0 \quad \forall x \in$ al intervalo de definición

$F(+\infty) = 1$

Función monotona no decreciente.

c) $F(x = m) = F(m) = \sum_{x \leq m} p(x) \quad \forall x \leq m.$

Para V. continua

$$F(x = m) = \int_{-\infty}^{m} f(x) \cdot dx.$$

a) $F(x) = \int_{-\infty}^{+x} f(x) \cdot dx \quad \forall x \in (-\infty, +\infty)$

b) Función monotona no decreciente o absolutall continua.

c) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$

$F(+\infty) = \lim_{x \to +\infty} F(x) = 1.$

Esperanza:

Para v. discreta

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

Para v. continua.

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

Varianza

$$V(x) = \sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot p(x_i)$$

$$V(x) = E(x^2) - (E(x))^2$$

$$V(x) = E(x_i^2) - (E(x_i))^2$$

Distribuciones de Probabilidad (Tema IV)

* MODELO → la palabra en sentido probabilístico, se refiere normal// a una expresión o fórmula matemática empleada para estudiar los resultados de un experimento o para predecir el comportamiento de futuras repeticiones de dicho experimento.

→ Clasificación:

Estudiamos estas...

* Distribuciones discretas finitas.

  El dominio de la variable aleatoria es un conjunto finito de nº.

* Distribuciones discretas contables.

  El dominio es una sucesión de numeros pero no limitada. con $\lim_{n \to \infty} \sum_{i=1}^{n} p_i = 1$.

* Distribuciones continuas.

  El dominio viene dado por un intervalo.

→ Modelo de Bernoulli

* El modelo probabilístico más sencillo es el que tiene solo 2 eventos simples en el espacio muestral. A los dos resultados posibles mutuamente excluyentes que completan el espacio muestral se los denomina exito y fracaso.

$$\left.\begin{array}{l}(1)\ \text{Exito} \to A \\ (0)\ \text{Fracaso} \to \bar{A}\end{array}\right\}$$ son de naturaleza cualitativa.

en estadística.

valores numericos.

* Podemos definir una variable aleatoria Bernoulli X que tiene los valores posibles 0 y 1. Es decir, x cuenta la factibilidad de A.

$$X = 0, 1.$$

x_i	$P_x(x_i)$
0	$P_x(0) = P(\bar{A}) = 1 - P(A) = 1 - \pi$.
1	$P_x(1) = P(A) = \pi$.

* Se obtienen 2 resultados posibles, mutuamente excluyentes: EXITO y FRACASO

* Las $P(A) = \pi$ se mantienen ctes.

* Los resultados son independientes.

Distribución Binomial

* El modelo de Bernoulli constituye la base, pues la distribución binomial es la repetición de n ensayos de bernoulli.

* Suponemos: (condiciones)

a) El resultado de cada experimento básico puede clasificarse dentro de una de 2 categorías: Exito y Fracaso (A y $\bar{A}$)

b) La probabilidad de un éxito, π, es la misma en cada repetición del experimento básico o ensayo.

$$P(A) = \pi$$

c) Se realizan n ensayos, donde n es un numero finito constante

d) Los resultados son independientes.

$\pi \rightarrow$ prob. del éxito
$1-\pi \rightarrow$ num. de fracasos
$x \rightarrow$ num. de exitos
$n \rightarrow$ num. total de experimentos.

$$\boxed{bin\,(x, n, \pi) = \binom{n}{x} \pi^{x} (1-\pi)^{n-x}}$$

Expresión que define a una familia de distribuciones de probabilidad donde cada miembro es caracterizado por valores específicos de n y π llamados parametros del modelo.

* Esperanza $\rightarrow$ $E(x) = n \cdot p$

* Varianza $\rightarrow$ $V(x) = n \cdot \pi \cdot (1-\pi)$

Distribución Multinomial o Polinomial

La distribución binomial es un caso particular de la multinomial, ya que se cumplen los supuestos b), c) y d) pero a) es:

a) Los resultados de cada experimento básico pueden clasificarse dentro de S categorías: $k_1, k_2, k_3 \dots k_s$. (si $S=2$ tenemos distrib. binomial

$$\boxed{\Pi_n = \frac{n!}{x_1! \cdot x_2! \dots x_n!} \cdot \pi_1^{x_1} \cdot \pi_2^{x_2} \dots \pi_n^{x_n}}$$

* Esperanza $\rightarrow$ $E(x_i) = n \cdot \pi_i$

* Varianza $\rightarrow$ $Var(x_i) = n \cdot \pi_i \cdot (1-\pi_i)$ para $i = 1, 2, 3 \dots m$

Distribución Geométrica.

→ Condiciones.

a) El experimento se realiza un numero finito de veces desconocido ⊛

b) Dos resultados posibles, E y F. Mutuamente excluyentes.

c) Los resultados son independientes.

d) Las probabilidades se mantienen constantes.

$$G(x) = \pi \cdot (1-\pi)^{x-1}$$

* Esperanza → $E(x) = 1/\pi$
* Varianza → $V(x) = \dfrac{(1-\pi)}{\pi^2}$

Distribución Hipergeométrica

→ conditions (bo!):

a) Existen dos resultados posibles, exito y fracaso.

b) Se realizan n ensayos, donde n es un numero finito cte, pero ojota! ↓

Las repeticiones del experimento basico se hacen sin reposición

∴ c) Los resultados son DEPENDIENTES y...

d) Las probabilidades no se mantienen constantes.

$$H(N, n, x, p) = \frac{\binom{N-k}{n-x}\binom{k}{x}}{\binom{N}{n}}$$

ya ni ganas de escribir tenía !!

K → exito de la pobl
X → " de la muestra
N → num de pobl
n → num de muestra

→ Finitos!

* $E(x) = n \cdot \left(\dfrac{k}{N}\right)$
* $V(x) = n\left(\dfrac{k}{N}\right) \cdot \dfrac{(N-k)}{N} \dfrac{(N-n)}{(N-1)}$

Distribución Hipergeométrica Generalizada.

* Condiciones:

a) El experimento básico se repite un número finito de veces.

b) Existen más de dos resultados posibles (como la Multinomial) mutuamente excluyentes.

c) Las respuestas son dependientes

d) Las probabilidades no se mantienen constantes.

$$\boxed{HG(N, k_1, k_n, x_1, x_n) = \frac{\binom{k_1}{x_1}\binom{k_2}{x_2}\cdots\binom{k_n}{x_n}}{\binom{N}{n}}}$$

* Esperanza → $E(x) = n\left(\frac{k_i}{N}\right)$

* Varianza → $V(x) = n\,\frac{k_i}{N}\,\frac{(N-k)}{N}\cdot\frac{(N-n)}{(N-1)}$

Distribución de Poisson:

* Condiciones (esto no tiene las bolas llenas... ??)

a) Es posible dividir el intervalo continuo considerado en subintervalos

b) La probabilidad de más de un acierto se mantiene constante a lo largo del intervalo.

c) La probabilidad de más de dos aciertos en un intervalo es lo suficientemente pequeña como para ignorarla.

d) Los aciertos son independientes entre intervalos.

e) La probabilidad de que ocurra un acierto es proporcional al tamaño del intervalo.

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

λ = Promedio de aciertos que ocurren en el intervalo continuo.

x = número de sucesos que ocurren en el intervalo continuo.

* Esperanza → $E(x) = \lambda$

* Varianza → $V(x) = \lambda$

Inferencia (Tema V)

* Concepto de distribución muestral:

Si una muestra aleatoria de tamaño n se elige de una población que tiene media μ y varianza σ^2, entonces $\bar{x}$ media es una variable aleatoria cuya distribución tiene media μ y:

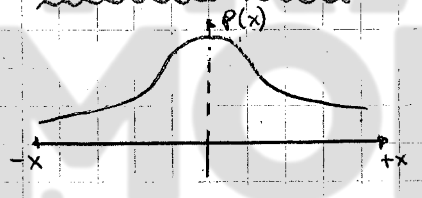
→ para muestras de poblaciones infinitas, la varianza de esta distribución es:

$$\sigma^2/n.$$

→ para muestras de poblaciones finitas de tamaño N, la varianza es:

$$\left(\frac{\sigma^2}{n}\right)\cdot\left(\frac{N-n}{N-1}\right)$$

* Distribución Normal



* simétrica
* asintótica al eje horizontal.

$$f(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot e^{-1/2\cdot\left(\frac{x-\mu}{\sigma}\right)^2}\; dx.$$

si llamamos → $z = \dfrac{x-\mu}{\sigma}$

$$z\cdot\sigma + \mu = x.$$

derivamos parcialmente → $\dfrac{\partial x}{\partial z} = \sigma$

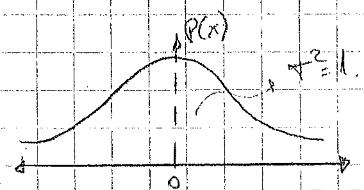
se transforma en derivada total → $\dfrac{dx}{dz} = \sigma$

$$dx = \sigma\cdot dz$$

∴ $$f(z) = \frac{1}{\sqrt{2\pi}\cdot\cancel{\sigma}}\cdot e^{-1/2\cdot z^2}\cdot\cancel{\sigma}\cdot dz$$

$$f(z) = \frac{1}{\sqrt{2\pi}}\cdot e^{-1/2\cdot z^2}\, dz$$

$X \sim N(\mu, \sigma^2) \xrightarrow{\text{transformada}} z \sim N(0,1) \rightarrow$

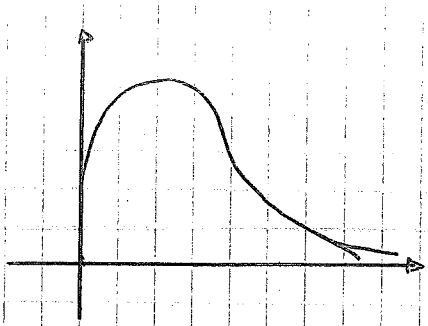


$\sim$ se distribuye normal//

* Distribución χ^2

$$V = \sum_{i=1}^{n} z_i^2$$

$$V = \sum_{i=1}^{n} \left(\frac{x-\mu}{\sigma}\right)^2 \Rightarrow \text{distribución}$$



$$\chi^2_{(gl)} = \frac{(n-1).\,S^2}{\sigma^2}$$ desviación estándar muestral.

* Distribución t

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

* Distribución F

$$F_{(n_1-1)(n_2-1)} = \frac{S_1/\sigma_1^2}{S_2/\sigma_2^2}$$

* TEOREMA CENTRAL DEL LIMITE

Para variables continuas el $\bar{x}$ de una muestra n es aproximado para el de la población. Cuanto mayor sea el n, es más cercano.

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

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

$\mu \rightarrow$ parametro. (población)

$\bar{x} \rightarrow$ estimador (muestra).

$$t = \frac{x-\mu}{S} \rightarrow t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$$

Error max de la estimación: para z.

$$E_{max} = \frac{z.\,\sigma}{\sqrt{n}}$$

Error max. de estimación para t

$E = t_{\alpha/2}.\,S/\sqrt{n}$

→ Para μ con σ^2 conocida:

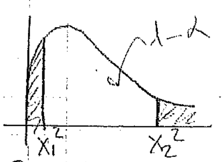
$$\bar{x} \pm \frac{z_{(1-\alpha/2)}.\,\sigma}{\sqrt{n}} \rightsquigarrow P\left(\bar{x} - z_{(1-\alpha/2)}.\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{(1-\alpha/2)}.\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

→ Para μ con σ^2 desconocida

$$\bar{x} \pm t_{n-1(1-\alpha/2)}.\frac{S}{\sqrt{n}} \rightsquigarrow P\left(\bar{x} - t_{(n-1)(1-\alpha/2)}.\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{(n-1).(1-\alpha/2)}.\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

→ Para σ^2

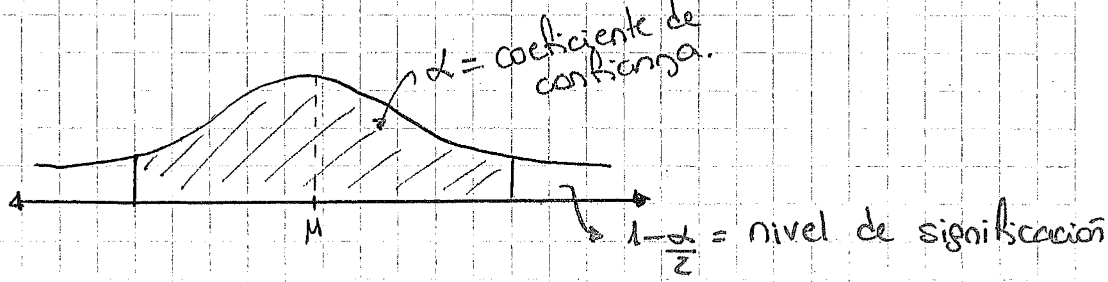
$$\chi^2_{(n-1)} = \frac{(n-1).\,S^2}{\sigma^2} \rightsquigarrow P\left(\frac{(n-1).\,S^2}{\chi^2_2} < \sigma^2 < \frac{(n-1).\,S^2}{\chi^2_1}\right) = 1-\alpha$$



→ Para π.

$$z = \frac{p-\pi}{\sqrt{p.(1-p)}} \rightsquigarrow P\left(p - z_2\sqrt{\frac{p.(1-p)}{n}} < \pi < p + z_2\sqrt{\frac{p.(1-p)}{n}}\right) = 1-\alpha$$

* Graficquillo:



• Estadística de Prueba

La estadística de prueba es una estadística que se deriva del estimador puntual del parámetro que estamos probando y en ella basamos nuestra decisión acerca de si rechazar o no la hipótesis nula

Ejemplo: $Z = \frac{\hat{\mu} - \mu_0}{\sigma / \sqrt{n}}$

Siempre se calcula considerando la hipótesis nula como si fuera verdadera.

* Para el caso específico de la media poblacional μ, el estimador es $\bar{X}$, cuya varianza es σ^2/n. Supondremos que conocemos la varianza σ^2

Hipótesis Nula: $H_0 : \mu = \mu_0$
Hipótesis Alternativa $H_1 : \mu < \mu_0$ ó $H_1 : \mu > \mu_0$ ó $H_1 : \mu \neq \mu_0$

Estadística de Prueba o V.P → $\boxed{Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}}$

* Si nuestro propósito está en la proporción de éxitos π, el estimador será p, cuya varianza es $p(1-p)/n$.

Hipótesis nula: $H_0 : \pi = \pi_0$
" alternativa: $H_1 : \pi < \pi_0$ ó $H_1 : \pi > \pi_0$ ó $H_1 : \pi \neq \pi_0$

Estadística de Prueba → $\boxed{Z = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}}}$

* Cuando la varianza poblacional no es conocida, sabemos que la podemos estimar con la varianza muestral, siendo la distribución una t-Student con n-1 grados de libertad.

(Hipótesis nula: $H_0 : \mu = \mu_0$. ~~tabla~~)

$$\boxed{t_{n-1} = \frac{\bar{X} - \mu}{S / \sqrt{n}}}$$

* Cuando nuestro interés está en el parámetro de variabilidad, en cuyo caso podemos hacer las pruebas sobre un valor específico de la varianza poblacional. Para ello nos basamos en el estimador del estimador de σ^2 que es una χ^2 con n-1 gl.

$$\boxed{\chi^2_{n-1} = \frac{(n-1) . S^2}{\sigma^2}}$$

σ^2

* Pasos a seguir en una prueba de Hipotesis

① - Planteo de la Hipotesis

② - Elección del nivel de significación

③ - Elección del estadístico de prueba

④ - Determinación de la región crítica y la regla de decisión

⑤ - Calculos

⑥ - Decisión y conclusión.

— o —

Usos de Chi² - ANALISIS DE FRECUENCIAS.

Cuando se tienen datos en FORMA DE FRECUENCIAS se puede realizar una prueba de hipótesis utilizando como distribución de contraste, Chi².

Consideraciones a tener en cuenta.

- FRECUENCIAS OBSERVADAS: es la frecuencia absoluta que se observa para cada categoría a considerar.
- FRECUENCIAS ESPERADAS: es la frecuencia que se esperaría tener en cada categoría si la hipótesis nula fuera cierta.

TABLA DE CONTINGENCIA: → presentación de datos en una clasif. a doble entrada. los datos están ubicados en celdas.

* ESTADISTICO DE CONTRASTE → $\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

Se pueden realizar tres tipos de pruebas diferentes, según la pregunta q' nos aqueja:

- BONDAD DE AJUSTE → 1 variable con sus categorías. y las frecuencias observadas para cada categoría.

- INDEPENDENCIA
- HOMOGENEIDAD } → 2 variables, con sus categorías y sus frecuencias observadas para c/ combinación posible.

* Bondad de Ajuste → gl → $k-1-r$ ← cantidad de parametros estimados. x lo gral siempre es cero. xq' la ultima categoría no es aleatoria.

* Independencia → $gl = (f-1).(c-1)$

Para frecuencia esperada → Se supone q' son independientes $P(A \cap B) = P(A).P(B)$

$E = n.p \rightarrow E = n.p(f_i \cap C_j) \rightarrow \boxed{E = n.\frac{f_i}{n}.\frac{c_i}{n}}$

* Homogeneidad → $gl = (f-1).(c-1)$ Se prueba la hipotesis de que la proporción de las filas es la misma. (homogeneas).

$E = F_i \times$ Probabilidad de columna $= n_i\left(\frac{C_i}{n}\right)$

* Estadistica de Prueba.

La estadistica de prueba es alguna estadistica que se puede calcular a partir de los datos de la muestra, sirve como un productor de decisiones; ya que la decisión de rechazar o no rechazar la hipotesis nula depende de la magnitud de la estadistica de prueba.
Siempre se calcula conciderando la hipotesis nula como si fuera verdadera.

- Para el caso especifico de la media poblacional μ y conociendo la varianza, la estadistica de prueba será:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$ (Distribución normal)

- En el caso que no se conozca la varianza poblacional, la estadistica de prueba será:

$$t_{n-1} = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$ (Distribución t-Student con n-1 grados de libertad)

- Para el caso especifico de proporción π, la estadistica de prueba será:

$$z = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}}$$ (distribución normal)

- Para el caso de querer estimar la varianza poblacional σ^2, la estadistica de prueba será:

$$\chi^2_{n-1} = \frac{(n-1)\, s^2}{\sigma^2}$$ (Distribución Chi cuadrado con n-1 grados de libertad)

Notas:

	Ho Verdadera	Ho Falsa
Rechazamos Ho	Error Tipo I P(error tipo I) = α	Decisión Correcta
No Rechazamos Ho	Decisión Correcta	Error Tipo II P(error tipo II) = β

¿? ↳ no * La probabilidad de cometer un error Tipo I se conoce como nivel de significación, se denota como α. ~~y es el tamaño de la región de rechazo.~~ ~~El complemento de la región de rechazo es $1-\alpha$ y es conocido como el coeficiente de confianza~~ (*)

~~* En una prueba de hipótesis de dos colas la región de no rechazo corresponde a un intervalo de confianza para el parámetro en cuestión.~~

• Región de Rechazo:

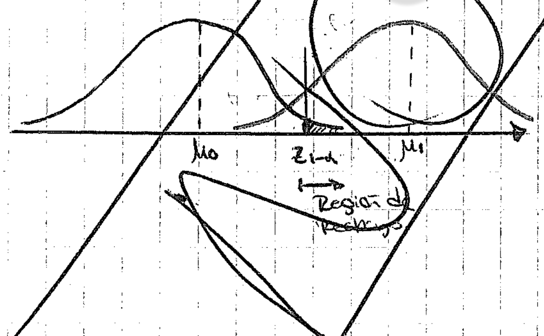
La región de rechazo es el conjunto de valores tales que si la prueba estadística cae dentro de este rango decidimos rechazar la hipótesis nula. Su localización depende de la forma de la hipótesis alternativa.

Por ejemplo: • si $H_1: \mu > \mu_0$ entonces la región se encuentra en la cola derecha de la distribución de la estadística de prueba.

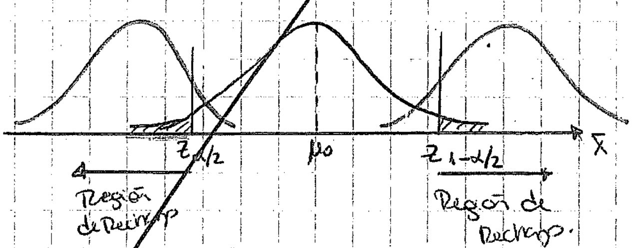
• si $H_1: \mu < \mu_0$ " " " " en la cola izq.

• si $H_1: \mu \neq \mu_0$ entonces la región se divide en dos partes, una parte estará en la cola derecha y la otra en la cola izquierda de la distribución estadística de prueba.

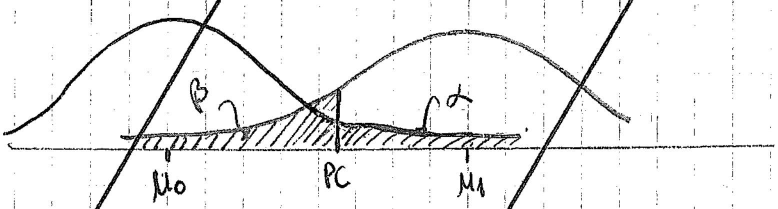
$H_1 : \mu > \mu_0$



$H_1 : \mu \neq \mu_0$



(*)



α = Probabilidad de cometer Error Tipo I

β = " " " Error Tipo II

$1-\beta$ = Potencia de la Prueba.

Regresión y Correlación (Tema VII)

↳ Las dos variables son medidas en forma ALEATORIA. Por lo que no hay una relación de dependencia de una a la otra

↳ Siempre hay una variable dependiente (y) que es medida en forma aleatoria y una variable independiente (x) que es considerada fija.

* Cuando se tienen datos bivariados, uno puede estar interesado en conocer:

- La forma de la relación entre las dos variables medidas (Regresión)
- Predecir una de las variables a partir de la otra (Regresión)
- La $\underbrace{\text{fuerza de la relación}}_{⊛}$ entre las dos variables (correlación)

* CORRELACIÓN

Tenemos dos variables x e y.

* Asumimos que las dos variables estan distribuidas normalmente con una media y una varianza

$$x \sim N(\mu_x; \sigma_x^2) \qquad y \sim N(\mu_y; \sigma_y^2)$$

Y por lo tanto la distribución bivariada que obtenemos también lo estara:

$$xy \sim N(\mu_x \mu_y; \sigma_x^2 \sigma_y^2; \rho).$$

* Tambien asumimos que cualquier relación entre x e y es una relación lineal

⊛ Covarianza: es la medida de la fuerza de la relación entre dos variables.

$$S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

estandarizando → COEFICIENTE DE CORRELACIÓN $\begin{cases} r \to \text{muestra} \\ \rho \to \text{población} \end{cases}$

$$-1 < r < 1$$

si $r = 0$ no hay relación entre las variables.

- Test de hipótesis para ρ

① $H_0: \rho = 0$
$H_1: \rho \neq 0$ } Para saber si existe o no una relación entre las variables. (grado de relación)

② α

③ $t_{n-2} = \dfrac{r - \cancel{\rho}^{\,0}}{\sqrt{\dfrac{1-r^2}{n-2}}}$

⊛ REGRESIÓN

* Supuestos:

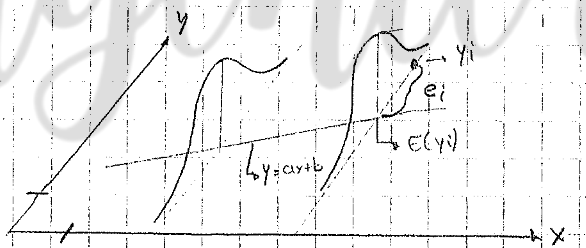
1- Los valores de x no son aleatorios, están fijos.

2- Para cada valor de x hay un conjunto de valores de y que están normal// distribuidos.

3- Las varianzas de las subpoblaciones de y son iguales.

4- Las medias de las subpoblaciones de y están sobre una recta.

5- Los valores de y son independientes entre sí.



* Modelos de Regresión:

Teórico → $E(Y/x) = \mu_{Y/x} = \alpha + \beta x$

Estadístico → $Y = \alpha + \beta x + e$

Estimado → $Y = a + bx$

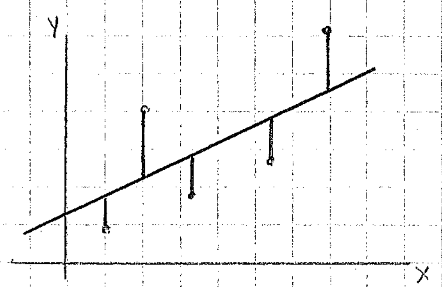
α y β → Parámetros de la regresión

a y b → Estimadores de la regresión

e → error.

* Estimación de los parametros de regresión.

METODO DE MINIMOS CUADRADOS.



distancia según lineas verticales.

$$b = \frac{\sum (x_i - \bar{x}).(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad a = \bar{y} - b.\bar{x}$$

* Test de hipotesis para β.

① $H_0: \beta = 0$
$H_1: \beta \neq 0$ } hay o no hay pendiente?

② $\alpha =$

③ $t_{n-2} = \frac{b - \beta^{\to 0}}{S_b}$

$$S_b = \frac{S_e^2}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \qquad S_e^2 = \frac{1}{n-2}\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n} - b^2\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\right)$$

* Test de hipotesis para α

① $H_0: \alpha = 0$
$H_1: \alpha \neq 0$ } corta a la ordenada al origen?

② Nivel de significación

③ $t_{n-2} = \frac{a - \alpha}{S_a}$

$$S_a = S_e^2 . \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}\right)$$

ANOVA (Tema VIII)

↳ Análisis de la varianza en un diseño de experimento.

* Principios básicos:
  - REPETICIÓN
  - ALEATORIZACIÓN
  - CONTROL LOCAL: tratar de disminuir la variabilidad.

* El error experimental se puede reducir con:
  - DISEÑO EXPERIMENTAL APROPIADO
  - SELECCIÓN DEL MATERIAL
  - INCREMENTANDO EL Nº DE EXPERIMENTOS
  - PERFECCIÓN DE LA TÉCNICA EXPERIMENTAL

* Identidad para Análisis de Varianza Sencilla:

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2 + n\cdot\sum_{i=1}^{k}(\bar{y}_i-\bar{y})^2$$

$$S.C_{TOTAL} = S.C_{error} + S.C_{tratamientos}.$$

* La tablita de la muerte:

Fuente de variación	S.C	g.l	C.M	F
Tratamientos	$\sum_{i=1}^{k}(\bar{y}_i-\bar{y})^2$	$k-1$	$S.C_{trat}/k-1$	CM_{trat}/CM_{error}
Error	$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2$	$k(n-1)$	$S.C_{error}/k(n-1)$	—
Total	$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y})^2$	$(n.k)-1$	—	—

k: cantidad de tratamientos

n: muestra total

* Modelo

$$y_{ij} = \mu + \tau_i + e_{ij}$$

y_{ij} = j-ésima observación del i-ésimo método

μ = promedio total de los métodos.

τ_i = efecto del i-ésimo tratamiento

e_{ij} = i-ésimo error del j-ésimo tratamiento

Dudas.

Probabilidad

Si los eventos son mutuamente excluyentes son independientes?

¿Ej. del dado?

Hay alguna relación entre mutuamente excluyentes e independientes?

Distribuciones de Probabilidad.

- ~~Las variables es discretas?~~
- Siempre los dos resultados serán mutuamente excluyentes? Pó?? No /
  excepto en distribución hipergeométrica generalizada.

Inferencia

α = coef. de confianza $1-\alpha$ → probabilidad de que se estime el parámetro   Prob. de cometer error

y en Dócimas es nivel de significación? xq? Y?

Usos de Chi^2 → Bondad de ajuste → cómo calcular las frecuencias esperadas? Hay un modelo? Lo da el problema, siempre?

Homogeneidad → los gl son igual q' en independencia?
$gl = (F-1)(C-1)$
→ cálculo de valores esperados?

Dócima → EL TEMA → Gráfico.....

Covarianza → variación en conjunto de las 2 variables

Regresión y Correlación

en correlación ρ (para pobl) y r (en muestra) son los coeficientes de correlación y varían entre -1 y 1. Es el grado de relación?

en regresión → tenemos R^2 varía entre 0 y 1 y da el % de los datos que se ajusten a una regresión lineal? Estandarizada?

	a	b	
S			
E			
	Ta		N

	a	b
S	E_1	
E		
	n_a	

Indep

$E_{11} = N \cdot P(a) \cdot P(s)$

$\cancel{N} \cdot \frac{Ta}{\cancel{N}} \cdot \frac{TS}{N}$

$E_{11} = n_s \cdot P(a)$

$= n_s \cdot \frac{TA}{N}$

$E_1 = n_a \cdot P(S)$

Prueba de Hipótesis.

Muchos problemas de ingeniería, ciencia y administración requieren que se tome una decisión entre aceptar o rechazar una proposición sobre algún parámetro. Esta proposición recibe el nombre de HIPOTESIS.

Este es uno de los aspectos más útiles de la inferencia estadística, puesto que muchos tipos de problemas sobre toma de decisiones, pruebas o experimentos en el mundo de la ingeniería pueden formularse como problemas de prueba de hipotesis.

Una hipotesis estadística es una proposición o supuesto sobre los parametros de una o más poblaciones.

Una prueba de hipotesis contiene 4 componentes principales:

- Hipotesis Nula
- Hipotesis Alternativa
- Región de Rechazo
- Estadística de Prueba.

- Hipotesis Nula:

Es la hipotesis que debe probarse, se denota como Ho. En general, se establece con el propósito expreso de ser rechazada. En consecuencia, el complemento de la conclusión que el investigador desea alcanzar se convierte en el enunciado de la hipotesis nula. En el proceso de prueba, puede ser rechazada o no. Si la hipotesis nula no se rechaza, se dirá que los datos sobre los cuales se basa la prueba no proporcionan evidencia suficiente que cause el rechazo.
Debe contener una proposición de desigualdad, ya sea: =, ≥ o ≤.

$Ho: \mu = \mu o$; $Ho: \mu \geq \mu o$; $Ho: \mu \leq \mu o$.

- Hipotesis Alternativa:

Es la que responde nuestra pregunta, la que se establece en base a la evidencia que tenemos. Se denota como H1.

$H_1: \mu \neq \mu o$; $Ho: \mu < \mu o$, $Ho: \mu > \mu o$.

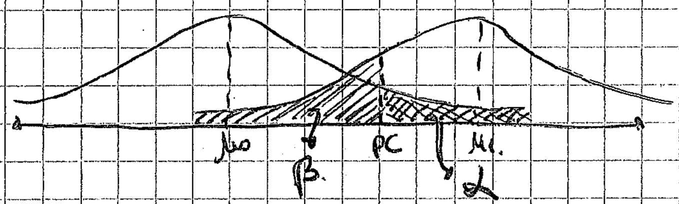
Como las conclusiones a las que llegamos se basan en una muestra, hay posibilidades de que nos equivoquemos:

- Dos decisiones correctas son posibles
  - Rechazar Ho cuando es falsa.
  - No rechazar Ho cuando es verdadera.
- Dos decisiones incorrectas son posibles:
  - Rechazar Ho cuando es verdadera.
  - No rechazar Ho cuando es falsa.

Notas:

	Ho verdadera	H_1 falsa
Rechaza	Error tipo I P(error tipo I) $= \alpha$	✓
No Rechaza.	✓	Error tipo II P(error tipo II) $= \beta$

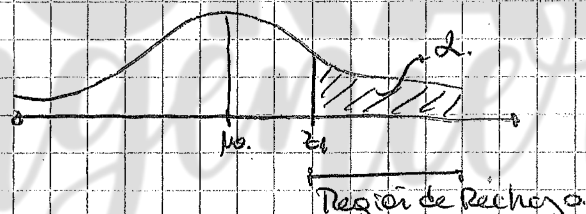
La probabilidad de cometer error tipo I se conoce como nivel de significación y se denota como α.
La probabilidad de cometer error tipo II se denota como β, y a $1-\beta$ se lo conoce como potencia de la prueba.



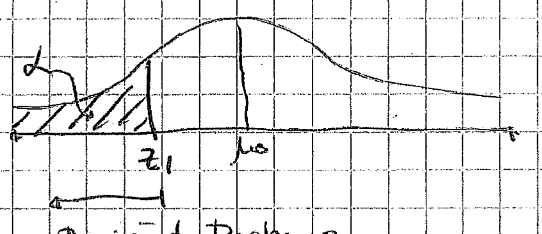
* Región de Rechazo

La región de rechazo es el conjunto de valores tales que si la prueba estadística cae dentro de este rango, decidimos rechazar la hipotesis nula. Su localización depende de la forma de la hipotesis alternativa.

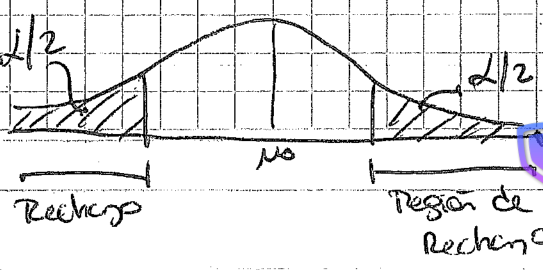
Por ejemplo: • Si H_a: $\mu > \mu_o$ entonces la región se encuentra en la cola derecha de la distribución.



• Si H_a: $\mu < \mu_o$ entonces la región de rechazo se encuentra en la cola izquierda de la distribución.



• Si H_a: $\mu \neq \mu_o$ entonces la región de rechazo se encuentra dividida en dos partes, una parte se encuentra en la cola derecha y la otra en la cola izquierda de la distribución.



Notas: