uc3m | Universidad Carlos III de Madrid

Grado en Ingeniería Informática

2019-2020

Apuntes

Estadística

Jorge Rodríguez Fraile[1]

[1] Universidad: 100405951@alumnos.uc3m.es | Personal: jrf1616@gmail.com

ÍNDICE GENERAL

Parte I

Teoría

Tema 1: Estadística descriptiva univariable.

Estadística: Es una herramienta de aprendizaje a partir de la observación.
Nos ayuda a sacar conclusiones generalizables a partir de un conjunto observable de datos ⟹ Inducción o Inferencia.

Datos cuantitativos: Valores numericos. Ordenables.
- Discretos: Valores finitos, naturales. Edad / nº de hermanos
- Continuos: Valores en intervalos. Altura / Temp

Datos cualitativos: No valores numericos. Sin orden. Estado civil
Categoricos o atributos

Ambos tipos son muy distintos, los cuantitativos aportan mucha más información.

Analisis básico: A la hora de enfrentarse a un conjunto de datos hay que comenzar realizando dos operaciones básicas. Ordenar y Resumir

Frecuencia absoluta (f): Número de veces que aparece cada dato de la variable.
Frecuencia total (n): Número total de datos de la variable.
Frecuencia relativa (fr): Cociente de la frecuencia absoltuta y la total. f/n
Frecuencia acumulada: Supuesta la ordenación de datos de menor a mayor.
- Absoluta (F): La suma de las anteriores y la propia clase.
- Relativa (Fr): La suma de las anteriores y la propia clase.

Todas las frecuencias se representan en la tabla de distribución de frecuencias.

Gráficos para variables cualitativas.

Diagrama de barras: Eje 1 Valor o categoria de variable
Eje 2 Altura proporcional a la frecuencia.


Diagrama de tarta: Circulo dividido en secciones proporcionales a la frecuencia.

Más de 4 o 5 sectores dificultan la lectura del diagrama.

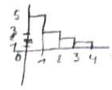
Variables cuantitativas: Las frecuencias son las mismas, pero las variables se agrupan. < clases / intervalo

Rango: Diferencia entre el mayor y menor valor de ésta.
Amplitud de un intervalo: Diferencia entre el extremo superior e inferior.
Marca de clase (m_i): Punto medio de cada intervalo o clase, valor representativo del intervalo.
El número de clases r debe oscilar entre 5 y 20; a menudo se escoge el entero más proximo a $\sqrt{n}$

Gráficos para variables cuantitativas:

Histograma: Representación para variables agrupadas en intervalos.

Abscisas: Intervalo de valor de la variable.

Ordenadas: Altura proporcional a la frecuencia. 

Muestra las tendencias generales de los datos:

Concentraciones: Más de una concentración => Datos heterogéneos.

Huecos: Indicio de que los datos proceden de poblaciones distintas.

Valores atípicos: Aquellos que se separan mucho del patron general.

Interesan para saber si hay error o investigar la aparición.

Asimetrías: Tendencia de los datos cuando nos alejamos de las zonas de concentración.

Positiva  Negativa  

Polinomio de frecuencias: Línea poligonal que resultar de unir los puntos centrales superiores.

Ambos se pueden contruir con la frecuencia acumulada.

Medidas características: Son aquellas que nos permiten resumir con un solo número los rasgos fundamentales de la distribución.

Medidas de tendencia central: Indican el valor medio de los datos.

Media aritmetica ($\bar{x}$): $\bar{x} = \frac{\sum_n x_j f(x_j)}{n}$ Muy sensible a los datos atípicos, por lo que si hay muchos usar mediana.

(Average)

1) $\sum_n (x_i - \bar{x}) = 0$

2) $\bar{y} = \frac{\sum_n (x_j + k)}{n} = \bar{x} + k$

3) $\bar{y} = \frac{\sum_n (x_j k)}{n} = \bar{x} k$

Mediana (Median): Valor que ocupa el centro de la distribución, dividiendolos en 2.

1° Ordenar los datos 2° Ocupa la posición central.

n par → la media aritmetica de los dos centrales. 1 2 3 4 6 7
3'5

n impar → el del medio. 1 2 3 4 5

En tabla buscar $F_r = 0'5$ ó $F = n/2$

Robustez: No Sensible a los datos atípicos.

Moda: El valor más frecuente de la distribución. Aporta poca información.

Para datos cualitativos y cuantitativos discretos. En los intervalos se llama intervalo modal

Estadística T2 9/9/19

Medidas de dispersión: Medida de la separación de los datos, generalmente, respecto a la mediana. Cuanto menos dispersión mejor.

Varianza (S_x^2): $S_x^2 = \dfrac{\sum_n (x_j-\bar{x})^2 \cdot f(x_j)}{n}$ Para calcular a mano: $S_x = \dfrac{\sum_n x_j^2 f(x_j)}{n} - \bar{x}^2$

1) Acotada y positiva. Unidades cuadradas
2) No se ve afectado por el cambio de origen $(x+k)$
3) Si se ve afectado por el cambio de escala (xk) $S_y^2 = k^2 S_x^2$

Cuasivarianza ($\hat{S}_x$) o varianza muestral: $\hat{S}_x = \dfrac{\sum_n (x_j-\bar{x})^2 f(x_j)}{n-1}$

Es la que calculan los programas.
Positiva y muy sensible a los datos atípicos Unidades cuadradas

Desviación típica (Standart Desviation) (S_x): La raíz cuadrada positiva de la Varianza. Unidades normales. $S_x = \sqrt{\dfrac{\sum_n (x_j-\bar{x})^2 f(x_j)}{n}} = \sqrt{S_x^2}$

Cuasidesviación típica ($\hat{S}_x$): $\hat{S}_x = \sqrt{\dfrac{\sum_n (x_j-\bar{x})^2 f(x_j)}{n-1}}$

Coeficiente de variación (CV): Es la medida de dispersión relativa. A estar en porcentaje está en una unidad adimensional

Nos permite: evaluar la representatividad de la media. A menor más representativa
Comparar la dispersión de distribuciones. $CV = \dfrac{S}{|\bar{x}|} \cdot 100$ si $\bar{x} \neq 0$

Coef. correlación = Mide la relación entre las variables

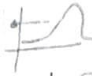
Cuartiles: Son los valores de la variable que dividen la distribución en c partes iguales. Cuartiles (Q) $c=4$ Percentiles (p) $c=100$
Quintiles (k) $c=5$ Q_2 = Mediana

Rango intercuartílicos (RI): Diferencia entre los cuartiles 3 y 1 (o percentil 75 y 25)

16/9/19

Medidas de forma:

Coeficiente de asimetría de Fisher (Skewness): $CA = \gamma_1 = \dfrac{\sum_n (x_i-\bar{x})^3 \cdot f(x_i)}{n S^3}$ ← Desviación típica

$\gamma_1 = 0$ Simetrica 
$\gamma_1 > 0$ Asimetria positiva o a derechas. 
$\gamma_1 < 0$ Asimetria negativa o a izquierdas. 

Coeficiente de apuntamiento o Curtosis (Kurtosis): $CAp = \dfrac{\sum_n (x_i-\bar{x})^4 f(x_i)}{n S^4} - 3$

$CAp = 0$ Curtosis media Mesocúrtica 
$CAp > 0$ Más apuntada Leptocúrtica. 
$CAp < 0$ Más achatada Platicúrtica 

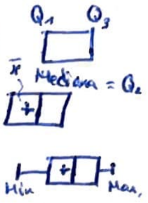
Diagrama de Caja: Muestra las caracteristicas principales y señala los posibles datos atipicos. Necesitamos: Maximo, Minimo y Cuartiles.

$LI = Q_1 - 1'5\,(Q_3 - Q_1)$ → Los que se salen de este rango se consideran
$LS = Q_3 + 1'5\,(Q_3 - Q_1)$ datos atipicos.

1° La caja, los lados son Q_1 y Q_3. (Altura da igual)

2° La Q_2 como recta vertical y la media como un +.

3° Los bigotes son el maximo y minimo no atipicos.



Tema 2: Descripción estadística de variables bidimensionales.

Estudio de 2 caracteres simultáneos en cada elemento de la población.

Distribución conjunta de frecuencias de dos variables: Valores observados y las frecuencias de aparación de cada par. $\sum_i \sum_j fr(x_i, x_j) = 1$

	c	d	$\sum f$
a	5 5/14	7 7/14	12/14
b	4 4/14	3 3/14	7 7/14
$\sum f$	9 9/14	10 10/14	19 1

Distribución marginal: Aparece en los margenes. Distribución de cada una de las variables, consideradas por separado.

$$f(x_i) = \sum_j f(x_i, j) \qquad f(y_j) = \sum_i f(x_i, y_j)$$

Distribución condicionada de y para $x = x_i$: Imponiendo la condición $x = x_i$.

$$f_r(y_j \mid x = x_i) = \frac{f(x_i, y_j)}{f(x_i)}$$

← Cada dato de esa x_i
← El total de esa columna

Representación grafica:

Diagrama de dispersión o nube de puntos.

Covariación: Relación de dependencia entre las variables estudiadas (x, y)

- Dependencia causal unilateral: X influye en Y, pero no a la inversa.
- Interdependencia: X influye en Y, y viceversa.
- Dependencia indirecta: Muestran una covariación mediante una tercera variable.
- Concordancia (Juez de Gimnasia)
- Covariación casual: La relación no tiene que ver, hay que verlo de otra manera.
  Es casualidad

Covarianza: Es una medida descriptiva de la relación lineal entre par de variables.

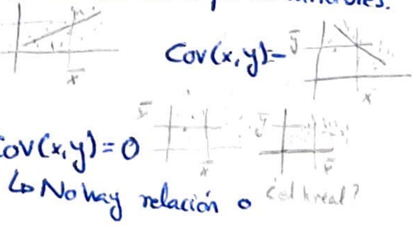
$$\frac{\sum x_i y_i f(x,y)}{n} - \bar{x}\bar{y} = S_{xy} = cov(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \cdot f(x_i)}{n}$$

$cov(x,y) = +$ $\quad$ $cov(x,y) = -$

Positiva.

Para evitar el inconveniente de las unidades se usa el Coef. de correlación lineal r

$cov(x,y) = 0$
↳ No hay relación o ¿el lineal?



Estadística T3 16/9/19

Coeficiente de correlación lineal: $r = \dfrac{S_{xy}}{S_x S_y}$ ← covarianza / ← Desv. típica. Varía entre -1 y 1

$r = -1$ Correlación lineal perfecta e inversa.
$r = 1$ Correlación lineal perfecta y directa.
$r = 0$ No existe correlación, $S_{xy} = 0$

Matriz de covarianzas: Las medidas de dependencia lineal de un conjunto de datos bidimensionales pueden presentarse en forma de matriz.

Matriz de covarianzas muestrales $M = \begin{pmatrix} S_x^2 & cov(x,y) \\ cov(y,x) & S_y^2 \end{pmatrix}$ Matriz de correlaciones muestrales $R = \begin{pmatrix} 1 & Corr(x,y) \\ Cor(y,x) & 1 \end{pmatrix}$

Recta de regresión: Recta que refleja, de la manera más aproximada posible, la evolución conjunta de dos variables. Cuanto más cercano al ± 1 el coef. de correlación será más explicativa la recta.

Para cada x_i tenemos: la y_i ordenada real y la $\hat{y}_i$ que es la de la recta de regresión.
El residuo es la diferencia entre y_i y $\hat{y}_i$; $e\,\big|^{y_i}_{\hat{y}_i}$ e sería el residuo. $e = y_i - \hat{y}_i$
Para minimizar la suma de errores se usa el Método de los mínimos cuadrados de Gauss: $\hat{y}_i = a + bx_i$ será $y - \bar{y} = \dfrac{S_{xy}}{S_x^2}(x - \bar{x})$ Recta de y sobre x; $x - \bar{x} = \dfrac{S_{xy}}{S_y^2}(y - \bar{y})$ x sobre y

Evaluación del modelo

Relación lineal X – Y: Ver los datos representados.

P1: Gráfico de dispersión. P2: Medidas características → * Covarianza S_{xy}; * Coef. de correlación. r_{xy} (*adimensional) — Lineal: 1 Perf. directa; 0 No lineal, pero puede haber otra; −1 Perf. inversa

Gráfico de valores previstos frente a valores observados ⇒ Linealidad: puntos distribuidos linealmente alrededor de la recta.

P5: El + potente * Gráfico de residuos frente a valores previstos ⇒ Linealidad: Puntos distribuidos al azar. Sin patrón

P3: Recta de regresión Ajuste: $\hat{y} = \hat{b}x + \hat{a}$ (pendiente, c/e)

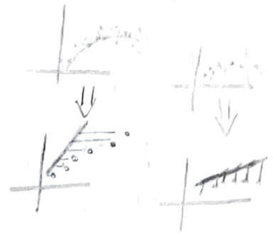
Nube de puntos estrecha:

P4: Bondad del ajuste: Nos indica la proporción de la dispersión de la variable respuesta y que es capaz de explicar la recta de regresión. Cuanto más % más ajustado a la recta de regresión está.
Cuanto más explicativa sea la regresión, menor será la variabilidad que queda en los residuos respecto a la de los datos, y R^2 será mayor.

Coeficiente de determinación $R_2^2 = 1 - \dfrac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = r^2$ coef. de correlación al cuadrado

Recta de regresión adecuada: Si la recta no se ajusta bien a los datos (asimetrica, se curvan), transformamos los datos
No hay / los punta

P6: Transformaciones: Cuando las hipotesis del modelo no se cumplen es necesario transformar los datos, de manera que los datos transformados cumplan la hipotesis. Las más utilizadas son:



Logaritmo

$y=\ln(x)$ $x=\ln(y)$ → Si los puntos estan muy apelotonados

Potencia

Ir probando

$y=x^c$ $x=y^c$

Cambiar los datos x por $\log(x)$ / x^c / $\frac{1}{x}$ / $\sqrt{x}$ y graficar el grafico de dispersión con $\log(x)$ vs y
ejes

Inversa

$y=1/x$ $x=1/y$

Raiz cuadrada

$y=\sqrt{x}$ $x=\sqrt{y}$

23/9/19

Tema 3: Probabilidad.

Experimento aleatorio: Observación de una propiedad de interés que proporciona distintos resultados, sin que pueda precisarse cuál de ellos aparecerá. Para analizarlo debe conocerse el espacio muestral.

Suceso elemental (A, B): Cada uno de los resultados posibles de un experimento aleatorio que verifican: Siempre ocurre alguno de ellos
Son mutuamente excluyentes.

Dado → 1, 2, 3, 4, 5, 6

Suceso compuesto: Aquel construido a partir de uniones de sucesos elementales.

Dado → Pares, Impares → {2,4,6}, {1,3,5}

Espacio muestral (Ω ó E): Unión de todos los sucesos elementales.

Suceso contrario o complementario ($\bar{A}$): Suceso que ocurre cuando no ocurre A 

Suceso seguro: El que siempre se observa.

Suceso imposible ($\emptyset$): Suceso que nunca se puede observar (Fuera del E)

Suceso unión de A y B ($A \cup B$): El suceso A o el suceso B. 

Suceso intersección de A y B ($A \cap B$): El suceso A y B a la vez. 

Sucesos mutuamente excluyentes o disjuntos: Sin elementos comunes. $A \cap B \neq \emptyset$

Leyes de De Morgan: $\overline{A \cup B} = \bar{A} \cap \bar{B}$ $\overline{A \cap B} = \bar{A} \cup \bar{B}$

Probabilidad: Medida de la incertidumbre antes de realizar un experimento aleatorio. Valoración de la verosimilitud de nuestras hipotesis.

Definición clásica (Laplace): $P(A) = \frac{\text{nº casos favorables}}{\text{nº casos probables}}$ Si todos los casos son igualmente posibles (equiprobables)

Limitaciones: Todos los casos igual de probables.

Número de casos posibles finitos.

Definición frecuentista: Probabilidad de un suceso es su frecuencia relativa de aparición si repetimos indefinidamente el experimento.

Propiedades fundamentales:

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

2) $P(E) = 1$ 3) $P(\emptyset) = 0$

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

5) Si $A \cap B = \emptyset$; $P(A \cup B) = P(A) + P(B)$

Si $A \cap B \neq \emptyset$; $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probabilidad condicionada: $P(A/\underset{\text{conocido}}{B}) = \frac{P(A \cap B)}{P(B)}$ La probabilidad de un suceso sabiendo (condicionada a) la ocurrencia de otro suceso.

Independencia: Dos sucesos A y B lo son si el conocimiento de la ocurrencia de uno no modifica la probabilidad del otro.

$P(A/B) = P(A)$ $P(B/A) = P(B)$ $P(A \cap B) = P(A) \cdot P(B)$ ← Condición de independencia.

Incompatibilidad: Dos sucesos A y B lo son si no pueden verificarse simultaneamente.

$P(A \cup B) = P(A) + P(B)$ ← $A \cap B = \emptyset$ Que sean incompatibles, no quiere decir que sean independientes. $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)}$

↳ Condicional es 0

Teorema de la probabilidad total: Sea $A_1, A_2, \ldots, A_k$ una partición del espacio muestral E / $A_1 \cup A_2 \cup \ldots \cup A_k = E$; $A_i \cap A_j = \emptyset \ \forall i \neq j$

Sea R un suceso cualquiera de ese espacio muestral

$$P(R) = \sum_{1}^{k} P(R/A_i) \cdot P(A_i)$$

Teorema de Bayes: Se aplica para calcular la probabilidad de cada una de las posibles causas, una vez observado el efecto.

$$P(A_i/R) = \frac{\overbrace{P(R/A_i)\,P(A_i)}^{P(A_i \cap R)}}{P(R)}$$

30/9/19

Tema 4: Variables aleatorias.

Ortodoxo: Variable cuyo valor numerico está determinado por el resultado de un experimento aleatorio.

Ligeramente heterodoxo: Variable que cuantifica la magnitud de interés, y cuya realización numerica concreta depende del azar (cada valor o intervalo de valores tendrá una probabilidad de aparición)

Se puede identificar cuando pide "Número de..." (Cantidad) o similar y depende de algo aleatorio. Necesitamos conocer el espacio muestral (E)

Se define mediante su función de probabilidad ($p(x)$) y una tabla con los valores y sus probabilidades.

x	$p(x)$
2	1/36
3	7/36
4	3/36
⋮	⋮

← Lanzar 2 dados prob sacar el valor 2 o 3 o 4

$p(x)$ Función de probabilidad, de cuantía o de masa de una variable aleatoria:

Es la función $p(x)$ de una variable discreta X que asigna a cada valor diferente de X: $x_1, x_2, \dots, x_k$ la probabilidad de ser obtenido en el experimento aleatorio.

$$p(x_0) = P(X = x_0) \qquad \boxed{\sum_E p(x_i) = 1}$$

$F(x)$ Función de distribución $F(x)$:

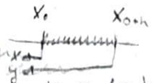
Función de distribución de la variable aleatoria X en el punto $x = x_0$ es la probabilidad de que X tome un valor menor o igual que x_0. Equivale a la frec. relativa acum.

$$F(x_0) = P(X \leq x_0) \qquad F(x_0) = P(-\infty < x \leq x_0) = P(-\infty, x_0]$$

Propiedades:

1) $F(+\infty) = 1 \quad F(-\infty) = 0$

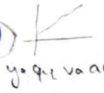
2) $P(x_0, x_{0+h}) = F(x_{0+h}) - F(x_0)$



$y - x =$ Lo del medio

3) Es una función monótona no creciente: $F(x_0) \leq F(x_{0+h})$

El anterior ≤ el proximo (ya que va acumulando hasta 1)



Variables aleatorias discretas: Toma un número de valores cuantitativos discretos.

$p(x)$ y $F(x)$ $F(x_0) = p(X \leq x_0) = \sum_{x_i \leq x_0} p(X = x_i)$

En una distribución discreta, la probabilidad se concentra en los puntos de discontinuidad x_i

Variable aleatoria continua: Puede tomar cualquier valor en un intervalo.

$$p(X = x_i) = 0 \qquad p(X \leq x_i) = p(X < x_i)$$

La probabilidad de cada punto concreto es nula.

$f(x)$ **Función de densidad:** Función que describe la densidad de probabilidad en cualquier intervalo. Haciendo las clases del histograma cada vez más pequeñas, éste tenderá a una curva $f(x)$, capaz de describir el comportamiento de la variable.

$$\boxed{f(x) = \frac{P(x_0 < X < x_0 + \Delta x)}{\Delta x}}$$

La probabilidad de cualquier intervalo vendrá dada por el área que $f(x)$ encierra en este intervalo.

$$\boxed{f(x) = F'(x)}$$

Tomando un intervalo tan pequeño como queramos ($\Delta x \to 0$)

Propiedades:

1) $f(x) \geq 0 \quad \forall x \in D_x$ — Dominio de definición — Hay aunque baja dentro de la función
2) $f(x) = 0 \quad \forall x \notin D_x$ — Fuera de la función, no hay probabilidad
3) $P(x \leq x_0) = \int_{-\infty}^{x_0} f(x)\,dx$
4) $\boxed{\int_{-\infty}^{\infty} f(x)\,dx = [F(x)]_{-\infty}^{\infty} = 1}$
5) $\boxed{P(a < x \leq b) = F(b) - F(a) = \int_a^b f(x)\,dx}$

Medidas de tendencia central.

Media o Esperanza matemática (μ ... $E(x)$)

Caso discreto: $\boxed{\mu = E(x) = \sum_i x_i P(x_i)}$

Caso continuo: $\boxed{\mu = E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx}$

Mediana:

Caso discreto: el m más pequeño que satisfaga $\boxed{F(m) \geq 0'5}$

Caso continuo: el m tal que $\boxed{F(m) = 0'5}$

Moda: Es el valor de mayor probabilidad o densidad.

Medidas de dispersión:

Varianza (σ^2, Var (x))

Caso discreto: $\boxed{\sigma^2 = Var(x) = \sum_{i=1}^{n} (x_i - \mu)^2 P(x)}$

Caso continua: $\boxed{\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\,dx}$

Equivale a $E(X-\mu)^2 = E(x^2) - \mu^2$

$$\boxed{\sigma^2 = E(x^2) - E(x)^2}$$

Percentil: p

Es el valor x_p que verifica: $F(x_p) = p$

$\mu(xy) - \mu(x)\mu(y)$

Covarianza: cov(x,y)

$cov(x,y) = E(xy) - E(y) \cdot E(x)$

La misma interpretación que las muestrales.

$$cov(x,y) = E\{[X - E[x]][Y - E[x]]\}$$

Correlación: $\rho = corr(x,y)$

La misma interpretación que las muestrales.

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

Transformaciones y medidas características.

En el caso de la media: $E(a+bx) = a + bE[x]$

$E(aX+bY) = aE(X) + bE(Y)$

En el caso de la varianza: $var(a+bx) = 0 + b^2 var(x)$

X e Y correladas: $var(aX+bY) = a^2 var(x) + b^2 var(Y) + 2ab\,cov(X,Y)$

Si no están correladas: $var(aX+bY) = a^2 var(x) + b^2 var(Y) = var(aX-bY)$

incorrelada · Hay independencia lo hay, pero si hay incorrelación no quiere decir que haya independencia

independencia ⇒ incorrelado

$var(aX \mp bY) = a^2 var(x) + b\,var(Y) \pm 2ab\,cov(X,Y)$

sólo

independencia ⇒ incorrelada

sólo →

11/10/19

Tema 5: Modelos probabilísticos.

de probabilidad

Para calcular probabilidades o medidas características.

El proceso der Bernoulli

Fenómeno aleatorio dicotómico: Que tiene 2 posibilidades, exito o fracaso.

La proporción es constante en la población: p: prob. de exito, $q = p-1$: prob. fracaso

Las observaciones son independientes entre sí.

Ley binomial (1,0) o de Bernoulli.

X: Exito → p / Fracaso → 1-p

Función de probabilidad: $P(x) = p^x q^{1-x}$; $x = 0,1$ $P(x=1) = p$ $P(x=0) = q$

Esperanza: $E(x) = p$

Varianza: $var(x) = p \cdot q$ Será maxima para $p = 0'5$

Distribución binomial. $B = (n,p)$

$X: \{0,1,2,\dots,n\}$

Modeliza una serie de fenómenos dicotómicos independientes entre sí.

(nº de veces que aparece x en Y repeticiones independientes con $P(x) = p$)

p = exito de x

Función de probabilidad, $P(x=r)$

Aparecen r veces en n · total · los que no nos interesan

$$P(X=r) = \binom{n}{r} p^r (1-p)^{n-r} \quad r = 0,1,\dots,n$$

nº que nos interesa.

$$C_n^r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

Esperanza:

$$E(x) = E\left(\sum_{j=1}^{n} x_j\right) = np = E(x)$$

Varianza:

$$Var(x) = npq = \sum_{j=1}^{n} var(x_j)$$

Desviación típica: $\sqrt{npq}$

$P(\text{zurdo}) = 0'1$

Prob. de haber más de 3 zurdos en 20 personas

$P(X>3) = 1 - P(X \leq 3) =$

$1 - [P(x=0) + P(x=1) + P(x=2) + P(x=3)]$

$1 - [\dots] = 0'13295$

$P(x=0) = \binom{20}{0} 0'1^0\, 0'9^{20} = 0'12158$

$P(x=2) = \binom{20}{2} 0'1^2\, 0'9^{18} = \dots$

Estadística T6 11/10/19

El proceso de Poisson: Contabiliza el nº de sucesos en un soporte continuo (tiempo, distancia)

Aparición de sucesos puntuales sobre un soporte continuo, suponiendo que el proceso generador de estos sucesos:
- es estable
- produce sucesos independientes.

$\mathbb{P}(\lambda)$ **Distribución de Poisson:** Modeliza la aparición de cierto número de sucesos sobre un soporte continuo en un intervalo de longitud fija.

Nº de personas en un tiempo. Unidades/tiempo

X: nº de sucesos en un intervalo de longitud T fija.

Obsevamos si aparece o no el suceso estudiado en cada segmento.

Es una distribución binomial en la que:

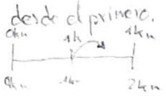
$X = \{0, 1, \ldots, \infty\}$

$n \to \infty$ — p pequeña ⇒ Probabilidad despreciable de aparición de 2 o más sucesos en uno de los n segmentos

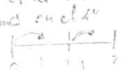
$p \to 0$

$np \to \lambda$ — parámetro. $\lambda = 5 \text{ hojas}/\text{día} = 5 \frac{\text{hojas}}{\text{día}} \cdot \frac{\text{día}}{24\text{hora}} \cdot \frac{5}{24} \text{ hojas}/\text{hora}$

Falta de memoria: Pasada una unidad del soporte continuo, es como si calculasemos la probabilidad desde el primero.



La probabilidad de accidente en el 2º kilometro es la misma que si calculásemos en el 1º



15/10/19

Características:

Función de probabilidad: $\boxed{P(X=r) = \frac{\lambda^r}{r!} e^{-\lambda} \quad r = 0, 1, 2, \ldots}$

Esperanza: $\boxed{E(x) = \lambda}$

Varianza: $\boxed{var(x) = \lambda}$

Es una distribución asimetrica, que tiende a la simetría al aumentar λ.

La suma de varias variables de Poisson independientes, tambien es una variable de Poisson. Sea $X_i \sim P(\lambda_i)$, $i = 1, \ldots, k$, un conjunto de variables de Poisson independientes. $Y = \sum_{i=1}^{k} X_i \to P(\lambda^*)$ $\quad \lambda^* = \sum_{i=1}^{k} \lambda_i$

Distribución exponencial ($Exp(\lambda)$) (no necesariamente)**:** Modela el tiempo entre la ocurrencia de dos sucesos consecutivos, siendo estos independientes y estables. Es una distribución continua.

Tiempo entre sucesos con un λ = sucesos/tiempo

Características:

Función de distribución: $F(t) = P(T \leq t) = 1 - e^{-\lambda t}$ $\quad f(t) = \lambda e^{-\lambda t}$

Esperanza: $E(T) = 1/\lambda$

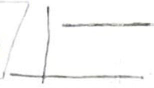
Varianza: $var(T) = 1/\lambda^2$

Si $\lambda = 7 \text{ accesos}/\text{min}$ $\quad E(T) = \frac{1}{7} = 0'143 \text{ minutos}/\text{acceso}$

Más de 15 seg entre accesos, probabilidad. $P(T > \underset{0'25\,min}{15s}) = 1 - P(T < 0'25) = 1 - (1 - e^{-\frac{1}{7} \cdot 0'25}) = 0'17$

Distribución uniforme: Es una distribución continua.

Características:

Función de densidad: $\boxed{f(x)=\frac{1}{b-a} \quad a\leq x\leq b}$ 

Función de distribución: $\boxed{F(x)=P(X\leq x)=\int_a^x \frac{1}{b-a}\,dx=\frac{x-a}{b-a} \quad a\leq x\leq b}$ 

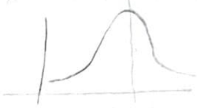
Esperanza: $\boxed{E(x)=\frac{a+b}{2}}$

Varianza: $\boxed{Var(x)=\frac{(b-a)^2}{12}}$

$N(\mu,\sigma)$

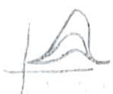
Distribución normal: Es la distribucion con función de densidad:

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



Depende de dos parámetros:

Media μ 

Desviación Típica σ 

$\left.\begin{matrix} N(\mu,\sigma) \\ N(\mu,\sigma^2) \end{matrix}\right\} N(3,5^2)$

Es uno de los modelos más frecuentes para describir variables reales continuas.

Simétrica, centrada en la media μ que es su mediana y su moda.

Forma de campana.

Coeficiente de apuntamiento igual a 3.

Se ajusta a lo observado en muchos procesos de medición, si no influyen los errores sistemáticos.

La normal de media 0 y desviación típica 1 se denomina:

Normal tipificada (Z)

Normal estándar

Normal (0,1) y su función está tabulada (en una tabla)

$$P(a<x<b)=P\left(\frac{a-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{b-\mu}{\sigma}\right) \rightarrow \text{Tipificar } Z\rightarrow N(0,1)$$

Distribución lognormal: Se llama lognormal a la variable aleatoria cuyo logaritmo neperiano es normal:

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

$$E(x)=e^{\mu+\frac{1}{2}\sigma^2} \qquad Var(x)=(e^{\sigma^2}-1)\,e^{2\mu+\sigma^2}$$

Si la variable y puede considerarse como un producto de variables aleatorias:

$y=x_1\cdot x_2\cdot x_3\cdot\ldots\cdot x_n$ su logaritmo neperiano seguirá una distribución normal.

Teorema central del Límite. ("Todo sigue una distrib. normal")

La suma de un conjunto de variables aleatorias se aproxima, al aumentar el número de variables, a una variable aleatoria normal independientemente de cual sea la distribución de esas variables.

Sea $x_1, x_2, \dots, x_n$ variables aleatorias independientes con media μ_i, desviación típica σ_i y distribución CUALQUIERA.

La desv. típica de Y es la raíz de la suma de varianzas. No la suma directa de σ

$Y = x_1 + x_2 + \dots + x_n$ Al crecer n: $\dfrac{Y - \sum \mu_i}{\sqrt{\sum \sigma_i^2}} \longrightarrow N(0,1)$ $\boxed{Y \to N\left(\sum \mu, \sqrt{\sum \sigma_i^2}\right)}$

Aplicaciones:

Distribución binomial → Normal:

$E(x_i) = p$ $Var(x_i) = pq$ $\boxed{Y \to N(\overset{\text{todos } \Sigma}{np}, \sqrt{npq})}$

Sirve para $n > 30$ y $npq > 5$

Distribución Poisson → Normal:

Ojo tipificar en la Normal para mirar la tabla. $P\left(Z < \dfrac{x - E(x)}{\sigma}\right)$ y es $P(Z < x)$

Sea $Y(0,T)$ una variable de Poisson que cuenta el nº de sucesos en el intervalo $(0,T)$. $Y(0,T) = x_1(0,t_1) + x_2(t_1,t_2) + \dots + x_n(t_{n-1},T)$

$\boxed{\mathbb{P}(\lambda) \to N(\lambda, \sqrt{\lambda})}$ $E(x_i) = \lambda$ $Var(x_i) = \lambda$

Para $\lambda > 5$ podemos aproximarla a una normal.

Distribución binomial → La Poisson.

La aproximación es valida para: $np > 1$ $p < 0{,}1 \Rightarrow \lambda = np$

Modelo de regresión simple (lineal):

La recta de regresión de y sobre x es de la forma: $\underset{\text{Observado en x/y para el i-ésimo}}{y_i} = a + bx_i + \underset{\text{error}}{e}$

Si queremos definir un tipo de relación válido para toda la población, encontramos numerosos factores que no controlamos.

Si fijamos el valor de $\underset{a+bx_i}{X = x_i}$ fijo obtendremos valores diferentes debido a $\underset{\text{variable}}{e}$ (ruido)

Si asumimos que todas las variables que influyen en Y lo hacen de forma lineal (aditiva):

$Y = a + bx_i + \underbrace{(c_1 z_1 + c_2 z_2 + \dots)}_{} \to e$ (error, ruido)

Por el teorema del límite central, la e (ruido, error) seguirá una distribución normal: $E(e) = 0$ Suponemos que se compensan unos con otros y la media es 0

Campana Gauss de la Normal

$var(e) = \sigma^2$ es constante. el ruido es homogeneo a lo largo de la recta

$\boxed{e \to N(0, \sigma^2)}$ $E(Y / \underset{\text{fija}}{X = x_i}) = E(\underset{\text{cte}}{a + bx_i} + \underset{\text{variable}}{e}) = a + b \cdot x_i + \underset{0}{E(e)} = a + bx_i$

$\boxed{Y \to N(a + bx_i, \underset{var(e)}{\sigma^2})}$ $var(Y / X = x_i) = var(a + bx_i + e) = 0 + var(e) = \sigma^2$

Cada punto y_i que observamos se interpreta como un valor al azar de la normal.

28/10/2019

Tema 6: Introducción a la inferencia estadística.

Proceso de inducción por el cual a partir de una muestra intentamos predecir cómo será el resto de la población que no se ha observado (variable aleatoria)

Muestra aleatoria simple: Conjunto de elementos de la población.

Elementos de la muestra independientes entre sí.

Elementos con las mismas características de la población

Al ser una variable aleatoria, cada muestra será distinta.

Distribución muestral de estimadores.

Cualquier estimador podrá tomar diferentes valores, cada uno con diferentes probabilidades dependiendo de la muestra.

Distribución muestral del estadístico

Cualquier función de los valores muestrales

Dependerá de: Tamaño de la muestra (n) / Distribución de la población.

Distribución muestral de la media.

Sea una variable aleatoria cualquiera, de tamaño n de media μ y desviación típica σ:

$\boxed{E(\bar{X}) = \frac{\cancel{n}\mu}{\cancel{n}} = \mu}$ $\quad E(\mu) = \mu$ ↳ Media poblacional

$E(\bar{X}) = \frac{E(\bar{X}_1) + E(X_1) + \dots}{n}$

Ya que en cada hueco de la muestra puede estar cualquier elemento $\Rightarrow E(X_1) = \mu$

$\boxed{Var(\bar{X}) = \frac{\cancel{n}\sigma^2}{n^{\cancel{2}}} = \frac{\sigma^2}{n}}$ $\quad \bar{X}$: Media muestral

$Var(\bar{X}) = E\left(\frac{\bar{X}_1 + X_1 \dots}{n}\right) = \frac{E(X_1) + E(\dots)}{n^2}$

al sacarlo a^2

Cuando (n>30) n es grande, la distribución de la media es asintoticamente normal, por el Teorema Central del Limite

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

Pero para cualquier n, si x es $N(\mu, \sigma)$

Media $= \mu$ Desviación típica $= \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

Estimación: 10/11/2019

Los parametros $(\mu, \sigma, \lambda, \dots)$ son valores numericos de la población (constantes)

Estimador es un estadístico que nos da con cierta exactitud el valor de los caracteres de la población inferir. Nomenclatura: ^

Ejem: $\hat{\mu}$ estimador de la media poblacional.

Se prefieren en general los que cumplen: $E(\hat{\theta}) = \theta$ ⇐ (Estimador insesgado o centrado)

$\boxed{Sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta}$ Diferencia entre la esperanza del estimador y el valor. ↳ propio del parametro.

Error estandar: La desviación típica de un estimador. $var(\hat{\theta}) = \frac{\sigma^2}{n}$ $\boxed{e(\hat{\theta}) = \frac{\sigma}{\sqrt{n}}}$

Error cuadratico medio (ECM) $\Rightarrow$ $\boxed{ECM(\hat{\theta}) = E(\hat{\theta} - \theta)^2}$ Mide el error sistematico de un estimador

Puede demostrarse: $\boxed{ECM(\hat{\theta}) = [sesgo(\hat{\theta})]^2 + Var(\hat{\theta})}$ ← Importante

Sesgo: Centrados entorno al original

ECM minimo $\Rightarrow$ sesgo minimo → insesgo

Varianza minima → eficiencia o precisión.

Sesgo positivo $\Rightarrow$ El estimador sobreestima el valor del parametro.

Sesgo negativo $\Rightarrow$ El estimador subestima el valor del parametro.

Eficiencia ó Precisión $\Rightarrow$ $\boxed{efic(\hat{\theta}) = \frac{1}{Var(\hat{\theta})}}$ Estan juntos los valores, pero no tienen por qué estar centrados.

Método de los momentos:

Metodo sencillo de construcción de estimadores.

Consiste en estimar una caracteristica poblacional con la relativa caracteristica muestral.

$\hat{\mu} = \bar{x}$ $\quad \sigma^2 = S^2$. Poblacional = muestral.

Diagnosis y crítica del modelo: Dados unos datos saber a que distribución se ajustan. modelo.

Test de la chi-cuadrado (χ^2): Consiste en comparar las frecuencias observadas con las esperadas.

Procedimiento: Muestra de tamaño $n \geq 25$

Se agrupan en $k \geq 5$ clases de almenos 3 datos cada una.

Calculamos la discrepancia entre O_i y E_i.

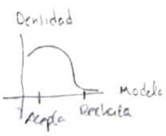
Estadístico chi-cuadrado:

$$\boxed{\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}}$$

$O_i \Rightarrow$ Frecuencias observadas de cada clase

$E_i \Rightarrow$ Frecuencias esperadas segun el modelo.

Resume la discrepancia entre datos y modelo:



Discrepancia alta: Rechazamos el modelo. Muy poco probable si el modelo es correcto.

Discrepancia baja: Aceptamos el modelo. La más probable de ocurrir.

Depende de los grados de libertad g. Hay k-1, si hay k intervalos.

p-valor: La probabilidad de obtener mi chicuadrado si el modelo es correcto. Area que queda a la derecha de χ_0^2 en la distribución.

Se rechaza un modelo si el p-valor < 0'05

Transformaciones para mejorar la normalidad:

Datos con asimetría positiva: La típica es $y = \ln(x)$

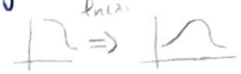
$y = 1/x$ — Más potente

$y = \ln(x)$

$y = \sqrt{x}$ — Menos potente

Comprimen la escala en los valores altos y la expanden en los bajos.

En general son: $y = e^c$ / $c<1$



Datos con asimetría negativa: La típica es $y = e^2$

Comprimen la escala en los ~~b~~ valores bajos y la expande en los valores altos. En general: $y = e^c$ / $c>1$

11/11/19

Tema 7: Inferencia con muestras grandes.

Intervalo de confianza: Para el parametro θ con nivel de confianza $(1-\alpha)$ es el intervalo $[\theta_1(x), \theta_2(x)]$ tal que:

$P[\theta_1(x) \leq \theta \leq \theta_2(x)]$ [Indica que si calculase 100 intervalos distintos, con muestras distintas, puedo asegurar que el $\underbrace{1-\alpha}_{\text{en }\%}$ está en el intervalo.]

No indica la probabilidad de que el θ esté en el intervalo.

[Dado un α, o nivel de significación se trata de encontrar un intervalo centrado que contenga su verdadero valor el $(1-\alpha)100\%$ de las veces.]

Intervalo de confianza para la media con:

Varianza (σ^2) poblacional conocida:

$\boxed{\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$ Por teorema central del límite. Tipificado $\boxed{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}} \to N(0,1) = Z$

Definimos $Z_{\alpha/2}$ como el valor Z tal que: $\boxed{P(Z > Z_{\alpha/2}) = \frac{\alpha}{2}}$ $\boxed{P(Z < -Z_{\alpha/2}) = \frac{\alpha}{2}}$

$P(-Z_{\alpha/2} < Z < Z_{\alpha/2})$

$\boxed{\mu \in \left(\bar{X} \pm Z_{\alpha/2} \cdot \sigma/\sqrt{n}\right)}$ $P\left(-Z_{\alpha/2} < \underbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}_{\sim Z} < Z_{\alpha/2}\right) = 1-\alpha$

Se pasa de un lado a otro

Se busca en la tabla una u otra depende ~~el valor~~ la tabla. Se busca el $\alpha/2$

Varianza (σ^2) poblacional desconocida:

σ^2 suele ser desconocida, con lo que la sustituimos por una estimación. $\hat{\sigma}^2 = S^2$

Pero es un estimador sesgado, pudiendose demostrar que: $\boxed{E(S^2) = \sigma^2 \frac{n-1}{n}}$

$\text{Sesgo}(S^2) = -\frac{\sigma^2}{n}$ Negativo $\Rightarrow S^2$ subestima la verdadera varianza.

Para corregir el sesgo:

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

Estadística T9 12/11/2019

Para saber si se acepta la hipótesis con intervalos tiene que ser para bilateral

Para unilateral no se puede mirar el intervalo de confianza

Si la muestra, n, es grande, el intervalo de confianza para la media poblacional μ es: $\boxed{\mu \in \left(\bar{x} \pm z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}\right)}$ $\hat{s}^2$ es la cuasivarianza.

Si la muestra no es grande, pero la población es normal:

$$\boxed{\mu \in \left(\bar{x} \pm t_{n-1,\alpha/2} \cdot \hat{s}/\sqrt{n}\right)}$$

$t = t$ de Student.

T de Student

$$\boxed{t_n = \frac{z}{\sqrt{\frac{1}{n}\chi^2_n}}}$$ Con $z \to N(0,1)$

Es una variable simétrica, con mayor dispersión que la normal estandar, a la que tiende rapidamente al aumentar n.

Intervalos de confianza para la varianza de poblaciones normales:

Lema de Fisher-Cochran $\frac{\hat{s}}{\sigma^2} \to \frac{\chi^2_{n-1}}{n-1} \rightsquigarrow \boxed{\frac{ns^2}{\sigma^2} \to \chi^2_{n-1}}$

Llamando χ^2_a y χ^2_b a los valores χ^2_{n-1} que dejan entre sí el $1-\alpha$ de la distribución:

$P(\chi^2_a \leq \frac{ns^2}{\sigma^2} \leq \chi^2_b)$ $\quad \frac{(n-1)\hat{s}^2}{\chi^2_a} \geq \sigma^2 \geq \frac{(n-1)\hat{s}^2}{\chi^2_b}$

$$\boxed{\sigma^2 \in \left[\frac{(n-1)\hat{s}^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)\hat{s}^2}{\chi^2_{n-1,1-\alpha/2}}\right]}$$

ojo

Intervalos de confianza para una proporción: (Bernuilli?)

$\bar{x} = p \quad \hat{s} = \hat{p}\hat{q} \quad \bar{x} \pm z_{\alpha/2} \cdot \frac{\hat{s}}{\sqrt{n}} \rightsquigarrow \sqrt{\hat{s}^2}$ $\quad \boxed{p \in \left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)}$

Se desea estimar la proporción p de elementos de la población que poseen un atributo determinado.

Intervalos de confianza para la λ de Poisson.

Sean $X_1, X_2, X_3 \ldots$ m.a.s de una distribución de Poisson.

$\frac{\bar{X}-\lambda}{\sqrt{s/n}} \to N(0,1) \quad \lambda = Var(x) = E(x) \quad P\left(-z_{\alpha/2} \leq \frac{\bar{x}-\lambda}{\sqrt{s/n}} \leq z_{\alpha/2}\right) = 1-\alpha$

Aprender → $\boxed{\lambda \in \left(\bar{x} \pm z_{\alpha/2} \cdot \sqrt{\frac{\bar{x}}{n}}\right)}$ ¿$\bar{x} = \lambda$?

20/11/2019

Determinación del tamaño muestral:

Función de la precisión que se quiera conseguir.

a) Tamaño muestral para la estimación de una proporción:

$L = Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$ Como p es desconocido tomamos $p = 0'5$ y $q = 0'5$ $\boxed{n = \frac{Z^2_{\alpha/2}}{4L^2}}$

$p \pm L$

b) Tamaño muestral para la estimación de la media:

$L = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \Rightarrow \boxed{n = \frac{Z^2_{\alpha/2} \cdot \sigma^2}{L^2}}$

$\bar{x} \pm L$

Contraste de hipotesis: Trata de comparar la hipotesis con los hechos.

Si la hipotesis coincide con los hechos $\Rightarrow$ Aceptamos la hipotesis.

Si no coincide con los hechos $\Rightarrow$ Rechazamos la hipotesis.

Se comparan 2 alternativas:

H_0: Hipótesis nula o "neutra" La más sencilla

La que mantendremos a no ser que los datos indiquen su falsedad.

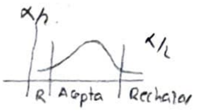
H_1: Hipotesis alternativa.

La que aceptamos si los datos parecen incompatibles con la hipotesis nula.

$H_0: \theta = \theta_0$ ⇝ Debe tener el signo igual

$H_1: \theta \neq \theta_0$ ⇝ Lo que nos preguntamos

} Contraste bilateral.

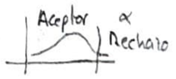


$H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$ ó

$H_1: \theta < \theta_0$

} Contrastes unilaterales.



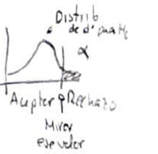
La hipotesis nula siempre es verdadera si no se demuestra lo contrario.

Cuándo una discrepancia se considera demasiado grande como para atribuirla al azar (rechazar H_0)

Se consideran discrepancias "demasiado grandes" aquellas que tienen una probabilidad "demasiado pequeña" de ocurrir si H_0 es cierta.

Si la discrepancia observada en la muestra, $\hat{d}$ cae dentro de la región de rechazo, rechazamos H_0. En caso contrario, la aceptaremos.

α = El 5%



Se pueden cometer dos tipo de errores:

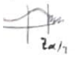
estadístico de contraste Z_0

Tipo 1: Rechazar H_0 siendo cierta. La probabilidad es $\alpha = P\left(\frac{\text{rechazar } H_0}{H_0 \text{ cierta}}\right)$

p-valor

Tipo 2: Aceptar H_0 siendo falsa.

p-valor = $P(Z > Z_0)$ ó $P(Z > Z_0)$ ó $P(T > T_0)$



Para determinar la región de rechazo, empleamos en lugar del nivel de significación α el nivel crítico p (p-value)

En unilateral

En bilateral

Estadístico de contraste → La Tabla

$2P(Z > Z_0)$

$2P(Z > Z_0)$

Es el mínimo nivel de significación que nos llevaría a rechazar la hipotesis nula.

$P_{valor} < 0'05$ (0'01): Existe muy poca evidencia en la muestra a favor de la hipotesis $\Rightarrow$ RECHAZAMOS.

↑ Discrepancia $\Rightarrow$ ↓ Probabilidad

Buscar en la borde tabla

* → α es la mínima area que debe tener el p-valor para ser aceptada la hipoteris.

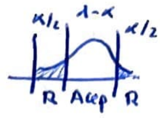
* → p-valor es el área hasta el area de rechazo (α), no es una probabilidad pero se calcula como una $P[\chi^2(k-v-1) > X_0^2]$ $P(Z > Z_0)$

unilat / bilateral ↳ estadístico de calculo

Contraste para la media.

Se desea contrastar la hipotesis de que la media de una distribución normal es μ_0

$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$
Contraste bilateral



Para cualquier variable X de media μ y varianza σ^2, si el tamaño de la muestra es suficientemente grande se cumple que:

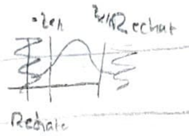
$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \to N(0,1) \quad \text{y además: } T = \frac{\bar{X} - \mu}{\hat{S} / \sqrt{n}} \to N(0,1)$$

Z se utiliza con σ y T se usa con $\hat{S}$

La T y Z pueden seguir una normal o una t de Student

Rechazamos H_0 si:

$$\frac{\bar{X} - \mu}{\hat{S}/\sqrt{n}} \begin{cases} > Z_{\alpha/2} & > Z_\alpha \\ < -Z_{\alpha/2} & < -Z_\alpha \text{ Rechazo} \end{cases}$$



$\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ Unilateral

$\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$ Unilateral.

Un intervalo de confianza con nivel $1-\alpha$ y un contraste usan la misma información. Hacer un contraste bilateral con nivel de significación α es equivalente a hacer un intervalo de confianza de nivel $(1-\alpha)$

Distribución para una regresión multiple específica $Y \sim N($valor para esa regr. $Y_{0,0,0}$, varianza residual $S_R^2)$

Parte II

Recursos

FORMULARIO DE ESTADÍSTICA

MODELOS DE PROBABILIDAD

Nombre	Símbolo	$p(x) = P(X = x)$; $F(x) = P(X \leq x)$; $f(x) = \dfrac{\partial F(x)}{\partial x}$	Media μ	Varianza σ^2
Bernoulli	$B(p)$	$p(x) = p^x q^{1-x}; x = 0, 1$	p	pq
Binomial	$B(n, p)$	$p(x) = \binom{n}{x} p^x q^{n-x}; x = 0, 1, ..., n$	np	npq
Geométrica	$G(p)$	$p(x) = pq^{x-1}; x = 1, 2, ...$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$
Poisson	$P(\lambda)$	$p(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}; x = 0, 1, ...$	λ	λ
Uniforme continua en (a, b)	$U(a, b)$	$f(x) = 1/(b-a); a \leq x \leq b$	$\dfrac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponencial	$\text{Exp}(\lambda)$	$f(x) = \lambda e^{-\lambda x}; x > 0$ $F(x) = 1 - e^{-\lambda x}; x > 0$	$1/\lambda$	$1/\lambda^2$
Normal	$N(\mu, \sigma^2)$	$f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}; x \in \Re$	μ	σ^2

REGRESIÓN SIMPLE

$$\hat{y} = a + bx$$

$$a = \bar{y} - b\bar{x}$$
$$b = \frac{s_{x,y}}{s_x^2} = \frac{\hat{s}_{x,y}}{\hat{s}_x^2}$$

INFERENCIA PARA UNA POBLACIÓN

Población	Contraste de hipótesis	Estadístico de contraste	Región de rechazo (p-valor $< \alpha$)	Intervalo de confianza $\text{IC}_{1-\alpha}$
Cualquier v.a. X con $E[X] = \mu$, $\text{Var}[X] = \sigma^2$ y $n \to \infty$	(1) $\text{H}_0 : \mu = \mu_0$; $\text{H}_1 : \mu \neq \mu_0$ (2) $\text{H}_0 : \mu \geq \mu_0$; $\text{H}_1 : \mu < \mu_0$ (3) $\text{H}_0 : \mu \leq \mu_0$; $\text{H}_1 : \mu > \mu_0$	$(a)\ z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ $(b)\ t_0 = \dfrac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}$	(1a) $\lvert z_0\rvert > z_{\alpha/2}$ (1b) $\lvert t_0\rvert > z_{\alpha/2}$ (2a) $z_0 < -z_\alpha$ (2b) $t_0 < -z_\alpha$ (3a) $z_0 > z_\alpha$ (3b) $t_0 > z_\alpha$	$\mu \in \left(\bar{x} \mp z_{\alpha/2}\sigma/\sqrt{n}\right)$ $\mu \in \left(\bar{x} \mp z_{\alpha/2}\hat{s}/\sqrt{n}\right)$
Normal $N(\mu, \sigma^2)$	(1)$\text{H}_0 : \mu = \mu_0$; $\text{H}_1 : \mu \neq \mu_0$ (2)$\text{H}_0 : \mu \geq \mu_0$; $\text{H}_1 : \mu < \mu_0$ (3)$\text{H}_0 : \mu \leq \mu_0$; $\text{H}_1 : \mu > \mu_0$	$(a)\ z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ $(b)\ t_0 = \dfrac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}$	(1a) $\lvert z_0\rvert > z_{\alpha/2}$ (1b) $\lvert t_0\rvert > t_{n-1;\alpha/2}$ (2a) $z_0 < -z_\alpha$ (2b) $t_0 < -t_{n-1;\alpha}$ (3a) $z_0 > z_\alpha$ (3b) $t_0 > t_{n-1;\alpha}$	$\mu \in \left(\bar{x} \mp z_{\alpha/2}\sigma/\sqrt{n}\right)$ $\mu \in \left(\bar{x} \mp t_{n-1;\alpha/2}\hat{s}/\sqrt{n}\right)$
Bernoulli $B(p)$ con $n \to \infty$	(1) $\text{H}_0 : p = p_0$; $\text{H}_1 : p \neq p_0$ (2) $\text{H}_0 : p \geq p_0$; $\text{H}_1 : p < p_0$ (3) $\text{H}_0 : p \leq p_0$; $\text{H}_1 : p > p_0$	$z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0 q_0/n}}$	(1) $\lvert z_0\rvert > z_{\alpha/2}$ (2) $z_0 < -z_\alpha$ (3) $z_0 > z_\alpha$	$p \in \left(\hat{p} \mp z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}}\right)$
Normal $N(\mu, \sigma^2)$	(1) $\text{H}_0 : \sigma^2 = \sigma_0^2$; $\text{H}_1 : \sigma^2 \neq \sigma_0^2$ (2) $\text{H}_0 : \sigma^2 \geq \sigma_0^2$; $\text{H}_1 : \sigma^2 < \sigma_0^2$ (3) $\text{H}_0 : \sigma^2 \leq \sigma_0^2$; $\text{H}_1 : \sigma^2 > \sigma_0^2$	$\chi_0^2 = \dfrac{(n-1)\hat{s}^2}{\sigma_0^2} = \dfrac{ns^2}{\sigma_0^2}$	(1) $\chi_0^2 > \chi_{n-1;\alpha/2}^2$ ó $\chi_0^2 < \chi_{n-1;1-\alpha/2}^2$ (2) $\chi_0^2 < \chi_{n-1;1-\alpha}^2$ (3) $\chi_0^2 > \chi_{n-1;\alpha}^2$	$\sigma^2 \in \left(\dfrac{(n-1)\hat{s}^2}{\chi_{n-1;\alpha/2}^2}; \dfrac{(n-1)\hat{s}^2}{\chi_{n-1;1-\alpha/2}^2}\right)$
Cualquier v.a. X con $\widehat{\theta}_{MV}$ y $n \to \infty$	(1) $\text{H}_0 : \theta = \theta_0$; $\text{H}_1 : \theta \neq \theta_0$ (2) $\text{H}_0 : \theta \geq \theta_0$; $\text{H}_1 : \theta < \theta_0$ (3) $\text{H}_0 : \theta \leq \theta_0$; $\text{H}_1 : \theta > \theta_0$	$t_0 = \dfrac{\hat{\theta}_{MV} - \theta_0}{\sqrt{\widehat{\text{Var}}(\hat{\theta}_{MV})}}$ $\widehat{\text{Var}}(\hat{\theta}_{MV}) = -\left(\dfrac{\partial^2 L(\hat{\theta}_{MV})}{\partial\theta^2}\right)^{-1}$	(1) $\lvert t_0\rvert > z_{\alpha/2}$ (2) $t_0 < -z_\alpha$ (3) $t_0 > z_\alpha$	$\theta \in \left(\hat{\theta}_{MV} \mp z_{\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{\theta}_{MV})}\right)$

INFERENCIA PARA DOS POBLACIONES

Dos poblaciones o v.a. X_1, X_2	Contraste de hipótesis	Estadístico de contraste	Región de rechazo (p-valor $< \alpha$)
Cualesquiera con $E(X_1) = \mu_1,\ E(X_2) = \mu_2$ $\mathrm{Var}(X_1) = \sigma_1^2$ $\mathrm{Var}(X_2) = \sigma_2^2$	(1) $\mathrm{H}_0 : \mu_1 = \mu_2$; $\mathrm{H}_1 : \mu_1 \neq \mu_2$ (2) $\mathrm{H}_0 : \mu_1 \geq \mu_2$; $\mathrm{H}_1 : \mu_1 < \mu_2$ (3) $\mathrm{H}_0 : \mu_1 \leq \mu_2$; $\mathrm{H}_1 : \mu_1 > \mu_2$	$(a)\ z_0 = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ $(b)\ t_0 = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}}$	(1a) $\lvert z_0 \rvert > z_{\alpha/2}$ (1b) $\lvert t_0 \rvert > z_{\alpha/2}$ (1b) $\lvert t_0 \rvert > t_{\nu;\alpha/2}$ (2a) $z_0 < -z_\alpha$ (2b) $t_0 < -z_\alpha$ (2b) $t_0 < -t_{\nu;\alpha}$ (3a) $z_0 > z_\alpha$ (3b) $t_0 > z_\alpha$ (3b) $t_0 > t_{\nu;\alpha}$ Bajo normalidad o si $n_1, n_2 \to \infty$ — Si $n_1, n_2 \to \infty$ — Bajo normalidad
Datos pareados $D = X_1 - X_2$	(1) $\mathrm{H}_0 : \mu_D = 0$; $\mathrm{H}_1 : \mu_D \neq 0$ (2) $\mathrm{H}_0 : \mu_D \geq 0$; $\mathrm{H}_1 : \mu_D < 0$ (3) $\mathrm{H}_0 : \mu_D \leq 0$; $\mathrm{H}_1 : \mu_D > 0$	$(a)\ z_0 = \dfrac{\bar{d}}{\sigma_D \,/\sqrt{n}}$ $(b)\ t_0 = \dfrac{\bar{d}}{\hat{s}_D \,/\sqrt{n}}$	(1a) $\lvert z_0 \rvert > z_{\alpha/2}$ (1b) $\lvert t_0 \rvert > z_{\alpha/2}$ (1b) $\lvert t_0 \rvert > t_{n-1;\alpha/2}$ (2a) $z_0 < -z_\alpha$ (2b) $t_0 < -z_\alpha$ (2b) $t_0 < -t_{n-1;\alpha}$ (3a) $z_0 > z_\alpha$ (3b) $t_0 > z_\alpha$ (3b) $t_0 > t_{n-1;\alpha}$ Bajo normalidad o si $n \to \infty$ — Si $n \to \infty$ — Bajo normalidad
Cualesquiera con $E(X_1) = \mu_1,\ E(X_2) = \mu_2$ $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \sigma^2$	(1) $\mathrm{H}_0 : \mu_1 = \mu_2$; $\mathrm{H}_1 : \mu_1 \neq \mu_2$ (2) $\mathrm{H}_0 : \mu_1 \geq \mu_2$; $\mathrm{H}_1 : \mu_1 < \mu_2$ (3) $\mathrm{H}_0 : \mu_1 \leq \mu_2$; $\mathrm{H}_1 : \mu_1 > \mu_2$	$(a)\ z_0 = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $(b)\ t_0 = \dfrac{\bar{x}_1 - \bar{x}_2}{\hat{s}_T\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ con $\hat{s}_T^2 = \frac{(n_1-1)\hat{s}_1^2 + (n_2-1)\hat{s}_2^2}{n_1+n_2-2}$	(1a) $\lvert z_0 \rvert > z_{\alpha/2}$ (1b) $\lvert t_0 \rvert > z_{\alpha/2}$ (1b)$\lvert t_0 \rvert > t_{n_1+n_2-2;\alpha/2}$ (2a) $z_0 < -z_\alpha$ (2b) $t_0 < -z_\alpha$ (2b) $t_0 < -t_{n_1+n_2-2;\alpha}$ (3a) $z_0 > z_\alpha$ (3b) $t_0 > z_\alpha$ (3b)$t_0 > t_{n_1+n_2-2;\alpha}$ Bajo normalidad o si $n_1, n_2 \to \infty$ — Si $n_1, n_2 \to \infty$ — Bajo normalidad
v.a. de Bernoulli $X_1 \sim B(p_1),\ X_2 \sim B(p_2)$	(1) $\mathrm{H}_0 : p_1 = p_2$; $\mathrm{H}_1 : p_1 \neq p_2$ (2) $\mathrm{H}_0 : p_1 \geq p_2$; $\mathrm{H}_1 : p_1 < p_2$ (3) $\mathrm{H}_0 : p_1 \leq p_2$; $\mathrm{H}_1 : p_1 > p_2$	$z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0\hat{q}_0\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ con $\hat{p}_0 = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$	(1) $\lvert z_0 \rvert > z_{\alpha/2}$ (2) $z_0 < -z_\alpha$ (3) $z_0 > z_\alpha$ Si $n_1, n_2 \to \infty$
v.a. Normales $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$ $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$	(1) $\mathrm{H}_0 : \sigma_1^2 = \sigma_2^2$; $\mathrm{H}_1 : \sigma_1^2 \neq \sigma_2^2$ (2) $\mathrm{H}_0 : \sigma_1^2 \geq \sigma_2^2$; $\mathrm{H}_1 : \sigma_1^2 < \sigma_2^2$ (3) $\mathrm{H}_0 : \sigma_1^2 \leq \sigma_2^2$; $\mathrm{H}_1 : \sigma_1^2 > \sigma_2^2$	$F_0 = \dfrac{\hat{s}_1^2}{\hat{s}_2^2}$	(1) $F_0 > F_{n_1-1;n_2-1;\alpha/2}$ ó $F_0 < F_{n_1-1;n_2-1;1-\alpha/2}$ donde $F_{n_1-1;n_2-1;1-\alpha/2} = 1\big/F_{n_2-1;n_1-1;\alpha/2}$ (2) $F_0 < F_{n_1-1;n_2-1;1-\alpha}$ (3) $F_0 > F_{n_1-1;n_2-1;\alpha}$

Dos poblaciones o v.a. X_1, X_2	Parámetro	Intervalo de confianza $\mathrm{IC}_{1-\alpha}$
Cualesquiera con $E(X_1) = \mu_1,\ E(X_2) = \mu_2$ $\mathrm{Var}(X_1) = \sigma_1^2$ $\mathrm{Var}(X_2) = \sigma_2^2$	$\mu_1 - \mu_2$	$\mu_1 - \mu_2 \in \left(\bar{x}_1 - \bar{x}_2 \mp z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}\right)$ Bajo normalidad o si $n_1, n_2 \to \infty$ $\mu_1 - \mu_2 \in \left(\bar{x}_1 - \bar{x}_2 \mp z_{\alpha/2}\sqrt{\dfrac{\hat{s}_1^2}{n_1} + \dfrac{\hat{s}_2^2}{n_2}}\right)$ Si $n_1, n_2 \to \infty$ $\mu_1 - \mu_2 \in \left(\bar{x}_1 - \bar{x}_2 \mp t_{\nu;\alpha/2}\sqrt{\dfrac{\hat{s}_1^2}{n_1} + \dfrac{\hat{s}_2^2}{n_2}}\right)$ con $\nu \approx \dfrac{\left(\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{\hat{s}_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{\hat{s}_2^2}{n_2}\right)^2}$ Bajo normalidad
Cualesquiera con $E(X_1) = \mu_1,\ E(X_2) = \mu_2$ $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \sigma^2$	$\mu_1 - \mu_2$	$\mu_1 - \mu_2 \in \left(\bar{x}_1 - \bar{x}_2 \mp z_{\alpha/2}\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\right)$ Bajo normalidad o si $n_1, n_2 \to \infty$ $\mu_1 - \mu_2 \in \left(\bar{x}_1 - \bar{x}_2 \mp z_{\alpha/2}\hat{s}_T\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\right)$ Si $n_1, n_2 \to \infty$ $\mu_1 - \mu_2 \in \left(\bar{x}_1 - \bar{x}_2 \mp t_{n_1+n_2-2;\alpha/2}\hat{s}_T\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\right)$ Bajo normalidad
v.a. de Bernoulli $X_1 \sim B(p_1), X_2 \sim B(p_2)$	$p_1 - p_2$	$p_1 - p_2 \in \left(\hat{p}_1 - \hat{p}_2 \mp z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}\right)$ Si $n_1, n_2 \to \infty$
v.a. Normales $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$ $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$	$\dfrac{\sigma_1^2}{\sigma_2^2}$	$\dfrac{\sigma_1^2}{\sigma_2^2} \in \left(\dfrac{\hat{s}_1^2}{\hat{s}_2^2}F_{n_2-1;n_1-1;1-\alpha/2};\ \dfrac{\hat{s}_1^2}{\hat{s}_2^2}F_{n_2-1;n_1-1;\alpha/2}\right)$ donde $F_{n_2-1;n_1-1;1-\alpha/2} = 1\big/F_{n_1-1;n_2-1;\alpha/2}$