

CURSO 2008/2009.	JUNIO. 2ª semana
Código de la Carrera 42	Código de la Asignatura 209.

PRIMERA PARTE: CUESTIONES TEÓRICO-CONCEPTUALES

1. Indicar las diferencias entre: función de densidad, de cuantía y de distribución.

Respuesta.-

Función de densidad: si X es una variable aleatoria continua y $f(x)$ es una función no negativa tal que $\int_{-\infty}^{+\infty} f(x)dx=1$, de manera que $F(x)= P[X\leq x] = \int_{-\infty}^x f(t)dt$, entonces f es una función de densidad.

Si X es una v.a. discreta, que toma los valores x_1, x_2, \dots, x_r , una función que a cada valor x_i le asigna un valor $P[x_i]$, cumpliéndose que

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

II) $\sum_{i=1}^r P(x_i) = 1$

recibe el nombre de función de cuantía o de probabilidad

En cualquier caso (tanto si X es discreta como continua), la función $F(x) = P[X\leq x]$ es la función de distribución.

2. Razone de forma intuitiva cuando un estimador es suficiente.

Respuesta.-

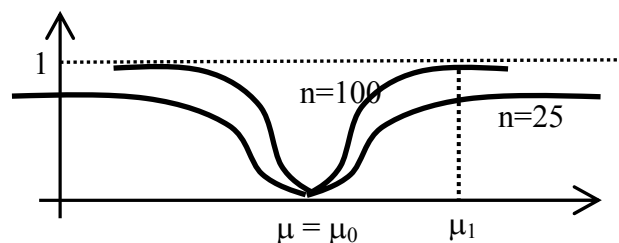
Un estimador $T = T(X_1, X_2, \dots, X_n)$ es suficiente para un parámetro θ si la distribución de (X_1, X_2, \dots, X_n) condicionada por $T = t$, no depende de θ .

Significa, de forma intuitiva que el estimador utiliza toda la información relevante contenida en la muestra sobre el parámetro a estimar y que ningún otro estadístico puede proporcionar más información sobre dicho parámetro

3. ¿Qué efecto tiene un incremento del tamaño muestral sobre la potencia de un contraste?

Respuesta.-

Para un nivel de significación α fijo, al aumentar el tamaño muestral disminuye el error de tipo II, $\beta = P[\text{Aceptar } H_0/H_0 \text{ falsa}]$, por tanto, **aumenta** la potencia del contraste $1 - \beta = P[\text{Rechazar } H_0/H_0 \text{ falsa}]$. Si representamos la curva de potencia para dos tamaños muestrales, p. ej. $n = 25$ y $n = 100$, se obtienen curvas del tipo:



4. Diferencia entre un contraste paramétrico y un contraste no paramétrico.**Respuesta.-**

Un contraste de hipótesis es paramétrico cuando las hipótesis se refieren al valor de un parámetro θ desconocido de la población (como por ejemplo μ , σ , ...).

Es no paramétrico cuando las hipótesis se refieren a características no paramétricas (forma de la distribución, localización, aleatoriedad de una muestra...)

PROBLEMAS

1.- Una página web recibe una media de 400 visitas diarias. Calcular:

- El porcentaje de días que recibe menos de 300 visitas.
- La probabilidad de que un día se reciban entre 350 y 600 visitas.
- Si queremos contratar publicidad en esa página web durante 90 días, queremos conocer la probabilidad de que en ese periodo se reciban menos de 25.000 visitas.

Justificar la distribución de probabilidad utilizada para la resolución del problema.

Solución.-

Suponemos que las visitas diarias a una página web son sucesos que ocurren de forma independiente durante un intervalo de tiempo. Podemos decir que se trata de un experimento que sigue un modelo de Poisson, donde $\lambda = 400$ (cuya media es 400 y cuya desviación típica es $\sqrt{400} = 20$).

Ahora bien, puesto que λ es suficientemente grande, podemos considerar que la variable $X =$ “número de visitas diarias” se distribuye, aproximadamente, normal $N(400, 20)$. Así pues:

$$\text{a) } P[X < 300] = (\text{corrección por continuidad}) = P[X \leq 299,5] = \\ = (\text{tipificación}) = P\left[Z \leq \frac{-100,5}{20}\right] = P[Z \leq -5,025] \cong 0.$$

$$\text{b) } P[350 < X < 600] = (\text{corrección por continuidad}) = P[350,5 \leq X \leq 599,5] = \\ = (\text{tipificación}) = P\left[\frac{-49,5}{20} \leq Z \leq \frac{199,5}{20}\right] = P[-2,475 \leq Z \leq 9,975] = (\text{tablas}) = 1 - 0,0067 = \\ = 0,9933$$

Si hubiésemos interpretado el número de visitas entre 350 y 600, ambos inclusive, hubiésemos obtenido un resultado cuya diferencia con el anterior es prácticamente insignificante.

c) Si X_i es el número de visitas el día i -ésimo ($i = 1, 2, \dots, 90$), la variable $Y = \sum_{i=1}^{90} X_i$ será de Poisson de parámetro $90 \cdot 400 = 36000$ y, por tanto, aproximadamente normal $N(36000, \sqrt{36000}) \cong N(36000; 189,74)$.

Con estos parámetros, la probabilidad $P[Y < 25000]$ es prácticamente nula puesto que la distancia de 25000 a 36000 es del orden de 58 “desviaciones típicas”

$$\left(\left|\frac{25000 - 36000}{189,74}\right| \cong 57,97\right).$$

2.- Un estudio sobre los sueldos mensuales de 12 trabajadores nos da los siguientes resultados:

1100 2000 1800 800 4200 1234 1354 1600 1900 1850 1200 1000

Determinar, (si sabemos de antemano que los salarios mensuales siguen una distribución normal), los intervalos de confianza para el precio medio y para la varianza del precio, con un nivel de significación del 0,05.

Solución.-

Para la muestra dada, obtenemos la media muestral $\bar{x} = \frac{20038}{12} \cong 1669,83$, la varianza

muestral $s^2 = \frac{1}{11} \sum_{i=1}^{12} (x_i - \bar{x})^2 \cong 787131,97$ y la desviación típica muestral $s \cong 887,20$

La variable $\frac{\bar{X} - \mu}{S} \sqrt{n}$ se distribuye t-Student con n-1 grados de libertad.

Para 11 grados de libertad, obtenemos de las tablas el intervalo al nivel de significación 0,05:

$$\left[-2,201 \leq \frac{\bar{X} - \mu}{S} \sqrt{12} \leq 2,201 \right] \leftrightarrow \left[\bar{X} - \frac{2,201 \cdot S}{\sqrt{12}} \leq \mu \leq \bar{X} + \frac{2,201 \cdot S}{\sqrt{12}} \right]$$

que, para los valores de la muestra dada proporciona:

$$[1669,83 - 563,70 \leq \mu \leq 1669,83 + 563,70] \leftrightarrow [1106,13 \leq \mu \leq 2233,53]$$

La variable $\frac{(n-1)S^2}{\sigma^2}$ se distribuye χ^2 con n-1 grados de libertad.

Para 11 grados de libertad, obtenemos de las tablas el intervalo al nivel de significación 0,05:

$$\left[3,816 \leq \frac{11S^2}{\sigma^2} \leq 21,92 \right] \leftrightarrow \left[\frac{11S^2}{21,92} \leq \sigma^2 \leq \frac{11S^2}{3,816} \right]$$

que, para los valores de la muestra dada proporciona:

$$[395002,36 \leq \sigma^2 \leq 2268986,29]$$