

ESTADÍSTICA II	ENERO 2011
Código de la Carrera 65	Código de la Asignatura 201

PRIMERA PARTE: CUESTIONES TEÓRICO-CONCEPTUALES

1. ¿Qué entendemos por Función de distribución?. Explique la respuesta.

Respuesta.-

Si X es una variable aleatoria y P es la función de probabilidad, se define la función de distribución $F: \mathbb{R} \rightarrow \mathbb{R} / F(x) = P[X \leq x]$.

Si X es discreta con valores $\{x_i, i= 1, 2, 3, \dots, r\}$ y función de probabilidad $P(x_i) = P[X = x_i]$, entonces $F(x) = \sum_{x_i \leq x} P(x_i)$.

Si X es continua y $f: \mathbb{R} \rightarrow \mathbb{R}$ es la función de densidad, entonces $F(x) = \int_{-\infty}^x f(t)dt$

Se deduce de la definición que $F(x)$ es una función creciente tal que $0 \leq F(x) \leq 1$.

2. ¿Por qué es interesante calcular la esperanza y la varianza de una variable aleatoria?. Explique la respuesta.

Respuesta.-

La esperanza y la varianza son las principales medidas de posición y de dispersión, respectivamente. En muchos casos, por ejemplo, si la variable aleatoria es normal, la esperanza y la desviación típica (raíz cuadrada de la varianza) caracterizan por completo la distribución de la variable.

Otro uso interesante para la esperanza y la varianza es que, mediante la desigualdad de Chebychev podemos acotar la probabilidad de que la variable aleatoria se aleje más o menos de su media (esperanza).

3. Explique si existe diferencia entre parámetro poblacional y estadístico. Ponga un ejemplo.

Respuesta.-

En una población con una variable aleatoria X cuya función de distribución sea $F(x)$, consideremos una muestra aleatoria simple $(X_1, X_2, X_3, \dots, X_n)$. Un parámetro poblacional es una constante característica de la población, como por ejemplo cualquier momento respecto del origen $\alpha_r = E(X^r)$ o respecto de la media $\mu_r = E[(X-E(X))^r]$, mientras que un estadístico es cualquier función de las variables aleatorias que integran la muestra, como por ejemplo, la media muestral, la varianza muestral, etc.

4. ¿Qué es una hipótesis estadística?. ¿Que tipos de hipótesis estadísticas podemos realizar en un contraste de hipótesis paramétrico?.

Respuesta.-

Es la hipótesis que se realiza respecto del valor de cierta característica desconocida de la distribución de una variable aleatoria en una población determinada. Si la característica desconocida es un parámetro poblacional θ , entonces, para confirmarla o rechazarla efectuamos el contraste entre dicha hipótesis (hipótesis nula H_0) y la contraria (hipótesis alternativa H_1).

Las hipótesis paramétricas pueden ser simples cuando se refieren a un único valor del parámetro (ej. $\theta = \theta_0$) o compuestas, cuando se refieren a una región no puntual del espacio paramétrico (ej.: $\theta \neq \theta_0$, $\theta < \theta_0$, etc.).

PROBLEMAS

1. Un contratista está interesado en conocer el coste total de un proyecto con el objeto de hacer una oferta. De trabajos anteriores sabe que los materiales le costarán 25.000 euros y la mano de obra 9.000 euros por día de trabajo. El contratista construye unas probabilidades subjetivas sobre la duración del proyecto como se ve en la siguiente tabla:

Duración en días	10	11	12	13	14
Probabilidad	0,1	0,4	0,3	0,1	0,1

- a) Cuál es la duración media del proyecto? ¿y su desviación típica?
- b) ¿Cuál será el coste total medio del proyecto? ¿y su desviación típica?
- c) ¿Qué oferta se podrá considerar como rentable para el contratista?

Solución.-

a) Sea X la variable aleatoria “duración en días del proyecto”. La duración media del proyecto es $E(X) = 10 \cdot 0,1 + 11 \cdot 0,4 + 12 \cdot 0,3 + 13 \cdot 0,1 + 14 \cdot 0,1 = 11,7$ días

Calculemos ahora $E(X^2) = 100 \cdot 0,1 + 121 \cdot 0,4 + 144 \cdot 0,3 + 169 \cdot 0,1 + 196 \cdot 0,1 = 138,1$.
luego $\text{Var}(X) = 138,1 - 11,7^2 = 1,21$. De donde $\text{DT}(X) = \sqrt{1,21} = 1,1$ días

b) Sea Y la variable aleatoria “coste total del proyecto”. Se cumple que $Y = 9000X + 25000$, luego el coste total medio será $E(Y) = 9000E(X) + 25000 = 130300$ euros. La desviación típica $\text{DT}(Y) = 9000 \cdot \text{DT}(X) = 9900$ euros.

c) Podría considerarse rentable ofertar el valor de y_i que haga máximo $y_i \cdot P(Y = y_i)$, que resulta ser $y_2 = 11 \cdot 9000 + 25000 = 124000$ euros.

- 2.- Una empresa dedicada a ofertar cursos de inglés para niños desea obtener un intervalo al 90% de confianza para el tiempo medio que tarda un niño en realizar uno de sus cursos. Se toma una muestra de 12 niños observando los tiempos invertidos en realizar el curso, y se obtiene que para la muestra seleccionada la media es de 90 días y la desviación típica 10 días. Suponiendo que la población se distribuye normalmente. Obtener:

- a) El intervalo indicado con un error del 5%
- b) El intervalo indicado considerando una longitud de 10 días.
- c) Explique los resultados obtenidos.

Solución.-

Observemos previamente que la variable $\frac{\bar{X} - \mu}{S} \sqrt{n}$ se distribuye como una t-Student con $n-1$ grados de libertad. El intervalo al $(1-\alpha)\%$ de confianza tiene la forma

$\left[\bar{X} - \frac{t_{\alpha} S}{\sqrt{n}}, \bar{X} + \frac{t_{\alpha} S}{\sqrt{n}} \right]$. En las tablas de la t-Student para 11 grados de libertad obtenemos que

$t_{0,05} = 1,796$, luego el intervalo al 90% de confianza sería $\left[90 - \frac{1,796 \cdot 10}{\sqrt{12}}, 90 + \frac{1,796 \cdot 10}{\sqrt{12}} \right] \cong [84,82$

; 95,18] cuya longitud aproximada es de 10,37 días y un error relativo $\frac{|\bar{X} - \mu|}{\bar{X}} \leq \frac{t_{\frac{\alpha}{2}} S}{\bar{X} \sqrt{n}} \leq 0,058$,

que equivale a un 5,8%.

Entonces:

- El intervalo de confianza, con error del 5% cumplirá que $\frac{|\bar{X} - \mu|}{\bar{X}} \leq 0,05$ es decir $\bar{X} - 0,05\bar{X} \leq \mu \leq \bar{X} + 0,05\bar{X}$ que, para $\bar{X} = 90$ es el intervalo [85,5 ; 94,5]
- El intervalo de confianza con una longitud de 10 días cumplirá que $|\bar{X} - \mu| = 5$ es decir $\bar{X} - 5 \leq \mu \leq \bar{X} + 5$ que, para $\bar{X} = 90$ es el intervalo [85 ; 95]
- En los apartados a y b hemos obtenido intervalos similares al obtenido aplicando la t-Student aunque esto se debe exclusivamente a los datos particulares del problema.