

<b>ESTADÍSTICA II</b>	<b>SEPTIEMBRE 2008</b>
<b>Código de la Carrera 65</b>	<b>Código de la Asignatura 201</b>

**PRIMERA PARTE: CUESTIONES TEÓRICO-CONCEPTUALES**

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

**Respuesta.-**

La esperanza y la varianza don 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).

2. Definición y características de una variable aleatoria binomial. Ponga un ejemplo.

**Respuesta.-**

Se dice que una variable aleatoria X es de Bernoulli si sólo toma los valores 0 y 1, con probabilidades  $P(X=1) = p$  y  $P(X=0) = 1-p = q$ . Se tiene entonces que  $\mu = E(X) = p$  y  $\sigma^2 = \text{Var}(X) = pq$ .

Se dice que una variable aleatoria X es binomial de parámetros (n,p) si existen n variables de Bernoulli  $X_1, X_2, \dots, X_n$ , independientes e idénticamente distribuidas (es decir que  $P(X_i) = p, \forall i$ ) tales que  $X = \sum_{i=1}^n X_i$ . X toma los valores 0, 1, 2, 3, ..., n y se tiene que

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, 2, \dots, n. \text{ Además } \mu = E(X) = \sum_{i=1}^n E(X_i) = np \text{ y } \sigma^2 = \text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) = npq.$$

Por ejemplo, se sabe que el 30% de las personas de una determinada población son lectores de un periódico P. Elegimos 50 personas al azar. Entonces la variable X = “nº de personas que leen el periódico P” se distribuye de forma binomial B(50, 0,3).

3. Explique el significado que tiene el valor crítico en un contraste de hipótesis paramétrico. Razone la respuesta.

**Respuesta.-**

Es aquel valor o valores que separan la región de aceptación de la región de rechazo de la hipótesis nula.

Supongamos por ejemplo que, respecto de determinado parámetro poblacional  $\theta$ , establecemos la hipótesis nula  $H_0: \theta < \theta_0$  y deseamos contrastarla con un nivel de significación  $\alpha$ . Si  $\hat{\theta}$  es el estimador de  $\theta$ , el valor crítico r es tal que, suponiendo cierta  $H_0$ ,  $P[\hat{\theta} > r] = \alpha$ . Si el estimador toma el valor  $\hat{\theta}_0$  para la muestra concreta obtenida y  $\hat{\theta}_0 > r$ , rechazaríamos  $H_0$ .

4. ¿Cuál es el objetivo de la estimación por intervalos de confianza? Razone la respuesta.

**Respuesta.-**

Supongamos que de la distribución de una variable aleatoria  $X$  se desconoce un determinado parámetro  $\theta$ . El objeto de la estimación por intervalos de confianza es hallar un intervalo de poca amplitud y alta probabilidad (coeficiente de confianza) de que en su interior se encuentre el parámetro  $\theta$ .

**PROBLEMAS**

1.- Una compañía de seguros posee 1000 pólizas. El siniestro que cubre cada póliza tiene una probabilidad de ocurrencia del 2% .

- Calcular la probabilidad de que la compañía tenga menos de dos siniestros.
- Calcular la probabilidad de que tenga al menos un siniestro.
- Explique los resultados obtenidos.

**Solución.-**

El número  $X$  de siniestros es una variable aleatoria binomial  $B(1000, 0,02)$ . Puesto que  $np = 1000 \cdot 0,02 = 20 > 5$  y  $p = 0,02 < \frac{1}{2}$ , la variable  $X$  es aproximadamente normal

$N(20, \sqrt{20 \cdot 0,98}) = N(20; 4,43)$ , luego:

$$\begin{aligned} \text{a) } P[X < 2] &= (\text{corrección por continuidad}) = P[X \leq 1,5] = (\text{tipificando}) = \\ &= P\left[Z \leq \frac{1,5 - 20}{4,43}\right] = P[Z \leq -4,18] \cong 0. \end{aligned}$$

$$\begin{aligned} \text{b) } P[X \geq 1] &= (\text{corrección por continuidad}) = P[X \geq 0,5] = (\text{tipificando}) = \\ &= P\left[Z \geq \frac{0,5 - 20}{4,43}\right] = P[Z \geq -4,4] \cong 1. \end{aligned}$$

c) Según hemos visto en el apartado a, el valor esperado del número de siniestros (la media de la variable  $X$ ) es de 20. Por tanto, con esos datos, se puede comprender de manera intuitiva que el hecho de tener solamente menos de dos siniestros es altamente improbable. Y, contrariamente, tener al menos un siniestro es prácticamente seguro.

2.- Una agencia de viajes supone que el 70% de la población ha salido al extranjero de vacaciones en alguna ocasión. Con el objetivo de obtener una información más precisa toma una muestra aleatoria simple.

- Obtener la distribución muestral del estadístico correspondiente, para una muestra aleatoria de 90 personas y de 450 personas.
- Realizar la representación gráfica de las funciones de densidad del estadístico correspondiente, para ambas muestras.
- Explique los resultados obtenidos.

**Solución.-**

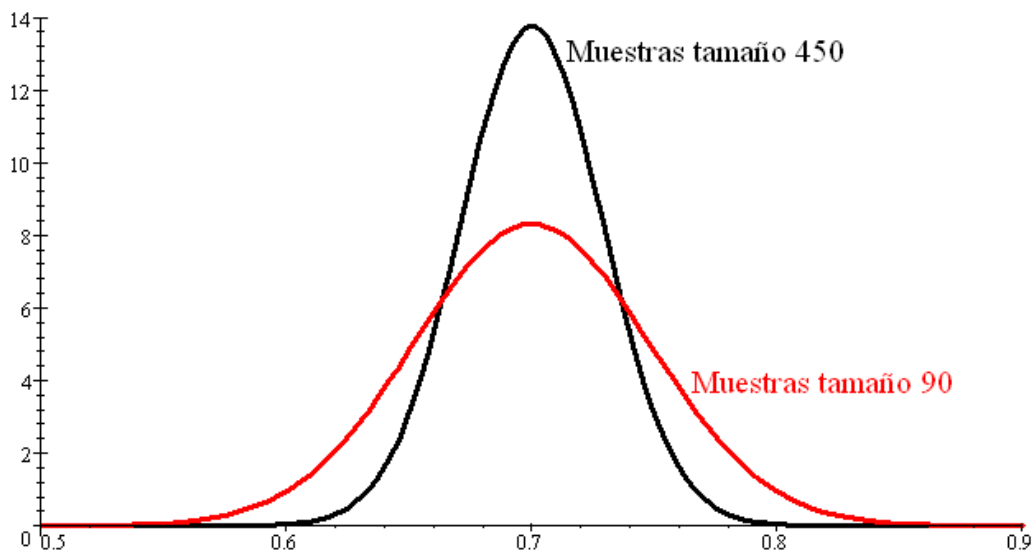
a) La hipótesis de la agencia supone que la proporción (poblacional) de personas que han salido al extranjero de vacaciones en alguna ocasión es 0,7. Para obtener una información más precisa de tal parámetro (es decir, para contrastar dicha hipótesis), tomaremos el estadístico proporción muestral :

$$\hat{p} = \frac{\text{n}^\circ \text{ de personas de la muestra que han salido al extranjero}}{\text{tamaño de la muestra}}$$

- Para muestras de 90 personas,  $E[\hat{p}] = \frac{90 \cdot 0,7}{90} = 0,7$  y desviación típica de  $\hat{p} = \sqrt{\frac{0,7 \cdot 0,3}{90}} = 0,048$ . Luego  $\hat{p}$  se distribuye aproximadamente normal  $N(0,7; 0,048)$ .

- De forma análoga, para muestras de 450 personas,  $\hat{p}$  se distribuye aproximadamente normal  $N\left[0,7; \sqrt{\frac{0,7 \cdot 0,3}{250}}\right] \cong N(0,7; 0,029)$

b) Las funciones de densidad son sendas campanas de Gauss con máximo en el punto de abscisa 0,7 y ordenada  $\frac{1}{\sigma\sqrt{2\pi}}$  cuyo valor es  $\frac{1}{0,048\sqrt{2\pi}} \cong 8,31$  para muestras de tamaño 90 y  $\frac{1}{0,029\sqrt{2\pi}} \cong 13,76$  para muestras de tamaño 450. Los puntos de inflexión tienen abscisas, en cada caso  $0,7 \pm \sigma$ .



c) Para un mismo nivel de significación, el tamaño de la región de aceptación es menor para muestras de tamaño 450 y por tanto más improbable cometer error de tipo II y en consecuencia el contraste efectuado con muestras de este tamaño tiene mayor potencia que el efectuado con muestras de tamaño 90.