UNIVERSIDAD CARLOS III DE MADRID
Ingeniería Técnica en Informática de Gestión
29 de Junio de 2007
ESTADÍSTICA

Nota: Justifica todas tus respuestas. Responde cada pregunta en hojas diferentes

1. Se desea medir la diferencia que hay entre dos grupos de estudiantes de ingeniería en el tiempo semanal dedicado al estudio. Tomamos una muestra de 45 estudiantes del grupo A y la media de horas semanales de estudio nos resulta 32 horas. Tomamos una media de 60 estudiantes del grupo B y la media resultante es de 25 horas. Las horas semanales de estudios presentan una cuasivarianza de 48 para el grupo A y de 56 para el grupo B.

   (a) Calcular un intervalo de confianza al 95% para la diferencia de medias. De acuerdo con el intervalo hallado, ¿hay evidencia suficiente para afirmar que las medias son iguales? **Justifica la respuesta (tanto el método como las conclusiones)**.

   (b) Usando exclusivamentee los resultados del apartado anterior, y sin realizar ninguna operación adicional, ¿cómo cambiarían las conclusiones si usásemos $\alpha = 0.10$? **Justifica la respuesta.**

   SOLUCIÓN:

   (a) Queremos comparar las medias de dos poblaciones usando muestras independientes. Los tamaños muestrales son suficientemente grandes, por tanto podemos utilizar lo intervalos basados en la normalidad de las medias muestrales aunque las poblaciones no lo sean. Las varianzas muestrales no son muy diferentes. Las cuasidesviaciones típicas son 6.9 y 7.5 respectivamente. Cuando una no es más del doble de la otra, podemos utilizar un estimador común de las varianzas en el contraste de igualdad de medias. Este estimador es

$$\hat{s}_T^2 = \frac{44 \times 48 + 59 \times 56}{44 + 59} = 52.6.$$

   Por tanto, el intervalo es

$$\begin{aligned} IC(0.95) : \mu_A - \mu_B \in & \left\{ \bar{x}_A - \bar{x}_B \pm z_{0.025} \hat{s}_T \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \right\} \\ = & \left\{ 32 - 25 \pm 1.96 \times 7.25 \times \sqrt{\frac{1}{45} + \frac{1}{60}} \right\} \\ = & \; 7 \pm 2.8 = (4.2; 9.8) \text{ horas.} \end{aligned}$$

   Un intervalo de confianza de nivel $1 - \alpha$ puede utilizarse para realizar un contraste de hipótesis bilateral. Como el cero no está dentro del intervalo se puede rechazar la hipótesis nula $H_0 : \mu_A - \mu_B = 0$ (de que ambos grupos estudien el mismo número medio de horas) con un nivel de significación del 5%.

(b) Un intervalo con $\alpha = 0.10$ será de un nivel de confianza del 90%, menor que el del apartado anterior. Por tanto será más estrecho. Será del tipo $\mu_A - \mu_B \in (7 \pm L)$ con $L < 2.8$. Por tanto, seguirá sin contener al cero. Con $\alpha = 0.10$, seguiremos rechazando la hipótesis nula de igualdad de medias.

2. Tenemos dos máquinas diferentes dedicadas a la fabricación de unas barras metálicas. La máquina 1 produce barras defectuosas en una proporción de un 3%. La máquina 2 produce barras cuya longitud, medida en cm, es una variable aleatoria cuya función de distribución $F(x) = P(X \leq x)$ es

$$F(x) = \begin{cases} 0 & ; \quad x \leq 0 \\ Kx^3 & ; \quad 0 < x < 2 \\ 1 & ; \quad x \geq 2 \end{cases}$$

(a) Tomamos una muestra aleatoria de 10 barras producidas por la máquina 1.¿Cuál es la probabilidad de que más de una sea defectuosa? Justifica la respuesta.

(b) Calcular el valor de K para que la función de distribución que rige la longitud de las barras de la máquina 2 esté bien definida.

(c) Las barras obtenidas por la máquina 2 cuya longitud esté fuera del intervalo [0.15;1.92] se consideran defectuosas. Calcular la proporción de piezas defectuosas.

SOLUCIÓN:

(a) Sea Y =número de barras defectuosas en un lote de 10. Suponiendo que el estado defectuoso/aceptable de cada pieza es independiente de las demás, tendremos que $Y \sim B(10, 0.03)$. Por tanto

$$\begin{aligned} P(Y > 1) &= 1 - P(Y \leq 1) = 1 - P\left\{(Y = 0) + P(Y = 1)\right\} \\ &= 1 - \left\{\left[\binom{10}{0} 0.03^0 0.97^{10}\right] + \left[\binom{10}{1} 0.03^1 0.97^9\right]\right\} \\ &= 1 - \left(0.97^{10} + 10 \times 0.03 \times 0.97^9\right) = 0.0345. \end{aligned}$$

(b) Para $x = 2$ tenemos que $F(2) = 1$. Por tanto

$$K2^3 = 1 \Rightarrow K = 1/8.$$

(c) La probabilidad de producir una pieza defectuosa es

$$\begin{aligned} P(\text{aceptable}) &= F(1.92) - F(0.15) = \frac{1}{8}1.92^3 - \frac{1}{8}0.15^3 = 0.884 \\ P(\text{defectuosa}) &= 1 - P(\text{aceptable}) = 1 - 0.884 = 0.116 \end{aligned}$$

3. Sea X una variable aleatoria de media μ y varianza σ^2 desconocidas, y sea $x_1, ..., x_{100}$ una muestra aleatoria simple de 100 datos de dicha variable. Decir si las siguientes afirmaciones son verdaderas o falsas, razonando la respuesta

(a) Como la muestra es grande, X es una normal por el teorema central del límite

(b) Si X fuese una Poisson, $\mu = \sigma^2$.

(c) Como la muestra es suficientemente grande, podemos hacer un intervalo de confianza para la varianza asumiendo condiciones de normalidad.

(d) Los diferentes valores que se obtienen al promediar las observaciones de muestras independientes de tamaño 100 siguen una ley normal, aunque X no lo sea.

(e) Si X es normal $\mu = \sigma^2/n$.

(f) La media muestral tiene una varianza que es sólo la centésima parte de la varianza de X.

SOLUCIÓN:

(a) FALSO. La distribución de una variable aleatoria no depende del número de realizaciones que tengamos de ella.

(b) VERDADERO. En una variable aleatoria de Poisson $E(X) = \lambda$ y $\text{Var}(x) = \lambda$.

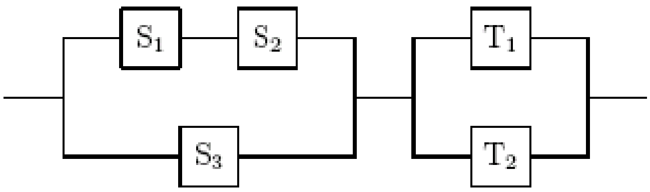
(c) FALSO. La inferencia sobre la varianza que hemos estudiado depende de que la población sea normal, independientemente del tamaño de la muestra.

(d) VERDADERO. Como el tamaño muestral es grande, la media muestral será, por el teorema central del límite, una variable aleatoria normal independientemente de cómo sea X.

(e) FALSO. La media y la varianza no tienen esa relación.

(f) VERDADERO. La varianza de la media muestral es $\text{var}(\bar{x}) = \sigma^2/n$, donde σ^2 es la varianza de la población. Por tanto, como $n = 100$, $\text{var}(\bar{x}) = \sigma^2/100$.

4. Como puede observarse en la figura, el siguiente sistema está formado por dos subsistemas conectados en serie, S y T, que están formados a su vez por los dispositivos S_1, S_2, S_3 y T_1, T_2, respectivamente.



Los cinco dispositivos fallan de manera independiente, siendo las probabilidades de fallo las siguientes:

$$S_1: \ 0.1 \quad S_2: \ 0.2 \quad S_3: \ 0.3 \qquad T_1: \ 0.1 \quad T_2: \ 0.2$$

Se pide:

(a) Calcula la probabilidad de que el subsistema S funcione.

(b) Calcula la probabilidad de que el subsistema T funcione

(c) Calcula la probabilidad de que el subsistema completo funcione.

SOLUCIÓN:

(a) En primer lugar llamemos: S_i= “falle dispositivo i” y $\bar{S}_i$= “funcione dispositivo i”. El sistema S funciona si funciona alguna de las dos ramas en paralelo que tiene. La rama superior, formada por S_1 y S_2 funcionará si funcionan ambos dispositivos. Se tiene entonces

$$\begin{aligned} P(\text{funcione S}) &= P(\bar{S}) = P\left[\bar{S}_3 \cup \left(\bar{S}_1 \cap \bar{S}_2\right)\right] \\ &= P(\bar{S}_3) + P\left(\bar{S}_1 \cap \bar{S}_2\right) - P\left(\bar{S}_3 \cap \bar{S}_1 \cap \bar{S}_2\right) \\ &= P(\bar{S}_3) + P\left(\bar{S}_1)P(\bar{S}_2\right) - P\left(\bar{S}_3)P(\bar{S}_1)P(\bar{S}_2\right) \\ &= 0.7 + 0.9 \times 0.8 - 0.7 \times 0.9 \times 0.8 = 0.916 \end{aligned}$$

(b) De forma similar al apartado anterior, llamemos T_i =el dispositivo i-ésimo falla. Al ser un sistema en paralelo, funcionará si lo hace alguno de ellos. Por tanto

$$\begin{aligned} P(\text{funcione T}) &= P(\bar{T}) = P(\bar{T}_1 \cup \bar{T}_2) \\ &= P(\bar{T}_1) + P(\bar{T}_2) - P\left(\bar{T}_1 \cap \bar{T}_2\right) \\ &= P(\bar{T}_1) + P(\bar{T}_2) - P\left(\bar{T}_1)P(\bar{T}_2\right) \\ &= 0.9 + 0.8 - 0.9 \times 0.8 = 0.98 \end{aligned}$$

(c) Como los sistemas S y T están conectados en serie para que funcione el sistema completo deben funcionar S y T

$$\begin{aligned} P(\text{funciona el sistema}) &= P(\bar{S} \cap \bar{T}) = P(\bar{S})P(\bar{T}) \\ &= 0.916 \times 0.98 = 0.89768 \end{aligned}$$