Universidad Carlos III de Madrid
Ingeniería Técnica en Informática de Gestión
30 de mayo de 2009
ESTADÍSTICA
Duración: 2 horas

Nota: Utiliza siempre $\alpha = 0{,}05$, **salvo que se indique lo contrario**. Justifica todas tus respuestas. Responde cada pregunta en hojas diferentes. El examen de prácticas es en las aulas 2.2.C03-04-05, después de este examen.

1. Una tarea de computación compleja es distribuida en 5 subtareas independientes y ejecutada por un conjunto de 5 ordenadores que operan en paralelo, realizando una subtarea independiente cada uno. La tarea se considera concluida cuando todos los ordenadores han terminado de ejecutar la subtarea que se le ha encomendado a cada uno. El tiempo necesario para ejecutar una subtarea elegida al azar puede modelizarse según una exponencial de media 10 segundos. Se pide:

   a) Si llamamos T_i al tiempo que el ordenador i-ésimo tarda en ejecutar una subtarea asignada al azar, ¿cuál es la probabilidad de que tarde más de 5 segundos en realizarla?

   b) Llamemos T_T al tiempo total que se tarda en ejecutar la tarea; es decir, desde que se lanzan simultaneamente las 5 subtareas en los 5 ordenadores hasta que acaba el último de ellos. ¿Cuál es la probabilidad de que T_T supere los 5 segundos?

   c) La ejecución de los programas demanda bastante memoria, por lo que es habitual que el ordenador falle. Si algún ordenador falla, el sistema informático vuelve a programar la ejecución de la tarea entre los ordenadores restantes (afectando por tanto los tiempos de ejecución). Si la probabilidad de que un ordenador falle durante la ejecución es de 0.25, ¿cuál es la probabilidad de que la tarea pueda realizarse con éxito?

   SOLUCIÓN:

   a) Según la información del enunciado, $T_i \sim Exp(\lambda)$, con $E(T_i) = 1/\lambda = 10$ segundos/subtarea. Por tanto $\lambda = 1/10 = 0{,}1$ subtareas/segundo. Se tiene entonces que

   $$P(T_i > 5) = e^{-\lambda 5} = e^{-0{,}1\times 5} = 0{,}6065.$$

   b) Al estar los ordenadores funcionando en paralelo y de forma independiente, se puede escribir que

   $$\begin{aligned} P(T_T &> 5) = P(\text{alguno dure más de 5 s.}) \\ &= 1 - P(\text{ninguno dure más de 5 s.}) \\ &= 1 - P\left[(T_1 \leq 5) \cap (T_2 \leq 5) \cap \cdots \cap (T_5 \leq 5)\right] \\ &= 1 - P(T_1 \leq 5)P(T_1 \leq 5)\cdots P(T_5 \leq 5) \\ &= 1 - (1 - 0{,}6065)^5 = 0{,}99057. \end{aligned}$$

   c) La tarea se puede realizar mientras funcione algún ordenador. Por consiguiente,

   $$\begin{aligned} P(\text{ejecución tarea con éxito}) &= P(\text{funcione algún ordenador}) \\ &= 1 - P(\text{fallen todos los ordenadores}) \\ &= 1 - P[(\text{Falla ord. 1}) \cap \cdots \cap (\text{Falla ordenador 5})] \\ &= 1 - P(\text{Falla ord. 1}) \cdots P(\text{Falla ord. 5}) \\ &= 1 - 0{,}25^5 = 0{,}999 \end{aligned}$$

2. El tiempo de ejecución de un programa desarrollado por la compañía A se distribuye según una Normal de media no superior a 1.2 horas, según declara la propia compañía. Sin embargo, la competencia, la compañia B, afirma que esto no es cierto ya que, después de probar el programa en 9 ocasiones, observó una duración media de 1.6 horas y una cuasivarianza de 0.36. Se pide:

a) Establece el contraste de hipótesis adecuado para contrastar la afirmación realizada por la competencia.

b) ¿Avalan estos datos la afirmación de la competencia?

c) La compañia B reta a la compañía A a comparar sus algoritmos en las mismas condiciones. En 10 simulaciones realizadas con cada uno de ellos se obtuvo; para el algoritmo de la compañia A, un tiempo medio de ejecución de 1.54 horas con una cuasivarianza de 0.6; y para el algoritmo de la compañía B, un tiempo medio de 1.44 horas con una cuasivarianza de 0.9 ¿Proporcionan los datos evidencia suficiente para concluir que hay diferencias entre los tiempos medios de ejecución de ambos algoritmos? Supóngase normalidad e igualdad de varianzas poblacionales en los tiempos de ejecución.

SOLUCIÓN:

a) la afirmación de la competencia supone que: o bien que el tiempo medio es superior a 1.2 horas, $\mu_A > 1{,}2$, o no lo es, $\mu_A \leq 1{,}2$. Por tanto, el contraste de hipótesis debe escribirse como:

$$\begin{aligned} H_0 &: \mu_A \leq 1{,}2. \\ H_1 &: \mu_A > 1{,}2. \end{aligned}$$

b) Para contrastar esta media poblacional a partir de la media muestral usamos el estadístico

$$t = \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}},$$

que en nuestra muestra toma el valor

$$t = \frac{1{,}6 - 1{,}2}{\sqrt{0{,}36/9}} = 2.$$

Al ser una población normal, la distribución de referencia es la $t_{n-1} = t_8$. La región de rechazo es para valores altos de t_8, por tanto rechazamos H_0 si $t > t_{8;0{,}05} = 1{,}85$. Como $t = 2$ rechazamos $\mathrm{H_0}$ y hay por tanto evidencia suficiente, con $\alpha = 0{,}05$, para dudar de la compañía A.

c) El contraste es

$$\begin{aligned} H_0 &: \mu_A = \mu_B \\ H_1 &: \mu_A \neq \mu_B \end{aligned}$$

y el estadístico de contraste es

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{S}_T \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}},$$

donde

$$\hat{S}_T^2 = \frac{9 \times 0{,}6 + 9 \times 0{,}9}{18} = 0{,}75 \Rightarrow \hat{s}_T = 0{,}866.$$

Se tiene entonces que

$$t = \frac{1{,}54 - 1{,}44}{0{,}866\sqrt{\dfrac{1}{10} + \dfrac{1}{10}}} = 0{,}258.$$

Rechazaremos para valores altos, en valor absoluto, de t. Como $|t_{18;0{,}025}| = 2{,}101 > 0{,}258$ nos encontramos en la región de No Rechazo. No podemos por tanto rechazar la hipótesis nula de igualdad en las duraciones medias de ambos algoritmos.

3. Un fabricante de transistores del tipo BC547B sabe por su información histórica que cuando su producción se mantiene en los niveles de calidad deseables, el valor de la variable X = *ganancia en corriente* de los transistores (conocida por coeficiente β, adimensional) sigue una distribución de media 290. Se toma una muestra de este tipo de transistores. El Statgraphics proporciona el siguiente resumen estadístico:

```
Summary Statistics for BC547B

Count = 100
Average = 282,29
Mode = 304,0
Variance = 766,854
Standard deviation = 27,6921
Skewness = 0,324111
```

```
Confidence Intervals for BC547B

95,0% confidence interval for mean:  [276,795;287,785]

95,0% confidence interval for standard deviation: [24,3139;32,1693]
```

Contestar, justificando la respuesta, si las siguientes afirmaciones son verdaderas o falsas:

a) El intervalo de confianza de la media no es fiable pues no sabemos si la variable X es normal.

b) El intervalo de confianza de la desviación típica no es fiable pues no sabemos si la variable X es normal.

c) Este resumen estadístico nos lleva a concluir que el proceso productivo no se mantiene en los niveles de calidad deseables.

d) Los intervalos de confianza de nivel 90 % serán más estrechos que los mostrados del 95 %

e) El intervalo de confianza de la desviación típica no está centrado en 27.69, lo que es indicio de falta de normalidad en X.

SOLUCIÓN:

a) Falso. Al ser n suficientemente grande, por el Teorema Central del Límite el intervalo para la media sigue siendo válido aunque la variable no sea nomal.

b) Verdadero. El intervalo de confianza de la desviación típica se basa en que la población sea normal

c) Verdadero. Como el 290 está fuera del intervalo del 95 %, tenemos que rechazar, con un nivel de significación del 5 %, que el proceso tenga de media 290.

d) Verdadero. A menor confianza, mayor valor de α, y menor valor de $z_{\alpha/2}$. Por tanto, tendremos intervalos más estrechos.

e) Falso. El intervalo de confianza de la desviación típica no está nunca centrado en el estimador.