UNIVERSIDAD CARLOS III DE MADRID
Ingeniería Técnica en Informática de Gestión
6 de Junio de 2008
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. A las 12:30 será el examen de prácticas en el aula 7.3 J06.

1. En una fábrica dedicada a la producción de tuberías, se decide sustituir una de las máquinas implicadas en el proceso productivo por una de tecnología más avanzada. El objetivo es acelerar el proceso productivo sin que las características del producto final se vean alteradas. La máquina se programa para que corte las tuberías en una longitud de 30 cm. Después de un mes de actividad productiva, tras la adquisición de la nueva maquinaria, se tomaron al azar 50 de las tuberías fabricadas durante el último mes y se obtuvieron los siguientes resultados:

$$\bar{x} = 28\ cm; \sum_{i=1}^{50} (x_i - \bar{x})^2 = 1000\ cm^2$$

Se pide:

a) A la vista de estos datos, ¿hay suficiente evidencia para asegurar que la longitud media actual de las tuberías ha disminuido con respecto al valor programado?

b) Calcula el p-valor asociado al contraste de hipótesis realizado en el apartado anterior e interpreta dicho resultado.

c) Calcula el número mínimo de tuberías que deberíamos seleccionar para obtener un intervalo de confianza para la media poblacional con una precisión de $\pm 0{,}4\ cm$ y un nivel de confianza del 95 %.

SOLUCIÓN:

a) Para responder a esta pregunta necesitamos realizar el siguiente contraste,

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

con un nivel de significación α y donde $\mu_0 = 30cm$. El estadístico de contraste es

$$t_0 = \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}.$$

De los datos se tiene que

$$\begin{aligned} \hat{s} &= \sqrt{\frac{\sum_{i=1}^{50} (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1000}{49}} = 4{,}52. \\ t_0 &= \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}} = \frac{28 - 30}{4{,}52/\sqrt{50}} = -3{,}13 \end{aligned}$$

Al tener una muestra grande, la distribución de referencia es la N(0,1). Usando $\alpha = 0{,}05$ rechazaremos la hipótesis nula si $t_0 < -z_{0,05} = -1{,}65$. Como $t_0 = -3{,}13 < -1{,}645 = z_{0,05}$, entonces los datos muestran que sí hay suficiente evidencia para rechazar la hipótesis nula H_0 a favor de la alternativa H_1, con un nivel de significación del 5 %.

b) El p-valor es, en este caso, el área que queda bajo la densidad de la normal estándar hacia la izquierda del valor $z_0 = -3{,}13$. El p-valor se interpreta como el mínimo nivel de significación que haría falta para que no se rechaza la hipótesis nula; esdecir, que deja el valor del estadístico de contraste ($z_0 = -3{,}13$) justo en la frontera. Así si el p-valor resulta ser menor que el nivel de significación α fijado de antemano, se rechazará la hipótesis nula. De las tablas de la normal estándar se tiene entonces que

$$p\text{-valor} = P(z < -3{,}13) = 0{,}0009.$$

c) Se quiere hacer un intervalo para la media tal que haya una confianza del 95 % de que

$$\mu \in (\bar{x} \pm 0{,}4)$$

por lo tanto, el número mínimo de tuberías que tenemos que medir para conseguir el intervalo de precisión deseado es:

$$L = z_{\alpha/2}\frac{\hat{s}}{\sqrt{n}} \Longrightarrow n = \left(\frac{z_{\alpha/2}\hat{s}}{L}\right)^2 = \left(\frac{1{,}96 \times 4{,}52}{0{,}4}\right)^2 \approx 491 \text{ tuberías.}$$

2. El trayecto desde una ciudad y su zona industrial viene recorrido por una línea de autobuses. Sea $n = 138$ el número de ciudadanos que trabajan en la zona industrial. Cada mañana todos los trabajadores deciden, de forma independiente con respecto a los demás, si tomar el autobús o ir en su coche. Sea p_i la probabilidad que el i-ésimo trabajador decida de tomar el autobús. Definimos la variable aleatoria X = número de trabajadores que en una mañana toman el autobús. Sea también Y_i la variable aleatoria que vale 1 si el trabajador i-ésimo toma el autobús y 0 en el caso contrario. Se pide:

a) Decir cómo está distribuida la variable aleatoria Y_i.

b) Escribir la variable aleatoria X en función de las Y_i y calcular $\mathbb{E}[X]$ y $\mathbb{V}\text{ar}[X]$ en función de los valores p_i.

Supongamos ahora que todas las p_i sean iguales a $p = 0{,}15$.

c) ¿Cuál es la distribución de X bajo estas hipótesis?

d) Calcular $\mathbb{E}[X]$ y $\mathbb{V}\text{ar}[X]$ en este caso.

e) ¿Cómo se puede aproximar la distribución de X? Calcula $\Pr(X < 28)$.

SOLUCIÓN:

a) Cada variable Y_i es una Bernoulli $Y_i \sim \mathrm{B}(1, p_i)$

b) $X = \sum_{i=1}^n Y_i$, $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[Y_i] = \sum_{i=1}^n p_i$ y, al ser las Y_i independientes se tiene que

$$\mathbb{V}\text{ar}[X] = \mathbb{V}\text{ar}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathbb{V}\text{ar}\,(Y_i) = \sum_{i=1}^n p_i(1-p_i)$$

c) Al ser ahora las Y_i independientes e idénticas se tiene que $X \sim \mathrm{B}(138, 0{,}15)$.

d) Siguiendo las propiedades de la binomial tenemos que $\mathbb{E}[X] = 138 \times 0{,}15 = 20{,}7$ y $\mathbb{V}\text{ar}[X] = 138 \times 0{,}15 \times 0{,}85 = 17{,}595$.

e) Como $np(1-p) = 138 \times 0{,}15 \times (1 - 0{,}15) = 17{,}6 > 5$ se puede utilizar el teorema central del límite y tomar como distribución de X la siguiente normal:

$$X \sim \mathrm{N}(20{,}7; 17{,}595) \tag{1}$$

Por tanto, la aproximación resulta

$$\begin{aligned}\Pr\{X &< 28\} = \Pr\{X < 28\} = \Pr\left(Z < \frac{28 - \mathbb{E}[X]}{\sqrt{\mathbb{V}\text{ar}[X]}}\right) \\ &= \Pr\left(Z < \frac{27 - 20{,}7}{\sqrt{17{,}595}}\right) = \Pr\{Z < 1{,}74\} = 0{,}959\end{aligned}$$

3. Una empresa fabrica chips, para lo que utiliza tres máquinas A, B y C. La máquina A fabrica el 60 % de los chips, la máquina B el 30 % y la máquina C el 10 % restante.

   En la fábrica se producen tres tipos de chips: grandes, medianos y pequeños. Uno de cada dos chips que fabrica la máquina A es grande, uno de cada cuatro mediano y uno de cada cuatro pequeño. Uno de cada tres chips que produce la máquina B es grande, mientras que dos de cada tres son pequeños. Por último, la máquina C sólo produce chips grandes.

   El porcentaje de defectuosos en la producción de la máquina A es del 1 % para los chips grandes, del 2 % para los medianos y del 4 % para los pequeños. Para la máquina B estos porcentajes son del 3 % para los chips grandes y del 1.5 % para los pequeños. Por último, la máquina C produce el 4 % de chips grandes defectuosos. Se pide:

   a) ¿Cuál es la probabilidad de que un chip elegido al azar sea defectuoso?

   b) Se escogió un chip al azar y resultó que no era defectuoso. ¿Cuál es la probabilidad de que se trate de un chip grande?

SOLUCIÓN:

a) Sean los siguientes sucesos: G=chip grande, M=chip mediano, y P=chip pequeño. Llamemos también a los siguientes sucesos A=chip fabricado en la máquina A, B=chip de la máquina B, y C=chip de la máquina C. Por último llamemos d=chip defectuoso. El enunciado nos dice que

$$\Pr(A) = 0{,}6; \Pr(B) = 0{,}3, \Pr(C) = 0{,}1.$$

El porcentaje de artículos defectuosos será

$$\Pr(d) = \Pr(d|A)\Pr(A) + \Pr(d|B)\Pr(B) + \Pr(d|C)\Pr(C).$$

Para calcular $\Pr(d|A)$ =porcentaje de chips defectuosos de la máquina A, hemos de diferenciar si los chips que se producen son G, M o P. Por comodidad llamemos d_A =chip defectuosos de la máquina A, es decir $d_A = d|A$. Llamaremos también G_A, M_A y P_A al tamaño del chip de la máquina A. Entonces, del enunciado tenemos que

$$\begin{aligned} \Pr(\mathrm{G}_A)\text{=Pr} \quad (G|A) =& \quad 0{,}5, \\ \Pr(\mathrm{M}_A)\text{=Pr} \quad (M|A) =& \quad 0{,}25, \\ \Pr(\mathrm{P}_A)\text{=Pr} \quad (P|A) =& \quad 0{,}25, \end{aligned}$$

y por tanto

$$\Pr(d_A) = \Pr(d_A|G_A)\Pr(G_A) + \Pr(d_A|M_A)\Pr(M_A) + \Pr(d_A|P_A)\Pr(P_A),$$

donde, del enunciado tenemos que

$$\Pr(d_A|G_A) = 0{,}01; \Pr(d_A|M_A) = 0{,}02; \Pr(d_A|P_A) = 0{,}04.$$

Por tanto

$$\Pr(d_A) = 0{,}01 \times 0{,}5 + 0{,}02 \times 0{,}25 + 0{,}04 \times 0{,}25 = 0{,}02.$$

Análogamente, para la máquina B tenemos

$$\begin{aligned} \Pr(\mathrm{G}_B)\text{=Pr} \quad (G|B) =& \quad 1/3, \\ \Pr(\mathrm{P}_B)\text{=Pr} \quad (P|A) =& \quad 2/3. \end{aligned}$$

$$\begin{aligned} P(d_B) &= P(d_B|G_B)P(G_B) + P(d_B|P_B)P(P_B) \\ &= 0{,}03 \times 1/3 + 0{,}015 \times 2/3 = 0{,}02 \end{aligned}$$

Para la máquina C tenemos:

$$\Pr(\mathrm{G}_C)=\Pr(G|C) = 1,$$
$$\Pr(d_C) = 0{,}04.$$

Finalmente tenemos

$$\begin{aligned}\Pr(d) &= \Pr(d|A)\Pr(A) + \Pr(d|B)\Pr(B) + \Pr(d|C)\Pr(C)\\ &= 0{,}02 \times 0{,}6 + 0{,}02 \times 0{,}3 + 0{,}04 \times 0{,}1 = 0{,}022.\end{aligned}$$

b) Lo que nos piden es, por el teorema de Bayes

$$\Pr(G|\bar{d}) = \frac{\Pr(\bar{d}|G)\Pr(G)}{\Pr(\bar{d})},$$

donde

$$\begin{aligned}\Pr(\bar{d}) &= 1 - 0{,}022 = 0{,}978\\ P(G) &= P(G|A)P(A) + P(G|B)P(B) + P(G|C)P(C)\\ &= 0{,}5 \times 0{,}6 + 1/3 \times 0{,}3 + 1 \times 0{,}1 = 0{,}5\\ P(\bar{d}|G) &= P(\bar{d}|G \cap A)P(A) + P(\bar{d}|G \cap B)P(B) + P(\bar{d}|G \cap C)P(C)\\ &= 0{,}99 \times 0{,}6 + 0{,}97 \times 0{,}3 + 0{,}96 \times 0{,}1 = 0{,}981.\end{aligned}$$

Se tiene entonces

$$\Pr(G|\bar{d}) = \frac{0{,}981 \times 0{,}5}{0{,}978} = 0{,}501$$

4. El trayecto desde una ciudad y su zona industrial viene recorrido por una línea de autobuses. Sea $n = 138$ el número de ciudadanos que trabajan en la zona industrial. Cada mañana todos los trabajadores deciden, de forma independiente con respecto a los demás, si tomar el autobús o ir en su coche. Sea p_i la probabilidad que el i-ésimo trabajador decida de tomar el autobús. Definimos la variable aleatoria X = número de trabajadores que en una mañana toman el autobús. Sea también Y_i la variable aleatoria que vale 1 si el trabajador i-ésimo toma el autobús y 0 en el caso contrario. Se pide:

a) Decir cómo está distribuida la variable aleatoria Y_i.

b) Escribir la variable aleatoria X en función de las Y_i y calcular $\mathbb{E}[X]$ y $\mathbb{V}\mathrm{ar}[X]$ en función de los valores p_i.

Supongamos ahora que todas las p_i sean iguales a $p = 0{,}15$.

c) ¿Cuál es la distribución de X bajo estas hipótesis?

d) Calcular $\mathbb{E}[X]$ y $\mathbb{V}\mathrm{ar}[X]$ en este caso.

e) ¿Cómo se puede aproximar la distribución de X? Calcula $\Pr(X < 28)$.

SOLUCIÓN:

a) Cada variable Y_i es una Bernoulli $Y_i \sim \mathrm{B}(1, p_i)$

b) $X = \sum_{i=1}^n Y_i$, $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[Y_i] = \sum_{i=1}^n p_i$ y, al ser las Y_i independientes se tiene que

$$\mathbb{V}\mathrm{ar}[X] = \mathbb{V}\mathrm{ar}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathbb{V}\mathrm{ar}\,(Y_i) = \sum_{i=1}^n p_i(1 - p_i)$$

c) Al ser ahora las Y_i independientes e idénticas se tiene que $X \sim \mathrm{B}(138, 0{,}15)$.

d) Siguiendo las propiedades de la binomial tenemos que $\mathbb{E}[X] = 138 \times 0{,}15 = 20{,}7$ y $\mathbb{V}\mathrm{ar}[X] = 138 \times 0{,}15 \times 0{,}85 = 17{,}595$.

e) Como $np(1-p) = 138 \times 0{,}15 \times (1 - 0{,}15) = 17{,}6 > 5$ se puede utilizar el teorema central del límite y tomar como distribución de X la siguiente normal:

$$X \sim \mathrm{N}(20{,}7; 17{,}595) \tag{2}$$

Por tanto, la aproximación resulta

$$\begin{aligned} \Pr\{X &< 28\} = \Pr\{X < 28\} = \Pr\left(Z < \frac{28 - \mathbb{E}[X]}{\sqrt{\mathbb{V}\mathrm{ar}[X]}}\right) \\ &= \Pr\left(Z < \frac{27 - 20{,}7}{\sqrt{17{,}595}}\right) = \Pr\{Z < 1{,}74\} = 0{,}959 \end{aligned}$$

Soluciones del Examen $\boxed{\text{C}}$

1. Verdadero/Falso

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:

_____False Como la muestra es grande, X es una normal por el teorema central del límite.

_____True Si X fuese una Poisson, $\mu = \sigma^2$.

_____True La media muestral tiene una varianza que es sólo la centésima parte de la varianza de X.

_____False Como la muestra es suficientemente grande ($n = 100$), podemos hacer un intervalo de confianza para la varianza aunque X no sea normal.

2. Elige la respuesta correcta

Un contraste de hipótesis bilateral, $H_0 : \mu = \mu_0$ y $H_1 : \mu \neq \mu_0$, con nivel de significación 5 % utilizando una muestra de tamaño 50^2 con media $\bar{x} = 30$ y cuasivarianza $\hat{\sigma}^2 = 15^2$ es equivalente a comprobar que μ_0 pertenezca a un intervalo de confianza:

(A) $IC95\,\%$: $[30 - 0.3z_{0.05}, 30 + 0.3z_{0.05}]$

(B) $IC5\,\%$: $[30 - 0.3z_{0.025}, 30 + 0.3z_{0.025}]$

(C) $IC95\,\%$:$[30 - 0.006z_{0.025}, 30 + 0.006z_{0.025}]$

$\boxed{\text{(D)}}$ $IC95\,\%$: $[30 - 0.3z_{0.025}, 30 + 0.3z_{0.025}]$

3. Elige la respuesta correcta

Un intervalo de confianza al 95 % para la media de una población con varianza σ^2 conocida construido a partir de una muestra de $n_1 = 200$ observaciones tiene amplitud 20. Si tomo una nueva muestra de tamaño $n_2 = 800$ y construyo un intervalo de confianza al 95 % para la media, ¿cuál será la amplitud del nuevo intervalo?

(A) 5

$\boxed{\text{(B)}}$ 10

(C) 80

(D) 40

4. Verdadero/Falso

Sea la ecuación de regresión de mínimos cuadrados $\hat{y} = 10 + 2x$ obtenida a partir de 20 pares de datos. Decir si son verdaderas o falsas las siguientes afirmaciones

_____False La correlación es 2.

_____True El valor previsto de y cuando $x = 5$ es 20.

_____True Tanto la covarianza como la correlación serán positivas.

_____False La regresión de mínimos cuadrados que relaciona x con y será $\hat{x} = -5 + 0.5y$.

5. Elige la respuesta correcta

Una muestra de 100 tornillos tiene media muestral igual a 2.5mm y cuasivarianza 0.09mm^2.

(Desde la tablas de la Normal Estándar se sabe que $z_{0.025} = 1.96$ y $z_{0.05} = 1.65$)

¿Cual será el intervalo de confianza IC95 % para la media μ?:

(A) $\mu \in [2.5 - 1.65 \times 0.03, 2.5 + 1.65 \times 0.03]$ mm

(B) $\mu \in [2.5 - 1.65 \times 0.09, 2.5 + 1.65 \times 0.09]$ mm

$\boxed{\text{(C)}}$ $\mu \in [2.5 - 1.96 \times 0.03, 2.5 + 1.96 \times 0.03]$ mm

(D) $\mu \in [2.5 - 1.65 \times 0.009, 2.5 + 1.65 \times 0.009]$ mm