Universidad Carlos III de Madrid

**Ingeniería Técnica en Informática de Gestión**

**18 de Septiembre de 2007**

**ESTADÍSTICA**

Nota: Justifica todas tus respuestas. Responde cada pregunta en hojas diferentes. Cuando en algún ejecicio haga falta un nivel de significación $\alpha$, utiliza $\alpha = 0.05$, salvo que se especifice expresamente otro valor.

1. Resuelve la siguientes cuestiones:

   (a) Sea $X$ una variable aleatoria discreta que toma los valores $X = \{1, 2, 3, a\}$ y con función de probabilidad $p(x) = x/10$. ¿Qué valor debe tomar $a$?

   (b) Sea $Y$ una variable continua con función de densidad $f(y) = y/18$, con $0 \leq y \leq b$. Calcula el valor de $b$.

   (c) Sean $B_1$ y $B_2$ dos sucesos disjuntos que se observan al realizar un experimento tal que $B_1 \cup B_2 = E$, siendo $E$ el espacio muestral. Al mismo tiempo de observar estos sucesos, se observa el suceso $A$, tal que se conoce que $P(A \cap B_1) = 0.27$ y que $P(A \cap B_2) = 0.23$. Calcula $P(A)$.

   (d) Sean $B_1$ y $B_2$ los sucesos mencionados en la cuestión anterior. Calcula $P(B_2|B_1)$.

   SOLUCIÓN:

   (a) Po definición de función de probabilidad tenemos que

   $$\sum_{i=1}^{4} p(x_i) = \frac{1}{10} + \frac{2}{10} + \frac{3}{10} + \frac{a}{10} = 1 \Rightarrow a = 4.$$

   (b) Por definición de función de densidad tenemos que

   $$\int_0^b \frac{y}{18} dy = \left[\frac{y^2}{36}\right]_0^b = \frac{b^2}{36} = 1 \Rightarrow b = \sqrt{36} = 6.$$

   (c) Por el teorema de la probabilidad total tenemos que $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$. Por otra parte, por la definición de probabilidad condicionada tenemos que

   $$P(A|B_1) = \frac{P(A \cap B_1)}{P(B_1)} \Rightarrow P(A|B_1)P(B_1) = P(A \cap B_1) = 0.27,$$
   $$P(A|B_2) = \frac{P(A \cap B_2)}{P(B_2)} \Rightarrow P(A|B_2)P(B_2) = P(A \cap B_2) = 0.23.$$

   Por tanto

   $$P(A) = 0.27 + 0.23 = 0.5.$$

   Ota forma de resolverlo es viendo que $A = (A \cap B_1) \cup (A \cap B_2)$. Entonces

   $$\begin{aligned} P(A) &= P\left[(A \cap B_1) \cup (A \cap B_2)\right] \\ &= P(A \cap B_1) + P(A \cap B_2) - P\left[(A \cap B_1) \cap (A \cap B_2)\right]. \end{aligned}$$

   Como $(A \cap B_1) \cap (A \cap B_2) = A \cap B_1 \cap B_2 = A \cap \varnothing = \varnothing$, tenemos que

   $$P(A) = P(A \cap B_1) + P(A \cap B_2) = 0.5.$$

(d) Al ser $B_1$ y $B_2$ disjuntos ($P(B_1 \cap B_2) = 0$) se tiene que $P(B_2|B_1) = P(B_2 \cap B_1)/P(B_1) = 0$.

2. Según el Instituto Nacional de Estadística, el 40% de los hogares tiene conexión a internet.

   (a) Si se eligen 10 hogares al azar, ¿cuál es la probabilidad de que haya más de 2 con conexión a internet?

   (b) Si se eligen 1000 hogares al azar, ¿cuál es la probabilidad de que haya más de 400 con conexión a internet?

   (c) Una empresa de estudios de mercado quiere entrevistar a un hogar que tenga conexión a internet. ¿A cuántos hogares tendrá que preguntar por término medio hasta que encuentre uno con conexión a internet (suponiendo que la empresa hace un muestreo aleatorio)?

   **SOLUCIÓN:**

   (a) Sea $X$ el número de hogares con internet en la muestra de tamaño 10. Entonces $X \sim B(10, 0.4)$. Utilizando ña función de probabilidad de la binomial tenemos que

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - \left[ \binom{10}{0} 0.4^0 (1 - 0.4)^{10-0} + \binom{10}{1} 0.4^1 (1 - 0.4)^{10-1} + \binom{10}{2} 0.4^2 (1 - 0.4)^{10-2} \right] \\ &= 1 - [0.6^{10} + 10 \times 0.4 \times 0.6^9 + 45 \times 0.4^2 \times 0.6^8] \\ &= 1 - 0.167 = 0.83. \end{aligned}$$

   (b) Sea $Y$ = número de hogares con internet en una muestra de 1000, donde cada uno tiene una probabilidad de 0.4 de tener internet. Al ser una muestra aleatoria de una población, se cumplen las condiciones de independencia y estabilidad. Se tiene entonces que $Y \sim B(1000, 0.4)$. El cálculo $P(Y > 400)$ utilizando la función de probabilidad de la binomial, tal y como se hizo en el apartado anterior, no es factible si sólo tenemos una calculadora y tiempo limitado. No obstante, como $npq = 1000 \times 0.4 \times 0.6 = 240 > 5$, podemos aproximar la binomial a la normal y usar que

$$Y \sim N(np, npq) = N(400, 240).$$

   Por tanto, utilizando la ssimetía de la normal es inmediato calcular que

$$P(Y > 400) = 0.5.$$

   (c) Sea $U$ = número de entrevistados hasta que uno tenga internet, donde cada entrevistado tiene probabilidad 0.4 de tenerlo. Como el muestreo es aleatorio, podemos considerar que observar el atributo tener/no tener acceso de internet de un entrevistado es independiente de los demás, y que además hay estabilidad en la proporción de personas con acceso a internet. Se tiene entonces que $U$ es una variable aleatoria geométrica $U \sim G(0.4)$ y por tanto el número medio de hogares entrevistados es

$$E(U) = \frac{1}{p} = \frac{1}{0.4} = 2.5 \text{ hogares}$$

3. La entada en vigor de la Ley 28/2005 por la que se prohibe fumar en lugares públicos ha suscitado opiniones muy encontradas. Con la aplicación de la ley, un porcentaje de fumadores ha decidido dejar el tabaco. En el mes siguiente a la entrada en vigor de la ley, se realiza una encuesta a una muestra aleatoria de 150 fumadores, de los que 30 manifestaron su intención de dejar de fumar.

(a) Calcula un intervalo de confianza del 95% para la proporción de fumadores que, en toda la población, quieren dejar de fumar a raíz de la ley.

(b) Calcula el tamaño muestral que se necesita si queremos estimar la proporción de fumadores que quiere dejar de fumar con una precisión de $\pm 0.02$ con un nivel de confianza del 95%.

(c) El CIS ha realizado encuestas antes y después de la implantación de la ley que permiten comparar la incidencia de tabaquismo. Cada encuesta se basa en muestras aleatorias de 1500 personas, siendo ambas muestras independientes. En la primera encuesta, realizada antes de la implantación de la ley, el 25.8% de los encuestados se declaraba fumador. En la segunda encuesta, realizada un año después de implantada la ley, el 23,7% se declaró fumador. ¿Hay evidencia suficiente para afirmar que ha disminuido la incidencia de tabaquismo?

**SOLUCIÓN**:

(a) Sea $p$ la proporción poblacional de fumadores que desea dejar de serlo. Al ser $n\hat{p}\hat{q} = 100 \times 0.2 \times 0.8 = 16 > 5$, podemos hacer infencia sobre $p$ utilizando que la proporción muestral $\hat{p}$ tiene una distribución normal. Bajo este supuesto, el intervalo de confianza de $p$ es

$$\text{IC}(0.95)\colon\ p \in \left\{ \hat{p} \pm z_{0.025}\sqrt{\frac{\hat{p}\hat{q}}{n}} \right\} = \left\{ 0.2 \pm 1.96\sqrt{\frac{0.2 \times 0.8}{150}} \right\} = \{0.2 \pm 0.064\}.$$

Por tanto, tenemos una confianza del 95% de que la proporción de fumadores que desan dejar de serlo están entre el 26.4% y el 13,3%.

(b) Si queremos que el intervalo de confianza del 95% tenga una semilongitud de 0.02, y utilizando la estimación de $\hat{p} = 0.2$, se tendrá que

$$1.96\sqrt{\frac{0.2 \times 0.8}{n}} = 0.02 \Rightarrow n = \left( \frac{1.96\sqrt{0.2 \times 0.8}}{0.02} \right)^2 = 1537 \text{ encuestados.}$$

(Nota: Si no aprovechásemos la estimación inicial de $\hat{p} = 0.2$, deberíamos ponernos en el caso más desfavorable de $p = q = 0.5$. En ese caso el tamaño muestral necesario sería

$$n = \left( \frac{1.96\sqrt{0.5 \times 0.5}}{0.02} \right)^2 = 2401 \text{ encuestados.)}$$

(c) Sea $p_1$ la proporción de individuos fumadores antes de la ley, y $p_2$ después de la ley. Las encuestas estiman que $\hat{p}_1 = 0.258$ y $\hat{p}_2 = 0.237$. Queremos saber si esas diferencias en las muestras se deben a que $p_1 > p_2$ o por el contrario la incidencia de tabaquismo no ha variado y lo que se observa se debe al azar de los individuos encuestados. El ejercicio de inferencia lo realizamos con un contraste unilateral

$$H_0 : p_1 \leq p_2,$$
$$H_1 : p_1 > p_2.$$

Los tamaños muestrales son lo suficientemente grandes y podemos por tanto hacer el contraste basado en la normalidad de los estimadores $\hat{p}_1$ y $\hat{p}_2$. El estadístico de contraste es

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0\hat{q}_0\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}},$$

donde

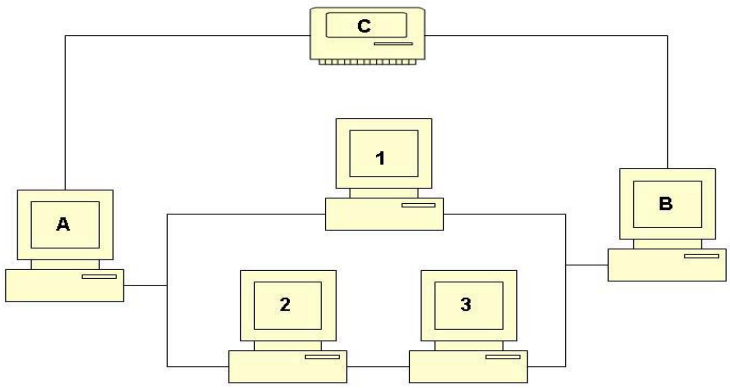
$$\hat{p}_0 = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{0.258 + 0.237}{2} = 0.248.$$

Por tanto

$$z_0 = \frac{0.258 - 0.237}{\sqrt{0.248 \times (1 - 0.248) \times \left(\frac{1}{1500} + \frac{1}{1500}\right)}} = 1.33$$

Rechazamos $H_0$ si $z_0 > z_{0.05} = 1.65$. Por tanto, los resultados de las encuestas no permiten rechazar la igualdad de proporción de fumadores antes y después de la ley. Las diferencias encontradas en las muestras no son significativas y pueden explicarse, con un 5% de significatividad, por la variabilidad asociada al muestreo.

4. Los ordenadores A y B están unidos a través de los ordenadores 1, 2 y 3 mediante una red como la de la figura. Además, están unidos a través del servidor C como se muestra también en la figura. El servidor C funciona con una probabilidad del 95%. Por otra parte, la probabilidad de que cada ordenador, 1, 2 ó 3, esté funcionando es la misma e igual a 90%. El estado funcionando/bloqueado de cada ordenador es independiente de los demás.




Se pide

(a) ¿Qué canal es más fiable para comunicarse A y B, la red formada por 1,2, y 3 o la conexión a través de C? Justifica la respuesta

(b) ¿Que probabilidad tiene de comunicarse A y B si se puede usar cualquier via de comunicación?

**SOLUCIÓN:**

(a) Si llamamos $D_i$, al suceso 'funciona el ordenador i.', i=1, 2 ó 3. Se tiene entonces que $P(D_1) = P(D_2) = P(D_3) = 0.90$. Por otra parte, si llamamos $D_c$ suceso 'funciona el ordenador C' tenemos que $P(D_c) = 0.95$.
Si llamamos $D_{2-3}$ al suceso 'hay comunicación en la rama 2-3', y $D_{AB-123}$ al suceso 'hay comunicación entre A y B usando los ordenadores 1, 2, y 3', se tiene que

$$P(D_{AB-123}) = P(D_1 \cup D_{2-3}) = 1 - P(\bar{D}_1 \cap \bar{D}_{2-3}) = 1 - P(\bar{D}_1)P(\bar{D}_{2-3}) = 1 - 0.1 \times P(\bar{D}_{2-3}).$$

Por otra parte tenemos que

$$P(D_{2-3}) = P(D_2 \cap D_3) = P(D_2)P(D_3) = 0.9^2 = 0.81.$$

Por tanto

$$P(D_{AB-123}\ ) = 1 - 0.1 \times 0.19 = 0.981$$

Por lo tanto, es más fiable usar la ruta formada por los ordenadores 1,2 y 3 que el ordenador $C$

(b) Llamemos $D_{AB}$ al suceso 'hay comunicación entre A y B. Si se pueden usar todos los ordenadores se tiene que

$$\begin{aligned} P(D_{AB}) &= P(D_C \cup D_{AB-123}\ ) = 1 - P(\bar{D}_C \cap \bar{D}_{AB-123}\ ) \\ &= 1 - P(\bar{D}_C)P(\bar{D}_{AB-123}\ ) \\ &= 1 - 0.05 \times 0.019 = 0.99905 \end{aligned}$$