ESTADÍSTICA
Segundo curso de Biología
Junio 2014

Apellidos: **Nombre:** **Grupo:**

(1) (2 puntos) Estamos estudiando el "número de gramos de potasa obtenidos por metro cúbico de serrín" en los residuos de la madera. A partir de una muestra de 10 explotaciones madereras, obtenemos la siguiente información muestral resumida:

$$\sum x_i = 3900 \qquad \sum x_i^2 = 1650000$$

(a) Asumiendo normalidad, halla un intervalo de confianza (al 90%) para el número medio de gramos de potasa por metro cúbico de serrín.
(b) ¿Cuántas explotaciones madereras necesitaríamos muestrear para estimar (al 90%) el número medio de gramos de potasa por metro cúbico de serrín con un error inferior a 30 gramos?

a) $\bar{X} = \frac{1}{10} \Sigma X_i = 390$,

$V_X = \frac{1}{10} \Sigma X_i^2 - (\bar{X})^2 = 12900$ = varianza muestral

Cuasi-varianza muestral:

$S_X^2 = \frac{10}{9} V_X = 14333.\bar{3}$, $S_X = 119'7219$

$I = \left(390 \pm t_{9,0'05} \frac{119'7219}{\sqrt{10}}\right)$

$\approx (390 \pm 1'833 \cdot 37'86)$

$\approx (390 \pm 69'396) \approx (320'603, 459'396)$

b) Hallar n t.q. $t_{n-1, 0'05} \frac{119'7219}{\sqrt{n}} \leq 30$

$\Longleftrightarrow t_{n-1, 0'05} \frac{119'7219}{30} \leq \sqrt{n}$. Como n es desconocido pero "grande" usamos la aproximación $t_{n-1, 0'05} \approx z_{0'05} \approx 1'64 \Rightarrow 6'5448 \leq \sqrt{n}$

$\Rightarrow n \geq 43$.

(2) (2 puntos) El serbal (*Sorbus aucuparia*) es un árbol que crece en un amplio rango de altitudes. Para estudiar cómo estos árboles se adaptan a diferentes hábitats, se recolectaron brotes en ramas de 12 árboles que crecían a diferentes metros de altitud (x_i) y se midió la correspondiente tasa de respiración (y_i) de los brotes recogidos en cada árbol (en μl de O_2 por hora y mg de tejido). A continuación aparecen algunos de los resultados obtenidos:

	Media muestral	Desviación típica muestral
Metros de altitud (x_i)	433.33	214.62
Tasa de respiración (y_i)	0.21	0.08

Además, el coeficiente de correlación muestral es $r = 0.89$.

(a) Calcula la recta de mínimos cuadrados que expresa la tasa de respiración en función de la altitud.
(b) Valora el grado de ajuste de la recta a los datos y predice la tasa de respiración de los brotes recogidos en un árbol situado a 0.5 kilómetros de altitud.

a) $y - \bar{y} = \dfrac{cov_{x,y}}{v_x}(x - \bar{x})$, $v_x = (214.62)^2 = 46061.7444$

$r = \dfrac{cov_{x,y}}{\sqrt{v_x}\sqrt{v_y}} = 0'89$

$cov_{x,y} = 0'89 \cdot 214.62 \cdot 0'08 \simeq 15,281$

$y \simeq 0'21 - 0'1438 + 0'00033175\, x$
$= 0'0662 + 0'00033175\, x$

b) Como $r = 0'89$ está muy proximo a 1, el ajuste es muy bueno.

$y(500) \simeq 0'0662 + 500 \cdot (0'00033175) \simeq 0'232$

(3) (3 puntos) El "número de gramos de potasa obtenidos por metro cúbico de serrín" varía de unas explotaciones madereras a otras y lo modelizamos mediante una $N(\mu = 400; \sigma = 40)$.

(a) Probabilidad de que, en una explotación elegida al azar, la cantidad de potasa sea inferior a 360 gramos por metro cúbico.

(b) Si elegimos 50 explotaciones al azar, ¿cuál es la probabilidad de que en más de 10 la cantidad de potasa por metro cúbico sea inferior a 360 gramos?

(c) Si elegimos 5 explotaciones madereras al azar, ¿cuál es la probabilidad de que la media de las cantidades de potasa de estas 5 explotaciones sea inferior a 360 gramos por metro cúbico?

a) X = nº de gramos de potasa por metro cúbico

$$P(X < 360) = P\left(\frac{X-400}{40} < -1\right) \approx P(Z < -1)$$

$Z \sim N(0,1)$

$$= P(Z > 1) \approx 0'15866$$

b) X_i = nº de gramos en la extracción i-ésima

$$\sum_1^{50} X_i = S_{50} \sim B(50, 0'15866)$$

corrección de De Moivre-Laplace

$$P(S_{50} > 10) = P(S_{50} > 10'5)$$

$$E(S_{50}) = 0'15866 \cdot 50 = 7'933$$

$$Var(S_{50}) = 50\,(0'15866)(1 - 0'15866)$$

$$\approx 6'6744, \quad \sqrt{Var\, S_{50}} \approx 2'5835$$

$$P(S_{50} > 10'5) \approx P\left(Z > \frac{10'5 - 7'933}{2'5835}\right) \approx P(Z > 0'9936)$$

$\approx 0'16109$ (sin corrección, $\approx 0'21186$)

c) Sea $M_5 = \frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)$

Entonces $M_5 \sim N(\mu_M, \sigma_M)$, con $\mu = 400$,

$$Var(S_5) = 5 \cdot (40)^2, \quad Var(M_5) = Var\left(\frac{1}{5} S_5\right)$$

$$= \frac{1}{5^2} \cdot 5 \cdot (40)^2 = 5 \cdot 8^2, \quad \sqrt{Var\, M_5} = 8\sqrt{5}$$

$$P(M_5 < 360) = P\left(Z < \frac{360-400}{8\sqrt{5}}\right)$$

$$\approx P(Z < -\sqrt{5}) = P(Z > \sqrt{5}) \approx P(Z > 2.24)$$

$$\approx 0'01255$$

(4) (3 puntos) El hematocrito es el porcentaje del volumen total de la sangre compuesto por glóbulos rojos y depende tanto del número de glóbulos rojos como de su tamaño. En una muestra de 21 hombres se obtuvo un valor medio de hematocrito de 48.8 y una cuasidesviación típica de 2.8. En otra muestra de 21 mujeres se obtuvo un valor medio de hematocrito de 40.6 y una cuasidesviación típica de 2.9. Asumiendo normalidad:

(a) Calcula un intervalo de confianza de nivel 95% para la varianza del nivel de hematocrito en la población de hombres.

(b) ¿Podemos aceptar igualdad de varianzas en la población de hombres y de mujeres (a nivel 0.10)?

(c) ¿Permiten los datos afirmar (a nivel 0.05) que el nivel medio de hematocrito en la población de hombres es superior al nivel medio en la población de mujeres? Determina razonadamente si el p-valor del contraste es mayor o menor que 0.05.

$n=21 \quad , \quad \mu_H = 48.8 \quad , \quad S_H = 2.8$

$m=21 \quad , \quad \mu_M = 40.6 \quad , \quad S_M = 2.9$

a) $$I = \left(\frac{20\,(2.8)^2}{\chi^2_{20,0'025}} , \frac{20\,(2.8)^2}{\chi^2_{20,0'975}} \right)$$

$$\simeq \left(\frac{156.8}{34.17} , \frac{156.8}{9.591} \right)$$

$$\simeq (4'59 , 16'349)$$

b) $$R = \left\{ \frac{2.8^2}{2.9^2} \notin \left(F_{20,20,0'95} , F_{20,20,0'05} \right) \right\}$$

$H_0 : \sigma_H = \sigma_M$. Como $\left(\frac{2.8}{2.9}\right)^2 \simeq 0'9322$

$$\in \left(\frac{1}{2.124} , 2.124 \right) \simeq (0'4708 , 2'124)$$

aceptamos H_0.

c) $H_1: \mu_H > \mu_M$, $H_0: \mu_H \leq \mu_M$

suponiendo $\sigma_M = \sigma_H$, por b).

$$R = \left\{ 40.6 - 48.8 < t_{40, 0'95} \, S_p \sqrt{\frac{1}{21} + \frac{1}{21}} \right\}$$

$$S_p = \frac{20 \, S_H + 20 \, S_M}{40} = \frac{S_H + S_M}{2} = 2'85$$

Como

$$-8'2 < -1'684 \, \frac{2.85}{\sqrt{21}} \sqrt{2} \simeq -1'4811$$

estamos en la región de rechazo, y por tanto aceptamos $\mu_H > \mu_M$ al nivel de significación 0'05.

Dado que rechazamos, el p-valor es $< \alpha = 0'05$.