EJERCICIO 1

(A) Es un problema de independencia de criterios y se tendrá que construir la tabla de contingencia de frecuencias teóricas (esperadas), n_{ij}^{t}, a partir de las frecuencias observadas, n_{ij}^{o} , que se dan en la tabla del ejercicio.

		COLOR DEL PELO (Y)				
		RUBIO	PELIRROJO	CASTAÑO	OSCURO	NEGRO
SEXO (X)	VARÓN	592	119	849	504	36
	HEMBRA	544	97	677	451	14

El contraste es entonces

$$\begin{cases} H_0 : \text{No existe relación entre el color del pelo (Y) y el sexo de las personas (X)} \\ H_a : \text{Sí existe relación entre el color del pelo (Y) y el sexo de las personas (X)} \end{cases}$$

que siempre es unilateral (de cola superior). A lo largo del ejercicio las unidades de las tablas de contingencia serán "personas". Por otra, parte la tabla con las frecuencias esperadas, n_{ij}^{t}, es

	COLOR DEL PELO (Y)					
SEXO (X)	RUBIO	PELIRROJO	CASTAÑO	OSCURO	NEGRO	
VARÓN	614,3703	116,8169	825,2897	516,4821	27,0409	2100,0000
HEMBRA	521,6297	99,1831	700,7103	438,5179	22,9591	1783,0000
	1136,0000	216,0000	1526,0000	955,0000	50,0000	3883,0000

donde todas las frecuencias teóricas son tales que $n_{ij}^{t} \geq 4$ (hipótesis de normalidad de la población). El estadístico del contraste es la prueba chi cuadrado de Pearson χ^2:

$$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{\left[n_{ij}^{t} - n_{ij}^{o}\right]^2}{n_{ij}^{t}} = 10.4674$$

cuyo cálculo se obtiene de la tabla

	COLOR DEL PELO (Y)					
SEXO (X)	RUBIO	PELIRROJO	CASTAÑO	OSCURO	NEGRO	
VARÓN	0,8145	0,0408	0,6812	0,3017	2,9683	4,8065
HEMBRA	0,9594	0,0481	0,8023	0,3553	3,4960	5,6610
	1,7739	0,0889	1,4835	0,6570	6,4643	10,4674

De las tablas de la distribución χ^2 (que es el modelo de probabilidad que rige en este tipo de problemas siempre y cuando se asuma que la probabilidad de referencia es normal) se deduce que el valor crítico, χ_1^2, que separa la zona de aceptación de la zona de rechazo de la H_0 es:

$$\chi_1^2 = \chi^2_{\alpha,\nu=(c-1)\times(r-1)} \equiv \chi^2_{\alpha,4\,gdl} \;/\; \text{Ï}\,(\chi^2 \leq \chi_1^2) = 1-\alpha$$

que para los distintos noveles de significación que se sugieren se obtiene la siguiente de referencia

α (%)	$\text{Ï}\,(\chi^2 \le \chi_1^2) = 1-\alpha$ (%)	χ_1^2	**¿Existe existencia estadística para rechazar la hipótesis nula?**
25	75	5.385269	SI
10	90	7.779440	SI
5	95	9.487729	SI
1	99	13.276704	**NO**

Resumiendo, a partir de la muestra suministrada SI existe evidencia estadística para rechazar la hipótesis nula; y en consecuencia, existe ALGUNA relación entre el tipo de color del pelo y el sexo de la persona con un nivel de significación $\alpha = 0.01$.

(B) Para calcular el nivel de significación extremo, α_{ext}, de esta prueba, que corresponde a

$$\alpha_{ext} \,/\, \text{Ï}\,(\chi^2 \ge \chi^2_{\alpha_{ext},4\,gdl} = 10.4674) = \alpha_{ext}$$

que, de las tablas, se corresponde con $\alpha_{ext} = 3.3249021\,\%$.

(C) Para obtener el tamaño mínimo de la muestra, para que (suponiendo resultados proporcionales a los observados en la tabla), la prueba de contraste de hipótesis sea válida, hay que tomar como referencia la frecuencia teórica que tiene el valor más pequeño. En este caso se trata de la celda: $n_{ij}^t = n^t_{X=HEMBRA,Y=NEGRO} = 22.9591\ personas$. Para ello se analiza la tabla de probabilidades marginales:

	COLOR DEL PELO (Y)					
SEXO (X)	RUBIO	PELIRROJO	CASTAÑO	OSCURO	NEGRO	
VARÓN	0,1582	0,0301	0,2125	0,1330	0,0070	0,5408
HEMBRA	0,1343	0,0255	0,1805	0,1129	0,0059	0,4592
	0,2926	0,0556	0,3930	0,2459	0,0129	1,0000

de donde se deduce que $n \times p_{X=HEMBRA,Y=NEGRO} \ge 4\ personas$, de donde se deduce que

$$n \ge \frac{4}{0.0059} = 677.9661017$$

Es decir, el tamaño de la muestra al menos ha de ser de n = 678 personas. Si se toma como referencia la condición $n_{ij}^t \ge 5$, análogamente, se deduce la condición n = 848 personas.

EJERCICIO 2

Para las muestras proporcionadas (n = 21)

M1	2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6	3.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7	2.5	4.3	2.3	3.4	3.6
M2	2.9	3.3	3.9	3.1	3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4	4.7	3.8	3.2	2.6	3.9	3.0	4.5	4.2	3.5

se deducen los siguientes valores intermedios

$$\sum_{i=1}^{n} x_{M1,i} = 69.6 \text{ } años; \sum_{i=1}^{n} x_{M2,i} = 73.7 \text{ } años;$$

$$\sum_{i=1}^{n} x_{M1,i}^2 = 242.84 \text{ } años^2; \sum_{i=1}^{n} x_{M2,i}^2 = 267.77 \text{ } años^2$$

La estadística descriptiva proporciona los siguientes resultados:

$$\begin{cases} \overline{x}_{M1} = 3.3143 \text{ años} \\ \hat{s}_{M1} = 0.7799 \text{ años} \end{cases} \qquad \begin{cases} \overline{x}_{M2} = 3.5095 \text{ años} \\ \hat{s}_{M2} = 0.6752 \text{ años} \end{cases}$$

(A) Se pide realizar un contraste de hipótesis del cociente de las desviaciones típicas de dos poblaciones con lo que se tiene que suponer que ambas son normales con el fin de aplicar los métodos paramétricos que se han considerado en el curso, y poder tomar la distribución de probabilidad F de Fisher-Snedecor como modelo de referencia.

El contraste se define bilateral (dos colas) a partir de la pregunta del enunciado:

$$\begin{cases} H_0 : \sigma_{M1} / \sigma_{M2} = 1 \; (\sigma_{M1} = \sigma_{M2}) \\ H_a : \sigma_{M1} / \sigma_{M2} \neq 1 \; (\sigma_{M1} \neq \sigma_{M2}) \end{cases}$$

El estadístico del contraste es:

$$F = \frac{\chi^2_{M2} \big/ \nu_{M2}}{\chi^2_{M1} \big/ \nu_{M1}} = \frac{\dfrac{(n_{M2}-1)\hat{s}^2_{M2}}{\sigma^2_{M2}} \Big/ \nu_{M2}}{\dfrac{(n_{M1}-1)\hat{s}^2_{M1}}{\sigma^2_{M1}} \Big/ \nu_{M1}} = \frac{\hat{s}^2_{M2}}{\hat{s}^2_{M1}} \frac{\cancel{\sigma^2_{M1}}}{\cancel{\sigma^2_{M2}}} = \frac{0.6752}{0.7799} = 0.86575$$

siendo los valores teóricos de referencia

$$F_{\alpha,\nu_{M2},\nu_{M1}} = \begin{cases} F_1 = F_{0.025,20,20} = F^{-1}_{0.975,20,20} = 2.46^{-1} = 0.406504065 \\ F_2 = F_{0.975,20,20} = 2.46 \end{cases}$$

Como $F_1 < F < F_2$, no existe evidencia estadística para rechazar la hipótesis nula con un nivel de significación $\alpha = 5\%$. O equivalentemente, no existe evidencia alguna a partir de las muestras proporcionadas de que las poblaciones de referencia tengan desviaciones típicas diferentes con un nivel de significación $\alpha = 5\%$.

(B) Se trata de un contraste de diferencia de medias (porque hay dos muestras) aritméticas con muestras independientes (porque las baterías implicadas son todas distintas), siendo el modelo de probabilidad la distribución de probabilidad t-Student ya que $n_{M1} = n_{M2} = 21 < 30$.

El contraste se define bilateral (de dos colas):

$$\begin{cases} H_0 : \mu_{M1} - \mu_{M2} = 0 \; (\mu_{M1} = \mu_{M2}) \\ H_a : \mu_{M1} - \mu_{M2} \neq 0 \; (\mu_{M1} \neq \mu_{M2}) \end{cases}$$

El error probable de la estimación es:

$$\sigma_{\mu_{M1}-\mu_{M2}} = \sqrt{\frac{(n_{M1}-1)\hat{s}^2_{M1} + (n_{M2}-1)\hat{s}^2_{M2}}{n_{M1}+n_{M2}-2}\left(\frac{1}{n_{M1}} + \frac{1}{n_{M2}}\right)} = \sqrt{\frac{\hat{s}^2_{M1} + \hat{s}^2_{M2}}{21}}$$

El estadístico del contraste deviene:

$$t = \frac{\left(\overline{x}_{M1} - \overline{x}_{M2}\right) - \cancel{\left(\mu_{M1} - \mu_{M2}\right)}}{\sigma_{\mu_A - \mu_B}} = \frac{3.3143 - 3.5095}{0.225112721} = \frac{-0.1952}{0.225112721} = -0.88275$$

siendo el valor teórico de referencia

$$t_1 = t_{\alpha,\nu} \equiv t_{97.5\%,40\,gdl} = \pm 2.02 \; con \; -t_1 < t < t$$

No existe evidencia estadística para rechazar la hipótesis nula con un nivel de significación $\alpha = 5\%$. Es decir, no existe suficiente evidencia de que las vidas medias estimadas a partir de ambas muestras sean distintas con un nivel de significación $\alpha = 5\%$.

Como conclusión final, se puede afirmar con un nivel de significación $\alpha = 5\%$ que ambas muestras corresponden a la misma población.

EJERCICIO 3

En efecto, se puede demostrar que $\lim\limits_{n\to\infty} \mathsf{P}(\lambda) = \mathsf{N}(\mu,\sigma)\Big|_{\substack{\mu=\lambda \\ \sigma=\sqrt{\lambda}}}$. Para este enunciado se tiene que $\lambda = 43.2\ pacientes$. Entonces, por tratarse de una variable discreta Y de Poisson se tiene que:

Y := número de pacientes que solicitan atención de urgencia en un día determinado

$$\text{Ï}\left(Y=n\right)=\frac{\lambda^n e^{-\lambda}}{n!}$$
$$\text{Ï}\left(Y\le n\right)=\sum_{k=1}^{n}\text{Ï}\left(Y=k\right)$$

Al realizar la aproximación de Poisson (discreta) por la normal (continua) se ha de tener en cuenta la corrección por continuidad.

(A)

$$\text{Ï}_1=\text{Ï}\left(Y=33\ pacientes\right)=\begin{cases} Poisson: & \dfrac{43.2^{33}e^{-43.2}}{33!}=0.018653329\ (resultado\ exacto) \\ Normal: & (ver\ debajo)\ (resultado\ aproximado) \end{cases}$$

Para este caso $\mathsf{N}\left(\mu=\lambda=43.2\ pacientes, \sigma=\sqrt{\lambda}=6.57267069\ pacientes\right)$.

$$\text{Ï}\left(Y=33\ pacientes\right)\underset{\mathsf{N}(\mu=\lambda,\sigma=\sqrt{\lambda})}{\cong}\text{Ï}\left(32.5\le Y\le 33.5\right)=\left(z_1=\frac{32.5-\mu}{\sigma}\le Z\le z_2=\frac{33.5-\mu}{\sigma}\right)=$$

$$\text{Ï}\left(z_1=-1.627953157\le Z\le z_2=-1.475808002\right)=\text{Ï}\left(Z\le z_2=-1.475808002\right)-\text{Ï}\left(Z\le z_1=-1.627953157\right)=$$

$$\text{Ï}_1=0.069997721\text{-}0.051767409\text{=}0.018230312$$

(B) Se pide calcular la probabilidad:

$$\text{Ï}_2=\text{Ï}\left(41\ pacientes<Y\le 53\ pacientes\right)=\text{Ï}\left(Y\le 53\ pacientes\right)-\text{Ï}\left(Y\le 41\ pacientes\right)$$

Si aplicamos Poisson directamente (método exacto) se tiene que

$$\text{Ï}_2=0.93758289-0.407192889\text{=}0.530390001$$

Aplicando la aproximación propuesta en el enunciado del enunciado se deduce que

$$\text{Ï}\left(41 < Y \le 53\right) \underset{\mathrm{N}\,(\mu=\lambda,\sigma=\sqrt{\lambda})}{\cong} \text{Ï}\left(41.5 \le Y \le 53.5\right) = \text{Ï}\left(z_1 = \frac{41.5-\mu}{\sigma} \le Z \le z_2 = \frac{53.5-\mu}{\sigma}\right) =$$

$$\text{Ï}\left(z_1 = \text{-}0.258646763 \le Z \le z_2 = 1.567095095\right) = \text{Ï}\left(Z \le z_2 = 1.567095095\right) - \text{Ï}\left(Z \le z_1 = \text{-}0.258646763\right) =$$
$$\text{Ï}_2 = 0.941453767\text{-}0.397953899\text{=}0.543499868$$

(C) La probabilidad de que el servicio de urgencia quede colapsado es $\text{Ï}_3 = \text{Ï}\left(Y > 50 \; pacientes\right) = 1 - \text{Ï}\left(Y \le 50 \; pacientes\right)$. Si aplicamos Poisson directamente (método exacto) se tiene que

$$\text{Ï}_3 = 1 - 0.86566799\text{=}0.13433201$$

Aplicando la aproximación propuesta en el enunciado del enunciado se deduce que

$$\text{Ï}\left(Y > 50\right) \underset{\mathrm{N}\,(\mu=\lambda,\sigma=\sqrt{\lambda})}{\cong} \text{Ï}\left(Y \ge 50.5\right) = \text{Ï}\left(Z \ge z_1 = \frac{50.5-\mu}{\sigma}\right) = \text{Ï}\left(Z \ge z_1 = 1.110659630\right)$$

$$\text{Ï}_3 = 1 - \text{Ï}\left(Z \le z_1 = 1.110659630\right) = 0.133357442$$

EJERCICIO 4

En lo que sigue A denota la población de los niños a los que se les suministro la medicina y B la de los que no se les suministró.

$$\pi_A \equiv \hat{\pi}_A = p_A = \frac{29}{120} = 0.2417\,;\ \pi_B \equiv \hat{\pi}_B = p_B = \frac{56}{280} = 0.2000$$

(A) Del enunciado del ejercicio el contraste se establece en términos de un contraste unilateral de cola inferior (ya que según el enunciado "…determinar si la nueva medicina de la compañía era efectiva …"):

$$\begin{cases} H_0: & \pi_{A,0} - \pi_{B,0} = 0\ (\pi_{A,0} = \pi_{B,0}) \\ H_a: & \pi_{A,0} - \pi_{B,0} < 0\ (\pi_{A,0} < \pi_{B,0}) \end{cases}$$

donde se aplica el modelo de probabilidad normal ya que es

$$\begin{array}{c} n_A p_A = 29 > 4 \\ n_A q_A = 91 > 4 \end{array} \wedge \begin{array}{c} n_B p_B = 56 > 4 \\ n_B q_B = 224 > 4 \end{array}$$

El error probable (desviación típica) de la estimación es:

$$\sigma_{\pi_A - \pi_B} = \sqrt{\frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}} = \sqrt{\frac{0.2417 \times 0.7583}{120} + \frac{0.2 \times 0.8}{280}} =$$
$$\sqrt{1.527342583 \times 10^{-3} + 5.714285714 \times 10^{-4}} = \sqrt{1.527342583 \times 10^{-3} + 5.714285714 \times 10^{-4}} =$$
$$= \sqrt{2.098771154 \times 10^{-3}} = 0.04581234718$$

El estadístico del contraste es

$$z = \frac{(p_A - p_B) - \cancel{(\pi_{A,0} - \pi_{B,0})}}{\sigma_{\pi_A - \pi_B}} = \frac{0.2417 - 0.2}{0.04581234718} = 0.9102349599$$

El valor que separa la región crítica de la zona de admisibilidad para un nivel de significación $\alpha = 2.5\%$ es

$$z_\alpha \equiv z_{2.5\%} = -1.959963985$$

utilizando la simetría de la distribución normal. Como $z_\alpha < z$, de la muestra proporcionada no existe evidencia estadística para rechazar la hipótesis nula; es decir, no existe evidencia suficiente de que la nueva medicina de la compañía sea efectiva en un plazo de dos días con un nivel de significación del 2.5 %.

(B) Para realizar la estimación confidencial (intervalar) propuesta se ha de trabajar tan solo con la población A, de donde se deduce que el error probable asociado será:

$$\sigma_{\hat{\pi}_A} = \sqrt{\frac{p_A q_A}{n_A}} = \sqrt{\frac{0.2417 \times 0.7583}{120}} = \sqrt{1.527342583 \times 10^{-3}} = 0.03908123373$$

Para el nivel de confianza propuesto se deduce que $z_\alpha \equiv z_{97.5\%} = \pm 1.959963985$ y el intervalo de confianza pedido es

$$\left[l_{\pi_A}, L_{\pi_A}\right] = \left[\hat{\pi}_A - z_\alpha \sigma_{\hat{\pi}_A}, \hat{\pi}_A + z_\alpha \sigma_{\hat{\pi}_A}\right] = \left[0.1651021894, 0.3182978106\right]$$

(C) Dado que la fórmula del error probable es

$$\sigma_{\hat{\pi}_A} = \sqrt{\frac{p_A q_A}{n_A}} = \sqrt{\frac{0.2417 \times 0.7583}{n_A}} \geq 0.01$$

de donde se sigue que

$$n_A \geq \frac{p_A q_A}{0.01^2} = \frac{0.2417 \times 0.7583}{0.01^2} = 1832.8111$$

de donde se sigue que n = 1833 personas (como mínimo).

EJERCICIO 5

El modelo de probabilidad que interviene es la distribución χ^2 de Pearson para la estimación/contraste de desviaciones típicas con n = 10 latas con (es decir, una población):

$$(n-1)\hat{s}^2 = 9\times 4.8^2 = 207.36\; mg^2$$

(A) En este caso se trata de un contraste de hipótesis de desviación típica de una sola poblacion, unilateral (de cola superior), según se deduce del enunciado; es decir:

$$\begin{cases} H_0: & \sigma_0 = \sqrt{18}\; mg \\ H_a: & \sigma > \sqrt{18}\; mg \end{cases}$$

El estadístico del contraste es

$$\chi^2 = \frac{(n-1)\hat{s}^2}{\sigma_0^2} = \frac{9\times 4.8^2}{18} = \frac{207.36}{18} = 11.52$$

La región de aceptación χ_1^2 para el contraste propuesto es

$$\chi_1^2 = \chi^2_{\alpha=95\%,\, \nu=9\; gdl} = 16.92$$

que comparado con el valor teórico de la distribución $\chi^2 < \chi_1^2$ (dentro de la región de admisibilidad), permite concluir que a partir de la muestra presentada no existe evidencia estadística para rechazar la hipótesis nula, y por lo tanto no existe evidencia suficiente para decir que Se desea que la desviación típica del contenido de azúcar no supere los $\sqrt{18}$ mg con un nivel de significación α = 5%.

(B) El intervalo de confianza para la desviación típica poblacional es:

$$[l_\sigma, L_\sigma] = \left[\sqrt{\frac{(n-1)\hat{s}^2}{\chi_2^2}}, \sqrt{\frac{(n-1)\hat{s}^2}{\chi_1^2}}\right]$$

donde

$$\chi^2_{\alpha,\nu} = \begin{cases} \chi^2_{2.5,9\; gdl} = 2.70 = \chi_1^2 \\ \chi^2_{97.5,9\; gdl} = 19.2 = \chi_2^2 \end{cases}$$

con lo que se obtiene el intervalo solicitado

$$[l_\sigma, L_\sigma] = \left[\sqrt{\frac{207.36}{19.2}}, \sqrt{\frac{207.36}{2.70}}\right] = [3.2863\; mg, 8.7635\; mg]$$

<u>EJERCICIO 6</u>

PARTE A:

La probabilidad de aceptación para un plan de muestreo PM(n, a) viene dada por:

$$\text{Ï}_{\text{aceptacion}} = \text{Ï}\left(Y \le a = 3\right) = \sum_{k=0}^{3} \text{Ï}\left(Y = k\right) \text{ siendo } \text{Ï}\left(Y = k\right) = \binom{60}{k} p^k (1-p)^{60-k}$$

Como se tiene que n = 60 no podemos utilizar las tablas de la distribución binomial que se tienen, ni utilizar la fórmula correspondiente (porque aparecen factoriales de números grandes), y hay que recurrir a aproximaciones.

Caso 1: Como $np = 60 \times 0.05 = 3 < 4$, entonces $\lim\limits_{\substack{n\to\infty \\ p\to 0}} \mathrm{B}(n,p) = \mathrm{P}(\lambda = np)$ con $\lambda = np = 3$ con lo que

$$\text{Ï}_{\text{aceptacion}}\Big|_{p=0.05} = \text{Ï}\left(Y \le 3\right) \underset{\substack{tablas \\ Poisson}}{=} 64.70\,\%$$

Caso 2: Ahora $\begin{matrix} np = 60 \times 0.20 = 12 > 4 \\ nq = 60 \times 0.80 = 48 > 4 \end{matrix}$, en consecuencia se ha de utilizar la aproximación:

$$\lim_{n\to\infty} \mathrm{B}(n,p) = \mathrm{N}\left(\mu = np, \sigma = \sqrt{npq}\right) \text{ tomando } \mu = 12, \sigma = \sqrt{npq} = 3.09839$$

con la corrección por continuidad. Así:

$$\text{Ï}_{\text{aceptacion}}\Big|_{p=0.20} = \text{Ï}\left(Y \le 3\right) \approx \text{Ï}\left(Y \le 3.5\right) = \text{Ï}\left(Z \le z_1 = \frac{3.5-12}{3.09839} = -2.743360261\right) =$$

$$\text{Ï}_{\text{aceptacion}}\Big|_{p=0.20} = 1 - \text{Ï}\left(Z \le z_1 = 2.743360261\right) = 1 - 0.997050653 = 0.2949347\ \%$$

PARTE B:

(A) El plan de muestreo es del tipo PM(n = 20, a) con probabilidad de aceptación

$$\text{Ï}_{\text{aceptacion}}\Big|_{p=0.05} = 70\% \le \text{Ï}_{\text{aceptacion}} = \text{Ï}\left(Y \le a\right) = \sum_{k=0}^{a} \text{Ï}\left(Y = k\right) \le \text{Ï}_{\text{aceptacion}}\Big|_{p=0.05} = 75\%$$

De las tablas de la distribución binomial para n = 20 y p = 0.05 se deduce que **a = 1** porque $\text{Ï}_{\text{aceptacion}}\Big|_{\substack{p=0.05\\ a=0}} = \text{Ï}\,\left(Y \leq 0\right) = 35.85\%$ (que se queda corto) y $\text{Ï}_{\text{aceptacion}}\Big|_{\substack{p=0.05\\ a=1}} = \text{Ï}\,\left(Y \leq 1\right) = 73.58\%$ (que se encuentra en el intervalo dado).

(B)

$$\text{Ï}_{\text{aceptacion}}\Big|_{\substack{n=20\\ a=1\\ p=0.02}} = \text{Ï}\,\left(Y \leq 1\right) = \sum_{k=0}^{1}\binom{20}{k}0.02^{k}(1-0.02)^{20-k} = \binom{20}{0}0.02^{0}\times 0.98^{20} + \binom{20}{1}0.02^{1}\times 0.98^{19}$$

$$= 0.98^{20} + 20\times 0.02\times 0.98^{19} = 94.01010215\%$$

(C) Del apartado (A) se deduce que la probabilidad de aceptación exigida se logra con una tasa de defectos p = 5 %, con lo que de esperar que se encuentren como máximo 50 piezas defectuosas.