Soluciones a los ejercicios propuestos del Tema 5

5.1. Vamos a aplicar el test de independencia de la χ^2 a las variables "color de ojos" y "mioía". Para ello usamos la siguiente tabla:

	Color de ojos			
Miopía	Azul	Verde	Castaño	$f_{i\cdot}$
No	75 / 72.85636	103 / 103.04909	183 / 185.09455	361
Sí	36 / 38.14364	54 / 53.95091	99 / 96.90545	189
$f_{\cdot j}$	111	157	282	n=550

El valor del estadístico D es $d = \frac{(75-72.85636...)^2}{72.85636...} + \cdots = 0.25258 < \chi^{2,2}_{0.95} = 5.99$, luego **no** se puede decir que haya relación entre las dos variables. **Nota:** hay que hacer los cálculos con todos los decimales que se pueda para evitar la propagación de errores al hacer los cuadrados y sumar.

5.2. Aplicar el test de independencia de la χ^2 a las variables "sexo" y "facultad de gustar". Para ello usamos la siguiente tabla:

	Sexo		
Facultad de gustar	Niño	Niña	Suma
Sí	a=60	b=40	$N_{x_1} = 100$
No	c=40	d=10	$N_{x_2} = 50$
Suma	$N_{y_1} = 100$	$N_{y_2} = 50$	n=150

El valor del estadístico $\tilde{D}$ (con la corrección de Yates de continuidad) es

$$\tilde{d} = \frac{150\left(|60 \times 10 - 40 \times 40| - \frac{150}{2}\right)^2}{100 \times 50 \times 100 \times 50} = 5.13375 > \chi^{2,1}_{0.95} = 3.84\,.$$

Luego sí podemos decir que hay relación. Las niñas "gustan" mejor la sustancia, pues el porcentaje de niñas que la detectan es mayor que el de niños (son, respectivamente, un 80 y un 60 %).
5.3.(a) % de tratados con A de entre los de talla mediana: $\frac{340}{540} \times 100 = 62.96\,\%$, % de talla grande de entre los tratados con A: $\frac{250}{700} \times 100 = 35.71\,\%$.
5.3.(b) Usamos el test de independencia de la χ^2 para las variables "talla" y "tratamiento", para lo que necesitamos la siguiente tabla (en ella redondeamos al quinto decimal, pero en realidad hemos de trabajar con todos los decimales que podamos para calcular el valor del estadístico, ya que si no los errores se propagan).

	Talla del perro			
Trat.	Pequeña	Mediana	Grande	$f_{i\cdot}$
A	110 / 92.89100	340 / 358.29384	250 / 248.8152	700
B	25 / 36.49289	150 / 140.75829	100 / 97.74882	275
C	5 / 10.61611	50 / 40.94787	25 / 28.43602	80
$f_{\cdot j}$	140	540	375	n=1055

El valor del estadístico D es

$$d = \frac{(110 - 92.890995\ldots)^2}{92.890995\ldots} + \cdots = 13.75635 > \chi^{2,4}_{0.99} = 13.28,$$

luego con $\alpha = 0.01$ **sí** se puede decir que hay relación entre las dos variables, aunque un "poco ajustado". Con $\alpha = 0.05$ también porque $0.05 > 0.01$ (el test es más arriesgado), y de hecho con "más holgura", ya que $\chi^{2,4}_{0.95} = 9.49$.
5.4. Como

$$Q = 0.13/0.23 = 0.56522 > Q_{0.95}(6) = 0.560\,,$$

con $\alpha = 0.05$ sí podemos decir que 2.07 es un "outlier". Como $Q_{0.99}(6) = 0.698 > 0.560$, con $\alpha = 0.01$ no podemos decirlo.
5.5. Como

$$Q = 0.2/0.29 = 0.68966 > Q_{0.95}(5) = 0.642\,,$$

con $\alpha = 0.05$ sí podemos decir que el valor 12.27 es un "outlier". Con $\alpha = 0.01$, no podemos decirlo, puesto que $Q_{0.99}(5) = 0.780 > Q$.

5.6. Como

$$Q = 0.99/1.02 = 0.97059 > Q_{0.99}(5) = 0.780\,,$$

sí que podemos decir que 13.14 es un "outlier".

5.7. Pareja "sospechosa": 0.421 y 0.423.

Para 0.423: calculamos el valor del estadístico $Q' = 0.020/0.026 = 0.76923 > Q'_{0.99}(8) = 0.710$. Entonces se puede decir que 0.423 es un "outlier" con $\alpha = 0.01$, luego también con $\alpha = 0.05$.

Para 0.421, eliminando 0.423, se obtiene: $Q = 0.018/0.024 = 0.75 > Q_{0.99}(7) = 0.637$; por tanto se puede decir que es un "outlier" con $\alpha = 0.01$, luego también con $\alpha = 0.05$. Así que podemos decir que ambos lo son.

5.8. Vamos a hacer el test de normalidad de Shapiro-Wilk con $\alpha = 0.05$ y $n = 11$ para ver si podemos decir que los datos no provienen de una normal:

i	$u_{(i)}$	$u_{(n-i+1)}$	$u_{(n-i+1)} - u_{(i)}$	a_{n-i+1}	$a_{n-i+1}\left(u_{(n-i+1)} - u_{(i)}\right)$
1	148	236	88	0.5601	49.2888
2	154	195	41	0.3315	13.5915
3	158	182	24	0.2260	5.424
4	160	170	10	0.1429	1.429
5	161	165	4	0.0695	0.278

La suma de la última columna es 70.0113, y $\sum_{i=1}^{11}(u_i - \bar{u})^2 = 6238.909093$. Entonces, el valor del estadístico es $W = \frac{(70.0113)^2}{6238.909093} = 0.785647$ menor que valor tabulado $W_{0.05}^{11} = 0.850$. Por tanto, sí se puede decir que **no** provienen de una normal.

5.9. Vamos a hacer el test de Kolmogorov-Smirnov para ver si podemos decir que los datos no provienen de una normal:

i	$u_{(i)}$	$\frac{u_{(i)}-80}{6}$	$F_0(u_{(i)})$	$\mid F_0(u_{(i)}) - \frac{i}{n} \mid$	$\mid F_0(u_{(i)}) - \frac{i-1}{n} \mid$
1	68	-2	0.02275	$0.00502\overset{\frown}{7}$	0.02275
2	68	-2	0.02275	$0.03280\overset{\frown}{5}$	$0.00502\overset{\frown}{7}$
3	72	$-1.\overset{\frown}{3}$	0.09176	$0.00842\overset{\frown}{6}$	$0.03620\overset{\frown}{4}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
26	81	$0.1\overset{\frown}{6}$	0.56749	$\boxed{0.15473\overset{\frown}{2}}$	$0.12695\overset{\frown}{4}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
36	92	2	0.97725	0.02275	$0.00502\overset{\frown}{7}$

El valor del estadístico es $K = 0.15473\overset{\frown}{2}$, que es menor que valor tabulado $\Delta^{36}_{0.95}$, que está entre 0.210 y 0.230. Por tanto, **no** se puede decir que las observaciones **no** provienen de la normal.

5.10. Hacemos el test de bondad de ajuste (de la χ^2) a que las proporciones de homicidios en las 4 estaciones son iguales a 1/4.

Categoría	O_i	E_i	$\frac{(O_i-E_i)^2}{E_i}$
Primavera	334	340.25	0.1148052902
Verano	372	340.25	2.962711242
Otoño	327	340.25	0.5159808964
Invierno	328	340.25	0.4410360029
Total	1361	1361	$d = 4.03453$

Entonces el valor del estadístico es $d = 4.03453 < \chi^{2,3}_{0.95} = 7.81$, luego **no** tenemos suficiente evidencia como para poder decir que haya relación entre la estación del año y la incidencia de homicidios.

5.11. Haremos el test de bondad de ajuste (de la χ^2) a que las proporciones son las dadas por la teoría, usando la tabla:

Categoría	O_i	E_i	$\frac{(O_i-E_i)^2}{E_i}$
Rojas	195	207	0.6956521739
Amarillas	73	69	0.231884058
Blancas	100	92	0.6956521739
Total	368	368	$d = 1.623188406$

El estadístico es $d = 1.623188406 < \chi^{2,2}_{0.95} = 5.99$. No se puede decir que los datos contradigan la teoría.

5.12. Hacemos el test de bondad de ajuste (de la χ^2) a una Poisson, tomando como estimación del parámetro $\hat{\lambda} = \bar{x} = 101/48$, y teniendo en cuenta que las frecuencias esperadas se obtienen así: $E_i = e^{-\hat{\lambda}} \frac{\hat{\lambda}^i}{i!}$, ayudándonos con la tabla:

Categoría	O_i	E_i	$\frac{(O_i-E_i)^2}{E_i}$
0	9	5.853468226	1.691418121
1	9	12.31667272	0.8931241563
2	10	12.95816609	0.6753074901
3	14	9.088713716	2.653921525
"≥ 4"	6	7.78297925	0.4084573406
Total	48	48	$d = 6.322228633$

El estadístico del test es $d = 6.322228633 < \chi^{2,3}_{0.95} = 7.81$. No hay evidencia como para poder decir que los datos no provienen de una Poisson, así que **no** podemos decir que **no** provienen de una Poisson.

5.13. Hacemos el test de homogeneidad (de la χ^2), para $s = 5$ poblaciones, usando la siguiente tabla:

	Variedad de gramínea				
	1	2	3	4	5
Germinadas	30	55	85	12	55
No germinadas	10	33	51	20	45
Total	$n_1 = 40$	$n_2 = 88$	$n_3 = 136$	$n_4 = 32$	$n_5 = 100$

Tenemos que

$$\hat{p}_1 = \frac{30}{40},\ \hat{p}_2 = \frac{55}{88},\ \hat{p}_3 = \frac{85}{136},\ \hat{p}_4 = \frac{12}{32},\ \hat{p}_5 = \frac{55}{100}$$

y también $\hat{p} = \frac{n_1\,\hat{p}_1 + \cdots + n_5\,\hat{p}_5}{n_1 + \cdots + n_5} = \frac{237}{396} = 0.59\overset{\frown}{84}$. El estadístico es:

$$d = \frac{1}{\hat{p}\,(1-\hat{p})} \sum_{j=1}^{5} n_j\,(\hat{p}_j - \hat{p})^2 = \frac{2.\overset{\frown}{90}}{0.59\overset{\frown}{84} \times (1 - 0.59\overset{\frown}{84})} = 12.10604251$$

y es $> \chi^{2,4}_{0.95} = 9.49$. Entonces, con una probabilidad de equivocarnos de un 5 % podemos decir que no hay diferencias, en cuanto a la germinación se refiere, entre las 5 variedades de gramínea.