Soluciones a los ejercicios propuestos del Tema 5

5.1. Vamos a aplicar el test de independencia de la $\chi^2$ a las variables “color de ojos” y “miopía”. Para ello usamos la siguiente tabla:

<table>
<thead>
<tr>
<th>Miopía</th>
<th>Color de ojos</th>
<th>Castaño</th>
<th>$f_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>75 / 72.85636</td>
<td>103 / 103.04909</td>
<td>183 / 185.09455</td>
</tr>
<tr>
<td>Sí</td>
<td>36 / 38.14364</td>
<td>54 / 53.95091</td>
<td>99 / 96.90545</td>
</tr>
<tr>
<td>$f_j$</td>
<td>111</td>
<td>157</td>
<td>282</td>
</tr>
</tbody>
</table>

El valor del estadístico $D$ es $d = \frac{(75-72.85636)^2 + \cdots}{72.85636} = 0.25258 < \chi^2_{0.95} = 5.99$, 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 $\chi^2$ a las variables “sexo” y “facultad de gustar”. Para ello usamos la siguiente tabla:

<table>
<thead>
<tr>
<th>Sexo</th>
<th>Facultad de gustar</th>
<th>Niño</th>
<th>Niña</th>
<th>Suma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sí</td>
<td>a=60</td>
<td>b=40</td>
<td></td>
<td>$N_{x_1} = 100$</td>
</tr>
<tr>
<td>No</td>
<td>c=40</td>
<td>d=10</td>
<td></td>
<td>$N_{x_2} = 50$</td>
</tr>
</tbody>
</table>

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

$$\tilde{d} = \frac{150 \left( |60 \times 10 - 40 \times 40| - \frac{150 \times 150}{2} \right)^2}{100 \times 50 \times 100 \times 50} = 5.13375 > \chi^2_{0.95} = 3.84.$$
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: \( \frac{340}{740} \times 100 = 62.96\% \), % 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 \( \chi^2 \) 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).

<table>
<thead>
<tr>
<th>Talla del perro</th>
<th>Trat.</th>
<th>Pequeña</th>
<th>Mediana</th>
<th>Grande</th>
<th>( f_i ).</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>110 / 92.89100</td>
<td>340 / 358.29384</td>
<td>250 / 248.8152</td>
<td>700</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>25 / 36.49289</td>
<td>150 / 140.75829</td>
<td>100 / 97.74882</td>
<td>275</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>5 / 10.61611</td>
<td>50 / 40.94787</td>
<td>25 / 28.43602</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>( f_j )</td>
<td>140</td>
<td>540</td>
<td>375</td>
<td>n=1055</td>
<td></td>
</tr>
</tbody>
</table>

El valor del estadístico \( D \) es

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

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 \( \chi^2_{0.95} = 9.49 \).

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:

<table>
<thead>
<tr>
<th>$i$</th>
<th>$u_i$</th>
<th>$u_{(n-i+1)}$</th>
<th>$u_{(n-i+1)} - u_i$</th>
<th>$a_{n-i+1}$</th>
<th>$a_{n-i+1}(u_{(n-i+1)} - u_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>148</td>
<td>236</td>
<td>88</td>
<td>0.5601</td>
<td>49.2888</td>
</tr>
<tr>
<td>2</td>
<td>154</td>
<td>195</td>
<td>41</td>
<td>0.3315</td>
<td>13.5915</td>
</tr>
<tr>
<td>3</td>
<td>158</td>
<td>182</td>
<td>24</td>
<td>0.2260</td>
<td>5.424</td>
</tr>
<tr>
<td>4</td>
<td>160</td>
<td>170</td>
<td>10</td>
<td>0.1429</td>
<td>1.429</td>
</tr>
<tr>
<td>5</td>
<td>161</td>
<td>165</td>
<td>4</td>
<td>0.0695</td>
<td>0.278</td>
</tr>
</tbody>
</table>

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 $W_{11}^{0.05} = 0.850$. 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:

| $i$ | $u(i)$ | $\frac{u(i) - 80}{6}$ | $F_0(u(i))$ | $| F_0(u(i)) - \frac{i}{n} |$ | $| F_0(u(i)) - \frac{i-1}{n} |$ |
|-----|--------|----------------------|-------------|------------------|-------------------|
| 1   | 68     | -2                   | 0.02275     | 0.005027         | 0.02275           |
| 2   | 68     | -2                   | 0.02275     | 0.032805         | 0.005027           |
| 3   | 72     | -1.3                 | 0.09176     | 0.008426         | 0.036204          |
|     |       |                      |             |                  |                   |
| 26  | 81     | 0.16                 | 0.56749     | 0.154732         | 0.126954           |
|     |       |                      |             |                  |                   |
| 36  | 92     | 2                    | 0.97725     | 0.02275          | 0.005027           |

El valor del estadístico es $K = 0.154732$, que es menor que valor tabulado $\Delta_{0.05}^{36}$, 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 $\chi^2$) a que las proporciones de homicidios en las 4 estaciones son iguales a $1/4$.

<table>
<thead>
<tr>
<th>Categoría</th>
<th>$O_i$</th>
<th>$E_i$</th>
<th>$(O_i - E_i)^2 / E_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primavera</td>
<td>334</td>
<td>340.25</td>
<td>0.1148052902</td>
</tr>
<tr>
<td>Verano</td>
<td>372</td>
<td>340.25</td>
<td>2.962711242</td>
</tr>
<tr>
<td>Otoño</td>
<td>327</td>
<td>340.25</td>
<td>0.5159808964</td>
</tr>
<tr>
<td>Invierno</td>
<td>328</td>
<td>340.25</td>
<td>0.4410360029</td>
</tr>
<tr>
<td>Total</td>
<td>1361</td>
<td>1361</td>
<td>$d = 4.03453$</td>
</tr>
</tbody>
</table>

Entonces el valor del estadístico es $d = 4.03453 < \chi^2_{0.05}^{2.3} = 7.81$, 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 $\chi^2$) a que las proporciones son las dadas por la teoría, usando la tabla:

<table>
<thead>
<tr>
<th>Categoría</th>
<th>$O_i$</th>
<th>$E_i$</th>
<th>$\frac{(O_i-E_i)^2}{E_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rojas</td>
<td>195</td>
<td>207</td>
<td>0.6956521739</td>
</tr>
<tr>
<td>Amarillas</td>
<td>73</td>
<td>69</td>
<td>0.231884058</td>
</tr>
<tr>
<td>Blancas</td>
<td>100</td>
<td>92</td>
<td>0.6956521739</td>
</tr>
<tr>
<td>Total</td>
<td>368</td>
<td>368</td>
<td>$d = 1.623188406$</td>
</tr>
</tbody>
</table>

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

5.12. Hacemos el test de bondad de ajuste (de la $\chi^2$) 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:

<table>
<thead>
<tr>
<th>Categoría</th>
<th>$O_i$</th>
<th>$E_i$</th>
<th>$\frac{(O_i-E_i)^2}{E_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
<td>5.853468226</td>
<td>1.691418121</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>12.31667272</td>
<td>0.8931241563</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>12.95816609</td>
<td>0.6753074901</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>9.088713716</td>
<td>0.653921525</td>
</tr>
<tr>
<td>“≥ 4”</td>
<td>6</td>
<td>7.78297925</td>
<td>0.4084573406</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>48</td>
<td>$d = 6.322228633$</td>
</tr>
</tbody>
</table>

El estadístico del test es $d = 6.322228633 < \chi^2_{0.95} = 7.81$. 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 $\chi^2$), para $s = 5$ poblaciones, usando la siguiente tabla:
Tenemos que

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

y también

$$\hat{p} = \frac{n_1 \hat{p}_1 + \cdots + n_5 \hat{p}_5}{n_1 + \cdots + n_5} = \frac{237}{396} = 0.5984.$$ El estadístico es:

$$d = \frac{1}{\hat{p} (1 - \hat{p})} \sum_{j=1}^{5} n_j (\hat{p}_j - \hat{p})^2 = \frac{2.90}{0.5984 \times (1 - 0.5984)} = 12.10604251$$

y es > $\chi^2_{0.05} = 9.49$. 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.