(b) Using \( m = b + c = 8 + 26 = 34 \), the large sample test statistic is

\[
z = \frac{b - c - 1}{\sqrt{b + c}} = \frac{8 - 26 - 1}{\sqrt{8 + 26}} = -3.258.
\]

Then the \( P \)-value is

\[
P = 2(1 - \Phi(|-3.258|)) = 2 \times 0.0006 = 0.0012.
\]

Since \( P < \alpha = 0.05 \), reject \( H_0 \) and conclude that there was a change in opinion of the students.

**Solutions to Section 9.3**

9.17  \( H_0 : p_1 = p_2 = \ldots = p_8 = 1/8 \) vs. \( H_1 : \) Not \( H_0 \). The expected frequencies in each case are \( np = 144 \times 1/8 = 18 \). Then

\[
\chi^2 = \sum \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}} = \frac{(29 - 18)^2}{18} + \frac{(19 - 18)^2}{18} + \ldots + \frac{(11 - 18)^2}{18} = 16.333.
\]

Since \( \chi^2 > \chi^2_{8-1,0.05} = 14.067 \), reject \( H_0 \) and conclude that the horse’s chances of winning are not the same for each starting gate.

9.18  \( H_0 : p_1 = p_2 = \ldots = p_{12} = 1/12 \) vs. \( H_1 : \) Not \( H_0 \). The expected frequencies in each case are \( np = 700 \times 1/12 = 58.333 \). Then

\[
\chi^2 = \sum \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}} = \frac{(66 - 58.333)^2}{58.333} + \frac{(63 - 58.333)^2}{58.333} + \ldots + \frac{(42 - 58.333)^2}{58.333} = 19.726.
\]

Since \( \chi^2 > \chi^2_{12-1,0.05} = 19.675 \), reject \( H_0 \) and conclude that the first births are not spread uniformly throughout the year.

9.19  (a) \( H_0 : p_i = \binom{i}{7}(0.5)^7 \) vs. \( H_1 : \) Not \( H_0 \). Using

\[
e_i = np_i = 98 \binom{7}{i}(0.5)^7,
\]

the results are summarized below:

<table>
<thead>
<tr>
<th>Sons</th>
<th>( n_i )</th>
<th>( e_i )</th>
<th>( \frac{[n_i-e_i]^2}{e_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.766</td>
<td>0.003</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>5.359</td>
<td>0.025</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>16.078</td>
<td>0.269</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>26.797</td>
<td>0.120</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>26.797</td>
<td>1.254</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>16.078</td>
<td>2.181</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>5.359</td>
<td>2.452</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.766</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>98</td>
<td><strong>6.278</strong></td>
<td></td>
</tr>
</tbody>
</table>


Note that cell 0 was combined with cell 1, and cell 7 was combined with cell 6, to satisfy the requirement that no cell can have $e_i < 1$ and no more than 1/5th of the $e_i$ can be < 5. Since $\chi^2 < \chi^2_{0.10} = 9.236$, do not reject $H_0$ and conclude that the binomial distribution with $p = 0.5$ is a plausible distribution.

(b) $H_0 : p_i$ is binomial vs. $H_1 : \text{Not } H_0$. Using

$$\hat{p} = \frac{\text{Number of Sons}}{\text{Number of Children}} = \frac{364}{7 \times 98} = 0.531$$

and

$$e_i = np_i = 98 \left( \binom{7}{i} \right) (0.531)^i (0.469)^{7-i},$$

the results are summarized below:

<table>
<thead>
<tr>
<th>Sons</th>
<th>$n_i$</th>
<th>$e_i$</th>
<th>$(\frac{n_i - e_i}{e_i})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.492</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>3.893</td>
<td>0.595</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>13.203</td>
<td>0.048</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>24.874</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>28.119</td>
<td>1.802</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>19.072</td>
<td>0.450</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>7.187</td>
<td>0.327</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1.161</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>98</td>
<td></td>
<td>$\chi^2 = 3.223$</td>
</tr>
</tbody>
</table>

Note that cell 0 was combined with cell 1, and cell 7 was combined with cell 6, to satisfy the requirement that no cell can have $e_i < 1$ and no more than 1/5th of the $e_i$ can be < 5. Since $\chi^2 < \chi^2_{0.05} = 7.779$, do not reject $H_0$ and conclude that the binomial distribution is a plausible distribution. This agrees with (a), since $\hat{p} = .531 \approx .5$.

9.20 (a) $H_0 : p_1 = 9/16, p_2 = p_3 = 3/16, p_4 = 1/16$ vs. $H_1 : \text{Not } H_0$.

(b) Using

$$e_i = np_i = 1611p_i,$$

the results are summarized below:

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>$n_i$</th>
<th>$e_i$</th>
<th>$(\frac{n_i - e_i}{e_i})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tall, cut-leaf</td>
<td>926</td>
<td>906.188</td>
<td>0.433</td>
</tr>
<tr>
<td>Dwarf, cut-leaf</td>
<td>293</td>
<td>302.063</td>
<td>0.272</td>
</tr>
<tr>
<td>Tall, potato-leaf</td>
<td>288</td>
<td>302.063</td>
<td>0.655</td>
</tr>
<tr>
<td>Dwarf, potato-leaf</td>
<td>104</td>
<td>100.688</td>
<td>0.109</td>
</tr>
<tr>
<td>Total</td>
<td>1611</td>
<td></td>
<td>$\chi^2 = 1.469$</td>
</tr>
</tbody>
</table>

Since $\chi^2 < \chi^2_{0.05} = 7.815$, do not reject $H_0$ and conclude that the proportion 9:3:3:1 fit the observed frequencies well.

9.21 Since

$$\hat{\lambda} = \frac{229 \times 0 + 211 \times 1 + \ldots + 7 \times 1}{229 + 211 + \ldots + 1} = 0.932,$$
then using the Poisson formula,

\[ p_i = \frac{e^{-0.932}(0.932)^i}{i!}, \]

and the expected frequencies are

\[ e_i = np_i = 576p_i. \]

The results are summarized below:

<table>
<thead>
<tr>
<th>Hits</th>
<th>( n_i )</th>
<th>( p_i )</th>
<th>( e_i )</th>
<th>( \frac{(n_i - e_i)^2}{e_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>229</td>
<td>0.394</td>
<td>226.743</td>
<td>0.022</td>
</tr>
<tr>
<td>1</td>
<td>211</td>
<td>0.367</td>
<td>211.390</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>93</td>
<td>0.171</td>
<td>98.539</td>
<td>0.311</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>0.053</td>
<td>30.622</td>
<td>0.626</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>0.012</td>
<td>7.137</td>
<td>0.003</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.002</td>
<td>1.331</td>
<td>0.204</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.000</td>
<td>0.207</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.000</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>576</td>
<td></td>
<td></td>
<td>( \chi^2 = 0.960 )</td>
</tr>
</tbody>
</table>

Note that cells 6 and 7 were combined with cell 5, to satisfy the requirement that no cell can have \( e_i < 1 \) and no more than 1/5th of the \( e_i \) can be < 5. Since \( \chi^2 < \chi^2_{0.05} = 9.488 \), do not reject \( H_0 \) and conclude that the Poisson distribution is a plausible model.

**9.22 (a)** Since \( \hat{\lambda} = 0.519 \), then

\[ p_i = \frac{e^{-0.519}(0.519)^i}{i!}, \]

and

\[ e_i = np_i. \]

The results are summarized below:

<table>
<thead>
<tr>
<th>Passengers</th>
<th>( n_i )</th>
<th>( p_i )</th>
<th>( e_i )</th>
<th>( \frac{(n_i - e_i)^2}{e_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>678</td>
<td>0.595</td>
<td>601.662</td>
<td>9.686</td>
</tr>
<tr>
<td>1</td>
<td>227</td>
<td>0.309</td>
<td>312.262</td>
<td>23.281</td>
</tr>
<tr>
<td>2</td>
<td>56</td>
<td>0.080</td>
<td>81.032</td>
<td>7.733</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>0.014</td>
<td>14.019</td>
<td>13.944</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>0.002</td>
<td>1.819</td>
<td>196.998</td>
</tr>
<tr>
<td>( \geq 5 )</td>
<td>14</td>
<td>0.000</td>
<td>0.206</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1011</td>
<td></td>
<td></td>
<td>( \chi^2 = 251.642 )</td>
</tr>
</tbody>
</table>

Note that cell \( \geq 5 \) was combined with cell 4, to satisfy the requirement that no cell can have \( e_i < 1 \) and no more than 1/5th of the \( e_i \) can be < 5. Since \( \chi^2 > \chi^2_{0.05} = 7.815 \), reject \( H_0 \) and conclude that the Poisson distribution is not a plausible distribution for the number of passengers.

(b) Since \( \hat{p} = 1/(1 + 0.519) = 0.658 \), then

\[ p_i = (1 - \hat{p})^{i-1} p = (0.342)^{i-1} (0.658) \]

and

\[ e_i = np_i. \]

The results are summarized below:
Since \( \chi^2 > \chi^2_{0.05} = 9.488 \), reject \( H_0 \) and conclude that the geometric distribution is not a plausible distribution for the number of occupants.

(c) While neither is a plausible distribution for the data, the geometric distribution seemed to fit much better, since the \( \chi^2 \) value is much smaller. Also note that the lack of fit of the geometric distribution comes primarily from the tail category (\( \geq 6 \)).

\[ 9.23 \] (a)

\[
\chi^2 = \sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i}
\]

\[ = \frac{(x - np_0)^2}{np_0} + \frac{(n - x - n(1 - p_0))^2}{n(1 - p_0)} \]

\[ = \frac{(x - np_0)^2}{np_0(1 - p_0)} = z^2. \]

(b) We reject \( H_0 \) if \(|z| > z_{\alpha/2} \) or if \( z^2 > z_{\alpha/2}^2 = \chi^2_{\alpha,\alpha} \). Hence the two tests are equivalent.

\[ 9.24 \] Using \( p = q = 0.5 \), then

\[ p_i = \binom{i - 1}{3} \left[ p^4 q^{i-4} + q^4 p^{i-4} \right] = \binom{i - 1}{3} (0.5)^{i-1} \]

and

\[ e_i = np_i. \]

The results are summarized below:

<table>
<thead>
<tr>
<th>Games (i)</th>
<th>( n_i )</th>
<th>( p_i )</th>
<th>( e_i )</th>
<th>( \frac{(n_i - e_i)^2}{e_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>6</td>
<td>0.125</td>
<td>6.500</td>
<td>0.038</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>0.250</td>
<td>13.000</td>
<td>0.308</td>
</tr>
<tr>
<td>6</td>
<td>21</td>
<td>0.313</td>
<td>16.250</td>
<td>1.388</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>0.313</td>
<td>16.250</td>
<td>0.312</td>
</tr>
<tr>
<td>Total</td>
<td>52</td>
<td></td>
<td></td>
<td>( \chi^2 = 2.046 )</td>
</tr>
</tbody>
</table>

Since \( \chi^2 < \chi^2_{0.05} = 7.815 \), we do not reject \( H_0 \) and conclude that this model does fit the data well.
Solutions to Section 9.4

9.25 (a) Multinomial sampling. $H_0 : p_{ij} = p_i p_j$ for all $i, j$, where $i$ refers to religious affiliation and $j$ refers to political party affiliation.

(b) Product Multinomial sampling. $H_0 : p_{ij} = p_j$ for all $i, j$, where $i$ refers to the type of mutual fund and $j$ refers to the return classification.

9.26 (a) Product Multinomial sampling. $H_0 : p_{ij} = p_j$ for all $i, j$, where $i$ refers to the age group and $j$ refers to the willingness to use the internet grocery service.

(b) Multinomial sampling. $H_0 : p_{ij} = p_i p_j$ for all $i, j$, where $i$ refers to the severity of injury and $j$ refers to the use of a safety restraint.

9.27 (a) Product Multinomial sampling.

(b) Using

$$\hat{e}_{ij} = \frac{n_i n_j}{n},$$

the results are summarized below:

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>412</td>
<td>1146</td>
<td>1497</td>
<td>1117</td>
<td>651</td>
<td>442</td>
<td>307</td>
<td>231</td>
<td>181</td>
<td>199</td>
<td>108</td>
<td>78</td>
<td>54</td>
<td>46</td>
<td>30</td>
<td>22</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>$e_{ij}$ Total</td>
<td>233</td>
<td>1213</td>
<td>1313</td>
<td>1102</td>
<td>663</td>
<td>454</td>
<td>400</td>
<td>275</td>
<td>205</td>
<td>122</td>
<td>64</td>
<td>40</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q.C.S. Total</td>
<td>503</td>
<td>2928</td>
<td>2933</td>
<td>2377</td>
<td>1420</td>
<td>860</td>
<td>860</td>
<td>594</td>
<td>441</td>
<td>263</td>
<td>138</td>
<td>85</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tot.</td>
<td>736</td>
<td>3881</td>
<td>4146</td>
<td>3479</td>
<td>2102</td>
<td>1434</td>
<td>1203</td>
<td>809</td>
<td>646</td>
<td>385</td>
<td>202</td>
<td>125</td>
<td>73</td>
<td>192</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Then

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

$$= \frac{(312 - 233)^2}{233} + \frac{(1146 - 1213)^2}{1213} + \ldots + \frac{(61 - 50)^2}{50}$$

$$= 101.494.$$

Since $\chi^2 > \chi^2_{0.01} = 26.217$, reject $H_0$ and conclude that the Q.C.S. letters do not match Mark Twain’s word length patterns.

9.28 (a) Product Multinomial sampling.

(b) $H_0 : p_{ij} = p_j$, where $i$ refers to type of city and $j$ refers to whether or not the wallet was returned.

(c) Using

$$\hat{e}_{ij} = \frac{n_i n_j}{n},$$

the results are summarized below:

<table>
<thead>
<tr>
<th>Type of Cities</th>
<th>Returned</th>
<th>Kept</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Cities</td>
<td>21</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Suburbs</td>
<td>18</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Medium Cities</td>
<td>17</td>
<td>13</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Small Cities</td>
<td>24</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>40</td>
<td>120</td>
</tr>
</tbody>
</table>
Then
\[ \chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \]
\[ = \frac{(21 - 20)^2}{20} + \frac{(9 - 10)^2}{10} + \ldots + \frac{(6 - 10)^2}{10} \]
\[ = 4.5. \]

Since \( \chi^2 < \chi^2_{(1,1),(2-1),0.10} = 6.251 \), do not reject \( H_0 \) and conclude that there are no differences in the return rates among the different types of cities.

9.29 (a) Multinomial sampling.

(b) \( H_0 : p_{ij} = p_i p_j \) for all \( i, j \), where \( i \) refers to gender and \( j \) refers to height above ground.

(c) Using
\[ e_{ij} = \frac{n_i n_j}{n}, \]
the results are summarized below:

<table>
<thead>
<tr>
<th>Gender</th>
<th>Height above ground</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 feet</td>
<td>35 feet</td>
</tr>
<tr>
<td>Males</td>
<td>173</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>184.75</td>
<td>113.25</td>
</tr>
<tr>
<td>Females</td>
<td>150</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>138.25</td>
<td>84.75</td>
</tr>
<tr>
<td>Total</td>
<td>323</td>
<td>198</td>
</tr>
</tbody>
</table>

Then
\[ \chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \]
\[ = \frac{(173 - 184.75)^2}{184.75} + \frac{(125 - 113.25)^2}{113.25} + \ldots + \frac{(73 - 84.75)^2}{84.75} \]
\[ = 4.593. \]

Since \( \chi^2 > \chi^2_{(2-1),(2-1),0.05} = 3.841 \), reject \( H_0 \) and conclude that gender and trap height are associated and are not independent.

9.30 (a)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Airsick</th>
<th>OK</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dramamine</td>
<td>31</td>
<td>77</td>
<td>108</td>
</tr>
<tr>
<td>Placebo</td>
<td>60</td>
<td>48</td>
<td>108</td>
</tr>
<tr>
<td>Column Total</td>
<td>91</td>
<td>185</td>
<td>216</td>
</tr>
</tbody>
</table>

(b) \( H_0 : p_{ij} = p_j \), where \( i \) refers to the treatment given and \( j \) refers to whether or not the volunteer became airsick.

(c) Using
\[ e_{ij} = \frac{n_i n_j}{n}, \]
the results are summarized below:
<table>
<thead>
<tr>
<th>Treatment</th>
<th>Airsick</th>
<th>OK</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dramamine</td>
<td>$n_{ij}$</td>
<td>31</td>
<td>77</td>
</tr>
<tr>
<td>*</td>
<td>$e_{ij}$</td>
<td>45.50</td>
<td>62.50</td>
</tr>
<tr>
<td>Placebo</td>
<td>$n_{ij}$</td>
<td>60</td>
<td>48</td>
</tr>
<tr>
<td>*</td>
<td>$e_{ij}$</td>
<td>45.50</td>
<td>62.50</td>
</tr>
<tr>
<td>Column Total</td>
<td>91</td>
<td>125</td>
<td>216</td>
</tr>
</tbody>
</table>

Then

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

$$= \frac{(31 - 45.50)^2}{45.50} + \frac{(77 - 62.50)^2}{62.50} + \ldots + \frac{(48 - 62.50)^2}{62.50}$$

$$= 15.970.$$  

Since $\chi^2 > \chi^2_{(2-1)(2-1),0.05} = 3.841$, reject $H_0$ and conclude that Dramamine is effective in reducing the chances of airsickness.

9.31 (a)

<table>
<thead>
<tr>
<th>Cholesterol Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\leq 250$</td>
</tr>
<tr>
<td>Personality Type</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
</tbody>
</table>

(b) $H_0 : p_{ij} = p_i.p_j$ for all $i,j$, where $i$ refers to the personality type and $j$ refers to the cholesterol level. Using

$$\hat{e}_{ij} = \frac{n_i.n_j}{n},$$

the results are summarized below:

<table>
<thead>
<tr>
<th>Personality Type</th>
<th>Cholesterol Level</th>
<th></th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\leq 250$</td>
<td>$&gt; 250$</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>$n_{ij}$</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>$e_{ij}$</td>
<td>14.5</td>
<td>5.5</td>
</tr>
<tr>
<td>B</td>
<td>$n_{ij}$</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$e_{ij}$</td>
<td>14.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Column Total</td>
<td>29</td>
<td>11</td>
<td>40</td>
</tr>
</tbody>
</table>

Then

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

$$= \frac{(12 - 14.50)^2}{14.50} + \frac{(8 - 5.50)^2}{5.50} + \ldots + \frac{(3 - 5.50)^2}{5.50}$$

$$= 3.135.$$  

Since $\chi^2 < \chi^2_{(2-1)(2-1),0.10} = 2.706$, reject $H_0$ and conclude that personality type and cholesterol level are associated and are not independent.

9.32 (a) Multinomial sampling.
(b) \( H_0 : p_{ij} = p_i p_j \) for all \( i, j \), where \( i \) refers to eye color and \( j \) refers to hair color. Using
\[
\hat{e}_{ij} = \frac{n_i n_j}{n},
\]
the results are summarized below:

<table>
<thead>
<tr>
<th>Eye Color</th>
<th>Hair Color</th>
<th>Black</th>
<th>Brown</th>
<th>Red</th>
<th>Blond</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td>( n_{ij} )</td>
<td>68</td>
<td>119</td>
<td>26</td>
<td>7</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td>( e_{ij} )</td>
<td>40.14</td>
<td>106.28</td>
<td>26.39</td>
<td>47.20</td>
<td></td>
</tr>
<tr>
<td>Blue</td>
<td>( n_{ij} )</td>
<td>20</td>
<td>84</td>
<td>17</td>
<td>94</td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>( e_{ij} )</td>
<td>39.22</td>
<td>103.87</td>
<td>25.79</td>
<td>46.12</td>
<td></td>
</tr>
<tr>
<td>Hazel</td>
<td>( n_{ij} )</td>
<td>15</td>
<td>54</td>
<td>14</td>
<td>10</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>( e_{ij} )</td>
<td>16.97</td>
<td>44.93</td>
<td>11.15</td>
<td>19.95</td>
<td></td>
</tr>
<tr>
<td>Green</td>
<td>( n_{ij} )</td>
<td>5</td>
<td>29</td>
<td>14</td>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>( e_{ij} )</td>
<td>11.68</td>
<td>30.92</td>
<td>7.68</td>
<td>13.73</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>108</td>
<td>286</td>
<td>71</td>
<td>127</td>
<td>592</td>
</tr>
</tbody>
</table>

Then
\[
\chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}
\]
\[
= \frac{(68 - 40.14)^2}{40.14} + \frac{(119 - 106.28)^2}{106.28} + \ldots + \frac{(16 - 7.68)^2}{7.68}
\]
\[
= 138.290.
\]
Since \( \chi^2 > \chi^2_{(4-1)(4-1),0.05} = 16.919 \), reject \( H_0 \) and conclude that eye color and hair color are associated and are not independent.

9.33 (a) \( H_0 : p_{ij} = p_i p_j \) for all \( i, j \) where \( i \) refers to the opinion on full evacuation and \( j \) refers to the distance from Three Mile Island.
(b) Using
\[
\hat{e}_{ij} = \frac{n_i n_j}{n},
\]
the results are summarized below:

<table>
<thead>
<tr>
<th>Full Evacuation</th>
<th>Distance (in miles) from Three Mile Island</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-3</td>
<td>4-6</td>
</tr>
<tr>
<td>Yes</td>
<td>( n_{ij} )</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>( e_{ij} )</td>
<td>7.04</td>
</tr>
<tr>
<td>No</td>
<td>( n_{ij} )</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>( e_{ij} )</td>
<td>8.96</td>
</tr>
<tr>
<td>Column Total</td>
<td>16</td>
<td>22</td>
</tr>
</tbody>
</table>

Then
\[
\chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}
\]
\[
= \frac{(7 - 7.04)^2}{4.04} + \frac{(11 - 9.68)^2}{9.68} + \ldots + \frac{(39 - 38.08)^2}{38.08}
\]
\[
= 0.449.
\]
Since $\chi^2 < \chi^2_{(2-1)(6-1), 0.10} = 9.236$, do not reject $H_0$ and conclude that one’s opinion on full evacuation is independent of one’s distance from Three Mile Island.

9.34 (a) $H_0 : p_{ij} = p_i p_{.j}$ for all $i, j$ where $i$ refers to tonsil size and $j$ refers to carrier status.

(b) Using 
\[
\hat{e}_{ij} = \frac{n_i n_{.j}}{n},
\]
the results are summarized below:

<table>
<thead>
<tr>
<th>Tonsil Size</th>
<th>Carrier Status</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Carrier</td>
<td>Non-carrier</td>
</tr>
<tr>
<td>Normal</td>
<td>$n_{ij}$</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>$e_{ij}$</td>
<td>26.58</td>
</tr>
<tr>
<td>Large</td>
<td>$n_{ij}$</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>$e_{ij}$</td>
<td>30.33</td>
</tr>
<tr>
<td>Very Large</td>
<td>$n_{ij}$</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>$e_{ij}$</td>
<td>15.09</td>
</tr>
<tr>
<td>Column Total</td>
<td></td>
<td>72</td>
</tr>
</tbody>
</table>

Then 
\[
\chi^2 = \sum_{i,j} \left( \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \right) 
\]
\[
= \frac{(19 - 26.58)^2}{26.58} + \frac{(497 - 489.42)^2}{489.42} + \ldots + \frac{(269 - 277.91)^2}{277.91} 
\]
\[
= 7.885. 
\]

Since $\chi^2 > \chi^2_{(3-1)(2-1), 0.05} = 5.991$, reject $H_0$ and conclude that tonsil size and carrier status are associated and are not independent.

9.35 (a) $H_0 : p_{ij} = p_i p_{.j}$ for all $i, j$, where $i$ refers to the age at diagnosis of breast cancer and $j$ refers to the frequency of breast self-exam.

(b) Using 
\[
\hat{e}_{ij} = \frac{n_i n_{.j}}{n},
\]
the results are summarized below:

<table>
<thead>
<tr>
<th>Age at Diagnosis Of Breast Cancer</th>
<th>Frequency of Breast Self-Exam</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monthly</td>
<td>Occasionally</td>
</tr>
<tr>
<td>&lt; 45</td>
<td>$n_{ij}$</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>$e_{ij}$</td>
<td>66.78</td>
</tr>
<tr>
<td>45 - 59</td>
<td>$n_{ij}$</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>$e_{ij}$</td>
<td>145.35</td>
</tr>
<tr>
<td>60+</td>
<td>$n_{ij}$</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>$e_{ij}$</td>
<td>137.87</td>
</tr>
<tr>
<td>Column Total</td>
<td></td>
<td>350</td>
</tr>
</tbody>
</table>
Then
\[
\chi^2 = \sum_{ij} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \frac{(91 - 66.78)^2}{66.78} + \frac{(90 - 93.11)^2}{93.11} + \ldots + \frac{(172 - 148.90)^2}{148.90} = 25.086.
\]

Since \(\chi^2 > \chi^2_{3-1,0.10} = 7.779\), reject \(H_0\) and conclude that age at diagnosis and frequency of breast self-exam are associated and are not independent.

**9.36** The estimate of the odds ratio is
\[
\hat{\psi} = \frac{n_{11}n_{22}}{n_{21}n_{12}} = \frac{31 \times 48}{60 \times 77} = 0.322.
\]

A 95% CI for \(\log_e \psi\) is
\[
[L, U] = \left[ \log_e \hat{\psi} \pm z_{0.025} \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \right]
\]
\[
= \left[ \log_e 0.322 \pm 1.96 \sqrt{\frac{1}{31} + \frac{1}{77} + \frac{1}{60} + \frac{1}{48}} \right]
\]
\[
= [-1.697, -0.569].
\]

Then a 90% CI for \(\psi\) is
\[
\left[ e^L, e^U \right] = \left[ e^{-1.697}, e^{-0.569} \right] = [0.183, 0.566].
\]

Since the CI is below 1, we can conclude that there is an association between use of dramamine and airsickness, and that people who take dramamine generally have lower odds of getting airsick than do people who do not take dramamine.

**9.37** The estimate of the odds ratio is
\[
\hat{\psi} = \frac{n_{11}n_{22}}{n_{21}n_{12}} = \frac{12 \times 3}{17 \times 8} = 0.265.
\]

A 90% CI for \(\log_e \psi\) is
\[
[L, U] = \left[ \log_e \hat{\psi} \pm z_{0.05} \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \right]
\]
\[
= \left[ \log_e 0.265 \pm 1.645 \sqrt{\frac{1}{12} + \frac{1}{8} + \frac{1}{17} + \frac{1}{3}} \right]
\]
\[
= [-2.604, -0.054].
\]

Then a 90% CI for \(\psi\) is
\[
\left[ e^L, e^U \right] = \left[ e^{-2.604}, e^{-0.054} \right] = [0.074, 0.947].
\]

Since the CI is below 1, we can conclude that there is an association between personality type and cholesterol level, and that type A personalities have higher odds of having higher cholesterol levels than do type B personalities.
Solutions to Chapter 9 Advanced Exercises

9.38  (a) Skilled Work: Using

\[ \hat{e}_{ij} = \frac{n_i n_j}{n}, \]

the results are summarized below:

<table>
<thead>
<tr>
<th>Home Repair</th>
<th>Age</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt; 30</td>
<td>31 - 45</td>
</tr>
<tr>
<td>Yes</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>53.17</td>
<td>60.21</td>
</tr>
<tr>
<td>No</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>11.80</td>
<td>16.45</td>
</tr>
<tr>
<td>Column Total</td>
<td>68</td>
<td>77</td>
</tr>
</tbody>
</table>

Then

\[ \chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \]

\[ = \frac{(56 - 53.17)^2}{53.17} + \frac{(56 - 60.21)^2}{60.21} + \ldots + \frac{(8 - 12.75)^2}{12.75} \]

\[ = 3.527. \]

Since \( \chi^2 < \chi^2_{(2-1)(3-1),0.05} = 5.991 \), do not reject \( H_0 \) and conclude that doing home repair and age are independent for skilled work.

Unskilled Work: Using

\[ \hat{e}_{ij} = \frac{n_i n_j}{n}, \]

the results are summarized below:

<table>
<thead>
<tr>
<th>Home Repair</th>
<th>Age</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt; 30</td>
<td>31 - 45</td>
</tr>
<tr>
<td>Yes</td>
<td>23</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>18.46</td>
<td>47.87</td>
</tr>
<tr>
<td>No</td>
<td>9</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>13.54</td>
<td>35.13</td>
</tr>
<tr>
<td>Column Total</td>
<td>32</td>
<td>83</td>
</tr>
</tbody>
</table>

Then

\[ \chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \]

\[ = \frac{(23 - 18.46)^2}{18.46} + \frac{(52 - 47.87)^2}{47.87} + \ldots + \frac{(51 - 42.33)^2}{42.33} \]

\[ = 6.568. \]

Since \( \chi^2 > \chi^2_{(2-1)(3-1),0.05} = 5.991 \), reject \( H_0 \) and conclude that doing home repair and age are associated for unskilled work.

Office Work: Using

\[ \hat{e}_{ij} = \frac{n_i n_j}{n}, \]

the results are summarized below:
<table>
<thead>
<tr>
<th>Home Repair</th>
<th>Age</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt; 30</td>
<td>31 - 45</td>
</tr>
<tr>
<td>Yes</td>
<td>$n_{ij}$</td>
<td>$e_{ij}$</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>191</td>
</tr>
<tr>
<td>No</td>
<td>$n_{ij}$</td>
<td>$e_{ij}$</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>76</td>
</tr>
<tr>
<td>Column Total</td>
<td>73</td>
<td>267</td>
</tr>
</tbody>
</table>

Then

\[
\chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}
\]

\[
= \frac{(54 - 50.36)^2}{50.36} + \frac{(191 - 184.19)^2}{184.19} + \ldots + \frac{(61 - 50.55)^2}{50.55}
\]

\[
= 4.789.
\]

Since $\chi^2 < \chi^2_{(2-1)(3-1).0.05} = 5.991$, do not reject $H_0$ and conclude that doing home repair and age are independent for office work.

(b) Summing over all types of work and using

\[
ed_{ij} = \frac{n_{i.n}n_{.j}}{n},
\]

the results are summarized below:

<table>
<thead>
<tr>
<th>Home Repair</th>
<th>Age</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt; 30</td>
<td>31 - 45</td>
</tr>
<tr>
<td>Yes</td>
<td>$n_{ij}$</td>
<td>$e_{ij}$</td>
</tr>
<tr>
<td></td>
<td>133</td>
<td>299</td>
</tr>
<tr>
<td>No</td>
<td>$n_{ij}$</td>
<td>$e_{ij}$</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>128</td>
</tr>
<tr>
<td>Column Total</td>
<td>173</td>
<td>427</td>
</tr>
</tbody>
</table>

Then

\[
\chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}
\]

\[
= \frac{(133 - 118.01)^2}{118.01} + \frac{(299 - 291.26)^2}{291.26} + \ldots + \frac{(120 - 97.27)^2}{97.27}
\]

\[
= 14.425.
\]

Since $\chi^2 > \chi^2_{(2-1)(3-1).0.05} = 5.991$, reject $H_0$ and conclude that doing home repair and age are associated.

(c) Just using the results from (b), we would conclude that doing home repair is associated with age. In actuality, doing home repair and age are only associated for unskilled work. For skilled and office work, there is no association.

9.39 (a)

\[
1 - \alpha = P \left( \hat{p} - p \right)^2 \leq \frac{z_{\alpha/2}^2 pq}{n}
\]
\[
\begin{align*}
= & \quad P \left[ \hat{p}^2 - 2p\hat{p} + p^2 \leq z_{\alpha/2}^2 \frac{p}{n} - z_{\alpha/2}^2 \frac{p^2}{n} \right] \\
= & \quad P \left[ p^2 \left( 1 + \frac{z_{\alpha/2}^2}{n} \right) + p \left( -2\hat{p} - \frac{z_{\alpha/2}^2}{n} \right) + \hat{p}^2 \right] \\
= & \quad P \left[ Ap^2 + Bp + C \leq 0 \right],
\end{align*}
\]

where
\[
\begin{align*}
A & = 1 + \frac{z_{\alpha/2}^2}{n} \\
B & = -2\hat{p} - \frac{z_{\alpha/2}^2}{n} \\
C & = \hat{p}^2.
\end{align*}
\]

(b) The solutions to this quadratic equation are
\[
p = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}
\]
\[
= \frac{2\hat{p} + \frac{z^2}{n} \pm \sqrt{\left(2\hat{p} + \frac{z^2}{n}\right)^2 - 4 \left(1 + \frac{z^2}{n}\right)(\hat{p}^2)}}{2 \left(1 + \frac{z^2}{n}\right)}
\]
\[
= \frac{2\hat{p} + \frac{z^2}{n} \pm \sqrt{4\hat{p}^2 + 4\hat{p} \frac{z^2}{n} + \frac{z^2}{n} - 4\hat{p}^2 - 4\hat{p} \frac{z^2}{n}}}{1 \left(1 + \frac{z^2}{n}\right)}
\]
\[
= \frac{\hat{p} + \frac{z^2}{n} \pm \sqrt{\hat{p} \left(1 - \frac{z^2}{n}\right) + \frac{z^2}{n^2}}}{1 + \frac{z^2}{n}}.
\]

(c) From the quadratic equation,
\[
Ap^2 + Bp + C \leq 0 \iff (p - p_L)(p - p_U) \leq 0.
\]

If \( p < p_L \), then
\[
p - p_L < 0, \quad p - p_U < 0, \quad \text{and} \quad Ap^2 + Bp + C > 0.
\]

If \( p > p_U \), then
\[
p - p_L > 0, \quad p - p_U > 0, \quad \text{and} \quad Ap^2 + Bp + C > 0.
\]

But if \( p_L \leq p \leq p_U \), then
\[
p - p_L \geq 0, \quad p - p_U \leq 0, \quad \text{and} \quad Ap^2 + Bp + C \leq 0.
\]

This CI is more accurate because it doesn’t use estimated proportions \( \hat{p} \) and \( \hat{q} \) to estimate the standard error in the formula
\[
\sqrt{\frac{p(1-p)}{n}}.
\]

(d) If we ignore the terms of order \( 1/n \), the CI simplifies to
\[
\left[ \frac{\hat{p} \pm \sqrt{\hat{p}\hat{q}}}{n} \right] = \left[ \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right].
\]
(a) \( H_0 : p_{ijk} = p_i \cdot p_j \cdot p_k \) for \( i = 1, \ldots, r \), \( j = 1, \ldots, c \), and \( k = 1, \ldots, t \).

(b) The estimate of the expected frequency \( e_{ijk} \) is

\[
\hat{e}_{ijk} = \frac{n}{n} \left( \frac{n_i}{n} \right) \left( \frac{n_j}{n} \right) \left( \frac{n_k}{n} \right) = \frac{n_i n_j n_k}{n^2}.
\]

(c) The test statistic is

\[
\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{t} \frac{(n_{ijk} - \hat{e}_{ijk})^2}{\hat{e}_{ijk}},
\]

with d.f. = \((r - 1)(c - 1)(t - 1)\).

9.41 (a)

<table>
<thead>
<tr>
<th>Defendent’s Race</th>
<th>Death Penalty</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>19</td>
<td>141</td>
</tr>
<tr>
<td>Black</td>
<td>17</td>
<td>149</td>
</tr>
</tbody>
</table>

Since \( \hat{p}_W = 19/160 = 0.119 \) and \( \hat{p}_B = 17/166 = 0.102 \), the test statistic is

\[
z = \frac{\hat{p}_W - \hat{p}_B}{\sqrt{\frac{\hat{p}_W(1-\hat{p}_W)}{n_W} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}} = \frac{0.119 - 0.102}{\sqrt{\frac{0.119(0.881)}{160} + \frac{0.102(0.898)}{166}}} = 0.470.
\]

Then the \( P \)-value is

\[P = 2(1 - \Phi(0.470)) = 2 \times 0.319 = 0.638.\]

Since \( P > \alpha = 0.05 \), do not reject \( H_0 \) and conclude that there is no racial difference between black and white defendants.

(b) White victims: Since \( \hat{p}_W = 19/151 = 0.126 \) and \( \hat{p}_B = 11/63 = 0.175 \), the test statistic is

\[
z = \frac{\hat{p}_W - \hat{p}_B}{\sqrt{\frac{\hat{p}_W(1-\hat{p}_W)}{n_W} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}} = \frac{0.126 - 0.175}{\sqrt{\frac{0.126(0.874)}{151} + \frac{0.175(0.825)}{63}}} = -0.888.
\]

Then the \( P \)-value is

\[P = 2(1 - \Phi(-0.888)) = 2 \times 0.187 = 0.374.\]

Since \( P > \alpha = 0.05 \), do not reject \( H_0 \) and conclude that there is no racial difference between black and white defendants when the victim is white.

Black victims: Since \( \hat{p}_W = 0/9 = 0.000 \) and \( \hat{p}_B = 6/103 = 0.057 \), the test statistic is

\[
z = \frac{\hat{p}_W - \hat{p}_B}{\sqrt{\frac{\hat{p}_W(1-\hat{p}_W)}{n_W} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}} = \frac{0.000 - 0.057}{\sqrt{\frac{0.000(1.000)}{9} + \frac{0.057(0.943)}{103}}} = -2.50.
\]

Then the \( P \)-value is

\[P = 2(1 - \Phi(-2.50)) = 2 \times 0.006 = 0.012.\]

Since \( P < \alpha = 0.05 \), reject \( H_0 \) and conclude that there are racial differences between black and white defendants when the victim is black.
(c) Black defendants are significantly more likely to receive the death penalty than white defendants when the victim is black, but not when the victim is white.

9.42 (a)

\[
1 - \alpha = P \left[ -z_{\alpha/2} \leq \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \leq z_{\alpha/2} \right] \\
= P \left[ (\bar{X} - \lambda)^2 \leq \frac{z_{\alpha/2}^2 \lambda}{n} \right] \\
= P \left[ \lambda^2 + \lambda \left( -2\bar{X} - \frac{z_{\alpha/2}^2}{n} \right) + \bar{X}^2 \leq 0 \right] \\
= P \left[ A\lambda^2 + B\lambda + C \leq 0 \right],
\]

where

\[
A = 1 \\
B = -2\bar{X} - \frac{z_{\alpha/2}^2}{n} \\
C = \bar{X}^2.
\]

The solutions to this quadratic equation are

\[
\lambda = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \\
= 2\bar{X} + \frac{z^2}{n} \pm \sqrt{\left(2\bar{X} + \frac{z^2}{n}\right)^2 - 4\bar{X}^2} \\
= \bar{X} + \frac{z^2}{2n} \pm \frac{1}{2} \sqrt{4\bar{X}^2 + 4\bar{X} \frac{z^2}{n} + \frac{z^4}{n^2} - 4\bar{X}^2} \\
= \bar{X} + \frac{z^2}{2n} \pm \sqrt{\frac{\bar{X}z^2}{n} + \frac{z^4}{4n^2}}.
\]

Then

\[
A\lambda^2 + B\lambda + C \leq 0 \iff (\lambda - \lambda_L)(\lambda - \lambda_U) \leq 0 \\
= \lambda_L \leq \lambda \leq \lambda_U.
\]

So \([\lambda_L, \lambda_U]\) is a \((1 - \alpha)\)-level CI for \(\lambda\).

(b) If we ignore the terms of order \(1/n\), the CI simplifies to

\[
\left[ \bar{X} \pm \frac{\bar{X}z^2}{n} \right] = \left[ \bar{X} \pm z_{\alpha/2} \sqrt{\frac{\bar{X}}{n}} \right].
\]

9.43 (a) The observed frequencies are

\[
n_{11} = n_1 \hat{p}_1, \quad n_{12} = n_1 (1 - \hat{p}_1) \\
n_{21} = n_2 \hat{p}_2, \quad n_{22} = n_2 (1 - \hat{p}_2)
\]
and the expected frequencies are

\[ e_{11} = n_1 \hat{p} \quad , \quad e_{12} = n_1 (1 - \hat{p}) \]
\[ e_{21} = n_2 \hat{p} \quad , \quad e_{22} = n_2 (1 - \hat{p}) \]

Then

\[
\chi^2 = \frac{(n_1 \hat{p}_1 - n_1 \hat{p})^2}{n_1 \hat{p}} + \frac{[n_1 (1 - \hat{p}_1) - n_1 (1 - \hat{p})]^2}{n_1 (1 - \hat{p})} \\
+ \frac{(n_2 \hat{p}_2 - n_2 \hat{p})^2}{n_2 \hat{p}} + \frac{[n_2 (1 - \hat{p}_2) - n_2 (1 - \hat{p})]^2}{n_2 (1 - \hat{p})}
\]
\[= \frac{n_1}{\hat{p} (1 - \hat{p})} \left[ (1 - \hat{p}) (\hat{p}_1 - \hat{p})^2 + \hat{p} (\hat{p}_1 - \hat{p})^2 \right] \\
+ \frac{n_2}{\hat{p} (1 - \hat{p})} \left[ (1 - \hat{p}) (\hat{p}_2 - \hat{p})^2 + \hat{p} (\hat{p}_2 - \hat{p})^2 \right]
\]
\[= \frac{n_1 (\hat{p}_1 - \hat{p})^2 + n_2 (\hat{p}_2 - \hat{p})^2}{\hat{p} (1 - \hat{p})}.
\]

Then \( z^2 = \chi^2 \) if

\[ (\ast) \quad n_1 (\hat{p}_1 - \hat{p})^2 + n_2 (\hat{p}_2 - \hat{p})^2 = \frac{n_1 n_2 (\hat{p}_1 - \hat{p}_2)^2}{n}.
\]

The left side of (\ast) is

\[
\text{LS of (\ast)} = n_1 \left( \frac{\hat{p}_1^2 - 2\hat{p}_1 \hat{p} + \hat{p}^2}{n} \right) + n_2 \left( \frac{\hat{p}_2^2 - 2\hat{p}_2 \hat{p} + \hat{p}^2}{n} \right)
\]
\[= \frac{n_1^2 \hat{p}_1^2}{n_1} - \frac{2n_1 m_1}{n} \hat{p} + \frac{n_2^2 \hat{p}_2^2}{n_2} - \frac{2n_2 m_2}{n} \hat{p} + \frac{m^2}{n}
\]
\[= \frac{n_1^2 \hat{p}_1^2}{n_1} + \frac{n_2^2 \hat{p}_2^2}{n_2} - \frac{m^2}{n}
\]
\[= \frac{n_2 (n_1 + n_2) \hat{p}_1^2 + n_1 (n_1 + n_2) \hat{p}_2^2 - 2n_1 m_1 n_2 + n_1 n_2 m_1^2}{n_1 n_2 n}
\]
\[= \frac{n_1 n_2 (\hat{p}_1^2 - \hat{p}_2^2)}{n},
\]

Therefore, \( z^2 = \chi^2 \).

(b) Our decision is to reject \( H_0 \) if \( |z| > z_{\alpha/2} \) or if \( \chi^2 = z^2 > z_{\alpha/2}^2 = \chi^2_{\alpha, 1} \), so the two tests are equivalent.

9.44 (a) No, one should not accept the explanation offered by the revenue service, because one should look closer at the different pass rates between men and women, separately for college educated and non-college educated employees.

(b) Case 1: For college educated men, \( \hat{p}_{cm} = 56/60 = 0.933 \) and for college educated women, \( \hat{p}_{cf} = 60/63 = 0.952. \) Then to test \( H_{0c}: p_{cm} = p_{cf} \) vs. \( H_{1c}: p_{cm} > p_{cf}, \) the test
The test statistic is

\[
z = \frac{\hat{P}_{cm} - \hat{P}_{cf}}{\sqrt{\frac{\hat{p}_{cm}\hat{q}_{cm}}{n_{cm}} + \frac{\hat{p}_{cf}\hat{q}_{cf}}{n_{cf}}}} = \frac{0.933 - 0.952}{\sqrt{\frac{(0.933)(0.067)}{60} + \frac{(0.952)(0.048)}{63}}} = -0.453.
\]

The \( P \)-value is

\[
P = 1 - \Phi(|-0.453|) = 0.326.
\]

Since \( P > \alpha = 0.05 \), do not reject \( H_{0c} \) and conclude that there is no difference in the pass rates for college educated men and women.

For non-college educated men, \( \hat{p}_{nm} = 12/55 = 0.218 \) and for non-college educated women, \( \hat{p}_{nf} = 8/188 = 0.043 \). Then to test \( H_{0n} : p_{nm} = p_{nf} \) vs. \( H_{1n} : p_{nm} > p_{nf} \), the test statistic is

\[
z = \frac{\hat{p}_{nm} - \hat{p}_{nf}}{\sqrt{\frac{\hat{p}_{nm}\hat{q}_{nm}}{n_{nm}} + \frac{\hat{p}_{nf}\hat{q}_{nf}}{n_{nf}}}} = \frac{0.218 - 0.043}{\sqrt{\frac{(0.218)(0.782)}{55} + \frac{(0.043)(0.957)}{188}}} = 3.041.
\]

The \( P \)-value is

\[
P = 1 - \Phi(3.041) = 0.001.
\]

Since \( P < \alpha = 0.05 \), reject \( H_{0n} \) and conclude that non-college educated men have a higher pass rate than non-college educated women. Hence, while there is no significant difference in the pass rates for college educated men and women, there is a significant difference in the pass rates for non-college educated men and women, invalidating the Revenue Service explanation.

**Case 2**: For college educated men, \( \hat{p}_{cm} = 56/60 = 0.933 \) and for college educated women, \( \hat{p}_{cf} = 27/63 = 0.429 \). Then to test \( H_{0c} : p_{cm} = p_{cf} \) vs. \( H_{1c} : p_{cm} > p_{cf} \), the test statistic is

\[
z = \frac{\hat{p}_{cm} - \hat{p}_{cf}}{\sqrt{\frac{\hat{p}_{cm}\hat{q}_{cm}}{n_{cm}} + \frac{\hat{p}_{cf}\hat{q}_{cf}}{n_{cf}}}} = \frac{0.933 - 0.429}{\sqrt{\frac{(0.933)(0.067)}{60} + \frac{(0.429)(0.571)}{63}}} = 7.187.
\]

The \( P \)-value is

\[
P = 1 - \Phi(7.187) \approx 0.
\]

Since \( P < \alpha = 0.05 \), reject \( H_{0c} \) and conclude that college educated men have a higher pass rate than college educated women. For non-college educated men, \( \hat{p}_{nm} = 12/55 = \)
0.2182, and for non-college educated women, \( \hat{p}_{nf} = 41/188 = 0.2181 \). Then to test \( H_0_n: p_{nm} = p_{nf} \) vs. \( H_1_n: p_{nm} > p_{nf} \), the test statistic is

\[
z = \frac{\hat{p}_{nm} - \hat{p}_{nf}}{\sqrt{\frac{\hat{p}_{nm}(1-\hat{p}_{nm})}{n_{nm}} + \frac{\hat{p}_{nf}(1-\hat{p}_{nf})}{n_{nf}}}}
\]

\[
= \frac{0.2182 - 0.2181}{\sqrt{\frac{0.2182(1-0.2182)}{188} + \frac{0.2181(1-0.2181)}{188}}}
\]

\[
= 0.002.
\]

The \( P \)-value is

\[
P = 1 - \Phi(0.002) = 0.500.
\]

Since \( P > \alpha = 0.05 \), do not reject \( H_0_n \) and conclude that there is no difference in the pass rates for non-college educated men and women. Hence, while there is no significant difference in the pass rates for non-college educated men and women, there is a significant difference in the pass rates for college educated men and women, invalidating the Revenue Service explanation.

9.45 (a)

\[
P(X = x|X + Y = m) = \frac{P(X = x, Y = m - x)}{P(X + Y = m)} = \frac{P(X = x)P(Y = m - x)}{\sum_i P(X = i)P(Y = m - i)}
\]

\[
= \frac{\binom{n_1}{i}p_1^{1-p_1}^{n_1-i}(n_2^{m-x}(1-p_2)^{n_2-m+x}}{\sum_i \binom{n_1}{i}p_1^{1-p_1}^{n_1-i}(n_2^{m-x}(1-p_2)^{n_2-m+x}}
\]

\[
= \frac{\binom{n_1}{i}p_1^{1-p_1}^{n_1-i}(n_2^{m-x}(1-p_2)^{n_2-m+x}}{\sum_i \binom{n_1}{i}p_1^{1-p_1}^{n_1-i}(n_2^{m-x}(1-p_2)^{n_2-m+x}}
\]

\[
i \text{ must be at least } 0 \text{ or } m - n_2, \text{ the maximum of } (0, n - 2), \text{ so that } \binom{n_1}{i} \text{ or } \binom{n_2}{m-x}\text{ are defined. Similarly, } i \text{ must be no larger than } n_1 \text{ or } m, \text{ the minimum of } (n_1, m), \text{ so that } \binom{n_1}{i} \text{ or } \binom{n_2}{m-x}\text{ are defined.}
\]

(b) \( H_0 : \psi = 1 \) vs. \( H_1 : \psi \neq 1 \). To find the power, first find the smallest \( x_1 \) and largest \( x_2 \) so that

\[
\sum_{i \geq x_1} \frac{\binom{n_1}{i} \binom{n_2}{m-i}}{\binom{n}{m}} \leq \alpha/2 \quad \text{and} \quad \sum_{i \leq x_2} \frac{\binom{n_1}{i} \binom{n_2}{m-i}}{\binom{n}{m}} \leq \alpha/2.
\]

Then the power is

\[
\text{Power} = \max(P(X \geq x_1|X + Y = m, \psi), P(X \leq x_2|X + Y = m, \psi))
\]

using the formula in (a).

(c) If \( \psi = 1 \), then

\[
P(X = x|X + Y = m) = \frac{\binom{n_1}{i} \binom{n_2}{m-x}}{\sum_i \binom{n_1}{i} \binom{n_2}{m-x}} = \frac{\binom{n_1}{i} \binom{n_2}{m-x}}{\binom{n}{m}}
\]