Chapter 5: Model solutions

1. Since we are interested in finding out whether or not there is an association between I.Q. and alcohol, we will perform a χ^2 test of independence. Our hypotheses are

$$\begin{aligned} H_0 &: \text{There is no association between alcohol and I.Q.} \qquad \text{versus} \\ H_1 &: \text{There } \textit{is} \text{ an association between alcohol and I.Q.} \end{aligned}$$

Recall that our test statistic is

$$X^2 = \sum \frac{(O-E)^2}{E},$$

where O and E are the observed and expected frequencies (respectively). Below are the calculations for our expected frequencies:

$$\begin{aligned} E_1 &= \frac{\text{row total for "Drinkers"} \times \text{column total for "Below norm"}}{\text{overall total}} \\ &= \frac{30 \times 21}{58} \\ &= 10.862. \end{aligned}$$

$$\begin{aligned} E_2 &= \frac{\text{row total for "Drinkers"} \times \text{column total for "Norm"}}{\text{overall total}} \\ &= \frac{30 \times 20}{58} \\ &= 10.345. \end{aligned}$$

$$\begin{aligned} E_3 &= \frac{\text{row total for "Drinkers"} \times \text{column total for "Above norm"}}{\text{overall total}} \\ &= \frac{30 \times 17}{58} \\ &= 8.793. \end{aligned}$$

$$\begin{aligned} E_4 &= \frac{\text{row total for "Non–drinkers"} \times \text{column total for "Below norm"}}{\text{overall total}} \\ &= \frac{28 \times 21}{58} \\ &= 10.138. \end{aligned}$$

$$\begin{aligned} E_5 &= \frac{\text{row total for "Non–drinkers"} \times \text{column total for "Norm"}}{\text{overall total}} \\ &= \frac{28 \times 20}{58} \\ &= 9.655. \end{aligned}$$

And finally,

$$
\begin{aligned}
E_6 &= \frac{\text{row total for “Non–drinkers”} \times \text{column total for “Above norm”}}{\text{overall total}} \\
&= \frac{28 \times 17}{58} \\
&= 8.207.
\end{aligned}
$$

Comparing our observed and expected frequencies, we have

O	E	$\frac{(O-E)^2}{E}$
12	10.862	0.1192
10	10.345	0.0115
8	8.793	0.0715
9	10.138	0.1277
10	9.655	0.0123
9	8.207	0.0766
		0.4188

Thus, our test statistic is $X^2 = \sum \frac{(O-E)^2}{E} = 0.4188$.

We now need to find our p–value. Our degrees of freedom is equal to

$$
\begin{aligned}
\nu &= (\text{number of rows} - 1) \times (\text{number of columns} - 1) \\
&= (2 - 1) \times (3 - 1) \\
&= 1 \times 2 \\
&= 2.
\end{aligned}
$$

Using table 3.1 from the lecture notes, we obtain the following critical values:

Significance level	10%	5%	1%
Critical value	4.61	5.99	9.21

Our test statistic $X^2 = 0.4188$ lies to the left of the first critical value, and so our p–value is bigger than 10%.

To conclude, since our p–value is so large, there is insufficient evidence to reject the null hypothesis, and so we retain H_0. It appears that there is no association between alcohol and I.Q., i.e. whether you are a “drinker” or a “non–drinker” has no effect on your I.Q. score! That’s good!

2. To see if the company has a discriminatory employment strategy, we need to perform a χ^2 test of independence to see if there's any association between "job type" and "gender". Thus,

$$\begin{aligned} H_0 &: \text{There is no association between job type and gender} \qquad \text{versus} \\ H_1 &: \text{There } \textit{is} \text{ an association between job type and gender.} \end{aligned}$$

As the question does not specify, we shall test the null hypothesis at the 5% level of significance.

There are 8 cells in the contingency table, so we have 8 expected frequencies to calculate. The first two calculations are shown below; the rest are given in the table below.

$$\begin{aligned} E_1 &= \frac{\text{row total for "Shelf Stackers"} \times \text{column total for "Male"}}{\text{Overall total}} \\ &= \frac{84 \times 95}{194} \\ &= 41.134. \end{aligned}$$

$$\begin{aligned} E_2 &= \frac{\text{row total for "Shelf Stackers"} \times \text{column total for "Female"}}{\text{Overall total}} \\ &= \frac{84 \times 99}{194} \\ &= 42.866. \end{aligned}$$

O	E	$\frac{(O-E)^2}{E}$
34	41.134	1.2373
50	42.866	1.1873
39	38.686	0.0025
40	40.314	0.0024
12	9.794	0.4969
8	10.206	0.4768
10	5.387	3.9502
1	5.613	3.7912
		11.1446

So our test statistic is $X^2 = 11.1446$.

A range for our p–value can be found from table 3.1. The degrees of freedom is equal to

$$\begin{aligned} \nu &= (\text{number of columns} - 1) \times (\text{number of rows} - 1) \\ &= (2-1) \times (4-1) \\ &= 1 \times 3 \\ &= 3. \end{aligned}$$

Using table 3.1 and with 3 degrees of freedom, we get the following critical values:

Significance level	10%	5%	1%
Critical value	6.25	7.82	11.34

Since our test statistic $X^2 = 11.1446$ lies in between the critical values 7.82 and 11.34, our p–value lies in the 1%–5% range.

Using table 2.1 to interpret our p–value, we see that there is moderate evidence against H_0, and so we reject the null hypothesis in favour of the alternative. It appears that "job type" and "gender" *are* associated, and so the type of job you do *does* depend on your sex. Thus, it appears that, based on the information in this sample anyway, the company *does* have a discriminatory employment strategy.