Exercise 6B.1

In a study of low birthweight and maternal smoking, the birth outcomes of 74 women who smoked during pregnancy were compared to the birth outcomes of 115 women who did not smoke during pregnancy. Of the 74 smokers, 30 had babies with low birthweight, and of the 115 non-smokers, 29 had babies with low birthweight.

(a) Construct a 2x2 contingency table for these data.

	Low Birthweight	Not Low Birthweight	Total
Smoker	30	44	74
Non-Smoker	29	86	115
Total	59	130	189

(b) What is the estimated probability that a baby is born with low birthweight among women who smoke? Among women who do not smoke?

Among women who smoke: $\frac{30}{74} = 0.405$

Among women who do not smoke: $\frac{29}{115} = 0.252$

(c) Calculate the expected counts for each cell in the table.

Common (overall) probability of low birthweight: $\hat{p} = \frac{30+29}{189} = \frac{59}{189} = 0.312$

Women who smoke:

  Expected # with low birthweight $= 0.312 \times 74 = 23.1$

  Expected # with not low birthweight $= (1 - 0.312) \times 74 = 50.9$

Women who do not smoke:

  Expected # with low birthweight $= 0.312 \times 115 = 35.9$

  Expected # with not low birthweight $= (1 - 0.312) \times 115 = 79.1$

(d) Calculate the chi-square test statistic using the expected counts you calculated in (c).

$$X^2 = \sum_{all\ cells} \frac{(Observed - Expected)^2}{Expected}$$

$$= \frac{(30 - 23.1)^2}{23.1} + \frac{(44 - 50.9)^2}{50.9} + \frac{(29 - 35.9)^2}{35.9} + \frac{(86 - 79.1)^2}{79.1}$$

$$= 2.06 + 0.935 + 1.33 + 0.602 = 4.93$$

(e) Test whether there is a significant difference in the probability of a low birthweight baby for women who smoke during pregnancy and women who do not smoke using a chi-square test. Make sure to write all steps of the hypothesis test and state a conclusion.

H_0 : smoking and having a low birthweight baby are not associated
H_a smoking and having a low birthweight baby are associated
(could also define p_1 and p_2 and write hypotheses in terms of these parameters)
test statistic $X^2 = 4.93$
Under H_0, $X^2 \sim \chi^2(1)$
Critical value method:
critical value $= 3.84$ (assuming $\alpha = 0.05$)
test statistic $4.93 >$ critical value $\rightarrow$ Reject H_0
P-value method:
p-value $= P(\chi^2(1) > 4.93) = 0.026$ from Stata   Stata code: display chi2tail(1, 4.93)
p-value $= 0.026 < 0.05 \rightarrow$ Reject H_0
Conclusion: There is evidence that there is a significant association between smoking during pregnancy and having a low birthweight baby. (Or: There is evidence that the proportion of mothers who smoke during pregnancy who have low birthweight babies is significantly different from the proportion of mothers who do not smoke during pregnancy.)

Exercise 6B.2

A study was done to compare the rate of chronic kidney disease (CKD) among diabetics and non-diabetics. The data from Stata are below. We are interested in whether the probability of CKD is the same for diabetics and non-diabetics.

```
. tabulate group ckd

               |          CKD
        group  |      yes         no  |     Total
---------------+----------------------+----------
      diabetics |       30         80  |       110
  non-diabetics |       35        190  |       225
---------------+----------------------+----------
         Total  |       65        270  |       335
```

(a) What is the estimated probability of CKD in each group (diabetics, non-diabetics)?

p_1 = probability of CKD among diabetics
p_2 = probability of CKD among non-diabetics
$$\hat{p}_1 = \frac{30}{110} = 0.273$$
$$\hat{p}_2 = \frac{35}{225} = 0.156$$

(b) Use a two-sample z-test to test whether the probability of CKD is different for diabetics and non-diabetics in the population.

$H_0 : p_1 = p_2$
$H_a : p_1 \neq p_2$

For test statistic we need the common (overall) proportion with CKD: $\hat{p} = \frac{65}{335} = 0.194$
And also the sample sizes: $n_1 = 110, n_2 = 225$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.273 - 0.156}{\sqrt{0.194(1-0.194)\left(\frac{1}{110} + \frac{1}{225}\right)}} = \frac{0.117}{0.0460} = 2.54$$

Under $H_0, Z \sim N(0,1)$
p-value $= 2 \times P(Z > |2.54|) = 2 \times 0.0055 = 0.011$
Since $0.011 < 0.05$, we reject H_0.
We have evidence that there is a significant difference in the probability of CKD for diabetics and non-diabetics.

(c) Perform a chi-square test to test the same hypothesis. Confirm that the conclusion is the same.

Hypotheses same as in (b), or we can write:
H_0 : diabetes and CKD are not associated (are independent)
H_a : diabetes and CKD are associated (are not independent)

Shortcut formula for test statistic: $X^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(c+d)(a+b)} = \frac{335(30 \times 190 - 80 \times 35)^2}{65 \times 270 \times 225 \times 110} = 6.49$

Under $H_0, X^2 \sim \chi^2(1)$
p-value $= P(\chi^2(1) > 6.49) = 0.011$ from Stata   Stata code: `display chi2tail(1,6.49)`
Same p-value as in (b), so we draw the same conclusion.