1. (17.4, p. 387)

   a) $\hat{p}_1 = \dfrac{40}{244} = 0.1639$, $\hat{p}_2 = \dfrac{87}{245} = 0.3551$

   b)
   - We will test the hypothesis $H_0 : p_1 = p_2$ against the two-sided alternative $H_A : p_1 \neq p_2$.
   - We set our significance level at $\alpha = 0.05$ and calculate the following:
   - $\bar{p} = \frac{40+87}{244+245} = \frac{127}{489} = 0.2597 \Rightarrow \bar{q} = 1 - \bar{p} = 0.7403$

     $$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{\frac{127}{489}\left(\frac{362}{489}\right)\left(\frac{1}{244} + \frac{1}{245}\right)} = 0.0397$$

   - $z_{\text{stat}} = \dfrac{\hat{p}_1 - \hat{p}_2}{SE_{\hat{p}_1 - \hat{p}_2}} = \dfrac{0.1639 - 0.3551}{0.0397} = \dfrac{-0.1912}{0.0397} = -4.816$
   - Using Table B from the book, we see that

     $$z_{\text{stat}} < -3.49 \Rightarrow p_{\text{1-sided}} < 0.0002 \Rightarrow p_{\text{2-sided}} < 0.0004$$

   - Our 2-sided p-value is much less than α, therefore we reject the null hypothesis that the two proportions are equal.
   - We have sufficient evidence to conclude that the proportion remaining smoke-free in the treatment group is significantly different that the proportion remaining smoke-free in the control group.

   c) Based on the results our randomized controlled double-blind trial, we have strong evidence indicating that use of sustained-release buproprion in combination with the nicotine patch helped more individuals remain smoke-free than did use of the nicotine patch alone.

2. (17.6, p. 388)

   (a) $\widehat{RD} = \hat{p}_1 - \hat{p}_2 = \dfrac{18}{41} - \dfrac{9}{36} = 0.4390 - 0.25 = 0.1890$

   (b)
   - We will test the hypothesis $H_0 : p_1 = p_2$ against the two-sided alternative $H_A : p_1 \neq p_2$.
   - We set our significance level at $\alpha = 0.05$ and calculate the following:
   - $SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\bar{p}\bar{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)} = \sqrt{\dfrac{27}{77}\left(\dfrac{50}{77}\right)\left(\dfrac{1}{41} + \dfrac{1}{36}\right)} = 0.109$
   - $z_{\text{stat}} = \dfrac{\hat{p}_1 - \hat{p}_2}{SE_{\hat{p}_1 - \hat{p}_2}} = \dfrac{0.189}{0.109} = 1.734$
   - Using Table B from the book, we obtain $p_{\text{1-sided}} = 0.0418 \Rightarrow p_{\text{2-sided}} = 0.0836$
   - Our 2-sided p-value is larger than α, therefore we cannot reject the null hypothesis that the two proportions are equal.

- We do not have sufficient evidence to conclude that the proportion of deaths among those receiving early bloodletting is significantly different from the proportion of deaths among those receiving late treatment.

3. (17.8, p. 389)

$\widehat{RD} = \hat{p}_1 - \hat{p}_2 = \frac{1}{111} - \frac{13}{117} = 0.009 - 0.1111 = -0.1021$

To calculate the 95% confidence interval for the risk difference, we use the plus-4 method.

- $\widetilde{RD} = \tilde{p}_1 - \tilde{p}_2 = \frac{2}{113} - \frac{14}{119} = 0.0177 - 0.1176 = -0.0999$
- $SE_{\tilde{p}_1 - \tilde{p}_2} = \sqrt{\frac{\tilde{p}_1\tilde{q}_1}{\tilde{n}_1} + \frac{\tilde{p}_2\tilde{q}_2}{\tilde{n}_2}} = \sqrt{\frac{\frac{2}{113}\left(\frac{111}{113}\right)}{113} + \frac{\frac{14}{119}\left(\frac{105}{119}\right)}{119}} = \sqrt{0.001026} = 0.0320$
- $\alpha = 0.05 \;\Rightarrow\; z_{1-\alpha/2} = 1.96$
- Therefore, by the plus-4 method, we obtain the following 95% CI for the difference in population proportions of coronary incidents between women with high and low serum cholesterol.

$$
\begin{aligned}
95\%\ CI \quad : \quad \widetilde{RD} \pm z_{1-\alpha/2} \cdot SE_{\tilde{p}_1 - \tilde{p}_2} &= -0.0999 \pm 1.96(0.0320) = -0.0999 \pm 0.0627 \\
&= (-0.1626, -0.0372)
\end{aligned}
$$

4. (17.18, p. 405)

$\widehat{RR} = \frac{\hat{p}_1}{\hat{p}_2} = \frac{182/2221}{256/2223} = \frac{0.082}{0.1152} = 0.712$

This suggests that the risk of death in the treatment group is 0.71 times the risk of death in the control group.

To calculate the 95% confidence interval for the risk ratio, we first work with its natural log, $\ln(\widehat{RR})$.

- $\ln(\widehat{RR}) = \ln(0.712) = -0.340$
- $SE_{\ln(\widehat{RR})} = \sqrt{\frac{1}{a_1} - \frac{1}{n_1} + \frac{1}{a_2} - \frac{1}{n_2}} = \sqrt{\frac{1}{182} - \frac{1}{2221} + \frac{1}{256} - \frac{1}{2223}} = 0.0922$
- $\alpha = 0.05 \;\Rightarrow\; z_{1-\alpha/2} = 1.96$
- Using the usual formula, we obtain the following 95% CI for $\ln(\widehat{RR})$.

$$
\begin{aligned}
95\%\ CI_{\ln(\widehat{RR})} \quad : \quad \ln(\widehat{RR}) \pm z_{1-\alpha/2} \cdot SE_{\ln(\widehat{RR})} &= -0.340 \pm 1.96(0.0922) = -0.340 \pm 0.1807 \\
&= (-0.521, -0.159)
\end{aligned}
$$

Exponentiating this CI, we obtain the 95% CI for $\widehat{RR}$.

$$95\%\ CI_{\widehat{RR}} : \left(e^{-0.521}, e^{-0.159}\right) = (0.5939, 0.8530)$$

We estimate the risk of death in the treatment group to be between 0.5939 and 0.853 times the risk of death in the control group.

Since the baseline (null) value 1 is not in the CI, then we would reject the null hypothesis that the risk of death is the same in both groups. That is, at $\alpha = 0.05$, we have evidence that the risk of death in the treatment group is significantly different (lower) than that of the control group.

5. (18.4, p. 420) We see that as anger-level increases, incidence of coronary heart disease appears to increase. There seems to be a positive association between anger-level and CHD.

   $$\hat{p}_{\text{Low}} = \frac{31}{3110} = 0.00997, \quad \hat{p}_{\text{Moderate}} = \frac{63}{4731} = 0.01332, \quad \hat{p}_{\text{High}} = \frac{18}{633} = 0.02844$$

6. (18.10, p. 431) For the χ^2 test, we will need both the observed and expected counts.

   Observed Counts:

Anger-Level	**CHD+**	**CHD-**	**Total**
Low	31	3079	3110
Moderate	63	4668	4731
High	18	615	633
Total	112	8362	8474

   Expected Counts:

Anger-Level	**CHD+**	**CHD-**	**Total**
Low	41.105	3068.895	3110
Moderate	62.529	4668.471	4731
High	8.366	624.634	633
Total	112	8362	8474

   - We will test the null hypothesis $H_0 : \{\text{no association between anger-level and CHD}\}$ against the two-sided alternative $H_A : \{H_0 \text{ is false}\}$.
   - We set our significance level at $\alpha = 0.05$ and calculate the following:

- $$\begin{aligned}
X^2_{\text{stat}} &= \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right] \\
&= \frac{(31-41.105)^2}{41.105} + \frac{(63-62.529)^2}{62.529} + \frac{(18-8.366)^2}{8.366} + \frac{(3079-3068.895)^2}{3068.895} \\
&\quad + \frac{(4668-4668.471)^2}{4668.471} + \frac{(615-624.634)^2}{624.634} \\
&= \frac{(-10.105)^2}{41.105} + \frac{(0.471)^2}{62.529} + \frac{(9.634)^2}{8.366} + \frac{(10.105)^2}{3068.895} + \frac{(-0.471)^2}{4668.471} + \frac{(-9.634)^2}{624.634} \\
&= 2.4842 + 0.00355 + 11.09419 + 0.03327 + 0.000047519 + 0.14859 \\
&\approx 13.76
\end{aligned}$$

- Using Table E from the book, we can calculate the p-value.
  Since $c = 2$, $r = 3 \Rightarrow df = (r-1)(c-1) = (2)(1) = 2$, we have that $10.60 < X^2_{\text{stat}} < 13.82$, therefore we know $0.001 < p < 0.005$.
- Our p-value is less than α, therefore we reject the null hypothesis of no association between anger-level and CHD.
- We have strong evidence to conclude that there is an association between one's anger level and the development of coronary heart disease.

7. (18.16, p. 444)

   $\widehat{OR} = \dfrac{61 \cdot 165}{93 \cdot 114} = 0.9493$

   To calculate the 95% confidence interval for the odds ratio, we first work with its natural log, $\ln(\widehat{OR})$.

   - $\ln(\widehat{OR}) = \ln(0.9493) = -0.0520$
   - $SE_{\ln(\widehat{OR})} = \sqrt{\dfrac{1}{a_1} + \dfrac{1}{b_1} + \dfrac{1}{a_2} + \dfrac{1}{b_2}} = \sqrt{\dfrac{1}{61} + \dfrac{1}{93} + \dfrac{1}{165} + \dfrac{1}{114}} = 0.2049$
   - $\alpha = 0.05 \Rightarrow z_{1-\alpha/2} = 1.96$
   - Using the usual formula, we obtain the following 95% CI for $\ln(\widehat{OR})$.

     $$\begin{aligned}
     95\%\ CI_{\ln(\widehat{OR})} \ :\ & \ln(\widehat{OR}) \pm z_{1-\alpha/2} \cdot SE_{\ln(\widehat{OR})} = -0.052 \pm 1.96(0.2049) = -0.052 \pm 0.4016 \\
     & = (-0.4536, 0.3496)
     \end{aligned}$$

     Exponentiating this CI, we obtain the 95% CI for $\widehat{OR}$.

     $$95\%\ CI_{\widehat{OR}} : \left(e^{-0.4536}, e^{0.3496}\right) = (0.6353, 1.4185)$$

8. (18.26, p. 459)

Exposure	Cases	Controls	Total
Used IUD	89	640	729
No IUD use	194	3193	3387
Total	283	3833	4116

$\widehat{OR} = \dfrac{89 \cdot 3193}{640 \cdot 194} = 2.2888$

This suggests that the odds of infertility among IUD users are 2.29 times that of non-IUD users.

To calculate the 95% confidence interval for the odds ratio, we first work with its natural log, $\ln(\widehat{OR})$.

- $\ln(\widehat{OR}) = \ln(2.2888) = 0.828$
- $SE_{\ln(\widehat{OR})} = \sqrt{\dfrac{1}{a_1} + \dfrac{1}{b_1} + \dfrac{1}{a_2} + \dfrac{1}{b_2}} = \sqrt{\dfrac{1}{89} + \dfrac{1}{194} + \dfrac{1}{640} + \dfrac{1}{3193}} = 0.1352$
- $\alpha = 0.05 \;\Rightarrow\; z_{1-\alpha/2} = 1.96$
- Using the usual formula, we obtain the following 95% CI for $\ln(\widehat{OR})$.

$$\begin{aligned} 95\%\ CI_{\ln(\widehat{OR})} \quad : \quad & \ln(\widehat{OR}) \pm z_{1-\alpha/2} \cdot SE_{\ln(\widehat{OR})} = 0.828 \pm 1.96(0.1352) = 0.828 \pm 0.265 \\ & = (0.563, 1.093) \end{aligned}$$

Exponentiating this CI, we obtain the 95% CI for $\widehat{OR}$.

$$95\%\ CI_{\widehat{OR}} : \left(e^{0.563}, e^{1.093}\right) = (1.756, 2.983)$$

That is, we estimate that the odds of infertility among IUD users is between 1.756 and 2.983 times the odds of infertility among non-IUD users.

Since the baseline (null) value 1 does not fall within the CI, we can reject the null hypothesis that the odds of infertility are the same in the two groups. Therefore, at $\alpha = 0.05$ we have evidence that the odds of infertility are significantly different (higher) in IUD users.