Homework 7 Solution
BIOS 600

1. (17.4) (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:
- $SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\bar{p}\bar{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)} = \sqrt{\dfrac{127}{489}\left(\dfrac{362}{489}\right)\left(\dfrac{1}{244} + \dfrac{1}{245}\right)} = \sqrt{0.001573} = 0.039657$
- $z_{\text{stat}} = \dfrac{\hat{p}_1 - \hat{p}_2}{SE_{\hat{p}_1 - \hat{p}_2}} = \dfrac{0.1639 - 0.3551}{0.039657} = \dfrac{-0.19117}{0.03089} = -4.8205$
- Using Table B from the book, we see that $z_{\text{stat}} < -3.49 \Rightarrow p_{\text{1-sided}} < 0.0002 \Rightarrow p_{\text{2-sided}} < 0.0004$ Exact p-values from software: $p_{\text{1-sided}} = 0.0000007 \implies p_{\text{2-sided}} = 0.0000014$.
- 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) (a) $\widehat{RD} = \hat{p}_1 - \hat{p}_2 = \dfrac{18}{41} - \dfrac{9}{36} = 0.4390244 - 0.25 = 0.1890244$

(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:
- $SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\bar{p}\bar{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)} = \sqrt{\dfrac{27}{77}\left(\dfrac{50}{77}\right)\left(\dfrac{1}{41} + \dfrac{1}{36}\right)} = \sqrt{0.0118784} = 0.1089879$
- $z_{\text{stat}} = \dfrac{\hat{p}_1 - \hat{p}_2}{SE_{\hat{p}_1 - \hat{p}_2}} = \dfrac{0.1890244}{0.1089879} = 1.734$
- 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. (17.16) $\widehat{RD} = \hat{p}_1 - \hat{p}_2 = \dfrac{30}{689} - \dfrac{8}{445} = 0.043541 - 0.017978 = 0.025563$

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

- $\widetilde{RD} = \tilde{p}_1 - \tilde{p}_2 = \dfrac{31}{691} - \dfrac{9}{447} = 0.044862 - 0.020134 = 0.024728$
- $SE_{\tilde{p}_1 - \tilde{p}_2} = \sqrt{\dfrac{\tilde{p}_1\tilde{q}_1}{\tilde{n}_1} + \dfrac{\tilde{p}_2\tilde{q}_2}{\tilde{n}_2}} = \sqrt{\dfrac{\frac{31}{691}\left(\frac{660}{691}\right)}{691} + \dfrac{\frac{9}{447}\left(\frac{438}{447}\right)}{447}} = \sqrt{0.0001061} = 0.0103028$
- $\alpha = 0.05 \Rightarrow z_{1-\alpha/2} = 1.96$
- 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.

$$\begin{aligned} 95\%\ CI \quad : \quad & \widetilde{RD} \pm z_{1-\alpha/2} \cdot SE_{\tilde{p}_1 - \tilde{p}_2} = 0.024728 \pm 1.96(0.0103028) = 0.024728 \pm 0.0201935 \\ & = (0.0045348, 0.0449218) \end{aligned}$$

4. (17.18) $\widehat{RR} = \dfrac{\hat{p}_1}{\hat{p}_2} = \dfrac{182/2221}{256/2223} = \dfrac{0.081945}{0.1151597} = 0.7115777$

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

- $\ln(\widehat{RR}) = \ln(0.7115777) = -0.3402707$
- $SE_{\ln(\widehat{RR})} = \sqrt{\dfrac{1}{a_1} - \dfrac{1}{n_1} + \dfrac{1}{a_2} - \dfrac{1}{n_2}} = \sqrt{\dfrac{1}{182} - \dfrac{1}{2221} + \dfrac{1}{256} - \dfrac{1}{2223}} = \sqrt{0.0085} = 0.092199$
- $\alpha = 0.05 \;\Rightarrow\; z_{1-\alpha/2} = 1.96$
- Using the usual formula, we obtain the following 95% CI for $\ln(\widehat{RR})$.

$$95\%\ CI_{\ln(\widehat{RR})} \;:\; \ln(\widehat{RR}) \pm z_{1-\alpha/2} \cdot SE_{\ln(\widehat{RR})} = -0.3402707 \pm 1.96(0.092199) = -0.3402707 \pm 0.18071$$
$$= (-0.52098, -0.15956)$$

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

$$95\%\ CI_{\widehat{RR}} : \left(e^{-0.52098}, e^{-0.15956}\right) = (0.59394, 0.85252)$$

5. (18.4) 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.

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

6. (18.10) 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 : {no association between anger-level and CHD} against the two-sided alternative H_A : {H_0 is false}.
- We set our significance level at $\alpha = 0.05$ and calculate the following:
-

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

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

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

- $\ln(\widehat{OR}) = \ln(0.9493) = -0.051979$
- $SE_{\ln(\widehat{OR})} = \sqrt{\dfrac{1}{a_1} + \dfrac{1}{b_1} + \dfrac{1}{a_2} + \dfrac{1}{b_2}} = \sqrt{\dfrac{1}{61} + \dfrac{1}{93} + \dfrac{1}{165} + \dfrac{1}{114}} = 0.204887$
- $\alpha = 0.05 \;\Rightarrow\; z_{1-\alpha/2} = 1.96$
- Using the usual formula, we obtain the following 95% CI for $\ln(\widehat{OR})$.

$$95\%\ CI_{\ln(\widehat{OR})} \;:\; \ln(\widehat{OR}) \pm z_{1-\alpha/2} \cdot SE_{\ln(\widehat{OR})} = -0.051979 \pm 1.96(0.204887) = -0.051979 \pm 0.401578$$
$$= (-0.453557, 0.349560)$$

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

$$95\%\ CI_{\widehat{OR}} : \left(e^{-0.453557}, e^{0.349560}\right) = (0.6354, 1.4185)$$