STA 100 HW 8 Solutions

Dr. Erin K. Melcon

1. (a) First, $\tilde{p}_1 = \frac{y_1+1}{n_1+2} = \frac{23+1}{99+2} = 0.2376$, and $\tilde{p}_2 = \frac{y_2+1}{n_2+2} = \frac{44+1}{105+2} = 0.4206$
   Then, the $(1-0.05)100\%$ (95)% confidence interval for the A vs. B group is:
   $\tilde{p_1} - \tilde{p_2} \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p_1}(1-\tilde{p_1})}{n_1+2} + \frac{\tilde{p_2}(1-\tilde{p_2})}{n_2+2}}$ or -0.183 $\pm$ (1.96)(0.0638) or (-0.308, -0.058).
   Thus the lower bound is -0.308.

   (b) Thus the upper bound is -0.058.

   (c) We are 95% confident that the true proportion of subjects with the disease in the B group is higher than that in the A group by between 0.058 and 0.308.

   (d) No. The interval is A - B, and the bounds are negative, which would suggest that group B has a higher proportion of disease.

   (e) The largest difference would be the most negative bound, or -0.308.

   (f) It would widen, since $Z_{\alpha/2}$ would increase to 2.576

---

2. (a) First, $\tilde{p}_1 = \frac{y_1+1}{n_1+2} = \frac{15+1}{100+2} = 0.1569$, and $\tilde{p}_2 = \frac{y_2+1}{n_2+2} = \frac{20+1}{100+2} = 0.2059$
   Then, the $(1-0.01)100\%$ (99)% confidence interval for the A vs. B group is:
   $\tilde{p_1} - \tilde{p_2} \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p_1}(1-\tilde{p_1})}{n_1+2} + \frac{\tilde{p_2}(1-\tilde{p_2})}{n_2+2}}$ or -0.049 $\pm$ (2.576)(0.0539) or (-0.1878, 0.0898).
   Thus the lower bound is -0.1878.

   (b) Thus the upper bound is 0.0898.

   (c) We are 99% confident that there is no significant difference in the true probability of having side effects for the New Drug vs. the Old Drug.

   (d) Since the confidence interval contains zero, the smallest difference we would expect is zero.

   (e) Yes, since it was stated a random sample was taken, the groups are independent, and we can see that there are at least 5 subjects in the New and Old drug group that showed side effects, and at least 5 that did not.

---

3. (a) The response variable is: peak flow.

   (b) The explanatory variable is: height.

   (c) The estimated intercept is: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 660 - (180.4118)(4.5114) = -153.9098$

   (d) The estimate slope is: $\hat{\beta}_1 = r\frac{s_y}{s_x} = 0.32725\frac{117.9952}{8.5591} = 4.5114$

   (e) When the height of a male increases by 1 cm, we expect peak flow to increase by 4.5114 liters/min on average.

   (f) The intercept has no practical meaning, since a male cannot be height 0 cm.

   (g) The predicted peak flow is: $\hat{y} = -153.9098 + (4.5114)174 = 631.0738$

---

4. (a) The 95% confidence interval is: $b_1 \pm t_{\alpha/2}\frac{s_e}{s_x\sqrt{n-1}}$ or $4.5114 \pm (2.131)\frac{3.6415}{8.5591\sqrt{16}}$ or $(4.2847, 4.7381)$.
   Thus, the lower bound is 4.2847

   (b) From (a) the upper bound is 4.7381

   (c) We are 95% confident that when height increases by 1 cm, peak flow increases by between 4.2847 and 4.7381 liters/min. on average.

   (d) $10\hat{\beta}_1$, or $10(4.5114) = 45.114$ liters/min.

   (e) Outliers can severely effect the estimates of the slope and intercept, and should be removed.

   (f) $s_e = \sqrt{\frac{SS(resid)}{n-2}} = \sqrt{\frac{198.909}{17-2}} = 3.6415.$

   (g) A typical error when estimating peak flow with height using linear regression is 3.6415 liters/min.

(h) $(r)^2 = 0.1071$, so the proportion of reduction of error when use the linear relationship between peak flow and height to predict peak flow (rather than using the sample mean of peak flow to predict peak flow) is 0.1071
OR
The reduction in the variance of peak flow when using the linear relationship between peak flow and height is 0.1071

---

5. (a) FALSE. When the value of X=0 takes on no practical meaning, the value of the intercept will not have a practical interpretation.

(b) FALSE. It means there is no linear relationship between the two variables, however there may be a non-linear relationship.

(c) TRUE. If the value of r is zero, since $b_1 = r\frac{s_y}{s_x}$, the value of the slope is also zero.

(d) TRUE. The regression line represents the average response of Y when X = x, not a particular response.
