EPIB-621 Solutions 2010 Final Exam

1. (a) Using usual methods for finding two sample confidence intervals, one calculates:

$$(12 - 3) \pm 1.96 \times \sqrt{\frac{5^2}{200} + \frac{4^2}{200}} \longrightarrow (8.113, 9.887)$$

(b) Similar to the above, one calculates:

$$(.6 - .4) \pm 1.96 \times \sqrt{\frac{.3^2}{200} + \frac{.3^2}{200}} \longrightarrow (0.1412, 0.2588)$$

(c) Both HDL and LDL show true increases that are <u>not</u> clinically relevant.

2. (a) When $X = 0$, $Y = 0.4$, so $\alpha = 0.4$. Approximate 95% interval would be given by:

$$0.4 \pm 1.96 \times \frac{0.3}{\sqrt{200}} \longrightarrow (0.3584, 0.4416)$$

(b) Yes, one can solve for $\hat{\beta}$ by isolating β in the regression equation $\overline{y} = \alpha + \beta * \overline{x}$, giving:

$$\hat{\beta} = \frac{0.5 - 0.4}{0.5} = 0.2$$

3. (a) Main reason is that several variables are giving similar information, i.e., collinearity. Possibly also a sample size issue, but that is a less important reason.

(b) From graph, can take any two points, such as (0, 1) and (40, 3.3), and calculate the slope from that as follows:

$$\frac{\Delta Y}{\Delta X} \approx \frac{2.3}{40} = 0.0575$$

4. (a) $OR = \exp(0.1536) = 1.166$.

(b) $OR = \exp(\beta_2) \times \exp(\beta_4) = \exp(0.1536) \times \exp(-.4656) = 0.732$.

5. (a) Yes, there is confounding as indicated by changes in coefficients from one model to the next for both X and Z, as one or the other enters/exits the model.

(b) Use model including Z, because it is confounded with X, so need to adjust. So, coefficient is $\beta_x = 0.001737$, with SE $= 0.0125127$. Using usual confidence interval formula for OR, one calculates:

$$OR = \exp(0.001737) = 1.0017$$

$$(\exp(0.001737 \pm 1.96 \times 0.0125127) \longrightarrow (0.97797, 1.02608)$$

(c) One cannot conclude no effect, since the range of X is large. For example, $1.026^{10} = 1.29$ which is a large OR for a change of 10 points on the x scale. So, inconclusive.

6. (a) Fits well, very few points far from the line, if any, all well within chance boundaries.

(b) If model fits well, then this is a binomial error, with maximum probability of 0.5 occurring at the point 0.5 (half above, half below). Calculating the SD of this binomial distribution gives (note that n= 1000 and 20 bins, sample size per bin is 50)

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.5(1-5)}{50}} = 0.0707$$

So expect that about 95% of points will be within about $2 \times 0.07 = 0.14$ of the line in either direction at this point.

7. (a) From the results for w, we find 0.008477 (0.0039, 0.01625).

(b) From results for pstep, it is a near certainty that Ontario differs from BC.

8. Using the notation from the course notes, we have $\beta^* = 3$ and $\tau = \frac{1}{3}$. So, plugging into the formula that adjusts for measurement error, we have:

$$\beta = \beta^*(1 + \tau^2) = 3(1 + 1/3^2) = 3.33$$