EPIB-621 Solutions 2010 Midterm Exam

1. (a) This answer will of course vary from researcher to researcher, here is one example: I do not have much idea of this proportion, but doubt it is less than 20%, or higher than 80%, and my best guess would be 50%. So, I will look for a beta density with a mean of 50%, and an SD of 15%, so that, approximately, 95% of my prior density will cover the range from 20% to 80%. Plugging $\mu = 0.5$ and $\sigma = 0.15$ into the equations that convert means and SDs into beta coefficients

$$\alpha = -\frac{\mu\,(\sigma^2 + \mu^2 - \mu)}{\sigma^2}$$

and

$$\beta = \frac{(\mu - 1)\,(\sigma^2 + \mu^2 - \mu)}{\sigma^2}$$

one finds $\alpha = 5.056$ and $\beta = 5.056$.

(b) One simply needs to add the number of successes $x = 60$ to α and the number of failures $(n - x) = 140$ to β from part (a), find find the posterior is beta(65.056, 145.056).

(c) Plugging into the formulae for beta density means and variances, we find:

$$\mu = \frac{\alpha}{\alpha + \beta} = \frac{65.056}{65.056 + 145.056} = 0.31$$

and

$$\sigma = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}} = 0.032$$

2. (a) Yes, the assumptions are generally satisfied, no pattern in the scatter plot to suggest non-linearity, no change in variance across the range of points, and only a slight skewness in the histogram. While this may have some small effect in the estimation of coefficients, it is probably not a large concern.

(b) One needs to calculate:

$$log(chol) = 5.957392 - 0.014243(30) = 5.524$$

so that

$$chol = exp(5.524) = 250.66$$

3. (a) From the intercept information, we calculate

$$6 \pm 1.96(0.026) \longrightarrow (5.95, 6.05)$$

(b) I would be concerned about reporting this since it is well outside the range of the data. One needs to assume linearity continues to hold in this range, which cannot be known from these data.

(c) We calculate:

$$-0.0163587 + 5 * 0.0021123 = -0.0057972$$

4. (a) There is much concern about confounding since X_1 and X_2 are both related to the outcome Y, which we can see from the regression coefficients, and to each other, which we can see from the way the mean of X_2 changes depending on the value of X_1.

(b) From the course notes on confounding, we know that

$$\beta_1^* = \beta_1 - \beta_2(\overline{X}_{2|X_1=1} - \overline{X}_{2|X_1=0})$$

So, after a small bit of algebra, we calculate:

$$\beta_1 = \beta_1^* + \beta_2(\overline{X}_{2|X_1=1} - \overline{X}_{2|X_1=0})$$

Plugging in, we find

$$\beta_1 = 8 - 2(6 - 3.5) = 3$$

5. (a) From out$postmean in the output, we find (approximately):

$$39.797 + 1.12age + 0.2033sex + 0.998cardio + 2.5hip + 2.6fit$$

(b) Plugging into the equation from (a), we find:

$$39.797 + 1.12(50) + 0.2033(1) + 0.998(0) + 2.5(0) + 2.6(5) = 109.04 \text{ seconds.}$$

(c) We would use model 1, which is the highest ranked model with two variables in it. Thus, the model is:

$$38.65 + 1.147age + 2.615fit$$