EPIB-621 Solutions 2015 Midterm Exam

1. (a) We use the formulae that convert means and SDs to beta coefficients:

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

and

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

Plugging in $\mu = 0.01$ and $\sigma = 0.05$, we get $\alpha = 0.0296$ and $\beta = 2.9304$.

(b) Sample size equavalence of the prior information is equal to the sum of prior parameters. Here we have $\alpha + \beta = 0.0296 + 2.9304 = 2.96$ or approximately 3.

(c) We observe 1 "success" and 49 "failures" so our posterior beta parameters are $\alpha = 0.0296 + 1 = 1.0296$ and $\beta = 49 + 2.9304 = 51.9304$.

(d) The formula for obtaining means and SDs from the beta parameters are:

$$\mu = \frac{\alpha}{\alpha + \beta}$$

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

Plugging in, we obtain mean = 0.01944109 and SD = 0.01879581.

2. We know that

$$\alpha = \overline{y} - \beta \times \overline{x}$$

We know that $\beta = 1.8$, and from the above table, we know $\overline{y} = 6.5$ and $\overline{x} = 3.5$. Therefore, plugging in, $\alpha = 6.5 - 1.8 \times 3.5 = 0.2$.

(a) 3. Mean with no CVD = intercept = 7.161. Mean with CVD adds the slope for CVD, so is equal to 7.161 + 3.231 = 10.392.

(b)We are 95% certain that the true slope is between (0.37, 6.09). Thus, there is likely to be at least a 0.37 increase in troponin with CVD, and the increase could be as large as 6.09.

(c) (i) Linearity holds, since there are only two points on the x-axis, and a straight line always connects two points. (ii) Normality of residuals is probably close enough for reasonable estimation, but a possibly slight skew to the right. There are only 50 data points, so do not expect perfect normality. (iii) Spread of points at zero is not too different from spread of points at 1, but perhaps somewhat higher variance for CVD = 1 compared to CVD = 0.

4. Likely no confounding, since x_2 has little to no effect on the outcome, one of the conditions for confounding. In addition, the point estimate for x_1 is stable whether x_2 is in the model or not. However, the SD for x_1 does increase in the presence of x_2, presumably because the high correlation creates some instability in the model.

5. (a) The design is balanced, in other words, there is the same sample size for noisy and quiet environments within each age range. Therefore there is zero correlation between environment and age, thus no confounding.

(b) On average, there is approximately 3.1 dB loss in noisy environments compared to quiet environments. Looking at the CI, we are 95% certain (because a Bayesian analysis would provide similar numerical estimates from flat priors) the true degree of loss from noisy environments is between 2.7 and 3.6 dBs.

(c) This would be similar to fitting the following regression:

0	0
1	-0.8
2	-1.5
3	-2.9

Here we see decreases of 0.8, 0.7 and 1.4, which averages out to a decrease of approximately $\frac{0.8+0.7+1.4}{3} = 2.9/3 \approx 1$. So I would expect the coefficient to be close to -1.

6. (a) This would imply that there is no unique effect of age or environ, but rather than the effect of age depends on the environment, and conversely, the effect of the environment varies with age.

(b) Not much evidence for an interaction term, but wide confidence intervals preclude definitive conclusions. One would need a larger sample size to narrow the CIs to derive a stronger conclusion, but no strong evidence so far.