EPIB-621 Solutions 2008 Midterm Exam

1. (a) Mean procedure time is found directly in the printout, as 90.520 minutes, with standard error 1.422. Hence the CI is calculated as:

$$\begin{aligned}\text{lower limit} &= 90.520 - 1.96 \times 1.422 = 87.73 \\ \text{upper limit} &= 90.520 + 1.96 \times 1.422 = 93.31\end{aligned}$$

(b) Directly from the printout, we find the effect for the placebo group is 90.520 minutes, while the difference from placebo to treatment is -4.86, so that the mean time for the treatment group is $90.520 - 4.86 = 85.66$ minutes.

(c) Between group difference = -4.86 minutes. From the printout, the SE for this difference is 2.012 minutes, so we can calculate the CI as follows:

$$\begin{aligned}\text{lower limit} &= -4.86 - 1.96 \times 2.012 = -8.804 \\ \text{upper limit} &= -4.86 + 1.96 \times 2.012 = -0.916\end{aligned}$$

From the above 95% CI, we c0onclude that the difference could be as much as 9 minutes (approximately), as as little than less than one minutes. Assuming that 9 minutes (about a 10% change) is considered as clinically meaningful, this interval shows that there is at least some effect, but the clinical meaningfulness of the effect is not clear, as we have not ruled out an effect of less than one minute difference. Therefore, overall, the CI is inconclusive, but there seems to be at least some small effect.

(d) From the histogram, residuals appear to be close to normality. As there are just two points in the domain of the x variable, the effect must be linear. Further, looking at the scatter plot, the variance seems to be similar at both $x = 0$ and $x = 1$, so variance is constant throughout the range. Therefore, a linear model seems reasonable here, all assumptions seem to hold.

2. (a) For each increase of one year in age, income increases by $300, with 95% CI $150 to $450. Therefore, there seems to be a strong effect of age on income, as even the lower limit of the CI is a "financially important" increase.

(b) Note that there is a difference of 3 years in age, two years in education, and a change from female to male. So, we can calculate:

$$\text{difference} = 0.3 \times 3 - 1.2 \times 2 + 1.9 = 0.4$$

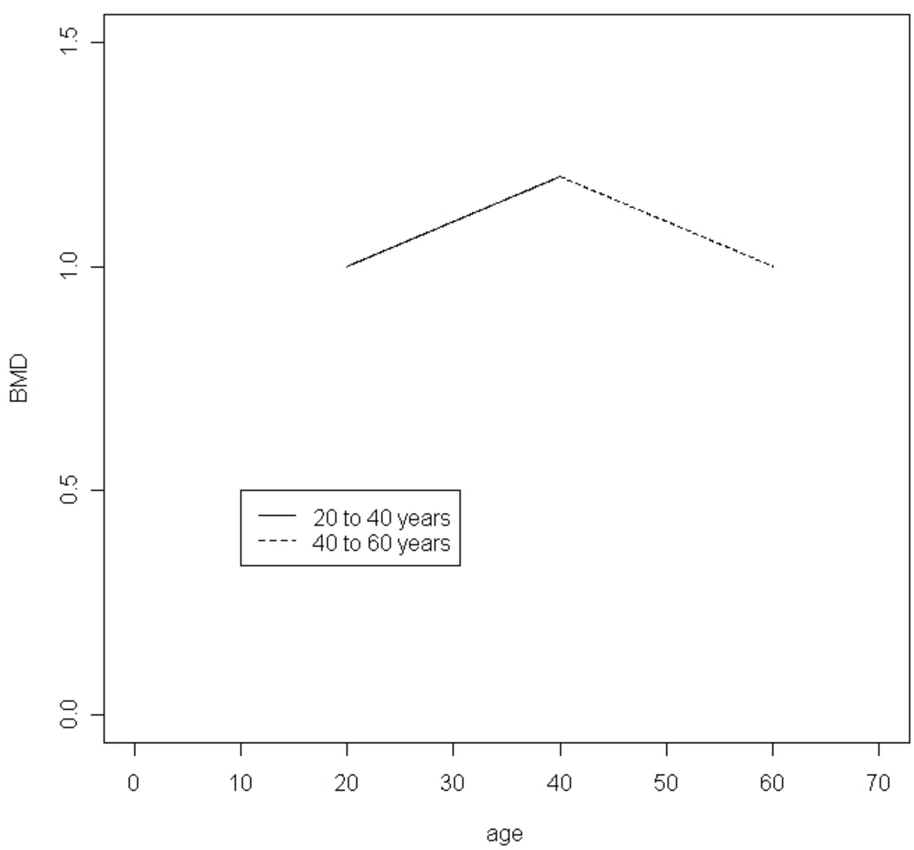
Therefore, the male would have a salary that is $400 higher than the females, on average, under these conditions.

(c) Evidence for confounding occurs when beta coefficients change as other variables enter or exit a model, and/or when the SE estimates become inflated when new variables enter the model. Clues about the relationships among independent and dependent variables can also be found in the correlation matrix.

In this case, the correlation matrix indicated relations between age and education, and both of these are also related to income, so there is potential for confounding between these two variables. Comparing the univariate to multivariate results, we see that both coefficients decrease in the multivariate model compared to the univariate for these two variables, and the confidence intervals are wider in the multivariate model, so indeed we find evidence of confounding between age and education.

The sex variable was much less related to age and education, and its coefficient and SE changed little in comparing univariate to multivariate models. Hence, sex does not appear to be strongly confounded with either age or education.

3. (a) Graph is:



(b) If a single line if fit to the data depicted in the graph from part (a), the slope would be close to zero, as that is the average change in BMD over the entire range from age 20 to 60.

(c) In part (a), we notice that there is a different slope between age and BMD depending on the age group, and within each age group, a straight line fits very well. One should therefore add an interaction term between age as a continuous variable and age group (dichotomous, 20 to 40, versus 40 to 60 years), which would then fit very well. This would fit much better than adding a quadratic term.

4. (a) We will use the beta distribution, since the range of our probability parameter here (probability of osteoarthritis) is 0 to 1. We would like to have a prior mean of $\mu = 0.15$, and a prior SD of $\sigma = 0.025$, so that 95% of our prior distribution covers the range from 10% to 20%, or 0.1 to 0.2 on the probability scale. Using the properties of the beta distribution, we calculate:

$$\alpha = -\frac{\mu\left(\sigma^2 + \mu^2 - \mu\right)}{\sigma^2} = 30.45$$

and

$$\beta = \frac{(\mu - 1)\left(\sigma^2 + \mu^2 - \mu\right)}{\sigma^2} = 172.58$$

Thus, our prior is beta(30.45, 172.58).

(b) We observe $x = 70$ cases and $n - x = 430$ non-cases of osteoarthritis, so our posterior distribution is:

$$beta(\alpha + x, n - x + \beta) = beta(30.45 + 70, 172.58 + 430) = beta(100.45, 602.58)$$

Thus, our posterior distribution is beta(30.45, 172.58).

(c) Using the posterior distribution calculated in (b) and the formulae for mean and SD of the beta distribution, we find:

$$\mu = \frac{\alpha}{\alpha + \beta} = 0.143$$

and

$$\sigma = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}} = 0.013$$

5. (a) For each increase of one unit in x, Y^* increases by 0.446 units.

(b) Recall the formula for the intercept from any simple linear model is:

$$\alpha = \overline{Y} + \beta \times \overline{x}$$

Now, since the variable is standardized, we know that $\overline{Y^*} = 0$, so we are left with, in this example, that

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

Now, β is non-zero, so we would "expect" α to be near zero only if we expected $\overline{x}$ to be near zero. So, this would be an expected result if we expected $\overline{x}$ to be near zero, and otherwise not. As no information was given about the distribution of x in this problem, the results would not particularly be expected from the information given.

(c) First, note that when $x = -1$, we have

$$Y^* = 0.01814 + 0.446 * (-1) = -0.42786$$

Next, from standardization, we have that

$$Y^* = \frac{Y - \overline{Y}}{s_y} = \frac{Y - 1}{5}$$

so by simple algebra, we can solve for Y, to get

$$Y = 5 \times Y^* + 1 = 5 \times (-0.42786) + 1 = -1.14$$