EPIB-621 Solutions 2019 Final Exam

1. From the information given, since the p-value was 0.05 exactly, we know that

$$t = 1.96 = \frac{\overline{x}}{s/\sqrt{n}}$$

Plugging in what we know so far, we have

$$t = 1.96 = \frac{5}{s/\sqrt{225}}$$

Solving for the only unknown parameter, s, we find

$$s = 5/1.96 \times \sqrt{225} = 38.26531$$

From the usual formula for a 95% confidence interval for a mean, we then calculate

$$5 \pm 1.96 \times 38.26531 \times \sqrt{225} = (0, 10)$$

Thus the 95% CI is (0, 10) [which we also could have just guessed from the information given, since the lower limit = 0 when p = 0.05 exactly].

2. (a) The intercept of -3.326355 represents, on the logit scale, the probability that a male aged 50 is experiencing depression, i.e., the prevalence of depression among 50 year old males. To find the probability we must convert this from the logit scale to the probability scale, using the inverse logit function. Thus we calculate:

$$P(\text{male aged 50 is depressed}) = \frac{exp(-3.326355)}{(1 + \exp(-3.326355))} = 0.03467804$$

Thus the intercept estimates a probability of 0.035 for a male aged 50.

(b) A similar calculation to part (a) is needed, but now adding in the coefficient for sex, since females are coded as 1, and the two coefficients for age and age squared, using age =5, since age is centered at 50. Thus we calculate:

$$\begin{aligned} & P(\text{female aged 55 is depressed}) \\ = & \frac{exp(-3.326355 + 0.629383 + 5 \times 0.050016 - 5^2 \times 0.017050)}{(1 + \exp(-3.326355 + 0.629383 + 5 \times 0.050016 - 5^2 \times 0.017050))} \\ = & 0.05349733 \end{aligned}$$

Thus we estimate the probability of depression for 55 year old females at 0.053.

(c) The coefficient for sex is 0.629383, with SE= 0.197050. Thus we calculate that OR = exp(0.629383) = 1.876452. For the confidence interval, we calculate:

$$\text{Lower Limit} = exp(0.629383 - 1.96 \times 0.197050) = 1.275279$$

and

$$\text{Upper Limit} = exp(0.629383 + 1.96 \times 0.197050) = 2.761022$$

Thus we have OR=1.88, with 95% CI (1.28, 2.76).

3. (a) The data from region 3 has 5 cases of RA in 500 subjects. Thus using these data alone, we have a point estimate of 5/500 = 0.01, with 95% CI calculated using the ususal formula for a binomial proportion, that is,

$$0.01 \pm 1.96 * \sqrt{0.01 \times 0.99/500} = (0.001278551, 0.01872145) \approx (0.0013, 0.019)$$

(b) Directly from the WinBUGS program, we see a point estimate of 0.01652, with 95% credible interval of (0.00751, 0.02579). The point estimate here, 0.0165, is higher than the point estimate from the data alone, which was 0.01. This is due to the effect of the hierarchical model, which will move the point estimate towards the average of the entire eight regions, which here is close to 0.023 (or 0.021, depending on whether you use the mean or median, both are correct). Thus the hierarchical model moved the initial estimate of 0.01 up towards the overall average of 0.023, resulting in an adjusted estimate for region 3 of 0.0165.

(c) The results for w were a point estimate of 0.02131 (or 0.023) with 95% credible interval of (0.00829, 0. 0.05048). This represents the prediction and 95% prediction interval for the prevalence in the "next similar region," or roughly the overall average and 95% interval over the eight regions, accounting for the between-region variability in prevalence rates.

4. (a) The outcome for the eighth data point, y[8], is missing, and so WinBUGS will automatically carry out multiple imputation to fill in this missing data item. Thus the results for y[8] represent a summary for the imputed missing y[8] data. Since the corresponding x value, x[8]= 34, it represents the prediction from the linear regression model of the outcome when age=34, together with its 95% prediction credible interval. Thus when age=34, we predict an outcome of about 79.3, with 95% interval 65.6 to 93.8.

(b) The data are identical to part (a), except that there is now a missing x or independent variable item. Unlike missing outcomes, where WinBUGS already has a model for prediction from the regression, there is no model in the program to predict (or impute) missing x values, so WinBUGS will stop with an error.

5. (a) The best prediction equation is taken from the EV (Expected Value) column, and therefore is

$$logit(p) = -3.898{+}0.048\mathrm{a}ge{+}0.973\mathrm{s}ex{+}0.669\mathrm{s}moke{+}0.418\mathrm{f}amhist{+}0.252\mathrm{p}revfrac$$

Plugging the values into this equation, we calculate

$$logit(p) = -3.898 + 0.048 \times 65 + 0.973 + 0.669 + 0.418 + 0.252 = 1.534$$

Taking the inverse logit of 1.534 we have $\exp(1.534)/(1{+}\exp(1.534)) = 0.8225908$. Thus this woman has a probablility of osteoporosis of about 82.3%.

(b) The overall rate from the summary statistics for osteo is 52%. The predicted probability of 82.3% is much higher than 52% because the woman in question has all of the risk factors, including being a woman, and previous history of fracture and family history of osteoporosis. Since all of these risk factors have positive coefficients in the model, such a person can be expected to have a higher probability compared to the average person in this population.