Name Jerry

Student Number

STA 302f 2015 Quiz 10

1. (6 points) In a study comparing the effectiveness of different exercise programmes, volunteers were randomly assigned to one of three exercise programmes (A, B, C) or put on a waiting list and told to work out on their own. Aerobic capacity is the body's ability to process oxygen. Aerobic capacity was measured before and after 6 months of participation in the program (or 6 months of being on the waiting list). The response variable was improvement in aerobic capacity. The independent variables were age (a covariate) and treatment group. *Note that the waiting list condition is one of the treatments.*

   Consider a regression model with an intercept.

   (a) Write the regression equation. Please use x for age, and make its regression coefficient β_1. You don't need to say how the dummy variables are defined. You will do that in the next part.

   $Y = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \beta_4 d_3 + \varepsilon$

   (b) Make a table with columns showing how you would set up indicator dummy variables for treatment condition. Waiting List should be the reference category. Make a wider column on the right in which you give $E(y|\mathbf{x})$ for each treatment condition.

	d_1	d_2	d_3	$E(Y\|x)$
A	1	0	0	$\beta_0 + \beta_2 + \beta_1 x$
B	0	1	0	$\beta_0 + \beta_3 + \beta_1 x$
C	0	0	1	$\beta_0 + \beta_4 + \beta_1 x$
Wait	0	0	0	$\beta_0 + \beta_1 x$

   (c) Suppose you wanted to know whether controlling for age, there is any difference among the treatment conditions in expected gain in aerobic capacity. In terms of β values, what null hypothesis would you test?

   $H_0: \beta_2 = \beta_3 = \beta_4 = 0$

   (d) Suppose you wanted to know whether, controlling for age, Exercise Programme A is better than the waiting list. In terms of β values, what null hypothesis would you test?

   $H_0: \beta_2 = 0$

   (e) Suppose you wanted to estimate the difference in average improvement between programmes A and C for a 27 year old participant. Give your answer in terms of $\widehat{\beta}$ values.

   $\hat{\beta}_2 - \hat{\beta}_4$

2. (4 points) In homework, you calculated a 95% prediction interval for the total amount of wood to be obtained from cutting down three particular trees. The final answer was a pair of numbers a lower prediction limit and an upper prediction limit. **Write the two numbers in the space below**. Circle them on your printout, and write "Question 2" beside the numbers.

(57.02, 86.14)

Attach your *complete* R printout to your quiz. Make sure your name and student number are written clearly on the printout.

This is based on Homework Question 2

(a) $W \sim N\left(\sum_{j=n+1}^{n+m} x_j'\beta,\ m\sigma^2\right)$

(b) $\hat{W} = \sum_{j=n+1}^{n+m} x_j'\hat{\beta} \sim N\left(\sum x_j'\beta,\ \left(\sum_{j=n+1}^{n+m} x_j\right)'\sigma^2(X'X)^{-1}\left(\sum_{j=n+1}^{n+m} x_j\right)\right)$ call this a'

(c) $W - \hat{W} \sim N(0,\ m\sigma^2 + \sigma^2 a'(X'X)^{-1}a)$

(d) $Z = \dfrac{W-\hat{W}}{\sqrt{\sigma^2(m + a'(X'X)^{-1}a}}$

(e) $T = \dfrac{W-\hat{W}}{\sqrt{\sigma^2(m+a'(X'X)^{-1}a}} \Big/ \sqrt{\dfrac{SSE}{\sigma^2}/(n-k-1)}$

$= \dfrac{W-\hat{W}}{\sqrt{MSE(m+a'(X'X)^{-1}a}} \sim t(n-k-1)$

(g) $1-\alpha = P_r\left\{\hat{W} - t_{\alpha/2}\sqrt{MSE(m+a'(X'X)^{-1}a} < W < \hat{W} + t_{\alpha/2}\sqrt{MSE(m+a'(X'X)^{-1}a}\right\}$

(f) W is based on a new set of independent data $\&$ so is independent of MSE based on $y_1, \ldots, y_n$.
$\hat{W}$ is a function of $\hat{\beta}$, which is independent of MSE