

Homework Assignment 3

(Due on 5/30 9:45am in class)

Problem 1. The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year (Y) can be predicted from the ACT test scores (x). The data for this study can be found on Canvas, you might find the function `read.table("C: file address here GPAdata.txt")` useful in reading the data into R. Note, in the analysis that follows you are expected to write all of your own functions for conducting the analysis, and provide them in your write up, of course checking your solutions with the built in functions is a good idea.

- (a) Obtain the least squares estimates of β_0 and β_1 , and state the estimated regression function. Be sure to interpret the meaning of these estimates in the context of this problem.
- (b) Plot the estimated regression function, along with the data. Does the estimated regression function fit the data well. In addition provide a 95% confidence band around your estimated regression function.
- (c) Obtain a point estimate of the mean freshman GPA for students with ACT test scores $x = 30$, also provide a 95% CI and 95% PI for this estimate, with their appropriate interpretation.
- (d) Obtain a 99% CI for both least squares estimates, and interpret.
- (e) Conduct hypothesis tests about both of the least squares estimates, using a t-statistic at the $\alpha = 0.01$ significance level, be sure to report a p-value. Interpret your results and comment on how these relate to the CIs you constructed in part (d).
- (f) Set up the ANOVA table.
- (g) Conduct an F-test, at the $\alpha = 0.01$ significance level, of whether or not $\beta_1 = 0$. Be sure to state the null and alternative hypothesis, the decision rule, and interpret your results. Comment on how this relates to one of the tests you performed in part (e).
- (h) Obtain R^2 and r , and interpret.
- (i) Prepare a box plot for the ACT scores. Are there any noteworthy features?

- (j) Plot the residuals against the fitted values, what does this tell you.
- (k) Prepare a QQ-plot of the residuals, and discuss appropriately.
- (l) Conduct the Brown Forsythe test to determine whether the assumption of constant error variance is appropriate, use $X = 26$ to split the data into two separate classes.

Problem 2. A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 15 women from each 10 year age group. Treating age as a predictor variable and muscle mass as your response, complete the following analysis. The data for this study can be found on Canvas. Again, you should use the function that you have written in this analysis.

- (a) Obtain the least squares estimates of β_0 and β_1 , and state the estimated regression function. Be sure to interpret the meaning of these estimates in the context of this problem.
- (b) Plot the estimated regression function, along with the data. Does the estimated regression function to fit the data well. In addition provide a 95% confidence band around your estimated regression function.
- (c) Obtain a point estimate of the mean muscle mass for women age $x = 55$, also provide a 95% CI and 95% PI for this estimate, with their appropriate interpretation.
- (d) Obtain a 99% CI for both least squares estimates, an interpret.
- (e) Conduct hypothesis tests about both of the least squares estimates, using a t-statistic at the $\alpha = 0.01$ significance level, be sure to report a p-value. Interpret your results and comment on how these relate to the CIs you constructed in part (d).
- (f) Set up the ANOVA table.
- (g) Conduct an F-test, at the $\alpha = 0.01$ significance level, of whether or not $\beta_1 = 0$. Be sure to state the null and alternative hypothesis, the decision rule, and interpret your results. Comment on how this relates to one of the tests you performed in part (e).
- (h) Obtain R^2 and r , and interpret.
- (i) Prepare a box plot for ages. Are there any noteworthy features?
- (f) Plot the residuals against the fitted values, what does this tell you.
- (g) Prepare a QQ-plot of the residuals, and discuss appropriately.
- (l) Conduct the Brown Forsythe test to determine whether the assumption of constant error variance is appropriate, use $X = 60$ to split the data into two separate classes.