Chapter 2 Questions 4, 13, and 23

Written Out Answers

\*\*I realized afterwards that I didn’t have to do some of these problems. I left them here anyways.

4) a) The 99% confidence interval for β1 is 0.005385614 to 0.07226864 GPA unit per ACT unit. This confidence interval does not include 0, which provides evidence to the director of admissions that there exists a relationship between ACT score and GPA after freshman year of college.

b) The two alternatives for the test are *H0*: β1 = 0 and *Ha*: β1 ≠ 0. The level of significance to be used is 0.01. Since the confidence interval for β1 from part (a) does not include 0, the conclusion for *Ha*. From the summary table in RStudio, the t\* value is 3.039777. The decision rule states that if |t\*| ≤ *t*(1 – α/2; n-2), conclude *H0* and if |t\*| > *t*(1 – α/2; n-2), conclude *Ha*. Here we have t(.995, 118) = 2.618137. Since 3.039777 > 2.618137, we can conclude *Ha*.

c) The P-value of the test is 0.00292 from the summary table, which is less than 0.01 (α), it supports concluding *Ha*.

13) a) The 95% confidence interval estimate of the mean freshman GPA for students whose ACT test score is 28 is 3.061384 to 3.341033. This means that with 95% confidence, we can say that a student with an ACT score of 28, will achieve a freshman GPA between 3.061 and 3.341. The Ŷ estimate is 3.201209.

b) The 95% prediction interval estimate of Mary’s GPA is between 1.959355 and 4.443063. The fit value remains 3.201209.

c) Yes, the prediction interval is wider than the confidence interval and it should be because the confidence interval is based on the mean response and the prediction interval is based on a new observation, because means are less variable than single observations.

d) The boundary values of the 95% confidence band for the regression line when X = 28 are computed after computing W2. W2 = 6.146181 and W = 2.479. From above, Ŷh = 3.20109, and s = 0.62317. Therefore, the confidence band for the regression line when X = 28 is 3.20109 – 0.62317(2.479) and 3.20109 + 0.62317(2.479) which is between 0.72209 and 5.68009. This interval is wider than the confidence interval in part (a), and it should be because the confidence band must include the whole regression line.

23) a) Nova table is set up. See knit section.

b) MSR is an estimate of 𝜎2 + 𝛽12 ∑(𝑋*i* − 𝑋̅)2 , which is an estimation of the variance plus the square of *β1* times the sum of the squared residuals. In this case the MSR = 0.3883. The MSE estimates 𝜎2 (the variance). In this case, the MSE = 0.3883. If *β1* = 0, then the second term of MSR is 0 and both the MSR and MSE estimate variance (𝜎2).

c) The F\* value is 9.2402 from the ANOVA table. The F test then follows with the two alternatives being *H0*: β1 = 0 and *Ha*: β1 ≠ 0. The decision rule states that if F\* ≤ F(1 – α; *dfR – dfF , dfF),* conclude *H0 ,* and if F\* > F(1 – α; *dfR – dfF , dfF),* conclude *Ha* . In RStudio, F(1 – α; *dfR – dfF , dfF),* = qf(0.99, 1, 118) = 6.854641. Therefore, F\* > F(1 – α; *dfR – dfF , dfF),* so the *H0* is rejected and *H1* is accepted.

d) The absolute magnitude of the reduction in the variation of Y when X is introduced into the regression model is called the SSE (error sum of squares). The SSE is 45.818 from the ANOVA table. The relative reduction is found by computing (SSTO – SSE = SSR) and then SSR/SSTO (or 1 – SSE/SSTO) and this value gives you R2 (*coefficient of determination)*. (1 – 45.818/ 49.40545) = 1 – 0.9274 = 0.0726 = R2.

e) r = 0.2694 found by taking the square root of R2 from above (d). This r value is confirmed by the cor(GPA, ACT) command in RStudio. The sign is positive (it is single regression, and the slope is positive).

f) r (*coefficient of correlation*) gives a clearer cut operational interpretation because it gives a clearer representation of the relationship between *X* and *Y*, however both r2 and r have limitations.