Quiz Practice Questions 5 (Attendance 10) for Statistics 512
Applied Regression Analysis
Material Covered: Chapter 11 Neter et al. and Kuhn

These are practice questions for the quiz. The quiz (not the practice questions) is worth 5% and marked out of 5 points. One or more questions is closely, but not necessarily exactly, related to one or more of these questions will appear on the quiz. These practice questions are *not* to be handed in. Quizzes are to be done *using Vista* on the Internet **before** 4am (West Lafayette time!) of the date of the quiz. Vista will *not* allow any quiz to be done late. It is *highly* recommended that you complete this practice quiz, by hand, *before* logging onto Vista. The quiz is an **individual** one which means that each student does this quiz by themselves without help from others.

Applied Linear Statistical Models
(Neter et al.) Questions.

Chapter	Problem(s)	hints
11, pages 490–496	11.4, 11.8	Calculator Maintenance
	11.14	Assessed Valuations

(11.4) Calculator Maintenance: qz5-11-4-calc-qualitative

(a) and (b) *Regression For Qualitative Variables.*
The regression without interaction is given by

$$\begin{aligned}
\hat{Y} &= b_0 + b_1X_1 + b_2X_2 \\
&= -2.3475 + 14.7234X_1 + 0.2766X_2 \\
&= -2.3475 + 14.7234X_1, \quad \text{if } X_2 = 0 \text{ (commerical)} \\
&= -2.0709 + 14.7234X_1, \quad \text{if } X_2 = 1 \text{ (student)}
\end{aligned}$$

where
b_0 average time of service call when no (0) calculators are serviced
X_1 is the number of calculators serviced
(since $b_1 = 14.7$, time increases 14.7 minutes per calculator serviced)
X_2 is calculator model, specifically,
either commerical calculators (when $X_2 = 0$)
or student calculators (when $X_2 = 1$).
It appears more service time is devoted to student calculators, rather than commercial calculators because both are associated with *parallel* regression lines, where the regression line associated with the student calculators crosses at a higher y–intercept.

(c) *(Bonferroni) Confidence Intervals of β_2*
$s\{b_2\} = 2.37811$, $t(0.975, 15) = ?$
(PRGM INVT ENTER ENTER 15 ENTER 0.975)
then 95% CI is $0.2766\pm\ ?(2.37811) = (-4.791, 5.344)$
since this includes zero (0), calculator model is not influencing the average service time, or, in other words, there is no significant difference between student and commercial service time.

(d) *Why include X_1?*
The number of calculators is included in the regression because this, as well as calculator model, influences service time.

(e) *Residuals Versus Interaction, X_1X_2*
Since the spread in the residuals for X_1X_2 ($X_1 = 0$) is the same as the spread in the residuals for X_1X_2 ($X_1 = 1$), this indicates there is no interaction.

(11.8) Calculator Maintenance: qz5-11-8-calc-interact

(a) *Regression With Interaction.*
$\hat{Y} = -1.56 + 14.54X_1 - 3.17X_2 + \,?X_1X_2$

(b) *Test if interaction term can be dropped.*

Source	Sum Of Squares	Degrees of Freedom	Mean Squares
Regression	16190	3	5396.57
X_1	16183	1	16183
$X_2 \mid X_1$	0.28959	1	0.28959
$X_1X_2 \mid X_1, X_2$	6.8299	1	6.8299
Error	314.276	14	22.44831
Total	16504	17	

1. *Statement.*
   $H_0 : \beta_{12} = 0$ versus $H_0 : \beta_{12} \neq 0$
   where we assume
   $$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon$$
2. *Test.*
   The partial F^* test statistic is
   $$\begin{aligned} F^* &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \\ &= \frac{SSE(X_1X_2) - SSE(X_1, X_2, X_1X_2)}{(n-3)-(n-4)} \div \frac{SSE(X_1, X_2, X_1X_2)}{n-4} \\ &= \frac{SSR(X_1X_2 \mid X_1, X_2)}{1} \div \frac{SSE(X_1, X_2, X_1X_2)}{n-4} \\ &= \frac{?}{1} \div \frac{?}{14} \;=\; 0.30 \end{aligned}$$
   The upper critical value at $\alpha = 0.10$,
   with $(1, 14)$ degrees of freedom is 3.10
   (Use PRGM INVF ENTER 1 ENTER 14 ENTER 0.90 ENTER)
3. *Conclusion.*
   Since the test statistic, 0.30, is smaller than the critical value, 3.10, we **accept** the null hypothesis that the interaction term is zero, $\beta_{12} = 0$.

(11.14) Assessed Valuation: qz5-11-14-value-qualitative

(a) *Scatterplot of Qualitative Regressions.*
Slopes of regression lines in attached graphs appear to be a little bit different

(b) *Regression and Test $H_0 : \beta_2 = \beta_{12} = 0$.*
The regression with interaction is
$\hat{Y} = 3.6513 +\ ?X_1 - 8.9100X_2 + 1.1335X_1X_2$
where $X_2 = 1$ for corner lots, $X_2 = 0$ for non–corner lots

Source	Sum Of Squares	Degrees of Freedom	Mean Squares
Regression	2203.17559	3	734.39186
X_1	1833.32787	1	1833.32787
$X_2 \mid X_1$	318.89195	1	318.89195
$X_1X_2 \mid X_1, X_2$	50.95577	1	50.95577
Error	76.06441	19	4.00339
Total	2279.24	22	

1. *Statement.*
   $H_0 : \beta_2 = \beta_{12} = 0$ versus H_0 : at least one $\beta_i \neq 0$
2. *Test.*
   The partial F^* test statistic is

$$
\begin{aligned}
F^* &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \\
&= \frac{SSE(X_2, X_1X_2) - SSE(X_1, X_2, X_1X_2)}{(n-2)-(n-4)} \div \frac{SSE(X_1, X_2, X_1X_2)}{n-4} \\
&= \frac{SSR(X_2, X_1X_2 \mid X_1, X_2)}{2} \div \frac{SSE(X_1, X_2, X_1X_2)}{n-4} \\
&= \frac{SSR(X_2 \mid X_1) + SSR(X_1X_2 \mid X_1, X_2)}{2} \div \frac{SSE(X_1, X_2, X_1X_2)}{n-4} \\
&= \frac{? + 50.95577}{2} \div \frac{76.06441}{19} = ?
\end{aligned}
$$

   The upper critical value at $\alpha = 0.10$,
   with $(2, 19)$ degrees of freedom is 2.61
   (Use PRGM INVF ENTER 2 ENTER 19 ENTER 0.90 ENTER)
3. *Conclusion.*
   Since the test statistic, ?, is larger than the critical value, 2.10, we **reject** the null hypothesis that $\beta_2 = \beta_{12} = 0$.

(c) *Regressions For Homes On Corners and Not On Corners.*

$$\begin{aligned}\hat{Y} &= -5.2587 + 3.6265X_1, \quad \text{if } X_2 = 1 \text{ (non–corner)} \\ &= 3.6513 + 2.4930X_1, \quad \text{if } X_2 = 0 \text{ (corner)}\end{aligned}$$

The two regressions are fairly different.
Both the y–intercepts and slopes are different[1].

(d) *CI For β_{12}.*
$s\{b_{12}\} = 0.31771$, $t(0.95, 19) = ?$
(Use PRGM INVT ENTER ENTER 19 ENTER 0.95 ENTER)
then 95% CI is $1.3347 \pm (?)0.31771 = (0.504, 1.693)$
since this interaction between different lots *ex*cludes zero (0),
this indicates a significant *difference* between non–corner and corner lots

(e) *Residual Plots For Homes On Corners and Not On Corners.*
The residual plots for both seem to indicate one outlier in the "on corners" case.

(f) *Test Equal variance For Homes On Corners and Not On Corners.*

1. *Statement.*
   $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_0 : \sigma_1^2 \neq \sigma_2^2$
   where 1 is homes on corners and 2 is homes not on corner
2. *Test.*
   The test statistic is

$$\begin{aligned}F &= \frac{SSE_1}{7} \div \frac{SSE_2}{12} \\ &= \frac{53.6822}{7} \div \frac{?}{12} \\ &= 4.11\end{aligned}$$

   The upper and lower critical values at $\alpha = 0.10$ are
   $F(0.025; 7, 12) = 0.21$, $F(0.975; 7, 12) = ?$
   (Use PRGM INVF ENTER 7 ENTER 12 ENTER 0.025 ENTER, for lower critical value, for example.)
3. *Conclusion.*
   Since the test statistic, 4.11, is outside (0.21, ?), we reject the null hypothesis that $\sigma_1^2 = \sigma_2^2$; that is, the data indicates the variances are *different*.

[1] By the way, these two regressions are *estimated* by splitting the data into two groups and *not* by setting $X_2 = 1$ or $X_2 = 0$ for the regression

$$\hat{Y} = 3.6513 + 2.4930X_1 - 8.9100X_2 + 1.1335X_1X_2$$

(g) *Residual and Normal Probability Plots.*
Attached residual plots (even the ones with a vertical "line") seem to have constant variance. Also, the normal probability plot is linear and so seems to indicate normality.