Quiz Practice Questions 1 (Attendance 2) for Statistics 512
Applied Regression Analysis
Material Covered: Chapters 1,2 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 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
1, pages 36–43	1.6	
	1.19,1.23	Grade point average data
	1.22,1.26	Plastic hardness data
2, pages 86–94	2.7, 2.16, 2.26	Plastic hardness data

(1.6) normal regression model

(a) *Understanding the assumptions in a normal regression model*
Use your calculators to sketch $E\{Y\} = 200 + 5X$ (use "Y=", WINDOW and GRAPH). Draw the positive sloped regression line you get from the calculator on a piece of paper. To the right of three vertical lines at $X = 10, 20, 30$, draw three "sideways" normal–shaped curves. Each "sideways" bell–shaped curves is centered where the vertical lines intersect the regression line. The width of each normal curve must have a width given by the standard deviation $\sigma = 4$.

(b) *Understanding the parameters in a normal regression model*
True / False
Parameter β_0 is the mean response at $X = 0$,
parameter β_1 is the change in the mean response for a unit change in the independent variable, X.

(1.19) Grade point average data: qz1-1-19,23-gpa-regress

(a) *Estimated linear regression*

$$b_0 = -1.69956, \qquad b_1 = ?, \qquad \hat{Y} = ? + ?X$$

(b) *Fit of linear regression to data*
An estimated regression line is a "good" fit to the data if the data clusters fairly closely and randomly around this line. From the SAS output, this (choose one) **does** / **does not** appear to be the case.

(c) *Linear regression and prediction*
Let $X = 5.0$ and so

$$\hat{Y}_h = -1.69956 + \ ?X = -1.69956 + \ ?(5.0) = \ ?$$

(d) *Linear regression and slope*
$b_1 = ?$

(1.22) Plastic hardness data: qz1-1-22,26-plastic

(a) *Estimated regression*
From SAS output,

$$\hat{Y} = 168.6 + \ ?X$$

(b) *Estimated regression and prediction*
At $X = 40$,

$$\hat{Y}_h = 168.6 + ?(40) = 249.975$$

(c) *Estimated slope*
$b_1 = \ ?$

(1.23) Grade point average data: qz1-1-19,23-gpa-regress

(a) *Residuals*
True / **False** The residuals sum to zero.

(b) *Error mean square*
The error mean square or residual mean square, MSE, is is given on the SAS output.

$$E\{MSE\} = \sigma^2 = ?, \qquad \sigma = \sqrt{?} = ?$$

The MSE is one rough way of measuring how well the linear regression fits the data: a smaller MSE indicates a good fit.

(1.26) Plastic hardness data: qz1-1-22,26-plastic

(a) *Residuals*
True / **False** The residuals sum to zero.

(b) *Error mean square*
The error mean square or residual mean square, MSE, is is given on the SAS output.

$$E\{MSE\} = \sigma^2 = ?, \qquad \sigma = \sqrt{?} = ?$$

The last estimate is expressed in Brinell units.

(2.7) Plastic hardness data, inference on β_1: qz1-2-7,16,26

(a) *Confidence*[1] *interval of slope,* β_1
From the SAS output, $b_1 = ?$ and $s\{b_1\} = 0.0904$
Also[2], $t(1-\alpha/2; n-2) = t(1-0.01/2; 16-2) = ?$

$$b_1 \pm t(1-\alpha/2; n-2)s\{b_1\} = 2.034 \pm (2.977)(0.09) = (1.77, 2.3)$$

(b) *Test*[3] *slope,* $\beta_1 = 2$
From the SAS output, $b_1 = ?$ and $s\{b_1\} = 0.0904$ and so

$H_0 : \beta_1 = 2$ versus $H_a : \beta_1 \neq 2$
$t^* = \frac{b_1 - \beta_{10}}{s\{b_1\}} = \frac{2.034-2}{0.09} = ?$, $t(1-\alpha/2; n-2) = 2.977$
since $t^* = ? < t = 2.977$, then data supports $\beta_1 = 2$
p–value[4] is 0.71

(c) *Power of the test*
If $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$
use table B.5 to determine the power of the test when $\beta_1 = 1.5$,
and $\delta = 0.3$ and $\sigma\{b_1\} = 0.1$
since $\delta = \frac{0.3}{0.1} = 3$, from table B.5, power is 0.50

---

[1] It is possible to use your calculator here.
Instead of SAS, it is possible to use PRGM REGINF, where {B0,B1,MN,PI} is {0,1,0,0}; NULL B1 can be whatever you want (let's make it 1) since we want a CI and not do a test; ALPHA is 0.01 (since we want a 99 percent confidence interval)

[2] Use PRGM INVT

[3] Use PRGM REGINF again, only let NULL B1 = 2

[4] SAS uses a F, rather than t, test statistic, where, notice $F^2 = 0.14 = t^2 = 0.38^2$; in both cases, though, the p–value is 0.71.

(2.16) Plastic hardness data, mean response $E\{Y_i\}$: qz1-2-7,16,26
The TI–83 calculator is better suited to deal with these questions than SAS.

(a) Use PRGM REGINF, where {B0,B1,MN,PI} is {0,0,1,0};
NULL MN can be whatever you want (let's make it 1) since we want a confidence interval and not do a test;
X is 30, the elapsed time, X_h
M can be whatever you want (let's make it 1) since we are not calculating a prediction interval;
ALPHA is 0.02 (since we want a 98 percent confidence interval) to determine

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{\hat{Y}_h\} = 229.631 \pm (2.624)(0.83) = (227.46, 231.81)$$

(b) Use PRGM REGINF, where {B0,B1,MN,PI} is {0,0,1,1};
NULL MN can be whatever you want (let's make it 1) since we want a confidence interval and not do a test;
X is 30, the elapsed time, X_h
M is one (1) since we want the prediction interval for one new observation;
ALPHA is 0.02 (since we want a 98 percent prediction interval) to determine

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{predmean\} = 229.631 \pm (2.624)(3.34) = (220.87, 238.39)$$

(c) Use PRGM REGINF, where {B0,B1,MN,PI} is {0,0,1,1};
NULL MN can be whatever you want (let's make it 1) since we want a confidence interval and not do a test;
X is 30, the elapsed time, X_h
M is ten (10) since we want the prediction interval for the mean of ten new observations;
ALPHA is 0.02 (since we want a 98 percent prediction interval) to determine

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{predmean\} = 229.631 \pm (2.624)(1.32) = (226.18, 233.09)$$

(d) Yes, the prediction interval in part (c) is narrower than the prediction interval in part (d). Yes, it should be narrower since we are determining the *mean* of ten observations, which are less variable than one observation.

(e) Use PRGM REGSCIPI,
X is 30, the elapsed time, X_h
M can be whatever you want (let's make it 1) since we are not calculating a prediction interval;
K can be whatever you want (let's make it 1) since we are not calculating

k simultaneous confidence intervals;
ALPHA is 0.02 (since we want a 98 percent prediction interval) to determine

$$\hat{Y}_h \pm W s\{\hat{Y}_h\} = (226.95, 232.31)$$

where we are using the second row of information: the WH part of the SIM CI MN A = 0.02 confidence interval

(2.26) Plastic hardness data; ANOVA and r^2: qz1-2-7,16,26

(a) *ANOVA of regression*
From the SAS output or REGANOVA from the calculator,

Source	Sum Of Squares	Degrees of Freedom	Mean Squares
Regression	5,297.5125	1	5,297.5125
Error	146.4250	14	10.4589
Total	5,443.9375	15	

(b) *significance test*
$H_0 : \beta_1 = 1$ versus $H_a : \beta_1 \neq 1$
$F^* = \frac{5{,}297.5125}{10.4589} = 506.51$, $F(0.99; 1, 14) = 8.86$ (use PRGM INVF or the SAS output)
since $F^* = 506.51 > F = 8.86$, then data rejects $\beta_1 = 0$ (there is a linear association)

(c) Use 2nd STATPLOT to plot both
$Y_i - \hat{Y}_i$ versus X_i (2nd STATPLOT 1:ENTER, choose scatter plot, L_1, L_4, "square" marks, then ZOOM ZoomStat)
and $\hat{Y}_i - \overline{Y}$ versus X_i (Define $L_5 = L_3 - 225.5625$, then 2nd STATPLOT 2:ENTER, scatter plot, L_1, L_5, "+" marks, then ZOOM ZoomStat)
Clearly, the *SSE* ($Y_i - \hat{Y}_i$) are smaller than the *SSR* ($\hat{Y}_i - \overline{Y}$) since the "square" marks are less spread out than the "+" marks
this implies that a large part of the variation in the observed data can be accounted for by the linear regression (associated with *SSR*)

(d) r^2

$$r^2 = \frac{5,297.5125}{5,443.9375} = 0.9731; \quad r = 0.9865$$