Quiz Practice Questions 2 (Attendance 4) for Statistics 512
Applied Regression Analysis
Material Covered: Chapters 4,5 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.

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

Chapter	Problem(s)	hints
4, pages 171–174	4.5, 4.9, 4.18	Plastic hardness
	4.14, 4.15	Grade point average
5, pages 211–214	5.1, 5.14, 5.17	various
	5.7, 5.26	Plastic hardness

(4.5) Plastic Hardness[1], family CIs for β_0, β_1, qz2-4-5-plastic-k-family-par-ci

(a) *Family Bonferroni Regression Parameters 90% CI*
From TI–83 (INVT 14 ENTER 0.975 ENTER)
$B = t(1 - \alpha/2g; n - 2) = t(1 - 0.10/2(2); 16 - 2) = t(0.975; 14) = 2.145$
From SAS,

1. Bonferroni CI for β_0:
   $b_0 = 168.6$ and $s\{b_0\} = 0.6633$,
   $b_0 \pm Bs\{b_0\} = 168.6 \pm 2.145(2.657) = (162.901, 174.299)$
2. Bonferroni CI for β_1:
   $b_1 = ?$ and $s\{b_1\} = ?$
   $b_1 \pm Bs\{b_1\} = ? \pm ?(?) = (?, ?)$

(b) Since the covariance[2] between intercept and slope is

$$\sigma\{b_0, b_1\} = -\overline{X}\sigma^2\{b_1\}$$

and, in this case, $\overline{X} = 28$, b_0 and b_1 are negatively correlated.
Also, the fact that the two parameters are negatively correlated is *not* reflected in the 2–family Bonferroni confidence interval calculated above.

(c) In 90 percent of the samples, the two CIs capture (include) both parameters, whereas in 10 percent of the samples, either one or both of the CIs do not capture (or include) their associated parameters.

[1] Both SAS and the TI–83 calculator need to be used *together* here; that is, SAS does *not* give you the complete answer.

[2] (4.5), page 155, Neter et al.

(4.9) Plastic Hardness, family CIs for expected responses,
qz2-4-9-plastic-k-family-exp-ci

(a) *Family Bonferroni Responses 90% CI*
From TI–83
$B = t(1 - \alpha/2g; n - 2) = t(1 - 0.10/2(3); 16 - 2) = t(0.9833; 14) = ?$
(INVT 14 ENTER 0.9833 ENTER)
From SAS,

1. $X_h = 20$:
   $\hat{Y}_h = 209.2875$ and $s\{\hat{Y}_h\} = 1.0847$
   $\hat{Y}_h \pm Bs\{\hat{Y}_h\} = 209.2875 \pm 2.360(1.0847) = (206.727, 211.847)$
2. $X_h = 30$:
   $\hat{Y}_h = ?$ and $s\{\hat{Y}_h\} = ?$
   $\hat{Y}_h \pm Bs\{\hat{Y}_h\} = ? \pm ?(?) = (?, ?)$
3. $X_h = 40$:
   $\hat{Y}_h = 249.9750$ and $s\{\hat{Y}_h\} = 1.3529$
   $\hat{Y}_h \pm Bs\{\hat{Y}_h\} = 249.9750 \pm 2.360(1.3529) = (246.782, 253.166)$

In 90 percent of the samples, the three CIs capture (include) all expected response parameters, whereas in 10 percent of the samples, either one or more of the CIs do not capture (or include) their associated parameters.

(b) *Family Bonferroni vs Working–Hotelling Responses 90% CI*
For Bonferroni, $B = 2.360$
For Working–Hotelling,
$W = \sqrt{2F(1 - \alpha; 2, n - 2)} = \sqrt{2F(1 - 0.10; 2, 16 - 2)} = ?$
(INVF 2 ENTER 14 ENTER 0.90 ENTER, then multiply by 2
and find the square root of the result)
So, since $B = 3.360 > W = ?$, the family Bonferroni gives wider (*less* efficient) CIs than the family Working–Hotelling 90% CIs.

(c) *Family Bonferroni vs Working–Hotelling Responses 90% CI*
From TI–83
$W = ?$
(same as in (b) above)
$B = t(1 - \alpha/2g; n - 2) = t(1 - 0.10/2(2); 16 - 2) = t(0.975; 14) = ?$
(INVT 14 ENTER 0.975 ENTER)
Since $W = 2.340 > B = ?$, use B; notice, in this case, that *two* X_h, rather than *three* X_h are used and this is why B changes.
From SAS,

1. $X_h = 30$:
   $\hat{Y}_h = 229.6312$ and $s\{\hat{Y}_h\} = 3.3385$
   $\hat{Y}_h \pm Bs\{\hat{Y}_h\} = 229.6312 \pm ?(3.3385) = (222.470, 236.792)$
2. $X_h = 40$:
   $\hat{Y}_h = 249.9750$ and $s\{\hat{Y}_h\} = 3.5056$
   $\hat{Y}_h \pm Bs\{\hat{Y}_h\} = 249.9750 \pm ?(3.5056) = (242.455, 257.495)$

(4.14) Grade point average; origin model: qz2-4-14,15-gpa-origin

(a) The regression through origin is

$$\hat{Y} = 0.5061X$$

(b) From the TI–83, $t(1-\alpha/2; n-1) = t(1-0.05/2; 20-1) = t = ?$ and so the 95% confidence interval of β_1 is

$$b_1 \pm ts\{b_1\} = 0.5061 \pm ?(0.0214) = (0.461, 0.551)$$

(c) At $X_h = 4.7$,

$$\hat{Y}_h \pm ts\{\hat{Y}_h\} = 2.37862 \pm ?(0.10071) = (2.17, 2.59)$$

(4.15) Grade point average; origin model: qz2-4-14,15-gpa-origin

(a) From the SAS output, the regression through the origin appears to be tilted too "low"; the actual intercept seems to be below the origin.

(b) The residuals are given on the SAS output. The residual plot appears to show that the regression is tilted too "low"; the actual intercept seems to be below the origin.

(c) Lack of fit test.

1. *Statement.*
   $H_0 : \mu_j = \beta_1 X_j \quad \text{versus} \quad H_1 : \ \mu_j \neq \beta_1 X_j.$
2. *Test.*
   The test statistic is

$$\begin{aligned} F^* &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \\ &= \frac{4.44112 - 1.3916667}{19 - 4} \div \frac{1.3917}{4} \\ &= \ ? \end{aligned}$$

   where $SSE(R) = 4.44112$ and $df_R = 19$ appears in 4.14(C) SS error and $SSE(F) = 1.3916667$, $df_F = 4$ appears in 4.15(C) SS error on the SAS output.
   The critical value is $F(0.995; 15, 4) = 20.4$
   (Use PRGM INVF ENTER 15 ENTER 4 ENTER 0.995 ENTER)
3. *Conclusion.*
   Since the test statistic, ?, is smaller than the critical value, 20.4, we **accept** the null hypothesis that the regression function is linear, $\mu_j = \beta_1 X_j$.

The p–value is 0.80.

(4.18) Plastic Hardness

The ordinary least square estimates will *not* be biased.
This is an example of the *Berkson* Model.

$$\delta_i = X_i^* - X_i$$

where X_i^* is the prerecorded (fixed) elapsed time,
and X_i is the unobserved (random variable) true time.

(5.1) Matrix manipulation

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \\ 2 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 3 & 8 & 1 \\ 5 & 4 & 0 \end{bmatrix}$$

From the TI–83 calculator,

(1) $\mathbf{A} + \mathbf{B} =$

$$\begin{bmatrix} 2 & 7 \\ 3 & 10 \\ 5 & 13 \end{bmatrix}$$

(2) $\mathbf{A} - \mathbf{B} =$

$$\begin{bmatrix} 0 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

(3) $\mathbf{AC} =$

$$\begin{bmatrix} 23 & 24 & 1 \\ 36 & ? & 2 \\ 49 & 56 & 3 \end{bmatrix}$$

(4) $\mathbf{AB}' =$

$$\begin{bmatrix} 13 & 17 & ? \\ 20 & 26 & 34 \\ 27 & 35 & 46 \end{bmatrix}$$

(5) $\mathbf{B}'\mathbf{A} =$

$$\begin{bmatrix} ? & 26 \\ 26 & 76 \end{bmatrix}$$

(5.7) Plastic Hardness, various matrices, qz2-5-7-plastic-matrices
From the SAS output[3]

(1) $\mathbf{Y}'\mathbf{Y} = 819,499$

(2) $\mathbf{X}'\mathbf{X} =$

$$\begin{bmatrix} 16 & 448 \\ ? & 13,824 \end{bmatrix}$$

(3) $\mathbf{X}'\mathbf{Y} =$

$$\begin{bmatrix} 3,609 \\ ? \end{bmatrix}$$

[3]The XPX option gives

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Y} \\ \mathbf{Y}'\mathbf{X} & \mathbf{Y}'\mathbf{Y} \end{bmatrix}$$

(5.14) simultaneous equations

(a) If

$$\begin{array}{rcrcr} 4y_1 & + & 7y_2 & = & 25 \\ 2y_1 & + & 3y_2 & = & 12 \end{array}$$

then $\boldsymbol{AY} = \boldsymbol{B}$ where

$$\mathbf{A} = \begin{bmatrix} 4 & 7 \\ 2 & 3 \end{bmatrix}$$

and

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

and

$$\mathbf{B} = \begin{bmatrix} ? \\ 12 \end{bmatrix}$$

(b) $\mathbf{Y} = \mathbf{A^{-1}B}$

$$\begin{bmatrix} ? \\ 1 \end{bmatrix}$$

(Hint: What is the intersection point, (y_1, y_2), of the system of equations? Use your calculator: type A, B into your calculator using 2nd MATRIX EDIT, then calculate $A^{-1}B$.)

(5.17) random vectors

(a) $\mathbf{W} = \mathbf{AY}$, or

$$\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \times \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

(b) expected value, $\mathbf{E}\{\mathbf{W}\} = \mathbf{AE}\{\mathbf{Y}\}$

$$\begin{bmatrix} E\{W_1\} \\ E\{W_2\} \\ E\{W_3\} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \times \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ E\{Y_3\} \end{bmatrix}$$

(c) variance–covariance, $\boldsymbol{\sigma^2}\{\mathbf{W}\} = \mathbf{A}\boldsymbol{\sigma^2}\{\mathbf{Y}\}\mathbf{A}'$, or

$$\boldsymbol{\sigma^2}\{\mathbf{W}\} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \times \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} & \sigma\{Y_1, Y_3\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} & \sigma\{Y_2, Y_2\} \\ \sigma\{Y_3, Y_1\} & \sigma\{Y_3, Y_2\} & \sigma^2\{Y_3\} \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

(5.26) Plastic Hardness, matrix manipulations, qz2-5-26-plastic-matrices.
Using the TI–83 or from the SAS output[4]

(a) In this case,

$$\mathbf{X} = \begin{bmatrix} 1 & 16 \\ 1 & 16 \\ \vdots & \vdots \\ 1 & 40 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} 199 \\ 205 \\ \vdots \\ 246 \end{bmatrix}$$

and so

(1) $(\mathbf{X'X})^{-1} =$

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 16 & 16 & \cdots & 40 \end{bmatrix} \times \begin{bmatrix} 1 & 16 \\ 1 & 16 \\ \vdots & \vdots \\ 1 & 40 \end{bmatrix} = \begin{bmatrix} 0.675000 & -0.021875 \\ -0.021875 & 0.00078125 \end{bmatrix}$$

(2) $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X}^{-1}\mathbf{Y} =$

$$\begin{bmatrix} 168.6 \\ 2.0343 \end{bmatrix}$$

(3) $\hat{\mathbf{Y}} = \mathbf{Xb} =$

$$\begin{bmatrix} 201.15 \\ 201.15 \\ \vdots \\ 249.975 \end{bmatrix}$$

(4) $\mathbf{H} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'} =$

$$\begin{bmatrix} 0.175 & \cdots & -0.050 \\ \vdots & & \vdots \\ -0.050 & \cdots & 0.175 \end{bmatrix}$$

(5) $SSE = \mathbf{Y'}\left[\mathbf{I} - \mathbf{H}\right]\mathbf{Y} = 146.425$

(6) $\mathbf{s^2\{b\}} = MSE \times (\mathbf{X'X})^{-1} =$

$$10.45893 \times \begin{bmatrix} 0.675000 & -0.021875 \\ -0.021875 & 0.00078125 \end{bmatrix} = \begin{bmatrix} 7.0598 & -0.2288 \\ -0.2288 & 0.008171 \end{bmatrix}$$

(7) At $X_h = 30$, $s^2\{pred\} = MSE(\mathbf{1} + \mathbf{X'_h}(\mathbf{X'X})^{-1}\mathbf{X_h}) = 11.1453$

[4]The I option gives

$$\begin{bmatrix} (\mathbf{X'X})^{-1} & \hat{\beta} \\ (\hat{\beta})^{-1} & SSE \end{bmatrix}$$

(b) variance–covariance
Since

$$\mathbf{s^2\{b\}} = \begin{bmatrix} s^2\{b_0\} & s\{b_0, b_1\} \\ s\{b_0, b_1\} & s^2\{b_1\} \end{bmatrix} = \begin{bmatrix} 0.675000 & -0.021875 \\ -0.021875 & 0.00078125 \end{bmatrix}$$

(1) $s^2\{b_0\} = 7.0598$
(2) $s\{b_0, b_1\} = -0.2288$
(3) $s\{b_1\} = \sqrt{0.008171} = 0.0904$

(c) matrix of quadratic form for *SSE* is $\mathbf{I} - \mathbf{H} =$

$$\begin{bmatrix} 0.825 & \cdots & 0.050 \\ \vdots & & \vdots \\ 0.050 & \cdots & 0.825 \end{bmatrix}$$