STAT 353 QUIZ 1 SOLUTION

Instructions: This is a closed book quiz. A Sharp EL-510-series calculator is permitted, no other material are allowed. To complete this quiz: [1] write your solution on paper first; show some work, a numerical answer alone is insufficient; [2] scan the paper into a PDF file; if you have two or more pages, be sure to scan them all into a single PDF file; [3] go to Brightspace → Assignments → Quiz 1 to submit your PDF file. Make sure the file is readable before submission. Unreadable files will get a score of zero.

Suppose $X_1 \sim \chi^2_3$, $X_2 \sim \chi^2_1$ and $Y = (Y_1, Y_2, Y_3)^\top \sim N(\mu, \Sigma)$ where

$Y_{(v)}^2 + \chi_1^2$

$$\mu = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and X_1, X_2 and Y are independent. Also, if $X \sim \chi^2_k$, then $E(X) = k$ and $V(X) = 2k$.

1. [1 mark] Find $E(3Y_1 + 2Y_2 + 1)$.

$$3E(Y_1) + 2E(Y_2) + 1 = 3 \times 2 + 2 \times 1 + 1 = 9$$

[Must show the 1st or 2nd step above and have the correct numerical answer 9 to receive 1 mark. Give 0 if either the step(s) or the numerical answer is missing. Same rule applies to Questions 2 to 4.]

2. [1 mark] Find $V(3Y_1 + 2Y_2 + 1)$.

$$3^2V(Y_1) + 2^2V(Y_2) + 2 \times 6 \times Cov(Y_1, Y_2) = 9 \times 3 + 4 \times 2 + 12 \times 1 = 47$$

3. [1 mark] $Cov(Y_1 + Y_2, Y_1 - Y_2)$.

$$Cov(Y_1, Y_1) - Cov(Y_1, Y_2) + Cov(Y_2, Y_1) - Cov(Y_2, Y_2) = 3 - 2 = 1$$

$Cov(Y_1, Y_2) = Cov(Y_2, Y_1)$

4. [1 mark] Find $E(Y_1 + X_1X_2)$.

$$E(Y_1) + E(X_1)E(X_2) = 2 + 3 \times 1 = 5$$

For the random variable W defined by each of the following equations, find its distribution if it is one of the following four distributions: $N(\mu, \sigma^2)$, t_k, χ^2_k and $F(m, n)$. If not, answer "distribution unknown". If your answer is one of the four, you must give values of the parameters or degrees of freedom.

5. [1 mark] $W = X_2 + Y_3^2$. [Answer: χ^2_2. No derivation required but give 0 if df not given or incorrect.]

$\sqrt{Y_3^2}$

6. [1 mark] $W = \frac{Y_1 - 2}{\sqrt{X_1}}$. [Answer: t_3. No derivation required. Give 0 mark if df not given or incorrect.]

$T = \frac{Z}{\sqrt{\frac{\chi}{k}}} \sim t_k$ $\quad \frac{Y-3}{\sqrt{3}}$

7. [1 mark] $W = \frac{X_1}{3X_2}$ [Answer: $F_{3,1}$. No derivation required. Give 0 if df's not given or incorrect.]

$\frac{\chi^2_3}{3\chi^2_1} = F = \frac{\chi^2_m / m}{\chi^2_n / n} = \frac{\chi^2_3 / 3}{\chi^2_1 / 1} = F_{3,1}$

8. [2 marks] $W = Y_1 + Y_2 + Y_3$. [Answer: $N(3, 8)$]

$E(W) = 2 + 1 + 0 = 3,$ [1 mark for mentioning normal, 1 for correct mean and variance]
$V(W) = V(Y_1) + V(Y_2) + V(Y_3) + 2Cov(Y_1, Y_2)$
$= 3 + 2 + 1 + 2 = 8$

9. [1 mark] $W = Y_1^2 + Y_2^2 + Y_3^2$. [Answer: distribution unknown]

The End

STAT 353 QUIZ 2 SOLUTION — Due by 11 pm, October 6

Instructions: This is a closed book quiz. A Sharp EL-510-series calculator is permitted, no other material are allowed. To complete this quiz: [1] write your solution on paper first; show some work, a numerical answer alone is insufficient; [2] scan the paper into a PDF file; if you have two or more pages, be sure to scan them all into a single PDF file; [3] go to Brightspace → Assignments → Quiz 1 to submit your PDF file. Make sure the file is readable before submission. Unreadable files will get a score of zero.

Problem 1. [6 marks] To study the impact of video gaming on academic performance, 20 students from a statistics course were randomly selected to write a test. Their test scores y and the average amount of time they each spent on playing video games per day x were recoded. Fitting a simple linear regression model to the data using command "fit = lm($y \sim x$)" in R yields the following output:

```
> summary(fit)
Call:  lm(formula = y ~ x)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  88.0000     1.6647   52.86  < 2e-16 ***
x            -2.9163     0.2268  -12.86 1.64e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.621 on 18 degrees of freedom
Multiple R-squared:  0.9018,    Adjusted R-squared:  0.8964
F-statistic: 165.4 on 1 and 18 DF,  p-value: 1.643e-10
```

$\hat{y} = 88 +$ [illegible] -2 [illegible]

(1a) [1 mark] Give the estimated regression line.

$\hat{y} = 88.00 - 2.92x$ [okay if un-rounded or other rounded estimates are used, e.g., $\hat{y} = 88 - 2.91x$]

(1b) [2 marks] Test for significance of regression at $\alpha = 0.01$ level [~~must state the null and alternative hypotheses, give the p-value, and state the conclusion; -1 for missing one of these 3 components, -2 for missing two or more of these components; i.e., 0 mark for missing two or more components~~].

H_0 : The regression not significant (or $\beta_1 = 0$) vs H_1 : The regression is significant (or $\beta_1 \neq 0$).

The p-value$= 1.64 \times 10^{-10} < 0.01$, so we reject H_0 and conclude that the regression is significant.

Ie is significare that there is a regression

0

(1c) [1 mark] Find the estimate of the error variance σ^2.

$\hat{\sigma}^2 = 3.621^2 = 13.112.$

(1d) [1 mark] What percentage of the total variability in test scores is explained by the linear model?

$R^2 = 0.9018$, so 90.18% of the total variability in the test scores are explained by the linear model.

(1e) [1 marks] Find the 95% confidence interval for the slope of the model β_1. [Some t critical values: $t_{0.025,18} = 2.101$, $t_{0.05,18} = 2.552$, $t_{0.025,19} = 2.093$, $t_{0.05,19} = 2.539$, $t_{0.025,20} = 2.086$, $t_{0.05,20} = 2.528$.]

$-2.9163 \pm t_{0.025,18} \times 0.2268 \Rightarrow (-3.3928, -2.4397)$. [okay if answer is rounded differently]

Problem 2. [4 marks] Consider linear model $y = \beta + \beta x + \varepsilon$ where the intercept and slope are both β, x is the predictor variable and ε is the random error satisfying $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma^2$.

(2a) [3 marks] With n data points (x_i, y_i) for $i = 1, 2, \ldots, n$, find the least squares estimator for β.

The error sum of squares is $s(\beta) = \sum_{i=1}^{n}(y_i - \beta - \beta x_i)^2$ [1 mark]. To find the minimizer of $s(\beta)$,

$$\frac{ds(\beta)}{d\beta} = \sum_{i=1}^{n} 2(y_i - \beta - \beta x_i)(-1 - x_i) = 0, \text{ so [1 mark]}$$

$= \sum_{i=1} (y_i - \beta(1+x_i))(1+x_i) = 0$

$$\sum_{i=1}^{n} y_i(1 + x_i) = \beta \sum_{i=1}^{n} (1 + x_i)^2.$$

Thus, the least squares estimator for β is

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i(1 + x_i)}{\sum_{i=1}^{n}(1 + x_i)^2}. \qquad \text{[1 mark]}$$

(2b) [1 mark] Denote by $\hat{\beta}$ the least squares estimator found in (2a). Find $V(\hat{\beta})$. [That is, express $V(\hat{\beta})$ in terms of the data values (x_i, y_i) and σ^2].

Since $V(y_i(1 + x_i)) = (1 + x_i)^2 V(y_i) = (1 + x_i)^2\sigma^2$ and y_i's are assumed to be independent,

$$V(\hat{\beta}) = \frac{V\left(\sum_{i=1}^{n} y_i(1 + x_i)\right)}{\left[\sum_{i=1}^{n}(1 + x_i)^2\right]^2} = \frac{\sum_{i=1}^{n}(1 + x_i)^2\sigma^2}{\left[\sum_{i=1}^{n}(1 + x_i)^2\right]^2} = \frac{\sigma^2}{\sum_{i=1}^{n}(1 + x_i)^2}.$$

The End

STAT 353 QUIZ 3 SOLUTION **Due by 11 pm on October 22**

Instructions: This is a closed book quiz. A Sharp EL-510-series calculator is permitted, no other material are allowed. To complete this quiz: [1] write your solution on paper first; show some work, a numerical answer alone is insufficient; [2] scan the paper into a PDF file; if you have two or more pages, be sure to scan them all into a single PDF file; [3] go to Brightspace → Assignments → Quiz 3 to submit your PDF file. Make sure the file is readable before submission. Unreadable files will get a score of zero.

Problem 1. [7 marks] A data set on gas consumption of 30 cars includes the following variables:

y – gas consumption (in miles/gallon),
x_1 – engine size (cylinder displacement in cubic inches),
x_2 – engine horsepower,
x_3 – engine torque,
x_4 – engine compression ratio,
x_5 – rear axle ratio.

The following two models are fitted to the data using the R code on page 2:

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$
Model 2: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$

(1a) [1 mark] Using the fitted Model 1 shown on page 2, compute by hand the predicted gas consumption for a car with engine size $x_1 = 300$, horsepower $x_2 = 200$ and torque $x_3 = 250$.

$\hat{y} = 32.621 - 0.078(300) + 0.007(200) + 0.040(250) = 20.621$ [rounding to 20.6 or 20.62 okay]

(1b) [1 mark] Is there evidence of multicollinearity in Model 1? Briefly explain.

Yes. VIF of all three variables are more than 10.

(1c) [3 marks] For Model 2, use a partial F test to assess the contribution of engine compression ratio x_4 and rear axle ratio x_5 given all other regressors are included. State the null and alternative hypotheses, compute the value of the test statistic and give the conclusion at $\alpha = 0.05$. [Some F distribution critical values: $F_{0.05,1,28} = 4.196$, $F_{0.05,2,24} = 3.403$, $F_{0.05,2,26} = 3.370$]

$H_0 : \beta_4 = \beta_5 = 0$ vs H_1 : at least of the β_4, β_5 is not zero. [1 mark]

$$F_{obs} = \frac{(263.31 - 195.49)/2}{195.49/24} = 4.163; \quad p\text{-value} = P(F_{2,24} > 4.163) < P(F_{2,24} > 3.403) = 0.05$$

Reject H_0 at the 5% level; contribution from x_4, x_5 significant. [-1 for wrong F_{obs}, -1 for wrong p-value bound, and -1 for wrong conclusion; i.e. give $0/2$ for two or more mistakes.]

(1d) [1 mark] For Model 2, suppose you wish to test hypothesis $H_0 : \beta_1 - \beta_2 = 0$ and $\beta_4 = 0$. Give the reduced model under this H_0. [either one of the following two expressions is okay]

$$y = \beta_0 + \beta_1(x_1 + x_2) + \beta_3 x_3 + \beta_5 x_5 + \varepsilon = \beta_0 + \beta_2(x_1 + x_2) + \beta_3 x_3 + \beta_5 x_5 + \varepsilon$$

(1e) [1 mark] Suppose you are to add a Part 3 to the R code about the reduced model in (1d). Give the R commands needed to fit the reduced model to the data.

```
xnew=x1+x2   [other variable names okay but must define new variables correctly]
mdlreduced=lm(y~xnew+x3+x5)      [fitted model must be correct]
```

Problem 2. [3 marks] The multiple linear regression model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\mathbf{y}$ is the vector of responses, $\mathbf{X}$ is the fixed matrix of regressor variable values and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ is the random error. The least squares estimate of the parameter vector $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$. The hat matrix is $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$.

(2a) [1 mark] Show that $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$.

$$E(\hat{\boldsymbol{\beta}}) = E((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top E(\mathbf{y}) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

(2b) [1 mark] Let $\mathbf{e}$ be the residual of the least squares fit. Show that $\mathbf{e} = (\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon}$.

The fitted value is $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, so $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}$. It follows that

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{y} = (\mathbf{I} - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = (\mathbf{X} - \mathbf{H}\mathbf{X})\boldsymbol{\beta} + (\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon} = (\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon}.$$

(2c) [1 mark] What is the function $S(\boldsymbol{\beta})$ that $\hat{\boldsymbol{\beta}}$ minimizes? Express $S(\boldsymbol{\beta})$ in terms of $\mathbf{X}$, $\mathbf{y}$ and $\boldsymbol{\beta}$.

$$S(\boldsymbol{\beta}) = \boldsymbol{\epsilon}^\top\boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

```
PROBLEM 1 R CODE AND OUTPUT

> ##### Part 1: Model 1 related R commands and output #####
>
> mdl1=lm(y~x1+x2+x3)
> summary(mdl1)
Call: lm(formula = y~x1+x2+x3)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 32.620594   2.050397  15.909 6.45e-15 ***
x1          -0.077808   0.036767  -2.116   0.0441 *
x2           0.007284   0.053111   0.137   0.8920
x3           0.039820   0.065881   0.604   0.5508
---
Residual standard error: 3.182 on 26 degrees of freedom
Multiple R-squared: 0.7688,     Adjusted R-squared: 0.7422
F-statistic: 28.83 on 3 and 26 DF,  p-value: 1.995e-08

> anova(mdl1)
Analysis of Variance Table
Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x1         1 866.50  866.50 85.5590 1.048e-09 ***
x2         1   5.60    5.60  0.5525    0.4639
x3         1   3.70    3.70  0.3653    0.5508
Residuals 26 263.31   10.13
---
> vif(mdl1)
      x1       x2       x3
```

```
52.30027 16.10358 85.76325


> ##### Part 2: Model 2 related R commands and output #####
>
> mdl2=lm(y~x1+x2+x3+x4+x5)
> anova(mdl2)
Analysis of Variance Table
Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x1         1 866.50  866.50 84.7132 2.421e-09 ***
x2         1   5.60    5.60  0.5471    0.4667
x3         1   3.70    3.70  0.3617    0.5532
x4         1  15.59   15.59  1.5243    0.2289
x5         1  52.24   52.24  6.4134    0.0182
Residuals 24 195.49    8.15
```

STAT 353 QUIZ 4 SOLUTION **Due by 11 pm on October 29**

Instructions: This is a closed book quiz. A Sharp EL-510-series calculator is permitted, no other material are allowed. To complete this quiz: [1] write your solution on paper first; show some work, a numerical answer alone is insufficient; [2] scan the paper into a PDF file; if you have two or more pages, be sure to scan them all into a single PDF file; [3] go to Brightspace → Assignments → Quiz 4 to submit your PDF file. Make sure the file is readable before submission. Unreadable files will get a score of zero.

Problem 1. [2 marks] Consider the following six statements about the multiple linear regression model $y = \beta_0 + \beta_1 x_2 + \cdots + \beta_k x_k + \varepsilon = \underline{x}^T \underline{\beta} + \varepsilon$:

(1) $\beta_0 \neq 0$;
(2) At least one of the β_j's is not zero;
(3) $Var(\varepsilon) = \sigma^2$ is a constant that does not depend on the x_i's;
(4) y and the x_i's are uncorrelated;
(5) For $i \neq j$, x_i and x_j are uncorrelated.
(6) $E(y|\underline{x}) = \beta_0 + \beta_1 x_2 + \cdots + \beta_k x_k$.

Which statements are part of the assumptions for the regression model? List the numbers of only and all such statements, e.g., if you think only the first two statements are, answer "(1) and (2)".

Answer: (3) and (6). [2 marks]

[-1 for each missing correct statement and -1 for each inclusion of a wrong statement, e.g., 1 mark for "(3)" or "(6)" or "(2), (3) and (6)", 0 mark for "(2) and (3)" or "(2), (3), (5) and (6)" or "all six".]

Problem 2. [3 marks] A linear regression model is fitted to a dataset (x_i, y_i) for $i = 1, 2, \ldots, n$. Denote by $\hat{y}_i$ the fitted value and by $\hat{\sigma}^2$ the residual mean squares (MS_{RES}).

(2a) [1 mark] Define the raw residual e_i and standardized residual d_i using y_i, $\hat{y}_i$ and $\hat{\sigma}^2$.

$$e_i = y_i - \hat{y}_i \quad \text{[give 0 if either } e_i \text{ or } d_i \text{ is not correctly defined.]}$$

$$d_i = \frac{e_i}{\hat{\sigma}} = \frac{y_i - \hat{y}_i}{\hat{\sigma}} = \frac{e_i}{\sqrt{MS_{RES}}} \quad \text{[any one of these expressions is fine.]}$$

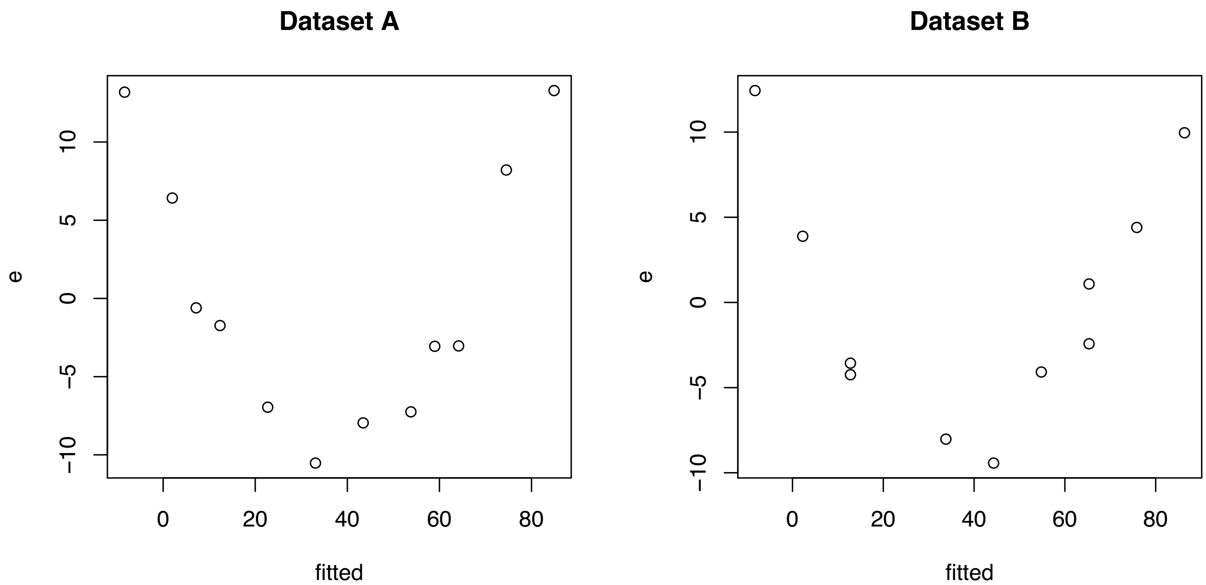
(2b) [1 mark] How do you use d_i to identify potential outliers? Only stating how, no need to explain.

Answer I: If $|d_i| > 3$, then the ith point is a potential outlier, or [give 1 mark to either I or II.]

~~Answer II: Plot d_i against i or $\hat{y}_i$. Those d_i's outside the ± 3 limits indicate potential outliers.~~

(2c) [1 mark] When plotting e_i against the fitted value $\hat{y}_i$, briefly describe the expected shape of the plot if there are no violations of the model assumptions and no outliers.

The plot is expected to look like a random/uniform band around the x-axis. [Give 1 mark so long as the term "uniform band" or "random band" is mentioned].



Problem 3. [3 marks] The simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ is fitted to Dataset A and Dataset B, respectively, and the resulting residual plots of e_i vs fitted value $\hat{y}_i$ are given above.

(3a) [2 marks] It is known that only one of the datasets allows test on the lack-of-fit of the simple linear regression model. Which dataset is it? Briefly explain.

Dataset B [1 mark]. This is because its residual plot shows there are replicates of residuals at two fitted values which means there are replicate observations of y at two x values [1 mark; okay if they simply say there are replicate observations of y at some x values in Dataset B based its residual plot].

(3b) [1 mark] Suppose the lack-of-fit test for the dataset in (3a) is significant (say, p-value < 0.05), what alternative model would you try to fit to this dataset? Briefly explain.

The residual plot has a strong quadratic shape. Try to fit model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$. [Give 1 mark for this answer even if their choice of dataset in (3a) is wrong as this point applies to both datasets.]

Problem 4. [2 marks] Let $y_1, y_2, \ldots, y_n$ be a random sample from an unknown distribution and $q_i = \Phi^{-1}(\frac{1}{n}(i - \frac{1}{2}))$ for $i = 1, 2, \ldots, n$ where $\Phi(z)$ is the CDF of $N(0, 1)$. Briefly describe (i) how to make a normal probability plot using the y_i's and the q_i's, and (ii) how to interpret this plot.

(i) Let $y_{[1]} \leq y_{[2]} \leq \ldots y_{[n]}$ be the ordered version of the sample. Plot points $(q_i, y_{[i]})$ for $i = 1, 2, \ldots, n$.

(ii) If the points form approximately a line, then it indicates the unknown distribution is a normal distribution. If they do not form a line, the distribution is not normal.

[1 mark for (i), 1 mark for (ii). For (ii), okay if only mention a line indicates a normal distribution.]

STAT 353 QUIZ 5 SOLUTION Due by 11 pm on November 5

Instructions: This is a closed book quiz. A Sharp EL-510-series calculator is permitted, no other material are allowed. To complete this quiz: [1] write your solution on paper first; show some work, a numerical answer alone is insufficient; [2] scan the paper into a PDF file; if you have two or more pages, be sure to scan them all into a single PDF file; [3] go to Brightspace → Assignments → Quiz 5 to submit your PDF file. Make sure the file is readable before submission. Unreadable files will get a score of zero.

Problem 1. [3 marks] Consider fitting the linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ to a dataset (x_i, y_i) for $i = 1, 2, \ldots, n$. Let $\hat{y}_i$ be the resulting fitted value and e_i be the residual.

(1a) [1 mark] When should we consider transforming y and fit the linear model $y' = \beta_0 + \beta_1 x + \varepsilon'$ to (x_i, y_i') where y_i' denotes the transformed y_i? Your answer must be based on the plot of $(\hat{y}_i, e_i)$.

When residual plot of $(\hat{y}_i, e_i)$ shows the variance of the residuals is not a “uniform band”. ~~For example, the plot has a funnel shape showing that the variance of the residuals is increasing with the fitted value. [1 mark for any mentioning of a non-random trend such as increasing or decreasing trend]~~

(1b) [1 mark] If y is Poisson or count data, which transformation is appropriate?

For Poisson/count data, the “square root transformation” is appropriate; that is $y' = \sqrt{y}$.

[1 mark for any mention of root transformation or $\sqrt{y}$ transformation.]

(1c) [1 mark] Name one or two ways for choosing a transformation. No need to explain.

(1) Use the data type, and (2) use the empirical relationship. The latter can be identified by studying the residual plot of $(\hat{y}_i, e_i)$. [1 mark for mentioning either (1) or (2), and no explanation required]

Problem 2. [2 marks] Give the linearized versions of the nonlinear models in (2a) and (2b). Write your answers in the form of, for example, $y' = \beta_0' + \beta_1' x_1' + \beta_2' x_2' + \beta_3' x_3' + \varepsilon'$, and clearly define the y', x_i', β_i' and ε' in the linearized model in terms of the y, x_i, β_i and ε in the original model.

(2a) [1 mark] $y = \beta_0 (x_1)^{\beta_1} (\beta_2)^{x_2} (e^{x_3 \beta_3}) \varepsilon$.

$$\ln y = \ln \beta_0 + \beta_1 \ln x_1 + (\ln \beta_2) x_2 + \beta_3 x_3 + \ln \varepsilon \quad \Rightarrow \quad y' = \beta_0' + \beta_1' x_1' + \beta_2' x_2' + \beta_3' x_3' + \varepsilon'$$

where $y' = \ln y, x_1' = \ln x_1, x_2' = x_2, x_3' = x_3$, $\beta_0' = \ln \beta_0, \beta_1' = \beta_1, \beta_2' = \ln \beta_2, \beta_3' = \beta_3$ and $\varepsilon' = \ln \varepsilon$.

[1 mark; y', x_i', β_i' and ε' must be correctly defined.]

(2b) [1 mark] $y = \frac{x^2}{\beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon}$

$$\frac{1}{y} = \beta_2 + \beta_1 \frac{1}{x} + \beta_0 \frac{1}{x^2} + \frac{1}{x^2}\varepsilon \quad \Rightarrow \quad y' = \beta_0' + \beta_1' x_1' + \beta_2' x_2' + \varepsilon'$$

where $y' = y^{-1}, x_1' = x^{-1}, x_2' = x^{-2}, \beta_0' = \beta_2, \beta_1' = \beta_1, \beta_2' = \beta_0$ and $\varepsilon' = \varepsilon x^{-2}$.

[1 mark; y', x_i', β_i' and ε' must be correctly defined.]

Problem 3. [5 marks] Consider linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $E(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{V}$ where $\mathbf{V}$ is a known symmetric positive definite matrix. The ordinary least squares estimator (LSE) for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_o = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, and the generalized least squares estimator (GLSE) is $\hat{\boldsymbol{\beta}}_g = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$.

(3a) [1 mark] Show that the GLSE $\hat{\boldsymbol{\beta}}_g$ is an unbiased estimator for $\boldsymbol{\beta}$.

$$E(\hat{\boldsymbol{\beta}}_g) = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}E(\mathbf{y}) = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

(3b) [1 mark] Show that $V(\hat{\boldsymbol{\beta}}_g) = \sigma^2(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$.

Since $\mathbf{V}$ is symmetric, $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$ and $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$ are also symmetric. Thus,

$V(\hat{\beta}$ is clean

$$\begin{aligned}
V(\hat{\boldsymbol{\beta}}_g) &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}V(\mathbf{y})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} \\
&= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\sigma^2\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}.
\end{aligned}$$

[okay if the symmetry of the matrices were not mentioned; assume they know this]

(3c) [1 mark] What is the advantage of the GLSE over the LSE? No need to explain.

$V(\hat{\boldsymbol{\beta}}_g) \leq V(\hat{\boldsymbol{\beta}}_o)$, or the GLSE has a "smaller" variance, or the GLSE is optimal.

[1 mark for either one of the above answers, no explanation required.]

(3d) [1 mark] When do we call the GLSE a weighted least squares estimator (WLSE)?

When the $\mathbf{V}$ matrix is diagonal $\mathbf{V} = diag(v_{11}, v_{22}, \ldots, v_{nn})$.

[okay if only mentioned $\mathbf{V}$ is diagonal but did not give $\mathbf{V} = diag(v_{11}, v_{22}, \ldots, v_{nn})$.

(3e) [1 mark] What is the function $S(\boldsymbol{\beta})$ that the WLSE minimizes?

Let $\mathbf{W} = \mathbf{V}^{-1} = diag(v_{11}^{-1}, v_{22}^{-1}, \ldots, v_{nn}^{-1}) = diag(w_1, w_2, \ldots, w_n)$. Then,

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T\mathbf{W}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) = \sum_{i=1}^n w_i(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2 = \sum_{i=1}^n w_i e_i^2 = \sum_{i=1}^n w_i\varepsilon_i^2$$

[any one of the above expressions would be okay, no need to have them all]

$V(\hat{\beta}_g) < V(\tilde{\beta})$

STAT 353 QUIZ 6 SOLUTION — Due by 11 pm on Sunday, November 12

Instructions: This is a closed book quiz. A Sharp EL-510-series calculator is permitted, no other material are allowed. To complete this quiz: [1] write your solution on paper first; show some work, a numerical answer alone is insufficient; [2] scan the paper into a PDF file; if you have two or more pages, be sure to scan them all into a single PDF file; [3] go to Brightspace → Assignments → Quiz 6 to submit your PDF file. Make sure the file is readable before submission. Unreadable files will get a score of zero.

Problem 1. [2 marks] Consider linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $E(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{V}$, and suppose $\mathbf{V}$ is known. The ordinary least squares estimator (OLSE) for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_o = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, and the generalized least squares estimator (GLSE) for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_g = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$.

(1a) [1 mark] What does the statement "the GLSE $\hat{\boldsymbol{\beta}}_g$ is the best linear unbiased estimator (BLUE) for $\boldsymbol{\beta}$" mean? Briefly explain.

Among all unbiased linear estimators, i.e~~., $\tilde{\boldsymbol{\beta}} = C\mathbf{y}$ such that $E(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$~~, the GLSE $\hat{\boldsymbol{\beta}}_g$ has the smallest variance, i.e., $V(\hat{\boldsymbol{\beta}}_g) \le V(\tilde{\boldsymbol{\beta}})$. [okay if only mention GLSE has the smallest variance among unbiased linear estimators without expressing it in terms of $\tilde{\boldsymbol{\beta}} = C\mathbf{y}$.]

(1b) [1 mark] Under what condition is the OLSE $\hat{\boldsymbol{\beta}}_o$ the BLUE?

The OLSE is the BLUE if $\mathbf{V} = \mathbf{I}$. [okay if saying errors are uncorrelated and of equal variance.]

Problem 2. [5 marks] Consider fitting multiple regression model $\mathbf{y}_i = \mathbf{x}_i^T\boldsymbol{\beta} + \varepsilon_i$ to a dataset $(\mathbf{x}_i, y_i)$ for $i = 1, 2, \ldots, n$ where there are k regressors and $p = k + 1$.

(2a) [2 marks] What is a leverage point? And how do we decide whether the ith point $(\mathbf{x}_i, y_i)$ is a leverage point by using the ith diagonal element of the $\mathbf{H}$ matrix h_{ii}?

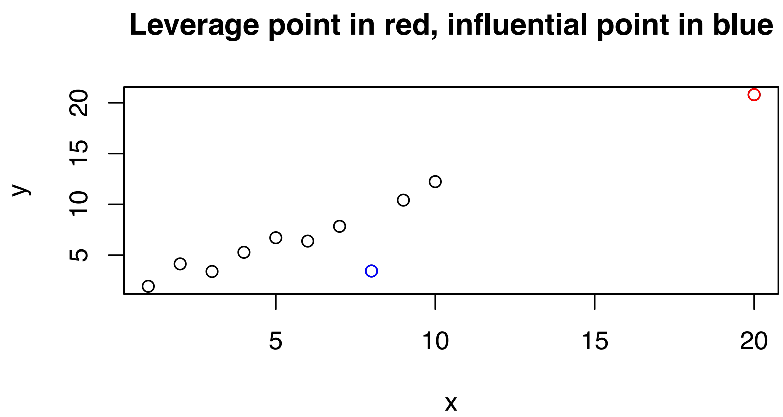
A point $(\mathbf{x}_i, y_i)$ is a leverage point if its $\mathbf{x}_i$ value is extreme relative to that of the other points but it has little impact on the $\hat{\boldsymbol{\beta}}$ [1 mark]. If $h_{ii} > 2p/n$, then $(\mathbf{x}_i, y_i)$ is a leverage point. [1 mark]

(2b) [2 mark] What is an influential point? And usually when do we call a point $(\mathbf{x}_i, y_i)$ an influential point by using its Cook's distance D_i?

A point $(\mathbf{x}_i, y_i)$ is influential if it has a big impact on $\hat{\boldsymbol{\beta}}$ [1 mark, okay with saying the $\hat{\boldsymbol{\beta}}$ value will change substantially when such a point is removed from the dataset]. If $D_i > F_{0.5,p,n-p}$, then $(\mathbf{x}_i, y_i)$ is influential [1 mark, okay if saying $D_i > 1$, then $(\mathbf{x}_i, y_i)$ is influential.]

(2c) [1 mark] Sketch a scatter plot of a dataset (x_i, y_i) with one leverage point and one influential point.

See the plot on the next page. Okay if used two plots, one to show a leverage point and another to show an influential point; see, e.g., Figures 6.1 and 6.2 in the textbook. Also okay if a combined leverage and influential point is shown.



Problem 3. [3 marks] Consider the one regressor polynomial regression model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \varepsilon$$

(3a) [2 mark] Describe the "forward selection" method for determining the model order p.

Step 1: fit the 1st order model $y = \beta_0 + \beta_1 x + \varepsilon$ and test $H_0 : \beta_1 = 0$. Suppose H_0 is rejected.

Step 2: fit the 2nd order model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ and test $H_0 : \beta_2 = 0$. Suppose H_0 is rejected

......

Step j: fit the jth order model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_j x^j + \varepsilon$ and test $H_0 : \beta_j = 0$ and suppose H_0 is rejected.

Step $(j+1)$: fit the $(j+1)$th order model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_j x^j + \beta_{j+1} x_{j+1}^{j+1} + \varepsilon$ and test $H_0 : \beta_{j+1} = 0$ and suppose H_0 is not rejected.

Then, the forward selection sets $p = j$.

[if "backward selection" is described instead, give 1 mark]

(3b) [1 mark] Briefly describe the ill-conditioning problem that may be caused by a high p and/or a narrow range of the x variable.

Let $\mathbf{X}$ be the design matrix of the polynomial regression model. When p is too high and/or the range of the x variable is narrow, $\mathbf{X}^T\mathbf{X}$ is nearly singular. This will lead to unreliable least squares estimate.

[okay if only mention a high p and/or a narrow range of the x variable generates multicollinearity, or only mention that columns of $\mathbf{X}$ are nearly linearly dependent.]

STAT 353 QUIZ 7 SOLUTION **Due by 11 pm on Friday, November 24**

Instructions: This is a closed book quiz. A Sharp EL-510-series calculator is permitted, no other material are allowed. To complete this quiz: [1] write your solution on paper first; show some work, a numerical answer alone is insufficient; [2] scan the paper into a PDF file; if you have two or more pages, be sure to scan them all into a single PDF file; [3] go to Brightspace → Assignments → Quiz 7 to submit your PDF file. Make sure the file is readable before submission. Unreadable files will get a score of zero.

Problem 1. [4 marks] Consider the following piecewise polynomial function

$$S(x) = x + (x-1)_+^2 + 2(x-2)_+ \qquad \text{for } x \in [0,3].$$

(1a) [1 mark] Find the value of $S(1.9)$.

$$S(1.9) = 1.9 + (1.9-1)_+^2 + 2(1.9-2)_+ = 1.9 + 0.9^2 = 2.71$$

(1b) [2 marks] Find the first derivative $S'(x)$.

$$S(x) = \begin{cases} x & \text{for } \ 0 \le x \le 1 \\ x + (x-1)^2 & \text{for } \ 1 < x \le 2 \\ x + (x-1)^2 + 2(x-2) & \text{for } \ 2 < x \le 3 \end{cases}$$

So

$$S'(x) = \begin{cases} 1 & \text{for } \ 0 \le x \le 1 \\ 1 + 2(x-1) = -1 + 2x & \text{for } \ 1 < x \le 2 \\ 1 + 2(x-1) + 2 = 1 + 2x & \text{for } \ 2 < x \le 3 \end{cases}$$

Alternatively, $S'(x)$ may be expressed in a single form as

$$S'(x) = 1 + 2(x-1)_+ + 2(x-2)_+^0 \qquad \text{for } \ 0 \le x \le 3$$

[1 mark if $S'(x)$ is not completely right but partially right for one or two intervals; for the single form answer, 1 mark if $S'(x)$ is not completely right but contains the $(x-1)_+$ or $(x-2)_+^0$ term.]

(1c) [1 mark] At what x value or values in interval $[0,3]$ is $S'(x)$ not continuous?

$S'(x)$ is not continuous at $x = 2$. [Give 0 for other answers, e.g., $x = 1, 2$ or $x = 0, 1, 3$].

Problem 2. [4 marks] Suppose you want to fit the piecewise polynomial model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 (x-1)_+ + \beta_4 (x-2)_+^2 + \varepsilon$$

to a dataset of n points (x, y) where x is the vector containing the n regressor variable values and y is the corresponding vector of n response variable values.

(2a) [3 marks] Give the R commands for fitting the model.

```
x1=x
x2=x^2
x3=pmax(x-1,0)
x4=pmax(x-2,0)^2
fit=lm(y~x1+x2+x3+x4)
```

[-1 for each missing command to the maximum of -3 (i.e., 0 for this part); okay if different variable names are used or x is used instead of x_1 (in this case the first line of command is unnecessary).]

(2b) [1 mark] How do you determine whether the quadratic model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ is sufficient (so that the piecewise polynomial model is unnecessary)? Briefly explain.

Fit the piecewise polynomial model and test $H_0 : \beta_3 = \beta_4 = 0$. If not rejected, then the quadratic model is sufficient. [okay if they also add testing and rejecting $H_0 : \beta_2 = 0$].

Problem 3. [2 marks] A dataset (x_i, y_i) for $i = 1, 2, \dots, n$ satisfies $x_i \le x_{i+1}$ and

$$(x_1, y_1) = (0, 1.7),\ (x_2, y_2) = (0.5, 2.1),\ (x_3, y_3) = (1, 2.2),\ (x_4, y_4) = (1.5, 2.1),\ (x_5, y_5) = (5, 2.5),\ \dots$$

(3a) [1 mark] Compute the 3-point moving average at x_2, $\hat{y}_2$.

$$\hat{y}_2 = \frac{y_1 + y_2 + y_3}{3} = \frac{1.7 + 2.1 + 2.2}{3} = \frac{6}{3} = 2$$

(3b) [1 mark] Is such a 3-point average at x_4 meaningful? Briefly explain.

No. From the dataset, we see that $(x_3, x_4, x_5) = (1, 1.5, 5)$. The three x values are not equally spaced, so the simple moving average is not meaningful.

The End

STAT 353 QUIZ 8 SOLUTION **Due by 11 pm on Friday, December 1**

Instructions: This is a closed book quiz. A Sharp EL-510-series calculator is permitted, no other material are allowed. To complete this quiz: [1] write your solution on paper first; show some work, a numerical answer alone is insufficient; [2] scan the paper into a PDF file; if you have two or more pages, be sure to scan them all into a single PDF file; [3] go to Brightspace → Assignments → Quiz 8 to submit your PDF file. Make sure the file is readable before submission. Unreadable files will get a score of zero.

Problem 1. [3 marks] We have discussed several criteria for comparing models, including R^2 and the adjusted R^2 (that is, R^2_{adj}). Consider comparing 2 models in the context of variable selection.

(1a) [1 mark] Is the model with a larger R^2 value usually the better model? Briefly explain.

Not necessarily (or no). (a) The full model has the largest R^2 value but in the context variable selection it is usually not the best model. (2) R^2 is an increasing function of the number of variables. We cannot just look at the R^2 value alone, need to take into consideration the number of variables in the model.

[Okay if only one of (a) and (b) is mentioned. Also okay if (a) and (b) are not mentioned but instead mentioning that the model at which the R^2 value does not increase much (or becomes flat) when adding more variables is the model of choice even though this answer is a bit off the target.]

(1b) [1 mark] Is the model with a larger R^2_{adj} value usually the better model? Briefly explain.

Yes. R^2_{adj} has taken into consideration the number of variables.

[Okay if mentioning R^2_{adj} initially increases as the number of variables increases, reaches a maximum at some point and then decreases as more variables are added. In class, we talked about the model with the highest R^2_{adj} being the best by this criterion.]

[For (1a) and (1b), if no explanation is given in both part but answered "No" to (1a) and "Yes" to (1b) correctly, give 1 mark for these two parts, i.e., deduct 1 mark for no explanations].

(1c) [1 mark] Name one or more criteria besides R^2 and R^2_{adj}. No need to explain.

C_p statistic, AIC, BIC [mention at least 1 from this list.]

Problem 2. [2 marks] One response variable y and k predictor variables $x_1, x_2, \ldots, x_k$ are available to build a linear model. Consider applying the method of all possible regression for variable selection.

(2a) [1 mark] What is the total number of subsets/models that are compared by this method?

$$\binom{k}{0} + \binom{k}{1} + \cdots + \binom{k}{k} = 2^k$$

[okay to just give 2^k as the answer]

(2b) [1 mark] Describe one potential problem of this method.

One problem with this method is that when k is large, the number of models to be examined 2^k is huge, leading to computational difficulties and it may even be impossible to compute.

[okay if only mention 2^k may be very large.]

Problem 3. [3 marks] One response variable y and k predictor variables $x_1, x_2, \ldots, x_k$ are available to build a linear model. Briefly describe the method of backward selection of predictor variables.

Step 1: fit the full model with all available predictor variables and compute the partial F statistics of all k parameters of the variables. Suppose the parameter of variable x_m has the smallest statistic, F_m.

Step 2: compare F_m with a pre-chosen F_{out} value. If $F_m > F_{out}$, keep all variables and stop the backward selection process. If $F_m \leq F_{out}$, drop variable x_m and redefine the full model as the model formed by the remaining variables, and then go back to step 1.

Iterate the two steps until none of the remaining variables can be dropped.

[1 mark for step 1 and 2 marks for step 2. Okay if the basic idea is clear even if not stated in 2 steps.]

Problem 4. [2 mark] Consider strategy for variable selection and model building. Do you need to worry about possible outliers and/or variables that may need to be transformed? If yes, when you do look into this issue, before variable selection or after variable selection? Briefly explain.

Yes, we need to worry about possible outliers and variables that may need to be transformed.

We should look into this issue before the variable selection because if there are outliers or variables that need to be transformed, they will affect the variable selection process. For example, an outlier can make all variables insignificant when applying the backward selection, leading to an empty model. A variable may be insignificant because it needs to be but not properly transformed.

[1 mark for answering yes to the first question, and 1 mark for indicating the issue has to be looked into before variable selection. Okay if no examples/explanation.]

The End