STAT 420 Spring 2017
2GR, 2UG

Homework 4

(Due Friday, February 24, by 4:00 p.m.)

Please submit your assignment *on paper*, following the Guidelines for Homework Write-Ups and Submissions.

Instruction: <u>DO</u> select or specify ONE of the following two options:

_____ I am registered for **3 credit hours** via section **2UG** for this course. So, I only need to work on **Problem 2 and Problem 3**. Only Problem 2 and Problem 3 need to be graded and contribute to my score of this homework.

_____ I am registered for **4 credit hours** via section **2GR** for this course. So, I need to work on ALL Problems (i.e. **Problem 1, Problem 2, and Problem 3**). All these THREE Problems need to be graded and contribute to my score of this homework.

---

1. For a constant matrix $\mathbf{A}$ and a random vector $\mathbf{z}$,

$$E(\mathbf{Az}) = \mathbf{A}\, E(\mathbf{z}); \qquad var(\mathbf{Az}) = A\, var(\mathbf{z})\, \mathbf{A}^T$$

(assuming expectations and variances all exist).

Consider the normal multiple linear regression model

$$\mathbf{Y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = var(\boldsymbol{\varepsilon}) = \sigma^2\, \mathbf{I}$.

Determine the mean vector and the variance-covariance matrix for each of the following vectors (in terms of $\mathbf{X}$, $\boldsymbol{\beta}$, and σ^2). Simplify, if possible.

(a) $\mathbf{Y}$

(b) $\hat{\boldsymbol{\beta}}$

(c) $\hat{\mathbf{Y}}$

2. **<u>Do NOT</u> use a computer for this problem** (a calculator is allowed).

We wish to examine the relationship between the age of a vehicle in years (x_1), the odometer mileage (x_2) (in thousands of miles), and the selling price (y) (in thousands of \$) for a particular brand of minivan at *Honest Harry's Used Car Dealership*. The data are as follows:

Consider the model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i,$$

where ε_i's are i.i.d. $N(0, \sigma^2)$.

x_1	x_2	y
1	35	17
2	15	20
3	50	11
3	30	14
4	35	14
5	75	8

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 6 & 18 & 240 \\ 18 & 64 & 820 \\ 240 & 820 & 11700 \end{bmatrix}, \quad \mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 84 \\ 228 \\ 2955 \end{bmatrix},$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 1.157576 & -0.20909 & -0.00909 \\ -0.20909 & 0.190909 & -0.00909 \\ -0.00909 & -0.00909 & 0.000909 \end{bmatrix} = \frac{1}{3300} \cdot \begin{bmatrix} 3820 & -690 & -30 \\ -690 & 630 & -30 \\ -30 & -30 & 3 \end{bmatrix}.$$

$\sum (y_i - \hat{y}_i)^2 = 7.65$ and $\sum (y_i - \bar{y})^2 = 90.$

(a) Find the least-squares estimates $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$.

(b) Perform the significance of the regression test at the 5% level of significance. State the null and alternative hypotheses. Report the value of the test statistic, the critical value(s), and the decision in the context of this problem.

(c) Test $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ at the 5% level of significance. Report the value of the test statistic, the critical value(s), and the decision in the context of this problem.

(d) Test $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 < 0$ at the 5% level of significance. Report the value of the test statistic, the critical value(s), and the decision in the context of this problem.

(e) Test $H_0 : \beta_0 = 25$ vs. $H_1 : \beta_0 < 25$ at a 10% level of significance. Report the value of the test statistic, the critical value(s), and the decision in the context of this problem.

(f) Construct a 95% confidence interval for β_1. Give an interpretation of the interval in the context of this scenario.

(g) Construct a 90% prediction interval for the selling price of a minivan that is 5 years old and has 60 thousand miles on its odometer. Given an interpretation of the interval in the context of this scenario.

(h) What proportion of observed variation in selling price is explained by a linear relationship with the age of a vehicle and the odometer mileage?

3. **DO** use a computer (i.e. R / R Studio) for this problem, and include any relevant R code and output used to achieve the solution.

   Double check or provide your answers in Problem 2 (a) - (h) (i.e. parts a – h in Problem 2). More specifically:

   - Provide the answer for 2(a);
   - For 2(b), 2(c), 2(d), and 2(e), you only need to give (1) the test statistic and (2) either the critical region or the p-value. You may state your decision simply as "Reject H_0" or "Do not reject H_0";
   - For 2(f) and 2(g), you only need to give the two intervals;
   - Provide the answer for 2(h).