STAT 420 Spring 2017
2GR, 2UG

Homework 2

(Due Friday, February 10, by 4:00 p.m.)

Please submit your assignment *on paper*, following the Guidelines for Homework Write-Ups and Submissions. Please include your name (with your last name underlined), and your NetID at the top of the first page. (Even if correct, answers might not receive credit if they are too difficult to read. No credit will be given without supporting work.)

1. Suppose a least-squares regression line is known to go through the origin, that is, the regression model is of the form

$$Y_i = \beta x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n, \quad \text{where } \varepsilon_i\text{'s are i.i.d. } N(0, \sigma^2),$$

and the associated equation of the regression line is $\hat{y} = \hat{\beta}\, x$. In this case, the least-square estimation for the parameter β reduces to finding the value $\hat{\beta}$ that minimizes the following function

$$f(\beta) = \sum_{i=1}^{n} [\, y_i - \beta x_i \,]^2 .$$

Use the derivative of f with respect to β to derive the formula for the slope of the least-squares regression line in this case.

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

Nancy claims that drinking beer has no effect on her running speed. The data on the right show how many seconds she took to run from her home to the grocery store where she usually goes for shopping, after consuming various amounts of beer, measured in ounces:

Beer Amount in ounces (x)	0	12	24	36	48	60	72
Running Time in seconds (y)	141	127	141	163	145	179	161

$$\sum x = 252, \quad \sum y = 1{,}057, \quad \sum x^2 = 13{,}104, \quad \sum y^2 = 161{,}447, \quad \sum xy = 40{,}068,$$

$$\sum (x - \bar{x})^2 = 4032, \quad \sum (y - \bar{y})^2 = 1{,}840, \quad \sum (x - \bar{x})(y - \bar{y}) = \sum (x - \bar{x})y = 2{,}016.$$

Consider the model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where ε_i's are i.i.d. $N(0, \sigma^2)$.

(a) Find the equation of the least-squares regression line.

(b) Calculate the residuals e_i. Does the sum of the residuals equal zero or not?

(c) Give an estimate for σ, the standard deviation of the observations about the true regression line.

(d) Calculate the multiple R-squared and the adjusted R-squared.

(e) How much time would you expect that Nancy would need to run from her home to this grocery store after consuming 40 and 160 ounces of beer, respectively? In which of these two predictions would you have more confidence?

(f) Use the F-test to test $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ (the significance of the regression test) at a 10% significance level. Report the value of the test statistic, critical value(s), and decision.

(g) Construct a 90% confidence interval for β_1. Does your answer for part (g) agree with your answer for part (f)?

(h) Nancy believes that when she is not drinking beer, it takes her on average 125 minutes to run from her home to this grocery store. What is the p-value (approximately) of the test $H_0 : \beta_0 = 125$ vs. $H_1 : \beta_0 \neq 125$? (You may give a range.)

(i) Construct a 90% confidence interval for σ^2.

(j) Construct 95% confidence interval for the average time Nancy needs to run from her home to this grocery store after consuming 136 ounces of beer.

(k) Construct 95% prediction interval for the time Nancy needs to run from her home to this grocery store after consuming 136 ounces of beer.