HW 4 (Due Oct 17, 2017) **Name:**

Problem 1. Suppose that we have postulated the model

$$Y_i = \beta_1 x_i + \epsilon_i, \quad , i = 1, 2, \ldots, n,$$

where the ϵ's are independent and identically distributed random variables with $E(\epsilon_i) = 0$ and $V(\epsilon_i) = 0$. $y = \beta_1 x$ describes a straight line passing through the origin. The model just described often is called the *no-intercept* model. **Find the least-square estimator of** β_1; i.e., the estimator is the minimizer of $\sum_{i=1}^{n}\{Y_i - \beta_1 x_i\}^2$ with respect to β_1.

Some data obtained on height x and diameter y of shells appear in the following table. If we fit the above model to the data, **compute the value of the least-square estimate of** β_1.

Specimen	Diameter (y)	Height (x)
OSU 36651	185	78
OSU 36652	194	65
OSU 36653	173	77
OSU 36654	200	76
OSU 36655	179	72
OSU 36656	213	76
OSU 36657	134	75
OSU 36658	191	77
OSU 36659	177	69
OSU 36660	199	65

Problem 2. Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

for $i = 1, 2, \ldots, n$, where $\epsilon_i \sim$ iid $N(0, \sigma^2)$. Let $\widehat{\beta}_0$ and $\widehat{\beta}_1$ be the least square estimators of β_0 and β_1, respectively. Denote $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$, $S_{yy} = \sum_{i=1}^n (Y_i - \bar{Y})^2$, and $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})$. We have

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{x}, \quad \widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}}.$$

1. Let $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$ be the fitted value of the ith response and $e_i = Y_i - \widehat{Y}_i$ be the ith residual. Show that

$$\sum_{i=1}^n (Y_i - \widehat{Y}_i) = 0.$$

2. Let $\mathrm{SSE} = \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2$. Prove that

$$\mathrm{SSE} = S_{yy} - \widehat{\beta}_1 S_{xy}.$$

   Further prove that $\mathrm{SSE} \leq S_{yy}$.

3. Prove that

$$e_i \sim N(0, \sigma^2(1 - m_{ii}))$$

   where

$$m_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}$$

   Note that this implies the residuals (unlike the errors) do not have constant variance. In an applied course, m_{ii} might be called the leverage associated with the ith observation. Leverages can be helpful in classifying observations as outliers or not.

Problem 3.

Using a chemical procedure called *differential pulse polarography,* a chemist measured the peak current generated (in microamperes, μA) when solutions containing different amounts of nickel (measured in parts per billion, ppb) are added to different portions of the same buffer.[8] Is there sufficient evidence to indicate that peak current increases as nickel concentrations increase? Use $\alpha = .05$.

x = Ni (ppb)	y = Peak Current (μA)
19.1	.095
38.2	.174
57.3	.256
76.2	.348
95	.429
114	.500
131	.580
150	.651
170	.722

In addition, complete the following parts:

1. Compute a 95 percent confidence interval for the mean peak current when the nickel concentration is $x^* = 100$ ppb. Interpret the interval.
2. Compute a 95 percent confidence interval for the peak current when the nickel concentration is $x^* = 100$ ppb. Interpret the interval.
3. Compare the two intervals. Which interval is wider? Why?