University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

Homework 2

EXERCISE 1
A new profit-sharing plan was introduced at an automobile parts manufacturing plant last year. Both management and union represenatives were interested in determining how a worker's years of experience influence his or her productivity gains. After the plan had been in effect for a while, the data shown below were collected:

Years of experience (x)	Number of units daily (y)
15.1	110
7.0	105
18.6	115
23.7	127
11.5	98
16.4	103
6.3	87
15.4	108
19.9	112

For your convenience:
$\sum_{i=1}^{9} y_i = 965$, $\sum_{i=1}^{9} x_i = 133.9$, $\sum_{i=1}^{9} y_i^2 = 104469$, $\sum_{i=1}^{9} x_i^2 = 2258.73$, $\sum_{i=1}^{9} x_i y_i = 14801.2$.

a. Use `R`: Construct a scatterplot of the number of units manufactured daily on the years of experience on the assebly line.

b. Find the least-squares regression line (perform all calculations by hand).

c. Use `R`: Verify your answer to part (b). Submit the `R` results.

d. Predict the number of units manufactured daily by an employee who has 10 years of experience on the assembly line.

EXERCISE 2
For the regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ show that the sum of the residuals is always equal to zero, i.e $\sum_{i=1}^{n} e_i = 0$.

EXERCISE 3
Suppose in the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where $i = 1, \cdots, n$, $E(\epsilon_i) = 0$, $\text{var}(\epsilon_i) = \sigma^2$ the measurements x_i were in inches and we would like to write the model in centimeters, say, z_i. If one inch is equal to c centimeters (c is known), we can write the above model as follows $y_i = \beta_0^* + \beta_1^* z_i + \epsilon_i$.

a. Suppose $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least squares estimates of β_0 and β_1 of the first model. Find the estimates of β_0^* and β_1^* in terms of $\hat{\beta}_0$ and $\hat{\beta}_1$.

b. Show that the value of R^2 remains the same for both models.

c. Find the variance of $\hat{\beta}_1^*$.

EXERCISE 4
Consider the regression model

$$y_i = (\beta_0 + \beta_1 \bar{x}) + \beta_1 (x_i - \bar{x}) + \epsilon_i$$

This model is called the *centered* version of the regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ that was discussed in class. If we let $\gamma_0 = \beta_0 + \beta_1 \bar{x}$ we can rewrite the *centered* version as $y_i = \gamma_0 + \beta_1 (x_i - \bar{x}) + \epsilon_i$. Find the least squares estimates of γ_0 and β_1.

EXERCISE 5
Consider the regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Show that $Cov(\bar{Y}, \hat{\beta}_1) = 0$ where $\bar{Y}$ is the sample mean of the y values, and $\hat{\beta}_1$ is the estimate of β_1.

EXERCISE 6
Consider the regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Find $cov(e_i, e_j)$.