University of California, Los Angeles
Department of Statistics

Statistics 13 **Instructor: Nicolas Christou**

Homework 2

EXERCISE 1

Consider the following data on `cadmium`, `copper`, `lead`, and `zinc` at 6 locations define by the coordinates `x` and `y`.

```
       x      y cadmium copper lead zinc
1 181072 333611    11.7     85  299 1022
2 181025 333558     8.6     81  277 1141
3 181165 333537     6.5     68  199  640
4 181298 333484     2.6     81  116  257
5 181307 333330     2.8     48  117  269
6 181390 333260     3.0     61  137  281
```

Compute the following by hand:

a. The standard deviation of `cadmium`.

b. The standard deviation of `lead`.

c. The estimates of β_0 and β_1 of the model

$$\texttt{cadmium}_i = \beta_0 + \beta_1 \texttt{lead}_i + \epsilon_i$$

d. The covariance between `cadmium` and `lead`.

e. The correlation coefficient between `cadmium` and `lead`.

EXERCISE 2

Consider the simple regression model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

a. Show that the sum of the residuals is always equal to zero:

$$\sum_{i=1}^{n} e_i = 0, \quad \text{where} \quad e_i = y_i - \hat{y}_i$$

b. Show that the estimate of β_1 can be computed also using:

$$\hat{\beta}_1 = r \frac{\texttt{sd(y)}}{\texttt{sd(x)}}$$

c. Use the result of part (b) to compute again $\hat{\beta}_1$ of exercise 1.

EXERCISE 3

Consider the simple regression model through the origin: $y_i = \beta_1 x_i + \epsilon_i$. Find the least squares estimate of β_1.

EXERCISE 4

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.

EXERCISE 5

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 6

Data have been collected for 19 observations of two variables, y and x, in order to run a regression of y on x. You are given that $s_y = 10$, $\sum_{i=1}^{19}(y_i - \hat{y}_i)^2 = 180$.

a. Compute the proportion of the variation in y that can be explained by x. [Ans. 0.90]

b. Compute the standard error of the estimate (s_e). [Ans. 3.25]

EXERCISE 7

Data on y and x were collected to run a regression of y on x. The intercept is included. You are given the following: $\bar{x} = 76, \bar{y} = 880, \sum_{i=1}^{n}(x_i - \bar{x})^2 = 6800, \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 14200, r_{xy} = 0.72, s_e = 20.13$.

a. What is the value of $\hat{\beta}_1$? [Ans. 2.088]

b. What is the value of $\hat{\beta}_0$? [Ans. 721.312]

c. What is the value of $\sum_{i=1}^{n}(y_i - \bar{y})^2$? [Ans. 57188]

d. What is the sample size n? [Ans. 70]

EXERCISE 8

Requires `R`. The two data sets (from *Mathematical Statistics and Data Analysis* by John Rice, Third Edition, Doxbury 2007) contain measurements on resistance to breathing in children with asthma (see asthma.txt data set) and in children with cystic fibrosis (see cystfibr.txt data set). In both data sets the predicor variable was height (cm). You can access the data set using the following command in `R`:

```
a1 <- read.table("http://www.stat.ucla.edu/~nchristo/statistics13/asthma.txt", sep=",", header=TRUE)
a2 <- read.table("http://www.stat.ucla.edu/~nchristo/statistics13/cystfibr.txt", sep=",", header=TRUE)
```

Is there a statistically significant relation between resistance to breathing and height in either group? Use the permutation test to test the hypothesis that the slope is equal to zero, or whether the correlation coefficient is equal to zero. You should also read the paper posted on the course website.