University of California, Los Angeles
Department of Statistics

Statistics 13 **Instructor: Nicolas Christou**

Exam 1
29 October 2014

Name: SOLUTIONS

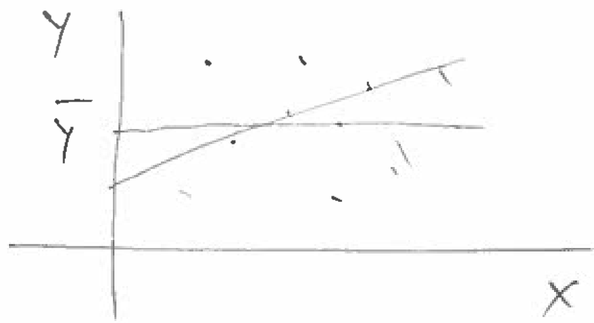
UCLA ID: ______________________

Section: ______

Problem 1 (20 points)
Answer the following questions:

a. Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Using data on y and x we find the least squares estimates $\hat\beta_0$ and $\hat\beta_1$ and fit the best line $\hat y_i = \hat\beta_0 + \hat\beta_1 x_i$. Is it possible that $var(y) < s_e^2$? Note: $var(y)$ is the variance of the y values and s_e^2 is the variability around the fitted line. If this is possible can you draw (approximately) the corresponding scatterplot of y on x?

YES.



VERY WEAK ASSOCIATION

b. Consider the simple regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. If $\beta_0 = 4$ what is the least squares estimate of β_1? Note: Use the procedure of least squares we discussed in class, but here the intercept is known!

$$Y_i = 4 + \beta_1 X_i + \varepsilon_i$$

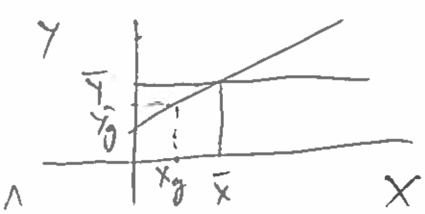
$$\text{MIN } S = \sum (Y_i - 4 - \beta_1 X_i)^2$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum (Y_i - 4 - \beta_1 X_i) X_i = 0$$

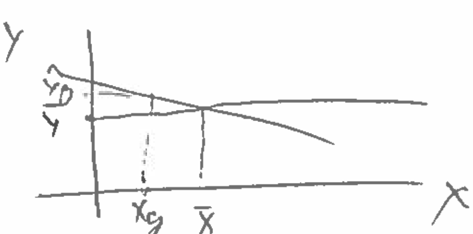
$$\sum X_i Y_i - 4 \sum X_i - \beta_1 \sum X_i^2 = 0 \Rightarrow \boxed{\hat\beta_1 = \frac{\sum X_i Y_i - 4 \sum X_i}{\sum X_i^2}}$$

c. Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. One of the goals of regression is to use the fitted regression line to make predictions for y when a new value of x is given. If a new value of x is below $\bar x$ is it true that the predicted value of y will be always below $\bar y$? Please explain your answer!

IT DEPENDS ON $\hat\beta_1$.



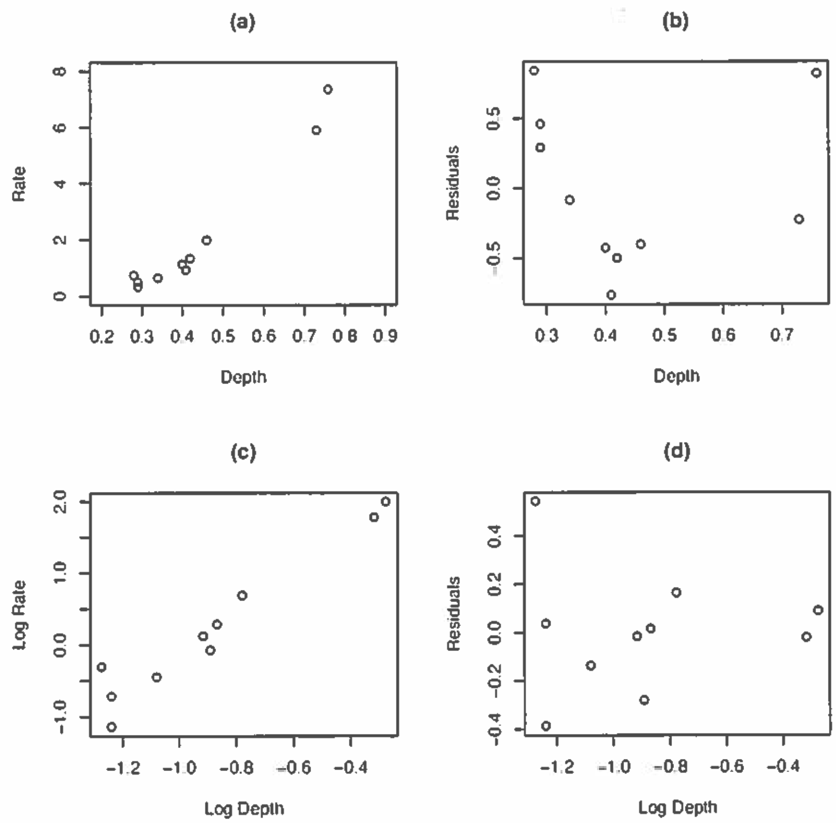
$\hat\beta_1 > 0 \Rightarrow X_g < \bar X \rightarrow \hat Y_g < \bar Y$



$\hat\beta_1 < 0 \Rightarrow X_g < \bar X \rightarrow \hat Y_g > \bar Y$

Problem 2 (20 points)
The data in the table below shows the depth of a stream and the rate of its flow.

Flow Rate	Depth
0.636	0.34
0.319	0.29
0.734	0.28
1.327	0.42
0.487	0.29
0.924	0.41
7.350	0.76
5.890	0.73
1.979	0.46
1.124	0.40



Plot (a) shows the scatterplot of `Rate` against `Depth`.
Plot (b) shows the scatterplot of `Residuals` against `Depth` from the simple regression of `Rate` on `Depth`.
Plot (c) shows the scatterplot of `Log(Rate)` against `Log(Depth)`.
Plot (d) shows the scatterplot of `Residuals` against `Log(Depth)` from the simple regression of `Log(Rate)` on `Log(Depth)`.

Answer the following questions:

a. Write the regression model equation based on plot (a).

$RATE = \beta_0 + \beta_1 DEPTH + \varepsilon_i$

b. Write the fitted line equation based on plot (c).

$\widehat{\log(RATE)} = \hat{\beta}_0 + \hat{\beta}_1 \log(DEPTH)$

c. Explain why it was necessary to transform the data using logarithms.

NON-LINEAR, ~~IT~~ TAKE LOGS TO LINEARIZE RELATIONSHIP.
IT SHOW ALSO ON PLOT (b)

Problem 3 (20 points)

Part A:

Data on cadmium (x) and cobalt (y) of soil at a certain area gave the following information:

$\sum_{i=1}^{6} x_i = 9.545$,

$\sum_{i=1}^{6} y_i = 61.668$,

$\sum_{i=1}^{6} x_i^2 = 15.78468$,

$\sum_{i=1}^{6} y_i^2 = 719.9573$,

$\sum_{i=1}^{6} x_i y_i = 96.43722$.

It was discovered that the last two x values were interchanged. The incorrect table is shown below:

x	y
.	.
.	.
.	.
.	.
1.145	16.320
1.565	3.508

$\Sigma x_i,\ \Sigma x_i^2,\ \Sigma y_i,\ \Sigma y_i^2$ DO NOT CHANGE

NEW $\Sigma x_i y_i = 96.43722 - (1.145)(16.320) - (1.565)(3.508) + (1.565)(16.320) + (1.145)(3.508) = 101.8183$

Find the corrected $\hat{\beta}_1$ from the regression of y on x.

$$\hat{\beta}_1 = \frac{\Sigma x_i y_i - \frac{1}{n}(\Sigma x_i)(\Sigma y_i)}{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}} = \frac{101.8183 - \frac{1}{6}(9.545)(61.668)}{15.78468 - \frac{(9.545)^2}{6}} \Rightarrow \underline{\hat{\beta}_1 = 6.1895}$$

Part B:

Data on cadmium (x) and cobalt (y) of soil at a certain area gave the following information:

$\sum_{i=1}^{6} x_i = 11.908$, NEW $\Sigma x_i = 11.908 - 3.508 + 1.145 = 9.545$

$\sum_{i=1}^{6} y_i = 59.305$, NEW $\Sigma y_i = 59.305 - 1.145 + 3.508 = 61.668$

$\sum_{i=1}^{6} x_i^2 = 26.77971$, NEW $\Sigma x_i^2 = 26.77971 - 3.508^2 + 1.145^2 = 15.78468$

$\sum_{i=1}^{6} y_i^2 = 708.9622$, NEW $\Sigma y_i^2 = 708.9622 - 1.145^2 + 3.508^2 = 719.9572$

$\sum_{i=1}^{6} x_i y_i = 101.8183$.

It was discovered that a pair of x and y was interchanged. The incorrect table is shown below:

x	y
.	.
.	.
.	.
.	.
.	.
3.508	1.145

$\Sigma x_i y_i$ WILL NOT CHANGE!

Find the corrected $\hat{\beta}_1$ from the regression of y on x.

$$\hat{\beta}_1 = \frac{101.8183 - \frac{1}{6}(9.545)(61.668)}{15.78468 - \frac{(9.545)^2}{6}} \Rightarrow \underline{\hat{\beta}_1 = 6.1895}$$

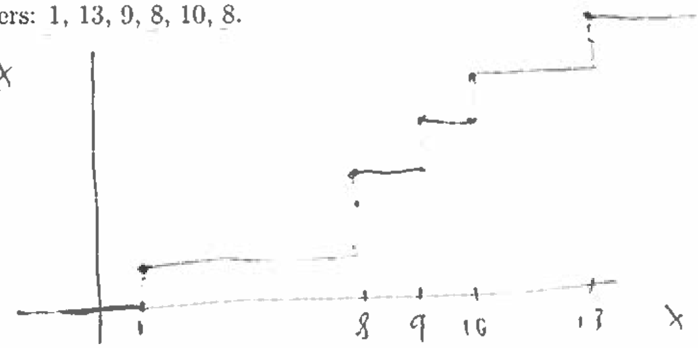
SAME DATA WERE USED FOR BOTH QUESTIONS!

$\frac{\#\text{ of obs}}{6}$

Problem 4 (20 points)
Answer the following questions:

a. Plot the ecdf of the following numbers: 1, 13, 9, 8, 10, 8.

x	ecdf
1	1/6
8	3/6
8	
9	4/6
10	5/6
13	6/6



b. Consider the body fat data (see handout "Introduction to R"). We use here the variables y, x_6, x_9, x_{10}. The variance covariance matrix and the sample means of the four variables are shown next:

```
#Variance covariance matrix:
         y     x6     x9    x10
y   70.036  9.980 37.483 24.587
x6   9.980  5.909 12.799  8.879
x9  37.483 12.799 51.324 33.715
x10 24.587  8.879 33.715 27.562

#Sample means:
     y     x6     x9    x10
19.151 37.992 99.905 59.406
```

Find the fitted line of the regression of y on x_9. Show all your work.

$$\hat\beta_1 = \frac{Cov(Y, X_9)}{VAR(X_9)} = \frac{37.483}{51.324} \Rightarrow \hat\beta_1 = 0.73$$

$$\hat\beta_0 = \bar y - \hat\beta_1 \bar x_9 = 19.151 - 0.73(99.905) \Rightarrow \hat\beta_0 = -53.78$$

$$\hat Y_i = -53.78 + 0.73 X_9$$

c. Refer to part (b). Compute the correlation coefficient between y and x_{10}.

$$Cor(Y, X_{10}) = \frac{Cov(Y, X_{10})}{SD(Y)\,SD(X_{10})} = \frac{24.587}{\sqrt{70.036}\sqrt{27.562}} = 0.5596$$

d. Refer to part (b). Consider the regression of y on x_6. Compute the value of the regression sum of squares.

$$SSR = \hat\beta_1^2 \sum (x_i - \bar x_6)^2 = 1.69^2(1477.25) \Rightarrow SSR = 4219.174$$

$$\hat\beta_1 = \frac{cov(Y, X_6)}{VAR(X_6)} = \frac{9.980}{5.909} = 1.69$$

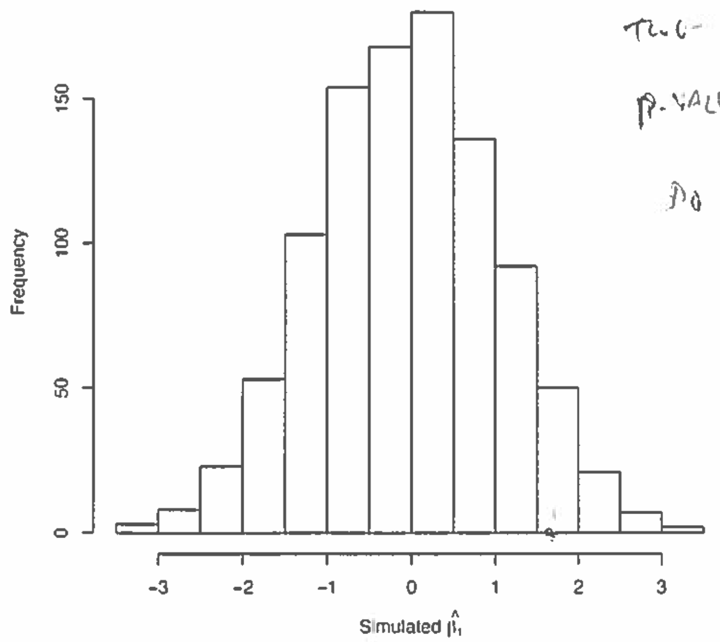
$$VAR(X_6) = \frac{\sum (x_i - \bar x_6)^2}{n-1}$$

$$\sum (x_i - \bar x_6)^2 = (n-1)5.909 = (250)(5.909) = 1477.25$$

$$n = 251$$

Problem 5 (20 points)
Normal temperature (degrees Fahrenheit) readings (x) and heart rates (beats per minute) of 65 males and 65 females were used in this problem.

a. Consider the data for males: The OLS estimate of β_1 is $\hat{\beta}_1 = 1.65$. A permutation test was run to test the hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_a : \beta_1 \neq 0$ of the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Under the null hypothesis the permutation test produced the following histogram. What is your conclusion?

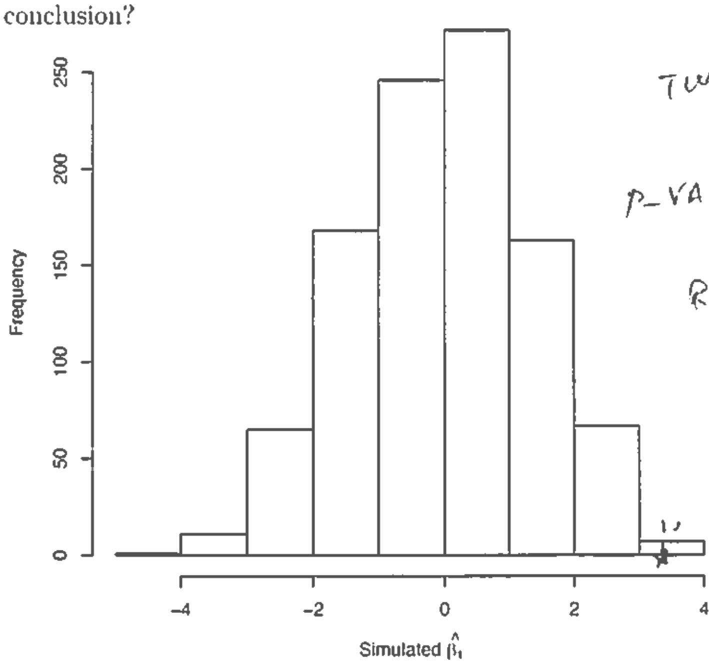


TWO-SIDED TEST

P-VALUE $\approx \frac{2 \times 50}{1000} = 0.10$

DO NOT REJECT Ho.

a. Consider the data for females: The OLS estimate of β_1 is $\hat{\beta}_1 = 3.13$. A permutation test was run to test the hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_a : \beta_1 \neq 0$ of the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Under the null hypothesis the permutation test produced the following histogram. What is your conclusion?



TWO-SIDED TEST

P-VALUE $\approx \frac{2 \times 10}{1000} = 0.02$

REJECT Ho.