Statistics and Bioinformatics
Problem Set 13
Due in class Tuesday, January 18, 2011

1) Calculate the predicted value of y in the following linear regressions:
   a) intercept = -2, slope = -3, for x = -1.

$y = -2 - 3(-1) = -2 + 3 = 1$

   b) slope = 5, intercept = 2, for x = 5.

$y = 2 + 5 \cdot 5 = 12$

   c) a = -4, b = 20, for x = 2.

$y = -4 + 20 \cdot 2 = 36$

   d) y = 3 - 8x, for x = 1.

$y = 3 - 8 = -5$

2) In a simple linear regression, the slope was found to be 5 and the intercept to be 1. If *x* is 3, then what is the predicted value of *y*?

$y = 1 + 5 \cdot 3 = 16$

3) Given values for the x and y variables, what command in R displays the slope and intercept of the best-fit line?

`lm(y~x)`

4) For particular vectors x and y, the R command lm(y~x) gives:

```
Call:

lm(formula = y ~ x)

Coefficients:

(Intercept)            x
    1.58025      0.08642
```

a) What kind of analysis was done?

(simple) linear regression

b) What is the intercept of the best-fit line relating x and y?

$1.58025 = a$

c) What is the slope of the best-fit line relating x and y?

$0.08642 = b$

d) If x = 2, then what is the predicted value of y?

$y = 1.58025 + 0.08642 \cdot 2 = 1.753$

5) A simple linear regression gives the following ANOVA table:

```
Analysis of Variance Table

Response: y

           Df     Sum Sq    Mean Sq F value    Pr(>F)
x           1 160511350 160511350  222.71 1.179e-07 ***
Residuals   9   6486356    720706
```

a) How much of the variation in y is explained by variation in x?

$R^2 = \frac{160511350}{160511350 + 6486356} = 0.96$

b) What is the coefficient of determination?

$R^2 = 0.96$

c) What is the correlation coefficient?

$r = \sqrt{R^2} = \sqrt{0.96} = 0.98$

d) What is the total number of points?

$n = df_{res} + df_{line} + 1 = 9 + 1 + 1 = 11$

6) For the following correlation coefficients, write down the strength of the correlation according to the table on page 109 of the Vasilj textbook.

a) 0.4 slaba
b) -0.4 slaba
c) -.7 jaka
d) 0 nema
e) -0.9 potpuna

7) For each of the following pairs of correlation coefficients, circle the one that represents the stronger relationship. If they are equally strong, circle both.
   a) 0.4 0.5
   b) -0.5 0
   c) 0.2 0.1
   d) -0.7 0.7

8) In a study of the relationship between length and mass, it is hypothesized that $M = aL^b$.
   a) What is the shape of this relationship if $b > 1$? Draw the graph.

   b) What transformation of M and L is necessary to linearize the relationship?

log (or ln)

   c) If a linear regression is performed on the transformed data, what is the intercept?

intercept = log (a)

   d) If a linear regression is performed on the transformed data, what is the slope?

slope = b

   e) In a particular study, a is found to be 2 and b is found to be 1.5. If length is 3, then what is the predicted value of mass?

$M = 2 \cdot 3^{1.5} = 10.39$

Do the following problem in a Word file that you email to tosaric@unizd.hr.

9) Length of 11 individuals is measured to be 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, and 20. Mass of these 11 individuals is measured to be 1447, 2558, 3096, 3401, 4810, 5870, 6842, 8335, 9910, 12090, and 13926. Assume the true relationship is $M = aL^b$.
   a) Using a transformation and the lm() function in R, find the best-fit curve for these data. Paste the results of lm() into your word file. What are your estimates of a and b?
   b) Using the raw, untransformed data, use nls() in R to estimate a and b. Paste the results of nls() in your Word file. What are your estimates of a and b?
   c) Plot a scatter diagram of L and M. Superimpose on this diagram the best-fit curve using curve(add=T) in R. Include this figure in your Word file.
   d) If an individual is observed to be 14.5 in length, what is its predicted mass?