11. Linear regression

Task 1. A biology student wishes to determine the relationship between temperature and heart rate in leopard frog, *Rana pipiens*. He manipulates the temperature in 2° increment ranging from 2 to 18°C and records the heart rate at each interval. His data are presented in table rana.xls http://edu.sablab.net/data/xls/rana.xls

1) Build the model and provide the p-value for linear dependency

2) Provide interval estimation for the slope of the dependency

3) Estimate 95% prediction interval for heart rate at 15°

Task 2. The height and arm span of 10 adults males were measured (***span.xls***). Is there a correlation between these two measurements? Carry out an appropriate analysis.

1) Determine correlation and its confidence intervals

2) Perform linear regression analysis

Task 3. Data are shown in the Table (http://edu.sablab.net/data/xls/leukemia.xls) for two groups of patients who died of acute myelogenous leukemia. Patients were classified into the two groups according to the presence or absence of a morphologic characteristic of white cells. Patients termed AG positive were identified by the presence of Auer rods and/or significant granulature of the leukemic cells in the bone marrow at diagnosis. For AG-negative patients, these factors were absent. Leukemia is a cancer characterized by an overproliferation of white blood cells; the higher the white blood count (WBC), the more severe the disease. Separately for each morphologic group, AG positive and AG negative:

1. Draw a scatter diagram with a regression line to show a possible association between the log survival time (take the log yourself and use as the dependent variable) and the log WBC (take the log yourself).
2. Build linear regression and check if a linear model is justified. Check the coefficient of determination and provide your interpretation. Is there the same effect of WBC for 2 groups?
3. What is the survival time for a patient with 20,000 WBC? Are estimates for different groups different or the same?

Task 4. Dataset **ActinGenes** contains log-ratios of expressions of actin-related genes. Log-ratio means the log of ratio expression in a sample and in universal human RNA solution. Samples are coming from cancer and healthy tissues of several organs (brain, breast, colon, liver, ovary, and uterus). The dataset can be downloaded from http://edu.sablab.net/data/xls/actingenes.xls

1. Build a linear regression model for expressions (here it means – log ratios) of two genes: transcription factor 3 (TCF3) and myosin IX A (MYO9A). Is the relation between these genes statistically significant? Provide an equation, describing the model and a numerical proof of your conclusion.
2. Now repeat the analysis for normal and cancer tissues. Draw the conclusions.