L3. Linear Models

http://edu.modas.lu/data or R

1. Four-o'clocks *Mirabilis jalapa*, are plants native to tropical America. Individual plants can have red, white, or pink flowers. Flower color in this species is thought to be controlled by a single gene locus with two alleles expressing incomplete dominance, so that heterozygotes are pink, while homozygotes are white or red. A horticulturist self-pollinates several pink-flowered plants and produces 240 progeny with 55 that are red-flowered, 132 that are pink-flowered and 53 that are white-flowered. **Are these data consistent with Mendelian ratios (1:2:1)?**

2. A study of kidney damage during organ retrieval for transplantation was conducted in the UK in 1992-1996. In many cases of organ donation, when the

Team	Damaged kidneys	Undamaged kidneys	Total
Renal retrieval	454	1692	2146
Liver retrieval	415	2054	2469
Total	869	3746	4615

kidneys are retrieved the liver is retrieved as well, in a single surgical procedure. When both types of organs were retrieved, the researchers categorized the surgical team, based on the operating surgeon's specialty, as either renal retrieval or liver retrieval team. **Was the reported kidney damage rate independent of the type of surgical team?**

3. R internal data "**chickwts**" contains the results of an experiment, conducted to measure and compare the effectiveness of various feed supplements on the growth rate of chickens. Check whether feeding influences the growth rate. **Define the most efficient feeding. Perform post hoc analysis to define significant differences.**

4. Data "**ToothGrowth**" contains the result of an experiment, conducted to measure the effect of various doses of vitamin C on the tooth growth (model animals – Guinea pigs). Vitamin C (VC) and orange juice (OJ) were given to animals in different quantities. **Compare the effects due to treatment and quantity. Provide a proper ANOVA model, post hoc analysis, and interpret the results.**

5. 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 **rana**. **(1) Build the model and provide the p-value for the linear dependency and interval estimate of its slope. (2) Estimate 95% prediction interval for heart rate at 15°**

6. The height and arm span of 10 adult males were measured (data **span**). **Is there a dependency between these two measurements? Carry out an appropriate analysis.**

7. Dataset **kidweights** contains a sample from the data used to form children's CDC Growth Charts (http://www.cdc.gov/growthcharts). Compare 2 regression models:

$$weight = b_1 \times height + b_0 \quad \text{and} \quad weight = b_1 \times height^2 + b_0 .$$

Which model gives a better prediction for the expected weight? Consider gender effect (use 2 models or gender as a factor). What is the expected weight for a boy of 33 inches?

8. Data are shown in the **leukemia** for two groups of patients who died of acute myelogenous leukemia. Patients were classified into 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. Analyze the data using a linear model (either 2 models for AP-pos and AG-neg or a single model with a factor). What is the survival time for a patient with 20,000 WBC?