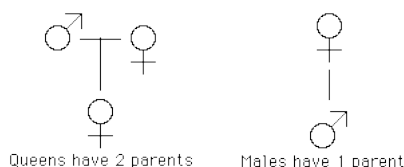
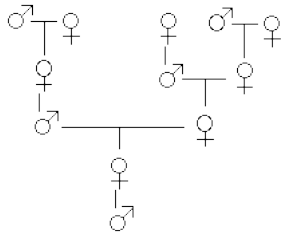


- (13 points) 1. Does the germination rate of *Primula* increase with increasing nutrient supply? To investigate this question the percentage of sprouted *Primula* was measured in 10 homogeneous pots provided with different defined nutrient levels. The nutrient concentrations involved were restricted to 10 different levels.
- (2 points) (a) Read the data (file: *primula.csv*), describe variables (in terms of dependent/independent, scale and effect type) and values (using statistical parameters) and visualize. If you observe problems with the data, discuss them.
- (1 point) (b) Given the variable types and effects, what is the suitable parameter to describe the correlation between nutrient level and germination ratio? What is an appropriate regression model?
- (1 point) (c) Compute and interpret the chosen parameter for description of correlation.
- (1 point) (d) Compute the regression and give the formula for the regression line.
- (3 points) (e) Visualize the regression line and the residuals. Perform residual analysis and discuss any problems you observe.
- (1 point) (f) Test whether the regression gradient is unequal zero on significance level  $\alpha = 0.01$ . Interpret the result.
- (1 point) (g) Predict the sprout ratio after application of nutrient level 7 and give the 99% CI for the predicted value.
- (2 points) (h) Visualize the dataset and the 99% confidence band of the regression as well as the 99% prediction band. Discuss how to interpret both bands. Discuss the dataset in this context.
- (1 point) (i) Answer the following questions: Does the dataset show that increasing nutrient supply increases the germination rate? Is there evidence for a linear relationship between nutrient level and germination rate?
- (4 points) 2. The influence of precipitation [rainfall measured in  $L/100m^2$ ] on the height of sunflowers [m] is to be examined. At 60 different locations precipitation and mean height of sunflowers has been measured.
- (1 point) (a) Read the data (file: *plantheight.csv*), describe variables (in terms of dependent/independent, scale and effect type) and values (using statistical parameters) and visualize. If you observe problems with the data, discuss them.
- (1 point) (b) Given the variable types and effects, what is the suitable parameter to describe the correlation between precipitation and sunflower height? What is an appropriate regression model?
- (1 point) (c) Compute and interpret the chosen parameter for description of correlation.
- (1 point) (d) Perform a statistical test to show that the computed parameter for description of correlation is significantly different from zero at level of significance 0.05. Also write down the appropriate hypothesis pair.
- (3 points) 3. In honeybees only females (queens and workers) have two parents. Male drones hatch from unfertilized eggs and thus only have a mother.



A look at a drone's family tree reveals it has 1 parent, 2 grand-parents, 3 great-grand-parents, 5 great-great-grand-parents and so forth.



So for every drone in a certain generation there was one queen in the previous one. And for every queen in a certain generation there was one queen and one drone in the previous generation.

Or in other words: The number of queens in an ancestral generation is the sum of drones and queens in the generation after it and the number of drones in an ancestral generation is equal to the number of queens in the generation that follows it.

Use a 'for' loop to find out how many bees there are in the 16th ancestral generation of a single drone. Assume that there are no overlaps in the ancestry of this drone and make sure to submit the R code you used to get to your answer.

This exercise has 3 questions, for a total of 20 points.