

Practicals in Quantitative Genetic Analyses

Rasmus Bak Stephansen & Peter Sørensen

2022-02-25

Introduction

In these practicals we will be analysing quantitative traits observed in a mice population. The mouse data consist of phenotypes for traits related to growth and obesity (e.g. body weight, glucose levels in blood), pedigree information, and genetic marker data. The practicals will be a mix of theoretical and practical exercises in R that are used for illustrating/applying the theory presented in the lectures and corresponding notes.

- Practical 1: Use R for Analysing Quantitative Traits
- Practical 2: Basic Quantitative Genetics illustrated in the mouse data
- Practical 3: Estimation of Genetic Parameters for traits in the mouse data
- Practical 4: Estimation of Breeding Values for traits in the mouse data
- Practical 5: Estimation of Genomic Breeding Values for traits in the mouse data

Mouse data

The **M16 mouse** was established as an outbred animal model of early onset polygenic obesity and diabetes. This was done by selection for 3- to 6-week weight gain for 27 generations from an outbred ICR base population. Breeding criterion was within-litter selection for the male and female with the largest weight gain from 3 to 6 weeks of age. An ICR control line was maintained in parallel, with random mating from generation to generation but maintaining a similar effective population size. Mice from the M16 line are larger at all ages and have increased body fat percentage, fat cell size, fat cell numbers, and organ weights when compared with ICR. Mice from the M16 line are larger at all ages and have increased body fat percentage, fat cell size, fat cell numbers, and organ weights when compared with ICR. These mice also exhibit hyperphagia, accompanied by moderate obesity, and are hyperglycemic, hyperinsulinemic, and hypercholesterolemic.

The **ICR mouse** is a strain of albino mice originating in SWISS and selected by Dr. Hauschka to create a fertile mouse line. Because mice of this strain have been sent to various places from the Institute of Cancer Research in the USA, the strain was named ICR after the initial letters of the institute. Mice of this strain are relatively large albinos with a gentle nature that grow well. The ICR mouse is a general-purpose model used for studies in a wide range of fields including toxicity, pharmacology, drug efficacy, and immunology.

A large **F2 population** (n=1181) was **established by crossing the M16 and ICR lines** (for a recent description of relevant phenotypes in the parental lines, see <https://onlinelibrary.wiley.com/doi/epdf/10.1038/oby.2004.176>). Twelve F1 families resulted from six pair matings of M16 males x ICR females and six pair matings of the reciprocal cross. A total of 55 F1 dams were mated to 11 F1 sires in sets of five F1 full sisters mated to the same F1 sire. These same specific matings were repeated in three consecutive replicates. Thus, the F2 population consisted of 55 full-sib families of up to 24 individuals each and 11 sire families of up to 120 individuals each. Actual numbers of mice within families varied slightly due to a small number of failed pregnancies. All litters were standardized at birth to eight pups, with approximately equal representation of males and females.

More information about the mouse data can be found in the following publications:

<https://onlinelibrary.wiley.com/doi/epdf/10.1038/oby.2004.176>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1449794/>

Practical 1: Use R for Analysing Quantitative Traits

Time schedule of practical session 1:

Time	Activity
11:15	Questions to lecture and multiple-choice questions
11:25	Introduction to R-studio
11:45	Assignments to groups – work with exercises
12:00	Break
12:35	Prepare final words of exercises in each group
12:45	Present final words
12:55	Repeat multiple-choice questions
13:00	End of practical session 1

Introduction:

In this practical we use R for explorative data analyses of two quantitative traits, body weight and blood glucose levels, observed in the F2 mouse population. These explorative data analyses includes computation of basic descriptive statistics such as mean, and variance used to describe each of these traits. Distribution plots (e.g., histogram) will be used to visualize whether the trait phenotypes follow a normal distribution. Boxplots will be used to spot potential effects of explanatory variables. Furthermore relationships between traits and variables will be characterized in terms of correlations and linear relationships.

Let's get started to explore our mouse data

One of the first thing to do is to explore the data used in the analysis. The goal is to understand the variables, how many records the data set contains, how many missing values, what is the variable structure, what are the variable relationships and more. Several commands/functions will be used. To read more about a specific function (e.g., 'str') write '?str'.

The mouse data set can be loaded using the following command:

```
mouse <- readRDS(url("https://github.com/psoerensen/bgcourse/raw/main/data/mouse.rds"))
```

Question 1: How many observations and which variables do we have in the data set? To get a fast overview of the data set you are working with you can use the 'str' function:

Answer:

The two quantitative traits we will be analysing are glucose levels in the blood (Gl) and body weight (BW) measured in the mice at 8 weeks of age. A more detailed view of the two quantitative traits in the 'data.frame' is provided by the 'summary' function:

```
summary(mouse[,5:6])
```

```
##           Gl           BW
## Min.      : 65.0   Min.    :23.04
## 1st Qu.:121.0   1st Qu.:34.06
## Median :139.0   Median :38.32
## Mean     :144.2   Mean     :38.72
## 3rd Qu.:164.0   3rd Qu.:43.40
## Max.     :292.0   Max.     :60.28
```

Question 2: What is the mean and variance of body weight and blood glucose levels? Use the ‘mean’ and ‘var’ functions to compute the mean and variance two traits:

Answer:

Question 3: How are the phenotypes of weight and glucose distributed? Use the ‘histogram’ and ‘boxplot’ functions to visualize the distribution the two traits:

Answer:

Question 4: Are the phenotypes of weight and glucose normally distributed? Use the ‘qqnorm’ function to create a quantile-quantile (QQ) plot of the trait values. Use the qqline function to add a line to a “theoretical,” by default normal, quantile-quantile plot:

Answer:

Question 5: Is there a relationship between the phenotypes of weight and glucose? Make a scatter plot of the the 2 traits using the ‘plot’ function. Compute the correlation using the ‘cor’ function and perform a statistical test to assess the significance of correlation between values of weight and glucose using the cor.test function:

Answer:

Let us explore the family structure. Use the ‘table’ function to determine the family size for sires and dams:

Question 6: What are the min and max family size? Use the 'table' and 'min' or 'max' functions to determine the min/max family size for sires and dams:

Answer:

Question 7: Does family influence the traits? Use the 'boxplot' function to visualize the potential effect of family on the two traits:

Answer:

Question 8: How many males and females?

Answer:

Question 9: Does gender influence the traits? Use the 'boxplot' function to visualize the potential effect of gender on the two traits:

Answer:

Question 10: How many observations in each replicate?

Answer:

Question 11: Does replicate influence the phenotype of weight and glucose? Use the 'boxplot' function to visualize the potential effect of replicate on the two traits:

Answer:

The exploratory data analysis is the process of analyzing and visualizing the data to get a better understanding of the data. It is not a formal statistical test.

Which factors should we include in the statistical model? To best answer these question we can fit a linear model that include these factors (sire, dam, sex, reps) in the model. This can be done using the 'lm' function:

```
fit <- lm(BW~sire+dam+sex+reps, data=mouse)
```

To test the effect of the variables in the model use the 'anova' function on the 'fit' object from the 'lm' function:

```
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: BW
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## sire       10  1536.6   153.7    9.2514 7.705e-15 ***
## dam        44  2020.9    45.9    2.7652 1.238e-08 ***
## sex         1 20637.7 20637.7 1242.5000 < 2.2e-16 ***
## reps        2  1723.6   861.8   51.8858 < 2.2e-16 ***
## Residuals 1119 18586.4    16.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 12: Do genetic factors influence the traits? Look at the output of the anova function.

Answer:

Some useful links explaining how to use R for basic concepts in statistics:

<http://www.r-tutor.com/elementary-statistics>

<http://www.r-tutor.com/elementary-statistics/probability-distributions/normal-distribution>

<http://www.r-tutor.com/elementary-statistics/numerical-measures>

<http://www.r-tutor.com/elementary-statistics/numerical-measures/mean>

<http://www.r-tutor.com/elementary-statistics/numerical-measures/variance>

<http://www.r-tutor.com/elementary-statistics/numerical-measures/standard-deviation>

<http://www.r-tutor.com/elementary-statistics/numerical-measures/covariance>

<http://www.r-tutor.com/elementary-statistics/numerical-measures/correlation-coefficient>

<http://www.r-tutor.com/elementary-statistics/simple-linear-regression>

<http://www.r-tutor.com/elementary-statistics/multiple-linear-regression>

<http://www.r-tutor.com/elementary-statistics/analysis-variance>

<https://antoinesoetewey.shinyapps.io/statistics-202/>