

## Assignment 2, ST2304

**Problem 1** From quantitative genetic theory it is known that the total phenotypic variance  $V_P$  in a given trait can be decomposed into the so called additive genetic variance  $V_A$  and the environmental variance  $V_E$ . The heritability of a trait  $h^2$  is then defined as the fraction

$$h^2 = \frac{V_A}{V_A + V_E} = \frac{V_A}{V_P}, \quad (1)$$

and is thus a number between 0 and 1. It is known that the slope in the regression of offspring on midparental value can be used as an estimate of this heritability.

Read the dataset `heights` into R using the command

```
heights <- read.table("https://www.math.ntnu.no/~jarlet/statmod/heights.dat")
attach(heights)
```

The dataset contain the height of former NTNU-students (the variable `offspring`), the height of their parents (the variables `mother` and `father`) and the sex of each student (the variable `sex`) encoded as values of 0 (female students) and 1 (male students).

1. Make scatter plots of all variable combinations using the `pairs` function.
2. Before estimating the heritability of height we will focus on sex differences. Estimate the difference in height between the sexes by fitting a linear regression model with `sex` as the only explanatory variable. What is the interpretation of the regression coefficient for `sex`. Read the output from `summary` to determine if this regression coefficient is significantly different from zero. Compare the result to the output from a two-sample  $t$ -test of difference between the sexes using the command

```
t.test(offspring[sex==0], offspring[sex==1], var.equal=TRUE)
```

Does the different or equal results of the two tests make sense?

3. Compute the midparental values, that is, the mean height of the parents of each student and add this as a second explanatory variable in the above linear regression model. What is the estimate of the heritability of height? Make a scatter plot of the heights of the students versus their midparental values. Use separate symbols for each sex by using the argument `pch=sex` (see the help page of `?points`). Add lines representing the estimated expected response as function of midparental value for each sex using `abline`.
4. Does including midparental value in the regression change the estimated sex difference in height? What happens to the standard error of the estimated sex difference?
5. Does the estimate of heritability (the regression coefficient for midparental value) change if you remove `sex` as an explanatory variable from the regression model?

**Problem 2** A spiral can be described as a parametric curve specified by the functions

$$\begin{aligned} x(t) &= at \cos(t) \\ y(t) &= at \sin(t) \end{aligned} \quad (2)$$

where the “parameter”  $t$  goes from zero to say,  $10 \cdot 2\pi$ .

Make a plot of this parametric curve in R by computing the necessary coordinates sufficiently densely spaced along the spiral. See `?plot`.

Phyllotaxis is a botanical phenomena by which for example each new seed appearing in the center of a growing sunflower moves away from the center at an angle increasing by  $\pi(3 - \sqrt{5})$  radians (or 137 degrees), the so called golden angle, for each new successive seed appearing. For a sunflower consisting of  $n$  seeds  $i = 1, 2, \dots, n$ , the angular direction of the  $i$ 'th seed from the center is thus

$$\theta(i) = \pi(3 - \sqrt{5})i. \quad (3)$$

Uniform spacing of the seeds implies that the  $i$ 'th seed in the sequence must be located a distance

$$r(i) = a\sqrt{i} \quad (4)$$

from the center, where  $a$  is a proportionality constant. Make a plot of a sunflower consisting of  $n = 1000$  seeds in R.