Design and Analysis of Experiments

- Lecture 1
Practical information

- Teachers: Jakob G. Rasmussen and Søren L. Buhl
- Textbook:
  Douglas C. Montgomery
- Evaluation:
  - At least 80% attendance (i.e. 8 lectures)
  - Complete 2 handin exercises
- Participant-list
- Computers & chords
- R
How not to do statistics – example 1

• This commercial claims that when you kiss a frog it will only very rarely turn into a prince, but you will always have succes with a Jägermeister
• Or does it… ?
How not to do statistics – example 2

Figure 2. Winning times for the Boston Marathon for men (1897 through 1992) and women (1972 through 1992) augmented with the best linear fit. Data from The 1993 Information Please Sports Almanac (p. 619–620).
How not to do statistics – example 3

- Do storks really bring babies?

![Graph showing a linear relationship between the number of storks and yearly births in Denmark.](image-url)
How not to do statistics – example 3

Graphical model:

Storks

Economy (drainage of wet areas)

Births

Economy (more women at work)

Storks

Graphical model:

Births

Economy
Introduction to R

- R demonstration, part 1 – no statistics so far…
- Exercises 1 + 2
Experiments

What is an experiment?
- Designed
- Observation, retrospective
- Observation, prospective
Experiments

1. Recognizing the problem
2. Selecting the response variable
3. Choice of factors, levels, ranges
4. Experimental design
5. Performing the experiment
6. Statistical analysis
7. Conclusions

We will focus on 6. in this course, but this has many implications for the other items.
Experiments

- Use your knowledge of the problem
- Keep it simple
- Practical and statistical significance are not the same
- Iterative experiments

Chapter 1 contains lots of other useful information, but now we turn to random variables.
Discrete random variables

- Discrete random variables take values in a countable set
- Examples of experiments leading to discrete random variables
  - Roll a die or flip a coin
  - Number of visits on a webpage
  - Number of failures in a production
- Examples of discrete distributions
  - Binomial distribution
  - Poisson distribution
- Discrete distributions are important, but not the focus of this course!
Continuous random variables

Continuous random variables take values in (an interval of) the real numbers

Examples of experiments leading to continuous random variables
- Height of a person in this room
- The amount of sugar in a patient's blood
- The time until the next accident in an airport

Examples of continuous distributions
- Normal (or Gaussian) distribution
- t-distribution
- $X^2$ (Chi-squared) distribution
- F-distribution
Normal distribution

- Density function for the normal distribution $X \sim N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- $\mu$ is the mean, and $\sigma^2$ is the variance

---

Normal Distribution: $\mu = 40$, $\sigma = 1$
Standard normal distribution

- $N(0,1)$ is called the standard normal distribution
- Standardization: if $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$
Central limit theorem

- Data can often be assumed to be normally distributed - however, when this does not hold, the central limit theorem is useful.

- Central limit theorem:
  If $X_1, \ldots, X_n$ are independent and identically distributed with (finite) mean $\mu$ and (finite) standard deviation $\sigma$, then

$$ \lim_{n \to \infty} \sum_{i=1}^{n} \frac{X_i - \mu}{\sqrt{n} \sigma} \sim N(0,1) $$

- Usually $X_1 + \ldots + X_n$ is approximately normally distributed for finite sums if $n$ is sufficiently big – as a rule of thumb $n > 30$

$$ \sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right) $$
R demonstration, part 2
Exercise 3