Methods 2 – Portfolio Assignment 1 – study group 14

Maja Gade Mortensen, Nanna Klitgaard Damgaard, Sára Fernezelyi

- *Type:* Group assignment
- *Due:* 10 March 2024, 23:59

In the following exercises, you will be asked to generate and summarize simulations from statistical models. You should use what you have learned so far (i.e. for loop, if else statements, sampling from continuous and discrete distributions. . .) to generate observations and summarize your samples using (one of) the appropriate methods. You can find examples of how to do that in Ch. 5. Note that here we will only focus on generative models, several aspects for inference and hypothesis testing discussed in Ch. 4 are not directly needed in this context.

In the first exercise, we will assume that the population of interest has a proportion of 0.51 men and 0.49 women. Your model should reflect that.

Please submit your answers on GitHub Classroom.

```
library(pacman)
p_load(ggplot2)
```

---

task 1 - by Nanna Klitgaard Damgaard

1. *(5.2 from ROS)* **Continuous probability simulation:** The logarithms of weights (in pounds) of men in the United States are approximately normally distributed with mean 5.13 and standard deviation 0.17; women's log weights are approximately normally distributed with mean 4.96 and standard deviation 0.20. Suppose 10 adults selected at random step on an elevator with a capacity of 1750 pounds. What is the probability that their total weight exceeds this limit?

```
# Set the parameters
men_mean <- 5.13
men_sd <- 0.17
proportion_men <- 0.51
women_mean <- 4.96
women_sd <- 0.20
proportion_women <- 0.49
n_adults <- 10
n_simulations <- 10000
weight_limit <- 1750
exceeds_limit_count <- 0

# Simulate
for (i in 1:n_simulations) {
  # Generate random genders for the adults
  genders <- sample(c("men", "women"), n_adults, replace = TRUE)
```

```r
  # Generate weights based on genders
  weights <- ifelse(
    genders == "men",
    rlnorm(n_adults, men_mean, men_sd),
    rlnorm(n_adults, women_mean, women_sd)
  )

  # Calculate total weight
  total_weight <- sum(weights)

  # Check if total weight exceeds the limit
  if (total_weight > weight_limit) {
    exceeds_limit_count <- exceeds_limit_count + 1
  }
}

# Calculate probability
probability <- exceeds_limit_count / n_simulations

# Print the result
cat(probability)
```

```
## 0.0551
```

---

task 2 - by Maja Gade Mortensen

2. *(5.6 from ROS)* **Propagation of uncertainty:** We use a highly idealized setting to illustrate the use of simulations in combining uncertainties. Suppose a company changes its technology for widget production, and a study estimates the cost savings at $5 per unit, but with a standard error of $4. Furthermore, a forecast estimates the size of the market (that is, the number of widgets that will be sold) at 40 000, with a standard error of 10 000. Assuming these two sources of uncertainty are independent, use simulation to estimate the total amount of money saved by the new product (that is, savings per unit, multiplied by size of the market).

```r
# Set the given parameters
save_perunit <- 5
save_perunit_se <- 4
market <- 40000
market_se <- 10000

# Number of Simulations
n_sims <- 10000
# Create empty vector to store simulated savings
savings <- rep(NA, n_sims)

# loop through each simulation
for (i in 1:n_sims) {
  savings[i] <-
    rnorm(1, save_perunit, save_perunit_se) * rnorm(1, market, market_se)
}

# Calculate the mean of the simulated savings
```

```r
mean(savings)
```

```
## [1] 200573.9
```

---

task 3 - Sára Fernezelyi

3. *(5.10 from ROS)* **Inference for a ratio of parameters:** A (hypothetical) study compares the costs and effectiveness of two different medical treatments.

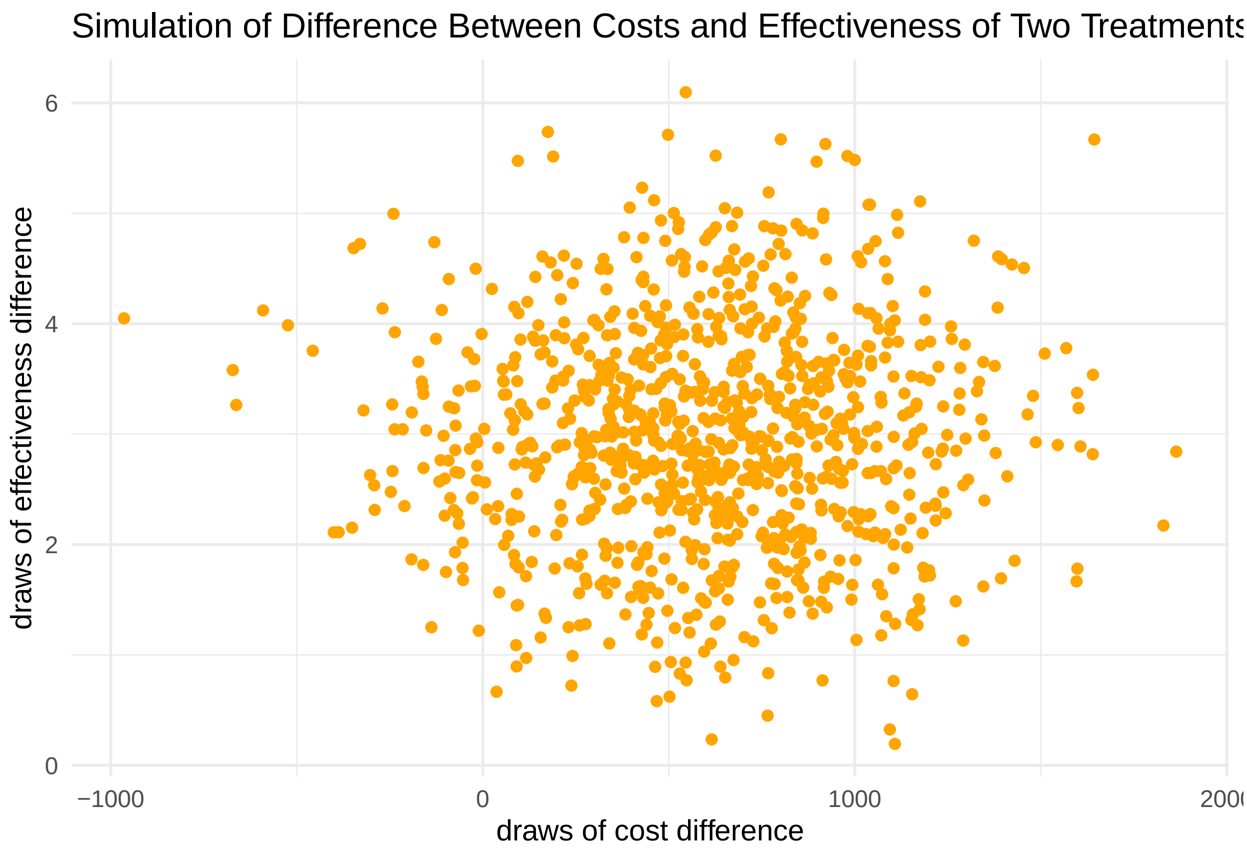
   - In the first part of the study, the difference in costs between treatments A and B is estimated at $600 per patient, with a standard error of $400, based on a regression with 50 degrees of freedom.
   - In the second part of the study, the difference in effectiveness is estimated at 3.0 (on some relevant measure), with a standard error of 1.0, based on a regression with 100 degrees of freedom.
   - For simplicity, assume that the data from the two parts of the study were collected independently.

   Inference is desired for the incremental cost-effectiveness ratio: the difference between the average costs of the two treatments, divided by the difference between their average effectiveness, a problem discussed further by Heitjan, Moskowitz, and Whang (1999).

   (a) Create 1000 simulation draws of the cost difference and the effectiveness difference, and make a scatterplot of these draws.
   (b) Use simulation to come up with an estimate, 50% interval, and 95% interval for the incremental cost-effectiveness ratio.
   (c) Repeat, changing the standard error on the difference in effectiveness to 2.0.

```r
#setting parameters

##parameters for estimated difference in costs
cost_diff_mean <- 600
cost_diff_se <- 400
cost_df <- 52

##parameters for estimated difference in effectiveness
effect_diff_mean <- 3
effect_diff_se <- 1
effect_df <- 102

##number of simulations
n_sims <- 1000

##vectors to store values from the simulation
cost <- rep(NA, n_sims)
effect <- rep(NA, n_sims)

#simulation

for (i in 1:n_sims){
  cost[i] <- rnorm(1, cost_diff_mean, cost_diff_se) #sampling from normal distribution with the given m
  effect[i] <- rnorm(1, effect_diff_mean, effect_diff_se) #sampling from normal distribution with the g
}

#making scatterplot for the simulation

costeffect_df <- data.frame(x = cost, y = effect)
```

```r
costeffect_plot <- ggplot(costeffect_df, aes(x = cost, y = effect)) +
  geom_point(color = "orange") +
  labs(
    x = "draws of cost difference",
    y = "draws of effectiveness difference",
    title = "Simulation of Difference Between Costs and Effectiveness of Two Treatments"
  ) +
  theme_minimal()
costeffect_plot
```




```r
#calculating cost-effectiveness ratio

cost_effect_ratio <- cost/effect

##calculating estimate
estimate <- mean(cost_effect_ratio)
print(paste("The estimate for the incremental cost-effectiveness ratio is:", estimate))
```

```
## [1] "The estimate for the incremental cost-effectiveness ratio is: 246.365532001611"
```

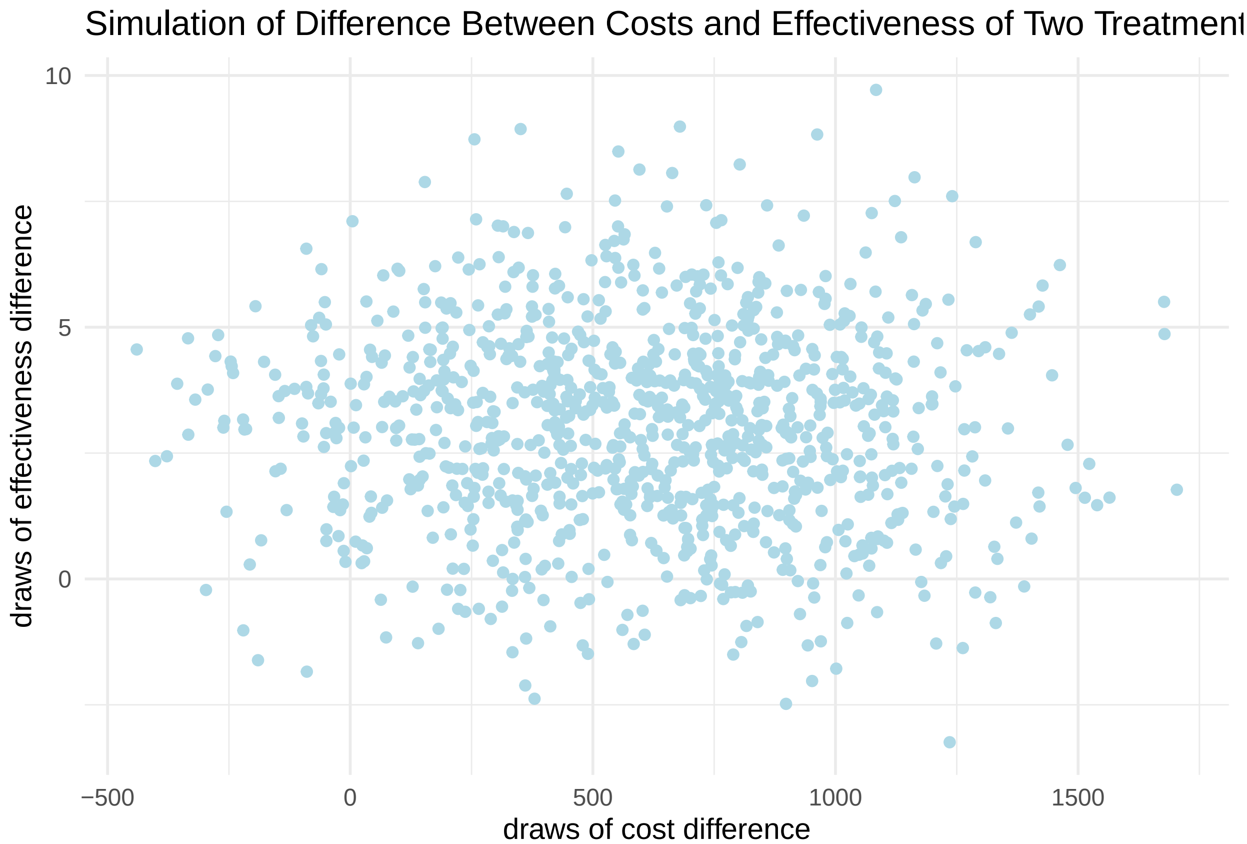
```r
##calculating intervals
intervals_50 <- quantile(cost_effect_ratio,c(0.25,0.75))
intervals_95 <- quantile(cost_effect_ratio,c(0.025,0.975))
print(paste("The 50% interval for the incremental cost-effectiveness ratio is: [", intervals_50[1], ","
```

```
## [1] "The 50% interval for the incremental cost-effectiveness ratio is: [ 110.192045880797 , 320.9918
```

```
print(paste("The 95% interval for the incremental cost-effectiveness ratio is:[", intervals_95[1], ",",
```

```
## [1] "The 95% interval for the incremental cost-effectiveness ratio is:[ -60.3515349982795 , 803.10935
```

```
#setting parameters

##parameters for estimated difference in costs
cost_diff_mean <- 600
cost_diff_se <- 400
cost_df <- 52

##parameters for estimated difference in effectiveness
effect_diff_mean <- 3
effect_diff_se <- 2 #changing standard error to 2 gives bigger spread in the estimated means
effect_df <- 102

##number of simulations
n_sims <- 1000

##vectors to store values from the simulation
cost <- rep(NA, n_sims)
effect <- rep(NA, n_sims)

#simulation

for (i in 1:n_sims){
  cost[i] <- rnorm(1, cost_diff_mean, cost_diff_se) #sampling from normal distribution with the given m
  effect[i] <- rnorm(1, effect_diff_mean, effect_diff_se) #sampling from normal distribution with the g
}

#making scatterplot for the simulation

costeffect_df <- data.frame(x = cost, y = effect)

costeffect_plot <- ggplot(costeffect_df, aes(x = cost, y = effect)) +
  geom_point(color = "lightblue") +
  labs(
    x = "draws of cost difference",
    y = "draws of effectiveness difference",
    title = "Simulation of Difference Between Costs and Effectiveness of Two Treatments"
  ) +
  theme_minimal()
costeffect_plot
```




```r
#calculating cost-effectiveness ratio

cost_effect_ratio <- cost/effect

##calculating estimate
estimate <- mean(cost_effect_ratio)
print(paste("The estimate for the incremental cost-effectiveness ratio is:", estimate))
```

```
## [1] "The estimate for the incremental cost-effectiveness ratio is: 294.468604742867"
```

```r
##calculating intervals
intervals_50 <- quantile(cost_effect_ratio,c(0.25,0.75))
intervals_95 <- quantile(cost_effect_ratio,c(0.025,0.975))
print(paste("The 50% interval for the incremental cost-effectiveness ratio is: [", intervals_50[1], ","
```

```
## [1] "The 50% interval for the incremental cost-effectiveness ratio is: [ 69.5894955107131 , 305.4779
```

```r
print(paste("The 95% interval for the incremental cost-effectiveness ratio is:[", intervals_95[1], ",",
```

```
## [1] "The 95% interval for the incremental cost-effectiveness ratio is:[ -1523.04110496801 , 1776.536
```
