Lab 4 - Chance and Probability Theory

Student Name

xx/xx/xx

Table of Contents

You can keep notes about the lab here and execute code like in the code block below:

# Type or copy/paste R code in an R code block to have the code executed and the  
# output printed in your final (knitted) document.  
10 + 20

## [1] 30

# Part one: correlation and random chance

## Chance exercises

#### 1. Estimate the range (minimum and maximum) of correlations that could occur by chance between two variables with n = 10.

saved\_value <- c() #make an empty variable  
for (i in 1:5) {  
 x <- runif(n = 10, min = 1, max = 10)  
 y <- runif(n = 10, min = 1, max = 10)  
 saved\_value[i] <- sum(x \* y)  
}  
  
min(saved\_value)

## [1] 291.5629

max(saved\_value)

## [1] 415.2185

#### 2. Estimate the range (minimum and maximum) of correlations that could occur by chance between two variables with n = 25.

saved\_value <- c() #make an empty variable  
for (i in 1:5) {  
 x <- runif(n = 10, min = 1, max = 10)  
 y <- runif(n = 10, min = 1, max = 10)  
 saved\_value[i] <- sum(x \* y)  
}  
  
min(saved\_value)

## [1] 287.2322

max(saved\_value)

## [1] 411.148

#### 3. What proportion of correlations (from question 1 and 2) is smaller than -0.3 and larger than +0.3 for n = 10 and n = 25? Create a histogram of the correlations for n = 10 and n = 25 and draw a vertical line at -0.3 and +0.3. Print the found proportions on the plots. Use ggarrange() from the ggpubr library to plot the histograms alongside each other in one figure.

library(ggpubr)  
# see ?ggarrange()   
# in short, you need to assign your ggplots to a variable like so:  
# plot1 <- ggplot(...)  
# plot2 <- ggplot(...)  
# ggarrange(plot1, plot2)

# Part two: generating data in R

## sample and binom() exercises

#### 1. Why are you unable to run the following bit of code? What could you do to fix this? Does this change the nature of your sampling procedure?

sample(c(1:10), 20)

## Error in sample.int(length(x), size, replace, prob): cannot take a sample larger than the population when 'replace = FALSE'

#### 2. How many sixes do you expect to roll when rolling 1 dice 10,000 times? And how many fives **and** sixes do you expect to roll when rolling 10 dice 10,000 times?

#### 3. What is the probability of obtaining exactly 4 heads when flipping 10 fair coins? And what is the probability to obtain **at least** 4 heads (so 4, 5, 6, 7, 8, 9 or 10 heads)? Cf. [the textbook](https://thomashulst.github.io/quantrma/chance.html#working-with-the-binomial-distribution-in-r) when unsure how to use the binom() functions.

## norm() exercises

Run the following bit of code, which samples 20 random numbers from a normal distribution, a couple of times and look at the results.

#### 1. What do you think the function set.seed() does? When do you think this could be useful?

set.seed(123)  
some\_numbers <- rnorm(20, 50, 25) # 20 numbers, mean = 50, s.d. = 25  
print(some\_numbers)

## [1] 35.9881088 44.2455628 88.9677079 51.7627098 53.2321934 92.8766247  
## [7] 61.5229051 18.3734691 32.8286787 38.8584507 80.6020449 58.9953457  
## [13] 60.0192863 52.7670679 36.1039716 94.6728284 62.4462620 0.8345711  
## [19] 67.5338975 38.1802148

#### 2. Suppose the mean of a normal distribution is 25 () and the standard deviation is 3 (). Calculate the probability of obtaining a value between 22 and 28 using R.

#### 3. Based on what you know about the standard normal distribution, could you have figured this out without calculations?

#### 4. Use R to calculate the probability of obtaining a value higher than 29.5 for this normal distribution.

## z-score exercises

[Right click and download this](https://github.com/thomashulst/quantrma_lab/blob/master/data/spssdata/StatsGrades.sav?raw=true) SPSS file containing 49 students’ exam grades (let’s say it’s the final exam for a statistics class). You have likely never worked with an SPSS (.sav) file before, but still, with a quick Google search, you should be able to find a library and function to load the data into R. Remember, you can install packages using install.packages() and load a library with library(). Make sure the .sav is in your working directory so you can load it into R.

Once you have successfully loaded the data into R:

#### 1. Create a table containing the mean and standard deviation for this sample of scores. Also produce a frequency histogram of the grades. How does the distribution of grades look?

#### 2. Transform each student’s score into a Z-score. Now, plot the frequency histogram of this Z-score distribution. Compare it to the raw score distribution. How are they the same? How are they different?

#### 3. Imagine you are a student in this class who received a 90 on this exam. However, the Professor has decided to grade on a curve, such that only the top 10% of the class receives an A (this professor only gives whole grades, no minuses or pluses). Calculate the z-score that corresponds to a raw score of 90 on this exam. Will you get an A with this grade? Why or why not?