

Practice for Exam, Answers

Misja Mikkers, Gertjan Verhoeven & Has van Vlokhoven

Packages needed

```
library(tidyverse)
library(readxl)
library(writexl)
library(palmerpenguins)
library(gapminder)
```

Introduction

In this pdf we will show you examples of exam questions.

Assignment 0

Pull the repository https://github.com/misjamikkers/Rprogramming21_student to find the data needed for the exam, the exam and the answers.

Question 1

In the Sourcedata folder (and on Canvas) you will find the excel file with the name dataQ1.xlsx

Your task is to

1. read this data into a R notebook
2. keep the rows of Group1
3. make a scatterplot with Variable_1 on the x-axis and Variable_2 on the y_axis

Which figure do you recognize in the graph?

- a. Dinosaur
- b. Heart
- c. the letter "L"
- d. the number "1"

Question 2

Your task is to

- use the penguins data from the palmerpenguins package
- keep the rows with year 2009
- create a boxplot with on the x-axis Species per sex and on the y-axis body_mass_g.

Which species per sex has the highest median body mass?

- a. females of the Adelie species
- b. males of the Adelie species
- c. females of the Chinstrap species
- d. males of the Chinstrap species
- e. females of the Gentoo species
- f. males of the Gentoo species
- g. the NA group

Question 3

Your task is

- use the data of the gapminder package
- keep the rows of year 2007
- calculate the medium life expectancy (lifeExp) per continent

What is the median life expectancy of Oceania? Round your answer to the nearest integer.

Question 4a

Your task is to

- Copy the code below (NOTE: include the set.seed!!!) into a notebook
- Use left_join to merge m1 and m2 (in that order!) in to a dataframe with the name “m3”

```
set.seed(123)

m1 <- tibble(id = sample(x= 1:100, size = 70, replace = TRUE)) %>%
  mutate(x = rnorm(n = length(id), mean = 0, sd = 1))

m2 <- tibble(id = sample(x= 1:100, size = 80, replace = TRUE)) %>%
  mutate(y = rnorm(n = length(id), mean = 10, sd = 5 ))
```

How many observations (rows) does m3 have?

Question 4b

Your task is to use m3 from question 4a to

- create a long dataframe m4 with a new column “Var” that indicates whether the row concerned shows a value of “x” or “y” and the values in column “Val”.

So your dataframe should look like:

| id | Var | Val |
|-----|-----|------------|
| 31 | x | -0.1639310 |
| 31 | y | 14.3968207 |
| ... | ... | ... |
| ... | ... | ... |

- Then delete all NA’s.

How many observations (rows) does m4 have?

Question 5

What does this code do?

```
knitr::opts_chunk$set(echo = FALSE)
```

- a. The code makes sure that this code chunk will not be executed
- b. The code makes sure that all the code chunks after this command in the Rmd will not be executed
- c. The code makes sure that this code chunk will not be shown in the output document (e.g. the pdf)
- d. The code makes sure that all the code chunks after this chunk will not be shown in the output document (e.g. the pdf)

Question 6

Your task is to

- open a notebook
- simulate some data:
 - USE `set.seed(123)` [This is really important!]
 - generate an x-variable in a tibble with a normal distribution with mean 100, and standard deviation 10. We need 1000 observations
 - make a new column y with the formula $\log(x) + 36 * x^5$
 - calculate the mean of the variable y.

Question 7

Create a function that for the numbers x to $x + 2 * \pi$ calls the function `sin(x)` and adds the results.

The answer when you execute your function with the argument 3 should be 0.9878401