

## Applied Statistics – Exercise 2

### Problems

1. The data set `pi2000` (UsingR) contains the first 2,000 digits of  $\pi$ . What is the percentage of digits that are 3 or less? What is the percentage of the digits are 5 or more?
2. Load the `starwars` dataset from the `dplyr` package. The dataset has 87 characters from the StarWars universe with 13 features. **Note** there are missing data in the database represented by `NA` values.
  - a) What is the homeworld of Mace Windu?
  - b) How many droids are in the dataset?
  - c) Who are the shortest and tallest humans in dataset?
  - d) What is the mean and standard deviation of the height all humans in the starwars database?
3. The following table shows the results of a survey of 10 pirates. In addition to some basic demographic information, the survey asked each pirate “What is your favorite superhero?” and “How many tattoos do you have?”

Name	Sex	Age	Superhero	Tattoos
Astrid	F	30	Batman	11
Lea	F	25	Superman	15
Sarina	F	25	Batman	12
Remon	M	29	Spiderman	5
Letizia	F	22	Batman	65
Babice	F	22	Antman	3
Jonas	M	35	Batman	9
Wendy	F	19	Superman	13
Niveditha	F	32	Maggott	900
Gioia	F	21	Superman	0

- a) Combine the data into a single dataframe. Complete all the following exercises from the dataframe!
- b) What was the mean age of female and male pirates separately?
- c) Add a new column to the dataframe called `tattoos.per.year` which shows how many tattoos each pirate has for each year in their life. Which pirate had the most number of tattoos per year?
- d) What was the median number of tattoos of pirates over the age of 20 whose favorite superhero is Spiderman?
- e) Make a boxplot of the age distribution of the pirates
- f) Make a piechart showing the number of pirates which has each superhero as their favorite.