Family_dataset

Khanin Sisaengsuwanchai

1/27/2022

0. Data check

Download **family.rda** from my local directory

```
load("/Users/khaninsi/Documents/Github/Khaninsi/Programming_skills/R/data/family.rda")
```

Check the correctness of the data.

```
head(fheight)
```

```
##  a  b  c  d  e  f
## 70 64 73 67 64 68
```

```
head(fweight)
```

```
## [1] 175 125 185 156 105 190
```

1. Let's begin by converting our measurements to the correct units

BMI is weight in kilograms divided by height in inches squared

I will initially define functions that convert lbs to kg and inches_to_m by using conversion formula from Google.

- lbs_to_kg
- inches_to_m

```
lbs_to_kg = function(weight_lbs) {
  weight_kg = weight_lbs/2.205
  return(weight_kg)
}

inches_to_m = function(height_inches) {
  height_m = height_inches/39.37
  return(height_m)
}
```

Convert 175 pounds to kilograms

```
weight_kg = lbs_to_kg(175)
weight_kg
```

```
## [1] 79.36508
```

Convert 69 inches to meters.

```
height_m = inches_to_m(69)
height_m
```

```
## [1] 1.752604
```

Compute BMI

```
bmi_1 = 79.36508/(1.752604^2)
bmi_1
```

```
## [1] 25.83818
```

This time do all the computations at once

```
bmi_2 = weight_kg/(height_m^2)
bmi_2
```

```
## [1] 25.83819
```

Why are the answers slightly different?

- This is because when I explicitly copy the values of weight_kg and height_m, they do not represent all decimal points of the data. In contrast, when we compute bmi_2, the weight_kg and height_m variables contain all decimal points in the data, leading to slightly different results.

2. Let's convert all of the values for weight

Convert fweight from lbs to kg

```
fweight_kg = lbs_to_kg(fweight)
fweight_kg
```

```
##  [1] 79.36508 56.68934 83.90023 70.74830 47.61905 86.16780 83.90023 56.23583
##  [9] 79.36508 97.50567 75.28345 63.49206 68.02721 56.68934
```

Convert fheight from inches to m

```
fheight_m = inches_to_m(fheight)
fheight_m
```

```
##        a        b        c        d        e        f        g        h
## 1.778004 1.625603 1.854204 1.701803 1.625603 1.727203 1.727203 1.651003
##        i        j        k        l        m        n
## 1.727203 1.803404 1.701803 1.676403 1.676403 1.574803
```

Assign the return value of the computations to a new variable called bmi

```
bmi = fweight_kg/(fheight_m^2)
bmi
```

```
##        a        b        c        d        e        f        g        h
## 25.10523 21.45222 24.40328 24.42852 18.01987 28.88404 28.12394 20.63085
##        i        j        k        l        m        n
## 26.60373 29.98085 25.99446 22.59240 24.20614 22.85856
```

We can summarize bmi by calling the summary function. Notice it provides quartiles, mean, min, and max

```
summary(bmi)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.02   22.66   24.42   24.52   26.45   29.98
```

3. Answering questions using the data

- What is the average weight for men over 70?

```
mean(fweight[fgender=="m" & fage>70])
```

```
## [1] 180
```

Average weight in lbs for men over 70 in lbs is 180

- What is the BMI for the tallest woman?

```
bmi[fheight == max(fheight[fgender=="f"]) & fgender=="f"]
```

```
##        f
## 28.88404
```

The BMI of the tallest woman is 28.88404.

- How many women are in family.rda? (use subsetting techniques)

```
length(fgender[fgender=="f"])
```

```
## [1] 6
```

There are 6 women in the family.rda.

- Using subsetting techniques, answer a new question you think a policy maker might find interesting about family.rda.

An interesting question that I try to inform the policy maker is about the profile of family members who are overweight, leading to a high risk of getting certain diseases and cancers. For example, does overweight likely to occur in particular gender or age?
With this information, it helps the policy maker to understand the group of people that has a high risk of obesity and thus makes policies to help this particular group with high priority.
But first let's trying to understand the overall bmi data for each group.
Find, summarize, and count the data that are overweight

```
overWeight = bmi[foverWt == TRUE]
summary(overWeight)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25.11   26.15   27.36   27.45   28.69   29.98
```

```
length(overWeight)
```

```
## [1] 6
```

Find, summarize, and count the data that are not overweight

```
not_overWeight = bmi[foverWt == FALSE]
summary(not_overWeight)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.02   21.25   22.73   22.32   24.26   24.43
```

```
length(not_overWeight)
```

```
## [1] 8
```

In this sample, we can clearly see the differences of these two groups. The average and median bmi of the overweight group is approximately 5 more than the average of the non-overweight group. Moreover, we have a slightly non-overweight individuals than overweight. Another interesting point is the cut-off of these two groups. It seems like that the bmi of 25 separates the overweight and non-overweight, which conforms with what CDC defines.

Let's check about the relationships of each gender and bmi

```
table(fgender[foverWt == TRUE])
```

```
##
## m f
## 4 2
```

```
table(fgender[foverWt == FALSE])
```

```
##
## m f
## 4 4
```

Based on the availability of the data, we cannot see any clear patterns of overweight with gender.

Another point is to examine the age of these two groups.

```
summary(fage[foverWt == TRUE])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    27.0    38.5    57.0    53.0    65.0    77.0
```

```
summary(fage[foverWt == FALSE])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    24.0    31.5    46.5    44.5    49.0    79.0
```

Both median and mean of age in the overweight group roughly exceed the non-overweight group by 10. As a result, I would inform the policy maker that the new policies should initially focus on the citizens who are more than 50 years old since they are more likely to be overweight.