Stat201 Assignment 1

Robert Ivill 46012819

RStudio Source Editor

bodyweight

Filter

	Workout	Bodyfat
1	Yes	13
2	Yes	6
3	Yes	20
4	Yes	8
5	Yes	14
6	Yes	19
7	Yes	18
8	Yes	25
9	Yes	16
10	Yes	24
11	Yes	15
12	Yes	10
13	Yes	15
14	No	22
15	No	16
16	No	22
17	No	21
18	No	30
19	No	26
20	No	12
21	No	23
22	No	28
23	No	23

Question 1:

After opening RStudio, I used the console to create the variable ‘Bodyfat’ and assigned the array of body fat percentages ‘c(13, 6, 20, 8, 14, 19, 18, 25, 16, 24, 15, 10, 15)’ to it. After using the function ‘mean(Bodyfat)’, we find out that the mean of these items is 15.61538, or 15.6 rounded to 1 d.p. Next, I cleared the environment to get rid of the variable and imported the bodyweight.csv dataset into R.

After importing the dataset, we could see what it looks like using the function ‘view(bodyweight)’. Attempting to use the function ‘mean(Bodyfat)’ did not work as it didn’t specify that it is from the bodyweight dataset. So, using the function ‘mean(bodyweight$Bodyfat)’ showed that the mean bodyfat from the dataset is 18.5%. Then I used the function ‘aggregate(bodyweight$Bodyfat, list(bodyweight$Workout), FUN=mean)’ to find out that the mean bodyfat of those who work out is 15.6% and 22.3% for those who don’t.

Question 2:

To create a histogram of the bodyfat measurements, I used the function ‘hist(bodyweight$Bodyfat, xlab="Bodyfat(%)", ylab = "Count", main="Bodyfat measurements")’, which adds labels to the graph as well. The histogram shows that most people were found to have a bodyfat between 20% and 25%, with 7 people being in this range. We can see that few people in the dataset had bodyfat measurements at the extreme ends, with 5-10% and 25-30% having 3 measurements each. Two more histograms with 10 and 20 bars were created as well by adding ‘breaks = 10’ and ‘breaks = 20’ into the function. These histograms show us that most measurements were around the 15% area. The graphs have two peaks at 15% and 20-24% and other areas of the graph only have 1 or 2 counts. The graph with 10 breaks is much more useful as it shows a good summary of the data, not being too vague or specific with the measurements compared to the other graphs. This shows how important it is to format histograms with the correct number of bars to show the full story behind the data.