Lecture 1.3 - Advanced distributions activity
DKU Stats 101 Spring 2025 S4

Professor MacDonald

2025-03-25

Lecture 1.3 in-class activity - more on distributions

We'll work with some items from the McDonald's menu to check our understanding of distributions.

Table 1: Calories of select McDonald's items

Category	Min	Q1	Median Calories	Q3	Max
Beef & Pork	240	430	500	570	750
Breakfast	150	410	470	570	1150
Chicken & Fish	190	410	480	610	1880

Boxplots

Boxplot by hand practice

1. Make side-by-side boxplots of these three distributions by hand
2. Which (if any) of these distributions have outliers?
3. Which are right skewed? Left skewed? How do you know?
4. What can you conclude from your hand-drawn graph?

Boxplot in ggplot

Now, make a boxplot using `ggplot` that replicates the ones you made by hand. Make sure to only display data on the main meal McDonald's menu items, not the other categories (hint: use the `filter` verb).

Information on how to make boxplots using `ggplot` can be found here.

What are the outliers for the three categories? Look them up in the dataset and decide with your partner if there is a justification for excluding the items from the analysis or not.

Normal distribution

Assuming normality

For the below exercises, we are going to assume `Carbohydrates` are normally distributed. We will relax that assumption in the following section.

1. What are the likely units of `Carbohydrates`? What are healthy ranges of carbohydrates per day (research on your own).
2. Calculate and display the mean of `Carbohydrates`. In what ways can using the mean be a problem here?
3. Calculate and display the standard deviation of `Carbohydrates` using the `sd()` command.

4. Then calculate the number of `Carbohydrates` that are -2, -1, 0, 1, and 2 standard deviations above and below the mean. Make a table using the Table menu.

Table ▾	≡ Ou
⊞ **Insert Table...**	
Insert Row Above	
Insert Row Below	
Insert Column Left	
Insert Column Right	
Delete Row	
Delete Column	
Delete Table	
Align Column	▶
Table Header	
Table Caption	

5. If McDonalds decides that they want to make their menu healthier and eliminate menu items that are greater than 75 carbohydrates, about what percent of their menu items would they need to eliminate?
6. How about if they wanted to eliminate all menu items that are less than 25 carbohydrates and greater than 75 carbohydrates?

Checking normality

1. Make a histogram using `ggplot` of carbohydrates. Does it match our assumptions of normality?
2. Investigate why there are deviations from normality. What factor(s) lead to this problem?
3. Check the outliers - are the outliers a problem or not?

3