**Exercise 1**

Open the file **World Happiness 2019.csv**. This dataset has a “happiness score” which was measured by simply asking the people in the country how happy they are on a scale of 0 to 10 and then calculating the average per country. It also contains many other variables which might be related to happiness. The dataset was compiled by the Sustainable Development Solutions Network (2020).

1. Create a scatterplot between the variables happiness score and GDP per capita. What do you see?

This should be your output:

A diagram of blue dots

Description automatically generated

You can see that there is a strong positive relationship. In other words, countries with higher GDPs per capita also tend to be happier.

1. Calculate the correlation between the variables happiness score and GDP per capita. Does it match with the scatterplot

The correlation coefficient is 0.79. This is in line with what the scatterplot shows.

1. Create a scatterplot matrix. Are there variables which show no clear relationship with happiness score?

Your output should be this:

A group of blue dots

Description automatically generated

It can be seen that the generosity variable does not correlate with happiness. Perceptions of corruption does, but it’s not very strong.

**A screenshot of a computer

Description automatically generated**

Most correlations are reasonably high, but the correlation between happiness score and generosity is not significant.

**Exercise 2**

Open the file **video\_games.xls**. This file has data from a large amount of video games sold by blockbuster. More information about this dataset can be found in the article written by Cox (2013). In this exercise you will try to predict the sales of the video games using linear regression. The sales are in thousands of units.

Create four regression models. All of the should have Sales as the dependent variable but a different independent variable.

What do the coefficients and p-values tell you about the relationship between the sales and the independent variables? Which independent variable has the most explanatory power?

*Solution*

Review score has a coefficient of 24.602. This means that as the review score goes up by 1 point, on average the sales increase by 24.6 thousand.

Used price has a coefficient of 55.799 which means that if the price goes up by 1 dollar, the sales increase by 55.8 thousand (this seems strange but probably it’s not because the videogames sell more because they are more expensive, but that they are more expensive because they sell more)

Year is not statistically significant so we ignore the coefficient (p-value is higher than 0.05)

Length has a coefficient of 1.7225 meaning that if the playtime goes up by 1 hour, the sales increase by 1.7 thousand.

**References**

Cox, J. (2014). What makes a blockbuster video game? An empirical analysis of US sales data. *Managerial and Decision Economics*, *35*(3), 189-198.

Sustainable Development Solutions Network (2020). *World Happiness Report*. Retrieved from: https://www.kaggle.com/unsdsn/world-happiness