Jessica Miron

In-Class Activity 7.2

BME 598: Applied Programming

Executive Summary

The dataset includes information on different coronary heart disease risk factors as well as whether patients have coronary heart disease or not. The goal of the analysis is to determine which risk factors are related and what each relationship means.

1. Load data from .tab file and Python packages
2. Create pair plots of all variables to determine which are most likely to have correlations and should be investigated further.
3. Compute pairwise Pearson correlation values and p-values for all variable pairs identified in the previous step. Convert to a Pandas data frame then export to a csv.
4. Create a heatmap to visualize the correlation between variables
5. Complete a biological interpretation on which relationships were significant and what the relationships mean.

Biological Interpretation

After analyzing the data, a heat map (Figure 1) and correlation and p-value table (Table 1) were created. In Table 1, correlation values above 0.3 were considered significant and p-values below 0.05 were considered significant. Eight relationships were found to be significant. These were adiposity and systolic blood pressure (SBP), adiposity and low density lipoprotein (LDL), adiposity and obesity, adiposity and age, age and SBP, age and tobacco, age and LDL, and LDL and obesity. Of these significant relationships all of them were positive. Looking at Figure 1, age, adiposity, and obesity have the most positive significant relationships which aligns with the data in Table 1.

Table 1: Pearson R-Correlation values and Pearson p-values for relationships between eight quantitative variables. Cells in red have values that are significant: above 0.3 for R-Correlation values and below 0.05 for p-values. Since each relationship is shown twice, there are eight significant relationships.

	adiposity R	adiposity pv	age R	age pv	alcohol R	alcohol pv	ldl R	ldl pv	obesity R	obesity pv	sbp R	sbp pv	tobacco R	tobacco pv	typea R	typea pv
sbp	0.36	2.72E-15	0.39	4.05E-18	0.14	2.54E-03	0.16	6.38E-04	0.24	2.25E-07	0.00	0.00E+00	0.21	4.18E-06	-0.06	2.18E-01
tobacco	0.29	3.46E-10	0.45	1.88E-24	0.20	1.37E-05	0.16	6.08E-04	0.12	7.37E-03	0.21	4.18E-06	0.00	0.00E+00	-0.01	7.54E-01
ldl	0.44	2.41E-23	0.31	7.11E-12	-0.03	4.74E-01	0.00	0.00E+00	0.33	3.09E-13	0.16	6.38E-04	0.16	6.08E-04	0.04	3.45E-01
adiposity	0.00	0.00E+00	0.63	1.26E-51	0.10	3.11E-02	0.44	2.41E-23	0.72	5.66E-74	0.36	2.72E-15	0.29	3.46E-10	-0.04	3.55E-01
typea	-0.04	3.55E-01	-0.10	2.74E-02	0.04	3.97E-01	0.04	3.45E-01	0.07	1.12E-01	-0.06	2.18E-01	-0.01	7.54E-01	0.00	0.00E+00
obesity	0.72	5.66E-74	0.29	1.61E-10	0.05	2.68E-01	0.33	3.09E-13	0.00	0.00E+00	0.24	2.25E-07	0.12	7.37E-03	0.07	1.12E-01
alcohol	0.10	3.11E-02	0.10	2.98E-02	0.00	0.00E+00	-0.03	4.74E-01	0.05	2.68E-01	0.14	2.54E-03	0.20	1.37E-05	0.04	3.97E-01
age	0.63	1.26E-51	0.00	0.00E+00	0.10	2.98E-02	0.31	7.11E-12	0.29	1.61E-10	0.39	4.05E-18	0.45	1.88E-24	-0.10	2.74E-02

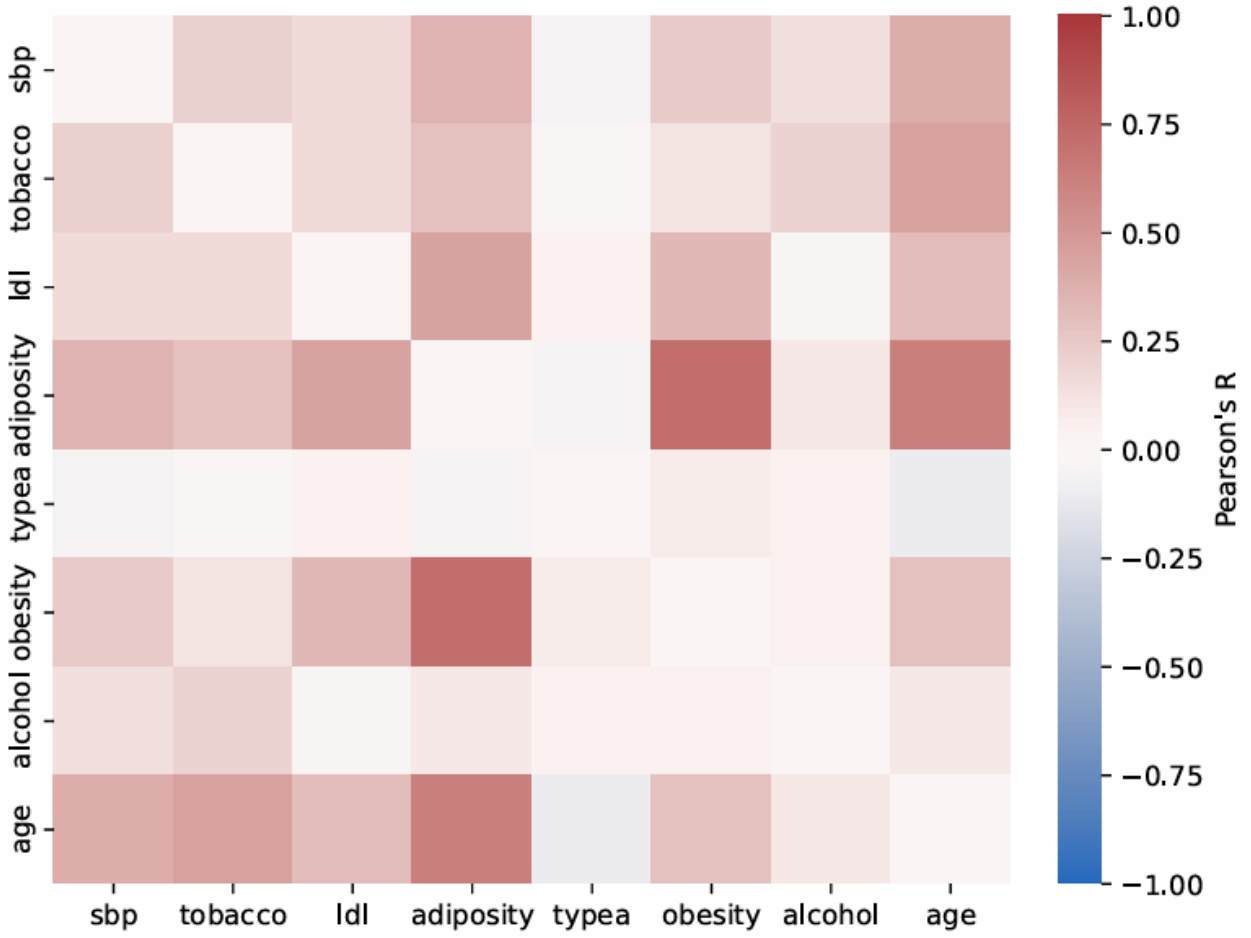

Figure 1: Heatmap of correlations between eight quantitative variables. An R value of one corresponding to red means a positive relationship. A R value of negative one corresponding to blue means a negative relationship. White represents very little to no correlation between variables.

The positive correlation that each of the significant relationships exhibit show that if a person has one risk factor, they are likely to have other risk factors. For example, if a patient has severe adiposity, they also likely have a high SBP, high LDL, severe obesity, and are older. This positive correlation between risk factors signifies that patients likely have many risk factors or few risk factors meaning that patients have greater risk of coronary heart disease beyond just one risk factor. Having the whole patient picture shows that many risk factors are related, increasing overall risk.

Diving deeper into two relationships, it was found that the positive correlations have been recorded in other studies. Adiposity, along with other cardiometabolic is known to increase SBP. Holmes et al. found that increases in BMI (which adiposity is related to) led to significant increases in both systolic and diastolic blood pressure. For the relationship between adiposity and age, Jura and Kozak showed that for both lean and obese individuals there are age related changes in adipose tissue deposition that increases adiposity.

References

Holmes, Michael V. et al., “Causal Effects of Body Mass Index on Cardiometabolic Traits and Events: A Mendelian Randomization Analysis”, *The American Journal of Human Genetics*, vol 94, 2 (2014): 198-208 https://doi.org/10.1016/j.ajhg.2013.12.014.

Jura, Magdalena, and Leslie P Kozak. “Obesity and related consequences to ageing.” *Age (Dordrecht, Netherlands)* vol. 38,1 (2016): 23. doi:10.1007/s11357-016-9884-3