Jessica Miron

Homework 7

BME 598: Applied Programming

Executive Summary

Ischemic stroke is caused by parts of the brain not getting blood flow due to a blockage somewhere in the vasculature. These blockages can be intracranial or extracranial. The paper presented examines differences in risk factors and pathophysiology in an attempt to correlate differences that may cause intracranial atherosclerosis ischemic stroke versus extracranial atherosclerosis ischemic stroke [1]. For this application, the data from the paper is being analyzed to find correlations between risk factors to determine significant relationships. The steps taken for analysis can be seen below.

1. Load data from the csv file and the Python packages
2. Make a pairs plot of all variables in the dataset to determine which variables should be further evaluated
3. Compute Pearson R-correlation values and p-values from the quantitative variables determined from the pairs plot and export to a csv file
4. Create a heatmap of the correlation values to visualize the relationships between variables
5. Write a biological interpretation analyzing which relationships are significant, their direction, and what the relationships signify for the dataset

Biological Interpretation

After analyzing the data and running pairwise Pearson correlation tests, a heatmap of the correlations and a table of R and p-values were created. Looking at the heatmap, Figure 1, there appears to be more significant positive correlations than negative correlations. The LDL/TC pair have the biggest positive correlation, and the AIP/HDL pair have the biggest negative correlation. Most pairs have small positive or negative correlations, but few have large correlations. Moving to the table of R and p-values, Table 1, the observations from Figure 1 can be verified. From Table 1 there are 13 significant pair correlations. These are AIP/HDL, AIP/TG, AIP/TyG, Age/Smoking, DBP/SBP, FBS/HbA1c, FBS/TyG, HbA1c/TyG, LDL/TC, Sex/Smoking, Smoking/Age, TC/TG, and TC/TyG. Of these 13 significant relationships, 11 are positive and two are negative. As noted from Figure 1, LDL/TC has the largest positive relationship at 0.93 and AIP/HDL has the largest negative relationship at -0.65.

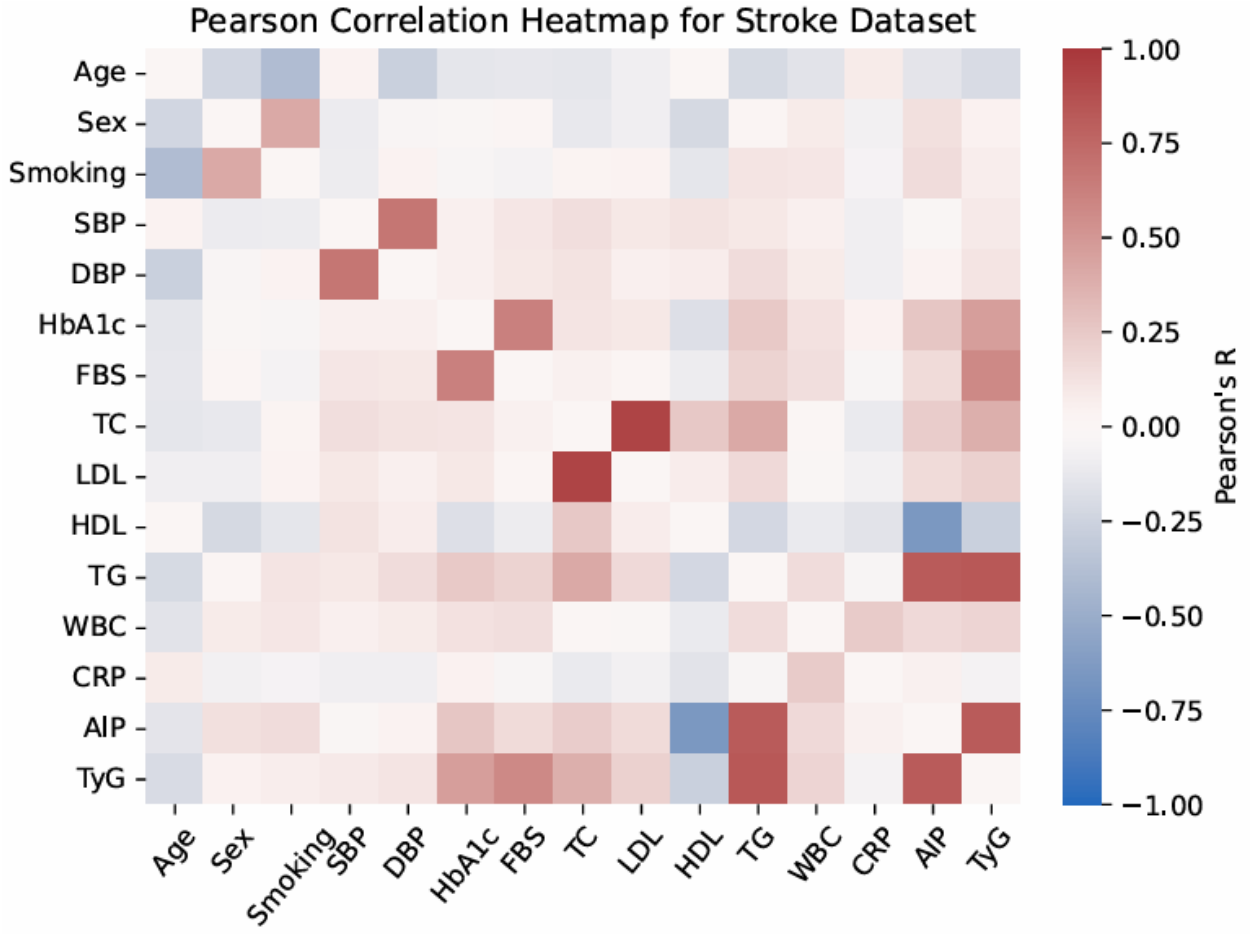

Figure 1: Heatmap of Pearson R-Correlation values across 15 variables. More red relationships are more positively related and more blue relationships are more negatively related. Values that are close to white have minimal relationship.

Looking at the meaning of each of these relationships, there are correlated risk factors that appear. The two negative relationships, AIP/HDL and Age/Smoking, both signify that as one factor increases the other decreases. For Age/Smoking, as age increases the rate of smoking decreases which makes sense as people in their youth are more likely to participate in risky behavior. For AIP/HDL, AIP is the atherogenic index of plasma which is the risk of developing atherosclerosis and HDL is high-density lipoprotein. AIP is calculated by taking the log of triglyceride levels divided by HDL, which makes sense why as HDL increases AIP would decrease due to the ratio of triglyceride to HDL getting smaller [1]. For the 11 positive relationships, an increase in one factor leads to an increase in the other. Meaning that if a patient has one risk factor, they most likely have the other. The largest positive relationships will be examined, which are LDL/TC and TG/TyG. LDL/TC had the largest positive relationship with an R correlation value of 0.93 and a p-value of 2.49e-222. This signifies a large positive relationship that is highly significant. LDL (low-density lipoprotein) and TC (total cholesterol) are directly related to each other as TC is a measure of LDL and HDL. This correlates to why as LDL increases TC increases. TG/TyG had the second largest positive relationship with an R correlation value of 0.83 and a p-value of 5.54e-130 which is also a large positive and highly

significant relationship. TG (triglycerides) and TyG (TG-glucose) are also directly related with TyG being a measure of TG. TyG is the natural log of the average of the TG and glucose levels [1]. Once again this directly relates to a positive relationship since an increase in TG will cause an increase in TyG.

Examining the two relationships of age/smoking and sex/smoking, there is a negative and positive correlation respectively. Lower ages have higher smoking rate and men have higher smoking rate than women. To validate these relationships, a study from Kranjac and Kranjac is examined. Their study included almost 46,000 participants across 21 years from age 18-74. They found that there is a negative correlation between age and smoking though the peak occurs at 27 years old [2]. After 27 years old, smoking continues to decline as age increases. They also found that men are more likely to smoke than women across all ages [2]. These findings match the correlated relationships calculated validating that these are true correlated relationships.

	AIP.R	AIP.pv	Age.R	Age.pv	CRP.R	CRP.pv	DBP.R	DBP.pv	FBS.R	FBS.pv	HDL.R	HDL.pv	HbA1c.R	HbA1c.pv	LDL.R	LDL.pv	SBP.R	SBP.pv	Sex.R	Sex.pv	Smoking.R	Smoking.pv	TC.R	TC.pv	TG.R	TG.pv	TyG.R	TyG.pv	WBC.R	WBC.pv
Age	-0.15	8.92E-04	0.00	0.00E+00	0.08	6.07E-02	-0.27	1.60E-09	-0.13	3.71E-03	0.01	8.82E-01	-0.13	2.76E-03	-0.08	6.36E-02	0.03	4.46E-01	-0.23	1.88E-07	**-0.39**	**6.77E-20**	-0.14	2.27E-03	-0.21	3.71E-06	-0.20	5.55E-06	-0.15	7.24E-04
Sex	0.13	2.63E-03	-0.23	1.88E-07	-0.08	9.28E-02	-0.03	5.02E-01	0.02	7.20E-01	-0.21	1.75E-06	0.00	9.62E-01	-0.08	7.48E-02	-0.10	1.96E-02	0.00	0.00E+00	**0.41**	**7.65E-22**	-0.12	7.73E-03	0.02	7.31E-01	0.05	3.15E-01	0.08	7.31E-02
Smoking	0.16	4.78E-04	**-0.39**	**6.77E-20**	-0.05	2.50E-01	0.03	4.62E-01	-0.06	2.07E-01	-0.14	1.83E-03	-0.04	3.98E-01	0.03	4.54E-01	-0.10	2.83E-02	**0.41**	**7.65E-22**	0.00	0.00E+00	0.03	5.17E-01	0.11	1.16E-02	0.07	1.26E-01	0.11	1.46E-02
SBP	0.00	9.71E-01	0.03	4.46E-01	-0.08	8.00E-02	**0.67**	**1.74E-67**	0.11	1.73E-02	0.12	6.86E-03	0.06	1.71E-01	0.09	3.47E-02	0.00	0.00E+00	-0.10	1.96E-02	-0.10	2.83E-02	0.14	1.18E-03	0.10	2.82E-02	0.09	3.61E-02	0.06	2.11E-01
DBP	0.03	4.43E-01	-0.27	1.60E-09	-0.08	7.30E-02	0.00	0.00E+00	0.10	2.47E-02	0.08	8.13E-02	0.06	1.80E-01	0.06	1.66E-01	**0.67**	**1.74E-67**	-0.03	5.02E-01	0.03	4.62E-01	0.12	5.84E-03	0.15	6.56E-04	0.11	1.30E-02	0.08	8.05E-02
HbA1c	0.26	3.40E-09	-0.13	2.76E-03	0.04	3.69E-01	0.06	1.80E-01	**0.62**	**2.50E-54**	-0.18	8.04E-05	0.00	0.00E+00	0.10	3.26E-02	0.06	1.71E-01	0.00	9.62E-01	-0.04	3.98E-01	0.12	8.88E-03	0.25	1.74E-08	**0.47**	**1.47E-28**	0.13	3.06E-03
FBS	0.16	4.28E-04	-0.13	3.71E-03	-0.03	4.56E-01	0.10	2.47E-02	0.00	0.00E+00	-0.10	3.09E-02	**0.62**	**2.50E-54**	0.02	7.26E-01	0.11	1.73E-02	0.02	7.20E-01	-0.06	2.07E-01	0.05	2.40E-01	0.20	7.96E-06	**0.57**	**5.30E-45**	0.15	1.06E-03
TC	0.23	1.85E-07	-0.14	2.27E-03	-0.12	9.06E-03	0.12	5.84E-03	0.05	2.40E-01	0.25	1.28E-08	0.12	8.88E-03	**0.93**	**2.49E-222**	0.14	1.18E-03	-0.12	7.73E-03	0.03	5.17E-01	0.00	0.00E+00	**0.41**	**8.99E-22**	**0.38**	**1.44E-18**	0.00	9.28E-01
LDL	0.16	4.10E-04	-0.08	6.36E-02	-0.08	8.60E-02	0.06	1.66E-01	0.02	7.26E-01	0.07	1.08E-01	0.10	3.26E-02	0.00	0.00E+00	0.09	3.47E-02	-0.08	7.48E-02	0.03	4.54E-01	**0.93**	**2.49E-222**	0.17	1.53E-04	0.21	2.18E-06	-0.02	7.05E-01
HDL	**-0.65**	**2.79E-61**	0.01	8.82E-01	-0.15	7.61E-04	0.08	8.13E-02	-0.10	3.09E-02	0.00	0.00E+00	-0.18	8.04E-05	0.07	1.08E-01	0.12	6.86E-03	-0.21	1.75E-06	-0.14	1.83E-03	0.25	1.28E-08	-0.22	3.88E-07	-0.27	5.80E-10	-0.11	1.02E-02
TG	**0.82**	**1.39E-121**	-0.21	3.71E-06	-0.03	4.39E-01	0.15	6.56E-04	0.20	7.96E-06	-0.22	3.88E-07	0.25	1.74E-08	0.17	1.53E-04	0.10	2.82E-02	0.02	7.31E-01	0.11	1.16E-02	**0.41**	**8.99E-22**	0.00	0.00E+00	**0.83**	**5.54E-130**	0.15	7.63E-04
WBC	0.17	1.44E-04	-0.15	7.24E-04	0.24	1.05E-07	0.08	8.05E-02	0.15	1.06E-03	-0.11	1.02E-02	0.13	3.06E-03	-0.02	7.05E-01	0.06	2.11E-01	0.08	7.31E-02	0.11	1.46E-02	0.00	9.28E-01	0.15	7.63E-04	0.19	2.27E-05	0.00	0.00E+00
CRP	0.05	2.76E-01	0.08	6.07E-02	0.00	0.00E+00	-0.08	7.30E-02	-0.03	4.56E-01	-0.15	7.61E-04	0.04	3.69E-01	-0.08	8.60E-02	-0.08	8.00E-02	-0.08	9.28E-02	-0.05	2.50E-01	-0.12	9.06E-03	-0.03	4.39E-01	-0.06	2.10E-01	0.24	1.05E-07
AIP	0.00	0.00E+00	-0.15	8.92E-04	0.05	2.76E-01	0.03	4.43E-01	0.16	4.28E-04	**-0.65**	**2.79E-61**	0.26	3.40E-09	0.16	4.10E-04	0.00	9.71E-01	0.13	2.63E-03	0.16	4.78E-04	0.23	1.85E-07	**0.82**	**1.39E-121**	**0.82**	**8.82E-121**	0.17	1.44E-04
TyG	**0.82**	**8.82E-121**	-0.20	5.55E-06	-0.06	2.10E-01	0.11	1.30E-02	**0.57**	**5.30E-45**	-0.27	5.80E-10	**0.47**	**1.47E-28**	0.21	2.18E-06	0.09	3.61E-02	0.05	3.15E-01	0.07	1.26E-01	**0.38**	**1.44E-18**	**0.83**	**5.54E-130**	0.00	0.00E+00	0.19	2.27E-05

Table 1: Table of Pearson R-Correlation and p-values. R values that have absolute values above 0.3 are considered significant and highlighted in red. p-values that are below 0.05 are considered significant and highlighted in red. Pairwise relationships that have both R-Correlation and p-values at a significant level are bolded to show significant relationships.