Assignment #4

 BIOINF 525: Module 2

Use the data set from the TROPHY study to answer the following questions. Return the R code, the output, and a brief description of the results.

1. Test whether LDL predict BMI24: H0: $ρ$ = 0vs. HA: $ρ$ ≠ 0.
2. Use a scatterplot to visually display the relationship between LDL and BMI24. Add a line (best linear fit) to the scatterplot to help identify any linear pattern between LDL and BMI24.
3. Calculate the Pearson correlation and the corresponding 95%CI between BMI24 and LDL. Is the correlation different from zero? H0: $ρ$ = 0 vs. HA: $ρ$ ≠ 0?
4. Use simple linear regression to predict BMI24 based on LDL. What is the R-square for the model?
5. Test whether the association between LDL and BMI24 is modified by treatment (Candesartan). That is, is the relationship between LDL and BMI24 different between treatment group and placebo group H0: $ ρ$ 1 = $ρ$ 2 vs. HA: $ ρ$ 1 ≠ $ρ$ 2.
6. Use one overlapping scatterplot to visually display the relationship between LDL and BMI24 for Candesartan and Placebo groups.Add a line (best linear fit) for each group and a legend for each line.
7. Test if the relationship between LDL and BMI24 is different for Candesartan and Placebo. Use multiple regression with BMI24 as outcome, and LDL, Trt, and LDL\*Trt as predictors to test this. Is the interaction term (Trt\*LDL) significant (p<.05)?
8. Repeat b), but adjusted for Age and Insulin (add Age and Insulin in the model in b. ). Is the interaction Trt\*LDL significant, after adjusting for Age and Insulin?

Only one of questions 3) and 4) is required (your pick), the other one is optional.

1. Principal Component Analysis
	1. Run a principal component to reduce the following multivariate data:

Insulin

Gluc\_fast

Triglyceride

HDL

LDL

* 1. Plot the variance of principal components (PCs) generated in a). Use the variance > 1 criteria to select PCs to use in the next analysis.
	2. What % of the variance is explained (alone and combined) by PCs chosen in part b)?
	3. Submit a biplot for PC1 and PC2. Based on the biplot identify which of the variables (Insulin, Gluc\_fast, Triglyceride, HDL, and LDL) contributes least in PC1, but most in PC2?
	4. Run a multiple regression to test whether PC1 and PC2 predict BMI24? Does PC1 or PC2 predict BMI24?
1. Cluster Analysis
	1. Run a hierarchical clustering to group subjects based on similarities on the following variables: BMI, Insulin, Gluc\_fast, Triglyceride, HDL, and LDL. Create the Dendrogram
	2. Identify two clusters on the Dendrogram by using “cutree” function. Save a cluster variable that indicates cluster membership for each subject.
	3. Describe the characteristics of each cluster by populating the following table

|  |  |  |
| --- | --- | --- |
|  | Cluster 1 (n=?) | Cluster 2 (n=?) |
| Insulin | mean±sd |  |
| Gluc\_fast |  |  |
| Triglyceride |  |  |
| HDL |  |  |
| LDL |  |  |

* 1. Use the t-test to test whether BMI24 different between two identified clusters?
1. Create a Heatmap for BMI, Insulin, Gluc\_fast, Triglyceride, HDL, and LDL. Create the Dendrogram. Use several option for optimal display.