**Analyzing Heart Disease Dataset**

Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_ Period: \_\_\_\_\_\_

**For each activity below:**

✔️ means to complete this task

✏️ means to write an answer here

**Investigating Heart Disease**

With access to a comprehensive heart disease dataset, you are part of a team that is investigating the relationships between various risk factors and the likelihood of a heart attack. Your team has created a dataset and the variables it includes are:

* age - Age of individual in years
* sex - Gender of the individual
  + 1 = Male
  + 0 = Female
* cp - Chest pain type from 4 different types
  + 1 = typical angina
  + 2 = atypical angina
  + 3 = non-anginal pain
  + 4 = asymptomatic
* trestbps - Resting blood pressure in mm Hg on admission to the hospital
* chol - Serum cholesterol measured in mg/dl
* fbs - Fasting systolic blood pressure > 120 mg/d.
  + 0 = False
  + 1 = True.
* restecg - Resting electrocardiographic results
  + 0 = normal
  + 1 = abnormal
  + 2 = probable or definite hypertrophy
* thalach - Maximum heart rate achieved measured in beats per minute
* exang - Exercise-induced angina
  + 1 = yes
  + 0 = no
* oldpeak - ST depression induced by exercise relative to rest. Does not have any levels.
* slope - Slope of peak exercise ST segment
  + 1 = upsloping
  + 2 = flat
  + 3 = downsloping.
* ca - Number of major vessels colored by flouroscopy, measured from 0 to 3.
* thal - Type of Thalassemia, a inherited blood disorder
  + 0 = normal
  + 1 = fixed defect
  + 2 = reversable defect.
* target - Likelihood of heart attack
  + 0 = low risk of heart attack
  + 1 = high risk of heart attack.

**You are tasked with studying 2 relationships:**

1. **The relationship between serum cholesterol level and resting blood pressure.**
2. **The relationship between serum cholesterol level and heart disease risk.**

# **Activity #1: Outline your analysis**

1. ✏️ Think about the task of studying if there is an association **between serum cholesterol level and resting blood pressure**. Using the knowledge from previous lessons, break down the task. Write down the steps you will take.

# **Activity #2: Use R to Perform Your Analysis**

1. ✔️ Save and load the Heart Disease dataset provided by your teacher into Posit Cloud.
2. ✔️ Plot serum cholesterol level and resting blood pressure.
3. ✏️ Write down your observations.
4. ✔️ Based on what you saw in step 2, test to see if there exists a meaningful association between serum cholesterol level and resting blood pressure. Use the techniques you learned in prior lessons.
5. ✏️ Write down your observations.

# **Activity #3: Analyze and interpret your tests**

1. ✏️ Analyze the results of the test(s) you performed. What did you conclude? Justify your conclusion(s).

# **Activity #4: Outline your analysis**

1. ✏️ Think about the task of studying if there is a relationship **between serum cholesterol level and heart disease risk**. Using the knowledge from previous lessons, break down the task. Write down the steps you will take.   
   Hint: Consider the differences between those who have a high risk of heart attack and those who have a low risk.

# **Activity #5: Use R to Perform Your Analysis**

1. ✔️ Plot and compare the serum cholesterol levels of those with heart disease risk and those without.

* Hint: You may need to create a new categorical variable based on the numerical “target” variable. If you are having trouble doing this, ask the teacher for help

1. ✏️ Write down your observations.
2. ✔️ Based on what you saw in step 1, test to see if there exists a meaningful difference in serum cholesterol levels between those with heart disease risk and those without. Use the techniques you learned in prior lessons.

* State your null and alternative hypothesis before you start.

1. ✏️ Write down your observations.
2. ✔️ Redo Activities #4-5 but use other variables that may be related to heart disease risk.

# **Activity #6: Analyze and interpret your tests**

1. ✏️ Analyze the results of the test(s) you performed. What did you conclude? Justify your conclusion(s).
2. ✏️ How do you think the knowledge and skills gained from this activity can be applied to real-world scenarios in the fields of medicine, research, or data analysis?
3. ✏️ Overall, how has this activity helped you develop a deeper appreciation for the importance of analyzing and interpreting data in the context of heart disease research? What other areas do you think data analysis could be valuable in the field of biology and medicine?