ACTIVITY 11

CAFFEINE AND HEART RATE

# DIRECTIONS

Have one group member upload this docx file to your Google Drive and share it with your groupmates and your instructor. Name your document “**GroupX\_Activity11”** (where X is your group number). Work together to type up your responses to each question. Download your document as a PDF and submit this to Canvas individually.

# INTRODUCTION

As you are likely aware, caffeine is a widely used stimulant and psychoactive drug found in many drinks that we consume, and has various effects on your body and health. Researchers collected data to attempt to quantify the effects caffeine has on resting heart rate. The researchers recruited individuals who drink a daily cup of coffee and were able to secure 50 volunteers. Each of these coffee drinkers were randomly assigned an amount of caffeine to be put into their drink. Their heart rate was recorded once before they were given their coffee and once again 1 hour after the drink was first consumed. Using this data, the researchers would like to answer the following research question:

*Is there a significant relationship between the amount of caffeine someone drinks and their heart rate an hour after drinking it?*

Data on the amount of caffeine given to the patients (mg) and the change in heart rate (bpm) can be found on the **caffeine.tp3** file.

# PART 1: EXPLORING THE DATA

1. Based on the context given above, which variable is the explanatory variable, and which is the response variable?
2. Based on your intuitions, do you believe there is a relationship between heart rate and caffeine? What do you believe the direction of the relationship is?
3. Create a scatterplot of caffeine and heart rate, putting the explanatory variable on the x-axis and the response variable on the y-axis. Describe the relationship between these two variables in terms of the overall shape of the relationship (linear, curved, etc.), the strength of the relationship, and the direction.

Recall that last class, we learned about using the **least squares** criterion to determine the best fitting line for a scatterplot. To find the slope and intercept of this line, we created two new columns in our data table, and then put the following functions into the formula editor to calculate each value:

linRegrSlope(Age, Distance)
linRegrIntercept(Age, Distance)

These formulas can be found in the formula editor under **Functions -> Statistical -> Two Attributes**. Remember to also be careful to choose the correct slope function, there is another function **linRegrSESlope** which is not the correct calculation!

1. Using the slope and intercept you found using the functions above, write out the equation of the least squares line below.
2. Use the diagonal reference line on your scatterplot to approximately draw the least squares line on your plot and paste that plot below. Does the least squares line here update or confirm your intuitions from question 2 about the relationship between caffeine consumption and heart rate?

# PART 2: CARRYING OUT THE HYPOTHESIS TEST

## CREATING A NULL MODEL

In order to determine if there is truly a relationship between caffeine consumption and heart rate, we will conduct a simulation and carry out a hypothesis test. To do this, we need to build a null model that assumes that there is no relationship between caffeine and heart rate in order to determine how likely it is to get data with as strong of a relationship as seen in our data set.

1. If these two variables were not related at all, what would you expect the slope of the least squares line to be?
2. Do you think the slope of the line we found signifies that the two variables are related, or do you think this could have just occurred by random chance? Explain.
3. Write out what the null and researcher’s hypotheses are for conducting this test.

## SIMULATING THE NULL MODEL USING TINKERPLOTS

1. Create a sampler in TinkerPlots that simulates this caffeine experiment under the null hypothesis and paste that model below. Be sure to explain the following:
	1. How you chose and named the attributes, an what labels you gave to the objects inside the devices.
	2. What replacement you chose for each device and why you made that choice.
	3. How your sampler is producing data assuming that the null hypothesis is true.
2. Simulate a data set based on your null model and use the *linRegrSlope* function again to get the slope from the line of best fit on this simulated data. Report that slope value here.

To collect statistics on any value, it must first appear on a plot. Currently, our slope value is in the data table, calculated from the formula editor, but we can get this onto a plot and collect it fairly easily.

Create a plot of the slope column that you calculated, and fully separate the dots. The plot itself shouldn’t be too interesting, as it’s just a bunch of dots with the same value. However, if you turn on the mean on your plot, you will be able to now collect on this slope value through this mean, as the mean of a set of identical values is just that value itself.



Right click on the mean triangle and then click “Collect Statistics” to begin repeating the simulation of your null model. **If you are using TinkerPlots 2, this method will not work directly – see the notes at the end of the activity for more assistance!**

1. Collect 500 statistics on your slope and then paste and image of the sampling distribution below. What does a dot in your plot represent?
2. What is the center of your sampling distribution? Does this value seem reasonable given the assumptions made in constructing your TinkerPlots sampler?
3. What is the *p*-value for the test? Paste your sampling distribution below with the *p*-value highlighted.
4. Interpret what this *p*-value means in the context of the problem.
5. Using this *p*-value, what do you conclude about the relationship between caffeine consumption and the change in heart rate?

# NOTE FOR TINKERPLOTS 2 USERS

Unfortunately, when trying to plot the slope, this happens on TinkerPlots 2:



In the plot to the right, you can’t get a number line or mean to appear on the x-axis, no matter how hard you try. When a column has all identical values, TinkerPlots thinks of this variable as categorical rather than numerical for some odd reason. Thanks for nothing, TinkerPlots. However, there is a way around this, and it involves tricking TinkerPlots a little bit.

**Step 1**: In the field that you want to collect statistics on, change the formula to something that varies. Here, I’ll just make it simple and use my **XVar**. (In this activity, you can use **Caffeine** or **Heart\_Rate** as your formula) The easiest thing to do will always be just to pick any of your attributes that have different values in each row.



**Step 2:** Make a new plot of your slope column. Since it varies now, TinkerPlots will plot it as a numerical variable. Drag one of the dots to the right, and a number line will appear on the x-axis. 

**Step 3:** Change the formula of your slope column back to the equation for the slope. And just like magic, TinkerPlots actually plots your equation as a numerical variable! You can click the mean triangle on the toolbar and collect statistics on the “mean” now.

