

NAME:

Statistics 111 Summer Session II

Homework Eight

You are allowed to discuss problems with other students, but the final answers must be your own work.

For all problems that require calculation, YOU MUST ATTACH SEPARATE PAGES, NEATLY WRITTEN, THAT SHOW YOUR WORK.

Please mark your answer in the space provided. As a general rule, each blank counts for one point unless otherwise specified. If necessary work is not shown, or if that work is substantially wrong, then you will not get credit even if the answer is correct. (The obvious purpose of this is to prevent students from mindlessly copying each other's answers.)

Report all numerical answers to at least two correct decimal places.

DUE DATE: START of class on Tuesday, August 7, 2018.

1. (**ANOVA**) Caffeine is the world's most widely used stimulant, with approximately 80% consumed in the form of coffee. Participants in a study investigating the relationship between coffee consumption and exercise were asked to report the number of hours they spent per week on moderate (e.g., brisk walking) and vigorous (e.g., strenuous sports and jogging) exercise. Based on these data the researchers estimated the total hours of metabolic equivalent tasks (MET) per week, a value always greater than 0. The table below gives summary statistics of MET for women in this study based on the amount of coffee consumed.¹

	<i>Caffeinated coffee consumption</i>					Total
	≤ 1 cup/week	2-6 cups/week	1 cup/day	2-3 cups/day	≥ 4 cups/day	
Mean	18.7	19.6	19.3	18.9	17.5	
SD	21.1	25.5	22.5	22.0	22.0	
n	12,215	6,617	17,234	12,290	2,383	50,739

- (a) Write the hypotheses for evaluating if the average physical activity level varies among the different levels of coffee consumption.

¹46M. Lucas et al. "Coffee, caffeine, and risk of depression among women". In: Archives of internal medicine 171.17 (2011), p. 1571.

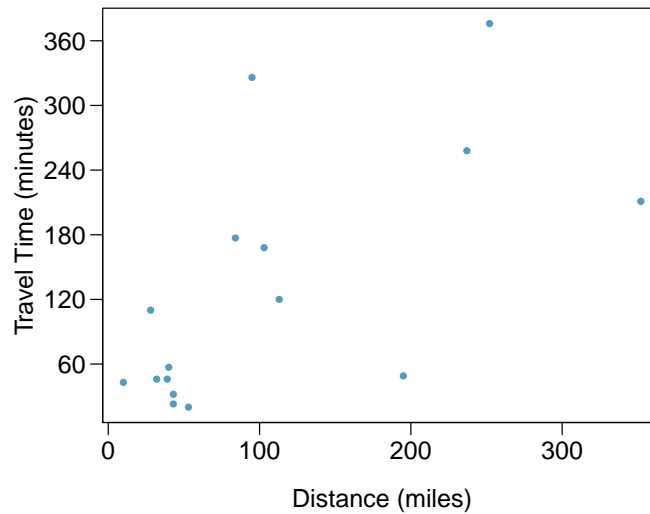
(b) Check conditions and describe any assumptions you must make to proceed with the test.

(c) Below is part of the output associated with this test. Fill in the empty cells (2 points).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
coffee	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	0.0003
Residuals	<input type="text"/>	25,564,819	<input type="text"/>		
Total	<input type="text"/>	25,575,327			

(d) What is the conclusion of the test?

2. **(Simple Linear Regression)** The Coast Starlight Amtrak train runs from Seattle to Los Angeles. The scatterplot below displays the distance between each stop (in miles) and the amount of time it takes to travel from one stop to another (in minutes). The mean travel time from one stop to the next on the Coast Starlight is 129 mins, with a standard deviation of 113 minutes. The mean distance traveled from one stop to the next is 108 miles with a standard deviation of 99 miles. The correlation between travel time and distance is 0.636.



- (a) Describe the relationship between distance and travel time.
- (b) How would the relationship change if travel time was instead measured in hours, and distance was instead measured in kilometers?

- (c) Correlation between travel time (in miles) and distance (in minutes) is $r = 0.636$. What is the correlation between travel time (in kilometers) and distance (in hours)?
- (d) Write the equation of the regression line for predicting travel time (2 points).
- (e) Interpret the slope and the intercept in this context.
- (f) Calculate R^2 of the regression line for predicting travel time from distance traveled for the Coast Starlight, and interpret R^2 in the context of the application.

(g) The distance between Santa Barbara and Los Angeles is 103 miles. Use the model to estimate the time it takes for the Starlight to travel between these two cities.

(h) It actually takes the Coast Starlight about 168 mins to travel from Santa Barbara to Los Angeles. Calculate the residual and explain the meaning of this residual value.

(i) Suppose Amtrak is considering adding a stop to the Coast Starlight 500 miles away from Los Angeles. Would it be appropriate to use this linear model to predict the travel time from Los Angeles to this point?

How hard was this homework for you and how many hours did it take you to finish it?
