Economics 30330: Statistics for Economics
Problem Set 2
University of Notre Dame
Instructor: Julio Garín
Spring 2012

Due Date: Beginning of class on Wednesday, February 8th. Please complete the assignment in the allotted space. You may work in groups, but you need to turn your own work. You may also draw all the figures by hand.

Descriptive Statistics (60 points)

1. Following a recent government shutdown, Minnesota Governor Mark Dayton proposes to give all state employees a $500 raise.

   (a) What would this do to the average monthly salary of state employees? To the SD? Explain mathematically as well as in your own words.

   (b) What would a 5 percent increase in salaries of all state employees do to the average monthly salary? To the SD? Explain mathematically as well as in your own words.

2. The Associated Press Team Marketing Report listed the Dallas Cowboys at the team with the highest ticket prices in the National Football League (*USA Today*, October 20, 2009). Data showing the average ticket price for a sample of 14 teams in the NFL are as follows.

Team	**Ticket Price** ($)	***z*-score**	**Team**	**Ticket Price** ($)	***z*-score**
Atlanta Falcons	72		Green Bay Packers	63	
Buffalo Bills	51		Indianapolis Colts	83	
California Panthers	63		New Orleans Saints	62	
Chicago Bears	88		New York Jets	87	
Cleveland Browns	55		Pittsburgh Steelers	67	
Dallas Cowboys	160		Seattle Seahawks	61	
Denver Broncos	77		Tennessee Titans	61	

(a) Compute the range, interquartile range, and the median ticket price.

(b) Compute the mean ticket price. The previous year, the mean ticket price was \$72.20. What was the percentage increase in the mean ticket price for the one-year period?

(c) Compute the sample variance and sample standard deviation. Interpret.

(d) Compute the standardized values (z-scores) for the listed ticket prices in the table above. Should the Dallas Cowboys' ticket price be considered an outlier? Explain.

(e) What are the mean and standard deviation of the standardized ticket prices? Explain.

3. The flashlight batteries produced by a Uruguayan manufacturer are known to have an average life of 60 hours with a standard deviation of 4 hours. Use Chebyshev's theorem to answer parts (a) through (c).

   (a) At least what percentage of batteries will have a life of 54 to 66 hours?

   (b) At least what percentage of the batteries will have a life of 52 to 68 hours?

   (c) Determine an interval for the battery lives that will be true for at least 80% of the batteries.

   (d) Suppose we know that the battery lives have a normal distribution. Approximately what percentage of batteries will have a life of 50 to 70 hours? Why?

   (e) Explain Chebyshev's Theorem in your own words.

4. The following is the frequency distribution for the speeds of a sample of Notre Dame students driving from South Bend to Chicago.

Speed (MPH)	f_i	M_i	f_iM_i	$(M_i - \bar{x})^2$	$f_i(M_i - \bar{x})^2$
50-54	2				
55-59	4				
60-64	5				
65-69	10				
70-74	9				
75-79	5				

(a) Complete the table.

(b) Compute the sample mean, variance, and standard deviation of the grouped sample.

(c) What does M_i represent and why do we use it?

5. The following table lists the study time and exam scores for a sample 5 students in a college statistics class.

Score	**Minutes Spent Studying**	**Hours Spent Studying**
60	60	1
80	180	3
75	90	1.5
95	240	4
85	225	3.75

(a) Calculate the sample covariance between test score and minutes spent studying.

(b) Calculate the sample covariance between test score and hours spent studying.

(c) Calculate the correlation coefficient between test score and minutes spent studying.

(d) Calculate the correlation coefficient between test score and hours spent studying.

(e) Why did we calculate two different measures to describe the relationship between time spent studying and test scores? Is one measure more useful than the other? Explain.

6. *You will need Excel, Stata, or another statistical programming software to complete this part of the assignment. Please calculate the answers in the software of your choice and report them here. You do not need to turn in your actual Excel spreadsheet, code, etc.*

   NBA teams are increasingly using statistics to inform their decisions on the court. One example is the analysis of "bad" players. David Berri defines a "bad" player as one that has half the Wins Produced per 48 minutes (WP48) value of a good player, or a WP48 of 0.050 or less; an average player has a WP48 of 0.100. The dataset named 'NBA' contains data on "bad" players for NBA teams in 2011.

   (a) Are we working with a sample or a population here? Justify your answer.

   (b) Calculate and interpret the following measures for *Bad Players*: mode, median, mean, and standard deviation.

   (c) Comment on the skewness of the data for *Bad Players* based on the measures you calculated in (a) (note: you do not actually need to calculate the skewness coefficient). What does the skewness tell us about the distribution of "bad" players?

   (d) Create a scatterplot of *Total Bad Player Minutes* and *Total Wins* (print and attach a copy of the scatterplot to your assignment; be sure to label the axes). Based on the plot, what do you expect the direction and approximate magnitude of the correlation between these two variables to be? Explain.

   (e) Calculate the covariance and correlation between *Bad Player Minutes* and *Total Wins*. What do you expect would happen to the covariance if you instead calculated *Bad Player Hours* instead of *Minutes*?.