**EDRS 6318**

**Statistical Data Analysis**

**Project # 1**

**Summer Session 2021**

**Chapters Material form 1-9 Olivárez**

1. Using the data file “*CarsDataSyn2921.sav*”, conduct the following analyses and interpretations of the results. Convert this Excel file into an SPSS file by importing at the SPSS application. There are 302 cases on this new data set with eight column variables with the following brief description:
   * 1. **Case#** Number of vehicle cases
     2. **MPG** Miles per gallon for each selected car
     3. **Weight** Vehicle’s total or gross weight (in pounds)
     4. **Year**  Vehicles year model (range from 1970 to 1982)
     5. **Origin** Vehicle manufacturing country of origin
     6. **Cylinder** Number of cylinders in the engine
     7. **ModelGRP** Years/Model Periods of vehicle’s manufacturing (1=Before Oil Embargo (<1975) and 2= After Oil Embargo (>1975))
     8. **Satis** Consumer satisfaction ratings (0=”Not satisfied at all” to 10=”Extremely satisfied”)

Note: You may use chart builder or other related procedures to do all you data screenings for the various task in this project.

1. Determine the level of measurement for each of the 8 variables for this dataset. Explain your reason for selecting the level of measurement given that some of the variables may play multiple measurement roles. Based on the definitions for continuous and discrete variables by the text authors, reclassified these variables into these two broad categories. In addition, based on this small pool of variables in this dataset, generate a hypothesis illustrating a correlational study and a hypothesis illustrating an experimental (comparative) study.
2. For the miles per gallon (MPG) variable, compute the most common central tendency measures (i.e., mean, median, and mode) for the entire data set (n = 302) as well as the same central tendency measures across country of origin (Origin), type of engine (Cylinder) and model period variables (ModelGRP). Which type of car produced the best miles per gallon results and from what country did this vehicle came from (Origin)? Interpret these results by making appropriate comparisons across country of origin and the other grouping variables. Assuming for this questions that the total sample (N = 302) is the population, perform the following tasks and interpret. For the customer satisfaction scores (Satis), compute the mean, standard deviation, and standard error of the mean for the total sample and across number of cylinders. Generate a random sample of size n = 75 cases (about 25 percent of the total sample). Perform the same analysis on the satisfaction variable on this new subsample (mean, std. dev. & std. error) and across number of cylinders. Compare the subsample statistics in relation to the total sample and number of cylinders variable with the results observed with the total sample parameters. What can you interpret between these statistical estimates (n = 75) versus the presumed parameters (n = 302) on the Satis variable and the breakdown of the Satis variable across number of cylinders. For SPSS use the Select Cases option under the Data menu and follow the steps. Here is where you tell SPSS to generate about 25 percent of the total sample.
3. Assuming now that the total dataset (n = 302) is just a sample, perform the following tasks where the results will be statistical estimates and not parameters. For the miles per gallon variable (MPG), compute the range, interquartile range variance and standard deviation for the entire data set(n = 302) as well as the same variability measures across country of origin, type of engine and model period variables. Which type of engine produced the more variability in the consumption of gasoline for these vehicles? Which car’s model period or group yielded the highest variability in MPG? Interpret these results in relation to the average values observed in the previous problem for the total sample and the country of origin variable.
4. Produce a frequency distribution (ungrouped) for the MPG variable for the entire set of cases (vehicles). Do the same analysis but now generate a frequency distribution for each case by country of origin. Provide the plot or plots for the MPG frequency distribution variable across country of origin. Do the plots or graphs provide any surprises or made these vehicle’s MPG levels a bit clearer? Interpret these results in terms of the concepts of symmetry, skewness and/or kurtosis as described in chapter 2. What is the percentile rank for a vehicle that has 28 miles per gallon gas consumption? Describe the vehicle(s) that are at the 90th percentile in terms of miles per gallon in relation to other group variables such as ModelGRP. How does this information help us decide on the type of car that we should be purchasing next?
5. Convert all the MPG scores to z-scores using the overall grand mean and standard deviation from the 302 cases in this dataset and treat them as parameters. Compare and contrast the distributions for the unstandardized (raw data) and standardized (z-scores). You may use any graphical device in the software you are using. Provide an explanation as to what does this standardization of scores help us explain for this variable? What has been gain by conducting this linear transformation? By using the standardized scores distribution for the MPG variable and splitting the file across car manufacturing countries, what are some general descriptive information (central tendency and variability) that you can report about them and the different potential differences (i.e., Ms or SDs) without using inferential statistical tools or procedures. Is there a car manufacturing country (Origin) where their produced cars appear to outperform other manufacturing companies in this dataset?
6. Assuming that the outcome variable MPG is derived from a normal distribution (You can transform it by using the square root function or any other function) and using the mean and the standard deviation from this dataset as the population parameters, what is the probability that a given manufactured car from the United States will yield a vehicle with a 25 miles per gallon (MPG) output? Conduct similar probability computation for a Japanese car? Interpret these results in the contest of the total dataset and these two countries.
7. Determine the overall sample dataset standard error of the mean for both MPG and Satis variables as well as the respective standard error means for the total sample across each car manufacturing country in this dataset. Interpret these results in the context of a respective central tendency descriptive statistics such as the means and standard deviations.
8. Build a variable using the RECODE option in SPSS with the variable **Origin** and dichotomize it into New Origin (new variable) with two categories USA and Non-USA. Obtain the mean, standard deviation, and the standard errors for these two groups and construct a graph incorporating the obtained distributional information. Examining the graph, do you think that these two group of cars’ MPG differ significantly from each other (Do not do a hypothesis test or any other procedure to explain these results).
9. Using the total dataset; i) determine if there is a statistical significance mean difference between expected miles per gallon MPG ratings for this sample of data (use the population parameters μ = 0 and μ = 22 mpg for your null hypothesis values; This is typical mpg for most cars and included city and highway driving) Are the cars’ MPG in this sample statistically significantly different from the population of cars’ mpg parameter of 22 mpg? What is the 99 percent confidence interval for this problem? Interpret along with the test of significance. What happens if the parameter is μ = 25 mpg? Use a level of significance of .05. How about if you were to use an alpha of 0.01?
10. Using the information from part i), perform the same analysis using a non-parametric procedure for these variables using the default null hypothesis of μ = 0 and a level of significance alpha equal to .05. Explain these results for both the .05 and .01 level of significance.
11. Using the total dataset compute and interpret the intercorrelations among the following variables: MPG, Satis, and Weight. Do we observe the same type of intercorrelations if we look at these variables across countries of origin? Interpret without making any causal links among the variables analyzed. Is the linearity assumption among these variables met? Explain. Can you make causal statements as to what you are observing from this extant results? Why or why not?
12. Using the level of efficiency in gas consumption (MPG) by these cars as a predictor of consumer satisfaction (Satis), perform a linear regression analysis and explain the level of prediction that the MPG variable has on someone planning to purchase a car in the future using the obtained linear equation given that the two new car options indicate having a 20 mpg and another one a 28 mpg. What is these consumers level of satisfaction? Check for assumptions required in this type of analysis. You may use scatter plots to talk about these assumptions. What is the difference between prediction and explanation in the context of this linear regression problem?
13. Run all these analyses using SPSS 25 or higher version or other similar package that you prefer so that you can apply the concepts learned in these chapters as you worked the many homework assignments performed by hand. Remember that SPSS uses (n – 1) for all calculations dealing with the variance or standard deviations as the default. You will need to instruct it to use (N) if the need arises. Some hand calculations may be needed but these are kept to a minimum. Additionally, you may submit an electronic version of your work or a hardcopy with selected output placed in the narrative of these items.
14. Caution: Although, this is the same set of problems for everyone and collaboration is expected and encouraged, however, the interpretation of the results for this project is more at the individual level. Let me know if you have any specific problems with the data set and program set up.

Check Blackboard for new due dates!!!!