

SPSS Project 5

- I. Using the data file “*CarsDataSum2021.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:

i. Case#	Number of vehicle cases
ii. MPG	Miles per gallon for each selected car
iii. Weight	Vehicle’s total or gross weight (in pounds)
iv. Year	Vehicles year model (range from 1970 to 1982)
v. Origin	Vehicle manufacturing country of origin
vi. Cylinder	Number of cylinders in the engine
vii. ModelGRP	Years/Model Periods of vehicle’s manufacturing (1=Before Oil Embargo (<1975) and 2= After Oil Embargo (>1975))
viii. 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.

- a) Using the total dataset and the new vehicle car manufacturing countries (USA v. Non-USA); i) determine if there is a statistical significance mean difference between expected miles per gallon MPG ratings for this sample of data across new country of origin grouping variable. Use a .05 level of significance. Would your observed results and conclusion have been different if the alpha level had been .01 instead? What are the confidence intervals for these two levels of significance? Do your results from the confidence intervals match your decisions using the test of significance? Why? Do check for the homogeneity of variances assumption on this problem. Interpret your results. Although SPSS uses a t-test command, you are in reality conducting a z-test of significance. Why? Perform same problem variables in this task using a non-parametric statistical procedure. Interpret both results.
- b) Use the total dataset test and omnibus F (GLM 1 or ANOVA) as to how different are the Miles Per Gallon in selected vehicles across manufacturing countries (USA, Germany and Japan). Use an alpha level of 0.025. However, conduct a check for the homogeneity of variances assumption ($H_0: \sigma^2 = \sigma^2 = \sigma^2$) using the Levene’s test. What kind of results does this test yielded and what can you do with them. If you checked the assumption of homogeneous variances, but did not meet it, proceed with the ANOVA test anyway and explain as to what can happen if this assumption is violated? What do we mean that the F-test is a robust test?
- c) If the F-test is significant in item (f) above, perform the Scheffé test of multiple comparisons to determine where the significance is located. Use an alpha of 0.05 to tests these mean differences. Which country appears to produce the most gas efficient vehicles and to what attributes on these vehicles do you attribute this efficiency? Does the effect size (eta-squared) agrees with the F-test results? Interpret overall results including a plot of these MPG means.

- d) If you were to compare USA vs. Non-USA vehicles using the t-test for independent samples and using a 0.025 level of significance, then compare this finding with the results observed when you ran the same variables using the GLM 1 or ANOVA approach for these two groups. GLM I does under SPSS does all analyses under a 0.05 level of significance. Compare the t-observed and the F-observed value as well as levels of significance and provide an overall interpretation of these two results.
- d) Again, test the significance or difference in MPG means using the Origin (all three car manufacturing countries) variable by comparing them with the following contrast: (C_1 : 1, -0.5, and -0.5). Test this contrast mean difference using an alpha of 0.05. Do you get the same results? Why or why not? Again, indicate or interpret the obtained results as to the most efficient car manufacturing country. Any other interesting contrast on mean differences that you can think of? You do not need to ran this contrast just provide one that may be of interest.
- e) Conduct the same GLM 1 or ANOVA analyses using the dependent variable Customer Satisfaction levels (Satis) across the three levels of the Origin independent variable. Use a .05 level of significance. Check the appropriate assumptions for this outcome variable and report accordingly (i.e., robust analyses) the results based on these preliminary finding. Conduct the non-parametric procedure Kruskal-Wallis (K-W) for the above problem and compare the results between GLM I and K-W. In addition with this new analysis, explain or interpret your results on the Satis variable by merging what you observed for the MPG (miles per gallon) variable in terms of efficiency and the customer satisfaction (see results from item (q)). Can we conclude that customers of more efficient cars were more satisfied with their vehicle purchase or not?
- f) Conduct a GLM III or factorial ANOVA for two independent variables using the Origin and Cylinders to determine the significance on these three hypotheses (i.e., two main effect + interaction effect), use a level of significance of 0.025 to reject any of these hypotheses. Remember to check for the appropriate model assumptions for GLM III and report regular or robust factorial ANOVA results accordingly. For the interaction effect, use graphs rather than simple main effect procedures to interpret this result. Interpret all these observed results incorporating the eta-squared and effect sizes (SPSS does not provide effect sizes, formulas will be discussed in class).
- g) (Optional problem). Conduct a GLM II or ANCOVA analyses for the dependent variable miles per gallon MPG in relation to the covariate variable Weight (overall car's weight) in relation to the grouping (IV) variable Origin. Explain what the limitations of the analysis are, if any of the typical assumptions are not met. Once you have conducted these tests, proceed to check for overall level of significance and determine which country appears to produce or

manufacture the most gas efficient automobiles. Is there a need to report results using robust ANCOVA procedures due to the lack of meeting some of the assumption? In addition, using GLM II provide evidence of where this significance is derived by using the following contrasts ($C_1: 1 -1 0$; $C_2: 1 0 -1$; and $C_3: 0 1 -1$). Use Tukey HSD to compare the above contrast results and explain these findings. What is the overall practical significance using eta-squared to describe the amount of variance explain on MPG by the adjusted groups. Determine if the homogeneity of slopes variables is met.

Run all these analyses using SPSS 25 or higher version 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.

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.