



BALL STATE
UNIVERSITY

Practical Assignment II

Analysis of Variance

Quinton Quagliano, M.S., C.S.P

Department of Educational Psychology

Table of Contents

1	Pre-assignment Reminders	2
2	Context	2
3	Instructions	3
3.1	Knowledge Checks (6 pts)	3
3.2	SPSS Applied Work (14 pts)	3

1 Pre-assignment Reminders

- Prior to attempting this homework, I strongly recommend that you have completed all other work through module 7, including:
 - Likely the most useful resource will be the video I specifically record for this assignment (near the assignment portal on Canvas)
 - Having watched all lectures thoroughly and taken notes (optionally with the guided notes)
 - Reading all assigned sections of the book, and completing practice examples as helpful and useful
 - Completing relevant lecture check-ins, and using answer key for review
 - Completing all quizzes, and reviewing correct answers at the end of the week
- This practical assignment will be graded for accuracy, please give yourself enough time to complete your best work
 - You may use your notes, book, SPSS software, provided data, and lectures to aid you in completing this
 - Where possible, include more detail to ensure you fully explain your rationale for a question
 - You should include ALL relevant SPSS output to support your work, please make clear where in the output you get your answers from
- This practical assignment is cumulative to the modules and practical assignment(s) prior, please review your previous work and notes to help you accomplish this work
- All syllabus and university policies on academic integrity, plagiarism, and other forms of misconduct apply to this assignment. Please review them if you are unfamiliar

2 Context

The following information will be useful for answer some questions on the homework, please ensure you read and understand the following context

You are gathering retrospective, archival data on graduates from a local large high school district. The population of interest is all the graduates from this district within the last 10 years. You have been given a dataset that should contain all of these individuals.

Your goal is to use your skills in SPSS and growing statistical knowledge to analyze the distributions of the variables.

Please use the provided dataset on Canvas, `hs_grad_data.sav`, to complete this assignment. Assume this data file contains records for any and all graduates in the population of interest.

3 Instructions

3.1 Knowledge Checks (6 pts)

1. Given that we have included data from ALL the members of the population, we do not do any sort of sampling. Instead, this is called a c_____ (hint: that “c” at the start is intentional, fill-in-the-blank; 1pt)
2. Describe the key differences between discrete and continuous numerical variable; make clear their distinct characteristics (1pt)
3. Identify the one special distribution of discrete variables and 3 special distributions of continuous variables that we have discussed in the lectures and write the formula notation for each of them (2pt)
4. Explain, in your own words, the concept of the central limit theorem. Please, at a minimum, touch on how it relates to statistics and parameters, and how it is impacted by sample size (2pt)

3.2 SPSS Applied Work (14 pts)

For each of the following questions, please provide any and all relevant SPSS output (or syntax when indicated). Follow the directions IN ORDER

1. The dataset accidentally includes some individuals who did not actually graduate, please filter and delete them out so that you only have the individuals or graduated. Show the SPSS syntax used to perform this filter and delete process. Assume that those left over after this process represent the entire population of interest. (2pt)
2. The salary variable was encoded with an error, where all of them are accidentally doubled. Compute a new variable that fixes this error by dividing them by half, and delete the original erroneous salary variable. Show the SPSS syntax used to perform this compute and deleting of a variable. (1pt)
3. Determine how many variables (if any) and how many constants (if any) are in the dataset. List all of them out. (1pt)
4. How many total records/rows are there in the data, i.e., the n of the sample? (1pt)
5. For each of the variables, identify their level/scale-of-measurement (1pts)
6. For each of the continuous variables, determine the mean, median, variance, standard deviation, please write out each of these with their correct notations (be mindful of correct ways to represent population parameters!) (2pts)
7. For each of the continuous variables, make a histogram to represent its distribution, and identify, based on the histogram, which of the continuous distributions it is best

represented by - please try to explain why - it okay to appeal to the shape of the graph (2pts)

8. For the variable you determine to be normally distributed, filter a random sample of 10 cases, and make a histogram of those cases (MAKE SURE YOU TURN OFF THE BUTTON TO DELETE RECORDS FROM THE FIRST STEP). Repeat this process for 20, 30, 40, and 50 cases, showing the histogram for each of these. Make a determination if/when one of these samples appears to be roughly normally-distributed. Explain why smaller samples do not show the same normality that the population does (4pts)