

Quantitative Methods for Educational Leadership
EEPL614
Fall 2021

Higher Education Homework #1

This assignment requires you to use EXCEL (or SPSS) to analyze data and interpret the results. The data is provided as an EXCEL file on the course's Blackboard site as **EEPL614 Homework_HE Data.xls**. The data you need will be in the sheet named **HE Homework F21**. If you do not see this sheet, you have the wrong version of the data. **It is recommended that you start each assignment with the original data!!!!**

For each assignment, your report must contain (a) a copy of EXCEL (or SPSS) output summarizing the results of your analysis, (b) your work calculating statistics when requested, and (c) the interpretations of your results. You may compare your EXCEL (or SPSS) OUTPUTS, but each person **MUST** do his/her own EXCEL (or SPSS) runs **AND** write up the results in his/her own words.

When you are asked to **CALCULATE** some simple statistics, it is important that you show your work. (If "calculate" is not bolded and in capital letters as above, you do not need to show your work although it might be helpful in getting full points on your assignment.) It is not required but recommended that, when possible, you also check your results against the same statistics obtained by using EXCEL. Regardless of whether you or the computer does the calculations, use at least **3** decimal points unless specified differently.

Give **BOTH** the empirical results **AND** substantive interpretations of your results. Also, it is not enough to say "see the results" in computer output or a table. Justify these interpretations by referring to specific statistics or results!!! Also, be sure to fully answer the questions!

Purpose of Analysis for this Assignment:

Recently, the faculty and administration union has approached Dr. Smith, President of Verbatim State Community College System (VSCCS), concerned about pay inequity between its male and female members. Before addressing their concerns, Dr. Smith decided it is important to determine the extent to which male and female faculty and professional staff in the system's three regions' districts receive different salaries, and what factors might explain those differences.

A description of the data compiled for this purpose is contained in *Higher Education Homework Dataset Codebook*. The purpose of the analysis for this assignment is to become familiar with the variables contained in the data set by obtaining basic descriptive statistics. The results will be suitable for the first preliminary monthly progress report summarizing the content of the data set.

Each of the 4 parts of this assignment are worth 25% of the assignment grade.

1. Examine the distribution of gender and levels of education in the sample.

Both of these variables are used to categorize a characteristic of individual faculty and professional staff members. Because we want to find out some basic information about those categories, these analyses require you to create pivot tables using EXCEL. (See the exercises for Class #2 for the procedures.).

Part A: Obtain the following EXCEL output (10 points):

I. The frequencies and relative percentages for the variable **MALE**.

	Data	
	Count of MALE	Count of MALE2
0 Females	266	66.83%
1 Males	132	33.17%
Grand To tal	398	100.00%

The frequency of the variable male is 132 and the relative percentage is 33.17%

II. The frequencies, relative percentages, and cumulative percentages for the variable **DEGREE**.

0 = bachelors, 1 = masters, 2 = certificate of advanced study, and 3 = doctorate.

	Data		
DEGREE	Count of DEGREE	Count of DEGREE2	Count of DEGREE3
0 Bach	37	9.30%	9.30%
1 Masters	257	64.57%	73.87%
2 Certi	80	20.10%	93.97%
3 Doc	24	6.03%	100.00%
Grand Total	398	100.00%	

Commented [OA1]: @Schiller, Kathryn S is this correct

Commented [SS2R1]: I assume the table is correct (I don't have the output here), which is what is asked for in Part A. However, the variable for "gender" is named "MALE" to indicate which value is "1." Remember, variables MUST VARY! You do answer the question correctly in Part B.

Part B: Answer the following questions based on your EXCEL output (10 points):

I. Based on the results for **MALE** obtained using EXCEL in Part A.I, describe the gender characteristics of the sample. In particular, answer the following questions: How many and what percentage of the 398 faculty and professional staff in the sample are male? What number and percentage are female?

a Out of the 398 Faculty and professional staff in the sample, 132 are male. The percentage of the 398 faculty and professional staff that are male in the sample is 33.17%

i $132 / 398 \times 100 = 33.17$

b Out of the 398 faculty and professional staff in the sample 266 are female and the percentage of females is 66.83%

i $266 / 398 \times 100 = 66.83$

II. Based on the results for **DEGREE** obtained using EXCEL in Part A.II, describe the distribution of degrees in the sample. In particular, answer the following questions: What does the number "80" in the output indicate? What degree do the largest number of faculty and professional staff in the sample have and what percentage of the sample do they make up? How many and what percentage of the sample have a degree higher than a master's? Be sure to state what numbers in the EXCEL output provided the needed information.

a The number "80" indicates the faculty and professional staff members that have earned a certificate of advanced study.

b The largest number of faculty and professional staff in the sample have a master's degree (257) and the percentage of the sample that has to master is 64.57%

i $257 / 398 \times 100 = 64.57$

c 104 (26.10%) of the faculty and professional staff have a higher degree than a master's degree.

i $80 + 24 = 104 / 398 \times 100 = 26.13\%$

Part C: (5 points) Write a brief (1 or 2 paragraphs) description of the gender composition and distribution of educational degrees in the sample that would be suitable for inclusion in a report to Dr. Smith. Be sure to describe the analysis in grammatically correct everyday language, include substantive interpretations of appropriate statistics (also giving the numbers in the text) for both variables and discuss the basis of your judgements

Commented [SS3]: Correct & Complete. All that is needed.

Commented [SS4]: Not needed because you are not asked to CALCULATE. But, ok to check your understanding of the numbers.

Commented [SS5]: Missing answer.

Commented [OA6R5]: Please check the bottom I've added the answer

Commented [SS7R5]: On any assignment, be sure to provide a "complete" answer by responding to all parts.

Commented [OA8R5]: Understood! I will be sure to do that. Thank you.

Commented [OA9]: Redo this again

concerning the importance of the results. For example, did you find the distribution of gender or education in this sample unusual? Explain your answer. How might this sample of VSCCS faculty and professional staff be similar to or different from those in New York or another area on these two variables? Finally, suggest at least one reason why each characteristic might be important for the union discussions.

First of all, Dr. Smith President of Vertim State Community College System has been approached due to a worry or concern regarding pay inequality between the male and female members of the school. Dr. Smith prioritizes analyzing the extent to which male and female faculty and professional staff that are present in the three regions district get different salaries, and what highlights those differences. Referring to the table with frequencies and relative percentages for the variable MALE. The number 0 indicates “FEMALE” in

2. Determine the overall distribution of salaries in the sample.

As a scale variable, **SALARY** is used to measure relatively fine-grained differences between cases, which can sometimes make seeing patterns and trends difficult. For this question, you will use two common statistical procedures to summarize and visualize the distribution in teachers’ salaries more clearly.

Part A: Obtain the following EXCEL output (8 points):

- I. Using “**Descriptive Statistics**” under Data/Data Analysis obtain basic descriptive statistics (using at least 3 decimal points) for **SALARY**.

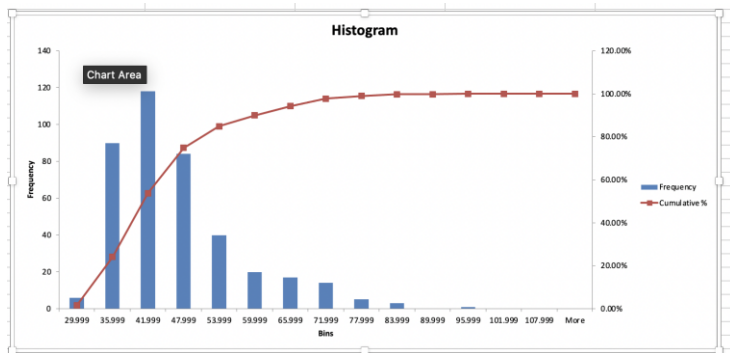
SALARIES	
Mean	43.8208794
Standard Error	0.534276538
Median	40.9425
	44.33
Standard Deviation	10.65878345
Sample Variance	113.6096646
Kurtosis	1.883731364
Skewness	1.360394416
Range	63.523

Minimum	26.842
Maximum	90.365
Sum	17440.71
Count	398

- II. Using “Histogram” under Data/Data Analysis, obtain a table showing the frequencies and cumulative percentages for **SALARIES** categorized by cutpoints every **\$5,000** with the lowest bin having a highest value of **29.999**. (Remember: the cut-points should be the highest value – e.g., 29.999 for the first one – to be put in that bin. Also, do not check “Pareto” in the dialogue box.)

<i>Bins</i>	<i>Frequency</i>	<i>Cumulative %</i>
29.999	6	1.51%
35.999	90	24.12%
41.999	118	53.77%
47.999	84	74.87%
53.999	40	84.92%
59.999	20	89.95%
65.999	17	94.22%
71.999	14	97.74%
77.999	5	98.99%
83.999	3	99.75%
89.999	0	99.75%
95.999	1	100.00%
101.999	0	100.00%
107.999	0	100.00%
More	0	100.00%

Commented [SS10]: Error. Should be 34.999 to be \$5,000 higher than \$29,999.



Part B: Based on the EXCEL output from above, answer the following questions (10 points):

- I. Answer the following questions based on the EXCEL output from Part A.I. How is the number “40,943” in the statistics box interpreted? What is the lowest and highest salary in the sample? Explain how you located the needed information in the output.

The number 40.943 in the statistics box is the median salary of the faculty and professional staff median is known as the middle value in a range of numbers.

The highest salary is 90.365 and the lowest is 26.842. This answer was found on the descriptive statistics and is labeled minimum and maximum.

- II. Answer the following questions based on EXCEL output from Part A.II. How many and what percentage of sample members earned salaries of \$55,000 or more but less than \$60,000? How many earn salaries of \$70,000 or more? CALCULATE the percentage of sample members who have salaries of \$30,000 or more but less than \$50,000? Explain how you located the needed information in the output.

Bins	Count
53,999	40
59,999	20
40+20=60/398x100=% 15.08	

Commented [SS11]: The bin cut-points not lining up with these numbers should indicated that the numbers are incorrect.

Part C: (7 points) Write a brief (1-2 paragraphs) description of your analyses that is suitable for inclusion in the first interim report to Dr. Smith. Be sure to describe the analysis in grammatically correct everyday language, include substantive interpretations of appropriate statistics (also giving the numbers in the text) obtained for this question, and discuss the basis of your judgments concerning the importance of the results for union

negotiations. Include a discussion of whether, in your opinion, this sample of VSCCS faculty and professional staff are well paid, clearly explaining the basis of your claim.

3. Determine the overall distribution of years of service to the consortium in the sample.

As a scale variable, **YEARS** is used to measure in the years that each faculty or professional staff member has worked in the community college system. Capturing relatively fine-grained differences between cases, patterns and trends in the distribution of scale variables are summarized using “descriptive statistics.”

Part A: Obtain the following EXCEL output (6 points):

Using “**Descriptive Statistics**” under Data/Data Analysis obtain basic descriptive statistics (using at least 3 decimal points) for **YEARS**.

Include in your output the lowest number of years of experience for the top quartile of the sample and the highest number of years of experience for the bottom quartile.

YEARS	
Mean	11.18341709
Standard Error	0.329546667
Median	10
Mode	10
Standard Deviation	6.574435366
Sample Variance	43.22320038
Kurtosis	1.064634286
Skewness	0.872893812
Range	38
Minimum	0
Maximum	38
Sum	4451
Count	398
Largest(100)	15
Smallest(100)	7

Part B: Based on the EXCEL results above, answer the following questions (12 points):

- I. The number “4451” in the row labeled “Sum” in the “Descriptive Statistics” output from Part A.I indicates what? [This number is used in calculating which measure of central tendency?

The number “4451”, the sum of years, stands for the total number of years that all faculty or/and professional staff members have worked in the community college system.

The number 4451 is used in calculating the mean. The mean is also known as the average. (Mean: Average; most widely used measure of location or central tendency) (Vogt, Johnson 2015).

Commented [OA12]: @Schiller, Kathryn S What does this mean I'm a little confused

Commented [SS13R12]: The question is asking, "total number of years" is used to calculate which of the following measures of central tendency -- mean, median, or mode?

- II. What percentage of the sample has “15” or more years of experience? The “7” next to “Smallest” indicates what about the sample?

27.1% of the sample has 15 or more years of experience. ($108/398 \times 100 = 27.1\%$)

Commented [SS14]: The question is asking you to indicate why 15 and 7 were used in the Descriptive Statistics procedure.

Commented [OA15R14]: 15 stands for the largest number of years of experience and 7 stands for the smallest number of years of experience.

- III. Identify and interpret for this variable three measures of central tendency from the EXCEL results in Part A?

Mean	11.18341709 On Average
Median	10 (Since Median is the middle value, in this situation 10 stands for the middle value of years of experience - Years worked in the system)
Mode	10 (Since Mode refers to the most commonly occurring value in a distribution, in this situation the most common years worked in the system is 10.)

Commented [SS16R14]: According to your output, the minimum is 0 years and maximum is 38. Recall how you decided on entering "15" and "7" into the EXCEL Descriptive Statistics dialog box to get this output.

Commented [OA17]: The 7 indicates the smallest number of years of experience.

Commented [SS18]: "Interpret" asks for you to use the term in a sentence -- e.g., on average, the sample had worked in the system for just over 11 years.

- IV. Identify and interpret for this variable three measures of spread, or diversity, from the EXCEL results in Part A? How can you calculate the standard deviation from the sample variance?

Standard Deviation 6.574435366

Sample Variance 43.22320038
Range 38

In order to calculate the standard deviation, you have to divide by the size of the data set minus 1, which is ($n - 1$).
The next step is to take the square root of the sample variance to find the standard deviation. That is how you get the standard deviation from the sample variance.

- V. How would you expect the mean and standard deviation to be affected by eliminating the cases with more than 30 years of experience (i.e., if staff members were forced to retire after 30 years)? Explain briefly your reasoning for the anticipated changes due to deleting these cases.

Commented [OA19]: Ask Professor

YEARS

Mean	10.85714286
Standard Error	0.306023698
Median	10
Mode	10
Standard Deviation	6.058960096
Sample Variance	36.71099744
Kurtosis	0.231063072
Skewness	0.593723373
Range	30
Minimum	0
Maximum	30
Sum	4256
Count	392

Part C: (7 points) Write a brief (1-2 paragraphs) description of your analyses that is suitable for inclusion in the first interim report to Dr. Smith. Be sure to describe the analysis in grammatically correct everyday language, include substantive interpretations of appropriate statistics (also giving the numbers in the text) obtained in this question, and discuss the basis of your judgements concerning the importance of the results. Discuss

both the central tendency and diversity in years of service among this sample of faculty and professional staff. Briefly discuss whether, in your opinion, this sample suggests that VSCCS may do a good job of attracting or retaining experienced staff members.

4. Examine the distribution of salaries for only faculty members.

Because upper-level administrators may have been included in this sample, Dr. Smith also wants analyses of annual salaries for only faculty members in the sample. She is particularly interested in the probability that faculty members have salaries in various ranges. Prior to your involvement in the project, another analyst found the following descriptive statistics for this subgroup: **mean = 41,579, standard deviation = 7,39, and $n = 316$** . Your predecessor estimated that approximately **26.8%** of sampled faculty members had salaries below **\$37,000** (based on a Z-score of **-.619**). She also estimated that **a third of faculty members had salaries below \$38,327 and a third had salaries above \$44,831**. (You are encouraged to test your ability to use the Z-score formulas by obtaining these numbers yourself.)

Now Dr. Smith wants estimates for different benchmarks: (a) the percentage of faculty members had salaries **above \$52,000** and (b) **the top and bottom 20%** of salaries. Be sure to show your work for any **CALCULATIONS** to **at least** the **third** decimal point.

Part A: Obtain the following output and **CALCULATE** the following numbers (8 points):

- I. **CALCULATE** the appropriate Z-score using at least **3** decimal points for estimating the probability that a faculty member has a salary over **\$52,000**. Find the appropriate probability using a normal probability table (e.g., Levine Table E.2).
- II. Use a normal probability table to find the appropriate Z-scores for **the bottom and top 20%** of salaries for faculty members. **CALCULATE** the cut-point for the upper boundary of the bottom 20% and the bottom boundary of the top 20%.

Part B: Based on Part A results, answer the following questions (10 points):

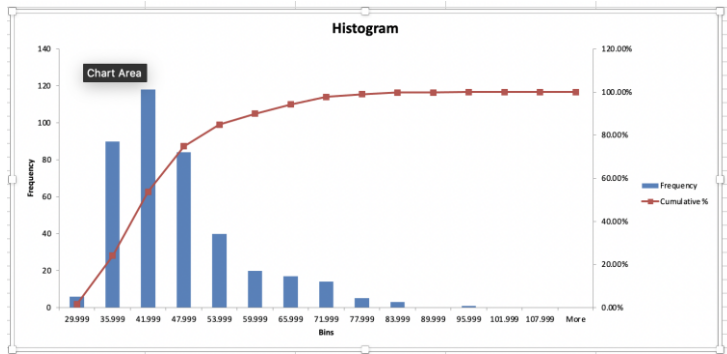
- I. What does the Z-score from A.II above indicate about how many standard deviations a salary of \$52,000 is above or below the mean salary? What is the probability that a faculty member had a salary less than \$52,000? Briefly explain.
- II. If the districts specify no minimum salary, what is the lowest possible salary if another sample of faculty members were drawn? If you wanted to make similar estimates about the distribution of years of service, what Z-scores would be used estimate the upper and lower 20% of years worked in VSCCS? Briefly explain.

Part C: (8 points) Write a brief (1-2 paragraphs) description of conclusions that might be drawn from analyses in Part A.I-II above concerning the distributions of salaries among this sample of VSCCS faculty members. Make this text suitable for inclusion in the first interim report to Dr. Smith by describing the analysis in grammatically correct everyday

language, include substantive interpretations of appropriate statistics (also giving the numbers in the text) obtained for this question, and discuss the basis of your judgements concerning the importance of the results for the union negotiations.

FOR EXTRA CREDIT:

For the four questions in this assignment, provide a graphical display (i.e., pie chart, histogram, side-by-side bar chart, etc.) to best illustrate each set of results. (Worth a total of 4 percentage points for the assignment.)



Note:

You may combine the Part C's from the questions into a report format, with at least one paragraph for each question. However, use sub-headings to indicate which text corresponds to the Part C for a particular question. I should have no difficulty in determining that all analyses have been discussed and all points that need to be addressed have been done so.