

## Report

(For questions 1 – 5): There are many variables in Stat\_Grades.sav that can be used to describe the students in the dataset. Use the variable Section and answer the following questions:

- Using the Variable View in SPSS (lower left), write out how Section is labeled, and note how many Section categories are included.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	6	0		None	None	8	Right	Scale
2	gender	Numeric	1	0	Gender	{1, FEMALE}...	None	8	Right	Nominal
3	ethnicit	Numeric	1	0	Ethnicity	{1, AMERICA}...	None	8	Right	Nominal
4	year	Numeric	1	0	Year in School	{1, FIRST-YEA}...	None	8	Right	Ordinal
5	lowup	Numeric	1	0	Lower or Uppe...	{1, LOWER DI}...	None	8	Right	Nominal
6	section	Numeric	1	0	Section of Class	None	None	8	Right	Nominal
7	prevgpa	Numeric	4	2	Previous GPA	None	None	8	Right	Scale
8	iq	Numeric	3	0	Short Form IQ ...	None	None	8	Right	Scale
9	review	Numeric	1	0	ATTENDED R...	{1, NO}...	None	8	Right	Ordinal
10	quiz1	Numeric	2	0	Quiz 1 Points	None	None	8	Right	Scale
11	quiz2	Numeric	2	0	Quiz 2 Points	None	None	8	Right	Scale
12	quiz3	Numeric	2	0	Quiz 3 Points	None	None	8	Right	Scale
13	quiz4	Numeric	2	0	Quiz 4 Points	None	None	8	Right	Scale
14	quiz5	Numeric	2	0	Quiz 5 Points	None	None	8	Right	Scale
15	final	Numeric	2	0	Final Exam Po...	None	None	8	Right	Scale
16	total	Numeric	3	0	Total Points	None	None	8	Right	Scale
17	percent	Numeric	4	0	Final Course P...	None	None	8	Right	Scale
18	grade	String	1	0	Numeric Grade	None	None	8	Left	Nominal
19	passfail	String	1	0	Pass or Fail	None	None	7	Left	Nominal

The variable Section is labeled "Section of Class". In the Values category, the response modalities are not indicated. In the Data View it is observed that this variable contains numbers. In order to know how many categories this variable has, the corresponding frequency distribution was obtained. This variable has three categories: 1, 2 and 3.

Section of Class				
		Frequency	Percent	Valid Percent
Valid	1	33	31.4	31.4
	2	39	37.1	37.1
	3	33	31.4	31.4
	Total	105	100.0	100.0

- What type of data and level of measurement does the Section variable represent: Is the data discrete or continuous? Is the data qualitative or quantitative? Is the data nominal, ordinal, ratio, or interval?

Section is a qualitative variable, with a nominal measurement scale. Although the response modalities are numbers, they are only used as category identifiers. The data is discrete.

3. Use SPSS to output the mean, median, and mode for the Section. Include the SPSS here. Which of the three descriptive numerical measures of central tendency (mean, median, or mode) **BEST** describes the variable Section? What information does the measure tell you about the students in the dataset (Stat\_Grades.sav) – be specific?

**Statistics**

Section of Class

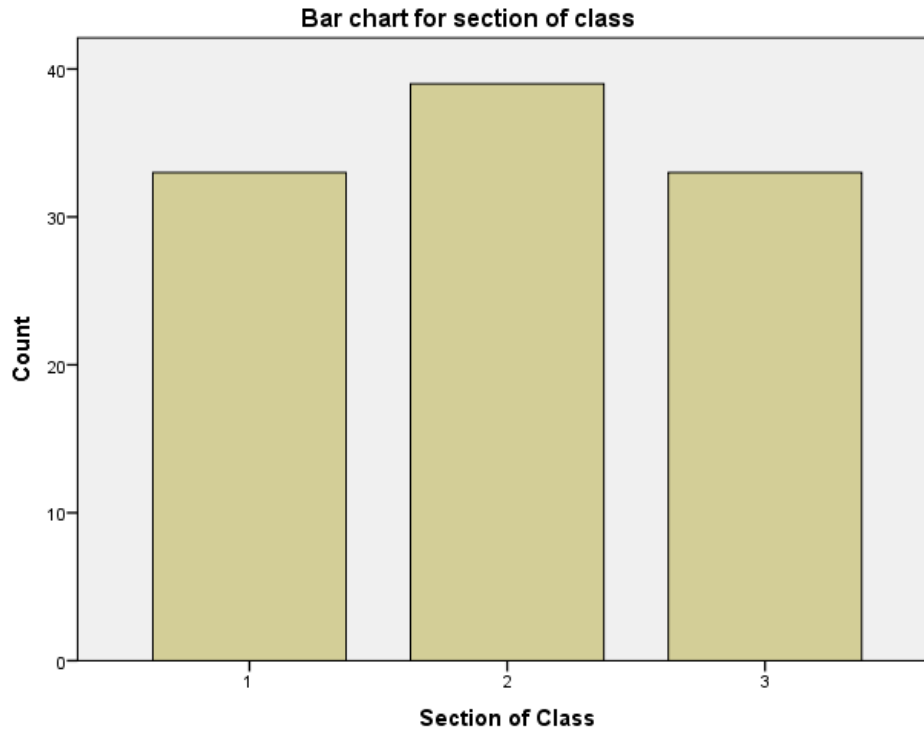
N	Valid	105
	Missing	0
Mode		2

The best measure described by this variable is the mode, due to its level of measurement. The arithmetic mean and the median have no meaning, since it is a qualitative variable. Mode tells us that section 2 is the one with the most students.

4. Which of the three descriptive numerical measures of central tendency (mean, median, or mode) **LEAST** (or most poorly) describes the variable Section? Explain.

The measure of central tendency that most poorly describes the variable section is the arithmetic mean, because it is a qualitative variable and it makes no sense to perform mathematical operations with this variable. The values of this variable are only category identifiers, which correspond to each of the sections.

5. Choose the most appropriate graph to represent the variable Section. (There is more than one possible answer). Create the graph in SPSS and include it here. Be sure the graph is *fully labeled*. Does this graph confirm what the mode tells you about this dataset? Explain.



The bar graph confirms that the group with the highest frequency is section 2, since the bar that corresponds to that category has greater height.

**(For questions 6 – 9): There are many variables in Stat\_Grades.sav that can be used to describe the students in the dataset. Use the variable Percept (Final Course Percent) and answer the following questions:**

- What type of data and level of measurement does the Percept variable represent: Is the data discrete or continuous? Is the data qualitative or quantitative? Is the data nominal, ordinal, ratio, or interval?

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	6	0		None	None	8	Right	Scale
2	gender	Numeric	1	0	Gender	{1, FEMALE}...	None	8	Right	Nominal
3	ethnicit	Numeric	1	0	Ethnicity	{1, AMERICAN I...	None	8	Right	Nominal
4	year	Numeric	1	0	Year in School	{1, FIRST-YEAR...	None	8	Right	Ordinal
5	lowup	Numeric	1	0	Lower or Upper Divi...	{1, LOWER DIVI...	None	8	Right	Nominal
6	section	Numeric	1	0	Section of Class	None	None	8	Right	Nominal
7	prevgpa	Numeric	4	2	Previous GPA	None	None	8	Right	Scale
8	iq	Numeric	3	0	Short Form IQ Test...	None	None	8	Right	Scale
9	review	Numeric	1	0	ATTENDED REVIE...	{1, NO}...	None	8	Right	Ordinal
10	quiz1	Numeric	2	0	Quiz 1 Points	None	None	8	Right	Scale
11	quiz2	Numeric	2	0	Quiz 2 Points	None	None	8	Right	Scale
12	quiz3	Numeric	2	0	Quiz 3 Points	None	None	8	Right	Scale
13	quiz4	Numeric	2	0	Quiz 4 Points	None	None	8	Right	Scale
14	quiz5	Numeric	2	0	Quiz 5 Points	None	None	8	Right	Scale
15	final	Numeric	2	0	Final Exam Points	None	None	8	Right	Scale
16	total	Numeric	3	0	Total Points	None	None	8	Right	Scale
17	percent	Numeric	4	0	Final Course Percent	None	None	8	Right	Scale
18	grade	String	1	0	Numeric Grade	None	None	8	Left	Nominal
19	passfail	String	1	0	Pass or Fail	None	None	7	Left	Nominal

The variable is not called percept, it was called percent. The variable percent is quantitative. The data is discrete, since the variable contains only integer values. The level of measurement of the variable is ratio.

7. Use SPSS to output the mean, median, and mode for the **Percept**. Include the SPSS here. Which of the three descriptive numerical measures of central tendency (mean, median, or mode) **BEST** describes the variable **Percept**? What information does the measure tell you about the students in the dataset (Stat\_Grades.sav)?

### Statistics

#### Final Course Percent

N	Valid	105
	Missing	0
Mean		80.34
Median		82.00
Mode		78

Since the asymmetry coefficient is between -1 and -1/2, the distribution is moderately asymmetric. Therefore, the median is the best measure of central tendency. It tells us that 50% of the students obtained a final course percent less than or equal to 82.

8. Which descriptive numerical measure of central tendency is **LEAST** appropriate to describe the variable **Percept**? Use SPSS to output ONLY that one numerical measure. Include it here. Why is that numerical measure the least appropriate? What information does that measure tell you about the students in the dataset (Stat\_Grades.sav)?

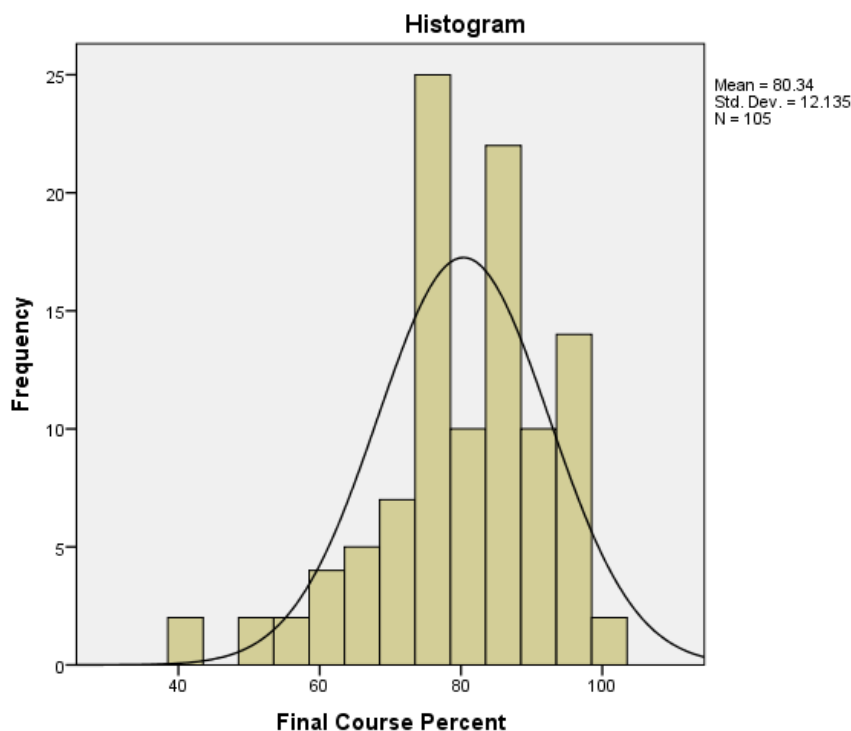
### Statistics

Final Course Percent

N	Valid	105
	Missing	0
Mode		78

The least appropriate measure could be the mode, since that measure is more appropriate for qualitative variables, to know which is the most frequent category. The final course percent is a quantitative variable, which can be best summarized by the mean or the median, depending on the symmetry of the distribution. In this case, the final percent that is most repeated is 78.

- Choose the most appropriate graph to represent the variable **Percent**. Create the graph in SPSS and include it here. Be sure the graph is *fully labeled*. Discuss why you chose this graph type and what it tells you about the students in the class.



The histogram was selected, since this is the best graph to represent quantitative variables. This graph allows to observe the range of the variable, the symmetry of the distribution and the mode. Additionally, a box with the mean, the standard deviation and the number of observations is included. The students in the class obtained values of final course percent approximately between 40 and 100, the grade that is repeated is slightly less than 80. The average of the students was 80.34. The distribution has a long left tail, due to its asymmetry. In addition, the curve of the normal distribution was superimposed, so that it would be compared with the distribution of the Final Course Percent. It is observed that this variable does not conform to the normal distribution.

**(For questions 10-13): Relationships (correlation) between variables.**

10. List all five quizzes, and include for each quiz, the numerical measures of mean, mode, and standard deviation. Explain what these measures tell you about the quizzes as compared to each other. On which quiz did students appear to perform the best? Which quiz had the lowest performance? Which quiz had the highest variance (variation in scores) and what does that tell you about that quiz? (100 word minimum, fully explain, include SPSS).

Fill in this table, include the SPSS, and then explain and answer the other questions.

	Mean	Mode	Standard Deviation
Quiz 1	7.47	10	2.481
Quiz 2	7.98	9	1.623
Quiz 3	7.98	10	2.308
Quiz 4	7.80	10	2.280
Quiz 5	7.87	10	1.765

		Statistics				
		Quiz 1 Points	Quiz 2 Points	Quiz 3 Points	Quiz 4 Points	Quiz 5 Points
N	Valid	105	105	105	105	105
	Missing	0	0	0	0	0
Mean		7.47	7.98	7.98	7.80	7.87
Median		8.00	8.00	9.00	8.00	8.00
Mode		10	9	10	10	10
Std. Deviation		2.481	1.623	2.308	2.280	1.765
Skewness		-.851	-.656	-1.134	-.919	-.713
Std. Error of Skewness		.236	.236	.236	.236	.236
Minimum		0	3	0	2	2
Maximum		10	10	10	10	10

The students obtained the highest mean in Quiz 2 and Quiz 3. The lowest mean occurred in Quiz 1. The most frequent grade in four of the five Quizzes was 10 points. The variability of the scores was higher in Quiz 1. This means that the marks obtained by the students are further away from their arithmetic mean. Distribution around the mean is in a wider range. Indeed, when detailing the minimum and maximum value of each Quiz, it is observed that the range of values of Quiz 1 (from 0 to 10) is greater than the ranges of Quiz 2, 4 and 5. The students seem to have the best performance in Quiz 2, because in that evaluation they obtained the highest mean and the smallest standard deviation.

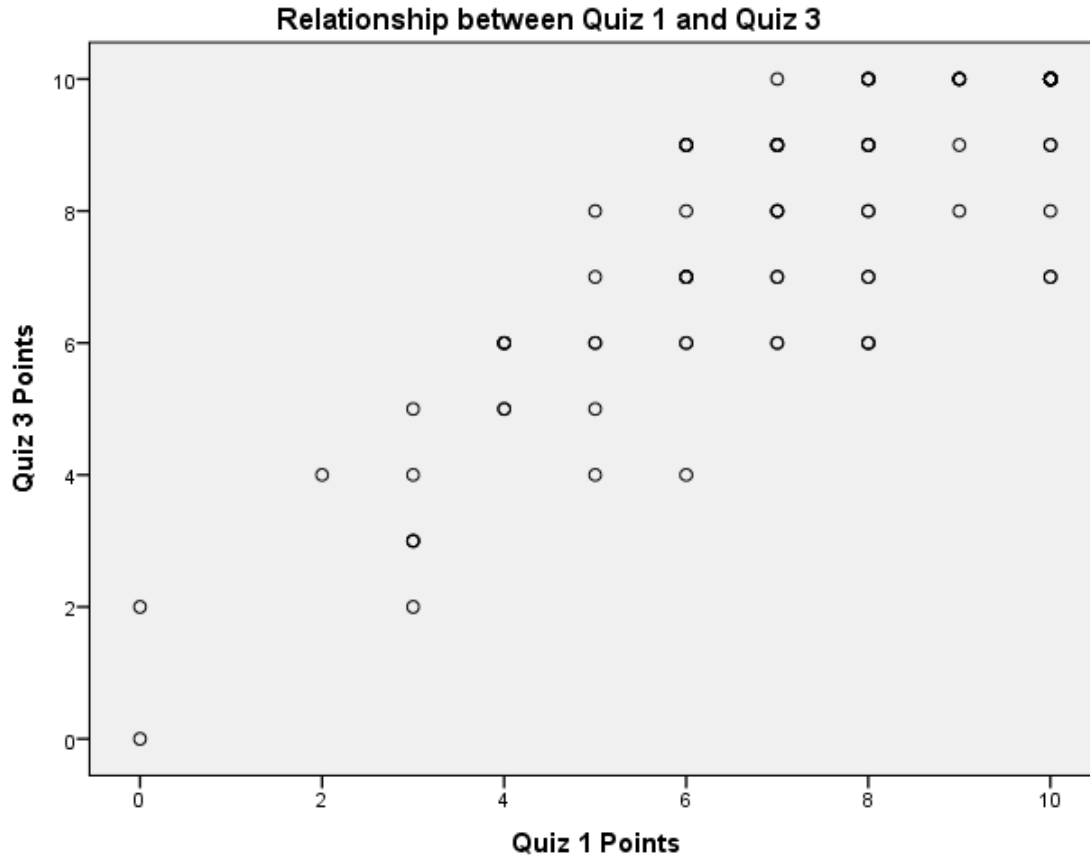
11. Looking at the relationship between each of the Five **Quizzes**: First, use (and include) SPSS to measure the relationship (correlation) between *each* Quiz. Which two of the five Quizzes has the strongest relationship (correlation) with each other, and how do you know? Which two of the five Quizzes has the weakest relationship with each other, and how do you know?

		Correlations				
		Quiz 1 Points	Quiz 2 Points	Quiz 3 Points	Quiz 4 Points	Quiz 5 Points
Quiz 1 Points	Pearson Correlation	1	.673**	<b>.858**</b>	.829**	.504**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	105	105	105	105	105
Quiz 2 Points	Pearson Correlation	.673**	1	.688**	.633**	.700**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	105	105	105	105	105
Quiz 3 Points	Pearson Correlation	.858**	.688**	1	.796**	.493**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	105	105	105	105	105
Quiz 4 Points	Pearson Correlation	.829**	.633**	.796**	1	<b>.445**</b>
	Sig. (2-tailed)	.000	.000	.000		.000
	N	105	105	105	105	105
Quiz 5 Points	Pearson Correlation	.504**	.700**	.493**	.445**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	105	105	105	105	105

\*\* . Correlation is significant at the 0.01 level (2-tailed).

The strongest correlation was presented in Quiz 1 and Quiz 3 (marked in blue, in the correlation matrix). A correlation coefficient  $r = .858$  was obtained, which is close to 1. The closer the correlation is to 1, the stronger it will be. The weakest correlation occurred between Quiz 4 and Quiz 5. A correlation coefficient  $r = .445$  was obtained. As long as the coefficient is closer to 0, the correlation between the variables will be weaker.

- Use SPSS to create the scatterplot for the two **Quizzes** that had the strongest relationship with each other. Paste the graph here. What does this graph show or tell you? Is it linear? Is the slope positive (going upwards) or negative (going downwards)? Does it seem like doing well on one of the two Quizzes here relates (or correlates) to doing well on the other Quiz?



There is a positive linear relationship between the scores obtained in Quiz 1 and Quiz 3. The slope of the line is positive which indicates that there is a positive linear relationship between these two variables. When the Quiz 1 points increases, the Quiz 3 points also increases. When the Quiz 1 points decreases, the Quiz 3 points also decreases. Students who scored high in Quiz 1 tend to get high scores in Quiz 3.

**13.** Will a low Quiz 1 grade cause a student to get a low Quiz 5 grade? Why or why not and explain? ( 50 word minimum)

No. A low Quiz 1 grade did not cause a student to get a low Quiz 5 grade. The correlation between variables only indicates the degree and direction of the linear relationship between two variables. It tells us that in this case, the low grades in Quiz 1 are related to low grades in Quiz 3. The linear correlation is not a cause-effect relationship. The occurrence of a score in Quiz 1 does not produce a score determined in Quiz 3. In short, the correlation does not imply causality.

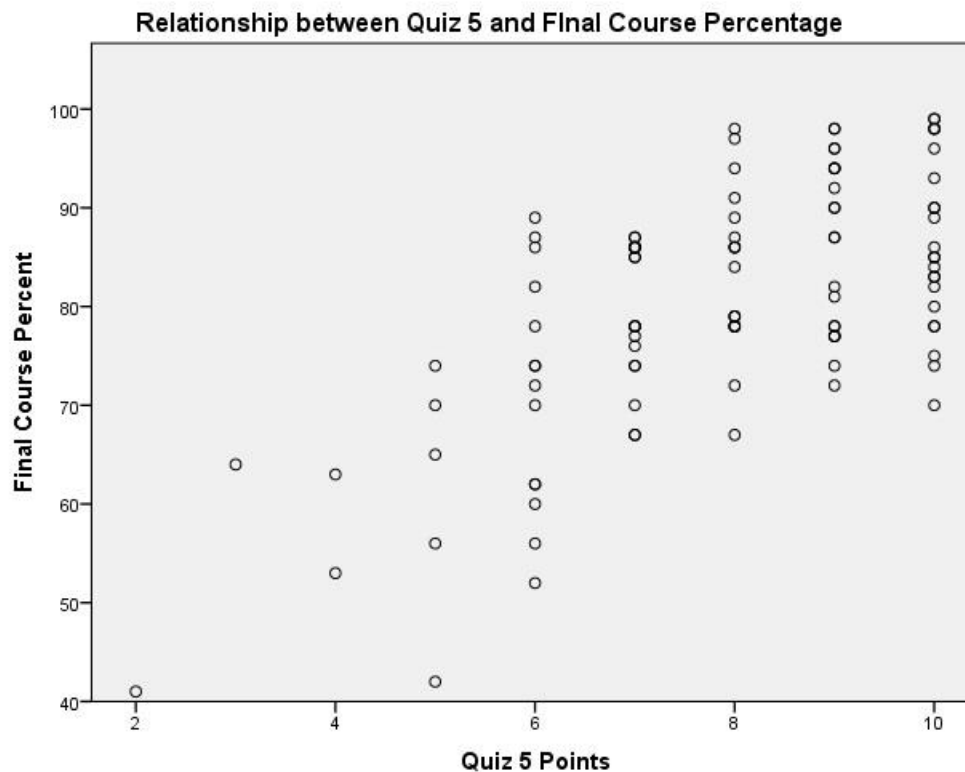
(For Questions 14-17 use the SPSS output below to determine the answers).

→ **Correlations**

**Correlations**

		Quiz 5 Points	Final Course Percent
Quiz 5 Points	Pearson Correlation	1	.651**
	Sig. (2-tailed)		.000
	N	105	105
Final Course Percent	Pearson Correlation	.651**	1
	Sig. (2-tailed)	.000	
	N	105	105

\*\* . Correlation is significant at the 0.01 level (2-tailed).



14. Which two variables are being investigated and label each variable as either dependent or independent.

The variables that are being investigated are Quiz 1 points (independent variable) and Final Course Percent (dependent variable).

15. The Correlation Coefficient (or r-value) is used to measure the relationship between any two variables. In this case, what is the r-value?

The correlation coefficient is  $r = .651$

16. What does the r-value and the scatterplot tell you about the relationship between **these two variables**? Is the relationship positive or negative? Is it strong, medium, or weak?

There is a moderate positive linear relationship between Quiz 1 points and Final Course Percent. The high values of Quiz 1 points are related to high values of Final Course Percent. The low values of Quiz 1 points are related to low values of Final Course Percent.

17. Suppose a student scored a 7 on the Quiz 5. Using the scatterplot, what is a good *estimate* of what their Final Course Percent?

A good estimate of Final Course Percent is about 77.

## References

Gravetter, F. J., & Wallnau, L. B. (2017). *Statistics for the behavioral sciences* (10th ed.). Boston, MA: Cengage Learning.

Green, S. B., & Salkind, N. J. (2014). *Using SPSS for Windows and Macintosh : analyzing and understanding data* (7th ed.). Boston: Prentice Hall.