

Name: Laura Lopez

Nyack College

Date: 26/01/2022

MAT 330

Course Pre-Assessment

Part I-Multiple Choice

1. A research calculated the percentages of students who are female at a sample of colleges and universities. The unit of analysis in this study was:

- a) females.
- b) gender.
- c) the college or university.
- d) percentages.
- e) a sample.

2. Generalizing from a survey of American colleges and universities to all American colleges and universities is an example of:

- a) measurement.
- b) a continuous variable.
- c) a census.
- d) inferential statistics
- e) descriptive statistics.

3. A parameter is to a statistic as:

- a) collectively exhaustive is to mutually exclusive.
- b) discrete is to continuous.
- c) a nominal variable is to an ordinal variable.
- d) a population is to a sample.
- e) descriptive statistics is to inferential statistics.

Name: _____

Date:

Nyack College

MAT 330

4. The General Social Survey asked respondents the meanings of each of ten words. The variable #CRCT.WORD reports the number of correct answers. The level of measurement of this variable is:

- a) nominal.
- b) ordinal.
- c) interval/ratio.
- d) none of the above.
- e) can't tell because there is not enough information.

5. The General Social Survey asked respondents to assess their own health as excellent, good, fair, or poor. The level of measurement of this variable is:

- a) nominal.
- b) ordinal.
- c) interval/ratio.
- d) none of the above.
- e) can't tell because there is not enough information.

6. The General Social Survey asked respondents what region of the country they lived in at age 16. The level of measurement of this variable is:

- a) nominal.
- b) ordinal.
- c) interval/ratio.
- d) none of the above.
- e) can't tell because there is not enough information.

Name: _____

Date:

Nyack College

MAT 330

7. Suppose that a researcher conducting a survey asks respondents their annual incomes using these values: \$20,000 or less, \$20,000 thru \$60,000; \$60,000 or more. A problem with this set of values is that:

- a) they are measured at the nominal level.
- b) they are not continuous.
- c) they are population data.
- d) they are not collectively exhaustive.
- e) they are not mutually exclusive.

8. Which of the following sets of categories is not collectively exhaustive?

- a) Frequency of newspaper reading: Daily; Few times a week; Once a week; Less often; Never.
- b) Favorite TV shows: Drama; Comedy; Sports.
- c) Employment status: Employed; Not employed.
- d) Self-reported health: Excellent; Good; Fair; Poor.
- e) Education: Years of school completed.

9. The CIA's World Factbook summarizes information about each country. Information for each country includes such variables as population size, gross national product, and mortality rates. Variables described in the World Factbook are:

- a) nominal variables.
- b) discrete variables.
- c) dichotomous variables.
- d) ecological variables.
- e) bivariate.

Name: _____

Nyack College

Date:

MAT 330

10. Consider the argument: "States with higher percentages of drivers under 20 years old have more pickup trucks per 1,000 population." Therefore, pickup trucks are mostly driven by teenagers. This argument:

- a) uses discrete data.
- b) uses inferential statistics.
- c) is not collectively exhaustive.
- d) involves an ecological fallacy.
- e) uses multivariate analysis.

11. We often use percentages because:

- a) percentages are more scientific.
- b) percentages are more precise.
- c) percentages are presentation quality.
- d) percentages make comparisons easier.
- e) percentages are collectively exhaustive.

12. Compared with a small sample, a large sample has percentages that are:

- a) more discrete.
- b) more ecological.
- c) more cumulative.
- d) more reliable.
- e) more interval/ratio.

13. We can assess the stability or reliability of percentages by considering:

- a) the total number of cases on which percentages are based.
- b) the size of the percentage.
- c) inferential statistics.
- d) all of the above.
- e) none of the above.

Name: _____

Date: _____

Nyack College

MAT 330

14. Consider this table:

Table- Self-Reported Health
(in percentages)

<u>Health</u>	<u>Percent</u>
Excellent	30.2
Good	47.4
Fair	17.0
Poor	5.4
Total	100.0
(N)	(493)

Which of the following is true?

- a) The cumulative percentage who report at least good health is 77.6.
- b) A spot map would be a good way to display these percentages.
- c) The frequency is too small for stable percentages.
- d) All of the above.
- e) None of the above.

15. If values are not mutually exclusive and collectively exhaustive:

- a) the variable is a subset.
- b) the researcher commits an ecological fallacy.
- c) percentages will not sum to 100.
- d) the variable is continuous.
- e) all of the above.

Name: _____

Date:

Nyack College

MAT 330

16. A researcher interested in variations in Asian Americans' political attitudes restricted her analysis to only Asian Americans in the General Social Survey. The Asian Americans that she analyzed are:

- a) a subset.
- b) outliers.
- c) discrete.
- d) missing data.
- e) ecological.

17. Generally, ordinal variables are best displayed visually with:

- a) cumulative percentages.
- b) collapsed values.
- c) pie charts.
- d) bar graphs.
- e) subsets.

18. In a bar graph:

- a) the width of bars should reflect the number of cases.
- b) values of ordinal variables should usually be listed from tallest to shortest.
- c) the vertical axis should usually begin at zero if possible.
- d) none of the above.
- e) all of the above.

19. A bar graph is a useful way to spot:

- a) significant digits.
- b) aggregate data.
- c) outliers.
- d) none of the above.
- e) any of the above.

Name: _____

Date:

Nyack College

MAT 330

20. In maps showing the distribution of a variable across the 50 American states, a researcher can reduce the visual effect of differences in geographic size by using:

- a) an area map.
- b) a spot map.
- c) a continuous variable.
- d) a discrete variable.
- e) raw data.

21. A mode:

- a) can be found only for discrete variables.
- b) is usually less than the mean.
- c) is the value of the most frequently occurring score.
- d) is usually less for ordinal than for interval/ratio variables.
- e) is useful only for symmetrical variables.

22. A distribution that has one "hump" is:

- a) symmetrical.
- b) dichotomous.
- c) unimodal.
- d) skewed.
- e) ecological.

23. Consider these scores: 0, 3, 1, 5, 1. The mean is:

- a) 1.
- b) 2.
- c) 3.
- d) 5.
- e) none of the above.

Name: _____

Date: _____

Nyack College

MAT 330

24. The median is less sensitive than the mean to:

- a) the skewness of a variable.
- b) the sum of squares.
- c) dichotomous variables.
- d) all of the above.
- e) none of the above.

25. Means can be greatly influenced by:

- a) outliers.
- b) the sum of squares.
- c) interval/ratio variables.
- d) all of the above.
- e) none of the above.

26. Which of the following always produces the smallest sum of squares?

- a) the mode.
- b) the median.
- c) the mean.
- d) a symmetrical distribution.
- e) a bimodal distribution.

27. The variable LIKESCI reports whether students like or do not like science. "Like science" is coded 1 and "does not like science" is coded 0. If 75 percent of the students at a school like science, the mean is:

- a) 0.
- b) .25.
- c) .50.
- d) .75
- e) 1.00.

Name: _____

Date:

Nyack College

MAT 330

28. In a symmetric, unimodal distribution:

- a) the median equals the mean.
- b) the mode equals the median.
- c) the mean equals the mode.
- d) all of the above.
- e) none of the above

29. The mean years of the GSS variable AGE KD BRN is higher than the median, so we know that the distribution of respondent's age when first child was born is:

- a) collapsed.
- b) normal.
- c) leptokurtic.
- d) bimodal.
- e) positively skewed.

30. Missing data should be included when finding:

- a) the mode.
- b) the median.
- c) the mean.
- d) all of the above.
- e) none of the above.

31. The variance measures deviation around the:

- a) mode.
- b) median.
- c) mean.
- d) sum of squares.
- e) standard deviation.

Name: _____

Date:

Nyack College

MAT 330

32. The variance:

a) is the square of the standard deviation.

b) can be 0.

c) is never negative.

d) all of the above.

e) none of the above.

33. The number of standard deviations a score lies from the mean is:

a) the case's Z-score

b) the case's kurtosis.

c) the standard error.

d) the confidence interval.

e) the sampling distribution.

34. Z-scores for a variable:

a) are normally distributed.

b) have a standard deviation of 0.

c) sum to zero.

d) all of the above.

e) none of the above.

35. All normal distributions:

a) have a mean of 1.

b) have sampling distributions.

c) are symmetrical.

d) all of the above

e) none of the above.

Name: _____

Date:

Nyack College

MAT 330

36. The standard deviation of a sampling distribution:

- a) applies only to population data.
- b) can be larger than the standard deviation for the population.
- c) is the sum of squares.
- d) is the standard error.
- e) is always greater than 1.

37. The central limit theorem tells us that the larger the size of a sample, then:

- a) the smaller the variance.
- b) the smaller the standard error.
- c) the less the mean.
- d) the greater the standard score.
- e) the more skewed the variable.

38. The central limit theorem tells us that the larger the size of a sample, the closer the standard error is to:

- a) the population's standard error.
- b) the mean.
- c) sN .
- d) 1.00.
- e) the standard score.

39. If the 95 percent confidence interval is between 2.5 and 2.7, we know that:

- a) the distribution is normal.
- b) the standard error is .2
- c) the distribution is symmetric.
- d) 95 percent of the sample's scores are between 2.5 and 2.7.
- e) the mean is 2.6.

Name: _____

Date:

MAT 330

Nyack College

40. A 99 percent confidence interval includes _____ number of scores than/as a 95 percent confidence interval.

- a) a smaller
- b) the same
- c) a larger
- d) a more normal
- e) a more standardized

Directions (Questions 41-45): Identify each of the following data sets as categorical (C) or numerical (N). Provide complete responses to each question.

41. Times of 30 marathoners N

42. Favorite T-shirt color for each of 100 volunteers C

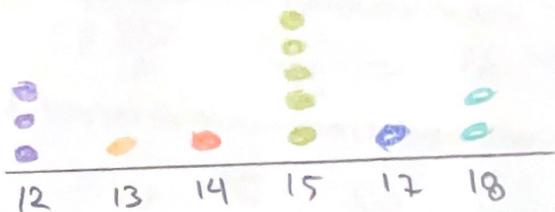
43. Type of music listened to by 100 middle school students C

44. Pants color worn by 30 officer workers N

45. Hours of sleep on a road trip for 360 NBA basketball players N

46. Create a dot plot for the following set of data:

{12, 12, 12, 13, 14, 15, 15, 15, 15, 15, 17, 18, 18}



Name: _____

Date: _____

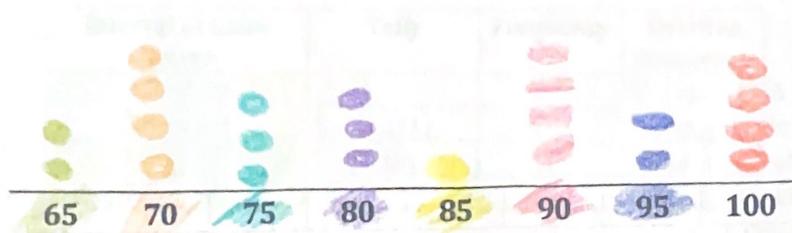
Nyack College

MAT 330

47. Mr. Brown recorded the following test scores for his math class:

100, 70, 75, 70, 80, 100, 90, 75, 95, 90, 70, 65, 80, 85, 100, 75, 100, 90, 65, 70, 90, 95, 80

a. Create a dot plot for the math test scores



b. Complete the tally mark and frequency columns in the table below:

Test Scores	Tally	Frequency	
65		2	= 130
70		4	= 280
75		3	= 225
80		3	= 240
85		1	= 85
90		4	= 360
95		2	= 190
100		4	= 400
Total		23	= 1910

c. What score describes the center of the data?

$$\bar{x} = \frac{\sum fx}{N} = \frac{1910}{23} = 83.0435$$

d. What was the most common (or frequent) test score?

70, 90, 100 has the highest Frequency

a. 4.

Name: _____

Date: _____

Nyack College

MAT 330

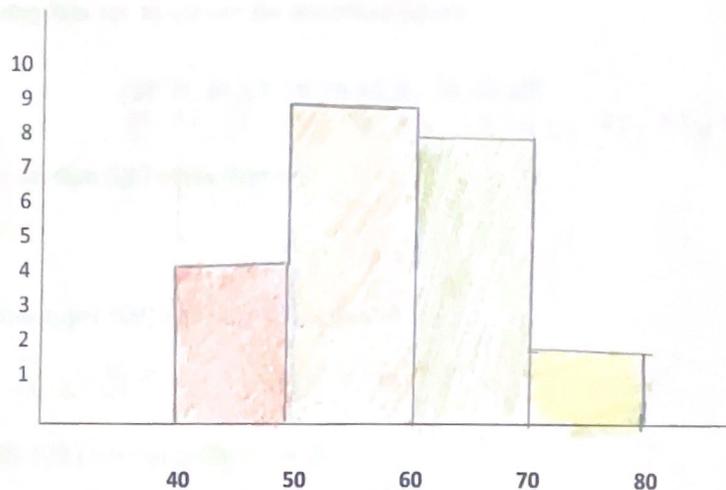
48. The table below represents the total points scored by the varsity basketball team for each of the 23 games played:

45, 50, 67, 72, 41, 62, 56, 59, 40, 60, 69, 47, 54, 76, 53, 64, 53, 57, 61, 64, 52, 56, 64

a. Complete the following table:

Interval of Game Scores	Tally	Frequency	Relative Frequency
40 < 50		4	$4/23 = 0.1739$
50 < 60	 	9	$9/23 = 0.3913$
60 < 70	 	8	$8/23 = 0.3478$
70 < 80		2	$2/23 = 0.087$
		23	= 1.0000

b. Plot the histogram for the game scores (make sure label the axes)



c. Based on the histogram, describe the center of the game scores

Midpoint	Frequency	fx
45	4	180
55	9	495
65	8	520
75	2	150
		= 1345

$$\bar{x} = \frac{\sum fx}{N}$$

$$\bar{x} = \frac{1345}{23} = 58.48$$

Name: _____

Nyack College

Date:

MAT 330

d. How would the histogram change if you added the game scores 73 and 74?

I have to add 73 and 74 in the last column of 70-80.

Directions (Questions 49-50): Find the mean, median, mode, and range of the following data sets. Provide complete responses to each question.

49. ~~1~~, 8, ~~8~~, 8, ~~8~~, 8, 5, ~~4~~, and 8

$$\text{mean} = \frac{1 + 2 + 3 + 4 + 5 + 7 + 8 + 8 + 8}{9} = \frac{46}{9} = \boxed{5.111} \quad \text{mode} = 8$$

$$\text{median} = \frac{n+1}{2} = \frac{9+1}{2} = \boxed{5} \quad \text{Range} = 8 - 1 = 7$$

50. 7, 10, 11, 12, 11, 11, 13, 8, 22, and 23

$$\text{mean} = 12.8$$

$$\text{mode} = 11$$

$$\text{median} = 11$$

$$\text{Range} = 16$$

51. Use the following data set to answer the questions below:

~~{58, 71, 51, 67, 58, 55, 53, 57, 58, 62, 65}~~

~~51, 53, 55, 57, 58, 58, 58, 62, 65, 67, 71~~

a. What is the median (Q2) of the data set?

$$Q_2 = 58$$

b. What are the upper (Q3) and lower (Q1) quartiles?

$$Q_3 = 55 \quad Q_1 = 65$$

c. Calculate the IQR (inter-quartile range)?

$$55 - 65 = -10$$

Name: _____

Date: _____

Nyack College

MAT 330

d. What are the minimum and maximum of the data set?

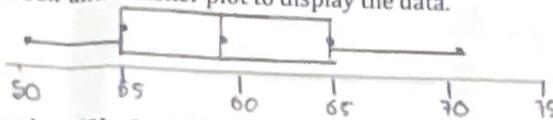
minimum

maximum

51

71

e. Draw a box-and-whisker plot to display the data.



Directions (Question 52): Complete the following stem-leaf tables.

52. Construct stem leaf plots for the following data sets:

a. {16, 22, 33, 41, 42, 41, 41, 40, 27, 18}

Stem	Leaf
1	6, 8
2	2, 7
3	3
4	0, 1, 1, 1, 2

b. {41, 38, 47, 48, 22, 28, 25, 40, 30}

Stem	Leaf
2	2, 5, 8
3	0, 8
4	0, 1, 7, 8

Name: _____

Date:

Nyack College

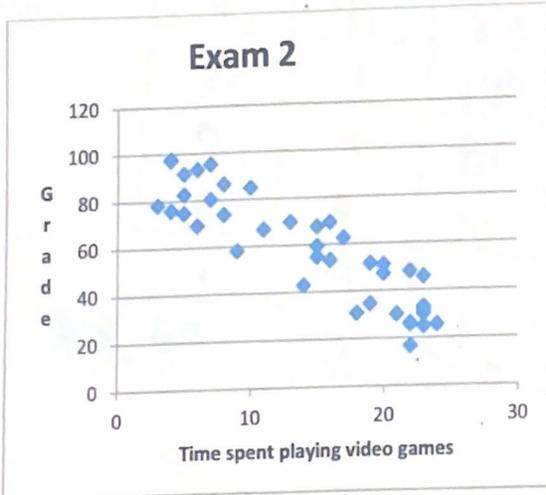
MAT 330

c. {76, 83, 54, 43, 44, 44, 103, 99, 94}

Stem	Leaf
4	3, 4, 4
5	4
7	6
8	3
9	4, 9
10	3

Part II

53. Below is a scatterplot of grade on exam 2 versus time spent playing video games in hours for the Exam 2 week for 40 students.



(a) Describe the important features of the relationship between two variables:

Form: linear or nonlinear?

Direction: negative (or no relation)?

Strength: strong, moderate, or weak?

Name: _____

Date: _____

Nyack College

MAT 330

54. Use the least square method to determine the equation of line of best fit for the data by completing the following table and then plot the line.

x	8	2	11	6	5	4	12	9	6	1
y	3	10	3	6	8	12	1	4	9	14

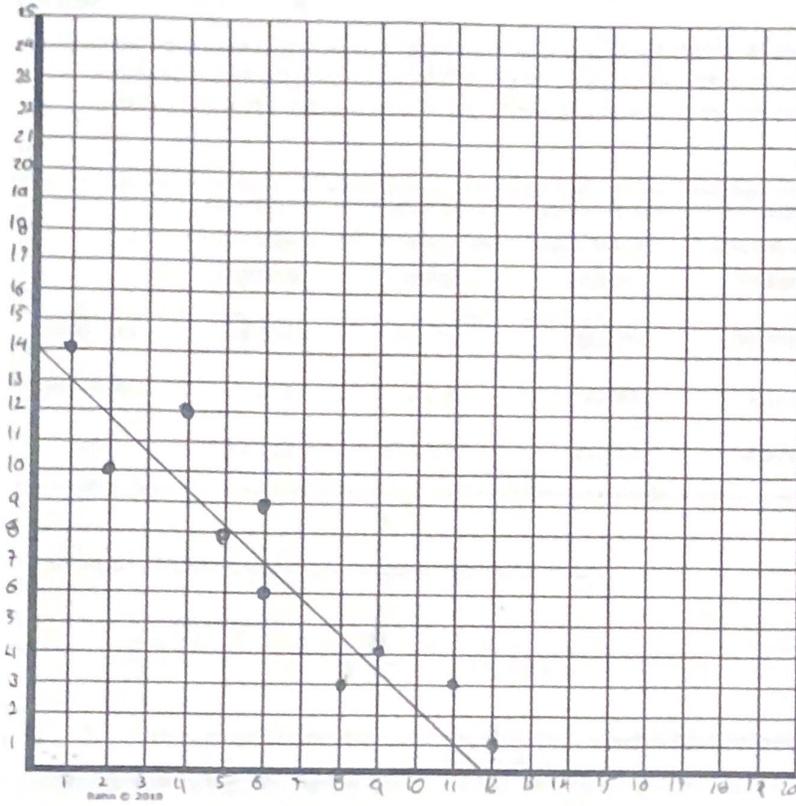
x	y	x^2	y^2	xy	
8	3	64	9	24	
2	10	4	100	20	
11	3	121	9	33	
6	6	36	36	36	
5	8	25	64	40	
4	12	24	144	48	
12	1	144	1	12	
9	4	81	16	36	
6	9	36	81	54	
1	14	1	196	14	
Total	64	70	528	656	317

Name: _____

Date: _____

Nyack College

MAT 330



Name: _____

Date: _____

Nyack College

MAT 330

Part III

55. Does education improve your chances of being employed? The following table provides data on the educational attainment and employment status of the U.S. population, 25 years of age and older, in the labor force as of October 2012. The data is in thousands of people. (Source: Bureau of Labor Statistics.)

	Less than a High School Diploma	High School Graduates, No College	Some College or Associate Degree	Bachelor's Degree and Higher	Total
Employed	9,886	33,702	35,176	46,849	125,613
Unemployed	1,377	3,076	2,599	1,827	8,879
Total	11,263	36,778	37,775	48,676	134,492

What is the probability that a randomly selected person:

a. Is unemployed?

$$8,879 / 134,492 = 0.066$$

b. Is unemployed given that he/she has less than a high school diploma?

$$1,377 / 11,263 = 0.1223$$

c. Are the events "unemployed" and "less than a high school diploma" independent?

$\frac{\text{Unemployed}}{\text{less than a high school diploma}}$ is not equal to $\frac{\text{Unemployed}}{\text{Total}}$

d. Is unemployed given that he/she is a high school graduate with no college education?

$$3,076 / 36,778 = 0.0836$$

e. Is unemployed given that he/she has some college education or an associate degree?

$$2,599 / 37,775 = 0.0688$$

f. Is unemployed given that he/she has at least a bachelor's degree?

$$1,827 / 48,676 = 0.0375$$

g. Based on your answers to parts b. through e., would you say that education improves your chances of being employed?

Yes, because the education make people to get more offers of jobs and no more unemployed.

Name: _____

Date: _____

Nyack College

MAT 330

h. Is unemployed and has at least a bachelor's degree?

$$1,827 / 134,492 = 0.01558$$

i. Has less than a high school diploma or is a high school graduate with no college education?

- Are the events "less than a high school diploma" and "high school graduate with no college education" mutually exclusive?

$$\frac{11,263}{134,492} + \frac{36,778}{134,492} = 0.35720 //$$

mutually exclusive = 0 //

56. Are women more likely to be unemployed than men? The following table provides data on the gender and employment status of the U.S. population, 16 years of age and older, in the labor force as of October 2012. The data is in thousands of people. (Source: Bureau of Labor Statistics.)

	Men	Women	Total
Employed	76,055	67,329	143,384
Unemployed	<u>6,642</u>	<u>5,616</u>	12,258
Total	82,697	<u>72,945</u>	<u>155,642</u>

What is the probability that a randomly selected person:

a. Is unemployed given that she is a woman?

$$\frac{5,616}{155,642} / \frac{72,945}{155,642} = \frac{5,616}{72,945} = 0.0770 //$$

b. Is unemployed given that he is a man?

$$\frac{6,642}{155,642} / \frac{82,697}{155,642} = \frac{6,642}{82,697} = 0.0803 //$$

c. Based on your answers to parts a. and b., would you say that women are more likely than men to be unemployed?

The results of questions a and b show that men have highest number of unemployment, but the number of women does not differ much from that of men.