EDUC/PSY 6600

Research Design & Analysis I

Name:

Unit 1 Assignment

Exploratory Data Analysis

---

*Please complete the following exercises. Feel free to work with classmates, but each student must turn in **UNIQUE** work, not photocopies or identical replicates. When applicable, use **APA format** in communicating your results in text. **Show your work!** If any question involves any math at all, show your work. When it doubt, write it out. Always show more than you think you need.*

1) WRITE-UP - Textbook Problems

Cohen Chap		Exercises	Pts	Off
2	A	1	1	
	C	2, 3, 4 *(no output, tables, or plots need to be included)*	2	
3	A	*1, 2, 3, 4, *5	5	
4	A	*1, 2, 4, 9a, 10a, 11, *12, 13, *14	11	
	B	3, 4, 5, 6, 7	5	

2) SUMMARY – Your Journal Article

Do NOT re-submit your HW 0 article!		Pts	Off
Half Page	Re-read your article you selected for Unit 0 HW (do not re-submit it). Summarize any mention or use of the concepts in the above chapters.	10	

3) R SYNTAX – Section C: Ihno's data set – add to the skeleton R notebook and knit to .pdf & upload

Cohen Chap		Exercises	Pts	Off
2	C	1, 2, 3, 4, 6, 9, 10	8	
3	C	1, 2, 3, 4, 5, 6	6	
4	C	1,	2	

Grading		Earned	Possible
CORRECTNESS	*a subset of spot-checked items: must show work, especially items from back of book or done in class*		50
COMPLETENESS	*more than one item is missing or skipped: 25/50 roughly half the assignment is completed: 10/50*		50
			100

2 A 1. Histograms

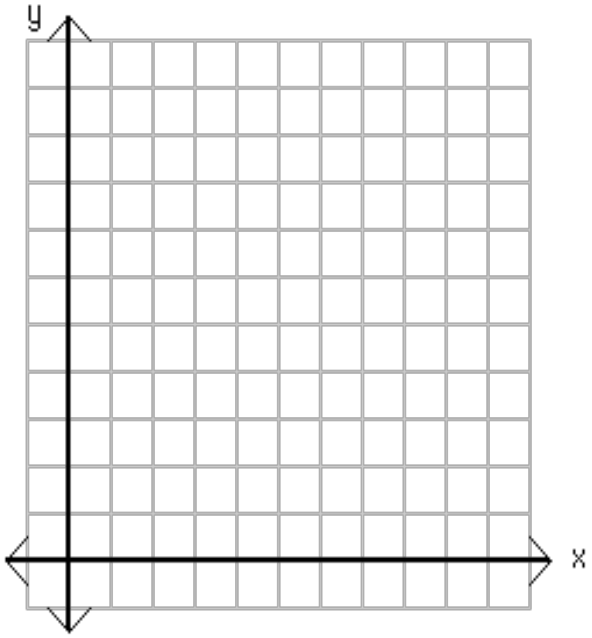
A psychotherapist has rated all 20 of her patients in terms of their progress in therapy, using a 7-point scale. The results are shown in the following table: (please use TWO decimal places)

	F	rf	cf	crf	cpf
Greatly improved	5				
Moderately improved	4				
Slightly improved	6				
Unchanged	2				
Slightly worse	2				
Moderately worse	1				
Greatly worse	0				

To answer the questions below, fill out the columns in the table above, such that:

- rf = relative frequency
- cf= cumulative frequency
- crf = cumulative relative frequency
- cpf = cumulative percentage frequency

a) Draw a bar graph to represent the above results.



b) What **proportion** of the patients was grealy improved?

c) How **many** patients did not improve?

d) What is the **percentile rank** of a patient who improved slightly?

What is **the percentile rank** of a patient who becomes slightly worse?

e) Which **category** of improvement corresponds to the third quartile?

Which **category** of improvement corresponds to the first quartile?

2 C	2. Distributions & Bar Plots

Request a frequency distribution (table) and a bar chart for the ***prevmath*** and ***phobia*** variables.

write code in R syntax file

(no output, tables, or plots need to be included here)

Would it **make sense** to request a histogram instead of a bar chart for *phobia*?
Discuss.

☐ yes ☐ no

2 C	3. Distributions & Bar Plots

Request a frequency distribution and a histogram for the ***statquiz*** variable.

write code in R syntax file

(no output, tables, or plots need to be included)

Describe the shape of this distribution.

2 C	4. Distributions & Bar Plots

Request a frequency distribution and a histogram for the ***anx_base*** *and* ***hr_base*** variables.

write code in R syntax file

(no output, tables, or plots need to be included)

Comment on R's choice of class intervals (# bins or binwidth) for each.

3 A	*1. Measures of central tendency

Select **the measure of central tendency** that would be most appropriate for describing each of the following hypothetical sets of data:

a. Religious preferences of delegates to the United Nations	☐ Mean ☐ Median ☐ Mode
b. Heart rates for a group of women before they start their first aerobics class	☐ Mean ☐ Median ☐ Mode
c. Types of phobias exhibited by patients attending a phobia clinic	☐ Mean ☐ Median ☐ Mode
d. Amounts of time participants spend solving a classic cognitive problem, with some of the participants unable to solve it	☐ Mean ☐ Median ☐ Mode
e. Height in inches for a group of boys in the first grade	☐ Mean ☐ Median ☐ Mode

3 A	2. Distribution descriptives

Describe a **realistic situation** in which you would expect to obtain each of the following:

a) A negatively skewed distribution

b) A positively skewed distribution

c) A bimodal distribution

3 A	*3. Distribution descriptives

A midterm exam was given in a large introductory psychology class. The **median** score was 85, the **mean** was 81, and the **mode** was 87.

What kind of distribution would you expect from these exam scores?

3 A	4. Measures of central tendency

A veterinarian is interested in the life span of golden retrievers. She recorded the age at death (in years) of the retrievers treated in her clinic. The ages were **12, 9, 11, 10, 8, 14, 12, 1, 9, 12.**

a) Calculate the mean, median, and mode of life span:

Mean	Median	Mode

b) After examining her records, the veterinarian determined that the dog that had died at 1 year was killed by a car. **Recalculate** the mean, median, and mode without that dog's data.

Mean	Median	Mode

c) Which measure of central tendency in part b **changed the most**, compared to the values originally calculated in part a?

☐ Mean ☐ Median ☐ Mode

Explain.

3	A	5. Measures of variability

Which of the three most popular **measures of variability** would you choose in each of the following situations?

a. The distribution is badly skewed with a few extreme outliers in one direction.	☐ Range ☐ SIQ range ☐ SD
b. You are planning to perform advanced statistical procedures (e.g., draw inferences about population parameters).	☐ Range ☐ SIQ range ☐ SD
c. You need to know the maximum width taken up by the distribution.	☐ Range ☐ SIQ range ☐ SD
d. You need a statistic that takes into account every score in the population.	☐ Range ☐ SIQ range ☐ SD
e. The highest score in the distribution is "more than 10."	☐ Range ☐ SIQ range ☐ SD

4	A	*1. Z-scores

If you **convert each score in a set of scores to a z score**, which of the following will be **true** about the resulting set of z scores?

a. The mean will equal 1.	☐ TRUE
b. The variance will equal 1.	☐ TRUE
c. The distribution will be normal in shape.	☐ TRUE
d. All of the above.	☐ TRUE
e. None of the above.	☐ TRUE

4	A	2. Z-scores

The distribution of body weights for adults is somewhat **positively skewed**— there is much more room for people to be above average than below.

If you take **the mean weights for random groups of 10 adults** each and form a new distribution, how will this new distribution **compare** to the distribution of individuals?

a. The new distribution will be **more symmetrical** than the distribution of individuals.	☐ TRUE
b. The new distribution will **more closely resemble the normal** distribution.	☐ TRUE
c. The new distribution will be **narrower** (i.e., have a smaller standard deviation) than the distribution of individuals.	☐ TRUE
d. All of the above.	☐ TRUE
e. None of the above.	☐ TRUE

4 A	4. z-scores

a) Calculate μ and σ for the following set of scores and then convert each score to a z score:
64, 45, 58, 51, 53, 60, 52, 49.

Mean: μ	Standard Deviation: σ
μ =	**σ =**

z-scores:

b) Calculate the mean and standard deviation of these z scores.

Mean: μ	Standard Deviation: σ
μ =	**σ =**

Did you obtain the values you **expected**?
Explain.

☐ yes ☐ no

4	A	*9a. Z-scores & the z-table

a) Use Table A.1 to find the **area** of the normal distribution **between the mean and z** , when **z equals 0.18**

4	A	10a. Z-scores & the z-table

a) Use Table A.1 to find the **area** of the normal distribution **beyond z** , when z equals when **z equals 0.09**

4	A	11. Z-scores & the z-table

Assuming that IQ is normally distributed with a **mean of 100** and a **standard deviation of 15**

describe completely the sampling distribution of the mean for a sample size (n) equal to 20.

4	A	*12. Standard error for the mean

If the **population** standard deviation (σ) for some variable equals 17.5, what is the value of the **standard error of the mean** when

a. N = 5 SE_{μ} =

b. N = 25 SE_{μ} =

c. N = 125 SE_{μ} =

d. N = 625 SE_{μ} =

If the sample size is cut in half, what happens to the standard error of the mean for a particular variable?

4 A	13. Standard error for the mean

a) In one college, freshman English classes always contain exactly **20 students**. An English teacher wonders how much these classes are **likely to vary** in terms of their verbal scores on the SAT. What would you expect for the standard deviation (i.e., **standard error) of class means** on the verbal SAT?

SE_{μ} =

b) Suppose that a crew for the space shuttle consists of **seven people**, and we are interested in the average weights of all possible shuttle crews. If the **standard deviation for weight is 30 pounds**, what is the **standard deviation for the mean weights** of shuttle crews (i.e., the standard error of the mean)?

SE_{μ} =

4 A	*14. Standard error for the mean

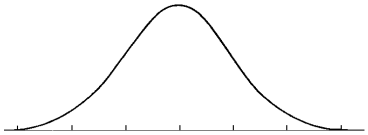
If for a particular sampling distribution of the mean we know that the **standard error is 4.6**, and we also know that **σ = 32.2**, what is the sample size (n)?

n =

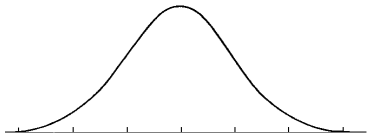
4 B	*3. Area under a normal curve

Assume that the resting heart rate in humans is normally distributed with $\mu = 72$ bpm (i.e., beats per minute) and $\sigma = 8$ bpm.

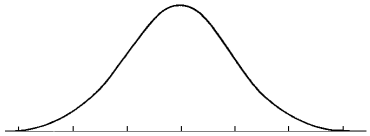
a) What proportion of the population has resting heart rates **above** 82 bpm?



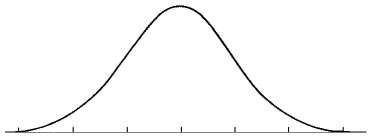
Above 70 bpm?



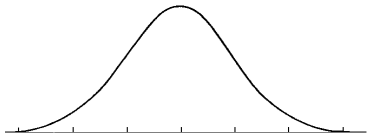
b) What proportion of the population has resting heart rates **below** 75 bpm?



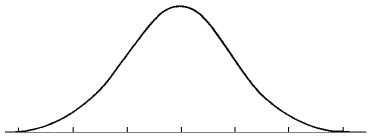
Below 50 bpm?



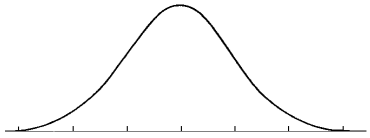
c) What proportion of the population has resting heart rates **between** 80 and 85 bpm?



Between 60 and 70 bpm?



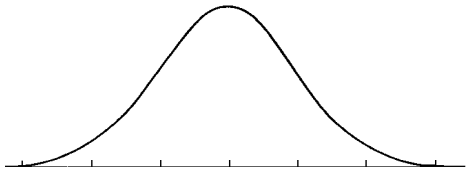
Between 55 and 75 bpm?



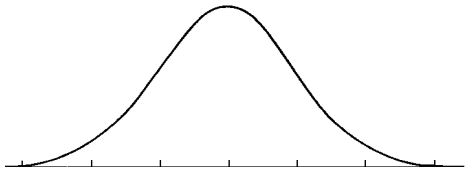
4	B	4. Area under a normal curve

Assume that the resting heart rate in humans is normally distributed with $\mu = 72$ bpm (i.e., beats per minute) and $\sigma = 8$ bpm.

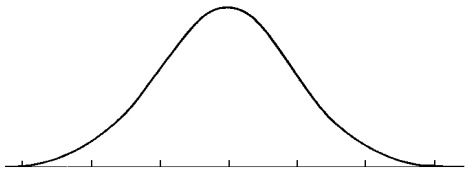
a) **Above** what heart rate do you find the upper 25% of the people? (That is, what heart rate is at the 75th percentile, or third quartile?)



b) **Below** what heart rate do you find the lowest 15% of the people? (That is, what heart rate is at the 15th percentile?)



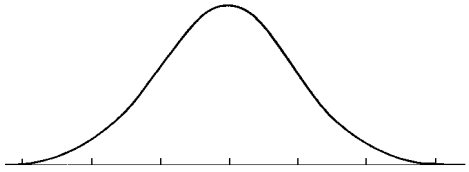
c) **Between** which two heart rates do you find the middle 75% of the people?



4	B	*5. Area under a normal curve

A new preparation course for the math SAT is open to those who have already taken the test once and scored in the **middle 90%** of the population.

In what **range** must a testtaker's previous score have fallen for the test-taker to be eligible for the new course?

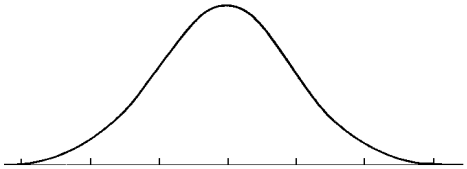


4 B	6. Area under a normal curve

A teacher thinks her class has an unusually high IQ, because her 36 students have an average IQ (X) of 108. If the population mean is 100 and σ = 15.

a) What is the **z score** for this class?

b) What **percentage** of classes (n = 36, randomly selected) would be even higher on IQ?

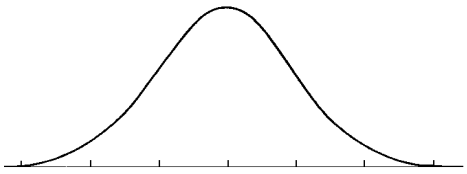


4 B	*7. Area under a normal curve

An aerobics instructor thinks that his class has an unusually low resting heart rate. If μ = 72 bpm and σ = 8 bpm, and his class of 14 pupils has a mean heart rate (X) of 66,

a) What is the **z score** for the aerobics class?

b) What is the **probability** of randomly selecting a group of 14 people with a mean resting heart rate **lower** than the mean for the aerobics class?



APA	Your HW-Unit 0 article revisited

Re-read your article you selected for Unit 0 HW (do not re-submit it).
Summarize any mention or use of the concepts in the above chapters.

*(you may choose to **type** this summary and include a printed copy here instead of hand writing OR upload a typed document to CANVAS)*