Math 217 Final Exam

Fall 2017

(150 pts possible)

[redacted] Name (4 pts)

1. For the data below compute the statistics requested.

1 pt ea. except "g" is 3 pts

You can use your calculator.

1.8, 2.2, 2.7, 3.4, 3.8, 3.9, 4.1, 4.3, 4.5, 4.8,
5.3, 5.6, 5.9, 6.3, 6.8, 7.4, 7.9, 8.5

a. mean: _____ ← 2 decimal places

b. sample standard deviation: _____

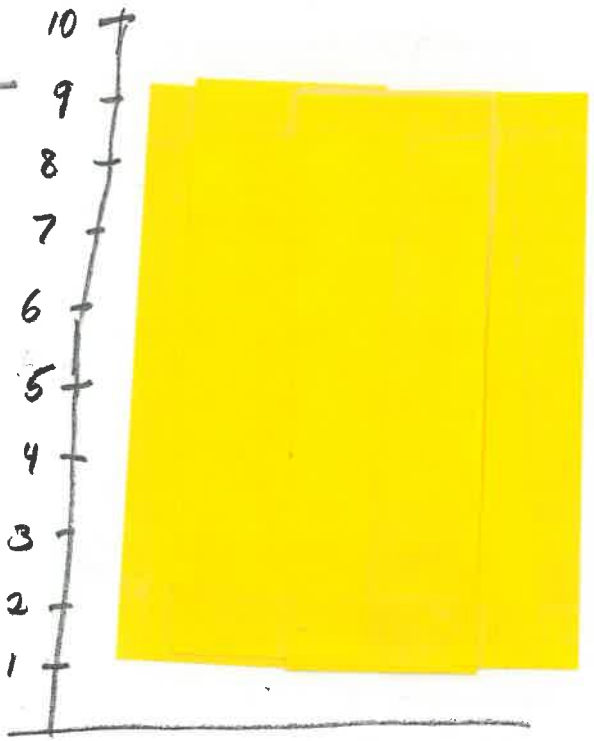
c. median: _____

d. Q1: _____

e. Q3: _____

f. IQR: _____

g. draw the boxplot and label the lines in the boxplot AND include the values ⟹



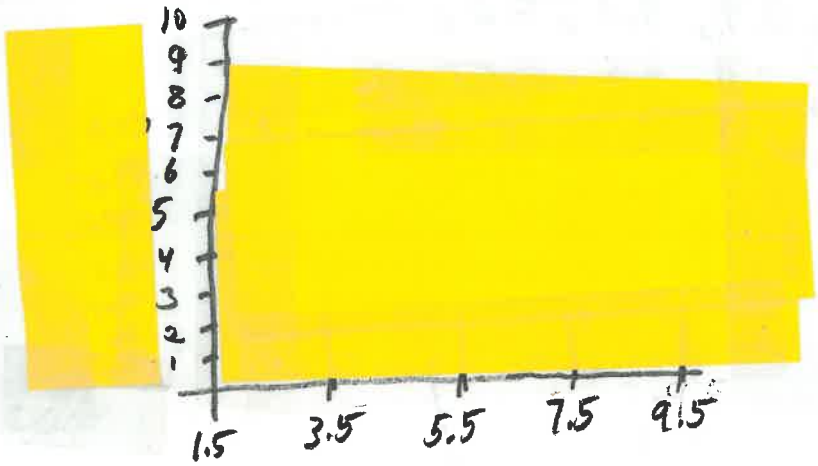
h. what rules of thumb are used to identify potential outliers?

i. What would be the values computed using the formulas in 1.h. be for this data?

2. For the data in problem 1, with two more values: 8.8, 9.1 make a frequency histogram using the intervals 1.5 to 3.5 : ____
3.5 to 5.5 : ____
5.5 to 7.5 : ____
7.5 to 9.5 : ____

a)

3pts



b) Change the frequency histogram above to a relative frequency histogram.

2pts

c) Using the relative frequency histogram in 2.a., compute

2pts $P(3.5 < X < 7.5) =$

2pts $P(X < 3.5) =$

3. Toss an unfair coin (where $p = .3$) 20 times and count the number of heads. If I repeat the process 100 times I could make a frequency histogram, and then a relative frequency histogram.

a) Why don't I always get 6 heads when I toss the coin 20 times? 2pts

b) There is a probability distribution for the number of heads that occur in 20 tosses. What is the relationship between the relative frequency histogram and that probability distribution AND what "Law" describes that relationship. 3pts

c. What is the probability of the following outcomes 2pts each

i) $P(X = 6 \text{ heads}) =$

ii) $P(X \leq 5 \text{ heads}) =$

iii) $P(X \geq 7 \text{ heads}) =$

iv) $P(3 \leq X \leq 9 \text{ heads}) =$

d. for the same binomial probability distribution where $n = 20$ and $p = .3$ compute by hand and show your work $P(X = 5)$

4pts

4. Compute the number of ways you can draw 4 balls out of a bag of 10 balls (each with a different number on the ball so you can tell them apart), if

a) order does not matter:

2pts

b) order matters:

2pts

5. For a binomial experiment/trial where $n = 100$ and $p = .92$

a) compute $\mu =$

2pts

b) compute $\sigma =$

2pts

c) Can you use the normal distribution to approximate probabilities for this distribution? Yes No ← circle one

3pts

If yes, use the normal distribution to approximate $P(X \geq 95)$

If no, explain in detail why not

6. For a normal distribution with $\mu = 200$ and $\sigma = 10$

2pts each

a. approximate $P(180 \leq X \leq 220)$ without using a calculator

b. approximate $P(X \geq 210)$ without using a calculator

c. approximate $P(X \leq 170)$ without using a calculator

d. If $X = 185$, compute its z-score and explain what it means.

7. For a normal distribution with $\mu = 250$ and $\sigma = 15$

(2pts ea) compute <u>with</u> your calculator

a. $P(225 \leq X \leq 275) =$

b. $P(X \leq 225) =$

c. $P(X \geq 280) =$

d. Suppose you take a sample of size $n = 16$ from this distribution and compute the sample mean, $\bar{x}$

i. What is the mean of the Sampling Distribution for $\bar{x}$?

$\mu_{\bar{x}} =$

ii. and what is the standard deviation of the probability distribution for $\bar{x}$?

$\sigma_{\bar{x}} =$

iii. Compute $P(\bar{x} \geq 260)$

iv. Compute $P(245 \leq \bar{x} \leq 255)$

v. What is another name for the standard deviation of the probability distribution for $\bar{x}$?

8. When working with a sample proportion, $\hat{p}$, taken from a population where the true proportion of "successes" is p.

20 pts extra

a. What is $\mu_{\hat{p}}$ =

b. What is $\sigma_{\hat{p}}$ =

c. So if $p = .8$, compute $\mu_{\hat{p}}$ =
and compute $\sigma_{\hat{p}}$ =

d. if $n = 100$ and $p = .8$ and $X = 75$,

i. compute $\hat{p}$ =

ii. compute a 95% Confidence Interval for p based on the $\hat{p}$ above.

iii. What is the margin of error for p based on the confidence interval above?

iv. Compute $P(\hat{p} \leq .75)$ if $p = .8$ (hint: you can do it by hand OR use a test of hypothesis)

9.a. What does the Central Limit Theorem state?

2 pts each

b. If X is normally distributed, how large does n need to be before you can use the normal distribution to compute probabilities for $\bar{X}$?

c. If X is ~~not~~ normally distributed, how large does n need to be before you can use the normal distribution to compute probabilities for $\bar{X}$?

~~d. If X is not skewed, how large does n need to be before you can use the normal distribution to compute probabilities for $\bar{X}$?~~

e. Before you can use the normal distribution to approximate probabilities for the binomial distribution, what needs to be true?

10. Suppose the average life expectancy is 80.2 years (2 pts each) ~~with a standard deviation of 15.6 years~~ (for adults who are 21 years old). Suppose this distribution is normally ~~not skewed~~ distributed. You believe that Seventh Day Adventists may have a longer lifespan because of their vegan diet and other healthy habits. You conduct a small study of 16 randomly selected deaths of Seventh Day Adventists with the following results: the mean age of the 16 subjects at death was 89.4 years with a sample standard deviation of 15.6 years. Conduct a test of hypothesis to analyze your results.

a. State your null and alternate hypotheses:

b. State the test you used on the TI-83/84 calculator to analyze the results:

c. What is the p-value of this test?

d. State your conclusion at the 1% significance level:

e. What is the probability you made a Type 1 error?

f. Construct a 99% Confidence Interval for the mean age at death for this group based on the data?

11. You would like to compare the average lifespan of Japanese to South Koreans. You randomly select 24 obituaries of Japanese and 29 of South Koreans with the following results.

(2 pts each)

Japanese	South Koreans
$\bar{x} = 88.7$	$\bar{x} = 82.9$
$s = 16.2$	$s = 15.6$

You do not know the population standard deviations of these groups, but in each case the sample data does not appear to be skewed, and is approximately normally distributed. Conduct a test of hypothesis to see if Japanese seem to live longer than South Koreans.

a. State your null and alternative hypotheses:

b. State the test you used on the TI-83/84 calculator.

c. What is the p-value for your test?

d. State your conclusion at 5% Significance Level:

e. Construct a 95% Confidence Interval for the difference in the population means.

12. You want to see if consuming caffeine affects performance of female swimmers in the 50 meter freestyle. You believe all the data will be approximately normally distributed. You test 18 swimmers by having them swim the 50 meter freestyle 20 minutes after taking a caffeine pill and without taking the pill. (You test them on two different days with half having caffeine one day and half not to help ensure randomness.) Here are your results.

Swimmer	Time w/ caffeine (group 1)	Time w/o caffeine (group 2)	Difference
1	33.29 (sec)	35.14 (sec)	1.85 (sec)
2	34.02 (sec)	33.79 (sec)	-0.23 (sec)
⋮	⋮	⋮	
18	31.06 (sec)	33.86 (sec)	2.80 (sec)

$\bar{x}_{diff} = 2.06$ (sec)

$s_{diff} = 2.37$ (sec)

You want to test if the difference is greater than 0; i.e., if group 1's times are less than group 2 (on the average).

a. State your null and alternative hypotheses:

b. Which test do you use on the TI-83/84 calculator?

c. What is the p-value for your test?

d. What is your conclusion?

e. Give a 95% Confidence Interval for the mean differences

13. You believe the proportion of natural red haired people is higher in Scotland than Ireland. You randomly select records from both countries with the following results:

(2 pts each)

Scotland	Ireland
$n = 1017$	$n = 987$
$x = 134$	$x = 102$
$\hat{p} =$ _____	$\hat{p} =$ _____

a. State your null and alternative hypotheses:

b. State the test on the TI 83/84 you used to analyze your results:

c. State your conclusion at the 5% significance level.

d. What was your p-value for this test?

e. Give a 95% Confidence Interval for the difference in the population proportions of red heads.

f. Would a 90% confidence interval be wider or more narrow?

14. Go to the computer lab and perform a linear regression/correlation of the following data (using Excel or any other tool you want to).

It has often been said that crickets chirp faster the warmer the temperature. You've collected data for crickets in your yard on various nights. The X values are the temperature in Fahrenheit and the Y values are the chirps per minute. (I made this up.) Plot the data, show a best-fit linear regression line on the plot, compute the formula for the line and for R^2, and also compute R.
Then use the regression line to compute the predicted # of chirps/min if the temperature is 80°F. The data set is on Moodle.

1pt. Best fit regression line: y = _____

1pt. How many chirps/min would you predict for a temperature of 80°F? _____

1pt. What is R^2: _____

1pt. What does R^2 represent?

1pt. What value is R (the correlation coefficient)? _____

3pts. Print off the plot with the regression line and R^2 and turn it in.

.14, (alternate problem) at a cadmium factory in India (this data is made up)

(2 pts each) community near a the incidence of pancreatic cancer has been 5 cases in 20,000 residents, so $\hat{p} = \frac{5}{20000} = .00025$. The incidence of pancreatic cancer throughout India is 2 cases per 100,000 residents, so $p = \frac{2}{100000} = .00002$. You would like to see if the incidence of pancreatic cancer is this community is significantly different than in the population as a whole.

(alternative #14, do one or the other)

a. State your null and alternate hypotheses:

b. Can you use the 1-Prop Z test on your calculator to compute the p-value for this data? Yes No ← circle one
Why or why not?

c. Compute $P(X \geq 5)$ for your sample of 20,000 residents if the true proportion is .00002.

d. You don't want there to be much chance of making a type 1 error, so you test this using a 1% significance level. State your conclusion.