

FINAL

There are 300 points on this exam. Be sure to show all necessary work on each problem and to give complete sentence answers to the discussion questions.

1. (18 pts) The high temperatures recorded on an May day for 22 U.S. cities are given below.

80 76 77 72 76 105 85 72 82 79 62 82 99 68 70 90 82 78 51 68 66 72

A. Give a stemplot for this data.

B. What was the median high temperature on this day?

2. (12 pts) The 30 individuals in the list below have qualified for the WXYZ-FM free trip to see Foot Fungus in concert on June 18. Use the random digits table, line 142, to select 4 winners. Be sure to show your work and indicate the individuals who were selected.

Bagwell	Hayes	Sanchez
Banks	Hernandez	Schilling
Bates	Johnson	Slocumb
Bell	Jordan	Sosa
Charlton	McRae	Swift
Cooper	Miller	Thompson
Daulton	Munoz	Valdes
Duncan	Parent	Walker
Eisenreich	Perry	Weiss
Gonzalez	Roberson	Wilkins

Work area:

Those selected are

3. (34 pts) (Real data!) Radioactive wastes from a plutonium production plant were stored in an open pots near a river. However some of the waste may have leaked into the river, causing adverse health reactions in those who lived downstream. To study this possibility, researchers led by Fadeley computed an index of exposure for each county (really its residents) and related that to cancer mortality (per 100,000 person-years) for that county. The data is as follows:

x - Exposure Index	8.3	6.4	3.4	3.8	2.6	11.6	1.2	2.5	1.6
y - Cancer Mortality	210	178	130	162	130	207	113	147	137

- A. Draw a scatterplot for this data.
- B. Why is the index of exposure the logical choice for the x -variable?

For the above data, the correlation r is 0.93 and the least-squares regression line is $\hat{y} = 114.7 + 9.23x$

- C. Sketch the regression line on the graph with your scatterplot.
- D. i. For each unit increase in the exposure index, by how much does the cancer mortality increase?
- ii. Interpret the "114.7" in the context of this problem.
- iii. What percent of the variation in counties' cancer mortality rates is explained by the least squares regression on exposure index?
4. (40 pts) Platelet count for normal, healthy four-month-old infants average 320 (billions/liter) with a standard deviation of 90. The distribution of counts follows a normal distribution.
- A. What is the probability that a randomly selected infant has platelet count less than 250?
- B. A group of 8 four-month-old infants with a certain genetic abnormality have an average count of 246 billions per liter. Is this sufficient evidence to abnormality have a lower platelet count? This calls for a statistical test of $H_0: \mu = 320$ versus $H_a: \mu < 320$.
- i. What exactly is " μ " in the above hypotheses? (Explain in a sentence.)
- ii. Assume that $\sigma = 90$ in this population. Carry out the statistical test described in (B): given your P-value and state your conclusion.
- iii. Suppose we don't assume a value for σ , but instead calculate the sample standard deviation s for the 8 genetically abnormal infants, getting $s = 80$. Carry out the statistical test described in (B): given your P-value and state your conclusion.

5. (12 pts) In a survey of Math 111 students, the students were asked "How far is it from Bradley Hall to your home (in miles)?" . The data from this question had a mean of 359 miles and a standard deviation of 943 miles.
- A. Will the distribution of this data be symmetric, right-skewed, or left-skewed? why?
- B. Will the median distance be less than, great than, or equal to the mean distance of 359 miles? Explain.
6. (22 pts) A clinic has a group of patients with high blood pressure some of whom are taking medication A and some of whom are taking medication B. A survey is conducted among patients of each medication to determine the extent of side effects they experiencing.
- A. Explain why the above study is not an experiment.
- B. What is the explanatory variable? What is the response variable(s)?
- C. Suppose that a second group of patients is available for whom no medication has yet been assigned. Suppose that these patients are eligible for either medication. Describe a good experiment, using appropriate principles, for testing whether side-effects differ with the type of medication.
7. (9 pts) Explain carefully what the phrase "95% confidence interval" means.
8. (23 pts) In a lifestyle survey of (a SRS of) 700 adult residents of the city of Metropolis, 22% of those surveyed used mass transit
- A. Verify that this situation meets the requirements to use the usual formulas for inference for a population.
- B. Find a 95% confidence interval for the true proportion p of adult Metropolis residents who use mass transit.

9. (30 pts) A traffic engineer wants to know the daily vehicle count on Highway One near the town of Metropolis. An automatic traffic counter records data for five weekdays as follows:

6000 5200 5000 4800 6500

A. For the data, find

i. the sample mean

ii. the sample standard deviation s (Show the necessary work using the algorithm from class.)

iii. the standard error of the mean

B. Find a 95% confidence interval for the true mean daily vehicle count on Highway One.

Problem 10-19 are multiple choice. Write the letter of the correct (or best) answer in the space provided to the left of the question. (7 pts each)

_____ 10. Which of the following confidence intervals for the mean in a normal population will have the smallest margin of error? (SRS with the same σ for each choice)

- A. 95% confidence interval with $n = 25$
- B. 95% confidence interval with $n = 50$
- C. 99% confidence interval with $n = 25$
- D. 99% confidence interval with $n = 50$
- E. all of the above will have the same margin of error

_____ 11. The number x of pets owned by Americans who own pets has the following distribution:

# of pets x	1	2	3	4	5
probability $p(x)$	0.40	0.45	0.10	0.04	0.01

What is μ , the mean number of pets owned by Americans who own pets?

- A. 1.8
- B. 2.0
- C. 2.3
- D. 2.5
- E. 3.0

- _____ 12. Suppose Judy's measured potassium level varies according to the normal distribution with mean $\mu = 3.7$ and standard deviation $\sigma = 0.25$. Judy will be diagnosed as hypokalemic if her average measured potassium over 4 days falls below 3.5. What is the probability that Judy will be diagnosed as hypokalemic? (one measurement per day.)
- A. .0548
 - B. .2119
 - C. .7881
 - D. .8000
 - E. .9452
- _____ 13. Considering the 8 observations 1, 2, 4, 5, 6, 8, 10, 20, find the first quartile of these observations.
- A. 1.5
 - B. 2.0
 - C. 2.5
 - D. 3.0
 - E. 3.5
- _____ 14. The entire group of individuals about whom information is desired is the
- A. distribution
 - B. statistic
 - C. population
 - D. parameter
 - E. sample
- _____ 15. The current government economic survey wishes to know, with 95% confidence, the average household income in Peoria to within \$500. How many randomly selected households must be used if the standard deviation of household income in Peoria is \$15,000?
- A. 30
 - B. 59
 - C. 333
 - D. 900
 - E. 3458

- _____ 16. Considering that the odds are slightly in the casino's favor, a gambler has the best chance of winning (i.e. leaving with more money than she arrived with) if she
- A. plays only once
 - B. plays only twice
 - C. plays 10 times
 - D. plays 50 times
 - E. plays 100 times
- _____ 17. In one of the Against All Odds videotapes, a new poem was discovered that might be attributed to William Shakespeare. A statistical analysis of this possibility was given, based on the number of new words used in this poem and in various Shakespeare works. Which of the following best expresses the conclusion for this analysis?
- A. Shakespeare wrote the poem.
 - B. Shakespeare didn't wrote the poem.
 - C. The number of new words is consistent with the statement that Shakespeare wrote the poem.
 - D. The number of new words is not consistent with the statement that Shakespeare wrote the poem.
 - E. Shakespeare knew more words than anyone else.
- _____ 18. "The reversal of the direction of an association when data from several groups is combined" describes ...
- A. conditional distribution
 - B. confounding
 - C. two-way table
 - D. Law of Large Numbers
 - E. Simpson's paradox
- _____ 19. Suppose we want to test $H_0: \mu = 0$ versus $H_a: \mu > 0$ with a small sample when the standard deviation is unknown. Based on a sample of $n = 3$ observations, we obtain a t statistic of 5.00 Then the P-value of the test is ...
- A. less than .0002
 - B. between .001 and .005
 - C. between .005 and .01
 - D. between .01 and .02
 - E. between .02 and .04

20. (30 pts) Matching! (5 pts each) Match each definition or description to the appropriate and the best term from the list below:

_____ the distribution of values taken by the statistic in all possible samples of the same size from the same population

_____ an observation with a large residual

_____ when the standard deviation of a statistic is estimated from data

_____ the use of impersonal chance to assign experimental units to treatments

_____ a sample consisting of people who choose to respond

_____ when the sample size (in an SRS) is large, the sample distribution of \bar{x} is approximately normal

A. Law of Large Numbers

H. confidence interval

B. Law of Averages

I. bias

C. sampling distribution

J. influential observation

D. Central Limit Theorem

K. outlier

E. stratified sample

L. randomization

F. voluntary response sample

M. standard deviation

G. standard error

N. histogram