

Statistical Methods Final Exam Practice

Chapters 2 – 8, 12, 13, 18 – 24

Use this data table for questions 3 – 9 of the practice exam. The data is not in numerical order since you do not have an in-class time constraint.

Geographic Latitude and Mean August Temperature

City	Latitude	Aug. Temp
Miami FL	26	83
Houston TX	30	82
Mobile AL	31	82
Phoenix AZ	33	92
Dallas TX	33	85
Los Angeles CA	34	75
Memphis TN	35	81
Norfolk VA	37	77
San Francisco CA	38	64
Baltimore MD	39	76
Kansas City MO	39	76
Washington DC	39	74
Pittsburgh PA	40	71
Cleveland OH	41	70
New York NY	41	76
Boston MA	42	72
Syracuse NY	43	68
Minneapolis MN	45	71
Portland OR	46	69
Duluth MN	47	64

I. Specify whether each of the following is a categorical variable or a quantitative variable.

1. Your favorite color.
2. Your shoe size.

II. Using both the Geographical Latitude and the August temperatures, find the following:

3. Create a stem and leaf display
4. Mean
5. Range
6. Standard Deviation
7. Five Number Summary
8. Draw a boxplot

III. Construct the following:

9. Use the August Temperatures to create a histogram with a bin width of five, describe its distribution.

Statistical Methods Final Exam Practice

IV. Adult IQ Scores have a normal distribution with a mean of 100 and a standard deviation of 15.

10. Using the Empirical Rule between what two IQ scores would 95% of adults fall?
11. Calculate the z-score for an adult with an IQ score of 67.

V. A survey of 1024 American households found that 48% of the households own a computer.

12. Identify the population.
13. Identify the sample.
14. Classify the “48% own a computer” as a parameter or a statistic.

VI. Data on salaries in the public school system are published annually in *National Survey of Salaries and Wages in Public Schools*. The mean annual salary for classroom teachers is \$40,133 with a standard deviation of \$8000.

15. Determine the standard error for samples of size 64.
16. Find the probability that the mean of the samples is greater than \$43,000.
17. Find the probability that the mean of the samples is between than \$41,000 and \$43,000.

VII. Do people lie about voting? In a survey of 1002 people, 701 people said that they voted in the recent presidential election (based on data from ICR Research Group). Voting records show that 71% of eligible voters actually did vote. Using these survey results, find the following:

18. The sample proportion of people who say they voted.
19. Construct a 95% confidence interval for the population proportion.
20. Determine whether the survey results are consistent with the actual voter turnout of 71%. Justify your answer with a statement of explanation.

VIII. In a survey of 2000 adults from the U.S. age 65 and over, 1320 received a flu shot.

21. Find the sample proportion of those receiving a flu shot.
22. Construct a 90% confidence interval for the population proportion.

IX. A recent study claimed that at least 15% of junior high students are overweight. In a sample of 160 students, 18 were found to be overweight. You think the claim is too high, perform an appropriate hypothesis test.

23. Write the null and alternative hypotheses.
24. Calculate $n\hat{p}$ and $n\hat{q}$.
25. Calculate the sample proportion, its z-score and find the P-value.
26. Write the conclusion.
 - a) Decision about hypothesis.
 - b) Conclusion with regards to context.

Statistical Methods Final Exam Practice

X. Construct the following confidence intervals.

27. There has been debate among doctors over whether surgery can prolong life among men suffering from prostate cancer, a type of cancer that typically develops and spreads slowly. In the summer of 2003, results of some Scandinavian research were published. Men diagnosed with prostate cancer were randomly assigned to either undergo surgery or not. Among 347 men, who had surgery, 16 eventually died, compared with 31 of 348 men who did not have surgery.

- Was this an experiment or observational study? Explain.
- Create a 95% confidence interval for the difference in rates of death for the two groups of men.
- Is there evidence that surgery may help?

28. In 1991, researchers at the National Cancer Institute released the results of a study that investigated the effect of weed-killing herbicides on house pets. They examined 827 dogs from homes where an herbicide was used on a regular basis, diagnosing malignant lymphoma in 473 of them. Of the 130 dogs from homes where no herbicides were used, only 19 were found to have lymphoma. Construct a 95% confidence interval for this difference.

XI. Perform the following hypothesis tests.

29. A survey of 430 randomly chosen adults found that 21% of the 222 men and 18% of the 208 women had purchased books online. Is there evidence that men are more likely than women to make online purchases of books?

- Write appropriate hypotheses.
- Find the P-value.
- Decision about hypothesis.
- Conclusion with regards to context.

30. A company with a fleet of cars hopes to keep gasoline costs down, and sets a goal of attaining a fleet average of at least 26 miles per gallon. To see if the goal is being met they check gasoline usage for 50 company trips chosen at random, finding a mean of 25.02 mpg and a standard deviation of 4.83 mpg. Is this strong evidence that they have failed to attain their fuel economy goal?

- Write appropriate hypotheses.
- Find the P-value.
- Decision about hypothesis.
- Conclusion with regards to context.

31. In 2001, one country reported that among 3132 white women who had babies, 94 were multiple births. There were also 70 multiple births to 606 black women. Does this indicate any racial difference in the likelihood of multiple births?

- Write appropriate hypotheses.
- Find the P-value.
- Decision about hypothesis.
- Conclusion with regards to context.

Statistical Methods Final Exam Practice

XII. Follow the directions provided.

32. The table shows the duration (in minutes) of Old Faithful's eruptions and the times (in minutes) until the next eruption.

Duration, x	Time, y	Duration, x	Time, y
1.8	56	3.78	79
1.82	58	3.83	85
1.88	60	3.87	81
1.9	62	3.88	80
1.92	60	4.1	89
1.93	56	4.27	90
1.98	57	4.3	84
2.03	60	4.3	89
2.05	57	4.43	84
2.13	60	4.43	89
2.3	57	4.47	86
2.35	57	4.47	80
2.37	61	4.53	89
2.82	73	4.55	86
3.13	76	4.6	88
3.27	77	4.6	92
3.65	77	4.63	91
3.7	82		

- Find the equation of the regression line for predicting the time until the next eruption.
- Explain in this context what the slope of this line means.
- Predict the time until the next eruption when the duration is 4.61 minutes.

33. Highway planners investigated the relationship between traffic density (# of automobiles per mile) and the average speed of the traffic on a moderately large city thoroughfare. The data were collected at the same location at 10 different times over a span of 3 months. They found a mean traffic density of 68.6 cars per mile (cpm) with standard deviation of 27.07 cpm. Overall the cars' average speed was 26.38 mph with a standard deviation of 9.68 mph. These researchers found the regression line for these data to be $\widehat{speed} = 50.55 - 0.352cars$.

- Predict the average speed of traffic on the thoroughfare when the traffic density is 50 cpm.
- What is the value of the residual for a traffic density of 56 cpm with an observed speed of 32.5 mph?

Statistical Methods Final Exam Practice

34. A Statistics instructor created a linear regression equation to predict students' final exam scores from their midterm exam scores. The regression equation was $\widehat{fin} = 10 + 0.9mid$.
- If Susan scored a 70 on the exam, what did the instructor predict for her score on the final?
 - Susan scored an 80 on the final. Using the predicted score from part a, find the residual.
 - How many points would someone need to score on the midterm to have a predicted final score of 73?
35. Marine biologists warn that the growing number of powerboats registered in Florida threatens the existence of manatees. The following data was reported in 2002:

Year	Manatees Killed	Powerboat Registrations (in 1000s)
1982	13	447.0
1983	21	459.6
1984	24	481.0
1985	16	497.9
1986	24	512.6
1987	20	512.3
1988	15	526.5
1989	34	558.6
1990	33	585.3
1992	33	613.5
1993	39	645.5
1994	43	675
1995	50	711
1996	47	719
1997	53	716
1998	38	716
1999	35	716
2000	49	735
2001	81	860

- In this context, which do you think is the explanatory variable?
- Make a scatterplot of these data and describe the association you see.
- Find the correlation between boat registrations and manatee deaths.
- Find the equation of the regression line.
- Interpret the slope of your model.
- How accurately did your model predict the high number of manatee deaths in 2001?
- Which is better for manatees, positive residuals or negative residuals?

Statistical Methods Final Exam Practice

36. The Minnesota Department of Transportation hoped that they could measure the weights of big trucks without actually stopping the vehicles by using a newly developed “weigh-in-motion” scale. After installation of the scale, a study was conducted to find out whether the scale’s readings correspond to the true weights of the trucks being monitored. The regression equation is $\widehat{Wt} = 10.85 + 0.64scale$, where both the scale reading and the predicted weight of the truck are measured in thousands of pounds.

- Estimate the weight of a truck if this scale read 31,200 pounds.
- If that truck actually weighed 32,120 pounds, what was the residual?
- If the scale reads 35,590 pounds, and the truck has a residual of -2440 pounds, how much does it actually weigh?

XIII. Analysis

37. What fraction of cars is made in Japan? The computer output below summarizes the results of a random sample of 50 autos. Explain carefully what it tells you.

z – Interval for proportion, with 90% confidence,

$$0.29938661 < p(\text{Japan}) < 0.46984416$$

38. A medical researcher measured the pulse rates (beats per minute) of a sample of randomly selected adults and found the following Student’s t-based confidence interval; explain carefully what it tells you:

With 95% Confidence,

$$70.887604 < \mu(\text{Pulse}) < 74.497011$$

39. A company’s old antacid formula provided relief for 70% of the people who used it. The company tests a new formula to see if it is better and gets a P-value of 0.27. Is it reasonable to conclude that the new formula and the old one are equally effective? Explain.

40. The United States Golf Association (USGA) sets performance standards for golf balls. For example, the initial velocity of the ball may not exceed 250 feet per second when measured by an apparatus approved by the USGA. Suppose a manufacturer introduces a new kind of ball and provides a sample for testing. Based on the mean speed in the test, the USGA comes up with a P-value of 0.34. Explain in this context what the “34%” represents.

ANSWERS:

- categorical
- Quantitative
- First is for latitude and the second is for temperature.

2		6										
3		0	1	3	3	4	5	7	8	9	9	9
4		0	1	1	2	3	5	6	7			

6		4	4	8	9							
7		0	1	1	2	4	5	6	6	6	7	
8		1	2	2	3	5						
9		2										

Statistical Methods Final Exam Practice

4. $\bar{x} = 37.95, \bar{y} = 75.4$ 5. 21, 28
 6. $s = 5.5864, s = 7.1994$
 7.

Min	Q1	Median	Q3	Max	Min	Q1	Median	Q3	Max
26	33.5	39	41.5	47	64	70.5	75.5	81.5	92

9. symmetric, unimodal, no outliers 10. 70 to 130
 11. $z = -2.2$ 12. All American households
 13. 1024 American households 14. Statistic
 15. $SE = 1000$ 16. $P(\bar{y} > 43000) = 0.002$
 17. $P(41000 < \bar{y} < 43000) = 0.191$ 18. $\hat{p} = 0.6996$
 19. (0.6712, 0.7280)
 20. It appears the survey is accurate since the percentage of actual voters falls within the confidence interval.
 21. $\hat{p} = 0.66$ 22. (0.6426, 0.6774)
 23. $H_0: p = 0.15$ 24. $n\hat{p} = 18$
 $H_A: p < 0.15$ $n\hat{q} = 142$
 25. $\hat{p} = 0.1125, z = -1.3284, P - value = 0.0920$
 26. a) Reject null hypothesis
 b) Conclude that less the 15% of junior high students are overweight
 27. a) Experiment because they were divided into groups having surgery or no surgery
 b) (0.0058, 0.0802)
 c) Yes there is evidence that the surgery helps, since 0 is not in the confidence interval.
 28. (0.35633, 0.49525)
 29. a) $H_0: p_m = p_w$ b) $P - value = 0.1883$
 $H_A: p_m > p_w$
 c) Retain null hypothesis
 d) There is no evidence to support that men are more likely than women to purchase books online.
 30. a) $H_0: \mu = 26$ b) $P - value = 0.0789$
 $H_A: \mu < 26$
 c) Reject null hypothesis
 d) There is strong evidence to say they have failed to attain their fuel economy goal.
 31. a) $H_0: p_w = p_b$ b) $P - value = 5.2 \times 10^{-21}$
 $H_A: p_w \neq p_b$
 c) Reject null hypothesis
 d) There is strong evidence to indicate a racial difference in the likelihood of multiple births.
 32. a) $\hat{y} = 35.301 + 11.824x$
 b) For each additional minute the eruption lasts the time until the next eruption is increased by 11.824 minutes.
 c) 89.810
 33. a) 32.95 mph b) 1.662
 34. a) 73 b) 7 c) 70
 35. a) Powerboat registrations b) Shows a strong, linear, and positive association
 c) 0.924 d) $\hat{y} = -45.893 + 0.132x$

Statistical Methods Final Exam Practice

- e) For each additional 1000 registered powerboats there is an increase in 0.132 manatee deaths.
- f) 67.627 which is under by 13.373
- g) Negative residuals because the actual number of deaths is less than the predicted.
36. a) 30,818
b) The model underestimated the weight by 1302 pounds.
c) 31,187.6
37. On the basis of this sample, we are 90% confident that the proportion of Japanese cars is between 29.9% and 47.0%.
38. Based on this sample, we can say, with 95% confidence, that the mean pulse rate of adults is between 70.9 and 74.5 beats per minute.
39. No, we can say only that there is a 27% chance of seeing the observed effectiveness just from natural sampling variation. There is no *evidence* that the new formula is more effective, but we can't conclude that they are equally effective.
40. If in fact this ball meets the velocity standard, then 34% of all samples of this size would have mean speeds at least as high as was recorded in this sample.

Level of Confidence	z^*
80%	1.282
90%	1.645
95%	1.96
98%	2.326
99%	2.576

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$CI = \hat{p} \pm ME$$