Final Exam for Stat 11

Solutions

Dots/Dash/Slash

Part 1: Multiple Choice

1/3/4. A candy company claims that its bags of mixed suckers are 20% strawberry, 30% cherry, 15% apple, 10% lemon, and 25% grape. A bag was purchased, and the number of each type of flavor was recorded in the chart below. Determine which statistical test is appropriate for this situation.

(a) Chi-square test of independence

(b) **Chi-square goodness-of-fit test**

(c) Chi-square test of homogeneity

(d) One sample t-test for an unknown mean

(e) Two sample t-test for a difference in means

Rubric items:

- little partial for (a) or (c)
- no credit for (d) or (e)

2/2/20. Doctors at a technology research facility randomly assigned equal numbers of people to use computer keyboards in two rooms. In one room a group of people typed a manuscript using standard keyboards, while in the other room people typed the same manuscript using ergonomic keyboards to see if those people could type more words per minute. After collecting data for several days the researchers tested the hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against the one-tail alternative and found a p-value of 0.22. Which of the following statement is TRUE?

(a) There's a 22% chance that people using ergonomic keyboards type more words per minute.

(b) There's a 22% chance another experiment will give these same results.

(c) There's a 22% chance that there's really no difference in typing speed.

(d) The people using ergonomic keyboards type 22% more words per minute.

(e) **None of the above are true.**

Rubric items:

- no partial credit

3\1\6. An entomologist writes an article in a scientific journal which claims that fewer than 6% of male fireflies are unable to produce light due to a genetic mutation. Identify the Type I error in this context.

(a) The error of failing to accept the claim that the true proportion is at least 6% when it is actually less than 6%.

(b) **The error of rejecting the claim that the true proportion is at least** 6% **when it really is at least** 6%**.**

(c) The error of failing to reject the claim that the true proportion is at least 6% when it is actually less than 6%.

(d) The error of rejecting the claim that the true proportion is less than 6% when it really is less than 6%.

(e) The error of accepting the claim that the true proportion is at least 6% when it really is at least 6%.

Rubric items:

- full credit for (b)
- half credit for (d)

4\4\1. In the past, the mean running time for a certain type of flashlight battery has been 8.4 hours. The manufacturer has introduced a change in the production method and wants to perform a hypothesis test to determine whether the mean running time has increased as a result. The hypotheses are:

$$H_0 : \mu = 8.4 \text{ hours} H_A : \mu > 8.4 \text{ hours}.$$

Explain the result of a Type II error.

(a) The manufacturer will decide the mean battery life is 8.4 hours when in fact it is 8.4 hours.

(b) The manufacturer will decide the mean battery life is greater than 8.4 hours when in fact it is greater than 8.4 hours.

(c) **The manufacturer will decide the mean battery life is 8.4 hours when in fact it is greater than 8.4 hours.**

(d) The manufacturer will decide the mean battery life is greater than 8.4 hours when in fact it is 8.4 hours.

(e) The manufacturer will decide the mean battery life is less than 8.4 hours when in fact it is greater than 8.4 hours.

Rubric items:

- no partial credit

5\6\3. Using a t-table, estimate the p-value for a test statistic of 1.76 with 24 degrees of freedom.

(a) 0.9088

(b) 0.0228

(c) **0.0456** (upper tail)

(d) 0.9772

(e) **0.9544** (lower tail)

Rubric items:

- full credit for either (c) or (e)

6\5\8. A certain population is strongly skewed to the right. We want to estimate its mean, so we will collect a sample. Which of the following statements should be TRUE if we use a large sample rather than a small one?

I. The distribution of our sample data will be closer to normal.

II. The sampling model of the sample means will be closer to normal.

III. The variability of the sample means will be greater.

(a) I only

(b) **II only**

(c) III only

(d) I and III only

(e) II and III only

Rubric items:

- half credit for (a)
- full credit for (b)

7\20\2. Based on a sample of 30 randomly selected years, a 90% confidence interval for the mean annual precipitation in one city is from 48.7 inches to 51.3 inches. Find the margin of error.

(a) 0.39 inches

(b) **1.3 inches**

(c) 0.10 inches

(d) 2.6 inches

(e) There is not enough information to find the margin of error.

Rubric items:

- no partial credit

8\19\17. A survey found that 79% of a random sample of 1024 American adults approved of cloning endangered animals. Find the margin of error for this survey if we want 90% confidence in our estimate of the percent of American adults who approve of cloning endangered animals.

(a) 5.43%

(b) 21.4%

(c) **2.09%**

(d) 4.56%

(e) 2.49%

Rubric items:

- no partial credit

9\10\13. A bicycle shop equips 60% of their bikes with a water bottle holder. 55% of the bikes they sell have a kickstand attached to the bike. 34% of the bikes sold have both features. Given that a randomly selected bike has a kickstand, what are the chances that it also has a water bottle holder?

(a) 26%

(b) 56.7%

(c) 34%

(d) **61.8%**

(e) 48%

Rubric items:

- no partial credit

10\12\12. A group of volunteers is recruited to test a new lip balm that prevents chapping in winter. The volunteers are paid $100 to participate. Which of the statements is TRUE?

(a) **The paying of volunteers may indicate that our volunteers are not similar to the general population.**

(b) The researcher should instead pick a SRS to experiment upon.

(c) The researcher needs to convince the volunteers to work for free.

(d) The paying of volunteers invalidates the experiment.

(e) There will no randomization in the experiment.

Rubric items:

- partial credit for (b)

11\9\9. A school district administrator sent a survey to all teachers in the district. Only 30% of the teachers responded to the survey. Which of the following is TRUE?

I. The people that did not responded are likely to be similar to those that did so he should use them as the sample.

II. This survey design suffered from non-response bias.

III. Because he sent the survey to everyone, this is a census and the results can be applied to the whole population.

(a) **II only**

(b) I and II only

(c) I, II, and III

(d) II and III only

(e) I only

Rubric items:

- no partial credit

12\11\11. Suppose a school district decides to randomly test high school students for attention deficit disorder (ADD). There are three high schools in the district, each with grades 9-12. The school board pools all of the students together and randomly samples 250 students. Is this a simple random sample?
(a) No, because we can't guarantee that students from each grade in the sample.

(b) No, because we can't guarantee that students from each school in the sample.

(c) Yes, because the students were chosen at random.

(d) **Yes, because they could have chosen any 250 students from the district.**

(e) Yes, because each student is equally likely to be chosen.

Rubric items:

- partial credit for (c) or (e)

13\13\10. The online MBA director at a large business school surveys a sample of current students to determine their level of satisfaction with the program. She finds that 67% are "very satisfied" with the online program. What is this?

(a) A parameter

(b) A target population

(c) **A statistic**

(d) A margin of error

(e) A sampling frame

Rubric items:

- partial credit for (a)

14\15\15. Shortly after the Sandy Hook Elementary School shooting in December 2012, a local television news program asked viewers to call in with their opinion about gun control. These results were likely biased because of what reason?

(a) **A voluntary response sample**

(b) Measurement error

(c) A bad sampling frame

(d) An undefined target population

(e) Leading questions

Rubric items:

- partial credit for (c) or (d)

15\14\14. When using midterm exam scores to predict a student's final grade in a class, the student would prefer to have which of the following?

(a) A positive residual, because that means the student's final grade is lower than we would predict with the model.

(b) **A positive residual, because that means the student's final grade is higher than we would predict with the model.**

(c) A negative residual, because that means the student's final grade is lower than we would predict with the model.

(d) A residual equal to zero, because that means the student's final grade is exactly what we would predict with the model.

(e) A negative residual, because that means the students final grade is higher than we would predict with the model.

Rubric items:

- partial credit for (d)

16\16\16. A small independent organic food store offers a variety of specialty coffees. To determine whether price has an impact on sales, the managers kept track of how many pounds of each variety of coffee were sold last month. Based on the summary statistics shown below, what percent of the variability in the number of pounds of coffee sold per week can be accounted for by a linear model on price?

	Price	Pounds
Mean	\$8.75	54.50
Standard Deviation	\$3.63	18.33
Correlation	0.927	

(a) **85.9%**

(b) 100%

(c) 92.7%

(d) 54.5%

(e) 4.68%

Rubric items:

- half credit for (c)

17\17\19. What are the residuals?

(a) Residuals are variation in the data that is explained by the model.

(b) Residuals are data collected from individuals that is not consistent with the rest of the group.

(c) None of these statements are correct.

(d) Residuals are possible models not explored by the researcher.

(e) **Residuals are the difference between observed responses and values predicted by the model.**

Rubric items:

- no partial credit

18\18\18. For families who live in apartments the correlation between the family's income and the amount of rent they pay is $r = 0.60$. Which of the following is/are possibly TRUE about this correlation?

I. In general, families with higher incomes pay more in rent.

II. On average, families spend 60% of their income on rent.

III. The regression line passes through 60% of the data points.

(a) **I only**

(b) I, II, and III

(c) I and II only

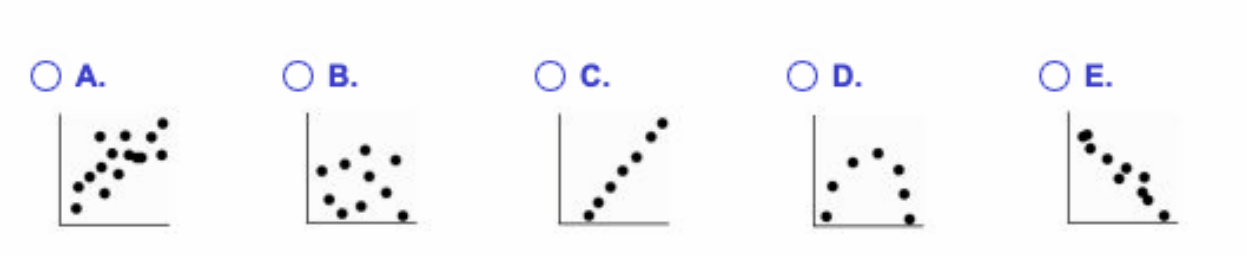
(d) II only

(e) I and III only

Rubric items:

- no partial credit

19\7\7. Which scatterplot shows a strong association between two variables even though the correlation is probably close to zero?



Solution: D

Rubric items:

- no partial credit

20\8\5. Consider the five number summary of hourly wages ($) for a sample of advertising/promotion managers.

Min	Q1	Median	Q3	Max
19.64	29.36	34.18	40.86	57.26

Suppose there had been an error and that the lowest hourly wage is actually $15.50 instead of $19.64. What would be the result caused due to the change?

(a) A decrease in the range

(b) An increase in the mean

(c) A decrease in the IQR

(d) **An increase in the standard deviation**

(e) An increase in the median

Rubric items:

- partial credit for (e)

Part 2: Fill in the blank

21/24/24 Vending machines on a college campus offer a variety of snacks. The purchasing agent believes that each type of snack is equally preferred by students and consequently orders equal quantities. The number of snacks sold from vending machines on this campus for the last six months is shown in the following table. What is the test statistic for determining if the purchasing agent's belief is supported?

$\chi^2 =$ _____.

Solution: 8.012

Rubric Items

- 1 - Correct
- used proportions instead of counts
- typo forgot to square
- first term correct
- 0 - incorrect

22/25/30. A manufacturing process has a 70% yield, meaning that 70% of the products are acceptable and 30% are defective. If three of the products are randomly selected, _____ is the probability that all three of them are acceptable.

Solution: 0.343

Rubric items:

- Correct
- typo in binomial formula o/w correct
- using correct binomial formula but not finding P(X=3) exactly
- Incorrect

23/23/23. A professor divided the students in her business class into three groups: those who have never taken a statistics class, those who have taken only one semester of a statistics class, and those who have taken two or more semesters of statistics. The professor randomly assigns students to groups of three to work on a project for the course. If 35% of the students have never taken a statistics class, 25% have taken only one semester of a statistics class, and the rest have taken two or more semesters of statistics, what is the probability that neither of the first two groupmates you meet has studied any statistics?

Solution: 0.123

Rubric items:

- Correct

- showing reasonable work
- correct answer shown but not selected
- Incorrect

24/21/29. Ten different families are tested for the number of gallons of water a day they use before and after viewing a conservation video. Given the summary statistics below, construct a 90% confidence interval for the mean of the differences.

μ_d	4.8
s_d	5.25

```
## [1] 4.8
```

```
## [1] 5.245104
```

Solution: CI is $[1.8, 7.8]$ so center is 4.8 and ME is 3

Rubric items:

- correct
- correct center
- used z critical value instead of t critical value, o/w correct ME
- used z critical value instead of t critical value and forgot to divide by $\sqrt{n}$
- wrong t critical value, o/w ME correct
- SE given instead of ME
- incorrect center and ME
- incorrect ME
- forgot to divide by $\sqrt{n}$ in SE term
- Comments
  - center counted as correct even though you gave the interval
  - counted as correct even though you didn't give the center

25/22/25. Based on the Normal model for yearly snowfall (in cm) in a certain town $N(57, 8)$, how many cm's of snow would represent the 80th percentile approximately?

Solution: 63.7 answers between 60-65 counted as correct

Rubric items:

- correct
- answer from 55-60 or from 65-70
- answer from 50-55 or from 70-75
- incorrect but work shown
- incorrect

- Comments
  - -0.5 deduction because question specified you should use the Z-table

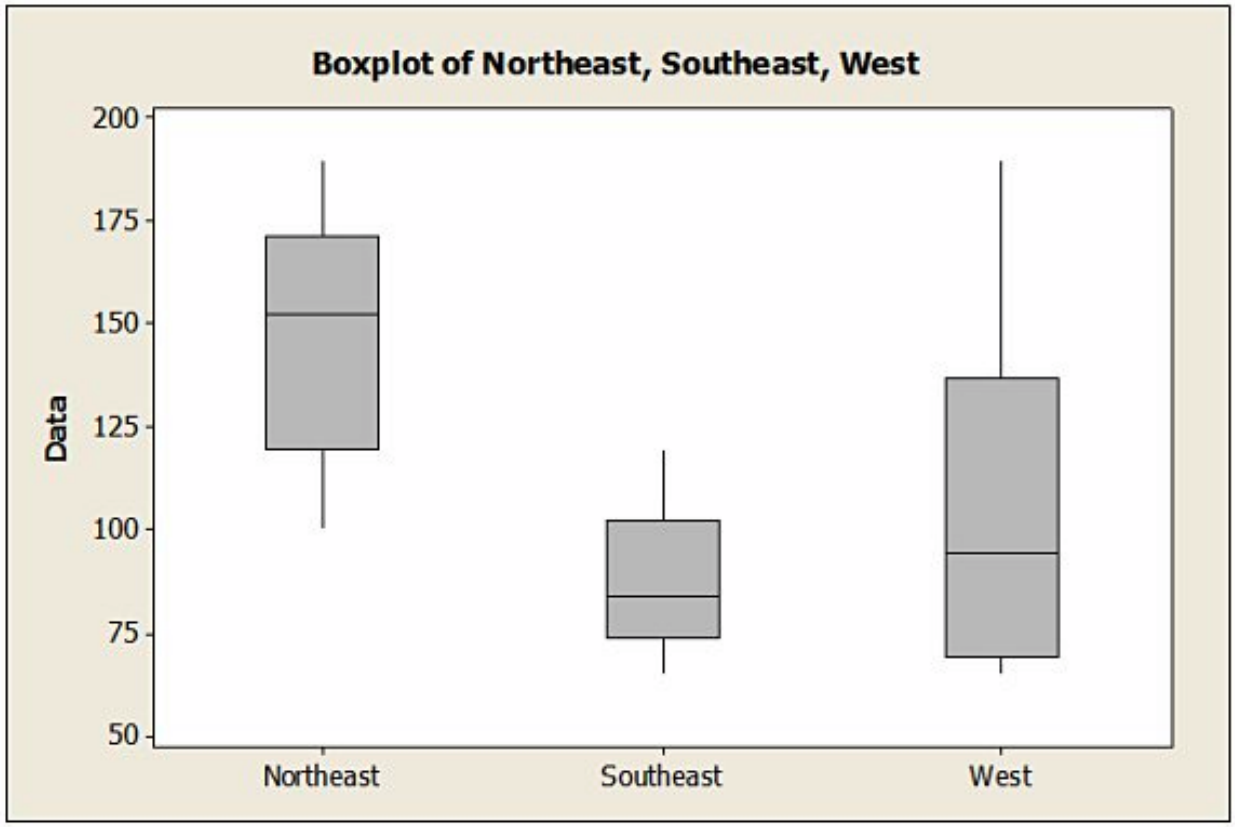
26. The ASQ (American Society for Quality) regularly conducts a salary survey of its membership, primarily quality management professionals. Based on the most recently published mean and standard deviation, a quality control specialist calculated the z-score associated with his own salary and found it was 2.50. This tells him that his salary is _____ standard deviations _____ the average salary.

Solution: 2.5, above

Rubric items:

- 0.5 2.5 st devs
- 0.5 higher than
- 0.4 away from
- 0 neither correct

27. The following boxplots show monthly sales revenue figures (in thousands of dollars) for a discount office supply company with locations in three different regions of the U.S. (Northeast, Southeast, and West).



_____ has the lowest median sales revenue.

Solution: southeast

28. If we transform a numeric data set by adding the same constant to each data value, the _____ and _____ will change but the _____ will stay the same.

Solutions: for first two - mean, median; for third - standard deviation, range, IQR

29. In words (not symbols), the parameters of a Binomial distributed random variable are _____ and _____.

Solution: number of trials, probability of success

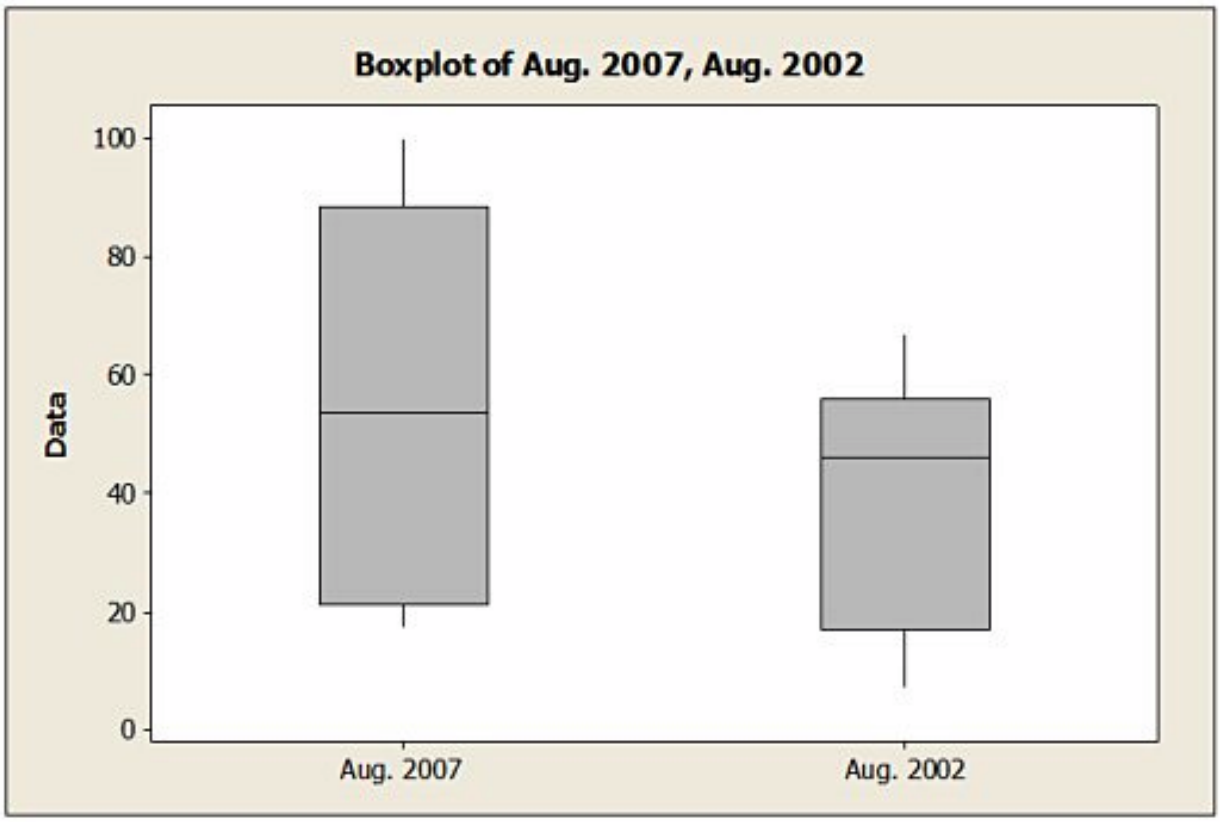
30. In words (not symbols), the parameters of a Normally distributed random variable are _____ and _____.

Solution: mean, variance/standard deviation

Part 3: Free response

To be eligible for partial credit, your answer must show all of your work and/or explain all of your reasoning.

31. The following boxplots show the closing share prices for a sample of technology companies on the first trading days in August 2007 and in August 2002.



Describe the story this plot tells about the closing share prices in these two years. Make sure your answer includes comparisons of variability, centrality, and the shape of the distributions.

Solution:

- Closing prices are more variable in 2007 than in the other year.
- The median closing share price is (slightly) higher in 2007 than in the other year.
- The distribution of closing prices in 2007 appears more symmetric than the distribution of closing prices in 2002 which appear to be skewed left.

Rubric items:

- 0.33 correct statement comparing variability
- 0.33 correct statement comparing centers
- 0.34 correct statement comparing distribution shapes

32. A survey of randomly selected college students found that 45 of the 99 freshmen and 57 of the 100 sophomores surveyed had purchased used textbooks in the past year. Construct a 98% confidence interval for the difference in the proportions of college freshmen and sophomores who purchased used textbooks.

Solution: $[-0.279, 0.049] = -0.115 \pm 2.32 \times 0.07$

Rubric items:

- 0.2 correct center
- 0.4 correct SE
- 0.4 correct critical value 2.32, 2.33, or 2.325
- 0.3 incorrect critical value of 2.05, 2.06, or 2.055

33/32/32. 5% of trucks of a certain model have needed new engines after being driven between 0 and 100 miles. The manufacturer hopes that the redesign of one of the engine's components has solved this problem. State the hypotheses to test the efficacy of the new design and **define the parameters in your hypotheses with words**.

H_0 : _____

H_A : _____

Solution: $H_1 : p = 0.05, H_A : p < 0.05$

Rubric items:

- 0.35 correct null
- 0.35 correct alternative
- 0.3 population parameter correctly defined
- incorrect

34. The number of hours per week that high school seniors spend on homework is normally distributed, with a mean of 10 hours and a standard deviation of 3 hours. 60 students are chosen at random. Let X represent the mean number of hours spent on homework for this group. Plot the probability that X is between 9.8 and 10.4 on the Normal curve below and then use the Z-table to approximate this probability.

Solution:

$$Pr(9.8 < X < 10.4) = Pr(\frac{9.8-10}{3/\sqrt{60}} < Z < \frac{10.4-10}{3/\sqrt{60}}) = Pr(Z < \frac{10.4-10}{3/\sqrt{60}}) - Pr(Z < \frac{9.8-10}{3/\sqrt{60}})$$

$$= P(Z < 1.033) - Pr(Z < -0.516) = 0.8485 - 0.3050 = 0.544$$

Note: $0.8485 - 0.3015 = 0.547$, or $0.8508 - 0.3015 = 0.549$, or $0.8508 - 0.3050 = 0.546$, are all acceptable answers as well.

Rubric items:

- correct center on curve
- correct st devs on curve
- correct shaded shape on curve
- correct SE
- correct probability arithmetic
- standardized upper and lower bounds

35/34/31. An education researcher was interested in examining the effect of the teaching method and the effect of the particular teacher on students' scores on a reading test. In a study, there are two different teachers (Juliana and Felix) and three different teaching methods (method A, method B, and method C). The number of students participating in the study is 258. Students are randomly assigned to a teaching method and teacher. What are the six different treatments?

Solution:

1. Juliana and method A
2. Juliana and method B
3. Juliana and method C
4. Felix and method A
5. Felix and method B
6. Felix and method C

Rubric items:

- correct
- correct answer given but not identified as the only answer
- incomplete but not incorrect
- identified 5 treatments
- identified 3 treatments
- identified 2 treatments
- incorrect

36. Write a question that can be answered with a hypothesis test for the difference in means.

Solutions will vary. Must not involve effect size estimation question. Must involve two populations/categorical varb with two levels. Must involve a numeric variable.