

STATISTICS ~ BUSINESS STATISTICS

SAMPLE TEST 3: ESTIMATION, HYPOTHESIS TESTS, & CORRELATION

NOTE: While the types of questions on the actual test will be similar to these, there may be more problems of each type than you see here.

PART ONE: ESTIMATION

In this class we have discussed z-intervals and t-intervals as ways to estimate averages. Answer these questions:

- _____ 1. Which statistic (**z** or **t**) is most often used when a large sample is available?
- _____ 2. How large does a sample have to be to be considered a “large” sample in statistics?
A. 10 B. 30 C. 100 D. 300 E. 1000

The TI-83 input screens for interval estimation look like the screen illustrations below:

ZInterval Inpt:Data Stats σ : \bar{x} : n: C-Level: Calculate	TInterval Inpt:Data Stats \bar{x} : Sx: n: C-Level: Calculate	1-PropZInt Inpt:Data Stats x: n: C-Level: Calculate
--	---	--

On these input screens ...

- _____ 3. Which variable stands for the **total number** in a sample?
- _____ and _____ 4. Which two variables stand for **standard deviation**?
- _____ 5. Which variable stands for the **mean** (average) of the sample?
- _____ 6. If a problem refers to 95% confidence, what would you enter for “C-Level”?
- _____ 7. In a 1-proportion z-test, which is larger, “x” or “n”?

The results from a TI-83 interval estimate screen are shown below:

```
TInterval  
(1249.1,1294.9)  
 $\bar{x}$ =1272  
Sx=37  
n=9
```

_____ 8. What is the **margin of error** in the t-interval problem shown on the previous page?

Solve these estimation problems.

You will most likely write your answers in the form { Min , Max }.

- _____ 9. The website *usatoday.com* traced the usage habits of 250 of its users. They found that users spent an average of 16.7 minutes each visiting the *usatoday.com* website, with a standard deviation of 7.9 minutes. Use this information to find an interval estimate (**z-interval**) of the population mean for length of visits to the website, with 95% confidence.
- _____ 10. A poll of 27 voters asked them to rate themselves on a scale of 1 to 10, where “1” was “extremely conservative” and “10” was “extremely liberal”. The average rating was 6.1, with a standard deviation of 1.5. Use this information to find an 85% confidence interval (**t-interval**) for the actual conservative/liberal rating of all voters
- _____ 11. ACT tests each question with a selected group of high school juniors before any question actually appears on the nationwide ACT examination. Suppose in a test group of 60 juniors, 39 of them (65%) answered a certain question correctly. Use this information to find an interval estimate (**1-proportion z-interval**) of the overall percentage of high school juniors who should answer this question correctly, with 80% confidence.
- _____ 12. A Gallup survey of 1,250 high school students found that 675 (54%) of them used the same definition of “sex” that President Clinton used in his famous case. Use this information to find an interval estimate (**1-proportion z-interval**) with 99% confidence of the actual percentage of high school students who have this definition of “sex”.
- _____ 13. NewsWatch, Utah-based media lobbying group, sampled 52 network news broadcasts in the year 2000. They found that the “average” network news broadcast included 3.2 visual clips that the group felt were “unacceptable” for family viewing. The standard deviation was 1.4 clips. Use this information to find an interval estimate (**z-interval**) with 90% confidence for the actual average amount of “unacceptable” clips per broadcast.
- _____ 14. Paul is a waiter at an elegant restaurant in New York. He counts his tips on 7 different nights. She finds she gets an average \$213.40 in tips a night, with a standard deviation of 31.73. Use this information to find an 90% confidence interval (**t-interval**) for Paul’s actual nightly tips.

You will use the formulas $n = \left(\frac{z_c \cdot s}{E}\right)^2$ and $n = p \cdot q \left(\frac{z_c}{E}\right)^2$ for the following problems.

- _____ 15. The Nielsen Company estimates the percentage of Americans who watch different TV shows. Suppose a new show comes on TV, and no one has any idea how many people would want to watch it. How many people would Nielsen have to sample to be 99% confident of their results and estimate the percentage of people who would watch the show with a margin of error of 2%?
- _____ 16. According to Nielsen, 68% of all American households own more than one television. If you wanted to estimate the percentage of people in Palo Alto County who own more than one TV, how many people would you have to sample to be 85% confident of your results, with a margin of error of 5%?
- _____ 17. Chad wants to estimate the average points scored in NBA games. If he wants to estimate with 95% confidence and with a margin of error of 1.5 points per game, how many games should he sample? (Assume the standard deviation is 12.5 points.)

PART TWO: Hypothesis Testing

18. List the main steps involved in the classical hypothesis testing method.
19. The goal of a hypothesis test is to see if results are **significant**. What does the word **significant** mean in statistics?
20. Explain in words what it means to say the “level of significance” is 5% (or .05).
- _____ 21. What variable is normally used to stand for **level of significance**?
22. Explain how you can tell, when doing a significance test, whether the result is significant or not.

State the hypotheses you would use for these significance tests.
(You do not actually have to perform the tests.)

23. Each serving of Sugar-Os cereal is supposed to provide 1250mg of riboflavin. In a sample of 12 bowls of cereal, an average of 1107mg was found, with a standard deviation of 71mg. According to this information, does Sugar-Os have significantly less riboflavin than claimed?

$H_1 =$ _____

$H_0 =$ _____

24. The Greater Chicagoland Convention and Visitors Bureau did a survey to find the average income of people who visited the city for various reasons. They found the average family income of people who attended sports events was \$87,900 with a standard deviation of \$12,970 and that the average family income of people who attended the theatre was \$127,400, with a standard deviation of \$28,950. Do these results indicate that people who attend sports events earn significantly less than those who go to the theatre?

$H_1 =$ _____

$H_0 =$ _____

Use the tables in your book to find these critical values.

_____ 25. What is the critical value of "z" when $\alpha = .10$?

_____ 26. What is the critical value of "z" when $\alpha = .01$?

_____ 27. What is the critical value of "t" for 27 degrees of freedom and $\alpha = .05$?

_____ 28. What is the critical value of "t" in a sample where $n = 16$, when $\alpha = .125$?

_____ 29. What is the critical value of "t" for a 2-sample test where one sample has 17, the other sample has 28, and you are using the 5% level of significance?

30. Which kind of test, z or t, is typically used with small samples?

Perform the following hypothesis tests.

31. According to the U.S. Census Bureau, the average American man marries at age 28. The standard deviation for marrying age is 3.3 years. A clerk at the courthouse keeps track of the data for people applying for marriage licenses. After she issues thirty licenses, she finds that the average age of men in those couples is 25.7 years. Is this significantly younger than the national average? (Do a 1-sample **z-test**, at the 10% level of significance.)

Identify these variables: $\bar{x} =$ _____ $\mu =$ _____ $\sigma =$ _____

$n =$ _____ $\alpha =$ _____

$z =$ _____ ... Use your table to find a critical value for "z".

$z =$ _____ ... Compute a test value for "z".

YES — NO (circle one) Is the age of marriage in this county significantly younger than the national average?

32. A study at Indiana University looked at whether toy dolls present girls with an unhealthy body image. According to medical records, the average waist size for adult women in America is 27.5 inches. The study looked at 20 toy dolls and found that if the dolls were blown up to the size of actual women they would have waists that averaged 17.0 inches, with a standard deviation of 2.9 inches. Do a 1-sample **t-test** at the 1% level of significance to see if the dolls really do have significantly smaller waists than real women.

Complete this blank screen:

T-Test	
Inpt: Data	Stats
$\mu_0:$	
$\bar{x}:$	
Sx:	
n:	
$\mu: \neq \mu_0$	< μ_0 $> \mu_0$
Calculate	Draw

$t =$ _____ Use your table to find a critical value for "t".

$t =$ _____ Compute a test value for "t".

YES — NO (circle one) Do Chicagoans travel to significantly fewer states than people nationwide?

33. On *Iron Chef*, different judges rate dishes prepared by celebrity chefs. Each judge can give a total of up to 20 points. Suppose the 3 judges give an average rating of 17.3 to one chef's food, with a standard deviation of 2.6. The same 3 judges give an average rating of 17.7 to a second chef's food, with a standard deviation of 4.3. Did the judges rate the first chef significantly higher than the second? Do a **2-sample t-test** at the 5% level of significant.

Find the value of these variables:

$\bar{x}_1 =$ _____ $s_1 =$ _____ $n_1 =$ _____

$\bar{x}_2 =$ _____ $s_2 =$ _____ $n_2 =$ _____

_____ How many degrees of freedom are there for this problem?

_____ What is the critical value of t ?

_____ On the graphing calculator, there is a line that says "Pooled". Should you select "YES" or "NO" on this line?

_____ What is the test statistic (calculated value)?

_____ YES/NO: Did the judges give significantly different ratings to the two chefs?

34. In a sample of 750 students at a suburban Texas high school, 548 of them (73%) said they drank alcohol "frequently". In a similar survey at a rural high school in Idaho, only 39 out of 72 students (54%) said they drank frequently. Do a **2-proportion z-test** to see if the proportion of students who drink frequently is higher in Texas than in Idaho. Use $\alpha = 5\%$.

_____ What is the critical value of z ?

_____ What is the test statistic (calculated value)?

_____ YES/NO: Is the rate of percentage of high school students who drink frequently significantly higher in suburban Texas than in rural Idaho?

35. *Fortune* magazine reported that 62% of Americans own some form of stocks or other securities. For her class project, Janelle surveyed 87 people, 43 of whom said they owned stocks or other securities. Do a **1-proportion z-test** with $\alpha = .01$ to see if the percentage who owned stocks was less than expected.

Identify these variables:

_____ x _____ n _____ P_0

_____ Use your table to find a critical value of "z"?

_____ Calculate a test value of "z"?

_____ YES or NO: Is the result significant?

_____ What was the percentage of Janelle's sample that owned stocks?

PART THREE: CORRELATION & REGRESSION

36. What does the term **correlation** mean in statistics?

_____ 37. TRUE or FALSE: When there is a correlation between two things, it means the first thing causes the second.

Would each of these things most likely be:

- + ... a positive correlation
- ... a negative correlation
- 0 ... no correlation

_____ 38. how hard a person exercises and their pulse rate immediately following the exercise

_____ 39. a person's income and the number of years until their retirement

_____ 40. the date of the month and the amount of rain that falls on that date

_____ 41. how many children a couple has and the amount of their weekly grocery bill

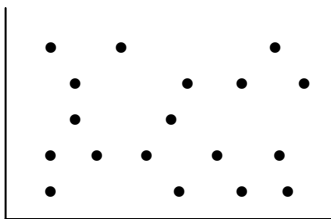
_____ 42. the population of a city and the number of movie theaters in the city

_____ 43. the number of files stored on a computer hard drive and the speed at which the computer operates

For each scatterplot, tell which value of "r" best describes the distribution.

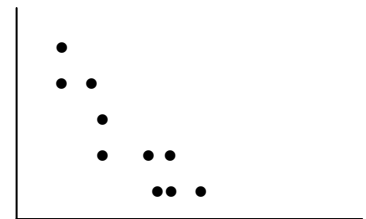
_____ 44.

- A. $r = 0$
- B. $r = 1$
- C. $r = -1$
- D. $r = .5$
- E. $r = -.5$



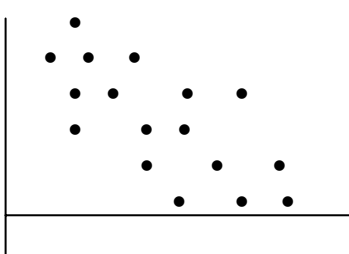
_____ 47.

- A. $r = 3.7$
- B. $r = -.3$
- C. $r = .2$
- D. $r = .8$
- E. $r = -.9$



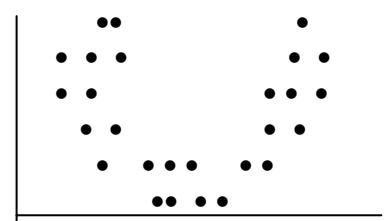
_____ 45.

- A. $r = -.4$
- B. $r = .8$
- C. $r = -.9$
- D. $r = .1$
- E. $r = 0$



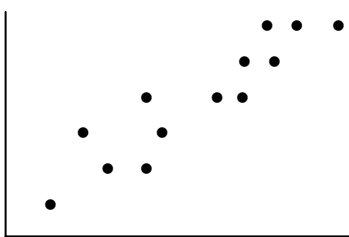
_____ 48.

- A. $r = -.4$
- B. $r = .4$
- C. $r = 0$
- D. $r = -4$
- E. $r = 4$



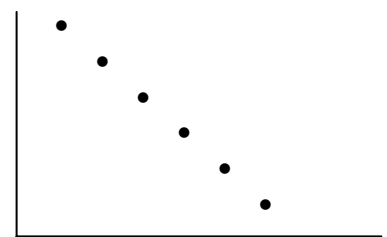
_____ 46.

- A. $r = .1$
- B. $r = .3$
- C. $r = .8$
- D. $r = -.2$
- E. $r = -.7$



_____ 49.

- A. $r = 0$
- B. $r = -1$
- C. $r = .6$
- D. $r = 1$
- E. $r = -3$



You are given the following set of data for the problems below:

The American Psychological Association did a study of divorced couples to find the relationship between length of courtship and length of marriage. Here is some of their data (all times in years):

Courtship	3	0.5	2	1.5	5	1	0.5
Marriage	9	6	14	10	20	6.5	2

- _____ 50. Calculate r for this data.
- _____ 51. Describe this correlation in words (e.g.: “strong negative”—there may be more than one correct answer.)
- _____ 52. How many **degrees of freedom** are there for this problem.
- _____ 53. Use the **r-table** below to find a critical value of “r”. Use $\alpha = .01$

d. f.	$\alpha' = .01$	$\alpha' = .05$	$\alpha' = .10$
<2	1.00	.99	.98
3	.98	.90	.88
4	.93	.81	.77
5	.88	.73	.67
6	.83	.67	.60
7	.79	.62	.55
8	.75	.58	.50
9	.72	.54	.47
10	.69	.52	.45
11	.66	.50	.43
12	.63	.48	.41
13	.61	.46	.39
14	.59	.44	.38
15	.57	.42	.37
16	.56	.41	.36
17	.54	.40	.35
18	.53	.39	.34
19	.52	.38	.33
20	.50	.37	.33
21	.49	.36	.32
22	.48	.35	.32
23	.47	.34	.31
24	.46	.34	.31
25	.45	.33	.30
26	.44	.32	.30

- YES — NO (circle one) 54. Is this a significant correlation?
- _____ 55. Find the **coefficient of determination**, which tells the percent of variation in length of marriage that can be predicted from length of courtship.

56. Suppose that in a correlation 35% of the difference in one variable can be determined from the difference in another variable. Use this information to calculate “r” for this correlation.

Answers

1. z
2. B ... 30
3. n
4. S_x and σ
5. \bar{x}
6. .95 (with a decimal is the best answer, although some calculators will accept the whole number 95)
7. n
8. 22.9 ... 1294.9 – 1272
9. (15.721 , 17.679)
10. (5.6718 , 6.5282)
11. (.57109 , .72891) or 57% to 73%
12. (.50369 , .57631) or 50% to 58%
13. (2.8807 , 3.5193)
14. (190.1, 236.7) or \$190.10 to \$236.70
15. $n = .5 \cdot .5 \left(\frac{2.58}{.02} \right)^2 = 4160.25$ so 4161
(always round up)
16. $n = .68 \cdot .32 \left(\frac{1.44}{.05} \right)^2 = 180.486144$ so 181
17. $n = \left(\frac{1.96 \cdot 12.5}{1.5} \right)^2 = 266.\bar{7}$ so 267
18. Understand problem → Find hypotheses → Choose level of significance → Look up critical value → Calculate test statistic → Compare & interpret
19. unlikely to happen by chance
20. there is a 5% chance you are wrong when you say the result is significant OR there is a 5% chance you say it's significant but it really just happened by chance
21. α (alpha)
22. If calculated value (test statistic) is bigger than table (critical value), then it is significant.
23. H_1 : Sugar-Os have significantly less riboflavin than claimed ... H_0 : Sugar-Os don't have significantly less riboflavin than claimed.

24. H_1 : People who attend sports events earn significantly less than those who go to the theatre. ... H_0 : People who attend sports events don't earn significantly less than those who go to the theatre.
25. 1.282
26. 2.326
27. 1.703
28. 1.197 (Remember n = 16 means 15 degrees of freedom.
29. $df = 17 + 28 - 2 = 43$... so use the row that is closest (45) ... $t = 1.679$
30. t
31. $\bar{x} = 25.7$... $\mu = 28$... $\sigma = 3.3$... $n = 30$... $\alpha = .10$... Critical Value of $z = 1.282$... Test value of $z = -3.81746$... YES—It is significant.
32. Screen shot:

```
T-Test
Inpt:Data Stats
μ0:27.5
x̄:17
Sx:2.9
n:20
μ:≠μ0 <μ0 >μ0
```

- Critical value (19 d.f., .01) = 2.539
Test value = -16.19221639 ... YES—significant
33. $\bar{x}_1 = 17.3$... $s_1 = 2.6$... $n_1 = 3$... $\bar{x}_2 = 17.7$... $s_2 = 4.3$... $n_2 = 3$... $d.f. = 3 + 3 - 2 = 4$... Critical value of “t” = 2.132 ... Pooled? = NO ... Test value of “t” = -.137876 ... NO—not significant
 34. Critical value of “z” = 1.645 ... Test value of “z” = 3.3903 ... YES—significant
 35. $x = 43$... $n = 87$... $P_0 = .62$... Critical value of “z” = 2.326 ... Test value of “z” = -2.4164 ... YES—significant ... $\hat{p} = .49425$ or about 49%

- 36. a relationship between two variables; as one increases, the other changes in a predictable way
- 37. FALSE
- 38. +
- 39. -
- 40. 0
- 41. +
- 42. +
- 43. -
- 44. A (0 because no pattern)
- 45. A (negative, but fairly spread out)
- 46. C (positive, and fairly close together)
- 47. E (negative, and quite close together)
- 48. C (pattern is not a line)
- 49. B (perfect negative correlation)

- 50. appx. .8794 ... **STAT** → EDIT → Put “Courtship” data in L1 and “Marriage” data in L2. QUIT (**2nd** **MODE**). **STAT** → CALC → Choice #4 “LinReg(ax+b)”
- 51. strong positive correlation
- 52. d.f. = 7 - 1 = 6
- 53. table value of “r” = .83
- 54. YES—significant
- 55. $r^2 = .773$, or 77.3%
- 56. We know that $r^2 = .35$, so $r = \sqrt{.35} = .5916$