

STATISTICS—Business Statistics

Sample Test Four

An analysis of all the players on a professional baseball team found that there was a positive correlation between the number of home runs players hit and the number of times they struck out. The regression equation was $\hat{y} = .9\hat{x} + 10.5$, where \hat{x} is the number of home runs a player hit and \hat{y} is the number of times he struck out.

- _____ 1. Early in the season a player hits 2 home runs. According to this formula, approximately how many times can he be expected to have struck out?
- _____ 2. In a given week, a team accumulates 15 home runs. According to the formula, approximately how many strike outs will the team have accumulated in the same week?
- _____ 3. Part way through the season, one player had struck out 25 times. According to the formula, how many home runs would that player have hit?

Admissions counselors at Michigan State University used to use a regression equation to help incoming students predict their college grade point average by using their ACT score. The equation is $\hat{y} = .15\hat{x} - .95$, where \hat{x} is the ACT score and \hat{y} is the estimated college GPA.

- _____ 4. The NCAA requires student athletes to have a minimum ACT score of 17. What grade point can a student with a 17 ACT score expect to get at Michigan State?
- _____ 5. Normally Michigan State accepts only students who are likely to earn at least a 2.0 GPA. What ACT score would predict this GPA?

Answer these questions about statistical tests.

6. What does **significant** mean in statistics?
- _____ 7. What variable is normally used to stand for the level of significance?
8. If you did a hypothesis test at the .05 level of significance, exactly what does that mean?

A machine makes auto parts. Each part is supposed to measure 21.0mm in diameter. The acceptable tolerance is for the standard deviation to be 0.2mm. A sample of 10 parts was tested, and the standard deviation was found to actually be 0.3mm. Is this too large? (Do a **standard deviation χ^2 test** at the .05 level of significance.)

9. What are the hypotheses for this test?

H_1 : _____

H_0 : _____

10. Identify these variables:

$n =$ _____ $s =$ _____ $\sigma =$ _____ $\alpha =$ _____

_____ 11. How many **degrees of freedom** are there in this test?

_____ 12. What is the critical (table) value of χ^2 ?

_____ 13. What is the test (calculated) value of χ^2 ?

YES – NO 14. Is the standard deviation too large?

An automatic teller machine normally dispenses around \$3,700 in cash each day. The bank needs to load the machine with enough cash to accommodate variations in demand. They have found from past experience that the standard deviation for cash dispensed is approximately \$1,100 per day. During the month of April they took data on 30 straight days and found out that the actual standard deviation was \$1,127. Does this indicate that the standard deviation has increased (which would mean they should load the machine with more money)? Do a **standard deviation χ^2 test** with $\alpha = .10$.

_____ 15. How many **degrees of freedom** are there for this test?

_____ 16. What is the critical (table) value of χ^2 ?

_____ 17. What is the test (calculated) value of χ^2 ?

YES – NO 18. Has the standard deviation increased?

A restaurant chain caters primarily to travelers. They have a total of 200 locations in every region of the country. The following table described the percent of travelers who vacation in each region of the country and the number of restaurants the chain has in each region:

	% of all travelers		# of restaurants
Northeast	15%		50
Southeast	35%		75
Midwest	7%		25
Northwest	13%		20
Southwest	30%		30

Do a **categorical χ^2 test** at the .10 level of significance to see whether the distribution of restaurants is significantly different from what would be expected, based on travel to the different regions.

_____ 19. How many **degrees of freedom** are there for this test?

_____ 20. What is the critical (table) value of χ^2 ?

21. In the center column of the table, figure the **expected** number of each restaurants in each region, based on the % of travelers. (The chain has a total of 200 restaurants.)

_____ 22. What is the test (calculated) value of χ^2 ?

YES – NO 23. Is the distribution significantly different than what would be expected?

_____ 24. Suppose you asked 75 people to name their favorite sport. They could choose between 10 sports: baseball, basketball, football, tennis, golf, track, swimming, wrestling, gymnastics, and jai-alai. How many **degrees of freedom** would there be if you did a categorical χ^2 test for this problem?

Census 2000 provides information on the size of households in America. The data for that sample is given below. Suppose for their project someone took a sample of 30 families in Spirit Lake, and the number of each size was recorded.

	Overall – Expected Values		Actual Totals for Sample
1 person	23%	—	7
2 people	28%	—	7
3 people	22%	—	7
4 people	16%	—	6
5 people	7%	—	2
6 or more people (use “6”)	4%	—	1

Total = 30 families

Do a **categorical χ^2 test** at the .05 level of significance to see whether the distribution of family size is different in Spirit Lake than it is in the nation as a whole.

_____ 25. How many **degrees of freedom** are there for this test?

_____ 26. What is the critical (table) value of χ^2 ?

_____ 27. What is the test (calculated) value of χ^2 ?

YES – NO 28. Is the distribution significantly different than what would be expected?

Andrew High School is a small school in eastern Iowa. Here is the male/female distribution for the various classes at Andrew during a recent school year:

	Boys	Girls	
Freshmen	13	8	21
Sophomores	12	14	26
Juniors	11	11	22
Seniors	9	12	21
	45	45	90

- _____ 29. How many **degrees of freedom** are there for this test?
- _____ 30. Use $\alpha = .05$ to find an appropriate table value for a **matrix χ^2 test**.
- _____ 31. Calculate a value of χ^2 for this data.
- YES – NO 32. Is the distribution of boys and girls significantly different among the different classes?

CALCULATORS USED FOR MATH CLASS

	TI-83	TI-85/TI-86	TI-92	Other Graphing	Scientific	4-Function
Math for General Education	2	1	0	3	8	2
Statistics	9	6	1	4	2	0
College Algebra	1	5	1	3	4	1
College Math	2	2	0	4	4	1
Calculus	3	8	1	2	1	0

- _____ 33. How many degrees of freedom would you have if you did a matrix χ^2 test for the above data? (You do not actually need to do the test.)
34. Explain the advantages and disadvantages of using **NON-PARAMETRIC** statistics.
35. Explain how you compare the critical value and test statistic for the **SIGN TEST** and **RUNS TEST**. How are these different than other statistical tests?

Sign Test →

Runs Test →

According to America Online, their average customer receives 7 e-mails per day. Tracey is an AOL subscriber. She keeps track of how many e-mails she receives over the course of two months. The results are given below:

8	13	11	6	17	0	12	8	5	7	24	12	2
6	13	10	6	5	9	47	8	10	3	1	0	15
85	8	11	6	1	5	17	15	32	0	5	7	0
4	12	3	8	9	12	10	6	8	9	12	5	22

Does Tracey receive significantly more than 7 e-mails per day? Answer this question by performing a sign test with $\alpha = .05$.

- _____ 36. How many days did Tracey receive more than 7 e-mails?
- _____ 37. How many days did Tracey receive fewer than 7 e-mails?
- _____ 38. What is "n" for the sign test?
- _____ 39. Use the sign test table to find the critical value for this problem.
- YES – NO 40. Does Tracey receive significantly more than 7 e-mails per day?

According to the September, 2001 state attendance audit, out of 14 classes from pre-kindergarten to twelfth grade at Bishop Garrigan High School and Seton Grade School, 10 of those classes had more boys than girls. (This is an interesting result, since until the mid 1980s almost every Catholic school in America had more girls than boys.) Do a sign test with $\alpha = .05$ to see if there are significantly more boy-heavy classes than girl-heavy classes at Garrigan and Seton today.

- _____ 41. How many of the classes had more girls than boys?
- _____ 42. Use the sign test table to find the critical value for this problem.
- YES – NO 43. Are there significantly more classes that have a majority of boys?

The most lopsided boy/girl ratio at Garrigan and Seton was a class with 31 boys and only 17 girls. Do a sign test with $\alpha = .10$ to see if this class had significantly more boys than girls.

- _____ 44. What is "n" for this problem?
- _____ 45. Use the sign test table to find a critical value for this problem.
- YES – NO 46. Are there significantly more boys than girls in the class?

A group of students were given the Graduate Record Exam (GRE)—which includes verbal and analytical scores. The results follow:

Verbal	630	590	590	350	720	690	750	400	280	530	500	580	375	730
Anal.	700	525	630	410	680	730	770	400	600	660	510	605	400	685

Do a sign test at the 2% level of significance to see if significantly more students scored higher on the analytical section than on the verbal section.

- _____ 47. How many students scored higher on the verbal section?
- _____ 48. How many students scored higher on the analytical section?
- _____ 49. How many students received the same score on both sections?
- _____ 50. What is “n” for this problem?
- _____ 51. Use the sign test table to find a critical value for this problem.
- YES – NO 52. Did significantly more students score higher on the analytical part?

The random number generator on a calculator was asked to find numbers between 1 and 100:

→ → →

94	91	15	51	41	73	4	34	100	20	80	95	12
22	37	1	94	11	1	55	86	98	28	28	5	72

The fact that two different numbers repeat themselves might lead you to suspect that the data is not random. Do a runs test with the characteristic ODD or EVEN to see if this data is random.

- _____ 53. How many odd numbers were there?
- _____ 54. How many even numbers were there
- _____ 55. How many “runs” were there?
- _____ 56. What is the critical range for the runs test?
- _____ 57. Is the data random?
58. Explain what a Spearman’s r test is used to do.

59. We briefly discussed six other topics that are often covered in statistics:

- P-value statistics
- non-linear regression
- multiple regression analysis
- calculus-based statistics
- analysis of variance
- high power tests

Choose two of these, and briefly explain what it involves.

Tell which test would be most appropriate for each of these problems.

- _____ 60. A cable TV network needs to show its advertisers that its ratings are significantly higher than they were last year. They compute the mean and standard deviation for the ratings from the 2001-02 season and the 2000-01 season.
- A. 1-sample t-test C. 2-proportion z-test
B. 2-sample t-test D. standard deviation χ^2 test
- _____ 61. The United Nations has compiled life expectancy data from the entire world. Among other things, they know the mean and standard deviation for age at death among people all over the world. Professor Gøttfreib believes that Danish people live longer than people in other countries do. He wants to compare the average age of death in Denmark with the U.N. statistics.
- A. 1-sample z-test C. 1-sample t-test
B. 1-proportion z-test D. Correlation r-test
- _____ 62. The Insurance Institute of America has data that show that the further a driver lives from work, the more likely he or she is to have an accident. They have tested to show that this is a significant result.
- A. 1-proportion z-test C. 1-sample t-test
B. Correlation r-test D. Standard deviation χ^2 -test
- _____ 63. The American Medical Association did a study on junk food and depression. They had people record how many times they ate junk food per week, and then they gave each subject a test to see how susceptible to depression they were. They found that as consumption of junk food increases, people become slightly less likely to be depressed.
- A. 2-proportion z-test C. Sign test
B. correlation r-test D. Runs test
- _____ 64. The Bureau of Alcohol, Tobacco, and Firearms regulates the gaming industry. They believe one of the keno dealers in a casino is cheating. They check surveillance cameras to see if the cards are being dealt in a random order.
- A. Sign test C. Matrix χ^2 test
B. Standard deviation χ^2 test D. Runs test
- _____ 65. In Algebra class, Mr. Burrow says he will occasionally collect homework. One student suspects he collects the homework significantly more often than he doesn't collect it. The student compares the number of days the homework is collected with the number of days it isn't collected.
- A. Sign test C. Categorical χ^2 test
B. 2-sample t-test D. Correlation r-test
- _____ 66. *People* magazine publishes a list of the most interesting people of the year. Joanne believes there are far more men than women on that list, and she counts to see if one sex is significantly more represented than the other..
- A. Correlation r-test C. Matrix χ^2 test
B. 2-proportion z-test D. Sign test

- _____ 67. The Los Angeles Metropolitan Transit Authority recently released the results of a study on rail ridership in southern California. They found that in 1999 about 1% of all L.A. commuters rode the MTA's subway and light-rail lines. In 2000 (after the subway had been extended to Hollywood and the San Fernando Valley), 8% of all LA commuters had taken the train. Is the percentage for 2000 significantly higher than it was in 1999?
- A. 2-proportion z-test C. categorical χ^2 -test
B. 1-proportion z-test D. sign test
- _____ 68. A union member believes that factory workers in Spencer are paid less than workers elsewhere. He finds data from the U.S. Bureau of Labor Statistics giving the average hourly pay and the standard deviation for all factory workers in America. He then takes a sample of factory workers in Spencer and compares the local average to the national average.
- A. 1-sample t-test C. 1-sample z-test
B. 2-sample t-test D. 1-proportion z-test
- _____ 69. For their Stats project, a group compared the movies shown on HBO and Showtime. They found out how many G, PG, R, and adult movies were shown on each network over the course of a month. Then they organized the data into a table, and they tested to see whether the distribution was different on HBO than it was on Showtime.
- A. Categorical χ^2 test C. Sign test
B. Matrix χ^2 test D. 2-sample t-test
- _____ 70. Iowa City has two public high schools: City High and West High. The administrators at West feel their school has a reputation for having smarter students. To see if this is true, you find the average and the standard deviation for the ITED scores at both City and West High.
- A. 2-sample t-test C. Correlation r-test
B. 2-proportion z-test D. Runs test
- _____ 71. ACT suspects that most of the students at the Hooterville High School test site cheated, because there is almost no variation in the scores at that site. They want to see if the standard deviation in Hooterville is less than it should be.
- A. Standard deviation χ^2 test C. Runs test
B. 1-sample t-test D. Sign test
- _____ 72. The Connecticut Department of Transportation reports that women are more likely to carpool than men. They compared the percentage of women commuters who carpooled with the percentage of men who carpooled.
- A. Sign test C. 1-proportion z-test
B. Categorical χ^2 -test D. 2-proportion z-test

Match the examples below to the methods of deception they illustrate.

- A. placebo effect
- B. NOYB effect
- C. comparing apples and oranges
- D. Non-representative samples

- _____ 73. A coffee manufacturer tried to test-market a new brand of coffee in two cities, using two different marketing strategies. When they looked at the final data, though, they found that even **before** the new marketing campaign, one of the cities had far more coffee drinkers than the other. Because the cities were **dissimilar**, they decided that the results could not be properly compared.
- _____ 74. A telephone poll asks people about the details of the last time they had sexual relations. Many of those called **refuse to answer** and just hang up the phone.
- _____ 75. A company wants to find out about the drinking habits of **all** American adults. They choose to survey a group of college students on Spring Break at South Padre Island about how much they drink.
- _____ 76. At the beginning of the semester a teacher tells her students they are part of a special study and will be learning by a special new method. Then the teacher proceeds to make **no changes** and teaches exactly the same way she always has. Even so, at the end of the semester, the students show significant improvement.
77. Besides the four given in Problems #73 — #76 above, give another example of statistical deception, and briefly tell what it's about.

Statistics--Business STATISTICS

FORMULAS

χ^2 Standard Deviation Test:

- $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$
- $df = n - 1$

χ^2 Multinomial/Categorical Test:

- $\chi^2 = \sum \frac{(O - E)^2}{E}$
- $E = p \sum O$
- $df = K - 1$

χ^2 Matrix/Independence Test:

- $df = (R - 1)(C - 1)$

ANSWERS:

1. $.9 * 2 + 10.5 = 12.3$, so about 12
2. $.9 * 15 + 10.5 = 24$
3. $25 = .9x + 10.5$... solving we get $x = 16.1$, so about 16
4. $.15 * 17 - .95 = 1.6$
5. $.15x - .95 = 2$... solving we get $19.\bar{6}$, so about 20
6. did not happen by chance
7. α
8. There is a 5% chance you will say a result is significant when it actually happened by chance.
9. H_1 : The standard deviation is significantly higher than 0.2mm. H_0 : The standard deviation is not significantly higher than 0.2mm.
10. $n = 10$, $s = 0.3$, $\sigma = 0.2$, $\alpha = .05$
11. $10 - 1 = 9$ d.f.
12. $\chi^2(9, .05) = 16.92$
13. $\chi^2 = \frac{9(.3)^2}{.2^2} = 20.25$
14. YES
15. d.f. = $30 - 1 = 29$
16. $\chi^2(29, .10) = 39.09$
17. $\chi^2 = \frac{29 * 1127^2}{1100^2} = 30.44$
18. NO
19. d.f. = $5 - 1 = 4$ (there are 5 categories)
20. $\chi^2(4, .10) = 7.78$
21. Table:

	EXPECTED
Northeast	$.15 * 200 = \mathbf{30}$
Southeast	$.35 * 200 = \mathbf{70}$
Midwest	$.07 * 200 = \mathbf{14}$
Northwest	$.13 * 200 = \mathbf{26}$
Southwest	$.30 * 200 = \mathbf{60}$

22. $\sum \frac{(O - E)^2}{E} = 13.33 + .36 + 8.64 + 1.38 + 15 = 38.71$ (the same can be calculated through the χ^2 -GOF test on a TI-84)
23. YES
24. $10 - 1 = 9$ d.f.
25. 5 d.f.
26. $\chi^2(5, .05) = 11.07$
27. $\chi^2 = .596$
28. NO
29. $(2 - 1) * (4 - 1) = 1 * 3 = 3$
30. 7.81
31. 1.772893773
32. NO
33. $5 * 4 = 20$
34. <See notes from class>

35. Sign test → # in smaller group should be LESS than critical value to have a significant result; Runs test → if # of runs is in the critical range, the data is random – outside the range is not random.
36. 30
37. 20
38. $30 + 20 = 50$
39. 18
40. NO
41. $14 - 10 = 4$
42. 3
43. NO
44. $17 + 31 = 48$
45. 19
46. YES
47. 3
48. 10
49. 1
50. $3 + 10 = 13$
51. 2
52. NO
53. 12
54. 14
55. 11
56. 8 to 20
57. Yes
58. To compare two different rankings of the same things
59. <See notes>
60. B
61. A
62. B
63. B
64. D
65. A
66. D
67. A
68. C
69. B
70. A
71. A
72. D
73. C
74. B
75. D
76. A
77. <See notes>