

Name _____

- 1) Distinguish between qualitative and quantitative data. Give an example for each.
- 2) Give an example of a random sample. Explain why this is important in design of experiments.
- 3) You plan to make a survey of 200 people. The plan is to talk to every 10th person coming out of the school library. Is there a problem with your plan?
- 4) You plan to make a survey of 200 people. How would you design your sample method to obtain a proper sample? Think of the different types of sampling methods we discussed in class i.e. "stratified sampling", "systematic sampling", "cluster sampling", and "convenience sampling".
- 5) A student surveyed a simple random sample of students at her college. Is this sample likely to be representative of all students at her college? Of all adults in the United States? Explain.
- 6) A researcher wants to obtain a sample of 100 school teachers from the 800 school teachers in a school district. Describe procedures for obtaining a sample of each type: random, systematic, convenience, stratified, cluster.

Form a conclusion about statistical significance. Do not make any formal calculations. Either use the results provided or make subjective judgments about the results.

- 7) In a random sample of 160 women, 78% favored stricter gun control laws. In a random sample of 220 men, 61% favored stricter gun control laws. Is there statistically significant evidence that a larger proportion of women than men favor stricter gun control laws?

Correlations

- 8) The table shows the weights (in pounds) and monthly incomes (in dollars) of nine randomly selected women between the ages of 18 and 65. Assume that the x-values are the weights and the y-values are the monthly incomes.

Weight (lb)	113	132	155	122	166	140	118	129	185
Monthly Income (dollars)	1420	3650	5475	2310	4710	2910	1720	2460	4115

If we use statistical methods to conclude that there is a correlation (or relationship or association) between the weights of women and their monthly incomes, can we conclude that by increasing her weight a woman can increase her monthly income?

- 9) A study shows that adults who work at their desk all day weigh more than those who do not. can we conclude that desk jobs cause people to gain weight. What possible explanation could you give to support this conclusion? What alternative conclusion can we make?

Weight (lb)	113	132	155	122	166	140	118	129	185
Time at desk per day (hrs)	1	7	10	2	4	5	3	6	8

If we use statistical methods to conclude that there

Provide an appropriate response.

- 10) A coach uses a new technique in training middle distance runners. The times, in seconds, for 8 different athletes to run 800 meters before and after this training are shown below.

Athlete	A	B	C	D	E	F	G	H
Before	115.2	114	116.4	119.8	110.9	112.4	111.5	117.3
After	112.9	112.7	114	120.6	109.1	109.1	107.9	113.4

Does the conclusion that the technique is effective appear to be supported with statistical significance? Does the conclusion that the technique is effective appear to have practical significance?

Use the given data to construct a frequency distribution.

- 11) Kevin asked some of his friends how many hours they had worked during the previous week at their after-school jobs. The results are shown below.

5 6 5 4 5 5 9 8 5 3 7 6
6 7 5 6 7 5 6 8 6 7 8 4

Construct a frequency distribution. Use 4 classes, a class width of 2 hours, and a lower limit of 3 for class 1.

- 12) A medical research team studied the ages of patients who had strokes caused by stress. The ages of 34 patients who suffered stress strokes were as follows.

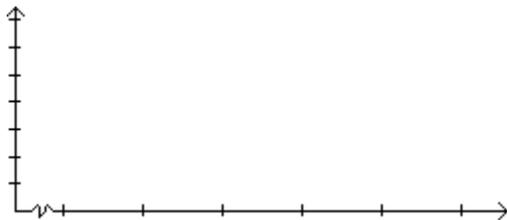
29 30 36 41 45 50 57 61 28 50 36 58
60 38 36 47 40 32 58 46 61 40 55 32
61 56 45 46 62 36 38 40 50 27

Construct a frequency distribution for these ages. Use 8 classes beginning with a lower class limit of 25.

Provide an appropriate response. (use class boundaries like we did in class for horizontal axis 13–14)

- 13) In a survey, 20 people were asked how many magazines they had purchased during the previous year. The results are shown below. Construct a histogram to represent the data. Use 4 classes with a class width of 10, and begin with a lower class limit of -0.5. What is the approximate amount at the center?

6 15 3 36 25 18 12 18 5 30
24 7 0 22 33 24 19 4 12 9



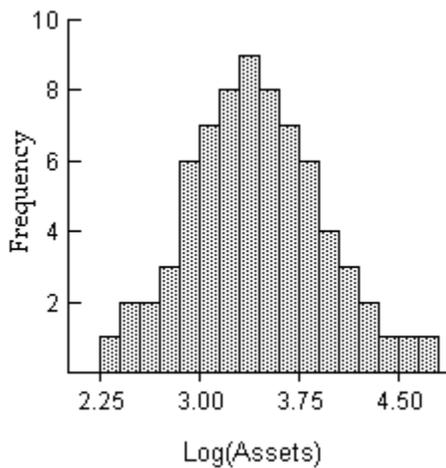
14) The frequency table below shows the number of days off in a given year for 30 police detectives.

Days off	Frequency
0-2	10
3-5	1
6-8	7
9-11	7
12-14	1
15-17	4

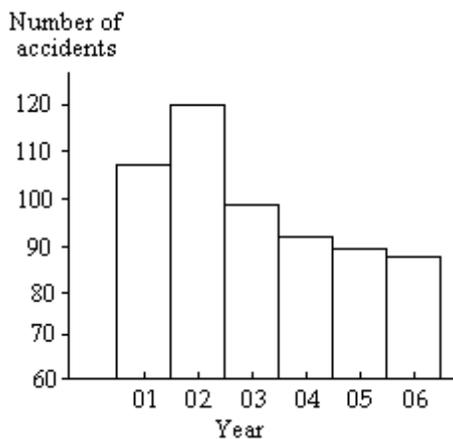
Construct a histogram. Use the class midpoints for the horizontal scale (**Again use class boundaries like we did in class for x axis**). Does the result appear to be a normal distribution? Why or why not?

Provide an appropriate response.

15) The histogram below shows the distribution of the assets (in millions of dollars) of 71 companies. Does the distribution appear to be normal?



16) The graph below shows the number of car accidents occurring in one city in each of the years 2001 through 2006. The number of accidents dropped in 2003 after a new speed limit was imposed. Does the graph distort the data? How would you redesign the graph to be less misleading?



- 17) The two most frequently used measures of central tendency are the mean and the median. Compare these two measures for the following characteristics: Takes every score into account? Affected by extreme scores? Advantages.
- 18) Explain how two data sets could have equal means and modes but still differ greatly. Give an example with two data sets to illustrate.
- 19) A city has 4 different area codes for phone numbers. Does it make sense to find the mean of these area codes? what would be the better measure of center mean, med or mode?

Find the mean for the given sample data.

- 20) Andrew asked seven of his friends how many cousins they had. The results are listed below. Find the mean number of cousins.

15 12 5 14 4 4 6

Find the mode for the given sample data.

- 21) 20 42 46 42 49 42 49

Find the median for the given sample data.

- 22) The temperatures (in degrees Fahrenheit) in 7 different cities on New Year's Day are listed below.

25 25 31 53 64 73 83

Find the median temperature.

Find the mean and median for each of the two samples, then compare the two sets of results.

- 23) A comparison is made between summer electric bills of those who have central air and those who have window units.

	May	June	July	Aug	Sept
Central	\$32	\$64	\$80	\$90	\$65
Window	\$15	\$84	\$99	\$120	\$40

Find the standard deviation for the given sample data. Round your answer to one more decimal place than is present in the original data.

- 24) 18 18 18 9 15 5 10 5 15

- 25) The top nine scores on the organic chemistry midterm are as follows.

47, 55, 71, 41, 82, 57, 25, 66, 81

Find the range, variance, and standard deviation for each of the two samples, then compare the two sets of results.

- 26) When investigating times required for drive-through service, the following results (in seconds) were obtained.

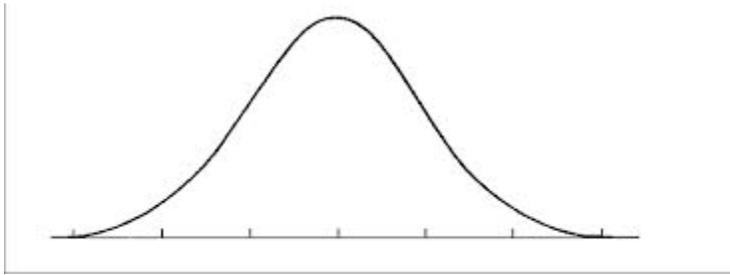
Restaurant A	120	123	153	128	124	118	154	110
Restaurant B	115	126	147	156	118	110	145	137

- 27) When investigating times required for drive-through service, the following results (in seconds) were obtained.

Restaurant A	120	67	89	97	124	68	72	96
Restaurant B	115	126	49	56	98	76	78	95

Empirical rule (26–29)

- 28) Fill out the graph for 1, 2 and 3 standard deviations. 68%, 95% and 99.7% Look at page 106 for LBCC and page 104 for SAC.



Use the empirical rule to solve the problem.

- 29) At one college, GPA's are normally distributed with a mean of 3 and a standard deviation of 0.6. What percentage of students at the college have a GPA between 2.4 and 3.6?
- 30) The systolic blood pressure of 18-year-old women is normally distributed with a mean of 120 mmHg and a standard deviation of 12 mmHg. What percentage of 18-year-old women have a systolic blood pressure between 96 mmHg and 144 mmHg?
- 31) The amount of Jen's monthly phone bill is normally distributed with a mean of \$55 and a standard deviation of \$12. What percentage of her phone bills are between \$19 and \$91?

Find the number of standard deviations from the mean. Round your answer to two decimal places.

- 32) The annual snowfall in a town has a mean of 35 inches and a standard deviation of 11 inches. Last year there were 60 inches of snow. How many standard deviations from the mean is that?
- 33) Mario's weekly poker winnings have a mean of \$353 and a standard deviation of \$67. Last week he won \$185. How many standard deviations from the mean is that?

Solve the problem. Round results to the nearest hundredth.

- 34) The mean of a set of data is 4.11 and its standard deviation is 3.03. Find the z score for a value of 10.86.
- 35) Scores on a test have a mean of 66 and a standard deviation of 9. Michelle has a score of 57. Convert Michelle's score to a z-score.

Find the z-score corresponding to the given value and use the z-score to determine whether the value is unusual. Consider a score to be unusual if its z-score is less than -2.00 or greater than 2.00. Round the z-score to the nearest tenth if necessary.

- 36) A time for the 100 meter sprint of 14.9 seconds at a school where the mean time for the 100 meter sprint is 17.6 seconds and the standard deviation is 2.1 seconds.
- 37) A body temperature of 96.7° F given that human body temperatures have a mean of 98.20° F and a standard deviation of 0.62°.

Answer Key

Testname: STATR1

- 1) Qualitative data can be separated into categories that are distinguished by nonnumeric characteristics. Quantitative data consist of numbers representing counts or measurements. Examples will vary.
- 2) In random sampling, each member of the population has an equal chance of being selected. Random sampling provides us with the best representative sample in which all groups of the population are approximately proportionately represented. Careless sampling can easily result in a biased sample which may be useless.
- 3) People who don't go to the library are excluded.
- 4) Stratified sampling subdivides the population into at least two different subpopulations and then draws a sample from each stratum. Systematic sampling selects a beginning point and then selects every kth element in the population. In cluster sampling, the population is divided into sections, then sections are randomly selected, and then all members of the randomly selected sections are surveyed. Convenience sampling uses readily available results. Examples will vary.
- 5) Yes. Since the sample is a simple random sample drawn from students at her college it is likely to be representative of this group. No. Since students at her college are not representative of all adults in the United States, a sample from this group, however well selected, is unlikely to be representative of all United States adults.
- 6) Answers will vary. One answer is as follows. (1) Random: List the names of the teachers in alphabetical order from 1 through 800. Select 100 teachers by a random number computer program. (2) Systematic: Blindly select from a box one of eight index cards, each of which has a number from 1 to 8 written on it. Sample from the alphabetized list, beginning with that number followed by all its integral multiples until 100 teachers are selected. (3) Convenience: Offer an incentive to the teachers, and select the first 100 volunteers. (4) Stratified: Prepare an alphabetized list of teachers by school (i.e., strata) and randomly select teachers in proportion to school size until 100 teachers are selected. (5) Cluster: Form 8 clusters from 8 consecutive blocks of 100 teachers in the alphabetized list. Blindly draw an index card from the box, and whichever card is drawn, all 100 teachers in that cluster will be the sample. Making clusters from the individual schools might not work, since the school or schools randomly selected might not have 100 teachers in total.
- 7) Yes. In these samples, the proportion of women favoring stricter gun control is substantially higher than the proportion of men favoring stricter gun control. If the true proportions were actually equal, there would be a very small likelihood of seeing such a large difference in the samples.
- 8) No. If a correlation (or relationship or association) is found, this doesn't mean that one variable is the cause of another. Larger weights do not cause higher incomes, but tend to be associated with higher incomes because both weight and income are associated with a third variable, age. Older women tend to be heavier and to have higher incomes than younger women.
- 9) Desk job workers are confined to their chairs for most of their work day. Other jobs require standing or walking around which burns calories. It is probably the lack of exercise that causes higher weights, not the desk job itself. Avoid causality altogether by saying lack of walking and exercise is associated with higher weights.
- 10) Yes. Almost all runners have considerably faster times after the training.
Yes. The differences appear to be substantial.

11)

<u>Hours</u>	<u>Frequency</u>
3-4	3
5-6	13
7-8	7
9-10	1

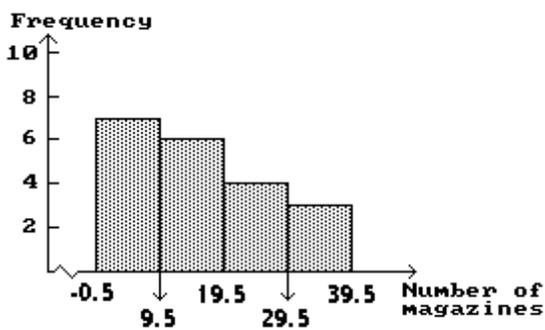
Answer Key

Testname: STATR1

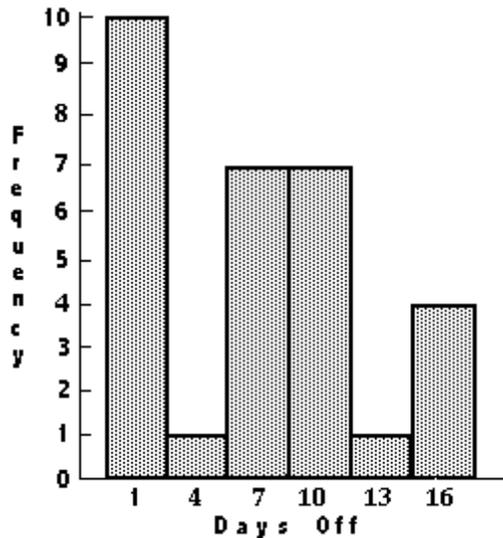
12)

Age	Frequency
25-29	3
30-34	3
35-39	6
40-44	4
45-49	5
50-54	3
55-59	5
60-64	5

13) The approximate amount at the center is 16 magazines.



14) The distribution does not appear to be normal. It is not bell-shaped and it is not symmetric.



15) Yes, it appears to be normal.

16) The graph distorts the data because the the vertical scale starts at 60 rather than 0, giving the impression of a large difference in the number of accidents, when actually the number of accidents only varies from 90 to 120. To make the graph less misleading, change the vertical scale so that it begins at 0 and increases in increments of 20.

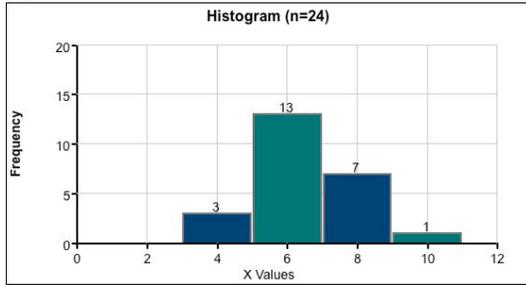
17) The mean takes every score of a data set into account. The median only takes into account the middle score of a ranked odd-numbered data set or the two middle scores of a ranked even-numbered data set. The mean is sensitive to extremes and can be drawn to very low or very high values, but the median is not affected by extremes. Since it uses all the values of a data set, the mean is the preferred average, unless there are extreme values. In the latter case, the median is preferred. An example of the latter is the comparison of salaries of occupations.

Answer Key

Testname: STATR1

- 18) The sets would have different sizes and standard deviations. Examples will vary. A general method for constructing examples is as follows: (1) compose several values and find the sum, (2) prepare another data set with, say, twice that sum and twice that size. Examples are Set A: 5, 10, 15, 2, 5, 8; Set B: 9, 10, 11, 3, 5, 7, 19, 1, 2, 5, 6, 12. Notice that Set A and Set B both have means = 7.5 and modes of 5. Yet, the n for A is 6, while the n for B is 12. Also, the s for A is 4.6, while the s for B is 5.1.
- 19) No. The area codes do not measure or count anything so the mean would be meaningless.
- 20) 8.6 cousins
- 21) 42
- 22) 53°F
- 23) Central air: mean = \$66.20; median = \$65
Window unit: mean = \$71.60; median = \$84
Window units appear to be significantly more expensive.
- 24) 5.4
- 25) 18.9
- 26) Restaurant A: 44 sec; 260.79 sec²; 16.15 sec
Restaurant B: 46 sec; 285.64 sec²; 16.90 sec
There is more variation in the times for restaurant B.
- 27) Restaurant A: 57 sec; 493.98 sec²; 22.23 sec
Restaurant B: 77 sec; 727.98 sec²; 26.98 sec
There is more variation in the times for restaurant B.
- 28) look at page 104 LBCC or page 106 SAC
- 29) 68%
- 30) 95%
- 31) 99.7%
- 32) 2.27 standard deviations above the mean
- 33) 2.51 standard deviations below the mean
- 34) 2.23
- 35) -1
- 36) -1.3; not unusual
- 37) -2.4; unusual

Review 1 additional graphs:
11)



12)

