STATISTICS

Sample Test 4

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

_____2. 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

_____3.	time spent in direct sunlight and severity of sunburn

_____4.	the unemployment rate in a community and spending at stores in that community

_____5.	time spent on a treadmill and calories burned

_____6.	the number of apps on someone's cell phone and the number of Facebook friends that person has

_____7.	the size of lakes in volume and the number of fish those lakes can support

_____8.	people's ages and their flexibility

_____9.	the time since last eating and the amount of hunger people report

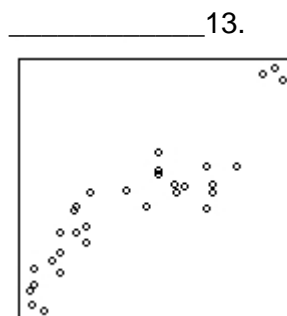_____10.	the number of wind turbines and the percentage of energy produced from wind power

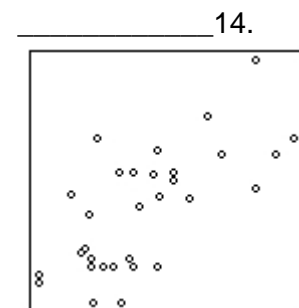_____11.	the day of the month and the gasoline sales at a convenience store

For each scatterplot, tell which value of "r" best describes the distribution.
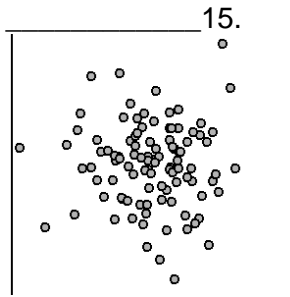
_____12.

A. .4
B. -.4
C. 4
D. -4
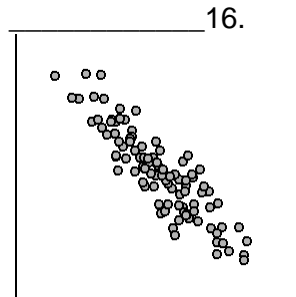
_____13.

A. .1
B. .3
C. .5
D. .7

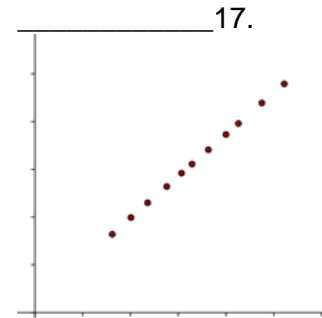_____14.
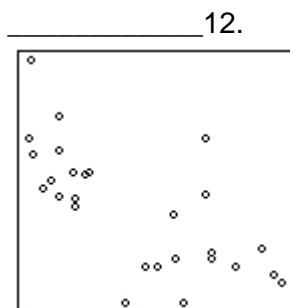
A. 5
B. -5
C. .5
D. -.5

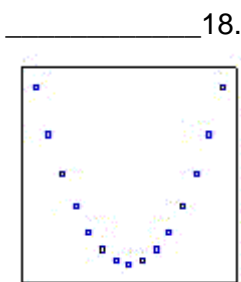For each scatterplot, tell which value of "r" best describes the distribution.

_____15.

A. 0
B. 1
C. .5
D. -.5

_____16.

A. -.2
B. -.4
C. -.6
D. -.8

_____17.

A. 1
B. -1
C. 0
D. 100

_____18.

A. 1
B. .4
C. 0
D. -.3

_____19.

A. -.2
B. -.8
C. .3
D. .9

_____20.
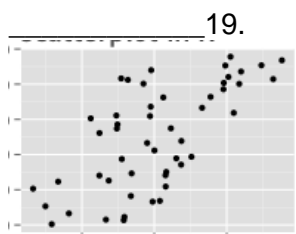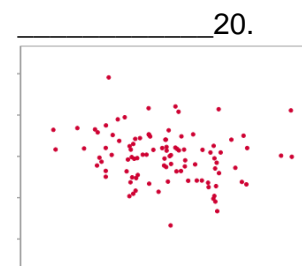
A. .7
B. -.2
C. 4
D. -.9

A sample of people were asked their annual income and the number of years of education they had. The results are shown at right.

You will do an r-test to see if there is a significant correlation between income and years of education. Use α = .05.

Income (thousands of $)	Education (years)
125	19
100	20
40	16
35	16
41	18
29	12
35	14
20	10
24	12
50	16
60	17
36	12
22	14
18	9
40	16

_____21.   What is **n** for this problem?

_____22.   Use the r-table (found on the next page) to find a **critical value** of "r".

_____23.   Calculate a **test statistic** for "r".

_____24.   YES or NO: Is there a significant relationship between income and education?

_____25.   What percentage of the difference in education can be predicted from income?

TABLE 9-6 Critical Values for Correlation Coefficient r

n	α = 0.05	α = 0.01	n	α = 0.05	α = 0.01	n	α = 0.05	α = 0.01
3	1.00	1.00	13	0.53	0.68	23	0.41	0.53
4	0.95	0.99	14	0.53	0.66	24	0.40	0.52
5	0.88	0.96	15	0.51	0.64	25	0.40	0.51
6	0.81	0.92	16	0.50	0.61	26	0.39	0.50
7	0.75	0.87	17	0.48	0.61	27	0.38	0.49
8	0.71	0.83	18	0.47	0.59	28	0.37	0.48
9	0.67	0.80	19	0.46	0.58	29	0.37	0.47
10	0.63	0.76	20	0.44	0.56	30	0.36	0.46
11	0.60	0.73	21	0.43	0.55			
12	0.58	0.71	22	0.42	0.54			

These are critical values for a one-tail r-test using the Pearson Product Moment Correlation Coefficient.

A study looked at the relationship between TV viewing and grade point average for high school students. The results are shown at right.

You will do an r-test to see if there is a significant correlation between TV viewing and grade point average. Use the 1% level of significance.

Hours of TV per week	G.P.A.
14	3.1
10	2.4
20	2.0
7	3.8
25	2.2
9	3.4
15	2.9
13	3.2
4	3.7
21	3.5
9	4.0

_____26. What is **n** for this problem?

_____27. Use the r-table (above) to find a **critical value** of "r".

_____28. Calculate a **test statistic** for "r".

_____29. YES or NO: Is the result significant?

There is a negative correlation between the weight of a car and the fuel economy that car gets. The correlation coefficient for this relationship is approximately r = -.6

_____30. What is the coefficient of determination, the percentage of variation in fuel economy that can be predicted from the weight of a car?

_____31. What amount of the variation in fuel economy is due to other factors besides the weight of the car?

A study found a positive correlation between the length of time college students had been at a party where alcohol was served and the amount of alcohol those students consumed. In the study the **coefficient of determination** was found to be .49.

_____32.   Use the information above to find "r".

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.

_____33.   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?

_____34.   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?

_____35.   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?

Andrew High School was a small school in eastern Iowa that recently closed. At the time they closed, this was the male/female distribution for the various classes at Andrew:

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

Use $\alpha = .05$ to do a **matrix χ^2 test**.

_____ X _____36.   What are the **dimensions** of the matrix you will enter?

_____37.   What is the calculated p-value?

_____38.   YES or NO: Is the distribution of boys and girls significantly different among the different classes?

The University of Michigan School of Medicine conducted a study to compare various treatments for heart attack patients. After each type of treatment they kept track of whether patients either died or suffered another heart attack over a five-year period. The results were:

	Died or suffered a heart attack	Did not die or suffer a heart attack
Directional Atherectomy	44	468
Balloon Angioplasty	23	477

Do a matrix χ^2-test to see if there is a significant difference in the two treatments.

_____39.   What is the calculated p-value for this test?

These questions refer to the heart attack data on the previous page.

_____40.   YES or NO: Is this test significant at the 10% level of significance?

_____41.   YES or NO: Is this test significant at the 5% level of significance?

_____42.   YES or NO: Is this test significant at the 1% level of significance?

_____43.   Given the nature of this experiment, which of these levels of significance would probably
          be most appropriate?
          A. .10 or 10%
          B. .05 or 5%
          C. .01 or 1%

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

Answer the questions on the next page about doing a **categorical (GOF) 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.

44.   If you did a categorical (GOF) test for this problem, tell what the numbers you would put into L2 (the
          expected values) would be

L1	L2
50	
75	
25	
20	
30	

_____45.   How many **degrees of freedom** are there for this
          test?

_____46.   What is the p-value for this problem?

_____47.   YES or NO: Is this a significant result?

```
    X²GOF-Test
Observed:L₁
Expected:L₂
df:█
Calculate Draw
```

_____48. 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?

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
  E. Moving the bullseye to fit the arrows
  F. Biased sources

_____49. 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.

_____50. A company is losing money overall. In their report to investors, they highlight their stores in China, the one country where their stores are profitable.

_____51. 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.

_____52. A study saying that chocolate may lower blood pressure was sponsored by Hershey.

_____53. 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.

_____54. 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.

We briefly discussed several other topics that are often covered in statistics:
  - standard deviation χ^2 test
  - runs test
  - Spearman's r-test
  - high power test
  - non-linear regression
  - multiple regression analysis
  - analysis of variance
  - calculus-based statistics

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

55.

56.

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

_____57.   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 2016-17 season and the 2017-18 season.
    A.   (1-sample) t-test   C.   2-proportion z-test
    B.   2-sample t-test   D.   matrix χ^2 test

_____58.   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 Gφttfreib 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

_____59.   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 as
the distance from work increases, so does the likelihood of accidents.
    A.   1-proportion z-test   C.   (1-sample) t-test
    B.   Correlation r-test   D.   Matrix χ^2 test

_____60.   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.   (1-sample) z-test
    B.   correlation r-test   D.   2-sample t-test

_____61.   The Los Angeles Metropolitan Transit Authority recently released the results of a study
on rail ridership in southern California. They found that in 2016 about 58% of all L.A.
commuters rode the MTA's subway and light-rail lines. In 2017 (after the subway had
been extended to Hollywood and the San Fernando Valley), 10% of all LA commuters
had taken the train. Is the percentage for 2017 significantly higher than it was in 2016?
    A.   2-proportion z-test   C.   categorical (GOF) χ^2 test
    B.   1-proportion z-test   D.   2-sample t-test

_____62.   A union officer 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 43 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

_____63.   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 (GOF) χ^2 test   C.   2-proportion z-test
    B.   Matrix χ^2 test   D.   2-sample t-test

_____64. 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.	Categorical (GOF) χ^2 test

_____65. 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.	2-saple t-test	C.	1-proportion z-test
B.	Categorical (GOF) χ^2 test	D.	2-proportion z-test

66. **SHORT ANSWER**: Consider the project you did for this class. **Briefly** answer these questions.

a.   What question did you attempt to answer?

b.   How did you go about gathering data?

c.   What test(s) did you perform on your data?

d.   Were the results of the project significant?

e.   What problems did you encounter that may have affected your results?