Stat 2610 Final Exam Review Name _______________

Your Stat 2610 instructor heard Minnesota college students learn better when humor is present. For an experiment, he is telling hilarious jokes during lectures in his statistics class and is collecting student achievement data. These data will later be compared to data from other MN college students.

1. Identify the population in this study:

2. Identify the sample in this study:

3. A sample of students is analyzed to make a specific statement concerning that sample from the population. This is an example of:

A) descriptive statistics

B) inferential random sampling analysis

C) inferential statistics

D) simple random sampling analysis

4. Are data at the ordinal level of measurement quantitative or qualitative?

A) Quantitative

B) Both quantitative and qualitative

C) Qualitative

D) Neither quantitative nor qualitative

5. What is the average miles per gallon (mpg) for all new cars? Using *Consumer Reports*, a random sample of 35 new cars gave an average of 21.1 mpg. What is the variable?

A) miles per gallon

B) all cars

C) new cars

D) total miles driven

6. Suppose you want to conduct a survey of benefits packages available in private businesses in Hawaii. You want a sample size of 100. Which type of sample is it if you use postal ZIP Codes to divide the state into regions. Pick a random sample of 10 ZIP Code areas and then include all the businesses in each selected ZIP Code area.

A) cluster sample

B) systematic sample

C) simple random sample

D) convenience sample

The following data represent all the scores for our Stat 2610 class after three quizzes.

13	55	63	75	83	88	93
17	58	68	77	83	88	95
22	60	70	77	83	90	
23	60	72	80	87	92	
47	60	73	80	87	92	
53	63	75	83	87	92	

7. Divide the data into eight classes. What is the class width: _____

8. Make a frequency table for the above data.

Class limits		Frequency	Cum. Freq.	

9. Using the data from #7, construct a histogram.

10. Using the data from #7, construct an Ogive graph.

11. Using the data from #7, construct a Pareto chart.

Let the top row (13, 55, 63, 75, 83, 88, 93) represent a sample of our class data.

12. Calculate the population mean to one decimal place: _____

13. Calculate the sample mean to one decimal ____

14. Which symbol best represents the finding in #12:

A) μ B) σ

C) $\bar{x}$ D) s

15. Which symbol best represents the finding in #13:

A) μ B) σ

C) $\bar{x}$ D) s

16. Calculate the population standard deviation to two decimals. If you need a formula, then try: $\sqrt{\frac{\sqrt{\Sigma(x-\mu)^2}}{n}}$

17. Which symbol best represents the finding in #16:

A) μ

B) σ

C) $\bar{x}$

D s

18. Give the five number summary and construct a Box and Whisker plot from the class data (population) identifying any outliers.

19. What attribute of a data set are being measured by mean, median, and mode?

20. What attribute of a data set are being measured by range, standard deviation, and variance?

21. Given the probability that the temperature tomorrow will be above freezing is 0.38 state the complement AND the probability this complement occurs.

Statement of the complement:

P(Complement) = __________

22. Your professor grades homework by randomly choosing 4 out of 20 homework problems to grade.

a) How many different groups of 4 problems can be chosen from the 20 problems?

b) Shannon did only 4 problems on the assignment. What is the probability that the problems she did comprised the group that was selected to be graded? (Round your answer to four decimal places.)

c) Beth did 8 problems. How many different groups of 4 did she complete?

d) What is the probability that one of the groups of 4 she completed comprised the group selected to be graded? (Round your answer to four decimal places.)

e) Taylor didn't do any of the homework problems, what is the probability she completed the four problems selected?

23. Approximately 73% of the wolves in the New Mexico and Arizona region are male, and 27% female due to efforts cattle ranchers in this area have made to exterminate wolves. Biologists suspect that male wolves are more likely than females to return to an area where the population has been greatly reduced. (Round your answers to three decimal places.) In a random sample of ten wolves spotted in the region, what is the probability that seven or more were male?

What is the probability that fewer than four were female?

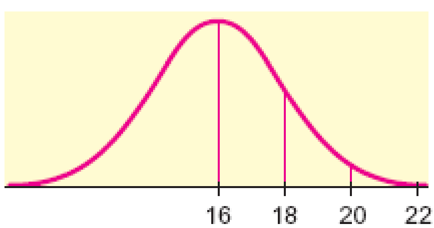
What is the probability that the first male observed occurs on the third wolf?

24. Much of Trail Ridge Road in Rocky Mountain National Park is over 12,000 feet high. Although it is a beautiful drive in summer months, in winter the road is closed because of severe weather conditions. Sustained gale-force winds (over 32 miles per hour and often over 90 miles per hour) occur on the average of 1.7 times every 52 hours at a Trail Ridge Road weather station. For an interval of four days (96 hours), what are the probabilities that $r = 2$, 3, and 4? What is the probability that $r < 2$? (Use 2 decimal places for λ. Use 4 decimal places for your answers.)

P(2) = 	P(3) =

P(4) = 	P(r<2) =

25. Find the mean and standard deviation from the normal curve pictured below.



μ = ________ σ = ________

26. I acquired measurements of every child at Greta's kindergarten class earlier this week when I was there for lunch. From the measurements I calculated the following:

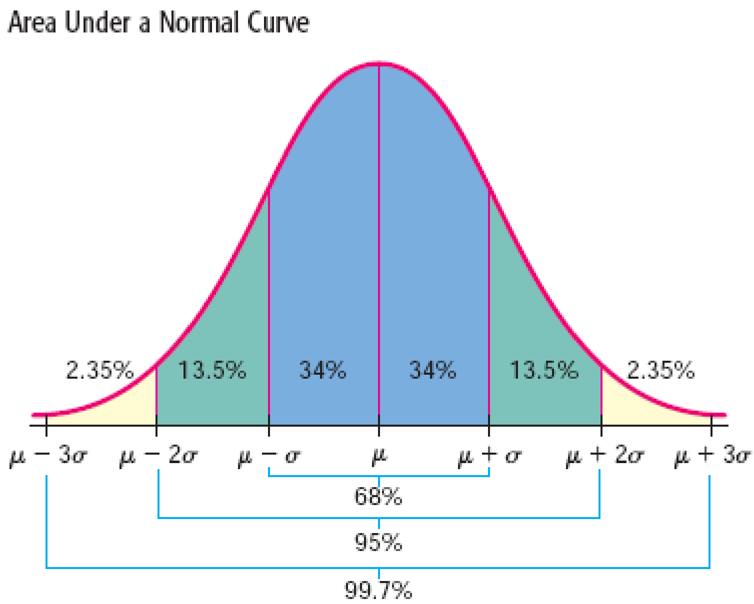
Min = 32"
Max = 47"
$\bar{x}$ = 41.79"
Median = 39"
s = 4.51

At lunch none of the kids could sit still or eat their lunches without spilling. This made me wonder if they're normal. Please determine if their heights are normally distributed using Pearson's index.

Pearson's index: $\frac{3(\bar{x}-Median)}{s}$

Conclusion: Normal or NOT Normal

27. Regardless of your conclusion from #26, assume the heights of Greta's classmates are normally distributed. Sketch their information onto the bell shaped curve.



What percent of his class is taller than 50.81"?
Sketch and shade a bell shaped curve to represent this.

Find the range of values that covers the middle 20% of students in her class. Sketch and shade a bell shaped curve to represent this.

28. If appropriate, use the normal distribution to estimate the binomial probability. After a concerted effort was made to remove lead from the environment only 11% of children in the United States are at risk of high blood-lead levels. In a random sample (our classroom) of 36 students, what is the probability that 12 or more had high blood-lead levels? (Round your answer to three decimal places.)

29. Suppose the heights of 18-year-old women are *normally distributed*, with mean 68.5 inches and standard deviation 1 inch. Calculate the probability that an 18-year-old woman selected at random is between 66.5 and 70.5 inches tall? (Round your answer to three decimal places.)

30. What's your favorite ice cream flavor? For people who buy ice cream, the all-time favorite is still vanilla. About 21% of ice cream sales are vanilla. Chocolate accounts for only 11% of ice cream sales. Suppose that 175 customers go to a grocery store in Cheyenne, Wyoming, today to buy ice cream. A customer who buys ice cream is not limited to one container or one flavor. Fine the probability that someone who is buying ice cream will buy chocolate **OR** vanilla then calculate the probability that between 50 and 60 customers will buy chocolate or vanilla ice cream. Round your answer to three decimal places and assume that the choice to buy one flavor is independent of the choice to buy another flavor.

31. Suppose Agbeko sampled 62 snow geese. The average weight for these birds was x = 6.9 pounds. Based on previous studies, we can assume that the weights of snow geese have a normal distribution, with σ = 1.37 pounds. Find a 90% confidence interval for the average weights of Agbeko's snow geese in the study region. Round your answers to two decimal places.

Calculate the margin of error in the snow goose calculation knowing $E = z_c(\frac{\sigma}{\sqrt{n}})$. Round your answer to two decimal places.

32. According to one source, the average weight for snow geese is 7.4 pounds. Write a null and alternative hypothesis to determine if Agbeko's geese were different than the reported population, then conduct the appropriate statistical test using α = 0.10.

H_0:

H_1:

Find the value of the sample test statistic: ________
(Round your answer to two decimal places)

Find the *P*-value of the test statistic: ________
(Round your answer to four decimal places)

33. Make a conclusion about the weights of Agbeko's geese being different from the population.

As you fondly recall from the homework: $E = z_c\sqrt{(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}$ and $E = t_c\sqrt{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})}$ and confidence intervals are found by using: $(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$.

34. The 2014 Ford Shelby GT 500 claims to produce 662 horsepower (insert man-grunt noise here). Modifications performed by Shelby American, called the Super Snake package, claimed to increase horsepower to 850 hp (more man-grunts and chest pounding). To check and see if the modifications actually increase horsepower as claimed, I sampled six 2014 Shelby GT 500's and recorded each vehicle's horsepower (the results appear below):

Car #	1	2	3	4	5	6
Stock HP	663	662	665	660	662	663

Next, technicians convert each of the six vehicles (all coupes) into the Super Snake and then record the horsepower of each (the results appear below):

Car #	1	2	3	4	5	6
Mod Hp	850	842	853	843	855	849

Let μ_1 be the population mean for x_1 and let μ_2 be the population mean for x_2. Find a 95% confidence interval for $\mu_1 - \mu_2$. (Round your answers to one decimal place.)

35. Does it appear the claimed horsepower increase *more than* the claimed 188 hp with the modifications? Calculate the correct p-value and decision below. (Use α = .05).

H_0:

H_1:

Find the value of the sample test statistic: ________
(Round your answer to two decimal places)

Find the *P*-value of the test statistic: ________
(Round your answer to four decimal places)

36. Make a conclusion about the hp claims of the Shelby American Super Snake package to really increase horsepower by 188 hp.

37. Case studies showed that out of 10,120 convicts who escaped from certain prisons, only 8005 were recaptured. Let p represent the proportion of all escaped convicts who will eventually be recaptured. Find a 99% confidence interval for p. (Round your answers to three decimal places.)

38. The U.S. Geological Survey compiled historical data about Old Faithful Geyser (Yellowstone National Park) from 1870 to 1987. Let x_1 be a random variable that represents the time interval (in minutes) between Old Faithful eruptions for the years 1948 to 1952. Based on 8900 observations, the sample mean interval was $x_1 = 64.9$ minutes. Let x_2 be a random variable that represents the time interval in minutes between Old Faithful eruptions for the years 1983 to 1987. Based on 24,989 observations, the sample mean time interval was $x_2 = 67.1$ minutes. Historical data suggest that $\sigma_1 = 9.19$ minutes and $\sigma_2 = 11.64$ minutes. Let μ_1 be the population mean of x_1 and let μ_2 be the population mean of x_2. Compute a 95% confidence interval for $\mu_1 - \mu_2$. (Use 2 decimal places.)

39. Using the data from #38, conduct the appropriate statistical test to determine if Old Faithful's eruptions are statistically significantly slower, at the $\alpha = 0.05$ level, in the 1980's than 1940's & 1950's.

H_0:

H_1:

Find the value of the sample test statistic: ________
(Round your answer to two decimal places)

Find the *P*-value of the test statistic: ________
(Round your answer to four decimal places)

Conclusion:

40. The table below lists the winning women's times for the Boston Marathon in five year increments since women first ran the course in 1972.

Year	2014	2012	2007	2002	1997	1992	1987	1982	1977	1972	
Winning (min)	138	151	149	140	146	143	145	149	168	190	

Construct a linear equation to predict the winning time in the marathon based on year. Explain what each term in your equation means then provide the r^2 value and its meaning.

Y = ________________

r^2 = ________________

Meanings:

41. A motion picture industry analyst is studying movies based on epic novels. The following data were obtained for ten Hollywood movies made in the past five years. Each movie was based on an epic novel. For these data, $\mathbf{x_1}$ = first-year box office receipts of the movie, $\mathbf{x_2}$ = total production costs of the movie, $\mathbf{x_3}$ = total promotional costs of the movie, and $\mathbf{x_4}$ = total book sales prior to movie release. All units are in millions of dollars.

$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$	$\mathbf{x_4}$	
85.1	8.5	5.1	4.7	
106.3	12.9	5.8	8.8	
50.2	5.2	2.1	15.1	
54.8	3.1	2.9	10.6	
30.3	3.5	1.2	3.5	
79.4	9.2	3.7	9.7	
91	9	7.6	5.9	
89.3	10.2	4.5	7.9	
73.30	7.70	4.11	8.28	Mean
25.52	3.44	2.08	3.70	Std Dev
________	_______	_______	_______	C.V. %

Here are some results from Excel:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.98074053
R Square	0.961852
Adjusted R Square	0.93324099
Standard Error	6.59430898
Observations	8

Calculate the coefficient of variation for each variable in the blanks above.

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 90.0%*	*Upper 90.0%*
Intercept	10.1279886	9.18438888	1.1027395	0.332024	-15.371963	35.6279401	-9.45172134	29.7076985
X Variable 1	4.56876731	1.093872487	4.1766909	0.0139555	1.5316904	7.60584422	2.236798767	6.90073586
X Variable 2	5.4768509	1.86556642	2.9357577	0.0425687	0.29720814	10.6564937	1.49974912	9.45395267
X Variable 3	0.66090076	0.706078414	0.9360161	0.4022566	-1.29948719	2.62128872	-0.84435024	2.16615176

Produce a regression equation predicting $\mathbf{x_1}$ based on an intercept, $\mathbf{x_2}$, $\mathbf{x_3}$, and $\mathbf{x_4}$. Cost = ____

If production costs and box sales were held fixed but promotional costs were increased by one million dollars, what would you expect for the corresponding change in box office receipts?

____

Find a 90% confidence interval for each variable: ________ ________ ________ ________

42. The following table shows age distribution and location of a random sample of 166 buffalo in Yellowstone National Park (based on information from *The Bison of Yellowstone National Park*, National Park Service Scientific Monograph Series).

Age	Lamar District	Nez Perce District	Firehole District	**Raw Total**
Calf	13	13	15	**41**
Yearling	10	11	12	**33**
Adult	34	28	30	**92**
Column Total	**57**	**52**	**57**	**166**

Use a Chi-square test to determine if age distribution and location are independent at the 0.05 level of significance.

H_0:

H_a:

Test statistic: _________

p-value: __________

Conclusion:

43. Consider the following sample: {-16, -8, 2, -2, 6, 4, 16, -4}. What is the third quartile?
    a. 3
    b. 4
    c. 5
    d. 6
    e. 7
44. The time it takes for you to complete this exam is an example of qualitative data?
    a. True
    b. False
45. Consider the following sample: {-16, -8, 2, -2, 6, 4, 16, -4}. What is the sample standard deviation?
    a. 9.65
    b. 10.20
    c. 9.19
    d. -0.25
    e. 9.78
46. Consider the following sample: {-16, -8, 2, -2, 6, 4, 16, -4}. What property best describes this sample?
    a. Skewed right
    b. Skewed left
    c. Symmetric
    d. Bimodal
    e. Trimodal

47. Which measure of central location can be strongly affected by extreme values?
    a. Mean
    b. Median
    c. Mode
    d. Interquartile range
    e. None of the above
48. For a data set, 70% of the observations are above the 30^{th} percentile.
    a. True
    b. False
49. Histograms are useful for displaying qualitative data.
    a. True
    b. False
50. The mode of the following data set is 23.

Class	Frequency
Toyota	20
Honda	23
Ford	12
Dodge	18

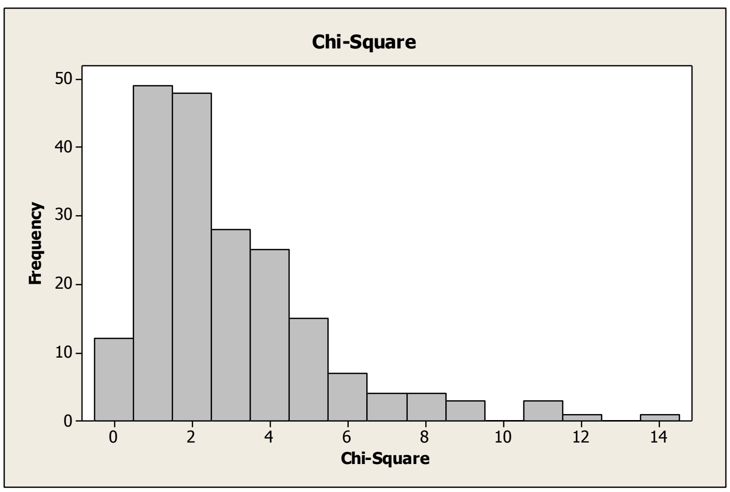
    a. True
    b. False

51. The following are all measures of variability: standard deviation, interquartile range, proportion, range.
    a. True
    b. False

52. What is the margin of error of the following confidence interval: (12.2, 13.8)?
    a. 1.6
    b. 0.8
    c. 13.8
    d. 13
    e. 12.2

53. Ten students in STAT 2610 need to randomly select two students from themselves to form a committee on self-initiative. Of the ten students, three are male and seven are female. What is the probability the committee consists of two males?
    a. 0.0238
    b. 0.0667
    c. 0.4900
    d. 0.4667
    e. 0.4200

54. It is possible to have the mean, median, and mode all be equal for a data set.
    a. True
    b. False
55. A bimodal distribution is always symmetric.
    a. True
    b. False
56. What is the standard deviation of the following data set: 4.32, 2.46, 1.01, 12.73, 6.66, 5.95, 2.73, 3.10?
    a. 3.68
    b. 4.87
    c. 3.71
    d. 6.48
    e. None of the above
57. The following graph is a histogram
    a. True
    b. False



58. A certain county health department has received 25 applicants for an opening that exists for a public health nurse. Of these applicants 10 are over 30 and 15 are under 30. Seventeen hold bachelor's degrees only, and eight have master's degrees. Of those under 30, six have master's degrees. If a selection from among these 25 applicants is made at random, what is the probability that a person over 30 or a person with a master's degree will be selected?
    a. 0.64
    b. 0.40
    c. 0.32
    d. 0.56
    e. None of the above
59. In a survey of college-age students, 42% reported binge drinking at least once per week (binge drinking is defined as having 5 or more drinks within a one-hour time period). If this percentage holds for the entire population, find, for a random sample of 22 college-age students, the probability that the random sample contains 10 to 12 students inclusive who binge drink at least once per week.
    a. 0.0839
    b. 0.1264
    c. 0.1601
    d. 0.3704
    e. None of the above

60. The mean amount spent by a family of four on sugar enhanced beverages (soda, punch, Cool-Aid, etc…) is $120 per month with a standard deviation of $12 per month. Assuming that the food costs are normally distributed, what is the probability that a family spends not more than $100 per month on sugar enhanced beverages?
    a. 0.04779
    b. 0.45221
    c. 0.95221
    d. 0.54779
    e. None of the above

61. The Burns Company produces inanimate carbon rods that the company claims have a mean length of 13 inches. Please test the company's claim by analyzing the lengths of a random sample of 10 inanimate carbon rods. Test the claim by using $\alpha = 0.05$ and choose the correct conclusion of the hypothesis test. The data follows: 11.1, 13.5, 11.7, 13.4, 11.5, 15.2, 12.0, 10.2, 13.5, 11.1. Assume the data was sampled from a normal distribution. What is the correct statement of your conclusion?
    a. The data supports the company's claim that the mean length is 13 inches
    b. The data does not support the company's claim that the mean length is 13 inches
    c. None of the above

62. Bottles of a popular cola are supposed to contain 300 milliliters of cola. There is some variation from bottle to bottle because the filling machinery is not perfectly precise. Conduct a one-sample *t* test to determine if there is statistical evidence to conclude the mean amount of cola in all bottles filled by this company is less than the advertised 300 milliliters based on the following sample of 6 observations: 299.4, 297.7, 301.0, 298.9, 300.2, 297.0. Use a significance level of $\alpha = 0.05$. Assume the population that was sampled is normally distributed.
    a. There is statistical evidence to conclude the mean is greater than 300 ml
    b. There is statistical evidence to conclude the mean is less than 300 ml
    c. Based on this data, the null hypothesis cannot be rejected
    d. Based on this data, the null hypothesis should be rejected
    e. None of the above

63. Which best describes some of the properties of the normal distribution?
    a. Symmetric with no mode
    b. Continues out forever in both the positive and negative directions
    c. Skewed left with one mode
    d. Symmetric with two modes

None of the above

Binomial	P(r) = ${}_nC_r p^r q^{n-r}$ $\mu = np$	$\sigma = \sqrt{npq}$
Geometric	P(n) = p q^{n-1} $\mu = \frac{1}{p}$	$\sigma = \frac{\sqrt{q}}{p}$
Poisson	P(r) = $\frac{e^{-\gamma}\gamma^r}{r!}$ $\mu = \gamma$	$\sigma = \sqrt{\gamma}$

TI-83 notes:

Binomialpdf(n, p, r)

Geometpdf(p, n)

Normalcdf(l, u, μ, σ)

Poissonpdf(μ, r)

Normalpdf(x, μ, σ)

invNorm(%, μ, σ)

Critical Values for Student's *t* Distribution

Significance level = α

Degrees of Freedom	.005 (1-tail) .01 (2-tails)	.01 (1-tail) .02 (2-tails)	.025 (1-tail) .05 (2-tails)	.05 (1-tail) .10 (2-tails)	.10 (1-tail) .20 (2-tails)	.25 (1-tail) .50 (2-tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	.765
4	4.604	3.747	2.776	2.132	1.533	.741
5	4.032	3.365	2.571	2.015	1.476	.727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711
8	3.355	2.896	2.306	1.860	1.397	.706
9	3.250	2.821	2.262	1.833	1.383	.703
10	3.169	2.764	2.228	1.812	1.372	.700
11	3.106	2.718	2.201	1.796	1.363	.697
12	3.054	2.681	2.179	1.782	1.356	.696
13	3.012	2.650	2.160	1.771	1.350	.694
14	2.977	2.625	2.145	1.761	1.345	.692
15	2.947	2.602	2.132	1.753	1.341	.691
16	2.921	2.584	2.120	1.746	1.337	.690
17	2.898	2.567	2.110	1.740	1.333	.689
18	2.878	2.552	2.101	1.734	1.330	.688
19	2.861	2.540	2.093	1.729	1.328	.688
20	2.845	2.528	2.086	1.725	1.325	.687
21	2.831	2.518	2.080	1.721	1.323	.686
22	2.819	2.508	2.074	1.717	1.321	.686
23	2.807	2.500	2.069	1.714	1.320	.685
24	2.797	2.492	2.064	1.711	1.318	.685
25	2.878	2.485	2.060	1.708	1.316	.684
26	2.779	2.479	2.056	1.706	1.315	.684
27	2.771	2.473	2.052	1.703	1.314	.684
28	2.763	2.467	2.048	1.701	1.313	.683
29	2.756	2.462	2.045	1.699	1.311	.683
Large	2.575	2.327	1.960	1.645	1.282	.675