STAT 201: Cumulative Practice Questions for the Final Exam

1. Review the following terms: sample, statistic, population, parameter, categorical, quantitative, discrete, continuous, descriptive, inferential

2. The number of points scored by the USC Gamecock football team for the 11 games in the 2001 season are displayed below.

32 14 16 37 42 7 46 10 38 17 20

For these data, $\sum x = 279$ and $\sum x^2 = 9007$.

a) Find the sample mean number of points scored.
b) Find the sample median number of points scored.
c) Make a boxplot of these data

3. Test scores for a history class a normally distributed with a mean of 86 and a standard deviation of 6. Find the z-score for Sam, who scored 60 on the test. Is Sam's score a potential outlier?

4. Review shapes of histograms.
Review the empirical rule.

5. A group of 300 people were surveyed. Their marital status was recorded along with answers to several questions. One question asked was whether "friends and social life" or "job or primary activity" contributes most to their general well-being. The results from this question appear in the table below.

	Single (never married)	Married	Widowed or divorced
Friends and social life	47	59	56
Job or primary activity	33	61	44

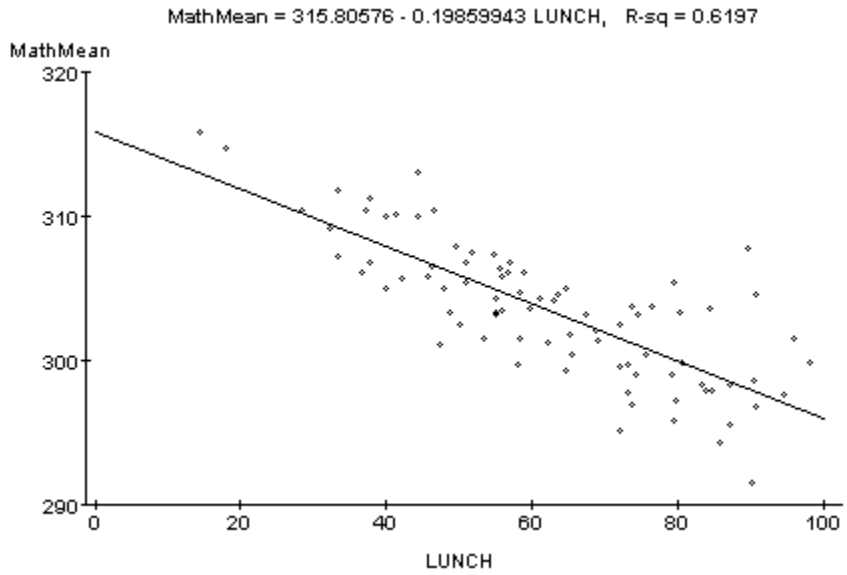
a) Find the probability that a randomly selected person chose "Friends and social life".
b) Find the probability that a randomly selected person is married and chose "friends and social life".
c) Find the probability that a randomly selected person chose "friends and social life" given the person is married.
d) Are the events chose "Friends and social life" and being married independent? Explain using probabilities.
e) Are the events chose "Friends and social life" and being married mutually exclusive? Explain using probabilities.

6. A pediatrician records the number of ear infections that his patients have during the winter. A distribution of probabilities for number of ear infections is displayed below.

Number of ear infections	0	1	2	3	4	5	6
Probability	0.05	0.16	0.25	0.33	0.15	0.04	0.02

a) What is the probability that a patient has more than three ear infections during the winter?
b) What is the mean of the probability distribution?

7. The Palmetto Achievement Challenge Test (PACT) is administered each year to students in South Carolina schools. Data were collected for the year 2000 test of third grade students for each school district in South Carolina. A regression analysis was conducted using the percent of students on free or reduced lunch (LUNCH) to predict the mean score on the math portion (MathMean). Output from a regression analysis in DoStat appears below.



a) Describe the relationship between the two variables based on the graph.
b) What is the value of r, the correlation coefficient?
c) Circle two outliers on the scatterplot. Label them "outlier 1" and "outlier 2".
d) What is the expected mean score in math for a district with 0% of its students on free or reduced lunch?
e) Interpret the slope in the context of the problem.
f) Suppose Lexington District 1 has 28% of its students on free or reduced lunch in 2001. Use the regression equation above to predict the mean score in math for Lexington District 1.

8. Experience has shown that 30% of rocket launchings at a NASA base have had to be delayed due to weather conditions. Consider ten rocket launchings at this NASA base. Let X = number of launchings among ten that are delayed due to weather conditions.

   a) X has a binomial distribution with n = _____ and p = _____ .
   b) Find the probability that at most three of ten launchings will have to be delayed due to weather conditions.
   c) Find the mean number of rocket launchings that will be delayed among ten.

9. a) If Z is a standard normal random variable, find $P(Z > 2.11)$
   b) If Z is a standard normal random variable, find $P(-1.15<Z<1.15)$

10. Gasoline consumption for all families in Rocktown has a normal distribution with a mean of 16.9 gallons per week with a standard deviation of 3.2 gallons per week.

    a) Find the probability that a randomly selected family consumes between 16 and 17 gallons of gasoline per week.
    b) Find the 80^{th} percentile for gasoline consumption
    c) Find the probability that a random sample of 6 families exceeds a mean of 17.5 gallons of gasoline per week.
    d) Would a sample mean of 17.5 be considered 'unusual'?

11. A movie theater is interested in the proportion of Columbia Pictures films that are R-rated. A random sample of 60 films are considered, of which 35 are R-rated.

    a) Compute a 95% confidence interval for the proportion of all Columbia pictures that are R-rated.
    b) Interpret the 95% confidence interval from a) in the context of the problem.
    c) If the level of confidence is changed to 90%, how will the width of the interval change?

12. A hotel manager is interested in the mean length of stay for his guests. A random sample of 16 guests produced a mean length of stay of 4.3 nights with a standard deviation of 1.2 nights. Assume the distribution of lengths of stay for guests follows a normal distribution.

    a) Compute a 95% confidence interval for the mean length of stay for all guests of this hotel.
    b) Interpret the interval from part a) in the context of the problem.

13. A report states that 35% of mothers of school age children are stay at home. A random sample of 200 mothers of school age children show that 75 are stay at home moms. Is there evidence at the .05 level of significance, that more than 35% of mothers of school age children are stay at home moms? Show all steps of the hypothesis test.

14. Suppose that in reality, more than 35% of mothers are stay at home moms. What kind of error, if any, occurred in problem 13? What is the probability of a type I error in problem 13? What would be the consequence if we lowered the probability of a type I error?

15. An obstetrician is studying twin pregnancies. He is interested in whether there is a difference in mean birth weight between the first born twin (twin A) and the second born twin (twin B). Data are available from 19 randomly selected sets of twins. A hypothesis test will be conducted using a 0.10 level of significance.

a) Do the data obtained from the twins represent dependent or independent samples?

Hypothesis test results:
$\mu A - \mu B$: mean of the paired difference between TwinA and TwinB
$H_0 : \mu_A - \mu_B = 0$
$H_A : \mu_A - \mu_B \neq 0$

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
TwinA - TwinB	-0.18092105	0.16087063	18	-1.124637	0.2755

b) Using the above output, give the five steps of the hypothesis test. Assume the weights are normally distributed.

16. A curriculum specialist is interested in whether instruction method A or instruction method B results in a greater mean score on a standardized test. A random sample of 23 students are taught using method A and a random sample of 21 students are taught using method B. Summary statistics and the inference is displayed below.

Summary statistics

Column	n	Mean	Variance	Std. Dev.	Std. Err.	Median	Range	Min	Max	Q1	Q3
A	23	41.52174	294.07904	17.148733	3.575758	42	75	10	85	28	54
B	21	51.47619	121.1619	11.007357	2.402002	53	47	24	71	44	58

99% Confidence interval results:

Difference	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
$\mu_B - \mu_A$	9.954452	4.307628	37.8554	-1.7282724	21.637175

a) Is this a dependent samples or an independent samples design?

b) What is the 99% confidence interval?

c) Interpret the confidence interval.