This examination has twenty problems. All are modifications of the recommended homework problems. Each problem will be graded on a three-point scale, with 3=fully correct, 2=mostly correct, 1=mostly incorrect, and 0=fully incorrect. Please draw a box or oval around each final answer so I can locate it easily. Of course, for me to assign partial credit, I will also need to see your steps. This is especially important when you use the statistical capabilities of your calculator. For example, if you use the normcdf function to solve a problem, one step should be to write down the function and its arguments exactly as they appear on your calculator screen (except you don’t need to wrap the text the way it appears on the small calculator screen).

1. Give one example each of the main four kinds of variables:

Categorical and Nominal: __________________________________________________________

Categorical and Ordinal: _________________________________________________________

Numerical and Discrete: _________________________________________________________

Numerical and Continuous: _____________________________________________________

2. Following are the average annual temperatures for the NOAA Climate Division containing St. Louis, from 1997 to 2006. Calculate their mean, median, SD, quartiles, and IQR. (You may use the “quick approximate” method described in the book for $Q_1$ and $Q_3$ if you wish.)

53.6  57.0  56.3  54.9  55.7  55.2  54.5  54.9  56.1  56.8

[Disclaimer: This is not intended as evidence for or against global warming.]
3. Use the grid below to do a normal plot of the following joint impairment scores:

8  8  22  9  18  3

Put the data values on the horizontal axis and the normal scores on the vertical axis, the way it is done in the book (and the way you made the plots in your SAS and STATA homework).

4. The correlation coefficient between the midterm scores and the final scores in a course is 0.80. A student scored 1.5 standard deviations below the mean on the midterm. If the final mean and SD are 80 and 16, what is the student’s predicted score on the final?
5. A computer network manager wants to model how access time in milliseconds \((y)\) for data files varies with the number of simultaneous users \((x)\) accessing the files. Based on 45 paired observations, the following summary statistics are obtained:

\[
\bar{x} = 8, \quad s_x = 2, \quad \bar{y} = 14, \quad s_y = 3, \quad r = 0.7
\]

Give the equation of the least-squares straight line to estimate the access time as a function of the number of simultaneous users. Having obtained the equation, compute the estimated access time if there are ten simultaneous users.

6. A soft drink company uses a filling machine to fill cans. Each 12 oz. can is to contain 355 milliliters of beverage. In fact, the amount varies according to a normal distribution with mean \(\mu = 355.5\) ml and standard deviation \(\sigma = 0.8\) ml. What is the probability that the mean content of a six-pack of cans is less than 355 ml?
7. To estimate the average number of years an employee stays with a company, a random sample of 15 employees is drawn from the company's past employee records and the sample mean $\bar{X}$ is calculated. Suppose that the true but unknown distribution of the length of stay of an employee is exponential with a mean of 4 years. What is the exact distribution of $\bar{X}$? What is its mean and standard deviation?

8. Let $S^2$ denote the sample variance computed from a random sample of size $n$ from an $N(\mu, \sigma^2)$ distribution. For sample sizes $n = 5, 10, \text{ and } 20$, find the probability that the sample variance $S^2$ exceeds the true variance $\sigma^2$ by a factor of two, i.e., $P(S^2 > 2\sigma^2)$. (Note: In the corresponding textbook exercise, this could only be done approximately, and then only for certain sample sizes. With the TI-83 it can be done exactly, and for all sample sizes.)
9. An engineer suspects that the temperature inside an oven is not as uniform as when it was new, at which time the temperature varied ±8°F around its setting. (Taking the range of a normal distribution to be roughly ±2σ, this translates into σ = 4°F.) To verify his suspicion, he takes 15 measurements in different parts of the oven. He wants a rule which decides that the true σ > 4 if the sample standard deviation exceeds 4c, where c is a suitably chosen constant. The rule must not have more than a 5% chance of making a wrong decision, i.e., deciding that σ > 4 when, in fact, σ = 4. Find the value of c. Based on this value of c, does the rule decide that σ > 4 if the sample standard deviation of the engineer's measurements is σ = 10°F?

10. Consider the Student's t-distribution with 7 degrees of freedom. Find the constant d so that \( P(|T| < d) = 0.95 \). Express the constant d in terms of the notation \( t_{0.025} \).
11. Consider the $F$-distribution with 4 and 6 degrees of freedom. Find the constant $c$ so that $P(F_{4,6} < c) = 0.95$. Express the constant $c$ in terms of the notation $F_{v_1,v_2,a}$.

12. Let $X_1, X_2, X_3, X_4$ be i.i.d. observations from a distribution with mean $\mu$ and variance $\sigma^2$. Calculate the variance of the estimator $0.2X_1 + 0.2X_2 + 0.2X_3 + 0.4X_4$. 
13. A random sample of 9 coffee cans is taken from a production line and the contents are weighed. The weights (in oz.) are as follows:

\[
26.3 \quad 25.9 \quad 26.9 \quad 26.7 \quad 24.8 \quad 25.3 \quad 26.4 \quad 25.7 \quad 26.3
\]

Calculate the sample mean as an estimate of the process mean, and calculate the standard error of the mean (SEM).

14. A random sample of size 17 from an \( N(\mu, \sigma^2) \) distribution has mean \( \bar{x} = 21.3 \). Calculate CI's for \( \mu \) for three levels of confidence: 50%, 95%, and 99%.
15. Let $p$ be the proportion of people in the population who prefer Gatorade over All Sport. We want to decide if more people prefer Gatorade. The hypotheses are set up as $H_0: p \leq \frac{1}{2}$ vs. $H_1: p > \frac{1}{2}$. Thirteen tasters participated in the taste-testing experiment. Suppose that 10 of the 13 tasters preferred Gatorade over All Sport. What is the $P$-value? Can you reject $H_0$ at $\alpha = 0.05$? Notes: Do not use the TI-83 1-PropZTest, as the sample size is too small for the normal approximation to the binomial to be accurate. You can treat the null hypothesis $H_0: p \leq \frac{1}{2}$ the same as if it were stated as $H_0: p = \frac{1}{2}$. The authors use the two forms interchangeably in one-sided tests. (See the next problem for instance.)

16. Consider testing $H_0: \mu = 0$ vs. $H_1: \mu > 0$ at the 5% level of significance based on a sample of size 16 from a normal distribution with unknown mean $\mu$ and known $\sigma = 2$. What is the $\beta$-risk of this rule if $\mu = 1$?
17. An EPA researcher wants to design a study to estimate the mean lead level of fish in a lake located near an industrial area. Based on past sample data, the researcher estimates that \( \sigma \) for the lead level in the fish population is approximately 0.019 mg/g. She wants to use a 99% CI having a margin of error no greater than 0.005 mg/g. How many fish does she need to catch?

18. A tire company has developed a new tread design. To determine if the newly designed tire has a mean life of 60,000 miles or more, a random sample of 10 prototype tires is tested. The mean tire life for this sample is 60,892 miles. Assume that the tire life is normally distributed with standard deviation \( \sigma = 1800 \) miles. Test the hypothesis \( H_0: \mu = 60,000 \) vs. \( H_1: \mu > 60,000 \). Compute the test statistic and its P-value. Based on the P-value, state whether \( H_0 \) can be rejected at \( \alpha = 0.05 \). Suppose that at least 80% power is needed to identify a tread design that has a mean life of 61,000 miles in the one-sided test described previously. How many tires should be tested?
19. A random sample of size 13 is drawn from a normal distribution with \( \mu = 70 \) and \( \sigma = 3 \). The mean of the sample is 68.7 and \( s = 2.9 \). (a) Calculate a 95\% z-interval for \( \mu \) assuming that you know \( \sigma = 3 \). (b) Calculate a 95\% t-interval for \( \mu \) assuming that you do not know \( \sigma \).

20. A thermostat used in an electrical device is to be checked for the accuracy of its design setting of 200°F. Eight thermostats were tested to determine their actual settings, resulting in the following data:

\[
202.2 \quad 203.4 \quad 200.5 \quad 202.5 \quad 206.3 \quad 198.0 \quad 208.7 \quad 200.8
\]

Perform a two-sided \( t \)-test to determine if the mean setting is different from 200°F. Use \( \alpha = 0.05 \).