(1) Why is the mean an inappropriate measure of central tendency in a very skewed distribution?

(2) a. What does $\sum_{i=1}^{N}\left(X_i - \overline{X}\right)$ mean?

b. What value does the above *always* equal?

(3) A researcher collected several sets of data. For each indicate which measure of central tendency would be best.

      a. The following personality scores
          0, 2, 3, 3, 8, 4, 9, 6, 7, 5, 6
      b. The following age scores
          10,15,18,15,14,13,42,15,12,14,42
      c. The following college years
          senior, junior, junior, freshman, freshman, junior, sophomore, junior
      d. The following the political affiliations
          Dem, Dem, Rep, Dem, Soc, Com, Com, Soc, Dem, Rep

(4) You have subtracted the mean from each score in a sample that is approximately normally distributed. Which of the following deviation from the mean score is likely to have the highest frequency?

    a.   −12
    b.   −7
    c.   +2
    d.   +18

(5) a. Distinguish among the symbols σ and s.
   b. When are the above symbols accompanied by the square sign (^{2})?

(6) Suppose you want to estimate population parameters from a sample of 30 scores. You decide to use the sample mean of 4.1 to estimate the population mean and then you estimate the population variance as

$$S_x^2 = \frac{1}{30}\left(597 - 30 \times (4.1)^2\right) = 3.09$$

    a.   What did you do wrong (in one sentence)?
    b.   Should the estimate be larger or smaller?

(7) For a distribution with a mean of 130 and a standard deviation of 15, approximately 68% of the scores will lie between which two scores? What about for approximately 95% of the scores?

(8) The new statistician for the football team at MIT has calculated the following player statistics. Player A's average running yards per carry is 5 with a standard deviation of 3. Player B's average running yards per carry is 5 with a standard deviation of 10.

    a.   Which player is the more consistent yardage gainer? Why?
    b.   Which player is more likely to break loose for a long yardage gain? Why?

(9) Dr. Jones has administered a test to her students. She calculated an average of 86 and a standard deviation of 12. Assuming she estimated the population parameters exactly,

    a.   What is the z-score of a student with a raw score of 80?
    b.   What is the z-score of a student with a raw score of 98?
    c.   What is the raw score for a student with a z-score of -1.5?
    d.   What is the raw score for a student with a z-score of $+1$?

10) Two psychologists, Tversky and Kahneman, asked a group of subjects to carry out the following task. They were told that:

- Linda is 31, single, outspoken, and very bright. She majored in philosophy college. As a student she was deeply concerned with racial discrimination and other social issues, and participated in anti-nuclear demonstrations.

The subjects were then asked to rank the likelihood of various alternatives, such as:
        (1) Linda is active in the feminist movement
        (2) Linda is a bank teller
        (3) Linda is a bank teller and active in the feminist movement

Tversky and Kahneman found that between 85% and 90% of subjects rated alternative (1) most likely, but alternative (3) more likely than alternative (2). Is it?

They call this phenomenon "conjunction fallacy", and note that it appears unaffected by prior training in probability and statistics. Explain why this is a fallacy. We are interested only in a mathematical explanation of the nature of the fallacy, not in your suppositions of psychological motivation.

(11) What proportion of the area under the standard normal curve would you expect to be? (You may use your z-tables. Round answers to the closest percent.)
    a.   Between z=-1.2 and z=+0.6
    b.   Below z=1.4
    c.   Below z=-2.6
    d.   Above z=-2.0

(12) What are the shape, mean, and standard deviation of a sampling distribution of means calculated from samples of size N taken from a normally distribution population with mean μ and variance σ^2?

(13) Given a standard deck of 52 playing cards determine the probability of the following events.
   a. of drawing a 7
   b. of drawing a club
   c. of drawing a 7 or a 10
   d. of drawing a 7 or a club
   e. of drawing a 7 then a 10 (without replacing the 7)

(14) In a sample with a mean of 46 and a standard deviation of 8, what is the probability of randomly selecting each of the following raw scores
   a. a score above 62
   b. a score between 30 and 54
   c. a score below 38

(15) What is the probability of getting 4 or 5 heads in 5 tosses of a fair coin?

(16) Consider a population for which the mean is 53 and standard deviation is 15. Using a criterion of p=0.05 and both tails of the sampling distribution, can either of the following samples (N=25) be called unrepresentative of the population?
   a. A sample with mean 56
   b. A sample with mean 47

(17) How many voters need to be polled to be 95% confident of having a 1% margin of error? Assume the actual percent of voters that favor each of the two candidates is 50%.

(19) For each of the following, note where the mean, median, and mode roughly fall.

a.

b.

19) Suppose 32,753 students nationwide take a standardized test for which the cutoff for passing is 1.2 standard deviations below the mean.
a) About how many students will pass?
b) What is the probability that all of the randomly selected 10 students will pass the test?

20) There is a theory that the anticipation of a birthday can prolong a person's life. While examining this notion statistically, experimenters found that only 60 of 747 people whose obituaries were published in Salt Lake City newspapers died in the three month periods preceding their birthdays. Test the appropriate hypothesis at the 0.01 significance level.

21) Is school performance getting worse? Each year the National Assesment of Educational Progress (NAEP) administers tests to nationwide sample of highschool graduates. The average score of students in 1973 was 55. In 1996 average score for a sample of 20 students was 52 with standard deviation of 10.

a) Is the difference of 3 real or due to chance? Use 5% significance level
b) Build a 95% confidence interval for true mean score in 1996.