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

The mean is not robust to outliers.

(2) a. What does $\Sigma(X-\overline{X})$ mean?

This notates the sum of the deviances from the mean.

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

Zero. This is the reason that we square the deviances when calculating the variance or sd.

(3) A researcher collected several sets of data. For each indicate which mean of central tendency she should compute. Why?

a. The following personality scores
0, 2, 3, 3, 8, 4, 9, 6, 7, 5, 6

Both the mean and median would be appropriate.

b. The following age scores
10,15,18,15,14,13,42,15,12,14,42

Both the mean and median would be appropriate.

c. The following college years
senior, junior, junior, freshman, freshman, junior, sophomore, junior

Either the mode or median would be appropriate.

d. The following the political affiliations
Dem, Dem, Rep, Dem, Soc, Com, Com, Soc, Dem, Rep

The mode would be appropriate.

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

a. −12
b. −7
c. +2
d. +18

+2. The closer the number to the mean, the more frequent that number.

(5) a. Distinguish among the symbols σ and s.

σ = population standard deviation, s = sample sd

b. When are the above symbols accompanied by the square sign (^{2})?

Variance.

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

$$ S^2_X = \frac{740 - \frac{(112)^2}{30}}{30} = 10.729 $$

a. What did you do wrong?

Should be N-1 in this denominator (population correction)

b. Should the estimate be larger or smaller?

Larger.

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

68%: 115,145
95%: 100, 160

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

a. Which player is the more consistent yardage gainer? Why?

Player A. Smaller SD

b. Which player is more likely to break loose for a big run? Why?

Player B. Larger SD.

(9) Dr. Jones has administered a test to her students. She calculated an average of 86 and a standard deviation of 12.

    a. What is the z-score of a student with a raw score of 80?
       -.5
    b. What is the z-score of a student with a raw score of 98?
       +1
    c. What is the raw score for a student for a student with a z-score of –1.5?
       68
    d. What is the raw score for a student for a student with a z-score of +1?
       98

(10) Of the z-scores –2.3,+1,+0.9,-0.6
    a. Which z-score corresponds to the smallest raw score?
    -2.3
    b. Which z-score reflects the raw score having the highest frequency?
    -0.6

(11) What proportion of the area under the standard normal curve would you expect to be (round answers to the closest percent)
    a. Between z=-1.2 and z=+0.6
    ~61%
    b. Below z=1.4
    ~92%
    c. Below z=-2.6
    ~.5%
    d. Above z=-2.0
    ~97.5%

(12) What are the shape, mean, and standard deviation of a sampling distribution of means?

Shape: Normal
Mean: Population Mean
SD: σ /sqrt(N) [N = # of samples, NOT # of trials]

(13) Given a standard deck of 52 playing cards determine the probability of the following events
    a. of drawing a 7
    4/52
    b. of drawing a club
    13/52
    c. of drawing a 7 or a 10
    4/52 + 4/52

d. of drawing a 7 or a club
4/52 + 13/52 – 1/52 (subtract for the 7 of clubs)
e. of drawing a 7 then a 10
4/52 * 4/51

(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
   2.5%
   b. a score between 30 and 54
   81.5%
   c. a score below 38
   16%

(15) What is the probability of getting 4 or 5 heads in 5 tosses of a fair coin?
$_5C_4 * .5^5 + _5C_5 * .5^5 = 5/32 + 1/32 = 6/32 = 3/16$

(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, which of the following samples (N=25) can be called unrepresentative of the population
   a. A sample with mean 56
   15/5 = 3 (sigma of x) ; z = 56 – 53 / 3 = 1 (< 1.96)

   b. A sample with mean 47
   53-47 / 3 = 2 (> 1.96)

(17) X is a random variable with mean 2 and standard deviation 4. You take 9 samples of this random variable and compute the sample mean m. What is the 99% confidence interval?
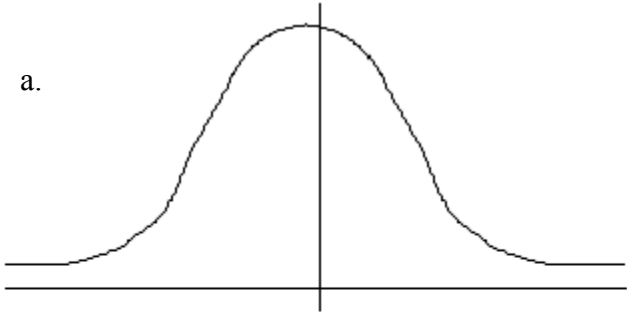
$t_{.005,df=8}$ = 3.36 ; 2 +/- 3.36 * 4/3

(18) 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%.

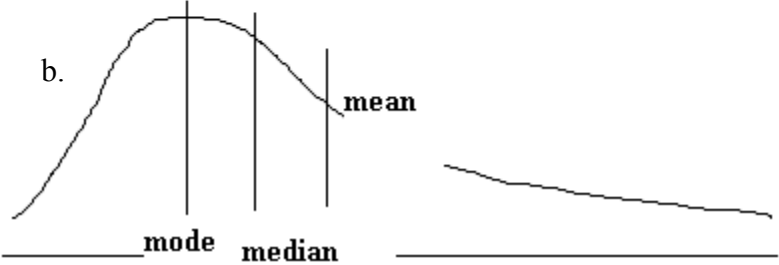margin of error e = σ_p*z = 1.96*$\sqrt{(0.5(1-0.5)/n)}$ = 0.01 ; n = 9604

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

a.

median,mean,mode

b.

mean

mode   median

c.

mean

median

mode