Biostat Exam 10/7/03

Coverage: *StatPrimer* 1–4

Part A (Closed Book)

INSTRUCTIONS

- Write your name in the usual location (back of last page, near the staple), and nowhere else.
- Turn in your Lab Workbook at the beginning of the exam. It is worth [10 pts] (see Workbook p. 2, ¶2).
- The code [M/C] means "multiple choice," indicating that you should select the single *best* response.
- Part A questions are worth ½ pt each unless otherwise noted.
- When you are done with Part A, please *turn it in* and pick up your Procedure Notebook

Chap 1 (Measurement and Sampling)

1. If statistics in *not* merely a compilation of computational techniques, what then *is* it? [1 pt]

2. What does *GIGO* stand for?

3. [M/C] Select the best definition of *measurement*.
   (a) a numerical value that can vary
   (b) a numerical value that can vary randomly
   (c) the assigning of numbers or codes according to prior set rules
   (d) none of the above

4. [M/C] What type of measurement is GRADE when coded (A, B, C, D, or F)?

   (a) Scale (b) Ordinal (c) Nominal

5. [M/C] What type of measurement is GRADE when based on percentage correct on an exam (i.e., 0%–100%)?

   (a) Scale (b) Ordinal (c) Nominal

6. [M/C] What type of measurement is "stage of renal failure" when coded (Stage I, Stage II, or Stage III)?

   (a) Scale (b) Ordinal (c) Nominal

7. [M/C] What type of measurement is "gestational age" when coded (1 = premature, 2 = not premature)?

   (a) Scale (b) Ordinal (c) Nominal

8. [M/C] When data are laid out in a data table, each *column* of data represents:

   (a) an observation (b) a variable (c) a value

9. SPSS calls each *row* of data a "case." Provide an alternative (better) term for this unit of data.

10. [M/C] Which of the following symbols is used to represent a *value*.

(a) x_i (b) X (c) $\overline{x}$ (c) N

11. [M/C] Select the best definition for *statistical* population.
(a) the collection of all possible values for a variable
(b) data in which each person has known probability of entering the sample
(c) data in which each person has equal probability of entering the sample
(d) all people living in a community

12. [M/C] What is the *ratio* of *n* to *N* called?
(a) a simple random sample
(b) a probability sample
(c) the sampling fraction
(d) the sampling frame

13. [M/C] Select the best definition for *simple random sample*.
(a) the collection of all possible values for a variable
(b) data in which each person has a known probability of entering the sample
(c) data in which each person has an equal probability of entering the sample
(d) all people living in a community

Chap 2 (Stem-and-Leaf Plots & Frequencies)

14. [M/C] A distribution with a long *left tail* (toward the lower numbers) is said to be

(a) platykurtotic (b) leptokurtotic (c) positively skewed (d) negatively skewed

15. [M/C] A relatively flat mound-shaped distribution with broad tails is said to be

(a) platykurtotic (b) leptokurtotic (c) positively skewed (d) negatively skewed

16. [M/C] When grouping data into class intervals (for descriptive purposes), how many class intervals should you try for?

(a) 1−3 (b) 3−12 (c) 12−15 (d) 15−20

17. [M/C] Statistical measures of spread quantify the *dispersion of points around* the distribution's:

(a) ends (b) edges (c) center (d) quartiles

18. The number of times a value occurs in a data set is called its

(a) frequency (b) relative frequency (c) cumulative frequency (d) none of the above

Chap 3 (Summary Statistics)

19. Other than the mean and mode, name a measure of central location. _____

20. Other than the variance and standard deviation, name a measure of spread. _____

21. [M/C] When the mean is much greater than the median, the distribution is

    (a) symmetrical (b) skewed positively (c) skewed negatively (d) none of the above

22. [M/C] Which of the following batches of numbers has the greatest *variability*?
    (a) −2 −1 0 1 2
    (b) −4 −2 0 2 4
    (c) −8 −4 0 4 8

23. [M/C] The height (from top to bottom) of the box in a boxplot provides a visual representation of a distribution's

    (a) location (b) spread (c) shape

24. Provide the symbol used to represent the sample mean. _____

25. Provide the symbol used to represent the population variance. _____

26. Provide the symbol used to represent the population standard deviation. _____

27. [M/C] The denominator of the *sample* variance is:

    (a) n (b) $n - 1$ (c) N (d) $N - 1$

28. When data are normally distributed, what % of the values lie within 2 standard deviations of the mean?

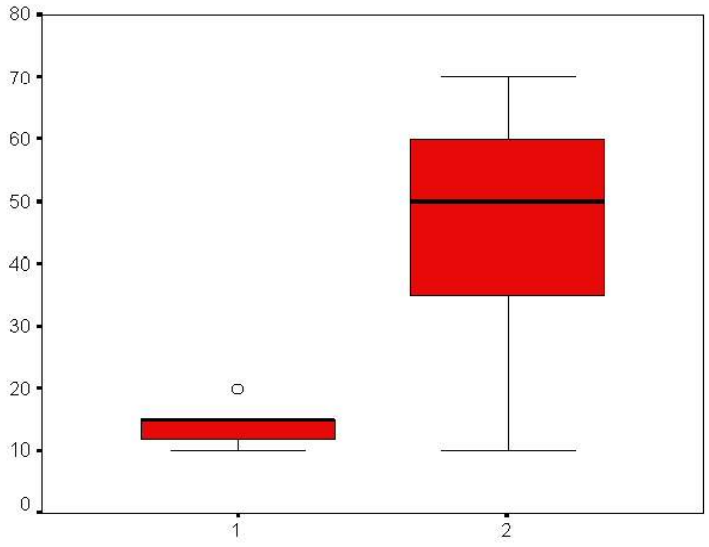
29. With respect to the boxplots shown in the Fig. to the right,

    (a) Which group has higher values, on the average?

    (b) Which group has greater variability?

    (c) What is the approximate value of group 1's median?

    (d) What is the approximate value of group 2's upper hinge?



Chap 4 (Intro to Probability)

30. [M/C] A random sample of 1000 people drawn from a population reveals 25 asthmatics. The investigator concludes that there is a 2.5% chance of being asthmatic. What type of probability is this?

    (a) logical (b) empirical (c) subjective (d) none of the above

31. Since the above survey suggests a 2.5% chance of being asthmatic, what is the chance of NOT being asthmatic?

32. The area under a probability curve sums to _____ (a numerical answer).

33. The two types of random variables are continuous and _____ (a word).

34. Normal distributions are identified by two parameters.

    (a) *Name* the parameter used that determines the distribution's central location:

    (b) Provide the symbol used to represent this parameter:

Part B (Procedure Notebook) *Write your name in the usual location!* **Place a box around final answers.**

Chap 4 (Probability)

1. Normal distribution problems [4 pts]

   (a) Draw a standard normal curve and shade the area corresponding to Pr(Z < −1.02).

   (b) Pr(Z < −1.02) =

   (c) $z_{.995}$ = _____

   (d) A random variable is normally distributed with a mean of 86 and standard deviation of 12. Convert a value of 80 from this distribution to a *z* score.

2. The prevalence of diabetes in a population is 1 in 80 (.0125). Suppose we select at random 4 people from this population. We want to determine probabilities of observing 0, 1, 2, 3, or 4 cases in a given sample. [6 pts]

   (a) What [specific] probability distribution describes this phenomenon?

   (b) What is the probability of observing 0 cases?

   (c) What is the probability of observing *at least one* (one or more) case?

   (d) Would it be unlikely to observe one or more cases if we define "unlikely" as occurring less than 5% of the time? (Yes/No)

Chap 3 (Summary Statistics) [10 pts]

3. A pediatric office sees patients of the following ages (years)

Males:	1	1	5	
Females:	4	4	4	5

(a) *Calculate* the mean age for males. (Show three decimal placed during calculations for this problem.)

(b) *Report* the mean for males. (Round appropriately. Report units.)

(c) *Calculate* the mean for females.

(d) *Report* the mean for females.

(e) *Calculate* the standard deviation for males.

(f) *Calculate* the standard deviation for females.

(g) In plain language, compare the groups.

Chap 3 (Summary Statistics) - *cont.* [10 pts]

4. Here is an ordered array of data:

10 25 30 35 36 40 40 75

(a) The 5-point summary for these data is:

min = _____

Q1 = _____

med = _____

Q3 = _____

max = _____

(b) IQR =

(c) Identify *lower* outside values, *if any*:

(d) Identify the *lower* inside value (bottom whisker point):

(e) Identify *upper* outside values, *if any*:

(f) Identify the upper inside value (upper whisker point):

Chap 2 (Stem-and-leaf plot and freq tables) [10 pts]

5. Scores on an exam are:

84	84	83	81	98
97	95	92	91	91
88	86	86	85	82
78	76	73	58	52

(a) Construct a stem-and-leaf plot of these scores. (Label your axis.)

(b) Describe the distribution.

(c) Fill in this frequency table:

Scores	**Freq**	**Rel Freq**	**Cum Rel Freq**
59 – 50			
69 – 60			
79 – 70			
89 – 80			
99 – 90			
Totals			--