Biostat Exam 1 F04 (10/5/04)
Coverage: Units 1–4

Part A (Closed Book) - Please record your answers to this part of the exam on your scantron.

1. Select valid definitions for *statistics*.

a. a means of learning from data
b. a compilation of computational techniques
c. a way of collecting, organizing, and interpreting information
d. A & C
e. A, B, & C

2. What type of measurement is "stage of cancer" when measured "Stage I, Stage II, Stage III, Stage IV"?

a. scale
b. ordinal
c. nominal

3. What type of measurement is "serum sodium level" when measured in "mg/dL"?

a. scale
b. ordinal
c. nominal

4. When data are laid out in a *data table*, each *column* represents:

a. an observation
b. a variable
c. a value

5. Select the best definition of "statistical population".

a. all people living in a community
b. all collection of all possible values for a variable
c. subset of people living in a community
d. a subset of possible values for a variable

6. What do you call the ratio of the sample size to population size?

a. a simple random sample
b. a probability sample
c. the sampling fraction
d. the sampling frame

7. The principle of "GIGO" implies that an analysis is only as good as the quality of data upon which it is based.

a. True
b. False

8. In a simple random sample, everyone in the population has _____ chance of entering the sample.

a. a zero
b. an equal
c. an unequal
d. an unknown

9. The number of times a value occurs in a data set is its:

a. frequency
b. relative frequency
c. cumulative relative frequency

10. Stem-and-leaf plots may be used to learn about a distribution's:

a. shape
b. location
c. spread
d. all of the above

11. The "plus-minus" value around a distribution's center provides clues as to its:

a. shape
b. location
c. spread

12. Which of the following is not a measure of spread?

a. inter-quartile range
b. standard deviation
c. median
d. variance

13. Which of the following symbols is used to represent a *population* mean?

a. μ
b. $\bar{x}$
c. s
d. σ

14. Which of the following symbols is used to represent a sample standard deviation?

a. μ
b. s
c. s^2
d. σ

15. Which of the following is true when a distribution is symmetrical?

a. mean > median
b. mean < median
c. mean = median

16. The height (from hinge to hinge) of a box in a boxplot provides visual clues as to:

a. central location
b. spread
c. shape
d. none of the above

17. The line inside a boxplot represents the distribution's:

a. mean
b. median
c. mode
d. inter-quartile range

18. The sample mean provides a reflection of:

a. a data point selected at random from the sample
b. a data point selected at random from the population
c. the population mean
d. all of the above

19. Upper and lower fences do NOT appear on boxplots.

a. True
b. False

20. The standard deviation is the _____ of the variance.

a. mean
b. average
c. square
d. square root

21. Select the technical term for "spread".

a. symmetry
b. modality
c. variability
d. kurtosis

22. The cumulative probability associted with a z score of 0 is

a. 50%
b. 68%
c. 95%
d. 99%

23. Random variables are numeric quantities that vary depending on chance.

a. True
b. False

24. Normal random variables are characterized by the following parameters:

a. mu & sigma
b. mu & p
c. n & p
d. n & sigma

25. The area under an entire probability curve always sums to:

a. 0
b. 1
c. 2
d. infinity

26. The mean of the standard normal (Z) distribution is

a. 0
b. 1
c. 2
d. infinity

27. If the probability of an event is .95, the probability of its complement is:

a. 0
b. .05
c. .5
d. 1

28. Binomial parameter p represents

a. the number of independent trials in a study
b. the probability of success for each trial
c. the probability of failure for each trial
d. none of the above

29. Before a study is conducted, a scientist believes a particular theory has only a 10% chance of being correct. This probability is conceived on:

a. logic
b. experience
c. subjectivity

30. Select the probability equivalent to $Pr(Z > 2.22)$:

a. $Pr(Z < 2.22)$
b. $Pr(Z < -2.22)$
c. $Pr(Z > -2.22)$
d. none of the above

Part B (Procedure Book) - Record your answers to this part of the exam in your blue book. Work carefully and neatly. *Please remember to number each response!*

(1) This set of questions refers to the data in this table [10 pts]:

ID	AGE	SEX	EXPOSED	CANCER	DOSE	SCR	WEIGHT
1	50	F	N	Y	36.0	.8	66
2	21	M	N	Y	29.0	1.1	68
3	35	F	Y	Y	16.2	.7	97
4	49	M	N	N	29.0	.8	83
5	38	F	Y	N	16.2	1.4	97
6	42	M	Y	N	18.0	1.0	82

(A) How many variables are there in this table?
(B) If you were creating an SPSS data set with these data, which variables would be set up as "string" types?
(C) Is DOSE a scale, ordinal, or nominal measurement?
(D) Construct a stem and leaf plot of the AGE data.
(E) Calculate the mean AGE.
(F) Calculate the standard deviation of AGE. Show all work.

(2) The next set of questions address this stem-and-leaf plot [8 pts]:

```
1|788
2|00112334458999
3|112355889
4|1122
5|0
6|12
(x 10)
```

(A) Describe the shape of the distribution.
(B) Report the approximate *gravitational* center of this distribution. [Answer need not be exact. No arithmetic necessary.]
(C) Describe the spread of the distribution.
(D) What is the relative frequency of values in the 10-19 range? (Note, $n = 33$.)
(E) What is the relative frequency of values in the 20-29 range?
(F) What is the cumulative (relative) frequency of values in the 20-29 range?
(G) What is the depth of the median?
(F) What is the median?

(3) Here is an ordered array [10 pts]:

```
       16      22      24      28      29      41
```

(A) Report the five point summary.
(B) IQR =
(C) Are there any outside values on top? If so, report them.
(D) Upper inside value =
(E) Are there any outside values on bottom? If so report them.
(F) Lower inside value =

(4) You select at random a sample of 4 people from a geriatric population. The prevalence of hypertension in this population is .25. [6 pts]
(A) What probability distribution describes the random number of hypertensives in a sample?
(B) What are the values of the parameters for the probability distribution?
(C) What is the probability of selecting a sample with 4 hypertensives?

(5) Questions about the normal distribution. [6 pts]
(A) $Pr(Z < 1.76) =$
(B) $Pr(Z > 0.56) =$
(C) $z_{.01} =$
(D) A population is normally distributed with a mean of 50 and standard deviation of 5. What is the z score associated with a value of 57.5 from this population?
(H) What is the probability of seeing a value that is less than 57.5 from the population described in part G?