

063:161 Introduction to Biostatistics
Fall 1992, Exam 1

I. True/False Questions (2 points each)

NOTE: If the statement is true, mark response A
If the statement is false, mark response B

1. One of the advantages of a histogram over a stem-and-leaf plot is that it is not necessary to first sort or group the data.
2. The sample mode is always a unique measure of central location.
3. The relative risk is the ratio of two conditional probabilities.
4. The sample mean is not changed if a constant is added to each observation in a sample.
5. The sample standard deviation is not changed if a constant is added to each observation in a sample.
6. Let x_1, x_2, \dots, x_n denote a sample of n observations. Then $\sum_{i=1}^n (x_i - \bar{x})$ is always equal to zero.
7. The sample range is a measure of central location.
8. One of the advantages of a stem-and-leaf plot over a histogram is that the actual data values are displayed.
9. $\Pr(A) + \Pr(\bar{A}) = 1$, for any event A .
10. If the events A and B are independent, then $\Pr(A \cap B) = \Pr(A) + \Pr(B)$.
11. If A and B are any events, then $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.
12. The coefficient of variation is a measure of central location.
13. For any two events A and B , the conditional probability of B given A is
$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$
14. If A and B are independent events, then $\Pr(A|B) = \Pr(A)$.
15. For any two events A and B , the events $\{A \cap B\}$ and $\{A \cap \bar{B}\}$ are independent.
16. For any two events A and B , $\Pr(A) = \Pr(A \cap B) + \Pr(A \cap \bar{B})$.
17. The relative risk is always between 0 and 1.

II. Multiple-Choice Questions (unless stated otherwise, each question is worth 3 points)

NOTE: Use the following responses for the next four questions.

- A – sample mean
- B – sample median
- C – sample variance
- D – sample standard deviation
- E – sample range

18. Which measure is completely determined by only two of the data values?
19. Which measure of central location is the easiest to calculate if the data values are listed in sorted order (from smallest to largest)?
20. Which measure of central location is unaffected if the largest data value in a sample of size n (with $n > 2$) is increased by 10?
21. Which measure of variability does not have the same units of measure as the data values?

The next two questions refer to the following frequency distribution of the age at onset for a sample of 70 newly-diagnosed diabetics:

<u>Age</u>	<u>Count</u>
0–19	5
20–29	8
30–39	12
40–49	17
50–59	23
60–79	5

22. The most appropriate graphical method for displaying these data is a:
 - A frequency distribution
 - B bar graph
 - C histogram
 - D stem-and-leaf plot
 - E cumulative frequency distribution
23. If these data were displayed in a histogram, and if the height of the bar for the category 30–39 years was 12 units, what would be the height of the bar for the category 60–79 years?
 - A 2.5
 - B 5
 - C 10

The next six questions refer to data collected from a sample of six men admitted to the hospital for a serious illness. The lengths of stay (weeks) were:

4 8 3 4 5 6

24. The sample mean is:

- A 4
- B 4.5
- C 5
- D 5.5
- E 6

25. The sample median is:

- A 4
- B 4.5
- C 5
- D 5.5
- E 6

26. The sample variance is:

- A 1.63
- B 1.79
- C 2.67
- D 3.2
- E 4

27. The coefficient of variation is:

- A 32.7%
- B 35.8%
- C 39.8%
- D 53.3%
- E 55.9%

28. If the data are expressed in terms of days (instead of weeks), the sample mean is:

- A 49 times larger than the sample mean calculated using weeks
- B 7 times larger than the sample mean calculated using weeks
- C the same as the sample mean calculated using weeks
- D 7 times smaller than the sample mean calculated using weeks
- E 49 times smaller than the sample mean calculated using weeks

29. If the data are expressed in terms of days (instead of weeks), the sample variance is:

- A 49 times larger than the sample variance calculated using weeks
- B 7 times larger than the sample variance calculated using weeks
- C the same as the sample variance calculated using weeks
- D 7 times smaller than the sample variance calculated using weeks
- E 49 times smaller than the sample variance calculated using weeks

The next six questions (two points each) refer to the following situation. Suppose that we have just discovered a new test for the AIDS virus which can be carried out easily using a machine that can be installed in shopping malls. For a randomly-selected person from the general population, let A denote the event that the test is positive for the AIDS virus, and let B denote the event that the individual truly has the AIDS virus (determined by a much more expensive test).

30. The predictive value positive of the new test is:

- A $\Pr(A|B)$
- B $\Pr(B|A)$
- C $\Pr(\bar{A}|\bar{B})$
- D $\Pr(\bar{B}|\bar{A})$
- E $\Pr(A|\bar{B})$

31. The predictive value negative of the new test is:

- A $\Pr(B|A)$
- B $\Pr(B|\bar{A})$
- C $\Pr(\bar{A}|\bar{B})$
- D $\Pr(\bar{B}|\bar{A})$
- E $\Pr(A|\bar{B})$

32. The sensitivity of the new test is:

- A $\Pr(A|B)$
- B $\Pr(A|\bar{B})$
- C $\Pr(\bar{A}|B)$
- D $\Pr(\bar{A}|\bar{B})$
- E $\Pr(\bar{B}|\bar{A})$

33. The specificity of the new test is:

- A $\Pr(A|B)$
- B $\Pr(A|\bar{B})$
- C $\Pr(\bar{A}|B)$
- D $\Pr(\bar{A}|\bar{B})$
- E $\Pr(\bar{B}|\bar{A})$

34. The probability of a false negative result is:

- A $\Pr(A|B)$
- B $\Pr(B|A)$
- C $\Pr(\bar{A}|\bar{B})$
- D $\Pr(\bar{B}|\bar{A})$
- E $\Pr(\bar{A}|B)$

35. The probability of a false positive result is:

- A $\Pr(A|B)$
- B $\Pr(\bar{A}|B)$
- C $\Pr(A|\bar{B})$
- D $\Pr(\bar{A}|\bar{B})$
- E $\Pr(\bar{B}|\bar{A})$

The next six questions refer to the following situation. In a high-risk population of elderly individuals, suppose that the probability of having Alzheimer's Disease is 0.3, that the probability of having an excessively high concentration of aluminum in the blood is 0.2, and that the probability of having Alzheimer's Disease and a high aluminum level is 0.1.

36. Are the events {Alzheimer's Disease} and {High Aluminum} independent?

- A yes
- B no

37. What is the probability that a randomly-selected individual from this population has a high aluminum level or Alzheimer's disease?

- A 0.06
- B 0.2
- C 0.4
- D 0.5
- E 0.6

38. If a randomly-selected individual from this population has Alzheimer's disease, what is the probability that the individual also has a high aluminum level?

- A 0.1
- B 0.143
- C 0.2
- D 0.33
- E 0.5

39. If a randomly-selected individual from this population has a high aluminum level, what is the probability that the individual also has Alzheimer's Disease?

- A 0.1
- B 0.125
- C 0.2
- D 0.33
- E 0.5

40. If a randomly-selected individual from this population has a normal aluminum level, what is the probability that the individual also has Alzheimer's Disease?

- A 0.125
- B 0.25
- C 0.375
- D 0.5
- E 0.875

41. What is the relative risk of Alzheimer's disease for individuals with high aluminum levels relative to individuals with normal aluminum levels?

- A 0.333
- B 0.5
- C 1.0
- D 2.0
- E 3.0