

Biost 517
Applied Biostatistics I

.....

Scott S. Emerson, M.D., Ph.D.
Professor of Biostatistics
University of Washington

First Quiz and Discussion

October 5, 2012

1

© 2002, 2003, 2005 Scott S. Emerson, M.D., Ph.D.

Question 1

.....

- A randomized clinical trial is conducted in cardiac arrest patients for whom 911 is called. Of interest is to investigate whether three minutes of CPR prior to delivery of shocks leads to better survival than immediate defibrillation.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

2

Question 2

.....

- Data on 250 healthy volunteers is analyzed in order to determine the “normal range” for the time it takes for a nerve impulse to get from the ankle to the brain.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

3

Question 3

.....

- Data on liver function tests from 312 patients with primary biliary cirrhosis is used to estimate the date that a patient should be placed on the transplantation list. The goal is to list the patient approximately one year prior to the time the patient might otherwise be expected to die.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

4

Question 4

- In a phase 1 clinical trial, the rates of absorption and elimination and volume of distribution for a new drug is studied in a group of healthy volunteers.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

5

Question 5

- Netflix offers a \$1 million prize to anyone who can improve their ability to recommend movies to their customers. Data is available on an average of 209 movie ratings supplied by each of 480,000 customers.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

6

Question 6

- Consider an **observational study** of the association between **hormone therapy and risk of cancer**.
- The investigator might consider analyses **involving socioeconomic status** because :
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

7

Question 7

- Consider an **observational study** of the association between **dietary calcium intake and growth**.
- The investigator might consider analyses **involving age** because :
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

8

Question 8

- Consider an observational study of the association between smoking and emphysema.
- The investigator might consider analyses involving bronchitis because:
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

9

Question 9

- Consider a randomized clinical trial of transcendental meditation versus exercise in lowering blood pressure.
- The investigator might consider analyses involving the blood pressure at the time of randomization because :
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

10

Question 10

- Consider a randomized clinical trial of dietary sucrose on patterns of gene expression.
- The R03 funded investigator might consider a design in which no measures of gene expression are made at the time of randomization because :
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

11

Question 11

- For the purposes of detecting errors in the data, the most useful descriptive statistic is
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

12

Question 12

- For the purposes of describing **materials and methods** for the study, the most useful descriptive statistic to describe the **central tendency (location)** of the data is
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

13

Question 13

- For the purposes of describing **materials and methods** for the study, the most useful descriptive statistic to describe the **spread (variability)** of the data is
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

14

Question 14

- For the purposes of assessing **the possibility of confounding** in the study, the most useful descriptive statistic to use in **stratified analysis** of the data is
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

15

Question 15

- For the purposes of assessing **validity of technical assumptions** for the study, the most useful descriptive statistic to use in **stratified analysis** of the data is
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

16

Question 16

- For the purposes of obtaining **preliminary estimates of association** for the study, the most useful descriptive statistic to use in **stratified analysis** of the data is
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

17

Question 17

- For the purposes of **exploring effects within subgroups** for the study, the most useful descriptive statistic to use in **stratified analysis** of the data is
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

18

Question 18

- The **sample mean** is an appropriate summary measure for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

19

Question 19

- The **sample geometric mean** is an appropriate summary measure for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

20

Question 20

- The **sample median** is an appropriate summary measure for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

21

Question 21

- The **sample quartiles** are appropriate summary measures for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

22

Question 22

- **Sample proportions** are appropriate summary measures for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

23

Question 23

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- Some patients are noted to **already be colonized at their very first clinic visit**
- In the resulting data, we would characterize the data on these subjects as
 - A. Missing
 - B. Left censored
 - C. Right censored
 - D. Interval censored
 - E. Any of the above

24

Question 24

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- Some patients are noted to be still not colonized at the time of data analysis
- In the resulting data, we would characterize the data on these subjects as
 - A. Missing
 - B. Left censored
 - C. Right censored
 - D. Interval censored
 - E. Any of the above

25

Question 25

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- Some patients were noted to be free of colonization at one visit, but colonized at their next visit one year later
- In the resulting data, we would characterize the data on these subjects as
 - A. Missing
 - B. Left censored
 - C. Right censored
 - D. Interval censored
 - E. Any of the above

26

Question 26

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- At some visits, the clinical staff forgot to obtain samples to see whether the patients were colonized
- In the resulting data, we would characterize the data for these visits as
 - A. Missing
 - B. Left censored
 - C. Right censored
 - D. Interval censored
 - E. Any of the above

27

Questions 27 through 32

- For each of the following continents, what is the first language of the most people and second most people living on that continent?

27. Asia
28. Europe
29. Africa
30. North America
31. South America
32. Australia

28

Answers and Discussion

.....

29

Question 1

.....

- A randomized clinical trial is conducted in cardiac arrest patients for whom 911 is called. Of interest is to investigate whether three minutes of CPR prior to delivery of shocks leads to better survival than immediate defibrillation.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups**
 - E. Predicting future observations

30

Question 1

.....

- A randomized clinical trial is conducted in cardiac arrest patients for whom 911 is called. Of interest is to investigate whether three minutes of CPR prior to delivery of shocks leads to better survival than immediate defibrillation.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups**
 - E. Predicting future observations

31

Question 2

.....

- Data on 250 healthy volunteers is analyzed in order to determine the “normal range” for the time it takes for a nerve impulse to get from the ankle to the brain.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

32

Question 2

- Data on 250 healthy volunteers is analyzed in order to determine the “normal range” for the time it takes for a nerve impulse to get from the ankle to the brain.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions**
 - D. Comparing distributions across groups
 - E. Predicting future observations**

33

Question 3

- Data on liver function tests from 312 patients with primary biliary cirrhosis is used to estimate the date that a patient should be placed on the transplantation list. The goal is to list the patient approximately one year prior to the time the patient might otherwise be expected to die.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

34

Question 3

- Data on liver function tests from 312 patients with primary biliary cirrhosis is used to estimate the date that a patient should be placed on the transplantation list. The goal is to list the patient approximately one year prior to the time the patient might otherwise be expected to die.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations**

35

Question 4

- In a phase 1 clinical trial, the rates of absorption and elimination and volume of distribution for a new drug is studied in a group of healthy volunteers.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

36

Question 4

- In a phase 1 clinical trial, the rates of absorption and elimination and volume of distribution for a new drug is studied in a group of healthy volunteers.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

37

Question 5

- Netflix offers a \$1 million prize to anyone who can improve their ability to recommend movies to their customers. Data is available on an average of 209 movie ratings supplied by each of 480,000 customers.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

38

Question 5

- Netflix offers a \$1 million prize to anyone who can improve their ability to recommend movies to their customers. Data is available on an average of 209 movie ratings supplied by each of 480,000 customers.
- Using the classification of scientific questions discussed in class, the goal of this study most closely aligns with:
 - A. Clustering cases
 - B. Clustering variables
 - C. Quantifying summary measures of distributions
 - D. Comparing distributions across groups
 - E. Predicting future observations

39

Question 6

- Consider an observational study of the association between hormone therapy and risk of cancer.
- The investigator might consider analyses involving socioeconomic status because :
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

40

Question 6

- Consider an **observational study** of the association between **hormone therapy and risk of cancer**.
- The investigator might consider analyses **involving socioeconomic status** because :
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding**
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

41

Question 7

- Consider an **observational study** of the association between **dietary calcium intake and growth**.
- The investigator might consider analyses **involving age** because :
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

42

Question 7

- Consider an **observational study** of the association between **dietary calcium intake and growth**.
- The investigator might consider analyses **involving age** because :
 - A. It is of interest to investigate effect modification**
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

43

Question 8

- Consider an **observational study** of the association between **smoking and emphysema**.
- The investigator might consider analyses **involving bronchitis** because:
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

44

Question 8

- Consider an observational study of the association between smoking and emphysema.
- The investigator might consider analyses involving bronchitis because:
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish**

45

Question 9

- Consider a randomized clinical trial of transcendental meditation versus exercise in lowering blood pressure.
- The investigator might consider analyses involving the blood pressure at the time of randomization because :
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

46

Question 9

- Consider a randomized clinical trial of transcendental meditation versus exercise in lowering blood pressure.
- The investigator might consider analyses involving the blood pressure at the time of randomization because :
- A. It is of interest to investigate effect modification
- B. It is of interest to address potential confounding
- C. It is of interest to obtain added precision of inference**
- D. He / she is foolish

47

Question 10

- Consider a randomized clinical trial of dietary sucrose on patterns of gene expression.
- The R03 funded investigator might consider a design in which no measures of gene expression are made at the time of randomization because :
- A. It is of interest to investigate effect modification
- B. It is of interest to address potential confounding
- C. It is of interest to obtain added precision of inference
- D. He / she is foolish

48

Question 10

- Consider a randomized clinical trial of dietary sucrose on patterns of gene expression.
- The R03 funded investigator might consider a design in which no measures of gene expression are made at the time of randomization because :
 - A. It is of interest to investigate effect modification
 - B. It is of interest to address potential confounding
 - C. It is of interest to obtain added precision of inference
 - D. He / she is foolish

49

Question 11

- For the purposes of detecting errors in the data, the most useful descriptive statistic is
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

50

Question 11

- For the purposes of detecting errors in the data, the most useful descriptive statistic is
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

51

Role of Minima, Maxima

- In monitoring clinical trials, need to distinguish between
 - Adverse experiences that affect nearly everyone
 - Adverse experiences that affect relatively few
- While not quite an error, minima and maxima are also the most important for detecting individual level toxicities
 - Descriptive statistics stratified by treatment group

52

Question 12

- For the purposes of describing materials and methods for the study, the most useful descriptive statistic to describe the central tendency (location) of the data is
 - Mean
 - Median
 - Geometric mean
 - Standard deviation
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

53

Question 12

- For the purposes of describing materials and methods for the study, the most useful descriptive statistic to describe the central tendency (location) of the data is
 - Mean
 - Median
 - Geometric mean
 - Standard deviation
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

54

Question 13

- For the purposes of describing materials and methods for the study, the most useful descriptive statistic to describe the spread (variability) of the data is
 - Mean
 - Median
 - Geometric mean
 - Standard deviation
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

55

Question 13

- For the purposes of describing materials and methods for the study, the most useful descriptive statistic to describe the spread (variability) of the data is
 - Mean (proportion above important thresholds)
 - Median
 - Geometric mean
 - Standard deviation
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

56

Question 14

- For the purposes of assessing the possibility of confounding in the study, the most useful descriptive statistic to use in stratified analysis of the data is
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

57

Question 14

- For the purposes of assessing possibility of confounding for the study, the most useful descriptive statistic to use in stratified analysis of the data is
 - A. Mean *(if adjust on additive scale)*
 - B. Median
 - C. Geometric mean *(if adjust multiplicatively)*
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

58

Linear Regression

- Difference in interpretation of slopes

Unadjusted Model : $E[Y_i | X_i] = \beta_0 + \beta_1 \times X_i$

- β_1 = Diff in mean Y for groups differing by 1 unit in X
 - (The distribution of W might differ across groups being compared)

Adjusted Model : $E[Y_i | X_i, W_i] = \gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i$

- γ_1 = Diff in mean Y for groups differing by 1 unit in X, but agreeing in their values of W

59

Relationships: True Slopes

- The slope of the unadjusted model will tend to be

$$\beta_1 = \gamma_1 + \rho_{XW} \frac{\sigma_W}{\sigma_X} \gamma_2$$

- Hence, true adjusted and unadjusted slopes for X are estimating the same quantity only if
 - $\rho_{XW} = 0$ (X and W are truly uncorrelated), OR
 - $\gamma_2 = 0$ (no association between W and Y after adjusting for X)

60

Conditional Confounding

$$g[\theta | \mathbf{X}] = \mathbf{X}\bar{\beta} \qquad g[\theta | \mathbf{X}, \mathbf{W}] = \mathbf{X}\bar{\gamma} + \mathbf{W}\bar{\delta}$$

$$\bar{\beta} = \bar{\gamma} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \bar{\delta}$$

$$\beta_1 = \gamma_1 + \sum_{j=1}^p (\bar{W}_{1j\bullet} - \bar{W}_{0j\bullet}) \delta_j$$

$$\bar{W}_{kj\bullet} = \frac{1}{n_k} \sum_{i=1}^n W_{ij} 1_{[X_i=k]}$$

51

Relationships: Estimated Slopes

- The estimated slope of the unadjusted model will be

$$\hat{\beta}_1 = \hat{\gamma}_1 \left(1 + \hat{\gamma}_2 r_{XW} \left[\frac{s_W}{s_X (r_{YX} - r_{YW} r_{XW})} \right] \right)$$

- Hence, estimated adjusted and unadjusted slopes for X are equal only if
 - $r_{XW} = 0$ (X and W are uncorrelated in the sample, which can be arranged by experimental design), OR
 - $\hat{\gamma}_2 = 0$ (which cannot be predetermined, because Y is random)
 - $s_W = 0$ (W is controlled at a single value in which case $r_{XW} = 0$)

52

Relationships: True SE

Unadjusted Model
$$[se(\hat{\beta}_1)]^2 = \frac{Var(Y | X)}{nVar(X)}$$

Adjusted Model
$$[se(\hat{\gamma}_1)]^2 = \frac{Var(Y | X, W)}{nVar(X)(1 - r_{XW}^2)}$$

$$Var(Y | X) = \gamma_2^2 Var(W | X) + Var(Y | X, W)$$

$$\sigma_{Y|X}^2 = \gamma_2^2 \sigma_{W|X}^2 + \sigma_{Y|X,W}^2$$

53

Relationships: True SE

Unadjusted Model
$$[se(\hat{\beta}_1)]^2 = \frac{Var(Y | X)}{nVar(X)}$$

Adjusted Model
$$[se(\hat{\gamma}_1)]^2 = \frac{Var(Y | X, W)}{nVar(X)(1 - r_{XW}^2)}$$

$$Var(Y | X) = \gamma_2^2 Var(W | X) + Var(Y | X, W)$$

Thus, $se(\hat{\beta}_1) = se(\hat{\gamma}_1)$ if

$$r_{XW} = 0$$

AND

$$\gamma_2 = 0 \quad \text{OR} \quad Var(W | X) = 0$$

54

Question 15

- For the purposes of assessing validity of technical assumptions for the study, the most useful descriptive statistic to use in stratified analysis of the data is
 - Mean
 - Median
 - Geometric mean
 - Standard deviation
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

65

Question 15

- For the purposes of assessing validity of technical assumptions for the study, the most useful descriptive statistic to use in stratified analysis of the data is
 - Mean
 - Median
 - Geometric mean
 - D. Standard deviation**
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

66

Question 16

- For the purposes of obtaining preliminary estimates of association for the study, the most useful descriptive statistic to use in stratified analysis of the data is
 - Mean
 - Median
 - Geometric mean
 - Standard deviation
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

67

Question 16

- For the purposes of obtaining preliminary estimates of association for the study, the most useful descriptive statistic to use in stratified analysis of the data is
 - A. Mean**
 - Median
 - C. Geometric mean**
 - Standard deviation
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

68

Inference for Means

- Most common parameter used as a basis for statistical inference is the mean
 - Proportions = mean of binary variable
 - log Geometric mean = mean of log transformed data
- Tends to reflect a wide variety of differences between distributions
 - E.g., extremely sensitive to changes in the tail of distributions
- Statistical theory allow us to know the sampling distribution, and thus allows us to do inference

69

Question 17

- For the purposes of exploring effects within subgroups for the study, the most useful descriptive statistic to use in stratified analysis of the data is
 - Mean
 - Median
 - Geometric mean
 - Standard deviation
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

70

Question 17

- For the purposes of exploring effects within subgroups for the study, the most useful descriptive statistic to use in stratified analysis of the data is
 - Mean
 - Median
 - Geometric mean
 - Standard deviation
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

71

Question 18

- The sample mean is an appropriate summary measure for which types of measurements?
 - Binary
 - Nominal (unordered categorical)
 - Ordered categorical
 - Counts
 - Interval continuous data
 - Ratio continuous data
 - Censored continuous data

72

Question 18

- The sample mean is an appropriate summary measure for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical (for direction of effect)
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

73

Question 19

- The sample geometric mean is an appropriate summary measure for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

74

Question 19

- The sample geometric mean is an appropriate summary measure for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data (if positive)
 - F. Ratio continuous data (if positive)
 - G. Censored continuous data

75

Question 20

- The sample median is an appropriate summary measure for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

76

Question 20

- The sample median is an appropriate summary measure for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

77

Question 21

- The sample quartiles are appropriate summary measures for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

78

Question 21

- The sample quartiles are appropriate summary measures for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

79

Question 22

- Sample proportions are appropriate summary measures for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

80

Question 22

- Sample proportions are appropriate summary measures for which types of measurements?
 - A. Binary
 - B. Nominal (unordered categorical)
 - C. Ordered categorical
 - D. Counts
 - E. Interval continuous data
 - F. Ratio continuous data
 - G. Censored continuous data

31

Question 23

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- Some patients are noted to already be colonized at their very first clinic visit
- In the resulting data, we would characterize the data on these subjects as
 - A. Missing
 - B. Left censored
 - C. Right censored
 - D. Interval censored
 - E. Any of the above

32

Question 23

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- Some patients are noted to already be colonized at their very first clinic visit
- In the resulting data, we would characterize the data on these subjects as
 - A. Missing
 - B. Left censored
 - C. Right censored
 - D. Interval censored
 - E. Any of the above

33

Question 24

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- Some patients are noted to be still not colonized at the time of data analysis
- In the resulting data, we would characterize the data on these subjects as
 - A. Missing
 - B. Left censored
 - C. Right censored
 - D. Interval censored
 - E. Any of the above

34

Question 24

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- Some patients are noted to be still not colonized at the time of data analysis
- In the resulting data, we would characterize the data on these subjects as
 - A. Missing
 - B. Left censored
 - C. **Right censored**
 - D. Interval censored
 - E. Any of the above

85

Question 25

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- Some patients were noted to be free of colonization at one visit, but colonized at their next visit one year later
- In the resulting data, we would characterize the data on these subjects as
 - A. Missing
 - B. Left censored
 - C. Right censored
 - D. Interval censored
 - E. Any of the above

86

Question 25

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- Some patients were noted to be free of colonization at one visit, but colonized at their next visit one year later
- In the resulting data, we would characterize the data on these subjects as
 - A. Missing
 - B. Left censored
 - C. Right censored
 - D. **Interval censored**
 - E. Any of the above

87

Question 26

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- At some visits, the clinical staff forgot to obtain samples to see whether the patients were colonized
- In the resulting data, we would characterize the data for these visits as
 - A. Missing
 - B. Left censored
 - C. Right censored
 - D. Interval censored
 - E. Any of the above

88

Question 26

- A cohort study is conducted in cystic fibrosis patients in order to assess the age at which the patients become colonized with *Pseudomonas*.
- At some visits, the clinical staff **forgot to obtain samples to see whether the patients were colonized**
- In the resulting data, we would characterize the data for these visits as
 - A. **Missing**
 - B. Left censored
 - C. Right censored
 - D. Interval censored
 - E. Any of the above

89

Questions 27 through 32

- For each of the following continents, what is the first language of the most people and second most people living on that continent?
27. Asia
 28. Europe
 29. Africa
 30. North America
 31. South America
 32. Australia

90

Questions 27 through 32

- For each of the following continents, what is the first language of the most people and second most people living on that continent?
- | | |
|-------------------|--|
| 27. Asia | <u>Chinese (Mandarin); Hindi</u> |
| 28. Europe | <u>Russian, German</u> |
| 29. Africa | <u>Arabic, Hausa</u> |
| 30. North America | <u>English, Spanish</u> |
| 31. South America | <u>Portuguese, Spanish</u> |
| 32. Australia | <u>English, (Chinese) Italian</u> |

91