

Biost 517
Applied Biostatistics I
.....
Scott S. Emerson, M.D., Ph.D.
Professor of Biostatistics
University of Washington

Quiz and Discussion

November 3, 2010

1

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

Question 1
.....

- 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

2

Question 2
.....

- 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

3

Question 3
.....

- 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

4

Question 4

-
- 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

5

Question 5

-
- 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
- Mean
 - Median
 - Geometric mean
 - Standard deviation
 - Minimum, maximum
 - 25th and 75th percentiles
 - Histogram

6

Question 6

-
- 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

7

Question 7

-
- 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

8

Question 8

- In the presence of **censored observations**, it is never **possible to estimate population**
- A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. 25th and 75th Percentiles
 - F. Probability of exceeding thresholds
 - G. All of the above
 - H. None of the above

9

Question 9

- In the presence of **censored observations**, it is **usually easiest to estimate population**
- A. Mean
 - B. Median and other percentiles
 - C. Geometric mean
 - D. Standard deviation
 - E. Probability of exceeding thresholds
 - F. All of the above
 - G. None of the above

10

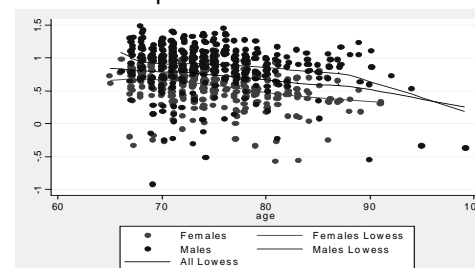
Question 10

- Consider the **correlation between age and salary** among full professors in the Arts aged 50 - 60 years old
 - Holding everything else constant, which of the following changes to study design might lead to a **higher sample correlation**
- A. Increasing sample size
 - B. Increasing the ages considered to 40 – 70 years old
 - C. Increasing the fields considered to include Professional schools
 - D. All of the above

11

Question 11

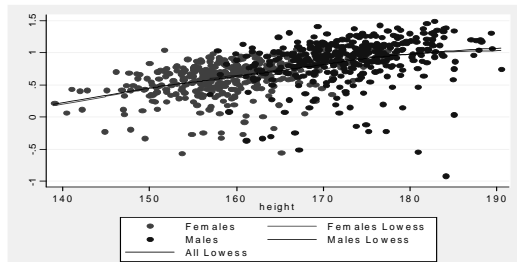
- In the scatterplot of log FEV versus age stratified by sex, we would **expect more extreme correlation** in
- A. Men
 - B. Women
 - C. Combined sample



12

Question 12

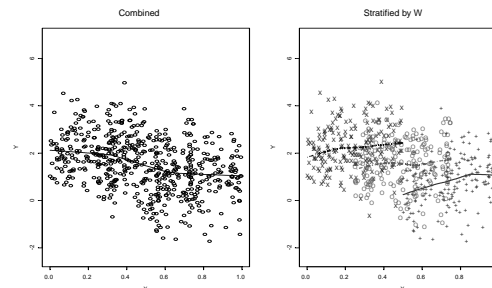
- In the scatterplot of log FEV versus height stratified by sex, we would **expect more extreme correlation** in
 - Men
 - Women
 - Combined sample



13

Question 13

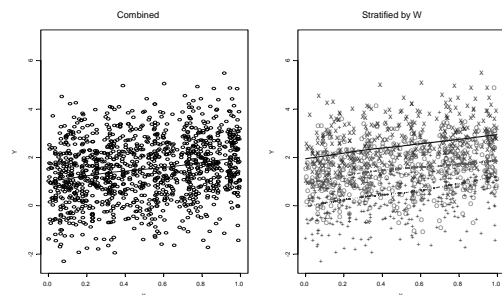
- Plot of Y vs X: **Stratification variable W appears to be**
 - An effect modifier
 - A confounder
 - A precision variable



14

Question 14

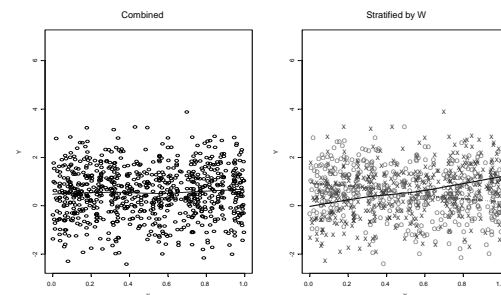
- Plot of Y vs X: **Stratification variable W appears to be**
 - An effect modifier
 - A confounder
 - A precision variable



15

Question 15

- Plot of Y vs X: **Stratification variable W appears to be**
 - An effect modifier
 - A confounder
 - A precision variable



16

Answers and Discussion

.....

17

Question 1

.....

- 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

18

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

19

Question 2

.....

- 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

20

Question 3

-
- 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 (**proportion above important thresholds**)
 - B. Median
 - C. Geometric mean
 - D. Standard deviation**
 - E. Minimum, maximum**
 - F. 25th and 75th percentiles**
 - G. Histogram

21

Question 4

-
- 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

22

Question 5

-
- 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**
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. Minimum, maximum
 - F. 25th and 75th percentiles
 - G. Histogram

23

Question 6

-
- 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

24

Inference for Means

.....

- Most common parameter used as a basis for statistical inference is the mean
 - Proportions = mean of binary variable
 - 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

25

Question 7

.....

- 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

26

Question 8

.....

- In the presence of **censored observations**, it is never **possible to estimate population**
 - A. Mean
 - B. Median
 - C. Geometric mean
 - D. Standard deviation
 - E. 25th and 75th Percentiles
 - F. Probability of exceeding thresholds
 - G. All of the above
 - H. None of the above**

27

Question 9

.....

- In the presence of **censored observations**, it is **usually easiest to estimate population**
 - A. Mean
 - B. Median and other percentiles**
 - C. Geometric mean
 - D. Standard deviation
 - E. Probability of exceeding thresholds**
 - F. All of the above
 - G. None of the above

28

Question 10

- Consider the **correlation between age and salary** among full professors in the Arts aged 50 - 60 years old
 - Holding everything else constant, which of the following changes to study design might lead to a **higher sample correlation**
- A. Increasing sample size
 - B. Increasing ages considered to 40 – 70 years old**
 - C. Increasing the fields considered to include Professional schools
 - D. All of the above

29

More Interpretable Formula for r

$$r \approx \beta \sqrt{\frac{\text{Var}(X)}{\beta^2 \text{Var}(X) + \text{Var}(Y | X = x)}}$$

β = (LS) slope between Y and X

$\text{Var}(X)$ = variance of X in sample

$\text{Var}(Y | X = x)$ = variance of Y in groups that have same value of X

30

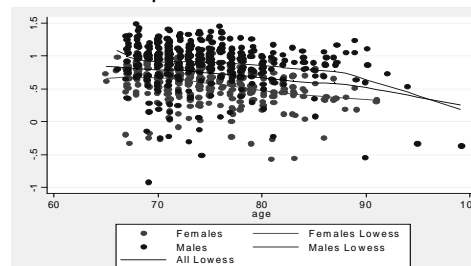
Properties of Correlation

- Correlation tends to increase in absolute value as
 - The absolute value of the slope of the line increases
 - The variance of data decreases within groups that share a common value of X
 - The variance of X increases
 - (Sample size is unimportant in tendencies toward lower or higher correlation)

31

Question 11

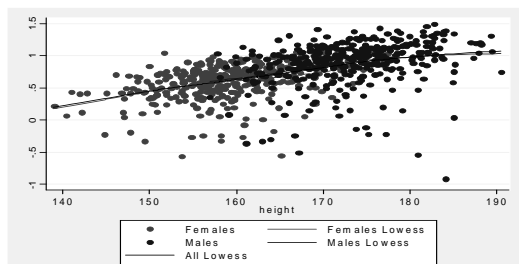
- In the scatterplot of log FEV versus age stratified by sex, we would **expect more extreme correlation** in
 - A. Men $r = -.22$
 - B. Women $r = -.37$**
 - C. Combined sample $r = -.22$



32

Question 12

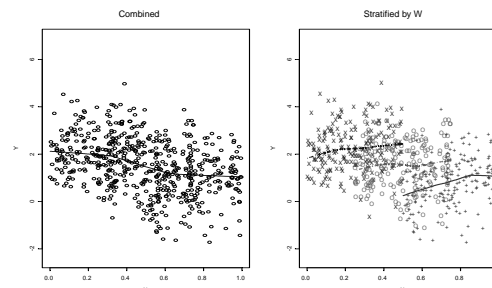
- In the scatterplot of log FEV versus height stratified by sex, we would **expect the higher correlation** in
 - A. Men $r = 0.34$
 - B. Women $r = 0.39$
 - C. Combined sample $r = 0.58$**



33

Question 13

- Plot of Y vs X: **Stratification variable W appears to be**
 - A. An effect modifier
 - B. A confounder**
 - C. A precision variable



34

Confounders

- Estimates of treatment effect the same across strata, AND
 - Confounder is causally associated with Response, AND
 - Confounder associated with POI in the sample
- When analyzing difference of means of continuous data
 - Stratified smooth curves of data are parallel
 - Distribution of POI differs across strata
 - Unadjusted, adjusted analyses give different estimates

35

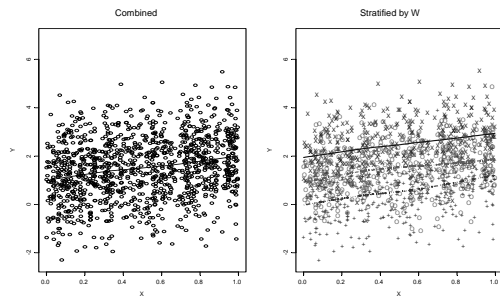
Interpreting Scatterplots

- W associated with Y independent of association between Y and X?
 - Vertical separation between lowess curves for strata
 - Causally?
 - Cannot get this from the data, much less the scatterplot
- W associated with X
 - Preponderance of X values different for the strata

36

Question 14

- Plot of Y vs X: **Stratification variable W appears to be**
 - A. An effect modifier
 - B. A confounder
 - C. A precision variable**



37

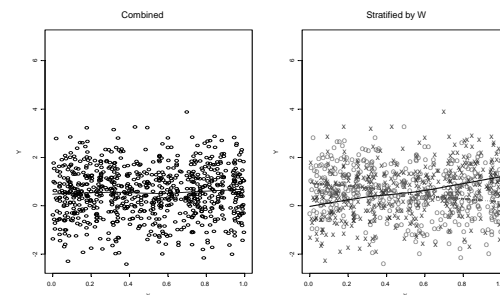
Precision Variables

- Estimates of treatment effect the same across strata, AND
 - Variable is causally associated with Response, AND
 - Variable not associated with POI in the sample
- When analyzing difference of means of continuous data
 - Stratified smooth curves of data are parallel
 - Distribution of POI same across strata
 - Unadjusted, adjusted analyses give similar estimates

38

Question 15

- Plot of Y vs X: **Stratification variable W appears to be**
 - A. An effect modifier**
 - B. A confounder
 - C. A precision variable



39

Effect modifiers

- Estimates of treatment effect differ among the strata
 - When analyzing difference of means of continuous data
 - Stratified smooth curves of data are nonparallel
 - (Graphical techniques difficult in other settings)
- We do not always care about effect modifiers
 - Sometimes we just average across it
 - In those cases, we may care about confounding effects

40