

Biost 518 / Biost 515
Applied Biostatistics II / Biostatistics II

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

First Quiz and Discussion
 January 22, 2014

1

Question 1

.....

- In considering the correlation between DSST and Age in two different studies:
 - Study 1: Least squares slope is -2
 - Study 2: Least squares slope is -1
- (1 pt) If all other aspects stay constant, which of the following statements is most likely to be true.
 - A. Study 1 has a higher sample correlation.
 - B. Study 2 has a higher sample correlation.
 - C. The sample correlations will tend to be equal.

2

Question 2

.....

- In considering the correlation between DSST and Age in two different studies:
 - Study 1: Age is sampled over the range of 70 to 80
 - Study 2: Age is sampled over the range of 60 to 90
- (1 pt) If all other aspects stay constant, which of the following statements is most likely to be true.
 - A. Study 1 has a more extreme sample correlation.
 - B. Study 2 has a more extreme sample correlation.
 - C. The sample correlations will tend to be equal.

3

Question 3

.....

- In considering the correlation between DSST and Age in two different studies:
 - Study 1: The study is restricted to college graduates
 - Study 2: The study takes all education levels
- (1 pt) If all other aspects stay constant, which of the following statements is most likely to be true.
 - A. Study 1 has a more extreme sample correlation.
 - B. Study 2 has a more extreme sample correlation.
 - C. The sample correlations will tend to be equal.

4

Question 4

- In considering correlation between DSST and Age in two different studies:
 - Study 1: The study sample size is 1,000
 - Study 2: The study sample size is 5,000
- (1 pt) If all other aspects stay constant, which of the following statements is most likely to be true.
 - A. Study 1 has a higher sample correlation.
 - B. Study 2 has a higher sample correlation.
 - C. The sample correlations will tend to be equal.

5

Question 5

- A researcher reports disease incidence of 30% in a population unexposed to DDT, and reports that disease incidence is 20% higher in an exposed population.
- (1 pt) What was the observed disease incidence in the exposed population?

6

Question 6

- (80 pts) Name the first four bands from the "British Invasion" who were inducted into the Rock and Roll Hall of Fame.

7

Answers and Discussion

8

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

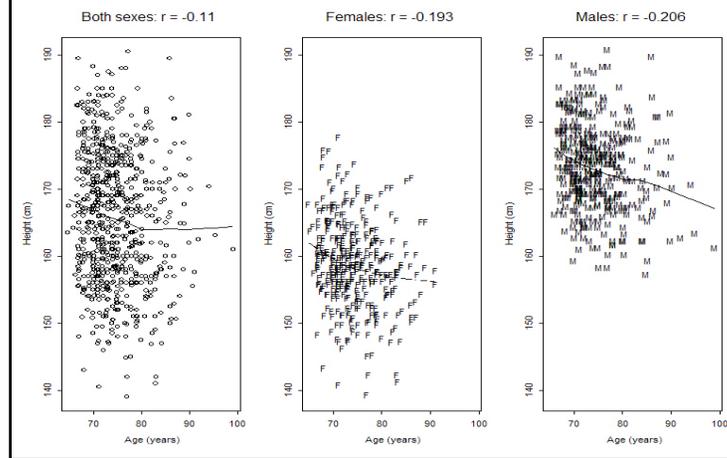
$\text{Var}(X)$ = variance of X in sample

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

Properties of Correlation

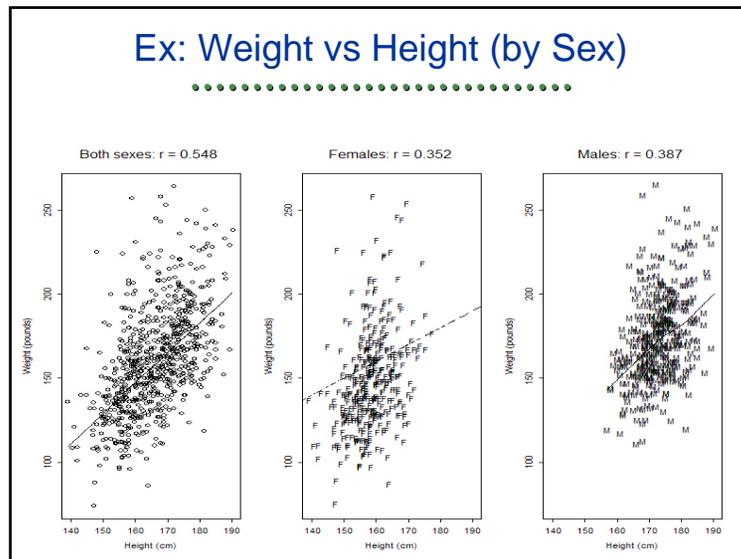
- Correlation tends to increase in absolute value as
 - The absolute value of the slope of the line increases
 - More negative the slope, the more negative the correlation
 - More positive the slope, the more positive the correlation
 - The variance of data decreases within groups that share a common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

Ex: Height vs Age (by Sex)



Ex: Height vs Age (by Sex)

- Correlation between Height and Age
 - Males: $r = -0.206$; Females: $r = -0.193$
 - Combined: $r = -0.110$
- Less extreme r in combined sexes
 - Approximately same slope in each sex and overall
 - Approximately same variance of age in each sex and overall
 - Combined group has higher within group variance of height by age (due to sex effect)



- ### Ex: Weight vs Height (by Sex)
-
- Correlation between Height and Weight
 - Males: $r = .387$; Females: $r = 0.352$
 - Combined: $r = 0.548$
 - More extreme r in combined sexes
 - Approximately same slope in each sex and overall
 - Approximately same within group variance (by height) for each sex and overall
 - Combined group has higher variance of height
- 14

- ### Scientific Relevance of r
-
- It should be noted that
 - the slope between X and Y is of scientific interest
 - the variance of $Y|X=x$ is partly of scientific interest, but it can be affected by restricting sampling to certain values of another variable
 - E.g., $\text{var}(\text{Height} | \text{Age})$ is less in males than when both sexes are included
 - the variance of X is often set by study design
 - This is often not of scientific interest
- 15

- ### Question 1
-
- In considering the correlation between DSST and Age in two different studies:
 - Study 1: Least squares slope is -2
 - Study 2: Least squares slope is -1
 - If all other aspects stay constant, which of the following statements is most likely to be true.
 - A. Study 1 has a higher sample correlation.
 - B. Study 2 has a higher sample correlation.**
 - C. The sample correlations will tend to be equal.
- 16

Question 2

- In considering the correlation between DSST and Age in two different studies:
 - Study 1: Age is sampled over the range of 70 to 80
 - Study 2: Age is sampled over the range of 60 to 90
- If all other aspects stay constant, which of the following statements is most likely to be true.
 - A. Study 1 has a more extreme sample correlation.
 - B. Study 2 has a more extreme sample correlation.**
 - C. The sample correlations will tend to be equal.

17

Question 3

- In considering the correlation between DSST and Age in two different studies:
 - Study 1: The study is restricted to college graduates
 - Study 2: The study takes all education levels
- If all other aspects stay constant, which of the following statements is most likely to be true.
 - A. Study 1 has a more extreme sample correlation.**
 - B. Study 2 has a more extreme sample correlation.
 - C. The sample correlations will tend to be equal.

18

Question 4

- In considering correlation between DSST and Age in two different studies:
 - Study 1: The study sample size is 1,000
 - Study 2: The study sample size is 5,000
- If all other aspects stay constant, which of the following statements is most likely to be true.
 - A. Study 1 has a higher sample correlation.
 - B. Study 2 has a higher sample correlation.
 - C. The sample correlations will tend to be equal.**

19

Question 5

- A researcher reports disease incidence of 30% in a population unexposed to DDT, and reports that disease incidence is 20% higher in an exposed population.
- (1 pt) What was the observed disease incidence in the exposed population?
- **It depends.**
 - **Was the researcher talking about an absolute difference? If so, 50% incidence in exposed.**
 - **Was the researcher talking about a relative difference? If so, 36% incidence.**
- **(Make your use of percentages clear– it may not read well, but if you want nice prose read Dickens instead of statistics.)**

20

Question 6

- (95 pts) Name the first four bands from the “British Invasion” who were inducted into the Rock and Roll Hall of Fame.

- **The Beatles (1988)**
- **The Rolling Stones (1989)**
- **The Who (1990)**
- **The Kinks (1990)**