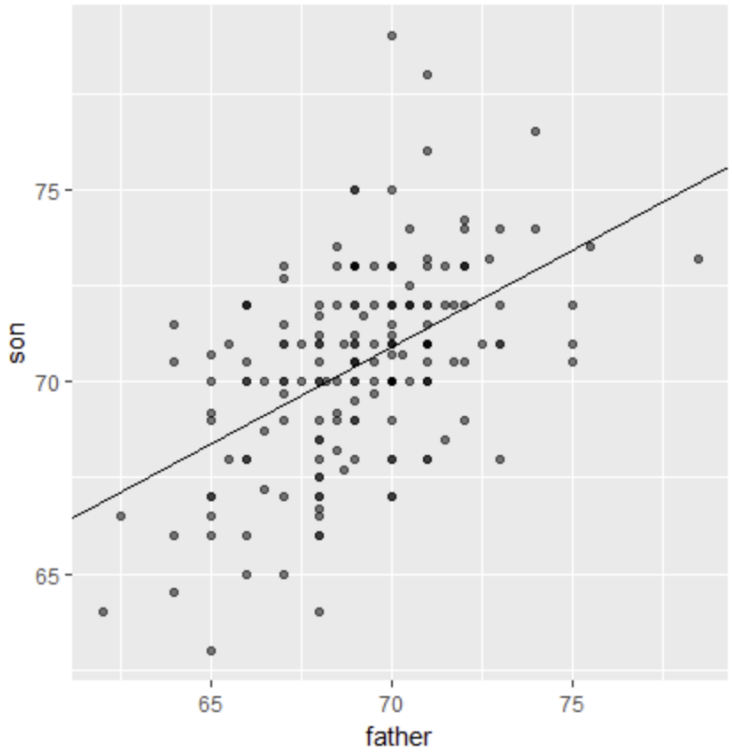
Assessment: Stratification and Variance Explained, Part 1

Question 1

1/1 point (graded)

Look at the figure below.



The slope of the regression line in this figure is equal to what, in words?

- (●) Slope = (correlation coefficient of son and father heights) * (standard deviation of sons' heights / standard deviation of fathers' heights) ✔
- ( ) Slope = (correlation coefficient of son and father heights) * (standard deviation of fathers' heights / standard deviation of sons' heights)
- ( ) Slope = (correlation coefficient of son and father heights) / (standard deviation of sons' heights * standard deviation of fathers' heights)

Slope = (mean height of fathers) - (correlation coefficient of son and father heights * mean height of sons).

Answer
Correct: Correct.

Explanation
From the notes, we know that the slope is $m = \rho\frac{\sigma_y}{\sigma_x}$. Expressed in words, the slope is the correlation coefficient of the son and father heights times the standard deviation of the sons' heights divided by the standard deviation of the fathers' heights.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Question 2

1/1 point (graded)
Why does the regression line simplify to a line with intercept zero and slope ρ when we standardize our x and y variables?
Try the simplification on your own first!

When we standardize variables, both x and y will have a mean of one and a standard deviation of zero. When you substitute this into the formula for the regression line, the terms cancel out until we have the following equation: $y_i = \rho x_i$.

When we standardize variables, both x and y will have a mean of zero and a standard deviation of one. When you substitute this into the formula for the regression line, the terms cancel out until we have the following equation: $y_i = \rho x_i$. ✔

When we standardize variables, both x and y will have a mean of zero and a standard deviation of one. When you substitute this into the formula for the regression line, the terms cancel out until we have the following equation: $y_i = \rho + x_i$.

Explanation
When we standardize variables, they have a mean of 0 and a standard deviation of 1, giving the equation $y_i = \rho x_i$. The equation $y_i = \rho + x_i$ is for a line with an intercept equal to the correlation coefficient, but the regression line simplifies to have an intercept of 0 when we standardize variables.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Question 3

1/1 point (graded)

What is a limitation of calculating conditional means?

Select ALL that apply.

- [x] Each stratum we condition on (e.g., a specific father's height) may not have many data points. ✔
- [x] Because there are limited data points for each stratum, our average values have large standard errors. ✔
- [x] Conditional means are less stable than a regression line. ✔
- [ ] Conditional means are a useful theoretical tool but cannot be calculated.

✔

Answer

Correct:
Correct, this is one disadvantage in using conditional means, but not the only one.
Correct, this is a disadvantage in using conditional means, but not the only one.
Correct, this is one disadvantage in using conditional means, but not the only one.

Explanation

Some limitations of calculating conditional means include: each specific stratum used for conditioning may not have data points, because there are limited data points for each stratum there may be large standard errors on the means, and conditional means are less stable than a regression line. Conditional means can be calculated, so it is not correct that they are only useful as a theoretical tool.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Question 4

1/1 point (graded)

A regression line is the best prediction of Y given we know the value of X when:

- (•) X and Y follow a bivariate normal distribution. ✔

- Both X and Y are normally distributed.
- Both X and Y have been standardized.
- There are at least 25 X-Y pairs.

Answer
Correct: Correct.

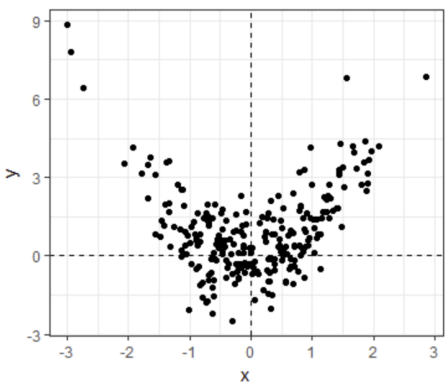
Explanation
In order for the regression line to be the best predictor of Y given a known value of X, X and Y must follow a bivariate normal distribution. It is insufficient for X and Y to each be normally distributed on their own; they must also have a joint bivariate normal distribution.

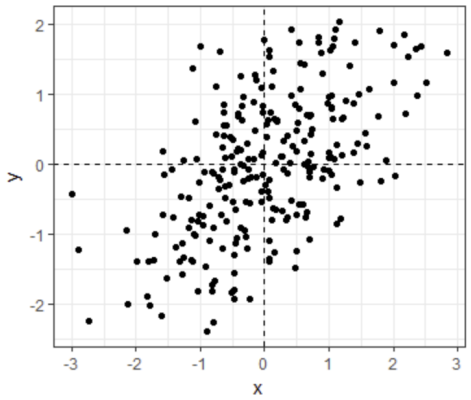
Submit You have used 1 of 2 attempts

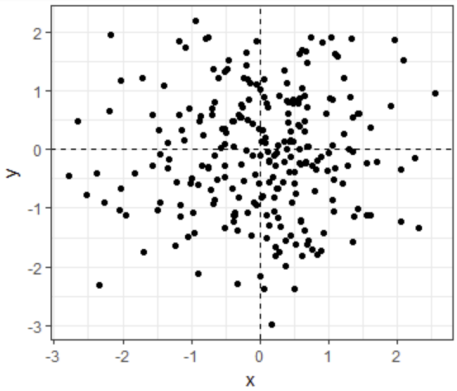
Answers are displayed within the problem

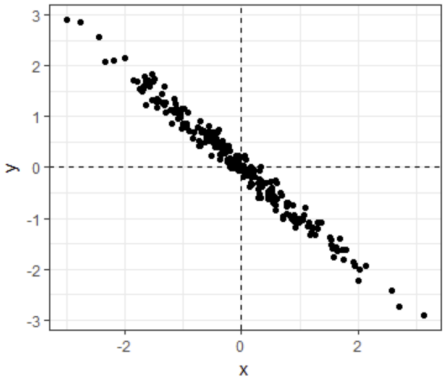
Question 5

1/1 point (graded)
Which one of the following scatterplots depicts an x and y distribution that is NOT well-approximated by the bivariate normal distribution?









Explanation

The v-shaped distribution of points from the first plot means that the x and y variables do not follow a bivariate normal distribution.

When a pair of random variables is approximated by a bivariate normal, the scatter plot looks like an oval (as in the 2nd, 3rd, and 4th plots) - it is okay if the oval is very round (as in the 3rd plot) or long and thin (as in the 4th plot).

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Question 6

1/1 point (graded)

We previously calculated that the correlation coefficient ρ between fathers' and sons' heights is 0.5.

Given this, what percent of the variation in sons' heights is explained by fathers' heights?

- 0%
- 25% ✔
- 50%
- 75%

Answer

Correct:

Correct. When two variables follow a bivariate normal distribution, the variation explained can be calculated as $\rho^2 \times 100$.

Explanation

When two variables follow a bivariate normal distribution, the variation explained can be calculated as $\rho^2 \times 100$. In this case, with a correlation coefficient of 0.5, then the the variance explained is $0.5^2 \times 100 = 25\%$.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Question 7

1/1 point (graded)

Suppose the correlation between father and son's height is 0.5, the standard deviation of fathers' heights is 2 inches, and the standard deviation of sons' heights is 3 inches.

Given a one inch increase in a father's height, what is the predicted change in the son's height?

- 0.333
- 0.5

- 0.667
- 0.75 ✔
- 1
- 1.5

Answer
Correct:
Correct! TThe slope of the regression line is calculated by multiplying the correlation coefficient by the ratio of the standard deviation of son heights and standard deviation of father heights: $\sigma_{son}/\sigma_{father}$.

Explanation
The slope of the regression line is calculated by multiplying the correlation coefficient by the ratio of the standard deviation of son heights and standard deviation of father heights: $\sigma_{son}/\sigma_{father}$. In this case, that is $0.5 \times \frac{3}{2}$.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

© All Rights Reserved