Assessment: Introduction to Linear Models

Question 1

1/1 point (graded)

Why is the number of home runs considered a confounder of the relationship between bases on balls and runs per game?

- ○ Home runs is not a confounder of this relationship.
- ○ Home runs are the primary cause of runs per game.
- ○ The correlation between home runs and runs per game is stronger than the correlation between bases on balls and runs per game.
- ● Players who get more bases on balls also tend to have more home runs; in addition, home runs increase the points per game.

✓

Answer

Correct: Correct.

Explanation

Number of home runs is a confounder of the relationship between bases on balls and runs per game because players who get more bases on balls also tend to have more home runs and home runs also increase the points/runs scored per game.

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Question 2

1/1 point (graded)

As described in the videos, when we stratified our regression lines for runs per game vs. bases on balls by the number of home runs, what happened?

● The slope of runs per game vs. bases on balls within each stratum was reduced because we removed confounding by home runs.

○ The slope of runs per game vs. bases on balls within each stratum was reduced because there were fewer data points.

○ The slope of runs per game vs. bases on balls within each stratum increased after we removed confounding by home runs.

○ The slope of runs per game vs. bases on balls within each stratum stayed about the same as the original slope.

✓

Answer

Correct: Correct.

Explanation

Home runs are a confounder in the runs per game vs. bases on balls regression analysis. When we removed confounding by home runs, the slope of runs per game vs. bases on balls within each stratum decreased.

Submit

You have used 1 of 2 attempts

ℹ Answers are displayed within the problem

Question 3

1/1 point (graded)

We run a linear model for sons’ heights vs. fathers’ heights using the Galton height data, and get the following results:

```
 > lm(son ~ father, data = galton_heights)

Call:
lm(formula = son ~ father, data = galton_heights)

Coefficients:
(Intercept)    father
    35.71       0.50
```

Interpret the numeric coefficient for "father."

○ For every inch we increase the son's height, the predicted father's height increases by 0.5 inches.

● For every inch we increase the father's height, the predicted son's height grows by 0.5 inches.

○ For every inch we increase the father's height, the predicted son's height is 0.5 times greater.

✓

Explanation

The coefficient for "father" gives the predicted increase in son's height for each increase of 1 unit in the father's height. In this case, it means that for every inch we increase the father's height, the son's predicted height increases by 0.5 inches.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 4

1/1 point (graded)

We want the intercept term for our model to be more interpretable, so we run the same model as before but now we subtract the mean of fathers' heights from each individual father's height to create a new variable centered at zero.

```
galton_heights <- galton_heights %>%
    mutate(father_centered=father - mean(father))
```

We run a linear model using this centered fathers' height variable.

```
> lm(son ~ father_centered, data = galton_heights)

Call:
lm(formula = son ~ father_centered, data = galton_heights)

Coefficients:
(Intercept)    father_centered
    70.45           0.50
```

Interpret the numeric coefficient for the intercept.

● The height of a son of a father of average height is 70.45 inches.

○ The height of a son when a father’s height is zero is 70.45 inches.

○ The height of an average father is 70.45 inches.

✓

Explanation

Because the fathers’ heights (the independent variable) have been centered on their mean, the intercept represents the height of the son of a father of average height. In this case, that means that the height of a son of a father of average height is 70.45 inches.

If we had not centered fathers’ heights to its mean, then the intercept would represent the height of a son when a father’s height is zero.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 5

1/1 point (graded)

Suppose we fit a multivariate regression model for expected runs based on BB and HR:

$$E\left[R|BB = x_1, HR = x_2\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Suppose we fix $BB = x_1$. Then we observe a linear relationship between runs and HR with intercept of:

○ β_0

○ $\beta_0 + \beta_2 x_2$

● $\beta_0 + \beta_1 x_1$

○ $\beta_0 + \beta_2 x_1$

✓

Explanation

If x_1 is fixed, then $\beta_1 x_1$ is fixed and acts as the intercept for this regression model. This is the basis of stratificaton.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Question 6

1/1 point (graded)

Which of the following are assumptions for the errors ϵ_i in a linear regression model?

☑ The ϵ_i are independent of each other

☑ The ϵ_i have expected value 0

☑ The variance of ϵ_i is a constant

✓

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem