

Assessment: Tibbles, do, and broom, part 2

We have investigated the relationship between fathers' heights and sons' heights. But what about other parent-child relationships? Does one parent's height have a stronger association with child height? How does the child's gender affect this relationship in heights? Are any differences that we observe statistically significant?

The **galton** dataset is a sample of one male and one female child from each family in the **GaltonFamilies** dataset. The **pair** column denotes whether the pair is father and daughter, father and son, mother and daughter, or mother and son.

Create the **galton** dataset using the code below:

```
library(tidyverse)
library(HistData)
data("GaltonFamilies")
set.seed(1) # if you are using R 3.5 or earlier
set.seed(1, sample.kind = "Rounding") # if you are using R 3.6 or later
galton <- GaltonFamilies %>%
  group_by(family, gender) %>%
  sample_n(1) %>%
  ungroup() %>%
  gather(parent, parentHeight, father:mother) %>%
  mutate(child = ifelse(gender == "female", "daughter", "son")) %>%
  unite(pair, c("parent", "child"))

galton
```

Question 8

2.0/2.0 points (graded)

Group by `pair` and summarize the number of observations in each group.

How many father-daughter pairs are in the dataset?

176

✓ Answer: 176

176

How many mother-son pairs are in the dataset?

179

✓ Answer: 179

179

Explanation

The following code will give the number of observations in each of the four groups:

```
galton %>%  
  group_by(pair) %>%  
  summarize(n = n())
```

Submit

You have used 1 of 10 attempts

i Answers are displayed within the problem

Question 9

2.0/2.0 points (graded)

Calculate the correlation coefficients for fathers and daughters, fathers and sons, mothers and daughters and mothers and sons.

Which pair has the **strongest** correlation in heights?

☐ fathers and daughters

☒ fathers and sons

☐ mothers and daughters

☐ mothers and sons



Explanation

The following code will give the maximum correlation:

```
galton %>%
  group_by(pair) %>%
  summarize(cor = cor(parentHeight, childHeight)) %>%
  filter(cor == max(cor))
```

Which pair has the **weakest** correlation in heights?

☐ fathers and daughters

☐ fathers and sons

☐ mothers and daughters

☒ mothers and sons



Explanation

The following code will give the minimum correlation:

```
galton %>%
  group_by(pair) %>%
  summarize(cor = cor(parentHeight, childHeight)) %>%
  filter(cor == min(cor))
```

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 10 has two parts. The information here applies to both parts.

Use **lm** and the **broom** package to fit regression lines for each parent-child pair type. Compute the least squares estimates, standard errors, confidence intervals and p-values for the **parentHeight** coefficient for each pair.

Question 10a

2.0/2.0 points (graded)

What is the estimate of the father-daughter coefficient?

✓ Answer: 0.345

Explanation

The estimate can be calculated using the following code:

```
library(broom)
galton %>%
  group_by(pair) %>%
  do(tidy(lm(childHeight ~ parentHeight, data = .), conf.int = TRUE)) %>%
  filter(term == "parentHeight", pair == "father_daughter") %>%
  pull(estimate)
```

For every 1-inch increase in mother's height, how many inches does the typical son's height increase?
Give your answer as a number with no units.

✓ Answer: 0.381

Explanation

The following code will give the decrease in height:

```
galton %>%
  group_by(pair) %>%
  do(tidy(lm(childHeight ~ parentHeight, data = .), conf.int = TRUE)) %>%
  filter(term == "parentHeight", pair == "mother_son") %>%
  pull(estimate)
```

You have used 1 of 10 attempts

i Answers are displayed within the problem

Question 10b

2.0/2.0 points (graded)

Which sets of parent-child heights are significantly correlated at a p-value cut off of .05?

Select ALL that apply.

☒ father-daughter

☒ father-son

☒ mother-daughter

☒ mother-son



Explanation

All of the parent-child heights are correlated with a p-value of <0.05 .

Which of the following statements are true?

Select ALL that apply.

☒ All of the confidence intervals overlap each other.

☐ At least one confidence interval covers zero.

☒ The confidence intervals involving mothers' heights are larger than the confidence intervals involving fathers' heights.

☐ The confidence intervals involving daughters' heights are larger than the confidence intervals involving sons' heights.

☒ The data are consistent with inheritance of height being independent of the child's gender.

☒ The data are consistent with inheritance of height being independent of the parent's gender.



Answer

Correct:

Correct. The confidence intervals all overlap.

Correct. The `std.error` values are higher for mothers than fathers, resulting in larger confidence intervals.

Correct. The confidence intervals overlap.

Correct. The confidence intervals overlap.

Explanation

The following code can be used to answer both questions:

```
galton %>%
  group_by(pair) %>%
  do(tidy(lm(childHeight ~ parentHeight, data = .), conf.int = TRUE)) %>%
  filter(term == "parentHeight" & p.value < .05)
```

All four of the confidence intervals overlap. The confidence intervals for mothers' heights are larger than those for fathers' heights, as observed from the standard errors. Because the confidence intervals overlap, the data are consistent with inheritance of height being independent of the child's or the parent's gender.

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

© All Rights Reserved