Assessment: Correlation

Question 1

1/1 point (graded)

While studying heredity, Francis Galton developed what important statistical concept?

- ○ Standard deviation
- ○ Normal distribution
- ◉ Correlation ✔
- ○ Probability

Explanation

Francis Galton developed the concept of correlation while study heredity.

Submit You have used 1 of 2 attempts

ℹ Answers are displayed within the problem

Question 2

1/1 point (graded)

The correlation coefficient is a summary of what?

- ◉ The trend between two variables ✔
- ○ The dispersion of a variable

- The central tendency of a variable
- The distribution of a variable

Answer
Correct: Correct.

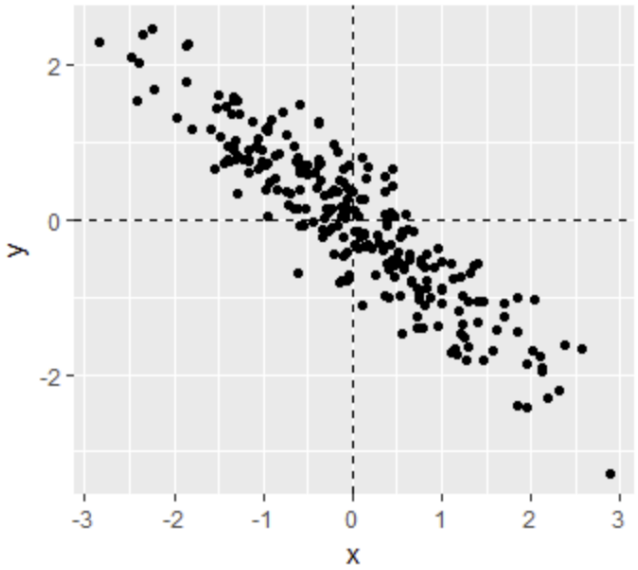
Explanation
The correlation coefficient is a summary of the trend between two variables.
The standard deviation describes the dispersion of a variable; the mean is a description of a variable's central tendency; the distribution of a variable (e.g., normal, log-normal) describes the possible values of your data and the probability of them occurring.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Question 3

1/1 point (graded)
Below is a scatter plot showing the relationship between two variables, x and y.



From this figure, the correlation between x and y appears to be about:

- -0.9 ✔
- -0.2

- 0.9
- 2

Answer
Correct:
Correct. The variables x and y have a strong negative relationship with each other; as x increases, y decreases.

Explanation
The correlation appears to be about -0.9. The variables x and y have a strong negative relationship with each other; as x increases, y decreases.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Question 4

1/1 point (graded)
Instead of running a Monte Carlo simulation with a sample size of 25 from the 179 father-son pairs described in the videos, we now run our simulation with a sample size of 50.

Would you expect the **mean** of our sample correlation to increase, decrease, or stay approximately the same?

- Increase
- Decrease
- Stay approximately the same ✔

Explanation
Because the expected value of the sample correlation is the population correlation, it should stay approximately the same even if the sample size is increased.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 5

1/1 point (graded)

Instead of running a Monte Carlo simulation with a sample size of 25 from the 179 father-son pairs described in the videos, we now run our simulation with a sample size of 50.

Would you expect the **standard deviation** of our sample correlation to increase, decrease, or stay approximately the same?

- Increase
- Decrease ✔
- Stay approximately the same

Explanation

As the sample size N increases, the standard deviation of the sample correlation should decrease.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 6

1/1 point (graded)

If X and Y are completely independent, what do you expect the value of the correlation coefficient to be?

- -1
- -0.5
- 0 ✔
- 0.5
- 1

Not enough information to answer the question

Answer
Correct: Correct. If X and Y are independent, then their correlation coefficient is 0.

Explanation
Variables that are independent of each other have a correlation coefficient of 0.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Question 7

1/1 point (graded)
Load the **Lahman** library. Filter the `Teams` data frame to include years from 1961 to 2001.

What is the correlation coefficient between number of runs per game and number of at bats per game?

0.658 ✔ **Answer:** 0.658

0.658

Explanation
The correlation coefficient can be calculated using the following code:

```
library(Lahman)
Teams_small <- Teams %>% filter(yearID %in% 1961:2001)
cor(Teams_small$R/Teams_small$G, Teams_small$AB/Teams_small$G)
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Question 8

1/1 point (graded)
Use the filtered `Teams` data frame from Question 7.

What is the correlation coefficient between win rate (number of wins per game) and number of errors per game?

-0.339 ✔ **Answer:** -0.340

-0.339

Explanation
The correlation coefficient can be calculated using the following code:

```
cor(Teams_small$W/Teams_small$G, Teams_small$E/Teams_small$G)
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Question 9

1/1 point (graded)
Use the filtered `Teams` data frame from Question 7.

What is the correlation coefficient between doubles (`X2B`) per game and triples (`X3B`) per game?

-0.011 ✔ **Answer:** -0.0116

-0.011

Explanation
The correlation coefficient can be calculated using the following code:

```
cor(Teams_small$X2B/Teams_small$G, Teams_small$X3B/Teams_small$G)
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem