|  |  |  |  |
| --- | --- | --- | --- |
| Participant | X | Y |  |
| 1 | 9 | 2 | 2.500 |
| 2 | 7 | 5 | 3.786 |
| 3 | 5 | 4 | 5.071 |
| 4 | 1 | 7 | 7.643 |
| 5 | 2 | 8 | 7.000 |
| Mean | 4.8 | 5.2 |  |
| SD | 3.347 | 2.387 |  |

For 5 participants, their scores on a predictor (X) and dependent variable (Y) were registered. The correlation between X and Y = -0.901. Use these values for the following questions:

1. Calculate the slope of the corresponding regression model, in which X predicts Y
2. Calculate the intercept of the corresponding regression model
3. Calculate the two-sided p-value of the slope. It’s true that you don’t have the standard error of the slope, so you’ll have to use a different t-formula that will get you the same answer in this situation.

DF = n – 2 = 5 – 2 = 3

Looking this t-value up in table D gives us

One-sided p: 0.02 > p > 0.01  
Two-sided p: 0.04 > p > 0.02

1. Calculate the 95% confidence interval around the correlation

Critical Z-value for 95% CI = 1.96

For the actual confidence interval, we need to transform back

1. Calculate the standardized slope of the corresponding regression model, in which X predicts Y

Because it is simple linear regression, the standardized slope is equal to the correlation, so b\* (sometimes also named ß) = -0.901

1. Calculate R, the correlation between the model and the dependent variable

Because it is simple linear regression, the correlation between model and dependent variable is equal to the absolute value of the correlation between X and Y, so it would be 0.901

1. Calculate R², the multiple correlation coefficient or coefficient of determination

R = 0.901, so R² = 0.901² = 0.812

1. Calculate R²W, the Adjusted R²
2. Fill in the entire ANOVA table below. You are given all observed values (Y), predicted values (y^), and the mean of y in the table. This should allow you to be able to calculate all sums of squares. N and number of predictors are also known, which should give all degrees of freedom.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Effect** | **SS** | **DF** | **MS** | **F** |
| Model | 18.514 | 1 | 18.514 | 12.965 |
| Error (Residual) | 4.284 | 3 | 1.428 |  |
| Total | 22.8 | 4 | 5.7 |  |

SSM = (2.500 – 5.2)² + (3.786 – 5.2)² + (5.071 – 5.2)² + (7.643 – 5.2)² + (7.000 – 5.2)² = 18.514  
SSE = (2 – 2.500)² + (5 – 3.786)² + (4 – 5.071)² + (7 – 7.643)² + (8 – 7.000)² = 4.284  
SST = (2 – 5.2)² + (5 – 5.2)² + (4 – 5.2)² + (7 – 5.2)² + (8 – 5.2)² = 22.800 (unrounded)

DFM = p = 1  
DFE = n – p – 1 = 5 – 1 – 1 = 3  
DFT = n – 1 = 5 – 1 = 4

MSM = 18.514 / 1 = 18.514  
MSE = 4.284 / 3 = 1.428   
MST = 22.8 / 4 = 5.7

F = MSM / MSE = 18.514 / 1.428 = 12.965, which is about the same as (-3.6)² from question 3

1. Why is the F-value in the ANOVA table in question 9 equal to the square of the t-value that you calculated for question 3? Is this always true for simple linear regression?

It is the same because the F-distribution is a squared t-distribution when the first degree of freedom = 1. So the square of any t-value will correspond to the F-value. Since simple linear regression always has DFM = 1, this is always true.

**A multiple regression was computed where “Speed” and “Hand-Eye coordination” were used to predict “Gaming skill”. N = 50. Use the following correlation table for questions 11 – 20.**

|  |  |  |  |
| --- | --- | --- | --- |
| Correlation | Speed | HandEye | Skill |
| Speed | 1 | 0.2 | 0.4 |
| HandEye | 0.2 | 1 | 0.6 |
| Skill | 0.4 | 0.6 | 1 |

1. Calculate the standardized slope (b\*) for Hand eye coordination
2. Calculate the partial correlation (pr) for Hand eye coordination
3. Calculate the semi-partial explained variance (sr²) for Speed
4. Calculate the R² for the model as a whole

R² = r²(HandEye) + sr²Speed = 0.6² + 0.082 = 0.442

1. Calculate R²Stein for the model
2. Calculate the semi-partial correlation for Hand eye coordination

**Additional information**: The standard deviation for Speed = 3, The standard deviation for Hand Eye coordination = 2, The standard deviation for Skill = 6

1. Calculate the unstandardized slope (b) for Hand eye coordination. Tip: the value you calculated for question 11 can help you here

b\* = 0.542

1. Assume that the standard error of the slope you just calculated is 1. Calculate the two-sided p-value of this slope

DF = DFE = n – p – 1 = 50 – 2 – 1 = 47. Round down to first available DF = 40

Looking up 1.626 for DF = 40 gives:

One-sided p-value: 0.10 > p > 0.05  
Two-sided p-value: 0.20 > p > 0.10

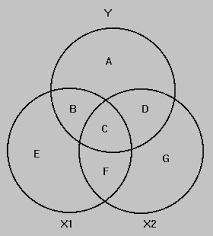
1. Calculate the 95% confidence interval for this slope

Critical t-value (DFE = 47, round down to 40) for 95% = 2.021

1. Now that you have so much information on the relationship between Hand-eye coordination and gaming skill, how would you interpret this slope? What does the slope itself mean? Is it significant? What is the effect size? How would you describe the effect size?

From the p-value and confidence interval we could see it is not statistically significant at the usual level of α = 0.05. The slope indicates that when your hand-eye coordination improves by 1, your gaming skill will increase by 1.626, given that speed is held constant. The effect size was high (sr² = 0.531² = 0.282), but due to the relatively large uncertainty in the standard error, we are not sure if this effect can be generalized to the population. It would be an interesting effect to research further with a larger sample size, and maybe looking for a method to reduce uncertainty, to further determine if this relationship also exists in the population. Due to the large practical significance (effect size), it has potential.

**Now, some ballentine puzzles. Since there are different ballentines out there, I will use the picture below for reference. Note that this ballentine shows explained variance, so for example, if you calculate sr with the formula of the formula sheet, never forget to square it.**



|  |  |  |  |
| --- | --- | --- | --- |
| Correlation | X1 | X2 | Y |
| X1 | 1 | 0.2 | -0.4 |
| X2 | 0.2 | 1 | 0.4 |
| Y | -0.4 | 0.4 | 1 |

1. What is the explained variance of X1 in Y, without partialing out X2 from either X1 or Y?

This refers to the zero-order correlation between X1 and Y, which is -0.4. Explained variance = -0.4² = 0.16, or 16%

1. What is the explained variance of X2 in Y, when X1 is partialed out of both X1 and Y?

This refers to the partial explained variance:

1. Calculate the area “C”

There are multiple ways to do this, but one way to do it is to take the zero-order explained variance between X1 (or X2) and Y, and then subtract the semi-partial explained variance of the same predictor:

Zero-order variance explained by X1 = 0.160

0.160 – 0.240 = –0.08 = C. Yes, it’s negative, and that is perfectly possible.

1. Calculate the area “C + F”

This is the overlap between X1 and X2, also known as the level of multicollinearity:

R²j = 0.2² = 0.04

1. Calculate the area “A”. What does this stand for?

There are multiple ways to do this. We know the whole circle of Y has an area of 1. A + B + C + D = 1 R² = B + C + D, so 1 – R² = A. This already tells us that A = proportion unexplained variance.

We already have “B” thanks to calculating the sr² in question 23, and we can find “C + D” by squaring the correlation between X2 and Y.

R² = 0.240 + 0.4² = 0.400

1 – R² = 0.600

1. What is the difference between partial and semi-partial correlation of X1?
2. Partial correlation removes overlap with X2 from both X1 and Y. Semi-partial correlation only removes overlap with Y
3. **Partial correlation removes overlap with X2 from both X1 and Y. Semi-partial correlation only removes overlap with X1**
4. Partial correlation removes overlap with X2 from X1. Semi-partial correlation only removes overlap with Y
5. Partial correlation removes overlap with X2 from Y. Semi-partial correlation only removes overlap with X1
6. Which of the following purely corrects R² for inflation due to number of predictors
7. Stein correction
8. **Wherry correction**
9. Both
10. Neither
11. Which value is the most appropriate for comparing unique contributions from predictors in explaining the variance of the dependent variable?
12. **Semi-partial explained variance**
13. Partial explained variance
14. R²
15. Adjusted R²
16. Which type of interval has the largest width?
17. Confidence interval for the slope
18. Confidence interval for mean response
19. **Prediction interval**
20. That differs for different samples
21. For simple linear regression, what is always true?
22. pr² > sr²
23. pr² < sr²
24. **pr² = sr²**
25. That differs for different samples