Chart, scatter chart

Description automatically generated

Table

Description automatically generatedTable

Description automatically generated

We expect an inverse relationship between ACEs and test scores.

The average test score among the N = 10 students in the 5th grade class is 100.4 (s.d. = 8.49) and the average number of ACEs is 3.90 (s.d. = 3.07). As expected, there is a strong relationship between test scores and ACEs that is negative and statistically significant (*r* = -.902, *p* < .001). This means that as the number of ACEs increases, test scores decrease (see Figure 1).

Let’s run a simple regression that predicts test score from ACEs score.

Analyze 🡪 Regression 🡪 Linear

Graphical user interface, application, Word

Description automatically generated

Table

Description automatically generated

**R** is the correlation between the independent variable(s) and the dependent variable. Note R x R = R square

**Adjusted R square** is an adjustment due to sampling variation that provides a more accurate estimate of R square for the population. The adjustment is made because as more independent variables are added to the model, each one will explain some of the variation in the dependent variable “by chance.”

**R square** tells us how much of the variance of the dependent variable can be explained by the independent variable(s). It compares the model with the independent variables to a model without independent variables, known as the null model. In this case, 81.3% of the variation in test scores is explained by differences in ACEs. This is extremely high!

**Standard error of the estimate** is the standard deviation of the error term. It is the square of the root of the Mean Square Residual or Error.

Table

Description automatically generated

**Residual** displays information about the variation that is not accounted for by the model.

**Regression** displays the information about the variation accounted for by the model.

Note: .813

And that 1 - .813 = .187

This means that 81.3% of the variation in test scores are explained by differences in ACEs and 18.7% is not explained. Note: this is the R square value from above.

A model with a large regression Sum of Squares in comparison to residual Sum of Squares indicates that the model accounts for most of the variation in the dependent variable (as here). Very high residual SS indicates the model fails to explain the variation in the dependent variable. This means that there are variables that are not part of the model that are important for explaining the variation in the dependent variable.

The mean square is the SS / df. The ratio of regression to residual Mean Square is the F-statistic, which has an accompanying significance level. If the significance of the F statistic is small (i.e., < .05) then the independent variables do a good job in explaining the dependent variable. If the significance is large (i.e., > .05) then it does not.

Table

Description automatically generated

The constant represents the Y-intercept or . Here, this is the value of the test score when the number of adverse child experiences is 0. This is also the height of the regression line when it crosses the Y axis. **Note:** in many cases the value of Y when X = 0 is meaningless (see the example in lecture that predicts home value from square footage, there is no such home that is 0 square feet).

**Unstandardized coefficients**: These are the values for the regression equation for predicting the dependent variable from the independent variable(s). These are unstandardized beta coefficients because they are measured in natural units, and therefore cannot be compared to one another to determine which is more influential.

**The standardized coefficients** (also confusingly referred to as betas) are an attempt to make the regression coefficients more comparable. **Note**: these will only be important when the model has more than one independent variable.

The t statistics can help you determine the relative importance of each variable in the model. The t-statistic is calculated by divided the variable’s unstandardized coefficient by its standard error. As a guide regarding useful predictors, look for t values well below -1.96 or above +1.96. As you can see, the ACEs variable has a t statistic = -5.904 (p < .001). This means that there is a statistically significant inverse relationship between ACEs and test score. The null hypothesis being tested is:

And the alternative is:

In words, w**hat is the null hypothesis being tested here?**

The significance column indicates whether or not a variable is a significant predictor of the dependent variable. The p-value (significance) is the probability that your sample could have been drawn from the population(s) being tested (or that a more improbable sample could be drawn) given the assumption that the null hypothesis is true. A p-value of .05, for example, indicates that you would have only a 5% chance of drawing the sample being tested if the null hypothesis was actually true. The standard in the social sciences is a p-value < .05.

Regression equation:

Test score = 100.120 – 2.492ACEs

Interpretation: (intercept ) The value of test score when there are no ACEs is 100.120. (slope For every one unit increase in ACEs score (in this case, each additional ACE), test score is reduced by 2.492 points. Note carefully: this interpretation is based on the measurement of the variables. Every “1 unit” increase in ACEs is equal to 1 additional ACE because of how the variables are measured.

|  |  |
| --- | --- |
| Predicted Test Scores for ACE values | |
| ACE | Predicted Score |
| 0 | 100.120 |
| 1 | 97.63 |
| 2 | 95.14 |
| 3 | 92.64 |
| 4 | ? |
| 5 | ? |
| 6 | ? |

Note that each additional ACE reduces test score by the same about, i.e. 2.492 points. This is the meaning of ‘linear’ in the parameters. Why might this assumption be unrealistic in this case?