**Linear Regression**

The relationship between two continuous variables can be explained with a straight line in the simplest way. In regression analysis, the relationship between a *dependent variable* and an *independent variable* can be explained with a simple linear regression model.

Regression requirements:

Number of variables

Measurement scale

linearity

### **Regression Equation**

Suppose Y is a dependent variable, and X is an independent variable.

𝑌=𝛽0+𝛽1𝑋Y=β0+β1X

We read this as “Y equals 𝛽1β1 times X, plus a constant 𝛽0β0.”

𝛽0β0*: constant*  
𝛽1β1*: regression coefficient*  
*X: independent variable*  
*Y: dependent variable*

In this formula,

* Y is the value of the dependent variable that we are trying to predict.
* 𝛽0β0 is the Y -intercept, the value of Y when X is equal to 0.
* 𝛽1β1 is the slope, for each increase in unit on the X -axis is a corresponding increase or decrease on the Y -axis.

Basically, we call Y as dependent variable, X as independent variable, 𝛽0β0 and 𝛽1β1 as coefficients.

The values must be calculated on the basis of means and variances, specifically sums of squares. Therefore, the three formulas we are working with to draw a regression line are as follows:

1. 𝑌=𝛽0+𝛽1𝑋Y=β0+β1X (formula for the predicted value of Y).
2. 𝛽1=𝑆𝑃𝑆𝑆𝑥=∑(𝑥𝑖−𝑥¯)(𝑦𝑖−𝑦¯)∑(𝑥𝑖−𝑥¯)2β1=SPSSx=∑(xi−x¯)(yi−y¯)∑(xi−x¯)2 (formula for the regression coefficient, or slope).
3. 𝛽0=𝑦¯−𝛽1𝑥¯β0=y¯−β1x¯ (formula for the Y -intercept).

𝑆𝑃SP*: Sum of products*

𝑆𝑆𝑥SSx*: Sum of squares for the independent variable*

### **Regression Example**

As television viewing hours ( X ) increase, GPA ( Y ) decreases. To test our hypothesis, we sample 10 students and gather the following data.

**Table

Description automatically generated**

Before we can begin to solve for our predicted values of Y (*Y*=*β*0+*β*1*X*), we must first determine the regression coefficient ( *β*1  ) and the Y-intercept ( *β*0 ). To determine these values, we must first calculate the sum of products (SP), the sum of squares for the independent variable (SSx), the mean for the dependent variable (Y), and the mean for the independent variable (X).

**Table

Description automatically generated**

**Text

Description automatically generated**

We can now use the values of the regression coefficient and the Y-intercept to calculate predicted values of y (Y) using given values of X to draw two points through which the fit line will pass. It is always a good idea to draw the fit line using the y -intercept as one point and the second point on the basis of using a large value of X in the formula for a line (*Y*=*β*0+*β*1*X*). This will produce two points that are far apart on the scatterplot, making the line easier to draw. First, we draw one point on the Y-axis at a GPA of 3.79 (the Y-intercept). Second, using an X value of 5, our highest value of X, we calculate the second point using the formula for a line as follows:

Text

Description automatically generated with medium confidence

This tells us that a student with an X value of 5 will have a Y value of 2.31.

### **Explaining Variance in the Dependent Variable (R-squared)**

The previous example we first calculate Pearson's r as follows,

**Text

Description automatically generated with medium confidence**

R-squared is nothing more than the square of Pearson’s r, but;it offers us unique insight on the relationship between the two variables in our analysis.

*R*2=(−.86)(−.86)=.74

The *R*2 value tells us what percent of the total variance in the dependent variable is explained by the independent variable.

**https://www.youtube.com/watch?v=w2FKXOa0HGA&t=1s**