OLS Regression – Bivariate

Making predictions about the value of one variable based on the value of another

When correlations are strong and significant, it is possible to make predictions about the value of one variable based on the value of another variable. The regression line is used to determine 1) the value of the dependent variable when the independent variable value is 0 and 2) the way in which the dependent variable varies as the independent variable varies. These two values – intercept and slope – are what make predictions possible using the regression equation.

Requirements:

- DV = Interval/ratio
- IV = interval/ratio or dummy variables (0/1)

Components of the Regression Line

a (intercept, constant)

The "a" is the value of the dependent variable when the value of the independent variable is 0. This is where the line would intersect with the Y axis. In essence, this is the starting point for the line and later predictions.

b (or slope)

The b coefficient tells us how much the dependent variable increases or decreases with each unit increase in the independent variable. One interprets the b coefficient by using 1) the sign and 2) the actual number. The sign indicates whether the value of the dependent variable increases (+) or decreases (-). The actual number is the amount (in units of the dependent variable) of that increase and decrease.

NOTE ABOUT DUMMY VARIABLES: If your independent variable is a dummy variable with values of 0 or 1, the b coefficient is telling you the value of the increase or decrease of the dependent variable based on whether someone has the "condition" included in the model compared to those who are not included. The regression line is comparing the means for the dependent variable between members of the two groups

For instance, you are looking at the relationship between gender (men and women) and income (in dollars). You included women in your regression and left men out so they can serve as a comparison group. The b coefficient is telling that if someone is a woman, the dependent variable will increase or decrease by the value of the b.

Determining significance

In OLS regressions, we determine whether the relationship between the dependent and independent variable is significant by looking at either the test statistic or the obtained probability (p value) associated with the b coefficient. It is the same process as in other tests.

Significance in this case can be understood as referring to the usefulness of the IV in making a prediction for the DV. If the b coefficient is significant, it is a useful predictor. If the b coefficient is not significant, it is not a useful predictor.

NOTE: Standard error (SE) is the standard deviation of the data points from the line. Smaller SEs lead to more accurate predictions because there is less variation between the data points and the line.

Test statistic and critical value

One compares the obtained test statistic with the critical value associated with the test (e.g. .05 two-tailed test has a critical value of 1.96). If the test statistic is not provided, one can calculate it with the following formula:

$$T_{n-k-1} = \frac{b}{SE_b}$$

n-k-1 = degrees of freedom where k is the number of independent variables
SE_b = Standard error for the b coefficient

Test statistic > critical value = significant, reject null
Test statistic < critical value = not significant, accept null

Probability and significance levels

When using SPSS, one can compare the obtained probability of the b coefficient with the pre-determined significance level:

Obtained probability < significance level = significant, reject null
Obtained probability > significance level = not significant, accept null

Making statements about the relationship between the IV and the DV

There are formulaic ways to discuss what you find in your regression analysis. You only interpret those coefficients that are significant.

Interval/ratio variables

Positive b
For every unit increase in [x/independent variable], it is predicted that [y/dependent variable] will increase by [b coefficient]

Examples:
- For every additional hour of counseling, an adolescent is predicted to increase her self-esteem score by 2.4.
- As the number of children increases in the house, the number of hours watching TV is predicted to increase by 4.3.

Negative b
For every unit increase in [x/independent variable], it is predicted that [y/dependent variable] will decrease by [b coefficient].

Examples:
- For every additional child, a mother's enjoyment of parenthood is predicted to decrease by .3.
- For every additional hour worked, amount of time spent with the family is predicted to decrease by 3.2 hours.

Dummy variables

Positive b
[The group(s) that is included in the model] will have [y/dependent] value that is [b coefficient] higher than [the group(s) not included in the model].

Examples:
- Women are predicted to earn $400 more than men.
- Those who are married or single are predicted to have a personal satisfaction score that is 3.45 points higher than those who are separated, divorced or widowed.

Negative b
[The group(s) that is included in the model] will have [y/dependent] value that is [b coefficient] lower than [the group(s) not included in the model].

Examples:

- Those who have graduated from high school are predicted to report .35 fewer health problems than those who did not graduate from high school.
- Adults are predicted to laugh 5 times less than children each day.

Making predictions about particular individuals based on their characteristics

To make predictions, one uses the regression equation by plugging in the a and b values and the appropriate value for the independent variable (X). The value for X comes from the person or client for whom you are interested in making a prediction. The resulting Y′ is the predicted value of the dependent variable for that particular individual.

REGRESSION EQUATION

$$Y' = a + bX$$

where Y' = Y value predicted from a specific X value
a = point where the line intersects the y-axis (the ***y*-intercept**)
b = **slope** of the line (the amount Y increases for each unit of increase in X)
X = X value used to predict Y'

Statements about the Y′ can be said in the following ways:

- An adolescent who smokes pot 2 times a week is predicted to be truant 3 times in the month.
- A parent who criticizes his or her child 5 times each day is predicted to hug that child 2 times each week.
- Those who own cars are predicted to walk 2 miles less than those who do not own cars.