

Interpreting Regression Results

Will Doyle

Components of a Regression Table

- Coefficients
- Standard Errors
- T statistics
- R^2
- N
- F (sometimes)

Interpreting Coefficients

- Coefficients are always expressed in terms of dependent variable
- Sign indicates direction of relationship between independent and dependent variable
- Number indicates the change in the dependent variable as a result of a one unit change in the independent variable

Interpreting Standard Errors

- The standard error of a coefficient is very similar to the standard error of the mean.
- The coefficient \pm a t value times the standard error will give you a 95% confidence interval for the coefficient
- The coefficient divided by the standard error will give you an obtained t-statistic

Interpreting t statistics

- T statistics have exactly the same interpretation as they did when comparing means
- Degrees of freedom are equal to the number of units minus the number of regressors: $n-k$

R^2

- R^2 is a measure of the amount of variance in the dependent variable accounted for by the independent variables in the model
- It is used to compare models that “nest” within one another
- It cannot be used to compare completely different models.

R^2

- Adjusted R^2 is a measure of the variance accounted for, with a penalty for more independent variables
- As we include more variables, unadjusted will always go up—we need to account for the fact that we have lots more stuff in the model.
- Many times, this is a good complement to the standard R^2

Calculation:

$$\text{Adj } R^2 = R^2 - \frac{k-1}{n-k} (1 - R^2)$$

F Statistic

- The F statistic in the context of regression is the ratio of explained sum of squares to residual sum of squares.
- Explained sum of squares: Sum of square differences between predictions and mean
- Residual sum of squares: Sum of square differences between prediction and actual value of case

The F Statistic

- The F statistic tests whether the combined coefficients on all covariates is 0
- Has the same interpretation as the F statistic for ANOVA
- DF, numerator= $k-1$
- DF, denominator= $n-k$

Regression on Dummy Variables

- Regression is known as a conditional mean function.
- Coefficients on dummy variables show how much higher or lower that group is than the other group, given other controls.
- Interpretation is always with reference to the comparison group.

Regression with Categorical Variables

- Categorical variables should NEVER be entered as their coded values.
- Categorical variables should ALWAYS be entered as a series of dummy variables, with one group omitted.
- Why? Otherwise, the groups perfectly predict each other, which causes regression to break down.