



Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

Introduction to Interpreting



Regression output in Stata

Regression Output in Stata

Objective: Learn to interpret key parts of regression output in Stata.

Prerequisites: Familiarity with importing data and running commands in Stata.



Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

Preparing to Run Regression

Launch Stata: Import your dataset to Stata.

View Your Data:

Command: Type browse in the command box to see all your variables and data points.



Running the Regression Command

Regression Command:

Command: reg PCE Income

Explanation: reg stands for regression. You type the dependent variable (PCE) first, followed by the independent variable (Income).

Purpose of Regression:

Helps to understand how changes in Income explain or predict changes in PCE.



Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

Observations

Observations:

Value: Shows the total sample size

Meaning: Confirms how many data points are being used for the regression.



F-Statistic

Value: Tells us how well the independent variable(s) explain the dependent variable.

Interpretation: Higher F-values mean the model fits the data well.

Significance: If the F-value is high, it suggests that Income is useful for predicting PCE.



P-Value of the F-Statistic

Explanation: Shows the significance level of the F-statistic.

Interpretation: If P-value is less than 0.05 (5%), the model is significant at the 5% level.

Practical Meaning: If significant, we can conclude the model is good at explaining variations in PCE with Income.



R-Squared

R-Squared:

Definition: Percentage of variation in the dependent variable (PCE) explained by the independent variable (Income).

Interpretation: Higher R-squared values indicate a better model fit. Example: An R-squared of 0.8 means 80% of PCE variation is explained by Income.



Adjusted R-Squared

Explanation: Adjusts for the number of independent variables to prevent overfitting.

Tip: Can decrease if you add too many explanatory variables without much effect on PCE.



Understanding the Coefficient Table



Indepth Research Institute
Transforming People and Organizations in Africa Since 2003

Coefficient

Meaning: Shows the estimated effect of Income on PCE.

Interpretation: Positive or negative sign indicates relationship direction (positive means increase in Income increases PCE).



Standard Error

Definition: Shows variability in the coefficient estimate.

Practical Use: Smaller standard errors suggest more precise estimates.



T-Statistic

Definition: Measures how many standard errors the coefficient is away from zero.

Calculation: Coefficient divided by its standard error.

Significance: Higher T-values suggest stronger evidence that Income affects PCE.



P-value

Interpretation: Tells whether the relationship between Income and PCE is statistically significant.

Meaning: Lower p-values (typically < 0.05) suggest a meaningful effect of Income on PCE.



Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

95% Confidence Interval

Definition: Range within which the true coefficient likely falls, assuming it's significant.

Example: If the confidence interval for Income is $[0.81, 0.83]$, it's likely the coefficient is within this range.

Interpretation: If the interval contains the coefficient, the effect is significant.



Break Down of the ANOVA Table



Indepth Research Institute
Transforming People and Organizations in Africa Since 2003

ANOVA Table Overview

The ANOVA table (Analysis of Variance) shows a breakdown of variability in your data, helping you understand how well your model fits.



Sum of Squares (SS)

The first column, SS, stands for "Sum of Squares," which measures different types of variance:

Model SS (Explained Sum of Squares or ESS): Represents the variation in the dependent variable explained by the model (notation: ESS or sometimes SSE in textbooks).

Residual SS (RSS): Measures the variation not explained by the model, also called the error term.

Total SS: This is the overall variance, combining both the model and residual sums of squares. Mathematically:

$$\text{Total SS} = \text{Model SS} + \text{Residual SS}$$



Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

Understanding Degrees of Freedom (DF)

Total DF: $n-1$, where n = sample size.

Residual DF: n - number of parameters (includes intercept and slope).

Model DF: Difference between Total DF and Residual DF.



Mean Square (MS) - Calculation

What is Mean Square (MS)?

MS is the average SS per degree of freedom.

Formula: $MS = SS \div DF$ (for each row: Model, Residual, Total).

Provides useful values for model diagnostics.



Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

Sigma Hat Squared (σ^2 or s^2)

Also called s^2 .

Found in the residual MS (RSS/DF).

Represents the variance of residuals, which helps assess model accuracy.



Root Mean Squared Error (MSE)

Square root of σ^2 (or s^2).

Interpretation: Indicates the typical size of errors in predictions.

Note: You can calculate σ^2 by squaring Root MSE, or vice versa.



Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

R-Squared (R^2) - Model Fit

Measures how well the model explains data variability.

Formula: $R^2 = \text{Model SS} / \text{Total SS}$

Higher R^2 means a better fit.



Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

Questions?

Open Floor for Questions/Discussions



Indepth Research Institute
Transforming People and Organizations in Africa Since 2003



Indepth Research Institute
Transforming People and Organizations in Africa Since 2003



(+254) 715 077 817 or (+254) 792 516 000



outreach@indepthresearch.org



www.indepthresearch.org



Runda-Nairobi, Tala Road, Off Kiambu Road