Reading Stata Regression Outputs

Introductory Econometrics I

Spring 2024

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

Source	SS	df	MS	Number of obs	=	100
				F(2, 97)	=	145.15
Model	22.5515771	2	11.2757885	Prob > F	=	0.0000
Residual	7.53516674	97	.077682131	R-squared	=	0.7496
				Adj R-squared	=	0.7444
Total	30.0867438	99	.303906503	Root MSE	=	.27872

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	1.529383	.0907477	16.85	0.000	1.349274	1.709492
x2	.2283016	.0965735	2.36	0.020	.0366299	.4199732
_cons	.3704453	.074674	4.96	0.000	.2222381	.5186526

Source	SS	df	MS
Model	22.5515771	2	11.2757885
Residual	7.53516674	97	.077682131
Total	30.0867438	99	.303906503

- **Source** – This is the source of variance, Model, Residual, and Total.
- **SS** –Sum of Squares

$$Total\ SS = Model\ SS + Residual\ SS,$$

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i \hat{u}_i^{\,2}.$$

- **Residual MS**: $\frac{1}{N-k-1}\sum_i \hat{u}_i^{\,2}$

Number of obs	=	100
F(2, 97)	=	145.15
Prob > F	=	0.0000
R-squared	=	0.7496
Adj R-squared	=	0.7444
Root MSE	=	.27872

- **F** and **Prob > F** – F statistics and p-values if all slope coefficients are zero. In this example, the F-stat associated with testing the null hypothesis: $\beta_1 = \beta_2 = 0$ against the alternative that it is not true.
- **R-squared** – R^2

$$R^2 = \frac{Model\ SS}{Total\ SS}.$$

- **Root MSE** – estimated standard deviation of the error term

$$Root\ MSE = \sqrt{Residual\ MS} = \sqrt{\frac{1}{N-k-1}\sum_i \hat{u}_i^{\,2}}.$$

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	1.529383	.0907477	16.85	0.000	1.349274	1.709492
x2	.2283016	.0965735	2.36	0.020	.0366299	.4199732
_cons	.3704453	.074674	4.96	0.000	.2222381	.5186526

- **Coef.** – Values of $\hat{\beta}s$
- **Std. Err.** – Standard errors associated with the coefficients.
- **t** and **P>|t|** – These columns provide the t-value and 2-tailed p-value used in testing the null hypothesis that the coefficient (parameter) is 0.
- **[95% Conf. Interval]** – This shows a 95% confidence interval for the coefficient.