

Standard Errors of Parameter Estimates

The standard errors of the estimated coefficients indicate their sampling variability, and hence their reliability. The estimated coefficient plus or minus one standard error is

0.190297 0.150341 0.172499

Std Error

approximately a 68% confidence interval for the true but unknown population parameter, and the estimated coefficient plus or minus two standard errors is approximately a 95% confidence interval, assuming that the estimated coefficient is approximately normally distributed.4 Thus large coefficient standard errors translate into wide confidence intervals.

58

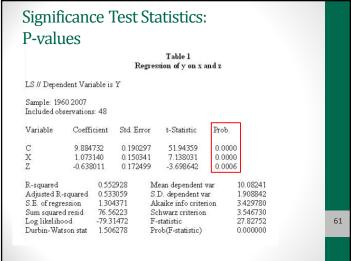
Significance Test Statistics: H₀: Parameter = 0

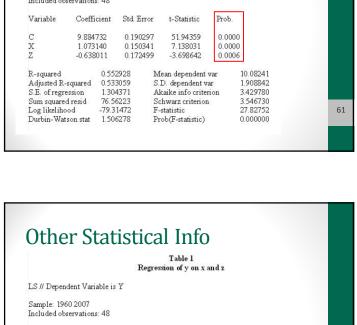
Each t statistic provides a test of the hypothesis variable irrelevance: that the true but unknown population parameter is zero, so that the corresponding variable contributes nothing and

51.94359 7.138031 -3.698642

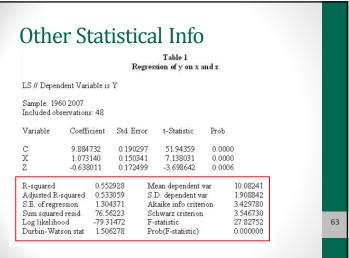
t-Statistic

can therefore be dropped. The t statistic is just the ratio of the estimated coefficient to its standard error.





Significance Test Statistics: P-values Prob. Associated with each t statistic is a probability value, or p-value, which is the probability of getting a value 0.0000 0.0000 of the t statistic at least as large in absolute value as the 0.0006 one actually obtained, assuming that the null hypothesis (irrelevance of the variable) is true. The smaller the p-value, the stronger the evidence against irrelevance.



R-squared Adjusted R-squared	0.552928 0.533059	Mean dependent var S.D. dependent var	10.08241 1.908842
E. of regression um squared resid	1.304371	Akaike info criterion Schwarz criterion	3.429780 3.546730
um squared resid og likelihood	-79.31472	F-statistic	27.82752
ındard measu		Prob(F-statistic) all goodness of fit i	
	re of over		s R2 (R-sqı
tandard measu hnically the pe iables in the m	re of over	all goodness of fit i	s R2 (R-sqı

Adjusted R²

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

Adjusts for the number of predictors in the model, penalizing (reducing the value) for each additional variable. A model with no "irrelevant" variables will have an adjusted R² almost as large as R².

$$R_{adj}^{2} = 1 - \frac{SSE/(n-k-1)}{SSTotal/(n-1)} = 1 - \frac{MSE}{MST}$$

65

S.E. of Regression

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

This is $\sqrt{\textit{MSE}}$. It is an unbiased estimator of σ , the standard deviation of the random errors (noise).

66

Sum Squared Residuals

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

This is SSE, the quantity we minimize to find the least-squares regression model.

67

Log Likelihood

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

The likelihood function is tremendously important in statistics, as it summarizes all the information contained in the data. It is simply the joint density function of the data, viewed as a function of the model parameters. A great method of optimization is called maximum likelihood estimation, and LS estimators are also maximum likelihood estimators. The given value is the maximum value of the likelihood function. It can be useful when comparing models, but on its own is kind of useless.

Durbin-Watson Stat

R-squared Mean dependent var 10.08241 Adjusted R-squared 0.533059 S.D. dependent var 1.908842 1.304371 S.E. of regression Akaike info criterion 3.429780 3 546730 Sum squared resid 76 56223 Schwarz criterion -79 31472 27 82752 Log likelihood F-statistic Durbin-Watson stat 1.506278 Prob(F-statistic) 0.000000

The Durbin-Watson statistic tests lag 1 serial correlation; i.e., is the last residual correlated to the present residual. In time series, we check for serial correlation at ALL lags, not just lag 1.

If there is no serial correlation at the tested lag, the DW stat should have a value of 2 (it is always between 0 and 4). Values less than 2 indicate positive correlation, and values greater than 2 indicate negative correlation.

09

Mean Dependent Variable

R-squared Mean dependent var Adjusted R-squared 0.533059 S.D. dependent var 1.908842 3.429780 S.E. of regression 1 304371 Akaike info criterion 76 56223 3 546730 Sum squared resid Schwarz criterion -79 31472 Log likelihood F-statistic 27 82752 Durbin-Watson stat 1.506278 Prob(F-statistic) 0.000000

This is just the value of \bar{y} .

, ,

S.D. Dependent Variable

Mean dependent var R-squared Adjusted R-squared 1.908842 0.533059 S.D. dependent var S.E. of regression 1.304371 Akaike info criterion 3.429780 Sum squared resid 76.56223 Schwarz criterion 3.546730 Log likelihood -79.31472 F-statistic 27.82752 Durbin-Watson stat 1.506278 Prob(F-statistic) 0.000000

This is just the value of s_v^2 .

Akaike Info Criterion (AIC) and Schwarz criterion (SIC)

0.552928 R-squared Mean dependent var Adjusted R-squared 0.533059 S.D. dependent var 1 908842 S.E. of regression 1.304371 Akaike info criterion 3.429780 76.56223 3.546730 Sum squared resid Schwarz criterion 27.82752 Log likelihood F-statistic Durbin-Watson stat Prob(F-statistic)

The Akaike and Schwarz criteria are used for model selection, and in certain contexts they have provable optimality properties in that regard. Both are penalized versions of MSE, where the penalties are functions of the degrees of freedom used in fitting the model. For both AIC and SIC, "smaller is better."

SIC is also called Bayes Info Criterion, or BIC.



0.552928 Mean dependent var 10.08241 R-squared Adjusted R-squared 0.533059 S.D. dependent var 1.908842 1.304371 3.429780 S.E. of regression Akaike info criterion 76 56223 3 546730 Sum squared resid Schwarz criterion -79.31472 Log likelihood 27 82752 F-statistic Durbin-Watson stat 1.506278 Prob(F-statistic) 0.000000

We use the F statistic to test the hypothesis that the coefficients of all variables in the regression except the intercept are jointly zero, versus the alternative that at least one of the coefficients is not zero.

/3

F-statistic

0.552928 Mean dependent var 10.08241 R-squared Adjusted R-squared 0.533059 S.D. dependent var 1.908842 1.304371 3.429780 S.E. of regression Akaike info criterion 76 56223 3 546730 Sum squared resid Schwarz criterion -79.31472 27.82752 Log likelihood F-statistic Durbin-Watson stat 1.506278 Prob(F-statistic) 0.000000

The F-statistic is computed from values sometimes given in an ANOVA table.

Source	d.f.	Sum of Squares	Mean Square	t.s.	p-value
Model	K	SSRegression	SSR/K	MSR	$F_{K,T-K-1}$
Error	T-K-1	SSE	SSE/(T-K-1)	MSE	
Total	<i>T</i> -1	SSTotal			

/4

Prob(F-statistic)

Mean dependent var 10.08241 R-squared Adjusted R-squared 0.533059 1.908842 S.D. dependent var S.E. of regression 1.304371 Akaike info criterion 3.429780 Sum squared resid 76.56223 Schwarz criterion 3.546730 Log likelihood -79.31472 F-statistic 27.82752 Durbin-Watson stat 1.506278 Prob(F-statistic) 0.000000

This is the p-value associated with the F-statistic.

In the case of only one right-hand-side variable, the t and F statistics contain exactly the same information, and one can show that $F=t^2$.

75

3.4 Regression From a Forecasting Perspective



The Key to Everything (or at Least Many Things)

Linear least squares regression, by construction, is consistent under very general conditions for "the linear function of x_t that gives the best approximation to y_t under squared-error loss."

This is a projection (like your shadow on a sunny day).

We project "reality" (the data) onto our model.

7

The Key to Everything (or at Least Many Things)

To forecast y_t for any given value of x_t , we can use the fitted line to find the value of y_t that corresponds to the given value of x_t .

This is the optimal forecast under quadratic loss, or the best linear approximation to it.

78

Why Take a Probabilistic Approach to Regression, as Opposed to Pure Curve Fitting?

We may want to test hypotheses regarding which variables.

We want to quantify the uncertainty associated with our forecasts (that is, we want interval and density forecasts).

79

Why Take a Probabilistic Approach to Regression, as Opposed to Pure Curve Fitting?

The regression function is the conditional expectation of y_t given x_t .

This is crucial for forecasting, because the expectation of future y conditional upon available information is a particularly good forecast.

In fact, under quadradic loss, it is the best possible forecast.

Residual Plots

Residual plots are useful for visually flagging neglected things that impact forecasting.

Residual serial correlation indicates that point forecasts could be improved.

81

Chapter 4

We will skip Chapter 4. It is mainly concerned with highlevel regression techniques to address specific problems in the FIC. These techniques are highly problem specific, and require a greater mathematical statistics understanding of the multiple regression model than is feasible in this course.