

Interpret Linear Regression Results

This example shows how to display and interpret linear regression output statistics.

On this page...

[Load sample data and define predictor variables.](#)
[Fit linear regression model.](#)
[Perform analysis of variance \(ANOVA\) for the model.](#)
[Decompose ANOVA table for model terms.](#)
[Display coefficient confidence intervals.](#)
[Perform hypothesis test on coefficients.](#)

Load sample data and define predictor variables.

```
load carsmall
X = [Weight,Horsepower,Acceleration];
```

Fit linear regression model.

```
lm = fitlm(X,MPG,'linear')
```

lm =

Linear regression model:
y ~ 1 + x1 + x2 + x3

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	47.977	3.8785	12.37	4.8957e-21
x1	-0.0065416	0.0011274	-5.8023	9.8742e-08
x2	-0.042943	0.024313	-1.7663	0.08078
x3	-0.011583	0.19333	-0.059913	0.95236

Number of observations: 93, Error degrees of freedom: 89
Root Mean Squared Error: 4.09
R-squared: 0.752, Adjusted R-Squared 0.744
F-statistic vs. constant model: 90, p-value = 7.38e-27

This linear regression outputs display shows the following.

<code>y ~ 1 + x1 + x2 + x3</code>	Linear regression model in the formula form using Wilkinson notation. Here it corresponds to: $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon.$
First column (under Estimated Coefficients)	Terms included in the model.
Estimate	Coefficient estimates for each corresponding term in the model. For example, the estimate for the constant term (<code>intercept</code>) is 47.977.
SE	Standard error of the coefficients.
tStat	<i>t</i> -statistic for each coefficient to test the null hypothesis that the corresponding coefficient is zero against the alternative that it is different from zero, given the other predictors in the model. Note that <code>tStat = Estimate/SE</code> . For example, the <i>t</i> -statistic for the intercept is 47.977/3.8785 = 12.37.
pValue	<i>p</i> -value for the F statistic of the hypotheses test that the corresponding coefficient is equal to zero or not. For example, the <i>p</i> -value of the F-statistic for <code>x2</code> is greater than 0.05, so this term is not significant at the 5% significance level given the other terms in the model.
Number of observations	Number of rows without any NaN values. For example, <code>Number of observations</code> is 93 because the <code>MPG</code> data vector has 6 NaN values and one of the data vectors, <code>Horsepower</code> , has one NaN value for a different observation.
Error degrees of freedom	$n - p$, where n is the number of observations, and p is the number of coefficients in the model, including the intercept. For example, the model has four predictors, so the <code>Error degrees of freedom</code> is $93 - 4 = 89$.
Root mean squared error	Square root of the mean squared error, which estimates the standard deviation of the error distribution.
R-squared and Adjusted R-squared	Coefficient of determination and adjusted coefficient of determination, respectively. For example, the <code>R-squared</code> value suggests that the model explains approximately 75% of the variability in the response variable <code>MPG</code> .
F-statistic vs. constant model	Test statistic for the F-test on the regression model. It tests for a significant linear regression relationship between the response variable and the predictor variables.
p-value	<i>p</i> -value for the F-test on the model. For example, the model is significant with a <i>p</i> -value of 7.3816e-27.

You can request this display by using `disp`. For example, if you name your model `lm`, then you can display the outputs using `disp(lm)`.

Perform analysis of variance (ANOVA) for the model.

```
anova(lm, 'summary')
```

```
ans =
```

	SumSq	DF	MeanSq	F	pValue
Total	6004.8	92	65.269		
Model	4516	3	1505.3	89.987	7.3816e-27
Residual	1488.8	89	16.728		

This ANOVA display shows the following.

SumSq	Sum of squares for the regression model, <code>Model</code> , the error term, <code>Residual</code> , and the total, <code>Total</code> .
DF	Degrees of freedom for each term. Degrees of freedom is $n - 1$ for the total, $p - 1$ for the model, and $n - p$ for the error term, where n is the number of observations, and p is the number of coefficients in the model, including the intercept. For example, <code>MPG</code> data vector has six <code>NaN</code> values and one of the data vectors, <code>Horsepower</code> , has one <code>NaN</code> value for a different observation, so the total degrees of freedom is $93 - 1 = 92$. There are four coefficients in the model, so the model <code>DF</code> is $4 - 1 = 3$, and the <code>DF</code> for error term is $93 - 4 = 89$.
MeanSq	Mean squared error for each term. Note that <code>MeanSq</code> = <code>SumSq</code> / <code>DF</code> . For example, the mean squared error for the error term is $1488.8/89 = 16.728$. The square root of this value is the <code>root mean squared error</code> in the linear regression display, or 4.09.
F	F-statistic value, which is the same as <code>F-statistic vs. constant model</code> in the linear regression display. In this example, it is 89.987, and in the linear regression display this F-statistic value is rounded up to 90.
pValue	p -value for the F-test on the model. In this example, it is 7.3816e-27.

If there are higher-order terms in the regression model, `anova` partitions the model `SumSq` into the part explained by the higher-order terms and the rest. The corresponding F-statistics are for testing the significance of the linear terms and higher-order terms as separate groups.

If the data includes replicates, or multiple measurements at the same predictor values, then the `anova` partitions the error `SumSq` into the part for the replicates and the rest. The corresponding F-statistic is for testing the lack-of-fit by comparing the model residuals with the model-free variance estimate computed on the replicates.

See the `anova` method for details.

Decompose ANOVA table for model terms.

```
anova(lm)
```

```
ans =
```

	SumSq	DF	MeanSq	F	pValue
x1	563.18	1	563.18	33.667	9.8742e-08
x2	52.187	1	52.187	3.1197	0.08078

x3	0.060046	1	0.060046	0.0035895	0.95236
Error	1488.8	89	16.728		

This anova display shows the following:

First column	Terms included in the model.
SumSq	Sum of squared error for each term except for the constant.
DF	Degrees of freedom. In this example, DF is 1 for each term in the model and $n - p$ for the error term, where n is the number of observations, and p is the number of coefficients in the model, including the intercept. For example, the DF for the error term in this model is $93 - 4 = 89$. If any of the variables in the model is a categorical variable, the DF for that variable is the number of indicator variables created for its categories (number of categories $- 1$).
MeanSq	Mean squared error for each term. Note that $MeanSq = SumSq / DF$. For example, the mean squared error for the error term is $1488.8 / 89 = 16.728$.
F	F-values for each coefficient. The F-value is the ratio of the mean squared of each term and mean squared error, that is, $F = MeanSq(x_i) / MeanSq(Error)$. Each F-statistic has an F distribution, with the numerator degrees of freedom, DF value for the corresponding term, and the denominator degrees of freedom, $n - p$. n is the number of observations, and p is the number of coefficients in the model. In this example, each F-statistic has an $F_{(1, 89)}$ distribution.
pValue	p -value for each hypothesis test on the coefficient of the corresponding term in the linear model. For example, the p -value for the F-statistic coefficient of x_2 is 0.08078, and is not significant at the 5% significance level given the other terms in the model.

Display coefficient confidence intervals.

```
coefCI(lm)
```

```
ans =  
  
40.2702  55.6833  
-0.0088 -0.0043  
-0.0913  0.0054  
-0.3957  0.3726
```

The values in each row are the lower and upper confidence limits, respectively, for the default 95% confidence intervals for the coefficients. For example, the first row shows the lower and upper limits, 40.2702 and 55.6833, for the intercept, β_0 . Likewise, the second row shows the limits for β_1 and so on. Confidence intervals provide a measure of precision for linear regression coefficient estimates. A $100(1-\alpha)\%$ confidence interval gives the range the corresponding regression coefficient will be in with $100(1-\alpha)\%$ confidence.

You can also change the confidence level. Find the 99% confidence intervals for the coefficients.

```
coefCI(lm, 0.01)
```

```
ans =  
  
    37.7677    58.1858  
   -0.0095   -0.0036  
   -0.1069    0.0211  
   -0.5205    0.4973
```

Perform hypothesis test on coefficients.

Test the null hypothesis that all predictor variable coefficients are equal to zero versus the alternate hypothesis that at least one of them is different from zero.

```
[p,F,d] = coefTest(lm)
```

```
p =  
  
    7.3816e-27
```

```
F =  
  
    89.9874
```

```
d =  
  
     3
```

Here, `coefTest` performs an F-test for the hypothesis that all regression coefficients (except for the intercept) are zero versus at least one differs from zero, which essentially is the hypothesis on the model. It returns `p`, the *p*-value, `F`, the F-statistic, and `d`, the numerator degrees of freedom. The F-statistic and *p*-value are the same as the ones in the linear regression display and ANOVA for the model. The degrees of freedom is $4 - 1 = 3$ because there are four predictors (including the intercept) in the model.

Now, perform a hypothesis test on the coefficients of the first and second predictor variables.

```
H = [0 1 0 0; 0 0 1 0];  
[p,F,d] = coefTest(lm,H)
```

```
p =  
  
    5.1702e-23
```

```
F =
```

96.4873

d =

2

The numerator degrees of freedom is the number of coefficients tested, which is 2 in this example. The results indicate that at least one of β_2 and β_3 differs from zero.

See Also

[anova](#) | [fitlm](#) | [LinearModel](#) | [stepwiselm](#)

Related Examples

- [Examine Quality and Adjust the Fitted Model](#)

More About

- [Coefficient Standard Errors and Confidence Intervals](#)
- [Coefficient of Determination \(R-Squared\)](#)
- [F-statistic and t-statistic](#)
- [Summary of Output and Diagnostic Statistics](#)

Was this topic helpful?