Multiple Linear Regression II: Inferences on the Model

Distribution of the Sum of Squares Error

Model assumptions:

- $Y \mid x$ follows a normal distribution with variance σ^2 and mean given by the model.
- $Y \mid x$ is independent of $Y \mid x'$ for $x \neq x'$.

(Here x may be a vector of several different factors or a single factor.)

Goal: Find the distribution of the error sum of squares

$$\mathrm{SS_E} = \langle \mathbf{Y}, (\mathbb{1}_n - H)\mathbf{Y} \rangle$$

where $\mathbf{Y} = (Y_1, \dots, Y_n)$ is the response vector and

$$H := X(X^T X)^{-1} X^T$$

is the hat matrix. Here X is the $(p+1) \times n$ model specification matrix.

Trace of $\mathbb{1}_n - H$

We first need a basic result from linear algebra:

27.1. Lemma. Let $P\colon \mathbb{R}^n \to \mathbb{R}^n$ be a projection, i.e., $P^2 = P$. Then the eigenvalues of P may only have values of 0 or 1.

Proof.
Suppose that $Pv = \lambda v$ for some $v \in \mathbb{R}^n$, $v \neq 0$, and $\lambda \in \mathbb{R}$. Then

$$\lambda v = Pv = P^2 v = P(\lambda v) = \lambda(Pv) = \lambda^2 v$$

so $\lambda = \lambda^2$, i.e., $\lambda = 0$ or $\lambda = 1$. $\square$

Trace of $\mathbb{1}_n - H$

Recall that the trace of a square $n \times n$ matrix $A = (a_{ij})$ is defined as

$$\operatorname{tr} A := \sum_{i=1}^{n} a_{ii}.$$

We will use the properties

$$\operatorname{tr}(A+B) = \operatorname{tr} A + \operatorname{tr} B, \qquad \operatorname{tr}(AB) = \operatorname{tr}(BA)$$

for square $n \times n$ matrices A, B. Furthermore,

$$\operatorname{tr} A = \text{sum of the eigenvalues of } A.$$

We have

$$\begin{aligned}\operatorname{tr} H &= \operatorname{tr}(X(X^T X)^{-1} X^T) = \operatorname{tr}((X^T X)^{-1} X^T X) \\ &= \operatorname{tr}(\mathbb{1}_{p+1}) = p + 1.\end{aligned}$$

so

$$\operatorname{tr}(\mathbb{1}_n - H) = \operatorname{tr} \mathbb{1}_n - \operatorname{tr} H = n - p - 1$$

Eigenvalues of $\mathbb{1}_n - H$

Since $\mathbb{1}_n - H$ is a projection, the sum of its eigenvalues is also equal to the number of eigenvalues that equal 1.

Hence, $n - p - 1$ eigenvalues of $\mathbb{1}_n - H$ are equal to 1 and $p + 1$ eigenvalues equal 0.

Since $\mathbb{1}_n - H$ is symmetric, we can apply the spectral theorem of linear algebra: there exists a matrix U (whose columns are eigenvectors of $\mathbb{1}_n - H$) such that

$$U^{-1} = U^T$$

and

$$U^T(\mathbb{1}_n - H)U = \begin{pmatrix} \mathbb{1}_{n-p-1} & 0 \\ 0 & 0 \end{pmatrix} =: D_{n-p-1} \tag{27.1}$$

Distribution of the Sum of Squares Error

Recall that in our model the response vector satisfies

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}$$

where $\mathbf{E}$ follows a normal distribution with mean 0 and variance σ^2.

Since $\mathbb{1}_n - H$ is an orthogonal projection and $(\mathbb{1}_n - H)X = 0$ (see (26.12)) and we find

$$\begin{aligned} \mathrm{SS_E} &= \langle (\mathbb{1}_n - H)\mathbf{Y}, (\mathbb{1}_n - H)\mathbf{Y} \rangle \\ &= \langle (\mathbb{1}_n - H)(X\boldsymbol{\beta} + \mathbf{E}), (\mathbb{1}_n - H)(X\boldsymbol{\beta} + \mathbf{E}) \rangle \\ &= \langle (\mathbb{1}_n - H)\mathbf{E}, (\mathbb{1}_n - H)\mathbf{E} \rangle \\ &= \langle \mathbf{E}, (\mathbb{1}_n - H)\mathbf{E} \rangle \end{aligned}$$

Distribution of the Sum of Squares Error

Since each E_j follows an independent normal distribution with mean zero and variance σ^2, we have

$$\frac{\mathrm{SS_E}}{\sigma^2} = \left\langle \frac{\mathbf{E}}{\sigma}, (\mathbb{1}_n - H)\left(\frac{\mathbf{E}}{\sigma}\right)\right\rangle = \langle \mathbf{Z}, (\mathbb{1}_n - H)\mathbf{Z}\rangle$$

where $\mathbf{Z} = (Z_1, \dots, Z_n)^T$ is a vector of i.i.d. standard normal random variables.

We now use the diagonalization (27.1),

$$\begin{aligned}\frac{\mathrm{SS_E}}{\sigma^2} &= \langle Z, U^T D_{n-p-1} UZ\rangle = \langle UZ, D_{n-p-1} UZ\rangle \\ &= \sum_{i=1}^{n-p-1} (UZ)_i^2\end{aligned}$$

Since each Z_j follows an independent standard normal distribution, so does each component of UZ. We conclude immediately that $\mathrm{SS_E}$ follows a chi-squared distribution with $n-p-1$ degrees of freedom.

Distribution of the Sum of Squares Error

We can apply analogous arguments to $\mathrm{SS_R}$ and $\mathrm{SS_T}$. In summary, we have

27.2. Theorem.

(i) $\mathrm{SS_E}/\sigma^2$ follows a chi-squared distribution with $n-p-1$ degrees of freedom.

(ii) If $\beta = (\beta_0, 0, \dots, 0)$, then $\mathrm{SS_R}/\sigma^2$ follows a chi-squared distribution with p degrees of freedom.

Furthermore, $\mathrm{SS_R}$ and $\mathrm{SS_E}$ are independent random variables.

27.3. Corollary. The estimator

$$S^2 := \frac{\mathrm{SS_E}}{n-p-1}$$

is unbiased for σ^2.

Practical Calculations

27.4. Lemma. The regression sum of squares can be expressed as

$$\mathrm{SS_R} = \langle \mathbf{b}, X^T Y \rangle - \frac{1}{n}\Big(\sum_{i=1}^{n} Y_i\Big)^2$$

In particular, in the case of the multilinear model,

$$\mathrm{SS_R} = b_0 \sum_{i=1}^{n} Y_i + \sum_{j=1}^{p} b_j \sum_{i=1}^{n} x_{ji} Y_i - \frac{1}{n}\Big(\sum_{i=1}^{n} Y_i\Big)^2,$$

and in the polynomial model,

$$\mathrm{SS_R} = b_0 \sum_{i=1}^{n} Y_i + \sum_{j=1}^{p} b_j \sum_{i=1}^{n} x_i^j Y_i - \frac{1}{n}\Big(\sum_{i=1}^{n} Y_i\Big)^2.$$

Estimated Variance and Correlation Coefficient

27.5. Example. In Example 26.2 we obtained the regression equation

$$\hat{\mu}_{Y|x_1,x_2} = 24.75 - 4.16x_1 - 0.015x_2.$$

for the mean gas mileage of cars as a function of weight x_1 and motor temperature x_2. We now want to find R^2 for our model.

It is convenient to write

$$\mathrm{SS_R} = \langle B, X^T Y \rangle - \frac{1}{n}\Big(\sum_{i=1}^{n} Y_i\Big)^2.$$

We first calculate

$$X^T y = \begin{pmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \end{pmatrix} = \begin{pmatrix} 170.00 \\ 282.405 \\ 8887.00 \end{pmatrix}, \qquad \sum_{i=1}^{n} y_i^2 = 2900.46.$$

Estimated Variance and Coefficient of Determination

These values give us

$$\mathrm{SS_T} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\Big(\sum_{i=1}^{n} y_i\Big)^2 = 10.46,$$

$$\mathrm{SS_R} = \left\langle \begin{pmatrix} 170.00 \\ 282.405 \\ 8887.00 \end{pmatrix}, \begin{pmatrix} 24.75 \\ -4.16 \\ -0.015 \end{pmatrix} \right\rangle - \frac{170.00^2}{25} = 10.32.$$

Hence,

$$R^2 = \frac{10.32}{10.46} = 0.9866.$$

We also note that $\mathrm{SS_E} = S_{yy} - \mathrm{SS_R} = 10.46 - 10.32 = 0.14$ and the estimated variance is

$$\hat{\sigma}^2 = s^2 = \frac{\mathrm{SS_E}}{n-p-1} = \frac{0.14}{10-2-1} = 0.02.$$

Estimated Variance and Coefficient of Determination

We can extract the estimated variance and R^2 directly from the model:

```
model = LinearModelFit[data, {x1, x2}, {x1, x2}];
model["EstimatedVariance"]
```

```
0.02005
```

```
model["RSquared"]
```

```
0.986582
```

F-Test for Significance of Regression

Since $\mathrm{SS_R}$ measures the variability associated with the model and $\mathrm{SS_E}$ measures "random variation", we will find the regression significant if $\mathrm{SS_R}$ is much larger than $\mathrm{SS_E}$. The basis for the test is Theorem 27.2.

27.6. *F*-Test for Significance of Regression. Let $x_1, \dots, x_p$ be the predictor variables in a multilinear model (26.1) for Y. Then

$$H_0\colon \beta_1 = \beta_2 = \cdots = \beta_p = 0,$$

is rejected at significance level α if the test statistic

$$F_{p,n-p-1} = \frac{\mathrm{SS_R}\,/p}{\mathrm{SS_E}\,/(n-p-1)} = \frac{\mathrm{SS_R}\,/p}{S^2} \tag{27.2}$$

satisfies $F_{p,n-p-1} > f_{\alpha,p,n-p-1}$.

Significance of Regression

We remark that

$$F_{p,n-p-1} = \frac{n-p-1}{p}\frac{\mathrm{SS_R}/S_{yy}}{\mathrm{SS_E}/S_{yy}} = \frac{n-p-1}{p}\frac{\mathrm{SS_R}/S_{yy}}{(S_{yy}-\mathrm{SS_R})/S_{yy}}$$
$$= \frac{n-p-1}{p}\frac{R^2}{1-R^2}$$

so the value of R^2 alone can be used to test for significance of regression.

27.7. Example. In Example 27.5, we obtained $R^2 = 0.986$. Since $n = 10$ and $p = 2$ the value of the test statistic for significance of regression is

$$\frac{n-p-1}{p}\frac{R^2}{1-R^2} = \frac{7}{2}\frac{0.986^2}{0.014} = 243.05.$$

The 95% point of the $F_{2,7}$-distribution is 4.74, so we can reject H_0 with $P < 0.05$. There is evidence that the regression is significant.

Expectation for Random Vectors

Goal: Derive distribution of the model parameters $\boldsymbol{\beta}$.

Recall: Let $Y = (Y_1, \dots, Y_n)^T$ be a random vector. Then

$$\mathrm{E}[Y] = \begin{pmatrix} \mathrm{E}[Y_1] \\ \vdots \\ \mathrm{E}[Y_n] \end{pmatrix}.$$

For random vectors Y, Z and a constant $m \times n$ matrix C:

(i) $\mathrm{E}[C] = C$,

(ii) $\mathrm{E}[CY] = C\,\mathrm{E}[Y]$,

(iii) $\mathrm{E}[Y + Z] = \mathrm{E}[Y] + \mathrm{E}[Z]$.

Expectation of the Least-Squares Estimators

We can calculate directly that the expectation of the response vector is

$$\mathrm{E}[\mathbf{Y}] = \mathrm{E}[X\boldsymbol{\beta} + \mathbf{E}] = \mathrm{E}[X\boldsymbol{\beta}] + \mathrm{E}[\mathbf{E}] = X\boldsymbol{\beta}.$$

Then

$$\begin{aligned} E[\mathbf{b}] &= \mathrm{E}[(X^TX)^{-1}X^T\mathbf{Y}] = (X^TX)^{-1}X^T\,\mathrm{E}[\mathbf{Y}] \\ &= (X^TX)^{-1}X^TX\boldsymbol{\beta} \\ &= \boldsymbol{\beta}. \end{aligned}$$

It follows that $\hat{\boldsymbol{\beta}} = \mathbf{b}$ is an unbiased estimator for $\boldsymbol{\beta}$.

Variance for Random Vectors

Recall: Let $Y = (Y_1, \dots, Y_n)^T$ be a random vector. Then

$$\operatorname{Var}[Y] = \begin{pmatrix} \operatorname{Var}[Y_1] & \operatorname{Cov}[Y_1, Y_2] & \dots & \operatorname{Cov}[Y_1, Y_n] \\ \operatorname{Cov}[Y_1, Y_2] & \operatorname{Var}[Y_2] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \operatorname{Cov}[Y_{n-1}, Y_n] \\ \operatorname{Cov}[Y_1, Y_n] & \dots & \operatorname{Cov}[Y_{n-1}, Y_n] & \operatorname{Var}[Y_n] \end{pmatrix}$$

and

$$\operatorname{Var}[CY] = C \operatorname{Var}[Y] C^T,$$

where C is a constant $m \times n$ matrix.

Variance of the Least-Squares Estimators

In our case, a random sample $(x_1, Y_1), \ldots, (x_n, Y_n)$ is given where the Y_i are independent and all Y_i have the same variance σ^2.

Therefore,

$$\mathrm{Var}[\mathbf{Y}] = \sigma^2 \mathbb{1}_n.$$

We then have

$$\begin{aligned}
\mathrm{Var}[\mathbf{b}] &= \mathrm{Var}[(X^T X)^{-1} X^T \mathbf{Y}] \\
&= (X^T X)^{-1} X^T \,\mathrm{Var}[\mathbf{Y}]((X^T X)^{-1} X^T)^T \\
&= \sigma^2 (X^T X)^{-1} X^T ((X^T X)^{-1} X^T)^T \\
&= \sigma^2 (X^T X)^{-1}
\end{aligned}$$

Variance of the Least-Squares Estimators

Let us write

$$X^T X = \begin{pmatrix} \xi_{00} & * & \cdots & \cdots & * \\ * & \xi_{11} & \ddots & & * \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & * \\ * & \cdots & \cdots & * & \xi_{pp} \end{pmatrix}$$

where the starred values are uninteresting for us.

Hence,

$$\mathrm{Var}[B_i] = \xi_{ii}\sigma^2, \qquad\qquad i = 0, \dots, p,$$

Note that the estimators $B_0, \dots, B_p$ are not independent of each other, but we will not investigate their covariance here.

Distribution of the Least-Squares Estimators

Since

$$\mathbf{b} = (X^T X)^{-1} X^T \mathbf{Y}$$

and the components of $\mathbf{Y}$ follow normal distributions, each b_i is a linear combination of independent normal distributions. Hence, each b_i must itself follow a normal distribution.

We have therefore proved the following result:

27.8. Theorem. The random vector $\mathbf{b}$ follows a normal distribution with mean $\boldsymbol{\beta}$ and variance-covariance matrix $\sigma^2 (X^T X)^{-1}$.

It is also possible to prove:

27.9. Theorem. The statistic $(n-p-1)S^2/\sigma^2 = \mathrm{SS_E}/\sigma^2$ is independent of $\boldsymbol{\beta}$.

Confidence Intervals for the Model Parameters

The variables

$$Z = \frac{b_j - \beta_j}{\sigma\sqrt{\xi_{jj}}}, \qquad\qquad j = 0, \dots, p,$$

are standard normal. Thus, for $j = 0, \dots, p$,

$$\frac{(\hat{\beta}_j - \beta_j)/(\sigma\sqrt{\xi_{jj}})}{\sqrt{(n-p-1)S^2/\sigma^2/(n-p-1)}} = \frac{\hat{\beta}_j - \beta_j}{S\sqrt{\xi_{jj}}}, \tag{27.3}$$

follows a T-distribution with $n-p-1$ degrees of freedom.

We immediately obtain the following $100(1-\alpha)\%$ confidence intervals for the model parameters:

$$\beta_j = \hat{\beta}_j \pm t_{\alpha/2,n-p-1} S\sqrt{\xi_{jj}}, \qquad\qquad j = 0, \dots, p.$$

Confidence Intervals for the Model Parameters

27.10. Example. Continuing from Example 27.5, we have $s^2 = 0.02005$, so the variance-covariance matrix is

```
MatrixForm[0.02005 Inverse[Transpose[X].X]]
```

$$\begin{pmatrix} 0.121719 & -0.060669 & -0.000344635 \\ -0.060669 & 0.0348589 & 0.0000434344 \\ -0.000344635 & 0.0000434344 & 5.17872\times 10^{-6} \end{pmatrix}$$

We can also obtain the matrix directly from the model:

```
MatrixForm[model["CovarianceMatrix"]]
```

$$\begin{pmatrix} 0.121719 & -0.0606689 & -0.000344635 \\ -0.0606689 & 0.0348589 & 0.0000434344 \\ -0.000344635 & 0.0000434344 & 5.17871\times 10^{-6} \end{pmatrix}$$

Confidence Intervals for the Model Parameters

Reading off from the diagonal, we find the variances of the estimators:

$$\begin{aligned}\widehat{\operatorname{Var} B_0} &= s^2\xi_{00} = 0.1217,\\ \widehat{\operatorname{Var} B_1} &= s^2\xi_{11} = 0.03485,\\ \widehat{\operatorname{Var} B_2} &= s^2\xi_{22} = 5.178 \cdot 10^{-6}.\end{aligned}$$

We hence have the following 95% confidence intervals:

$$\begin{aligned}\beta_0 &= \hat{\beta}_0 \pm t_{0.025,7}\sqrt{s^2\xi_{00}} = 24.75 \pm 2.365\sqrt{0.1217}\\ &= 24.75 \pm 0.825\\ \beta_1 &= \hat{\beta}_1 \pm t_{0.025,7}\sqrt{s^2\xi_{11}} = -4.16 \pm 2.365\sqrt{0.03485}\\ &= -4.16 \pm 0.44\\ \beta_2 &= \hat{\beta}_2 \pm t_{0.025,7}\sqrt{s^2\xi_{22}} = -0.15 \pm 2.365\sqrt{5.178 \cdot 10^{-6}}\\ &= -0.15 \pm 0.0054\end{aligned}$$

Confidence Intervals for the Model Parameters

Mathematica can directly give the confidence intervals and the standard deviations of the estimators (the square roots of the diagonal elements of the variance-covariance matrix).

```
model["ParameterConfidenceIntervalTable"]
```

	Estimate	Standard Error	Confidence Interval
1	24.7489	0.348882	{23.9239, 25.5738}
x_1	−4.15933	0.186705	{−4.60082, −3.71785}
x_2	−0.014895	0.00227568	{−0.0202761, −0.00951389}

Confidence Intervals for the Estimated Mean

Let us write

$$\mathbf{x_0} = \begin{pmatrix} 1 \\ x_{10} \\ \vdots \\ x_{p0} \end{pmatrix} \qquad \text{or} \qquad \mathbf{x_0} = \begin{pmatrix} 1 \\ x \\ \vdots \\ x^p \end{pmatrix}$$

depending on whether we are considering a multilinear or a polynomial model. Of course, any combination of the two may be considered analogously.

Our goal is to make inferences on the estimated mean at $\mathbf{x_0}$. We write

$$\widehat{\mu}_{Y|\mathbf{x_0}} = \mathbf{x_0}^T \mathbf{b} = \mathbf{x_0}^T (X^T X)^{-1} X^T \mathbf{Y}$$

We see that $\widehat{\mu}_{Y|\mathbf{x_0}}$ is a linear combination of the independent and normally distributed Y_i and therefore follows a normal distribution.

Confidence Intervals for the Estimated Mean

Furthermore,

$$\mathrm{E}[\widehat{\mu}_{Y|\mathbf{x_0}}] = \mathrm{E}[\mathbf{x_0}^T\mathbf{b}] = \mathbf{x_0}^T\,\mathrm{E}[\mathbf{b}] = \mathbf{x_0}^T\boldsymbol{\beta} = \mu_{Y|x_{10},\dots,x_{p0}}$$

and

$$\operatorname{Var}\widehat{\mu}_{Y|\mathbf{x_0}} = \mathbf{x_0}^T(\operatorname{Var}\hat{\beta})\mathbf{x_0} = \sigma^2\mathbf{x_0}^T(X^TX)^{-1}\mathbf{x_0}.$$

It follows that

$$\frac{\widehat{\mu}_{Y|\mathbf{x_0}} - \mu_{Y|\mathbf{x_0}}}{\sigma\sqrt{\mathbf{x_0}^T(X^TX)^{-1}\mathbf{x_0}}}$$

is standard normal and, after dividing by $\sqrt{(n-p-1)S^2/\sigma^2}/\sqrt{n-p-1}$ that

$$\frac{\hat{\mu}_{Y|\mathbf{x_0}} - \mu_{Y|\mathbf{x_0}}}{S\sqrt{\mathbf{x_0}^T(X^TX)^{-1}\mathbf{x_0}}} \tag{27.4}$$

follows a T distribution with $n-p-1$ degrees of freedom.

Confidence Intervals for the Estimated Mean

We thus have the following $100(1-\alpha)\%$ confidence interval for $\mu_{Y|\mathbf{x_0}}$:

$$\mu_{Y|\mathbf{x_0}} = \widehat{\mu}_{Y|\mathbf{x_0}} \pm t_{\alpha/2,n-p-1} S \sqrt{\mathbf{x_0}^T (X^T X)^{-1} \mathbf{x_0}}$$

27.11. Example. Following on from Example 27.5, the estimate for the average gasoline mileage for a car weighing 1.5 tons being operated at 70° F is

$$\hat{\mu}_{Y|1.5,70} = 24.75 - 4.16 \cdot 1.5 - 0.14897 \cdot 70 = 17.47.$$

We want to find a 95% confidence interval for this mean. The vector x_0 is given by

$$\mathbf{x_0} = \begin{pmatrix} 1 \\ 1.5 \\ 70 \end{pmatrix}.$$

Prediction Intervals

Then

$$\mu_{Y|1.5,70} = 17.47 \pm 2.365 \cdot S\sqrt{\mathbf{x_0}^T(X^TX)^{-1}\mathbf{x_0}} = 17.47 \pm 0.16.$$

This agrees with Mathematica's built-in functionality:

```
model["MeanPredictionBands"] /. {x1 → 1.5, x2 → 70}
```

```
{17.3105, 17.6239}
```

As in the previous section, we can obtain a similar $100(1-\alpha)\%$ ***prediction interval*** for the value of $Y \mid x_{10}, \dots, x_{p0}$,

$$Y \mid \mathbf{x_0} = \widehat{\mu}_{Y|\mathbf{x_0}} \pm t_{\alpha/2,n-p-1} S\sqrt{1 + \mathbf{x_0}^T(X^TX)^{-1}\mathbf{x_0}}.$$

We omit the (completely analogous) details.

Hypothesis Testing on the Model Parameters

Based on the T-distributions of (27.3) and (27.4) we can of course perform tests on the model parameters β and the predicted mean $\widehat{\mu}_{Y|x}$.

Since such tests should be routine by now, we omit the details. However, a special case is of interest:

27.12. T-Test for Model Sufficiency. Suppose that a regression model using the parameters $\beta_0, \dots, \beta_p$ is fitted to Y. Then for any $j = 0, \dots, p$

$$H_0 \colon \beta_j = 0$$

is rejected at significance level α if the test statistic

$$T_{n-p-1} = \frac{b_j}{S\sqrt{c_{jj}}}.$$

satisfies $|T_{n-p-1}| > t_{\alpha/2,n-p-1}$.

T-Test for the Model Parameters

If we are able to reject H_0, there is evidence that the predictor is needed for the model.

If we fail to reject H_0, there is no evidence that the predictor is needed and we may proceed to fit a model without this predictor.

27.13. Example. Suppose we are given the data

x	5	7.5	10	12.5	15	17.5	20
y	1	2.2	4.9	5.3	8.2	10.7	13.2

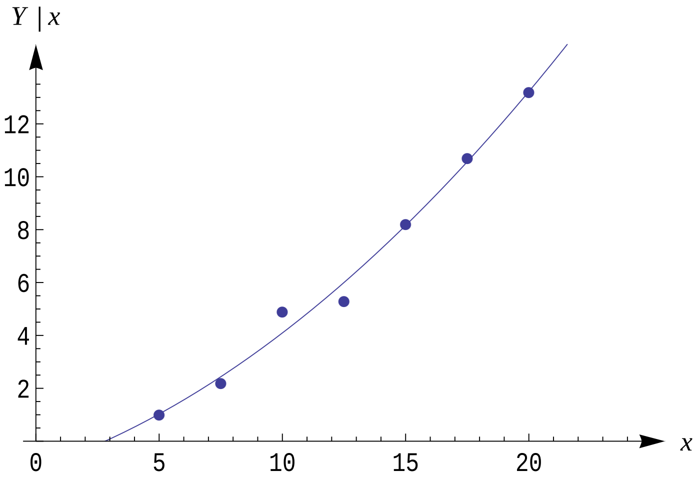
We would like to find a quadratic model for the data:

```
Data = {{5, 1}, {7.5, 2.2}, {10, 4.9`}, {12.5, 5.3`},
   {15, 8.2`}, {17.5, 10.7}, {20, 13.2`}};
model = NonlinearModelFit[Data, b0 + b1 x + b2 x^2 {b0, b1, b2}, x];
model["BestFit"]
```

$-1.03571 + 0.312857\,x + 0.02\,x^2$

T-Test for the Model Parameters

The data and the model curve is plotted below.



T-Test for the Model Parameters

We can find confidence intervals for all model parameters:

```
model["ParameterConfidenceIntervalTable",
 ConfidenceLevel → 0.975]
```

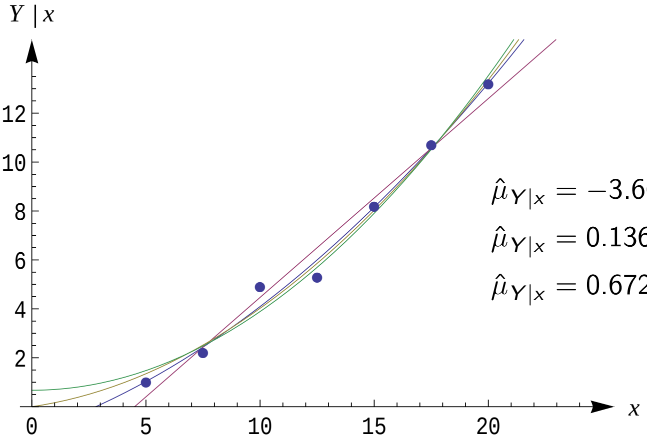
	Estimate	Standard Error	Confidence Interval
b_0	−1.03571	1.3838	{−5.87265, 3.80122}
b_1	0.312857	0.244475	{−0.541682, 1.1674}
b_2	0.02	0.00963554	{−0.0136801, 0.0536801}

Based on these 95% confidence intervals, we can not reject $H_0 : \beta_j = 0$ for any $j = 0, 1, 2$. This means that there is no evidence that any single β_j is non-zero.

However, not all coefficients will be zero. The regression is clearly significant (as can be seen by conducting a test for significance of regression; see Example 27.7).

T-Test for the Model Parameters

We can eliminate any one of the there predictors simply be deleting the corresponding column from the model specification matrix X. This yields the alternative models



$$\hat{\mu}_{Y|x} = -3.66071 + 0.812857x,$$
$$\hat{\mu}_{Y|x} = 0.136315x + 0.0265909x^2,$$
$$\hat{\mu}_{Y|x} = 0.67285 + 0.0321498x^2.$$

General Test for Model Sufficiency

It is of course not clear which of these three models is best; this is a question we will return at a later point.

The T-test 27.12 can be used to determine whether a single predictor may be eliminated from the model. It is often practical, however, to compare a general subset of predictors with a full model of $p + 1$ predictor variables,

$$\mu_{Y|x_1,\dots,x_p} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p. \tag{27.5}$$

After possibly renumbering the variables we compare with a ***reduced model*** of $m + 1 < p + 1$ predictor variables

$$\mu_{Y|x_1,\dots,x_m} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \dots + \tilde{\beta}_m x_m. \tag{27.6}$$

We define the sums of squares errors for the two models by

$$\begin{aligned} \mathrm{SS_{E;full}} &= \text{sum of squares error } \mathrm{SS_E} \text{ for full model,} \\ \mathrm{SS_{E;reduced}} &= \text{sum of squares error } \mathrm{SS_E} \text{ for reduced model.} \end{aligned}$$

Partial F-Test for Model Sufficiency

We will base our test on the principle that there is evidence that the full model is needed if $\mathrm{SS}_{\mathrm{E;full}} \ll \mathrm{SS}_{\mathrm{E;reduced}}$.

27.14. Partial F-Test for Model Sufficiency. Let $x_1, \dots, x_p$ be possible predictor variables for Y and (27.5) and (27.6) the full and reduced models, respectively. Then

$$H_0 \colon \text{the reduced model is sufficient}$$

is rejected at significance level α if the test statistic

$$F_{p-m,n-p-1} = \frac{n-p-1}{p-m} \frac{\mathrm{SS}_{\mathrm{E;reduced}} - \mathrm{SS}_{\mathrm{E;full}}}{\mathrm{SS}_{\mathrm{E;full}}} \tag{27.7}$$

satisfies $F_{p-m,n-p-1} > f_{\alpha,p-m,n-p-1}$.

Partial F-Test for Model Sufficiency

27.15. Example. In the context of Example 27.13 we can compare the linear and quadratic models

$$\widehat{\mu}_{Y|x;\text{full}} = -1.03571 + 0.312857x + 0.02x^2, \qquad \mathrm{SS}_{\mathrm{E;full}} = 1.21857,$$
$$\widehat{\mu}_{Y|x;\text{reduced}} = -3.66071 + 0.812857x, \qquad \mathrm{SS}_{\mathrm{E;reduced}} = 2.53107.$$

Here, $n = 7$, $p = 2$, $m = 1$, so

$$F_{p-m,n-p-1} = \frac{n-p-1}{p-m} \frac{\mathrm{SS}_{\mathrm{E;reduced}} - \mathrm{SS}_{\mathrm{E;full}}}{\mathrm{SS}_{\mathrm{E;full}}} = 4.30832.$$

The critical point $f_{0.05,1,4} = 7.71$, so we can not reject H_0 at the 5% level of significance. There is no evidence that the full model is needed.

Partial F-Test for Model Sufficiency

27.16. Example. Continuing with Example 27.13 we can also compare the general quadratic model with a square monomial model:

$$\begin{aligned} \widehat{\mu}_{Y|x;\text{full}} &= -1.03571 + 0.312857x + 0.02x^2, & \text{SS}_{\text{E;full}} &= 1.21857, \\ \widehat{\mu}_{Y|x;\text{reduced}} &= 0.0346414x^2, & \text{SS}_{\text{E;reduced}} &= 1.83967. \end{aligned}$$

Here, $n = 7$, $p = 2$, $m = 0$, so

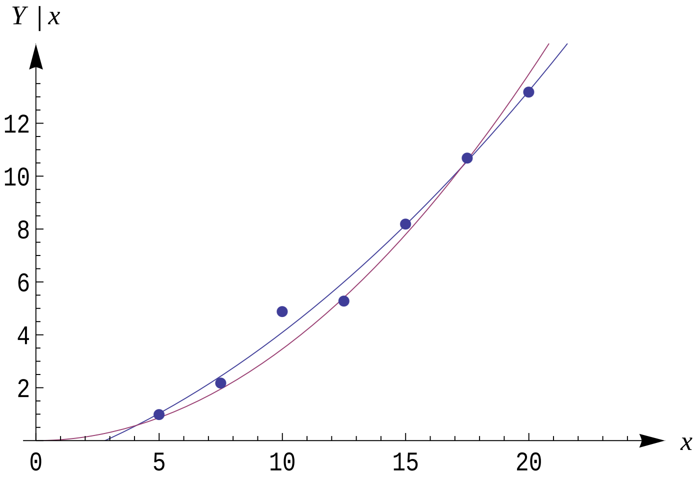
$$F_{p-m,n-p-1} = \frac{n-p-1}{p-m} \frac{\text{SS}_{\text{E;reduced}} - \text{SS}_{\text{E;full}}}{\text{SS}_{\text{E;full}}} = 1.01939.$$

The critical point $f_{0.05,2,4} = 6.94$, so we can not reject H_0 at the 5% level of significance. There is no evidence that the full model is needed.

Comparing the sum of squares errors with the previous example, we can furthermore conclude that a square monomial model gives a better fit than the linear model.

Partial F-Test for Model Sufficiency

The graph below shows the quadratic and the square monomial models.



T-Test and Partial F-Test for Single Predictors

While the T-test can be used to determine whether a single predictor is necessary for a given model, the F-test can be applied to an arbitrary subset of predictors.

The question arises whether there is a difference between the two tests when considering a single predictor, i.e., whether the F-test applied to a single variable (as in Example 27.15) always yields the same result as the T-test.

It is possible to prove that, indeed, the T-test for a single variable is equivalent to a partial F-Test when applied to a reduced model lacking only that single variable.

Interpretations of the Partial F-Test

Furthermore, since $\mathrm{SS_T} = \mathrm{SS_R} + \mathrm{SS_E}$, the test statistic (27.7) can be re-written as

$$\begin{aligned} F_{p-m,n-p-1} &= \frac{n-p-1}{p-m} \frac{\mathrm{SS_{E;reduced}} - \mathrm{SS_{E;full}}}{\mathrm{SS_{E;full}}} \\ &\frac{n-p-1}{p-m} \frac{\mathrm{SS_{R;full}} - \mathrm{SS_{R;reduced}}}{\mathrm{SS_{E;full}}}. \end{aligned}$$

This shows that the F-test for significance of regression based on the statistic (27.2),

$$F_{p,n-p-1} = \frac{n-p-1}{p} \frac{\mathrm{SS_R}}{\mathrm{SS_E}}$$

may be regarded as a partial F-test where the reduced model contains no regressors.

Moreover, the partial F-test can be formulated in terms of the determination coefficients R^2 for the full and reduced models.