Math 423/533 Midterm
The Theory So Far

Notation

- sample size n, data index i
- number of predictors, p ($p = 2$ for simple linear regression)
- y_i: response for individual i
- $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ – $(1 \times p)$ row vector
- $\mathbf{X}$ - $(n \times p)$ matrix containing all predictors for all individuals $i = 1, \ldots, n$.
- $\mathbf{y} = (y_1, \ldots, y_n)^\top$ – $(n \times 1)$ column vector
- Y_i and $\mathbf{Y}$: random variables corresponding to responses

Linear model assumptions

For $i = 1, \ldots, n$,

$$\mathbb{E}_{Y_i|\mathbf{x}_i}[Y_i|\mathbf{x}_i] = \mathbf{x}_i\beta = \sum_{j=0}^{p-1} \beta_j x_{ij}$$

and

$$\mathbb{V}\mathrm{ar}_{Y_i|\mathbf{x}_i}[Y_i|\mathbf{x}_i] = \sigma^2$$

where

$$\beta = (\beta_0, \ldots, \beta_{p-1})^\top$$

is the $(p \times 1)$ vector of regression coefficients, and $\sigma^2 > 0$ is the error variance.

We assume also that $Y_1, \ldots, Y_n$ are independent given $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

Linear model assumptions

In vector form

$$\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathbf{Y}|\mathbf{X}] = \mathbf{X}\beta \qquad (n \times 1)$$

and

$$\mathbb{V}\mathrm{ar}_{\mathbf{Y}|\mathbf{X}}[\mathbf{Y}|\mathbf{X}] = \sigma^2 \mathbf{I}_n \qquad (n \times n).$$

This is equivalent to a model specification of

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

where

$$\mathbb{E}_{\epsilon|\mathbf{X}}[\epsilon|\mathbf{X}] = \mathsf{0}_n \qquad\qquad \mathbb{V}\mathrm{ar}_{\epsilon|\mathbf{X}}[\epsilon|\mathbf{X}] = \sigma^2 \mathbf{I}_n.$$

The intercept

We usually consider including the 'special' predictor

$$x_{i0} \equiv 1 \qquad i = 1, \ldots, n$$

and specify the model

$$\mathbb{E}_{Y_i|\mathbf{x}_i}[Y_i|\mathbf{x}_i] = \mathbf{x}_i\beta = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ij} = \sum_{j=0}^{p-1} \beta_j x_{ij}$$

This model has p β parameters.

We will let p count the total number of predictors, including the intercept term.

Simple linear regression

We specify

$$\mathbb{E}_{Y_i|\mathbf{x}_i}[Y_i|\mathbf{x}_i] = \beta_0 + \beta_1 x_{i1} = [1\ x_{i1}] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \mathbf{x}_i \beta$$

with $p = 2$ parameters in the regression model.

This model posits a straight line relationship between x and y.

Least squares estimation in simple linear regression

On the basis of data (x_{i1}, y_i), $i = 1, \ldots, n$, we choose the line of best fit according to the least squares principle. We estimate parameters $\beta = (\beta_0, \beta_1)^\top$ by $\widehat{\beta}$ where

$$\widehat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1})^2 = \arg\min_{\beta} S(\beta)$$

where we may also write, in vector form,

$$S(\beta) = (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta)$$

We achieve the minimization by calculus.

The Normal Equations

We solve

$$\frac{\partial S(\beta)}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1}) = 0$$

$$\frac{\partial S(\beta)}{\partial \beta_1} = -2\sum_{i=1}^{n}x_{i1}(y_i - \beta_0 - \beta_1 x_{i1}) = 0$$

These two equations can be written

$$n\beta_0 + \beta_1 \sum_{i=1}^{n} x_{i1} = \sum_{i=1}^{n} y_i$$

$$\beta_0 \sum_{i=1}^{n} x_{i1} + \beta_1 \sum_{i=1}^{n} x_{i1}^2 = \sum_{i=1}^{n} x_{i1} y_i$$

or, in matrix form

$$(\mathbf{X}^\top \mathbf{X})\beta = \mathbf{X}^\top \mathbf{y}$$

The Normal Equations

These equations are termed the *Normal Equations*. If the symmetric $p \times p = 2 \times 2$ matrix

$$\mathbf{X}^\top \mathbf{X}$$

is non-singular, then we may write the solution

$$\widehat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

which yields a $p \times 1 = 2 \times 1$ vector of least squares estimates.

Explicitly, we have

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \begin{bmatrix} n & \sum x_{i1} \\ \sum x_{i1} & \sum x_{i1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i \\ \sum x_{i1} y_i \end{bmatrix}$$

Estimates

$$\left[\begin{array}{cc} n & \sum x_{i1} \\ \sum x_{i1} & \sum x_{i1}^2 \end{array}\right]^{-1} = \frac{1}{n\sum x_{i1}^2 - \{\sum x_{i1}\}^2}\left[\begin{array}{cc} \sum x_{i1}^2 & -\sum x_{i1} \\ -\sum x_{i1} & n \end{array}\right]$$

We write

$$\begin{aligned} S_{xx} &= \sum x_{i1}^2 - \left\{\sum x_{i1}\right\}^2 = \sum (x_{i1} - \bar{x}_1)^2 \\ S_{xy} &= \sum x_{i1} y_i - \frac{1}{n}\left\{\sum x_{i1} \sum y_i\right\} = \sum y_i (x_{i1} - \bar{x}_1) \end{aligned}$$

Thus, after some algebra

$$\begin{aligned} \widehat{\beta}_0 &= \bar{y} - \widehat{\beta}_1 \bar{x}_1 \\ \widehat{\beta}_1 &= \frac{S_{xy}}{S_{xx}} \end{aligned}$$

Residuals and fitted values

Define for $i = 1, \ldots, n$,

$$e_i = y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_{i1}) = y_i - \widehat{y}_i$$

- e_i – ith residual
- $\widehat{y}_i$ – ith fitted value.

Statistical properties of least squares estimators

It is evident from the formula

$$\widehat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \mathbf{A}\mathbf{y}$$

say, where

$$\mathbf{A} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$$

that the least squares estimates are merely linear combinations of the observed responses $\mathbf{y} = (y_1, \ldots, y_n)^\top$.

Specifically in the simple linear regression

$$\widehat{\beta}_0 = \sum \left(\frac{1}{n} - \bar{x}_1 c_i \right) y_i \qquad \widehat{\beta}_1 = \sum c_i y_i$$

where, for $i = 1, \ldots, n$,

$$c_i = \frac{x_{i1} - \bar{x}_1}{S_{xx}}.$$

Statistical properties of the estimators

In random variable form, we have the estimators

$$\widehat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \mathbf{A}\mathbf{Y}$$

and thus, under the model assumptions

$$\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathbf{Y}|\mathbf{X}] = \mathbf{X}\beta$$

and

$$\mathbb{V}\mathrm{ar}_{\mathbf{Y}|\mathbf{X}}[\mathbf{Y}|\mathbf{X}] = \sigma^2 \mathbf{I}_n$$

we can study distributional properties of the estimators.

Statistical properties of the estimators (cont.)

We have, using elementary properties of expectation and variance,

$$\begin{aligned}
\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\widehat{\beta}|\mathbf{X}] &= \beta & (p \times 1) \\
\mathbb{V}\mathrm{ar}_{\mathbf{Y}|\mathbf{X}}[\widehat{\beta}|\mathbf{X}] &= \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1} & (p \times p)
\end{aligned}$$

with $p = 2$. Explicitly

$$\begin{aligned}
\mathbb{V}\mathrm{ar}_{\mathbf{Y}|\mathbf{X}}[\widehat{\beta}_0|\mathbf{X}] &= \sigma^2 \frac{\sum x_{i1}^2}{n S_{xx}} = \sigma^2 \left(\frac{1}{n} + \frac{(\overline{x}_1)^2}{S_{xx}} \right) \\
\mathbb{V}\mathrm{ar}_{\mathbf{Y}|\mathbf{X}}[\widehat{\beta}_1|\mathbf{X}] &= \frac{\sigma^2}{S_{xx}}
\end{aligned}$$

Residuals and Fitted values

(i) $\mathbf{1}_n^\top \mathbf{e} = \sum_{i=1}^n e_i = 0.$

so that

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \widehat{y}_i.$$

(ii) $\underset{\sim}{\mathbf{x}}_1^\top \mathbf{e} = \sum_{i=1}^n x_{i1} e_i = 0$

(iii) $\widehat{\mathbf{y}}^\top \mathbf{e} = \sum_{i=1}^n \widehat{y}_i e_i = 0,$

that is, the observed residual vector $\mathbf{e}$ is orthogonal to the observed $n \times 1$ vectors

$$\underset{\sim}{\mathbf{x}}_1 = (x_{11}, \ldots, x_{n1})^\top$$

and

$$\widehat{\mathbf{y}} = (\widehat{y}_1, \ldots, \widehat{y}_n)^\top.$$

Estimating σ^2

Let

$$\begin{aligned}
\mathrm{SS_{Res}} &= \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \\
&= \sum_{i=1}^{n} y_i^2 - n(\overline{y})^2 - \widehat{\beta}_1 S_{xy} \\
&= \mathrm{SS_T} - \widehat{\beta}_1 S_{xy}
\end{aligned}$$

say, where

$$\mathrm{SS_T} = \sum_{i=1}^{n} y_i^2 - n(\overline{y})^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

Estimating σ^2 (cont.)

We study the statistical properties of the random variable

$$\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - \mathbf{x}_i\widehat{\beta})^2 = (\mathbf{Y} - \mathbf{X}\widehat{\beta})^\top(\mathbf{Y} - \mathbf{X}\widehat{\beta})$$

where

$$\widehat{\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$$

is the vector of least squares estimators.

Estimating σ^2 (cont.)

But

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\beta} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} = \mathbf{H}\mathbf{Y}$$

say, where

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$$

is the 'hat matrix'.

We can show that $\mathbf{H}$ is symmetric, and that

$$\mathbf{H}^\top\mathbf{H} = \mathbf{H}$$

so $\mathbf{H}$ is idempotent.

Estimating σ^2 (cont.)

Now consider the simpler model where dependence on x_{i1} is omitted, and we merely have an intercept term. Predictions in this model use the $(n \times 1)$ matrix

$$\mathbf{X} = (1, 1, \ldots, 1)^\top = \mathbf{1}_n$$

yielding the corresponding hat matrix

$$\mathbf{H}_1 = \mathbf{1}_n(\mathbf{1}_n^\top \mathbf{1}_n)^{-1}\mathbf{1}_n^\top = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$$

which is merely the $(n \times n)$ matrix with all elements equal to $1/n$.

Estimating σ^2 (cont.)

We have that

$$\mathrm{SS}_{\mathrm{Res}} = (\mathbf{Y} - \mathbf{X}\widehat{\beta})^\top(\mathbf{Y} - \mathbf{X}\widehat{\beta}) = \mathbf{Y}^\top(\mathbf{I}_n - \mathbf{H})\mathbf{Y}$$

where the $(n \times n)$ matrix $(\mathbf{I}_n - \mathbf{H})$ is symmetric and idempotent.

Now, we have the sum of squares decomposition

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 + \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2$$

or

$$\mathrm{SS}_{\mathrm{T}} = \mathrm{SS}_{\mathrm{Res}} + \mathrm{SS}_{\mathrm{R}}$$

Similarly to the previous result we have

$$\mathrm{SS_T} = \mathbf{Y}^\top(\mathbf{I}_n - \mathbf{H}_1)\mathbf{Y}$$

and

$$\mathrm{SS_R} = \mathbf{Y}^\top(\mathbf{H} - \mathbf{H}_1)\mathbf{Y}$$

yielding the representation

$$\mathbf{Y}^\top(\mathbf{I}_n - \mathbf{H}_1)\mathbf{Y} = \mathbf{Y}^\top(\mathbf{I}_n - \mathbf{H})\mathbf{Y} + \mathbf{Y}^\top(\mathbf{H} - \mathbf{H}_1)\mathbf{Y}$$

where the $(n \times n)$ matrices $(\mathbf{I}_n - \mathbf{H}_1)$ and $(\mathbf{H} - \mathbf{H}_1)$ are also symmetric and idempotent.

Estimating σ^2 (cont.)

Using the result for the expectation of a quadratic form that if $\mathbf{V}$ is a k-dimensional random vector with

$$\mathbb{E}[\mathbf{V}] = \mu \qquad \mathbb{V}\mathrm{ar}[\mathbf{V}] = \Sigma$$

then for $k \times k$ matrix $\mathbf{A}$, we have

$$\mathbb{E}[\mathbf{V}^\top \mathbf{A} \mathbf{V}] = \mathrm{trace}(\mathbf{A}\Sigma) + \mu^\top \mathbf{A} \mu$$

it follows that

$$\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathbf{Y}^\top (\mathbf{I}_n - \mathbf{H}) \mathbf{Y}] \;\; = \;\; (n-p)\sigma^2$$

Hence an unbiased estimator of σ^2 is

$$\widehat{\sigma}^2 = \frac{\mathrm{SS}_{\mathrm{Res}}}{n-p} = \mathrm{MS}_{\mathrm{Res}}$$

with $p = 2$.

Estimating σ^2 (cont.)

Using similar methods

$$\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathrm{SS}_{\mathrm{R}}|\mathbf{X}] = \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathbf{Y}^\top(\mathbf{H} - \mathbf{H}_1)\mathbf{Y}|\mathbf{X}] = (p-1)\sigma^2 + \beta_1^2 S_{xx}$$

and

$$\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathrm{SS}_{\mathrm{T}}|\mathbf{X}] = \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathbf{Y}^\top(\mathbf{I}_n - \mathbf{H}_1)\mathbf{Y}|\mathbf{X}] = (n-1)\sigma^2 + \beta_1^2 S_{xx}$$

Standard errors for the estimators

We have that

$$\mathbb{V}\mathrm{ar}_{\mathbf{Y}|\mathbf{X}}[\widehat{\beta}|\mathbf{X}] = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$$

which is estimated by

$$\widehat{\sigma}^2(\mathbf{X}^\top\mathbf{X})^{-1}.$$

The standard errors of the estimators are estimated by the square roots of the diagonal elements of this matrix; denote them by

$$\mathrm{e.s.e}(\widehat{\beta}_j) \quad j = 0, 1.$$

Hypothesis Testing

We can formulate hypothesis tests for the parameters provided we make the normality assumption

$$\epsilon | \mathbf{X} \sim \text{Normal}(\mathtt{0}_n, \sigma^2 \mathbf{I}_n).$$

For $j = 0, 1$, to test

$$\begin{aligned} & \mathrm{H}_0 \quad : \quad \beta_j = 0 \\ \text{vs} \quad & \mathrm{H}_1 \quad : \quad \beta_j \neq 0 \end{aligned}$$

we use the test statistic

$$t_j = \frac{\widehat{\beta}_j}{\text{e.s.e}(\widehat{\beta}_j)}.$$

Hypothesis Testing (cont.)

If H_0 is true, we have by standard distributional results that corresponding statistic

$$T_j \sim \text{Student}(n-p)$$

with $p = 2$. We reject H_0 at significance level α if

$$|t_j| > t_{\alpha/2,n-p}$$

where $t_{\alpha,\nu}$ is the $1-\alpha$ quantile of the Student-t distribution with ν degrees of freedom.

Confidence Intervals

A $(1-\alpha) \times 100\%$ confidence interval for β_j is

$$\widehat{\beta}_j \pm t_{\alpha/2, n-p} \times \text{e.s.e}(\widehat{\beta}_j) \qquad j = 0, 1.$$

Global Model Adequacy

The R^2 statistic is defined by

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}$$

and is a measure of the global adequacy of x as a predictor of y.

The adjusted R^2 statistic is defined by

$$R^2_{Adj} = 1 - \frac{SS_{Res}/(n-p)}{SS_T/(n-1)}$$

and is a measure that acknowledges that SS_{Res} decreases in expectation as p increases.

Local Model Adequacy

Residual plots are used to assess 'local' model adequacy.

If the model assumptions are correct, then the residual plots

- e_i vs i
- e_i vs x_{i1}
- e_i vs $\widehat{y}_i$

should be 'patternless' that is, should not exhibit systematic patterns in either mean-level or variability.

The residuals should form a horizontal 'band' around zero, with equal variability around zero everywhere.

Prediction

Predictions from the model at value of x are formed by using the estimated regression coefficients; at $x = x_{i1}$ observed in the sample, we have the prediction equal to the fitted value

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1}.$$

In vector form, we have

$$\widehat{\mathbf{y}} = \mathbf{X}\widehat{\beta}.$$

At $x = x_1^{\text{new}}$, we have the prediction

$$\widehat{y}^{\text{new}} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1^{\text{new}}.$$

Confidence and Prediction Intervals

In the random variable form we have predictions

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\beta} = \mathbf{H}\mathbf{Y}$$

so that

$$\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\widehat{\mathbf{Y}}|\mathbf{X}] = \mathbf{X}\beta$$

and

$$\mathbb{V}\text{ar}_{\mathbf{Y}|\mathbf{X}}[\widehat{\mathbf{Y}}|\mathbf{X}] = \sigma^2\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \sigma^2\mathbf{H}$$

Therefore a $(1-\alpha) \times 100\%$ **confidence interval** for the prediction at $x = x_{i1}$ is

$$\widehat{y}_i \pm t_{\alpha/2,n-p} \times \sqrt{\widehat{\sigma}^2 h_{ii}}$$

where h_{ii} is the (i,i)th diagonal element of $\mathbf{H}$.

Confidence and Prediction Intervals (cont.)

For a prediction at $x = x_1^{\text{new}}$, we have that

$$\mathbb{V}\text{ar}_{\mathbf{Y}|\mathbf{X}}[\widehat{Y}^{\text{new}}|\mathbf{X}] = \sigma^2 x^{\text{new}}(\mathbf{X}^\top \mathbf{X})^{-1}(x^{\text{new}})^\top = \sigma^2 h^{\text{new}}$$

and a $(1-\alpha) \times 100\%$ **confidence interval** for the prediction at $x = x_1^{\text{new}}$ is

$$\widehat{y}^{\text{new}} \pm t_{\alpha/2,n-p} \times \sqrt{\widehat{\sigma}^2 h^{\text{new}}}$$

Confidence and Prediction Intervals (cont.)

A **prediction interval** at $x = x_1^{\text{new}}$ incorporates the random variation that is present in the observations. Let

$$\widehat{Y}_{\text{O}}^{\text{new}} = \widehat{Y}^{\text{new}} + \epsilon^{\text{new}}$$

where ϵ^{new} is a zero mean, variance σ^2 random residual error, independent of all other random quantities. Then

$$\begin{aligned} \mathbb{V}\text{ar}_{\mathbf{Y}|\mathbf{X}}[\widehat{Y}_{\text{O}}^{\text{new}}|\mathbf{X}] &= \mathbb{V}\text{ar}_{\mathbf{Y}|\mathbf{X}}[\widehat{Y}^{\text{new}}|\mathbf{X}] + \mathbb{V}\text{ar}_{\mathbf{Y}|\mathbf{X}}[\epsilon^{\text{new}}|\mathbf{X}] \\ &= \sigma^2 h^{\text{new}} + \sigma^2 \\ &= \sigma^2(1 + h^{\text{new}}). \end{aligned}$$

Thus a $(1 - \alpha) \times 100\%$ **prediction interval** for the prediction at $x = x_1^{\text{new}}$ is

$$\widehat{y}^{\text{new}} \pm t_{\alpha/2,n-p} \times \sqrt{\widehat{\sigma}^2(1 + h^{\text{new}})}$$

The Analysis of Variance

The sums-of-squares decomposition

$$\mathrm{SS_T} = \mathrm{SS_{Res}} + \mathrm{SS_R}$$

forms the basic component of the Analysis of Variance (ANOVA) as it describes how overall observed variability in response y ($\mathrm{SS_T}$) is decomposed into

- a component corresponding to the residual errors ($\mathrm{SS_{Res}}$) and
- a component corresponding to the regression ($\mathrm{SS_R}$).

The Analysis of Variance (cont.)

Under the assumption of Normality of residual errors,

$$\epsilon|\mathbf{X} \sim \text{Normal}(\mathtt{0}_n, \sigma^2\mathbf{I}_n),$$

and the hypothesis that $\beta_1 = \mathtt{0}$, we have the result that for the sums-of-squares random variables

$$\begin{aligned}
\frac{\mathrm{SS_T}}{\sigma^2} &= \frac{\mathbf{Y}^\top(\mathbf{I}_n - \mathbf{H}_1)\mathbf{Y}}{\sigma^2} \sim \chi^2_{n-1} \\
\frac{\mathrm{SS_{Res}}}{\sigma^2} &= \frac{\mathbf{Y}^\top(\mathbf{I}_n - \mathbf{H})\mathbf{Y}}{\sigma^2} \sim \chi^2_{n-p} \\
\frac{\mathrm{SS_R}}{\sigma^2} &= \frac{\mathbf{Y}^\top(\mathbf{H} - \mathbf{H}_1)\mathbf{Y}}{\sigma^2} \sim \chi^2_{p-1}
\end{aligned}$$

with $\mathrm{SS_{Res}}$ and $\mathrm{SS_R}$ independent.

Consequently we can show that under the hypothesis, the random variable

$$F = \frac{\mathrm{SS_R}/(p-1)}{\mathrm{SS_{Res}}/(n-p)}$$

has a Fisher-F distribution with $p-1$ and $n-p$ degrees of freedom

$$F \sim \text{Fisher}(p-1, n-p).$$

We can construct a test of $\mathrm{H_0} : \beta_1 = 0$ based on this result: we reject $\mathrm{H_0}$ at significance level α if

$$F > \mathrm{F}_{\alpha,p-1,n-p}$$

where $\mathrm{F}_{\alpha,\nu_1,\nu_2}$ is the $(1-\alpha)$ quantile of the Fisher-F distribution with ν_1 and ν_2 degrees of freedom

This test is equivalent to the test of $\mathrm{H}_0 : \beta_1 = 0$ based on the t-statistic; we have that

$$t_1^2 = \left\{ \frac{\widehat{\beta}_1}{\text{e.s.e}(\widehat{\beta}_1)} \right\}^2 = F$$

and the two-tailed test based on t_1 is equivalent to the one-tailed test based on F.