Distribution of Quadratic Forms

Quadratic Forms

If $\mathbf{A}$ is a symmetric matrix and $\mathbf{y}$ is a vector, the product $\mathbf{y^T A y} = \sum_i a_{ii}y_i^2 + \sum_{i\neq j} a_{ij}y_i y_j$ is called a **quadratic form**.

Theorem 4.4a

Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{a} = [a_1, ..., a_p]^T$. Then, $\mathbf{a^T x} = \sum_{i=1}^p a_i x_i \sim \mathcal{N}_1(\mathbf{a^T}\boldsymbol{\mu}, \mathbf{a^T}\boldsymbol{\Sigma}\mathbf{a})$.

Remark: In this form, we can show that certain sum of squares have X^2 distributions.

Remark: It can be shown that $\mathbf{y^T I y} = \mathbf{y^T}(\mathbf{I}-\frac{1}{n}\mathbf{J})\mathbf{y} + \mathbf{y^T}(\frac{1}{n}\mathbf{J})\mathbf{y}$.

Sum of Squares Properties

(1) $\mathbf{I} = (\mathbf{I}-\frac{1}{n}\mathbf{J}) + (\frac{1}{n}\mathbf{J})$, **(2)** $\mathbf{I}, \mathbf{I}-\frac{1}{n}\mathbf{J}, \frac{1}{n}\mathbf{J}$ are idempotent; and **(3)** $(\mathbf{I}-\frac{1}{n}\mathbf{J})(\frac{1}{n}\mathbf{J}) = \mathbf{O}$.

Theorem 5.2a (Mean and Variances)

If $\mathbf{y}$ is a random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\mathbf{A}$ is a symmetric matrix of constants, then $\mathbb{E}(\mathbf{y^T A y}) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\mathbf{T}\mathbf{A}\boldsymbol{\mu}$.

Theorem 5.2b-d (Mean and Variances)

If $\mathbf{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$M_{\mathbf{y^T A y}}(t) = |\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\boldsymbol{\mu}^\mathbf{T}[\mathbf{I} - (\mathbf{I}-2t\mathbf{A}\boldsymbol{\Sigma})^{-1}]\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right\}$

$\mathrm{Var}(\mathbf{y^T A y}) = 2\mathrm{tr}[(\boldsymbol{A\Sigma})^2] + 4\boldsymbol{\mu^T A\Sigma A\mu}$

$\mathrm{Cov}(\mathbf{y}, \mathbf{y^T A y}) = 2\boldsymbol{\Sigma A\mu}$

Corollary

Let $\mathbf{B}$ be a $k \times p$ matrix of constants. Then $\mathrm{Cov}(\mathbf{By}, \mathbf{y^T A y}) = 2\boldsymbol{B\Sigma A\mu}$.

Theorem 5.2e (Mean and Variances)

Let $\mathbf{v} = \begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix}$ be a partitioned random vector with mean vector and covariance vector matrix given by $\begin{bmatrix}\boldsymbol{\mu_y}\\ \boldsymbol{\mu_x}\end{bmatrix}$ and $\begin{bmatrix}\boldsymbol{\Sigma_{yy}} & \boldsymbol{\Sigma_{yx}}\\ \boldsymbol{\Sigma_{xy}} & \boldsymbol{\Sigma_{xx}}\end{bmatrix}$, respectively, where $\mathbf{y}$ is a $p \times 1$, $\mathbf{x}$ is $q \times 1$, $\boldsymbol{\Sigma_{yx}}$ is $p \times q$. Let $\mathbf{A}$ be a $q \times p$ matrix of constants. Then $\mathbb{E}(\mathbf{x^T A y}) = \mathrm{tr}\boldsymbol{(A\Sigma_{yx})} + \boldsymbol{\mu_x^T A\mu_y}$.

Central Chi-Square Distribution

Let $z_1, ..., z_n$ be iid $\mathcal{N}(0,1)$ random variables. Since the z_i's are independent, the random vector $\mathbf{z} = [z_1, ..., z_n]^T$ is distributed as $\mathcal{N}_n(0, \mathbf{I}_n)$. Furthermore, $\sum_{i=1}^n z_i^2 = \mathbf{z^T z}$ is $\chi^2(n)$.

Theorem 5.3a

If $u \sim \chi^2(n)$, then **(1)** $\mathbb{E}(u) = n$, **(2)** $\mathrm{Var}(u) = 2n$, and **(3)** $M_u(t) = \frac{1}{(1-2t)^{n/2}}$.

Non-Central Chi-Square Distribution

Let $y_1, ..., y_n$ be random variables independently distributed as $\mathcal{N}(\mu_i, 1)$, so that $\mathbf{y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{I}_n)$, where $\boldsymbol{\mu} = [\mu_1, ..., \mu_n]^T$. The density of $v = \sum_{i=1}^n y_i^2 = \mathbf{y^T y}$ is called the **non-central chi-square distribution** and is denoted by $\chi^2(n, \lambda)$, where λ is the noncentrality parameter given by $\lambda = \frac{1}{2}\sum_{i=1}^n \mu_i^2 = \frac{1}{2}\boldsymbol{\mu^T\mu}$. **Remark: λ is not an eigenvalue.**

Theorem 5.3b

If $v \sim \chi^2(n, \lambda)$, then **(1)** $\mathbb{E}(v) = n + 2\lambda$, **(2)** $\mathrm{Var}(v) = 2n + 8\lambda$, and **(3)** $M_v(t) = \frac{1}{(1-2t)^{n/2}}\exp\left\{-\lambda\left(1 - \frac{1}{1-2t}\right)\right\}$

Remark: The mean of $v = \sum_{i=1}^n y_i^2$ is greater than the mean of $u = \sum_{i=1}^n (y_i - \mu_i)^2$.

Theorem 5.3c

If $v_1, \ldots, v_k$ are independently distributed as $\chi^2(n_i, \lambda_i)$, then $\sum\limits_{i=1}^k v_i$ is distributed as $\chi^2\left(\sum\limits_{i=1}^k n_i, \sum\limits_{i=1}^k \lambda_i\right)$.

Non-Central Chi-Square Distribution
Corollary If $\lambda = 0$, then $\mathbb{E}(v)$, $\mathrm{Var}(v)$, and $M_v(t)$ reduce to $\mathbb{E}(u)$, $\mathrm{Var}(u)$, and $M_u(t)$ for the central chi-square distribution. Thus, $\chi^2(n,0) = \chi^2(n)$
Corollary If $u_1, \ldots, u_k$ are independently distributed as $\chi^2(n_i)$, then $\sum_{i=1}^k u_i$ is distributed as $\chi^2(\sum_{i=1}^k n_i)$
Central F-distribution If $u \sim \chi^2(p)$ and $v \sim \chi^2(q)$, where u and v are independent, then $w = \frac{u/p}{v/q}$ has a (central) F distribution with p and q degrees of freedom.
Central t-distribution If $z \sim \mathcal{N}(0,1)$ and $u \sim \chi^2(q)$, where z and u are independent, then $t = \frac{z}{\sqrt{u/p}}$ has a (central) t distribution with p degrees of freedom.
Theorem If $u \sim \chi^2(p)$ and $v \sim \chi^2(q)$, and u and v are independent, then $w = \frac{u/p}{v/q} \sim F(p,q)$. The mean and variance of w are given by: $\mathbb{E}(w) = \frac{q}{q-2}$, $\mathrm{Var}(w) = \frac{2q^2(p+q-2)}{p(q-1)^2(q-4)}$
Non-Central F-distribution If $u \sim \chi^2(p,\lambda)$ and $v \sim \chi^2(q)$, where u and v are independent, $z = \frac{u/p}{v/q}$ has the non-central F distribution with noncentrality parameter λ, denoted by $F(p,q,\lambda)$, where λ is the same noncentrality parameter as in the non-central chi-square distribution. Remark: The mean of z, $\mathbb{E}(z) = \frac{q}{q-2}(1 + \frac{2\lambda}{p})$, is greater than the mean for the central F distribution.
Non-Central t-distribution If $y \sim \mathcal{N}(\mu,1), u \sim \chi^2(p)$, where y and u are independent, then $t = \frac{y}{\sqrt{u/p}}$ is said to have a non-central t distribution with p degrees of freedom and noncentrality parameter μ, denoted by $t(p,\mu)$. Furthermore, if $y \sim \mathcal{N}(\mu,\sigma^2)$, then $t = \frac{y/\sigma}{\sqrt{u/p}} \sim t(p,\frac{\mu}{\sigma})$ and $\frac{y}{\sigma} \sim \mathcal{N}(\frac{\mu}{\sigma},1)$
Distribution of Quadratic Forms - Theorem 5.5 Let $y \sim \mathcal{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma}), \mathbf{A}$ be a symmetric matrix of constants of rank r, and $\lambda = \frac{1}{2}\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$. Then, $\mathbf{y^TAy} \sim \chi^2(r,\lambda)$ iff $A\boldsymbol{\Sigma}$ is idempotent.
Independence of Linear and Quadratic Forms - Theorem 5.6 Suppose that $\mathbf{B}$ is a $k \times p$ matrix of constants, $\mathbf{A}$ is a $p \times p$ symmetric matrix of constants, and $y \sim \mathcal{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Then $\mathbf{By}$ and $\mathbf{y^TAy}$ are independent iff $\mathbf{B\Sigma A} = \mathbf{O}$. Let $\mathbf{A}$ and $\mathbf{B}$ be symmetric matrix of constants. If $y \sim \mathcal{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, then $\mathbf{y^TAy}$ and $\mathbf{y^TBy}$ are independent iff $\mathbf{A\Sigma B} = \mathbf{O}$.
Multiple Linear Regression Model $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$ Remark: To estimate β_i's, we will use a sample of n observations on y and the associated x variables. The model for the i^{th} observation is: $y_i = \beta_0 + \beta_1 x_i 1 + \cdots + \beta_p x_i p + \varepsilon_i$, $i = 1, ..., n$
Assumptions $\mathbb{E}(\varepsilon_i) = 0$ for $i = 1, \ldots, n$, or equivalently, $\mathbb{E}(y_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ $\mathrm{Var}(\varepsilon_i) = \sigma^2$ for $i = 1, \ldots, n$, or equivalently, $\mathrm{Var}(y_i) = \sigma^2$ $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for all $i \neq j$, or equivalently, $\mathrm{Cov}(y_i, y_j) = 0$

Linear Regression Model

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \cdots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \cdots + \beta_p x_{2p} + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \cdots + \beta_p x_{np} + \varepsilon_n \end{cases} \Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Multiple Linear Regression Model

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad ; \mathbb{E}(\varepsilon) = \mathbf{0},\ \mathrm{Var}(\varepsilon) = \sigma^2 \mathbf{I}.$$

Remark: Equivalently, $\mathbb{E}(\mathbf{y}) = \mathbf{X}\beta, \quad \mathrm{Var}(\mathbf{y}) = \sigma^2 \mathbf{I}$.

Remark:

1. We generally assume $n > p + 1$ and $\mathrm{rank}(\mathbf{X}) = p + 1$.
2. If $n < p + 1$ or if there's a linear relationship among the x's, then $\mathbf{X}$ will not have full column rank.
3. If the values of x_{ij}'s are planned/chosen, then $\mathbf{X}$ contains the experimental design and is sometimes called the **design matrix**.
4. The β parameters are called the **(partial) regression coefficients**.

Estimating : Least Squares Approach

Goal: Find estimators that minimize the sum of the squared deviations of the n observed y's from their predicted values $\hat{y}$.

We seek the estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ that minimizes $S(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) = \|\mathbf{y} - \mathbf{X}\beta\|^2$

Remark: The predicted value $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$ estimates $\mathbb{E}(y_i)$, not y_i.

OLS Estimator: Theorem 7.3a

If $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times (p+1)$ of rank $(p+1) < n$, then the value of $\hat{\boldsymbol{\beta}} = \left[\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p\right]^\top$ that minimizes the sum of the squared deviations $S(\boldsymbol{\beta})$ is $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y}$.

Remark:

1. Note that $\mathbf{X}^\mathbf{T}\mathbf{X}$ contains products of columns of $\mathbf{X}$, while $\mathbf{X}^\mathbf{T}\mathbf{y}$ contains products of the columns of $\mathbf{X}$ and $\mathbf{y}$

$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} n & \sum_i x_{i1} & \sum_i x_{i2} & \cdots & \sum_i x_{ip} \\ \sum_i x_{i1} & \sum_i x_{i1}^2 & \sum_i x_{i1} x_{i2} & \cdots & \sum_i x_{i1} x_{ip} \\ \sum_i x_{i2} & \sum_i x_{i1} x_{i2} & \sum_i x_{i2}^2 & \cdots & \sum_i x_{i2} x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_i x_{ip} & \sum_i x_{i1} x_{ip} & \sum_i x_{i2} x_{ip} & \cdots & \sum_i x_{ip}^2 \end{bmatrix} \quad \mathbf{X}^\top \mathbf{y} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_{i1} y_i \\ \vdots \\ \sum_i x_{ip} y_i \end{bmatrix}$$

2. If $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y}$, then $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \hat{\mathbf{y}}$ is known as the **residual vector**.
3. The residual vector estimates ε in the model and can be used to check the assumptions of the model.
4. We can also write $\hat{\boldsymbol{\varepsilon}}$ as $\hat{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X}\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y} = \left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top\right] \mathbf{y} = \mathbf{H}\mathbf{y}$, where $\mathbf{H} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top$ is both symmetric and idempotent.

Properties of $\hat{\beta}$: Theorem 7.3b-c

1. If $\mathbb{E}(\mathbf{y}) = \mathbf{X}\beta$, then $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for β.
2. If $\mathrm{Var}(\mathbf{y}) = \sigma^2\mathbf{I}$, the covariance matrix for $\hat{\boldsymbol{\beta}}$ is given by: $\sigma^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$.

Theorem (7.3d Gauss-Markov Theorem)

Let $\mathbf{c}^\top\boldsymbol{\beta}$ be a linear function of $\boldsymbol{\beta}$, where $\mathbf{c}$ is a $(p+1)\times 1$ vector of constants. If $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{Var}(\mathbf{y}) = \sigma^2\mathbf{I}$, then the best linear unbiased estimator (BLUE) of $\mathbf{c}^\top\boldsymbol{\beta}$ (that is, with minimum variance) is $\mathbf{c}^\top\hat{\boldsymbol{\beta}}$.

Normal Model

Recall: The OLS estimator $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{y}$ has no assumptions on the distribution of $\mathbf{y}$

Assumption: $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{I})$, or equivalently, $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$.

Maximum Likelihood Estimator: Theorem 7.6a

If $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ (or equivalently, $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \sigma^2\mathbf{I})$), where $\mathbf{X}$ is $n \times (p+1)$ of rank $(p+1) < n$, the maximum likelihood estimator of $\boldsymbol{\beta}$ and σ^2 are $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{y}$ and $\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$.

Properties of $\hat{\beta}$ and $\hat{\sigma}^2$: Theorem 7.6b

Suppose that $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times (p+1)$ of rank $(p+1) < n$ and $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_p]^\top$. Then the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\sigma}}^2$ for $\boldsymbol{\beta}$ and σ^2 in the normal model have the following distributional properties:

1. $\hat{\boldsymbol{\beta}} \sim \mathcal{N}_{p+1}\left(\beta, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\right)$.
2. $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p-1)$, or equivalently, $\frac{(n-p-1)s^2}{\sigma^2} \sim \chi^2(n-p-1)$.
3. $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ (or s^2) are independent.

Theorem (7.6c-d)

If $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, then:

1. $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are jointly sufficient for $\boldsymbol{\beta}$ and σ^2.
2. $\hat{\boldsymbol{\beta}}$ and s^2 have minimum variance among all unbiased estimators.

Corollary

If $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, then the minimum variance unbiased estimator of $\mathbf{c}^\top\boldsymbol{\beta}$ is $\mathbf{c}^\top\hat{\beta}$, where $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator for $\boldsymbol{\beta}$.

Estimating Mean Response

Let $\boldsymbol{\theta} = \mathbb{E}(\mathbf{y})$. Since $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, the best linear unbiased estimator (BLUE) for $\boldsymbol{\theta}$ is: $\hat{\boldsymbol{\theta}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$. Under the assumption that $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, we have: $\mathbf{X}\hat{\boldsymbol{\beta}} \sim \mathcal{N}_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\right)$.

Recall: Estimators

Under the assumption that $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, the MLE are: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$ $\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$.

Given values of the predictors $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$, we can estimate the value $\hat{y}_i$ of the response.

F-Test for the General Linear Hypothesis

Recall: By the Gauss-Markov Theorem, the BLUE for $c^\top\boldsymbol{\beta}$ is $c^\top\hat{\boldsymbol{\beta}}$.

Goal: Test $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ **vs.** $H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{d}$, where $\mathbf{C}$ is $k \times (p+1)$ of rank k, and $\mathbf{d} \in \mathbb{R}^k$.

Theorem (8.4a) If $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, then

- $\mathbf{C}\hat{\boldsymbol{\beta}} \sim \mathcal{N}_k\left(\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top\right)$.
- $\frac{(n-p-1)s^2}{\sigma^2} = \frac{\mathbf{y}^\top\mathbf{H}\mathbf{y}}{\sigma^2} \sim \chi^2(n-p-1, 0)$.

Remarks: Recall that $\mathbf{H} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$.

Theorem (8.4f)

If $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$,

- $\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}\right)^\top \left[\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top\right]^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}\right) \Big/ \sigma^2 \sim \chi^2(k, \lambda)$, where

$$\lambda = \frac{(\mathbf{C}\boldsymbol{\beta} - \mathbf{d})^\top \left[\mathbf{C}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}\right]^{-1} (\mathbf{C}\boldsymbol{\beta} - \mathbf{d})}{2\sigma^2} = 0 \text{ when } H_0 \text{ is true.}$$

- $\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}\right)^\top \left[\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top\right]^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}\right)$ and $\mathbf{y}^\top H\mathbf{y}$ are independent.

Remarks: $(n-p-1)s^2 = \mathbf{y}^\top H\mathbf{y}$, where $H = I - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$.

Theorem (8.4g)

Let $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. If $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ is true, then

$$F = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^\top \left[\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top\right]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})/\sigma^2}{k} \Bigg/ \frac{\mathbf{y}^\top H\mathbf{y}/\sigma^2}{n-p-1}$$

$$F = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^\top \left[\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top\right]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})}{k \cdot \mathbf{y}^\top H\mathbf{y}/(n-p-1)} \sim F(k, n-p-1, 0).$$

Generalized Linear Hypotheses

Test for the joint effect of the explanatory variables: $H_0 : \begin{bmatrix}\beta_1\\ \vdots \\ \beta_p\end{bmatrix} = \mathbf{0}$ vs $H_1 : \begin{bmatrix}\beta_1\\ \vdots \\ \beta_p\end{bmatrix} \neq \mathbf{0}$

Effect of the ith variable: $H_0 : \beta_i = 0$ vs $H_1 : \beta_i \neq 0, \quad i = 1, \dots, p$

F-Test

Decision Rule: Given significance level α, reject H_0 if $F > F_{1-\alpha}(k, n-p-1)$.

Coefficient of Determination R^2

The total sum of squares, SST, can be written as: $\text{SST} = \sum_{i=1}^n (y_i - \bar{y})^2 = \mathbf{y}^\top\mathbf{y} - n\bar{y}^2 = (\hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{y} - n\bar{y}^2) + (\mathbf{y}^\top\mathbf{y} - \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{y}) = \text{SSR} + \text{SSE}$.

Remark: $\hat{\boldsymbol{\varepsilon}}^\top\hat{\boldsymbol{\varepsilon}} = \mathbf{y}^\top\mathbf{y} - \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{y} = \text{SSE}$.

Coefficient of Determination R^2

The proportion of the total sum of squares due to regression

$$R^2 = \frac{SSR}{SST} = \frac{\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y} - n\bar{y}^2}{\mathbf{y}^T\mathbf{y} - n\bar{y}^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

is known as the **coefficient of determination**, or the squared multiple correlation.

Remark:

1. The positive square root R is called the (multiple) correlation coefficient.
2. The ratio is a measure of model fit and provides an indication of how well the x's predict y.

Properties

1. The range of R^2 is $0 \leq R^2 \leq 1$.
2. $R = r_{yy}$; that is, the multiple correlation is equal to the simple correlation.
3. Adding a variable x to the model **increases** the value of R^2.
4. If $\beta_1 = \beta_2 = \cdots = \beta_p = 0$, then $\mathbb{E}(R^2) = \frac{p}{n-1}$.

Note that the estimated coefficients $\hat{\beta}_j$ will not be zero even when the β_j's are zero.

Remarks:

1. R^2 can also be written as

$$R^2 = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^\top \left[\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top\right]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})}{(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^\top \left[\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top\right]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}) + \mathbf{y}^\top\mathbf{H}\mathbf{y}}$$

2. If all β_j's were zero, $R^2 = 1$. If $y_i = \hat{y}_i$ for all i, $R^2 = 1$.
3. When p is a relatively large fraction of n, the resulting R^2 will be large, but not meaningful. To compensate for this, the adjusted R^2 can be used instead: $R^2_{adj} = \frac{(n-1)R^2 - p}{n-p-1}$

Regression Concerns: Variable Selection

Selection Metrics

1. **Mallow's** C_p, $C_p = \frac{\text{SSR}}{s^2} - (n + 2p)$ where $R^2 = \frac{\text{SSR}}{\text{SST}}$ and $s^2 = \frac{\text{SSE}}{n-p-1} = \frac{\boldsymbol{y}^T H \boldsymbol{y}}{n-p-1}$
2. **Akaike Information Criterion (AIC)**, $\text{AIC} = -2\ln L(\hat{\beta}_p, \hat{\sigma}^2) + 2p$
3. **Bayesian Information Criterion (BIC)**, $\text{BIC} = -2\ln L(\hat{\beta}_p, \hat{\sigma}^2) + p\ln p$

Remark: Here, $L(\cdot)$ represents the likelihood function.

Regression Concerns: Multicollinearity

1. Some of the predictors are functions of one or more other predictors.
2. This problem, known as **multicollinearity**, can distort the standard error of estimate and can lead to incorrect conclusions as to which independent variables are statistically significant.
3. The **variance inflation factor (VIF)** can be used to check this. The VIF for the j^{th} predictor is the multiple R^2 when the j^{th} predictor is regressed to other predictors: $VIF_j = \frac{1}{1-R_j^2}$

   Interpretation: Retain if $VIF_j < 5$, while remove if $VIF_j > 10$.

Regression Concerns: Residuals

1. Residuals should be uncorrelated (that is, no patterns on the plot).
2. Normality can be verified using a **normal Q-Q plot**, and the statistical tests **Shapiro-Wilk** test and **Anderson-Darling** test.

   In these tests, the null hypothesis is that the data comes from a normal distribution

Motivation

1. Several treatments or treatment combinations are applied to randomly selected experimental units.
2. **Goal**: Compare treatment means

   We can use analysis-of-variance and linear models to compare means!

Illustration: Case 1: $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

In matrix form, $\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{bmatrix}$ where $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

Remarks:

1. In most ANOVA applications, there are more than two treatments considered.
2. We focus on **balanced models**: each treatment has the same number of observations.
3. The data matrix $\mathbf{X}$ is not full-rank. As such, $\mathbf{X}^\top\mathbf{X}$ is also not full rank and $(\mathbf{X}^\top\mathbf{X})^{-1}$ does not exist.
4. $\boldsymbol{\beta} = [\mu, \alpha_1, \alpha_2]^\top$ is not unique and *not estimable*. That is, individual parameters μ, α_1, α_2 cannot be *uniquely* estimated unless they are subject to constraints or side conditions.

One-Way ANOVA

The one-way balanced ANOVA model can be expressed as (Case 1) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

where $\mathbf{y} = \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_k} \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} \mathbf{j} & \mathbf{j} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{j} & \mathbf{0} & \mathbf{j} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{j} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{j} \end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}$, and $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$.

Remark: Here, $\mathbf{X}$ is $kn \times (k+1)$, k is the number of treatments, n is the number of observations per treatment, $\mathbf{y}_i = [y_{i1}, \ldots, y_{in}]^\top$ and $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \ldots, \varepsilon_{in}]^\top$, $\mathbf{j}$ and $\mathbf{0}$ are $n \times 1$, and $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are $kn \times 1$

Note: for **Case 2:** $y_{ij} = \mu_i + \varepsilon_{ij}$, $\boldsymbol{\beta} = [\mu_1, ..., \mu_k]^T$ and $\boldsymbol{X}$ does not have the first column of 1's.

Assumptions

- $\mathbb{E}(\varepsilon_{ij}) = 0$ for all i, j
- $\text{Var}(\varepsilon_{ij}) = \sigma^2$ for all i, j
- $\text{Cov}(\varepsilon_{ij}, \varepsilon_{rs}) = 0$ for all i, j
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$
- We often use the constraint (side condition) $\sum_{i=1}^k \alpha_i = 0$

Estimating β and σ^2

A generalized inverse of $\mathbf{X}^\top\mathbf{X}$ in our model is given by $(\mathbf{X}^\top\mathbf{X})^- = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{n} \end{bmatrix}$ Using this, an estimator for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^-\mathbf{X}^\top\mathbf{y} = \begin{bmatrix} 0 \\ \bar{y}_{1\cdot} \\ \vdots \\ \bar{y}_{k\cdot} \end{bmatrix}$. where $\bar{y}_{i\cdot} = \frac{1}{n}\sum_{j=1}^n y_{ij}$.

Moreover, an estimator for σ^2 is $s^2 = \frac{\text{SSE}}{k(n-1)}$, where $\text{SSE} = \mathbf{y}^\top\mathbf{y} - \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{y} = \mathbf{y}^\top[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^-\mathbf{X}^\top]\mathbf{y}$.

Remarks:

1. The rank of the idempotent matrix $[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^-\mathbf{X}^\top]$ is $k(n-1)$.
2. s^2 is an unbiased estimator of σ^2.
3. SSE can be written as

$$\text{SSE} = \mathbf{y}^\top\mathbf{y} - \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{y} = \sum_{i=1}^k\sum_{j=1}^n y_{ij}^2 - \sum_{i=1}^k \bar{y}_{i\cdot}y_{i\cdot} = \sum_{ij}(y_{ij} - \bar{y}_{i\cdot})^2 = \sum_{ij} y_{ij}^2 - \sum_i \frac{y_{i\cdot}^2}{n},$$

where $y_{i\cdot} = \sum_{j=1}^n y_{ij}$

Motivation

Goal: Test H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$ vs H_1: at least two means are unequal.

One-way ANOVA Model: Hypothesis Testing

Recall: One-way ANOVA model can be written as a linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

Using $\mu_i = \mu + \alpha_i$, the null hypothesis for the test of equality of means can be written as H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_k$. Assuming $\mathbf{y} \sim \mathcal{N}_{kn}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, a hypothesis test for H_0 can be constructed using the F-distribution.

Linear Models in ANOVA

The general linear model can be expressed as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where $\mathbf{y}$ is the vector of observations, $\mathbf{X}$ is the design matrix, $\boldsymbol{\beta}$ is the vector of parameters, and $\boldsymbol{\varepsilon}$ is the error term.

One-way ANOVA Model: General Linear Hypothesis

Theorem (12.7b)

If $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $N \times p$ of rank $k < p \leq N$, if $\mathbf{C}$ is $m \times p$ of rank $m \leq k$ such that $\mathbf{C}\boldsymbol{\beta}$ is a set of m linearly independent estimable functions, and if $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, then

1. $\mathbf{C}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{C}^T$ is nonsingular.
2. $\mathbf{C}\hat{\boldsymbol{\beta}} \sim \mathcal{N}_m(\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-}\mathbf{C}^T)$.
3. $\dfrac{SSH}{\sigma^2} = \dfrac{(\mathbf{C}\hat{\boldsymbol{\beta}})^T[\mathbf{C}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-}\mathbf{C}^T]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}})}{\sigma^2} \sim \chi^2(m, \boldsymbol{\lambda})$, where
$$\lambda = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}})^T[\mathbf{C}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-}\mathbf{C}^T]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}})}{2\sigma^2}.$$
4. $\dfrac{SSE}{\sigma^2} = \dfrac{\boldsymbol{y}^T[\mathbf{I} - \mathbf{X}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-}\mathbf{X}^T]\boldsymbol{y}}{\sigma^2} \sim \chi^2(N-k)$.
5. SSH and SSE are independent.

Theorem (12.7c)

Let $\boldsymbol{y} \sim N_N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $N \times p$ with rank k, where $k < p \leq N$. Let $\mathbf{C}$, $\mathbf{C}\boldsymbol{\beta}$, and $\hat{\boldsymbol{\beta}}$ be defined as in the previous theorem. If $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is true, then

$$F = \frac{SSH/m}{SSE/(N-k)} = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}})^T[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{C}^T]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}})/m}{SSE/(N-k)} \sim F(m, N-k, 0).$$

One-way ANOVA Model: Hypothesis Testing

Note that the hypothesis $H_0 : \alpha_1 = \cdots = \alpha_k$ can be written as

$$H_0 : \begin{bmatrix} \alpha_1 - \alpha_2 \\ \alpha_1 - \alpha_3 \\ \vdots \\ \alpha_1 - \alpha_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow H_0 : \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$$

That is, $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{C} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & 0 & 0 & \cdots & -1 \end{bmatrix}$ is $(k-1) \times (k+1)$, and $\boldsymbol{\beta} = [\mu, \alpha_1, \ldots, \alpha_k]^T$ is $(k+1) \times 1$. For **Case 2:** $y_{ij} = \mu_i + \varepsilon_{ij}$, $\mathbf{C}$ does not have the first column of 0's.

Since $y \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, by previous theorem, we can use the test statistic $F = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}})^T[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{C}^T]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}})/(k-1)}{\mathrm{SSE}/k(n-1)} \sim F(k-1, k(n-1))$ to test the hypothesis $H_0 : \alpha_1 = \cdots = \alpha_k$.

One-Way ANOVA Table

Source of Variation	df	SS	MS $= \frac{SS}{df}$	F-statistic
Treatments	$k-1$	$SSH = \frac{1}{n}\sum_i y_{i.}^2 - \frac{y_{..}^2}{kn}$	$\frac{SSH}{(k-1)}$	$F = \frac{\frac{SSH}{(k-1)}}{\frac{SSE}{k(n-1)}}$
Error	$k(n-1)$	$SSE = \sum_{ij} y_{ij}^2 - \frac{1}{n}\sum_i y_{i}.^2$	$\frac{SSE}{k(n-1)}$	
Total	$kn-1$	$SST = \sum_{ij} y_{ij}^2 - \frac{y_{..}^2}{kn}$		

Testing Contrasts

A linear contrast is any linear combination of the individual group means such that the linear coefficients sum to 0: $L = \sum_{i=1}^k c_i\alpha_i, \quad \text{where} \quad \sum_{i=1}^k c_i = 0.$

Remark:

1. For the one-way model, the contrast $\sum_i c_i\alpha_i$ is equivalent to $\sum_i c_i\mu_i$.
2. The hypothesis of interest is $H_0 : \sum_i c_i\alpha_i$ or $H_0 : \sum_i c_i\mu_i = 0$, which represents a comparison of means if $\sum_i c_i = 0$.

Note that the null hypothesis can be written as $H_0 : \mathbf{c}^T\boldsymbol{\beta}$, where $\mathbf{c} = [0, c_1, \ldots, c_k]^T$ and $\boldsymbol{\beta} = [\mu, \alpha_1, \ldots, \alpha_k]^T$. Assuming $\mathbf{y} \sim \mathcal{N}_{kn}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, H_0 can be tested using the previous theorem with $m = 1$ and test statistic $F = \frac{(\mathbf{c}^T\hat{\boldsymbol{\beta}})^T[\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{c}]^{-1}(\mathbf{c}^T\hat{\boldsymbol{\beta}})}{\mathrm{SSE}/k(n-1)} = \frac{(\mathbf{c}^T\hat{\boldsymbol{\beta}})^2}{s^2[\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{c}]} \sim F(1, k(n-1))$.