Linear Models in Statistics:
A Collection of Core Theorems

From Rencher & Schaalje (2008)

November 25, 2024

Contents

1 Foundational Matrix Theorems (Ch. 2) 2

2 Multivariate Normal Distribution Theorems (Ch. 4) 2

3 Statistical Distribution Theorems of Quadratic Forms (Ch. 5) 3

4 Simple Linear Regression (Ch. 6) 3

5 Regression Theory Theorems (Ch. 7) 5

6 Multiple Regression: Tests of Hypotheses and Confidence Intervals (Ch. 8) 8

7 Multiple Regression: Model Validation and Diagnostics (Ch. 9) 11

8 Multiple Regression: Random x's (Ch. 11) 12

9 Analysis of Variance Models (Ch. 12) 13

Preface

This document presents a comprehensive collection of theorems from fundamental concepts in linear models, organized by chapter. Each result is presented with its original theorem number from Rencher & Schaalje (2008) for easy reference.

1 Foundational Matrix Theorems (Ch. 2)

Theorem 1.1 (2.6a)**.** *(i) If* $\mathbf{A}$ *is positive definite, then all its diagonal elements* a_{ii} *are positive.*

(ii) If $\mathbf{A}$ *is positive semidefinite, then all* $a_{ii} \geq 0$*.*

Theorem 1.2 (2.6b)**.** *Let* $\mathbf{P}$ *be a nonsingular matrix.*

(i) If $\mathbf{A}$ *is positive definite, then* $\mathbf{P}^T\mathbf{AP}$ *is positive definite.*

(ii) If $\mathbf{A}$ *is positive semidefinite, then* $\mathbf{P}^T\mathbf{AP}$ *is positive semidefinite.*

Theorem 1.3 (2.13a)**.** *The only nonsingular idempotent matrix is the identity matrix* $\mathbf{I}$*.*

Theorem 1.4 (2.13b)**.** *If* $\mathbf{A}$ *is singular, symmetric, and idempotent, then* $\mathbf{A}$ *is positive semidefinite.*

Theorem 1.5 (2.13c)**.** *If* $\mathbf{A}$ *is an* $n \times n$ *symmetric idempotent matrix of rank* r*, then* $\mathbf{A}$ *has* r *eigenvalues equal to 1 and* $n - r$ *eigenvalues equal to 0.*

Theorem 1.6 (2.13d)**.** *If* $\mathbf{A}$ *is symmetric and idempotent of rank* r*, then* $rank(\mathbf{A}) = tr(\mathbf{A}) = r$*.*

Theorem 1.7 (2.13e)**.** *If* $\mathbf{A}$ *is an* $n \times n$ *idempotent matrix,* $\mathbf{P}$ *is an* $n \times n$ *nonsingular matrix, and* $\mathbf{C}$ *is an* $n \times n$ *orthogonal matrix, then:*

(i) $\mathbf{I} - \mathbf{A}$ *is idempotent*

(ii) $\mathbf{A}(\mathbf{I} - \mathbf{A}) = \mathbf{O}$ *and* $(\mathbf{I} - \mathbf{A})\mathbf{A} = \mathbf{O}$

(iii) $\mathbf{P}^{-1}\mathbf{AP}$ *is idempotent*

(iv) $\mathbf{C}^T\mathbf{AC}$ *is idempotent*

2 Multivariate Normal Distribution Theorems (Ch. 4)

Theorem 2.1 (4.3)**.** *If* $\mathbf{y}$ *is distributed as* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, its moment generating function is given by:* $[M_{\mathbf{y}}(\mathbf{t}) = e^{\mathbf{t}^T\boldsymbol{\mu} + \mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}/2}]$

Theorem 2.2 (4.3a)**.** *If* u *is distributed as* $\chi^2(n)$*, then:*

1. $E(u) = n$
2. $var(u) = 2n$
3. $M_u(t) = \frac{1}{(1-2t)^{n/2}}$

Theorem 2.3 (4.3b)**.** *If* v *is distributed as* $\chi^2(n, \lambda)$*, then:*

1. $E(v) = n + 2\lambda$
2. $var(u) = 2n + 8\lambda$
3. $M_v(t) = \frac{1}{(1-2t)^{n/2}} e^{-\lambda(1 - \frac{1}{(1-2t)})}$

Theorem 2.4 (4.3c)**.** *If* $v_1, v_2, ..., v_k$ *are independently distributed as* $\chi^2(n_i, \lambda_i)$*, then:* $\sum_{i=1}^k v_i$ *is distributed as* $\chi^2(\sum_{i=1}^k n_i, \sum$

Theorem 2.5 (4.4a)**.** *Let the* $p \times 1$ *random vector* $\mathbf{y}$ *be* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, let* $\mathbf{a}$ *be any* $p \times 1$ *vector of constants, and let* $\mathbf{A}$ *be any* $k \times p$ *matrix of constants with rank* $k \leq p$*. Then: (i)* $z = \mathbf{a}^T\mathbf{y}$ *is* $N(\mathbf{a}^T\boldsymbol{\mu}, \mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})$ *(ii)* $\mathbf{z} = \mathbf{Ay}$ *is* $N_k(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$

Theorem 2.6 (4.4b)**.** *If* $\mathbf{y}$ *is* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, then any* $r \times 1$ *subvector of* $\mathbf{y}$ *has an* r*-variate normal distribution with the same means, variances, and covariances as in the original* p*-variate normal distribution.*

Theorem 2.7 (4.4c)**.** *If* $\mathbf{v} = \begin{pmatrix}\mathbf{y} & \mathbf{x}\end{pmatrix}$ *is* $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, then* $\mathbf{y}$ *and* $\mathbf{x}$ *are independent if* $\boldsymbol{\Sigma}_{yx} = \mathbf{O}$*.*

Theorem 2.8 (4.4d)**.** *If* $\mathbf{y}$ *and* $\mathbf{x}$ *are jointly multivariate normal with* $\boldsymbol{\Sigma} yx \neq \mathbf{O}$*, then the conditional distribution of* $\mathbf{y}$ *given* $\mathbf{x}$*,* $f(\mathbf{y}|\mathbf{x})$*, is multivariate normal with mean vector and covariance matrix:* $[E(\mathbf{y}|\mathbf{x}) = \boldsymbol{\mu} y + \boldsymbol{\Sigma} yx \boldsymbol{\Sigma} xx^{-1}(\mathbf{x} - \boldsymbol{\mu} x)][cov(\mathbf{y}|\mathbf{x}) = \boldsymbol{\Sigma} yy - \boldsymbol{\Sigma} yx \boldsymbol{\Sigma} xx^{-1} \boldsymbol{\Sigma}_{xy}]$

Theorem 2.9 (4.4e)**.** *If* $\mathbf{y}$ *and* $\mathbf{x}$ *are jointly multivariate normal with mean vector and covariance matrix:* $[E(\mathbf{y}, \mathbf{x}) = \begin{pmatrix}\boldsymbol{\mu} y & \boldsymbol{\mu} x\end{pmatrix}, \quad cov(\mathbf{y}, \mathbf{x}) = \begin{pmatrix}\boldsymbol{\Sigma} yy & \boldsymbol{\Sigma} yx \, \boldsymbol{\Sigma} xy & \boldsymbol{\Sigma} xx\end{pmatrix}]$ *Then the conditional distribution of* y *given* $\mathbf{x}$ *is normal with:* $[E(y|\mathbf{x}) = \mu_y + \boldsymbol{\sigma} yx^T \boldsymbol{\Sigma} xx^{-1}(\mathbf{x} - \boldsymbol{\mu} x)][var(y|\mathbf{x}) = \sigma_y^2 - \boldsymbol{\sigma} yx^T \boldsymbol{\Sigma} xx^{-1} \boldsymbol{\sigma} yx]$

3 Statistical Distribution Theorems of Quadratic Forms (Ch. 5)

Theorem 3.1 (5.2a)**.** *Let* $\mathbf{y}$ *be a random vector with mean* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$*, and let* $\mathbf{A}$ *be a symmetric matrix of constants. Then:*

$$E(\mathbf{y}^T \mathbf{A} \mathbf{y}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}$$

Theorem 3.2 (5.2b)**.** *If* $\mathbf{y}$ *is* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, then the moment generating function of* $\mathbf{y}^T\mathbf{A}\mathbf{y}$ *is:*

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

Theorem 3.3 (5.2c)**.** *If* $\mathbf{y}$ *is* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, then:*

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

Theorem 3.4 (5.2d)**.** *If* $\mathbf{y}$ *is* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, then:*

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

Theorem 3.5 (5.5)**.** *Let* $\mathbf{y}$ *be distributed as* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, let* $\mathbf{A}$ *be a symmetric matrix of constants of rank* r*, and let* $\lambda = \frac{1}{2}\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$*. Then* $\mathbf{y}^T\mathbf{A}\mathbf{y}$ *is* $\chi^2(r, \lambda)$ *if and only if* $\mathbf{A}\boldsymbol{\Sigma}$ *is idempotent.*

Theorem 3.6 (5.6a)**.** *Suppose that* $\mathbf{B}$ *is a* $k \times p$ *matrix of constants,* $\mathbf{A}$ *is a* $p \times p$ *symmetric matrix of constants, and* $\mathbf{y}$ *is distributed as* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*. Then* $\mathbf{B}\mathbf{y}$ *and* $\mathbf{y}^T\mathbf{A}\mathbf{y}$ *are independent if and only if* $\mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{O}$*.*

Corollary 3.7 (to Theorem 5.6a)**.** *If* $\mathbf{y}$ *is* $N_p(\boldsymbol{\mu}, \sigma^2\mathbf{I})$*, then* $\mathbf{B}\mathbf{y}$ *and* $\mathbf{y}^T\mathbf{A}\mathbf{y}$ *are independent if and only if* $\mathbf{B}\mathbf{A} = \mathbf{O}$*.*

Theorem 3.8 (5.6b)**.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be symmetric matrices of constants. If* $\mathbf{y}$ *is* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, then* $\mathbf{y}^T\mathbf{A}\mathbf{y}$ *and* $\mathbf{y}^T\mathbf{B}\mathbf{y}$ *are independent if and only if* $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{O}$*.*

Theorem 3.9 (5.6c)**.** *Let* $\mathbf{y}$ *be* $N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$*, let* $\mathbf{A}_i$ *be symmetric of rank* r_i *for* $i = 1, 2, \ldots, k$*, and let* $\mathbf{y}^T\mathbf{A}\mathbf{y} = \sum_{i=1}^k \mathbf{y}^T\mathbf{A}_i\mathbf{y}$*, where* $\mathbf{A} = \sum_{i=1}^k \mathbf{A}_i$ *is symmetric of rank* r*. Then:*
(i) $\mathbf{y}^T\mathbf{A}_i\mathbf{y}/\sigma^2$ *is* $\chi^2(r_i, \boldsymbol{\mu}^T\mathbf{A}_i\boldsymbol{\mu}/2\sigma^2)$*,* $i = 1, 2, \ldots, k$
(ii) $\mathbf{y}^T\mathbf{A}_i\mathbf{y}$ *and* $\mathbf{y}^T\mathbf{A}_j\mathbf{y}$ *are independent for all* $i \neq j$
(iii) $\mathbf{y}^T\mathbf{A}\mathbf{y}/\sigma^2$ *is* $\chi^2(r, \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}/2\sigma^2)$

4 Simple Linear Regression (Ch. 6)

Theorem 4.1 (6.1 - Linear Regression Model)**.** *The simple linear regression model is given by:*

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n$$

with assumptions:

1. *$E(\epsilon_i) = 0$ for all i, or equivalently, $E(y_i) = \beta_0 + \beta_1 x_i$*

2. $Var(\epsilon_i) = \sigma^2$ *for all* i*, or equivalently,* $Var(y_i) = \sigma^2$

3. $Cov(\epsilon_i, \epsilon_j) = 0$ *for all* $i \neq j$*, or equivalently,* $Cov(y_i, y_j) = 0$

Theorem 4.2 (6.2 - Least Squares Estimators)**.** *The least squares estimators that minimize* $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ *are:*

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

These estimators have the following properties:

$$E(\hat{\beta}_1) = \beta_1$$

$$E(\hat{\beta}_0) = \beta_0$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

Theorem 4.3 (6.3 - Error Sum of Squares)**.** *The error sum of squares can be expressed as:*

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

The unbiased estimator of σ^2 *is:*

$$s^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}$$

with

$$E(s^2) = \sigma^2$$

Theorem 4.4 (6.4 - Distributional Results)**.** *Under normality assumption* $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$*:*

1. $\hat{\beta}_1$ *is* $N(\beta_1, \sigma^2 / \sum_{i=1}^n (x_i - \bar{x})^2)$

2. $(n-2)s^2/\sigma^2$ *is* $\chi^2(n-2)$

3. $\hat{\beta}_1$ *and* s^2 *are independent*

4. *The test statistic*

$$t = \frac{\hat{\beta}_1}{s/\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

   follows $t(n-2, \delta)$ *with* $\delta = \beta_1 / [s/\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}]$

Theorem 4.5 (6.5 - Confidence Interval)**.** *A* $100(1-\alpha)\%$ *confidence interval for* β_1 *is given by:*

$$\hat{\beta}_1 \pm t_{\alpha/2, n-2} s / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Theorem 4.6 (6.6 - Coefficient of Determination)**.** *The coefficient of determination* r^2 *is defined as:*

$$r^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where the total sum of squares can be partitioned as:

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

or

$$SST = SSR + SSE$$

The t statistic can be expressed in terms of r as:

$$t=\frac{\hat{\beta}_1}{s/\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}=\sqrt{n-2}\frac{r}{\sqrt{1-r^2}}$$

5 Regression Theory Theorems (Ch. 7)

Theorem 5.1 (7.3a)**.** *If $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$, where $\mathbf{X}$ is $n\times(k+1)$ of rank $k+1<n$, then the value of $\hat{\boldsymbol{\beta}}=(\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_k)'$ that minimizes (7.5) is*

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Theorem 5.2 (7.3b)**.** *If $E(\mathbf{y})=\mathbf{X}\boldsymbol{\beta}$, then $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$.*

Theorem 5.3 (7.3c)**.** *If $cov(\mathbf{y})=\sigma^2\mathbf{I}$, the covariance matrix for $\hat{\boldsymbol{\beta}}$ is given by $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.*

Theorem 5.4 (7.3d)**.** ***(Gauss-Markov Theorem)*** *If $E(\mathbf{y})=\mathbf{X}\boldsymbol{\beta}$ and $cov(\mathbf{y})=\sigma^2\mathbf{I}$, the least-squares estimators $\hat{\beta}_j$, $j=0,1,\ldots,k$, have minimum variance among all linear unbiased estimators.*

Theorem 5.5 (7.3e)**.** *If $\mathbf{x}=(1,x_1,\ldots,x_k)'$ and $\mathbf{z}=(1,c_1x_1,\ldots,c_kx_k)'$, then $\hat{y}=\hat{\boldsymbol{\beta}}'\mathbf{x}=\hat{\boldsymbol{\beta}}_z'\mathbf{z}$, where $\hat{\boldsymbol{\beta}}_z$ is the least squares estimator from the regression of $\mathbf{y}$ on $\mathbf{z}$.*

Theorem 5.6 (7.3f)**.** *If s^2 is defined by (7.22), (7.23), or (7.24) and if $E(\mathbf{y})=\mathbf{X}\boldsymbol{\beta}$ and $cov(\mathbf{y})=\sigma^2\mathbf{I}$, then*

$$E(s^2)=\sigma^2$$

Theorem 5.7 (7.3g)**.** *If $E(\boldsymbol{\epsilon})=0$, $cov(\boldsymbol{\epsilon})=\sigma^2\mathbf{I}$, and $E(\epsilon_i^4)=3\sigma^4$ for the linear model $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$, then s^2 in (7.23) or (7.24) is the best (minimum variance) quadratic unbiased estimator of σ^2.*

Theorem 5.8 (7.4a)**.** *For the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{X}'\mathbf{y}$:*

1. *The coefficient matrix $\mathbf{X}'\mathbf{X}$ is symmetric and positive definite*
2. *The normal equations have a unique solution $\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$*
3. *The fitted values $\hat{\mathbf{y}}=\mathbf{X}\hat{\boldsymbol{\beta}}$ are unique*

Theorem 5.9 (7.4b)**.** *For any vector $\mathbf{c}$:*

1. *$\mathbf{c}'\hat{\boldsymbol{\beta}}$ is a linear function of $\mathbf{y}$*
2. *If $E(\mathbf{y})=\mathbf{X}\boldsymbol{\beta}$, then $\mathbf{c}'\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\mathbf{c}'\boldsymbol{\beta}$*
3. *If $cov(\mathbf{y})=\sigma^2\mathbf{I}$, then $var(\mathbf{c}'\hat{\boldsymbol{\beta}})=\sigma^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$*

Theorem 5.10 (7.5)**.** *The model $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$ can be transformed to $\mathbf{z}=\mathbf{W}\boldsymbol{\gamma}+\boldsymbol{\delta}$ by the transformation $\mathbf{z}=\mathbf{T}\mathbf{y}$ and $\mathbf{W}=\mathbf{T}\mathbf{X}$, where $\mathbf{T}$ is any nonsingular matrix. The estimators $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}$ are related by $\hat{\boldsymbol{\gamma}}=\hat{\boldsymbol{\beta}}$.*

Theorem 5.11 (7.6a)**.** *If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n\times(k+1)$ of rank $k+1<n$, the maximum likelihood estimators of $\boldsymbol{\beta}$ and σ^2 are*

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$\hat{\sigma}^2=\frac{1}{n}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$$

Theorem 5.12 (7.6b). *Suppose that* $\mathbf{y}$ *is* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, *where* $\mathbf{X}$ *is* $n \times (k+1)$ *of rank* $k+1 < n$ *and* $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$. *Then the maximum likelihood estimators* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\sigma}^2$ *have the following distributional properties:*

1. $\hat{\boldsymbol{\beta}}$ *is* $N_{k+1}[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}]$
2. $n\hat{\sigma}^2/\sigma^2$ *is* $\chi^2(n-k-1)$, *or equivalently,* $(n-k-1)s^2/\sigma^2$ *is* $\chi^2(n-k-1)$
3. $\hat{\boldsymbol{\beta}}$ *and* $\hat{\sigma}^2$ *(or* s^2*) are independent*

Theorem 5.13 (7.6c). *If* $\mathbf{y}$ *is* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, *then* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\sigma}^2$ *are jointly sufficient for* $\boldsymbol{\beta}$ *and* σ^2.

Theorem 5.14 (7.6d). *If* $\mathbf{y}$ *is* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, *then* $\hat{\boldsymbol{\beta}}$ *and* s^2 *have minimum variance among all unbiased estimators.*

Corollary 5.15 (to Theorem 7.6d). *If* $\mathbf{y}$ *is* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, *then the minimum variance unbiased estimator of* $\mathbf{a}'\boldsymbol{\beta}$ *is* $\mathbf{a}'\hat{\boldsymbol{\beta}}$ *where* $\hat{\boldsymbol{\beta}}$ *is the maximum likelihood estimator.*

Theorem 5.16 (7.7a). *In the model* $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ *with* $cov(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$, *the fitted values have the following properties:*

1. $E(\hat{\mathbf{y}}) = \mathbf{X}\boldsymbol{\beta}$
2. $cov(\hat{\mathbf{y}}) = \sigma^2\mathbf{H}$
3. $cov(\hat{\mathbf{y}}, \mathbf{y}) = \sigma^2\mathbf{H}$

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

Theorem 5.17 (7.7b). *The residuals* $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \hat{\mathbf{y}}$ *have the following properties:*

1. $E(\hat{\boldsymbol{\epsilon}}) = \mathbf{0}$
2. $cov(\hat{\boldsymbol{\epsilon}}) = \sigma^2(\mathbf{I} - \mathbf{H})$
3. $cov(\hat{\boldsymbol{\epsilon}}, \mathbf{y}) = \sigma^2(\mathbf{I} - \mathbf{H})$
4. $cov(\hat{\boldsymbol{\epsilon}}, \hat{\mathbf{y}}) = \mathbf{0}$

Theorem 5.18 (7.8a). *Let* $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, *let* $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, *and let* $cov(\mathbf{y}) = cov(\boldsymbol{\epsilon}) = \sigma^2\mathbf{V}$, *where* $\mathbf{X}$ *is a full-rank matrix and* $\mathbf{V}$ *is a known positive definite matrix. Then:*

1. *The best linear unbiased estimator (BLUE) of* $\boldsymbol{\beta}$ *is:*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

2. *The covariance matrix for* $\hat{\boldsymbol{\beta}}$ *is:*

$$cov(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

3. *An unbiased estimator of* σ^2 *is:*

$$s^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n-k-1}$$

Theorem 5.19 (7.8b). *If* $\mathbf{y}$ *is* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$, *where* $\mathbf{X}$ *is full-rank and* $\mathbf{V}$ *is a known positive definite matrix, where* $\mathbf{X}$ *is* $n \times (k+1)$ *of rank* $k+1$, *then the maximum likelihood estimators for* $\boldsymbol{\beta}$ *and* σ^2 *are:*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$
$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Theorem 5.20 (7.8c). *Under the assumption that* $\mathbf{y}$ *is* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$:

1. $\hat{\boldsymbol{\beta}}$ *is* $N_{k+1}[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}]$
2. $(n-k-1)s^2/\sigma^2$ *is* $\chi^2(n-k-1)$
3. $\hat{\boldsymbol{\beta}}$ *and* s^2 *are independent*

Theorem 5.21 (7.9a). *If we fit the model* $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\epsilon}^*$ *when the correct model is* $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$ *with* $cov(\mathbf{y}) = \sigma^2\mathbf{I}$, *then:*

1. $E(\hat{\boldsymbol{\beta}}_1^*) = \boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2$ *where* $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$
2. $cov(\hat{\boldsymbol{\beta}}_1^*) = \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}$

Theorem 5.22 (7.9b). *Let* $\hat{y}_0^1 = \mathbf{x}_{01}'\hat{\boldsymbol{\beta}}_1^*$ *where* $\hat{\boldsymbol{\beta}}_1^* = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$, *then if* $\boldsymbol{\beta}_2 \neq \mathbf{0}$:

$$
\begin{aligned}
E(\mathbf{x}_{01}'\hat{\boldsymbol{\beta}}_1^*) &= \mathbf{x}_{01}'(\boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2) \\
&= \mathbf{x}_0'\boldsymbol{\beta} - (\mathbf{x}_{02} - \mathbf{A}'\mathbf{x}_{01})'\boldsymbol{\beta}_2 \neq \mathbf{x}_0'\boldsymbol{\beta}
\end{aligned}
$$

Theorem 5.23 (7.9c). *Let* $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ *from the full model be partitioned as* $\hat{\boldsymbol{\beta}} = \begin{pmatrix}\hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2\end{pmatrix}$ *and let* $\hat{\boldsymbol{\beta}}_1^* = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$ *be the estimator from the reduced model. Then:*

1. $cov(\hat{\boldsymbol{\beta}}_1) - cov(\hat{\boldsymbol{\beta}}_1^*) = \sigma^2\mathbf{A}\mathbf{B}^{-1}\mathbf{A}'$ *which is a positive definite matrix, where* $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ *and* $\mathbf{B} = \mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1\mathbf{A}$
2. $var(\mathbf{x}_0'\hat{\boldsymbol{\beta}}) \geq var(\mathbf{x}_0'\hat{\boldsymbol{\beta}}_1^*)$

Theorem 5.24 (7.9d). *If* $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ *is the correct model, then for the reduced model* $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\epsilon}_1^*$ *(underfitting), where* $\mathbf{X}_1$ *is* $n \times (p+1)$ *with* $p < k$, *the variance estimator:*

$$
s_1^2 = \frac{(\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta}_1^*)^T(\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta}_1^*)}{n-p-1}
$$

has the expected value:

$$
E(s_1^2) = \sigma^2 + \frac{\boldsymbol{\beta}_2^T\mathbf{X}_2^T[\mathbf{I} - (\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T]\mathbf{X}_2\boldsymbol{\beta}_2}{n-p-1}
$$

Theorem 5.25 (7.9f). *For the model* $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\epsilon}^*$ *when the true model is* $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$:

$$
MSE(\hat{\boldsymbol{\beta}}_1^*) = \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1} + \mathbf{A}\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{A}'
$$

where $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$

Theorem 5.26 (7.10). *If* $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{O}$, *then the estimator of* $\boldsymbol{\beta}_1$ *in the full model* $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$ *is the same as the estimator of* $\boldsymbol{\beta}_1^*$ *in the reduced model* $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\epsilon}^*$.

Theorem 5.27 (7.9e). *If* $\mathbf{y}$ *is* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ *and* $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{O}$, *then:*

1. $\hat{\boldsymbol{\beta}}_1$ *and* $\hat{\boldsymbol{\beta}}_2$ *are independent*
2. *The test statistics for testing* $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ *and* $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ *are independent*

Theorem 5.28 (7.11). *For the model* $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ *with* $cov(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$:

$$
R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = corr^2(\mathbf{y}, \hat{\mathbf{y}})
$$

where SSR is the regression sum of squares, SSE is the error sum of squares, and SST is the total sum of squares.

6 Multiple Regression: Tests of Hypotheses and Confidence Intervals (Ch. 8)

Theorem 6.1 (8.1a). *For the linear model* $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ *with* $cov(\mathbf{y}) = \sigma^2\mathbf{I}$*:*

1. $SSE = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$
2. $SSR = \mathbf{y}'\mathbf{H}\mathbf{y} - n\bar{y}^2$
3. $SST = \mathbf{y}'\mathbf{y} - n\bar{y}^2$

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

Theorem 6.2 (8.1b). *If* $\mathbf{y}$ *is* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$*, then:*

1. SSE/σ^2 *is* $\chi^2(n-k-1)$
2. SSR/σ^2 *is* $\chi^2(k, \lambda)$ *where* $\lambda = \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}/2\sigma^2$

Theorem 6.3 (8.1c). *If* $\mathbf{y}$ *is* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$*, then SSR and SSE are independent.*

Theorem 6.4 (8.1d). *Let* $\mathbf{y}$ *be* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ *and define the F statistic*

$$F = \frac{SSR/k}{SSE/(n-k-1)} = \frac{SSR/k}{s^2}$$

The distribution of F is:

1. *If* $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ *is false, then F is distributed as* $F(k, n-k-1, \lambda_1)$*, where* $\lambda_1 = \boldsymbol{\beta}_1'\mathbf{X}_c'\mathbf{X}_c\boldsymbol{\beta}_1/2\sigma^2$
2. *If* $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ *is true, then* $\lambda_1 = 0$ *and F is distributed as* $F(k, n-k-1)$

Theorem 6.5 (8.2a). *The matrix* $\mathbf{H} - \mathbf{H}_1 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$ *is idempotent with rank h, where h is the number of elements in* $\boldsymbol{\beta}_2$*.*

Theorem 6.6 (8.2b). *If* $\mathbf{y}$ *is* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ *and* $\mathbf{H}$ *and* $\mathbf{H}_1$ *are as defined previously, then:*

1. $\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/\sigma^2$ *is* $\chi^2(n-k-1)$
2. $\mathbf{y}'(\mathbf{H} - \mathbf{H}_1)\mathbf{y}/\sigma^2$ *is* $\chi^2(h, \lambda_1)$*, where*
$$\lambda_1 = \boldsymbol{\beta}_2'[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2]\boldsymbol{\beta}_2/2\sigma^2$$
3. $\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$ *and* $\mathbf{y}'(\mathbf{H} - \mathbf{H}_1)\mathbf{y}$ *are independent*

Theorem 6.7 (8.2c). *Let* $\mathbf{y}$ *be* $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ *and define an F statistic as:*

$$F = \frac{\mathbf{y}'(\mathbf{H} - \mathbf{H}_1)\mathbf{y}/h}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/(n-k-1)} = \frac{SS(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/h}{SSE/(n-k-1)}$$

The distribution of F is:

1. *If* $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ *is false, then F is distributed as* $F(h, n-k-1, \lambda_1)$
2. *If* $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ *is true then* $\lambda_1 = 0$ *and F is distributed as* $F(h, n-k-1)$

Theorem 6.8 (8.2d). *If the model is partitioned as in (8.7), then* $SS(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}_2'[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2]\hat{\boldsymbol{\beta}}_2$*, where* $\hat{\boldsymbol{\beta}}_2$ *is from the partition of* $\hat{\boldsymbol{\beta}}$ *in the full model:*

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Theorem 6.9 (8.3)**.** *The F statistics for testing $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ and $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ can be written in terms of R^2 as:*

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

and

$$F = \frac{(R^2 - R_r^2)/h}{(1-R^2)/(n-k-1)}$$

where R^2 and R_r^2 are the coefficients of determination for the full and reduced models respectively.

Theorem 6.10 (8.4a)**.** *If $\mathbf{y}$ is distributed $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and $\mathbf{C}$ is $q \times (k+1)$ of rank $q \leq k+1$, then:*

1. $\mathbf{C}\hat{\boldsymbol{\beta}}$ *is* $N_q[\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']$
2. $SSH/\sigma^2 = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/\sigma^2$ *is* $\chi^2(q, \lambda)$, *where* $\lambda = (\mathbf{C}\boldsymbol{\beta})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\boldsymbol{\beta}/2\sigma^2$
3. SSE/σ^2 *is* $\chi^2(n-k-1)$
4. *SSH and SSE are independent*

Theorem 6.11 (8.4b)**.** *Let $\mathbf{y}$ be $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and define the statistic:*

$$F = \frac{SSH/q}{SSE/(n-k-1)} = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/q}{SSE/(n-k-1)}$$

where $\mathbf{C}$ is $q \times (k+1)$ of rank $q \leq k+1$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. The distribution of F in (8.27) is as follows:

1. *If $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is false, then*

   $$F \text{ is distributed as } F(q, n-k-1, \lambda)$$

   where $\lambda = (\mathbf{C}\boldsymbol{\beta})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\boldsymbol{\beta}/2\sigma^2$
2. *If $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is true, then*

   $$F \text{ is distributed as } F(q, n-k-1)$$

Theorem 6.12 (8.4c)**.** *Under $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, a $100(1-\alpha)\%$ confidence region for $\mathbf{C}\boldsymbol{\beta}$ consists of all vectors that satisfy:*

$$(\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}} \leq qs^2 F_{\alpha,q,n-k-1}$$

Theorem 6.13 (8.4d)**.** *The F test in Theorem 8.4b for the general linear hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a full-reduced-model test.*

Theorem 6.14 (8.5)**.**

1. *The maximum value of F in (8.44) is given by:*

   $$\max_{\mathbf{a}} \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})^2}{\sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}} = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}}{\sigma^2}$$

2. *If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, then $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}/(k+1)\sigma^2$ is distributed as $F(k+1, n-k-1)$. Thus*

   $$\max_{\mathbf{a}} \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})^2}{\sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}(k+1)}$$

   is distributed as $F(k+1, n-k-1)$.

Theorem 6.15 (8.7a)**.** *If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, the likelihood ratio test for $H_0 : \boldsymbol{\beta} = \mathbf{0}$ can be based on:*

$$F = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}/(k+1)}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k-1)}$$

We reject H_0 if $F > F_{\alpha,k+1,n-k-1}$.

Theorem 6.16 (8.7b (continued)). *Under H_0:*

$$LR = \frac{|S_0^*|^{n/2}}{|\hat{S}|^{n/2}} = (1 - R^2)^{n/2}$$

When H_0 is false:

$$LR = \frac{\max_{H_0} L(\boldsymbol{\beta}, \sigma^2)}{\max_{H_1} L(\boldsymbol{\beta}, \sigma^2)}$$

Theorem 6.17 (8.8). *Under the normal model with $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, a $100(1-\alpha)\%$ confidence region for $\boldsymbol{\beta}$ consists of all vectors $\boldsymbol{\beta}$ that satisfy:*

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq (k+1)s^2 F_{\alpha,k+1,n-k-1}$$

Theorem 6.18 (8.9). *A $100(1-\alpha)\%$ confidence interval for β_j is given by:*

$$\hat{\beta}_j \pm t_{\alpha/2,n-k-1} s\sqrt{g_{jj}}$$

where g_{jj} is the jth diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$.

Theorem 6.19 (8.10). *A $100(1-\alpha)\%$ confidence interval for $\mathbf{a}'\boldsymbol{\beta}$ is given by:*

$$\mathbf{a}'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1} s\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$$

Theorem 6.20 (8.11). *A $100(1-\alpha)\%$ confidence interval for $E(y_0) = \mathbf{x}_0'\boldsymbol{\beta}$ is given by:*

$$\mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1} s\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$$

Theorem 6.21 (8.12). *A $100(1-\alpha)\%$ prediction interval for a future observation y_0 at $\mathbf{x}_0$ is given by:*

$$\mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1} s\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$$

Theorem 6.22 (8.13). *A $100(1-\alpha)\%$ confidence interval for σ^2 is given by:*

$$\frac{(n-k-1)s^2}{\chi^2_{\alpha/2,n-k-1}} \leq \sigma^2 \leq \frac{(n-k-1)s^2}{\chi^2_{1-\alpha/2,n-k-1}}$$

Theorem 6.23 (8.14). *A $100(1-\alpha)\%$ confidence interval for σ is given by:*

$$\sqrt{\frac{(n-k-1)s^2}{\chi^2_{\alpha/2,n-k-1}}} \leq \sigma \leq \sqrt{\frac{(n-k-1)s^2}{\chi^2_{1-\alpha/2,n-k-1}}}$$

Theorem 6.24 (8.15). *For simultaneous confidence intervals with familywise confidence coefficient $1-\alpha$, Bonferroni confidence intervals for $\beta_1, \beta_2, \ldots, \beta_k$ are given by:*

$$\hat{\beta}_j \pm t_{\alpha/2k,n-k-1} s\sqrt{g_{jj}}, \quad j = 1, 2, \ldots, k$$

Theorem 6.25 (8.16). *For d linear functions $\mathbf{a}_1'\boldsymbol{\beta}, \mathbf{a}_2'\boldsymbol{\beta}, \ldots, \mathbf{a}_d'\boldsymbol{\beta}$, Bonferroni confidence intervals with familywise confidence coefficient $1-\alpha$ are given by:*

$$\mathbf{a}_i'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2d,n-k-1} s\sqrt{\mathbf{a}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}_i}, \quad i = 1, 2, \ldots, d$$

Theorem 6.26 (8.17). *Scheffé simultaneous confidence intervals for all possible linear combinations $\mathbf{a}'\boldsymbol{\beta}$ with confidence coefficient $1-\alpha$ are given by:*

$$\mathbf{a}'\hat{\boldsymbol{\beta}} \pm s\sqrt{(k+1)F_{\alpha,k+1,n-k-1}\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$$

Theorem 6.27 (8.18). *Scheffé simultaneous prediction intervals for d future observations with confidence coefficient $1-\alpha$ are given by:*

$$\mathbf{x}_i'\hat{\boldsymbol{\beta}} \pm s\sqrt{dF_{\alpha,d,n-k-1}[1 + \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i]}, \quad i = 1, 2, \ldots, d$$

7 Multiple Regression: Model Validation and Diagnostics (Ch. 9)

Theorem 7.1 (9.2). *If* $\mathbf{X}$ *is* $n \times (k+1)$ *of rank* $k+1 < n$*, and if the first column of* $\mathbf{X}$ *is* $\mathbf{j}$*, then the elements* h_{ij} *of* $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ *have the following properties:*

1. $\frac{1}{n} \le h_{ii} \le 1$ *for* $i = 1, 2, \ldots, n$
2. $-0.5 \le h_{ij} \le 0.5$ *for all* $j \ne i$
3. $h_{ii} = \frac{1}{n} + (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)'(\mathbf{X}_c'\mathbf{X}_c)^{-1}(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)$*, where* $\mathbf{x}_{1i}' = (x_{i1}, x_{i2}, \ldots, x_{ik})$*,* $\bar{\mathbf{x}}_1' = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k)$*, and* $(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)'$ *is the ith row of the centered matrix* $\mathbf{X}_c$
4. $tr(\mathbf{H}) = \sum_{i=1}^{n} h_{ii} = k + 1$

Theorem 7.2 (9.3a). *For the mean-shift outlier model* $E(y_i) = \mathbf{x}_i'\boldsymbol{\beta} + u$*, where* $\mathbf{x}_i'$ *is the ith row of* $\mathbf{X}$*, the test statistic* t_i *in (9.26) or (9.31) has a* $t(n-k-1)$ *distribution, and can be used to test* $H_0 : u = 0$.

Theorem 7.3 (9.3b). *For the deleted residual* $\hat{\epsilon}_{(i)}$ *defined in (9.27), we have:*

$$\hat{\epsilon}_{(i)} = \frac{\hat{\epsilon}_i}{1 - h_{ii}}$$

where $\hat{\epsilon}_i$ *is the ith residual and* h_{ii} *is the ith diagonal element of* $\mathbf{H}$.

Theorem 7.4 (9.4a). *Cook's distance* D_i *can be expressed as:*

$$D_i = \frac{r_i^2}{k+1} \frac{h_{ii}}{1 - h_{ii}}$$

where r_i *is the standardized residual and* h_{ii} *is the ith diagonal element of* $\mathbf{H}$.

Theorem 7.5 (9.4b). *For the estimator* $\hat{\boldsymbol{\beta}}_{(i)}$ *obtained by deleting observation i:*

$$\hat{\boldsymbol{\beta}}_{(i)} = \hat{\boldsymbol{\beta}} - \frac{\hat{\epsilon}_i}{1 - h_{ii}} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$$

where $\hat{\boldsymbol{\beta}}$ *is the full sample estimator,* $\hat{\epsilon}_i$ *is the ith residual,* h_{ii} *is the ith diagonal element of* $\mathbf{H}$*, and* $\mathbf{x}_i$ *is the ith row of* $\mathbf{X}$.

Theorem 7.6 (9.4c). *The deleted sample variance* $s_{(i)}^2$ *can be computed as:*

$$s_{(i)}^2 = \frac{SSE_{(i)}}{n-k-2} = \frac{SSE - \hat{\epsilon}_i^2/(1 - h_{ii})}{n-k-2}$$

where SSE *is the error sum of squares from the full sample and* $SSE_{(i)}$ *is the error sum of squares with observation i deleted.*

Theorem 7.7 (9.4d). *For the deleted observations* $(y_i, \mathbf{x}_i')$*, the studentized residual* t_i *can be computed as:*

$$t_i = \frac{\hat{\epsilon}_{(i)}}{\sqrt{\widehat{var}(\hat{\epsilon}_{(i)})}} = \frac{\hat{\epsilon}_i}{s_{(i)}\sqrt{1 - h_{ii}}}$$

where $s_{(i)}^2$ *is the deleted sample variance.*

Theorem 7.8 (9.5). *The prediction sum of squares (PRESS) statistic is given by:*

$$PRESS = \sum_{i=1}^{n} \hat{\epsilon}_{(i)}^2 = \sum_{i=1}^{n} \left(\frac{\hat{\epsilon}_i}{1 - h_{ii}} \right)^2$$

where $\hat{\epsilon}_{(i)}$ *is the deleted residual and* h_{ii} *is the ith diagonal element of* $\mathbf{H}$.

Theorem 7.9 (9.6)**.** *The standardized distance* $(x_{1i} - \bar{x}_1)'(\mathbf{X}_c'\mathbf{X}_c)^{-1}(x_{1i} - \bar{x}_1)$ *in the expression for* h_{ii} *can be written as:*

$$\sum_{r=1}^{k} \frac{1}{\lambda_r} \cos^2 \theta_{ir}$$

where λ_r *is the* r*th eigenvalue of* $\mathbf{X}_c'\mathbf{X}_c$ *and* θ_{ir} *is the angle between* $\mathbf{x}_{1i} - \bar{\mathbf{x}}_1$ *and* $\mathbf{a}_r$*, the* r*th eigenvector of* $\mathbf{X}_c'\mathbf{X}_c$*.*

Theorem 7.10 (9.7)**.** *For any point* $(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)$*, the leverage* h_{ii} *is large if either:*

1. $(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)'(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)$ *is large*
2. θ_{ir} *is small for some* r *corresponding to a small eigenvalue* λ_r

Theorem 7.11 (9.8)**.** *The covariance matrix of the residual vector* $\hat{\boldsymbol{\epsilon}}$ *is:*

$$cov(\hat{\boldsymbol{\epsilon}}) = \sigma^2(\mathbf{I} - \mathbf{H})$$

where $\mathbf{H}$ *is the hat matrix.*

8 Multiple Regression: Random x's (Ch. 11)

Theorem 8.1 (10.2a)**.** *If* $(y_1, \mathbf{x}_1'), (y_2, \mathbf{x}_2'), \ldots, (y_n, \mathbf{x}_n')$ *is a random sample from* $N_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, with* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *as given in (10.2) and (10.3), the maximum likelihood estimators are:*

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} \hat{\mu}_y \\ \hat{\boldsymbol{\mu}}_x \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \bar{\mathbf{x}} \end{pmatrix}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S} = \frac{n-1}{n}\begin{pmatrix} s_{yy} & \mathbf{s}_{yx}' \\ \mathbf{s}_{yx} & \mathbf{S}_{xx} \end{pmatrix}$$

Theorem 8.2 (10.2b)**.** *The maximum likelihood estimator of a function of one or more parameters is the same function of the corresponding estimators; that is, if* $\hat{\boldsymbol{\theta}}$ *is the maximum likelihood estimator of the vector or matrix of parameters* $\boldsymbol{\theta}$*, then* $g(\hat{\boldsymbol{\theta}})$ *is the maximum likelihood estimator of* $g(\boldsymbol{\theta})$*.*

Theorem 8.3 (10.2c)**.** *If* $(y_1, \mathbf{x}_1'), (y_2, \mathbf{x}_2'), \ldots, (y_n, \mathbf{x}_n')$ *is a random sample from* $N_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, the maximum likelihood estimators for* β_0*,* $\boldsymbol{\beta}_1$*, and* σ^2 *are:*

$$\hat{\beta}_0 = \bar{y} - \mathbf{s}_{yx}'\mathbf{S}_{xx}^{-1}\bar{\mathbf{x}}$$

$$\hat{\boldsymbol{\beta}}_1 = \mathbf{S}_{xx}^{-1}\mathbf{s}_{yx}$$

$$\hat{\sigma}^2 = \frac{n-1}{n}s^2 \text{ where } s^2 = s_{yy} - \mathbf{s}_{yx}'\mathbf{S}_{xx}^{-1}\mathbf{s}_{yx}$$

Theorem 8.4 (10.5)**.** *If* $(y_1, \mathbf{x}_1'), (y_2, \mathbf{x}_2'), \ldots, (y_n, \mathbf{x}_n')$ *is a random sample from* $N_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, the likelihood ratio test for* $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ *or equivalently* $H_0 : \rho^2_{y|x} = 0$ *can be based on* F *in (10.44). We reject* H_0 *if* $F \geq F_{\alpha,k,n-k-1}$*.*

Theorem 8.5 (10.6)**.** *The increase in* R^2 *due to* z *can be expressed as:*

$$R^2_{yw} - R^2_{yx} = \frac{(\hat{r}_{yz} - r_{yz})^2}{1 - R^2_{zx}}$$

where $\hat{r}_{yz} = \hat{\boldsymbol{\beta}}_{zx}'\mathbf{r}_{yx}$ *is a "predicted" value of* r_{yz} *based on the relationship of* z *to the* x*'s.*

Theorem 8.6 (10.7)**.** *For the random vector* $(y, \mathbf{x}')$*, the function* $t(\mathbf{x})$ *that minimizes the mean squared error* $E[y - t(\mathbf{x})]^2$ *is given by* $E(y|\mathbf{x})$*.*

Theorem 8.7 (10.7b). *The linear function $t(\mathbf{x})$ that minimizes $E[y - t(\mathbf{x})]^2$ is given by $t(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}_1'\mathbf{x}$, where:*

$$\beta_0 = \mu_y - \boldsymbol{\sigma}_{yx}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x$$

$$\boldsymbol{\beta}_1 = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx}$$

Theorem 8.8 (10.7c). *If $(y_1, \mathbf{x}_1'), (y_2, \mathbf{x}_2'), \ldots, (y_n, \mathbf{x}_n')$ is a random sample with mean vector and covariance matrix $\hat{\boldsymbol{\mu}}$ and $\mathbf{S}$, then the estimators $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}_1$ that minimize $\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}_1'\mathbf{x}_i)^2/n$ are given by:*

$$\hat{\beta}_0 = \bar{y} - \mathbf{s}_{yx}'\mathbf{S}_{xx}^{-1}\bar{\mathbf{x}}$$

$$\hat{\boldsymbol{\beta}}_1 = \mathbf{S}_{xx}^{-1}\mathbf{s}_{yx}$$

Theorem 8.9 (10.8a). *The expression for $r_{12|3}$ in (10.66) is equivalent to an element of $\mathbf{R}_{y|x}$ in (10.65) and is also equal to $r_{y_1-\hat{y}_1, y_2-\hat{y}_2}$ from (7.94), where $y_1 - \hat{y}_1$ and $y_2 - \hat{y}_2$ are residuals from regression of y_1 on y_3 and y_2 on y_3.*

Theorem 8.10 (10.8b). *The sample covariance matrix of the residual vector $y_i - \hat{y}_i(\mathbf{x})$ is equivalent to $\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}$ in (10.65), that is, $\mathbf{S}_{y-\hat{y}} = \mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}$.*

9 Analysis of Variance Models (Ch. 12)

Theorem 9.1 (12.2a). *If $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, the system of equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ is consistent.*

Theorem 9.2 (12.2b). *In the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, the linear function $\mathbf{l}'\boldsymbol{\beta}$ is estimable if and only if any one of the following equivalent conditions holds:*

1. *$\mathbf{l}'$ is a linear combination of the rows of $\mathbf{X}$; that is, there exists a vector $\mathbf{a}$ such that $\mathbf{a}'\mathbf{X} = \mathbf{l}'$*
2. *$\mathbf{l}'$ is a linear combination of the rows of $\mathbf{X}'\mathbf{X}$ or $\mathbf{l}$ is a linear combination of the columns of $\mathbf{X}'\mathbf{X}$*
3. *$\mathbf{l}$ or $\mathbf{l}'$ is such that $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{l} = \mathbf{l}$ or $\mathbf{l}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{l}'$*

Theorem 9.3 (12.2c). *In the non-full-rank model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, the number of linearly independent estimable functions of $\boldsymbol{\beta}$ is the rank of $\mathbf{X}$.*

Theorem 9.4 (12.3a). *Let $\mathbf{l}'\boldsymbol{\beta}$ be an estimable function of $\boldsymbol{\beta}$ in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$. Let $\hat{\boldsymbol{\beta}}$ be any solution to the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$, and let $\mathbf{r}$ be any solution to $\mathbf{X}'\mathbf{X}\mathbf{r} = \mathbf{l}$. Then the two estimators $\mathbf{l}'\hat{\boldsymbol{\beta}}$ and $\mathbf{r}'\mathbf{X}'\mathbf{y}$ have the following properties:*

1. *$E(\mathbf{l}'\hat{\boldsymbol{\beta}}) = E(\mathbf{r}'\mathbf{X}'\mathbf{y}) = \mathbf{l}'\boldsymbol{\beta}$*
2. *$\mathbf{l}'\hat{\boldsymbol{\beta}}$ is equal to $\mathbf{r}'\mathbf{X}'\mathbf{y}$ for any $\hat{\boldsymbol{\beta}}$ or any $\mathbf{r}$*
3. *$\mathbf{l}'\hat{\boldsymbol{\beta}}$ and $\mathbf{r}'\mathbf{X}'\mathbf{y}$ are invariant to the choice of $\hat{\boldsymbol{\beta}}$ or $\mathbf{r}$*

Theorem 9.5 (12.3b). *Let $\mathbf{l}'\boldsymbol{\beta}$ be an estimable function in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$ and $cov(\mathbf{y}) = \sigma^2\mathbf{I}$. Let $\mathbf{r}$ be any solution to $\mathbf{X}'\mathbf{X}\mathbf{r} = \mathbf{l}$, and let $\hat{\boldsymbol{\beta}}$ be any solution to $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$. Then the variance of $\mathbf{l}'\hat{\boldsymbol{\beta}}$ or $\mathbf{r}'\mathbf{X}'\mathbf{y}$ has the following properties:*

1. *$var(\mathbf{r}'\mathbf{X}'\mathbf{y}) = \sigma^2\mathbf{r}'\mathbf{X}'\mathbf{X}\mathbf{r} = \sigma^2\mathbf{r}'\mathbf{l}$*
2. *$var(\mathbf{l}'\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{l}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{l}$*
3. *$var(\mathbf{l}'\hat{\boldsymbol{\beta}})$ is unique, that is, invariant to the choice of $\mathbf{r}$ or $(\mathbf{X}'\mathbf{X})^{-}$*

Theorem 9.6 (12.3c). *If $\mathbf{l}_1'\boldsymbol{\beta}$ and $\mathbf{l}_2'\boldsymbol{\beta}$ are two estimable functions in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$ and cov$(\mathbf{y}) = \sigma^2\mathbf{I}$, the covariance of their estimators is given by:*

$$cov(\mathbf{l}_1'\hat{\boldsymbol{\beta}}, \mathbf{l}_2'\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{r}_1'\mathbf{l}_2 = \sigma^2\mathbf{l}_1'\mathbf{r}_2 = \sigma^2\mathbf{l}_1'(\mathbf{X}'\mathbf{X})^-\mathbf{l}_2$$

where $\mathbf{X}'\mathbf{X}\mathbf{r}_1 = \mathbf{l}_1$ and $\mathbf{X}'\mathbf{X}\mathbf{r}_2 = \mathbf{l}_2$.

Theorem 9.7 (12.3d). *If $\mathbf{l}'\boldsymbol{\beta}$ is an estimable function in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, then the estimators $\mathbf{l}'\hat{\boldsymbol{\beta}}$ and $\mathbf{r}'\mathbf{X}'\mathbf{y}$ are BLUE.*

Theorem 9.8 (12.3e). *For s^2 defined in (12.22) for the non-full-rank model, we have the following properties:*

1. *$E(s^2) = \sigma^2$*
2. *s^2 is invariant to the choice of $\hat{\boldsymbol{\beta}}$ or to the choice of generalized inverse $(\mathbf{X}'\mathbf{X})^-$*

Theorem 9.9 (12.3f). *If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, then the maximum likelihood estimators for $\boldsymbol{\beta}$ and σ^2 are given by:*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$$
$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Theorem 9.10 (12.3g). *If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, then the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and s^2 have the following properties:*

1. *$\hat{\boldsymbol{\beta}}$ is $N_p[(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-]$*
2. *$(n-k)s^2/\sigma^2$ is $\chi^2(n-k)$*
3. *$\hat{\boldsymbol{\beta}}$ and s^2 are independent*

Theorem 9.11 (12.3h). *If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, and if $\mathbf{l}'\boldsymbol{\beta}$ is an estimable function, then $\mathbf{l}'\hat{\boldsymbol{\beta}}$ has minimum variance among all unbiased estimators.*

Theorem 9.12 (12.6a). *If $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, and if $\mathbf{T}$ is a $(p-k) \times p$ matrix of rank $p-k$ such that $\mathbf{T}\boldsymbol{\beta}$ is a set of nonestimable functions, then there is a unique vector $\hat{\boldsymbol{\beta}}$ that satisfies both $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ and $\mathbf{T}\hat{\boldsymbol{\beta}} = \mathbf{0}$.*

Theorem 9.13 (12.7a). *Consider the partitioned model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$. If $\mathbf{X}_2'\mathbf{X}_1 = \mathbf{O}$, any estimate of $\boldsymbol{\beta}_2^*$ in the reduced model $\mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2^* + \boldsymbol{\epsilon}^*$ is also an estimate of $\boldsymbol{\beta}_2$ in the full model.*

Theorem 9.14 (12.7b). *If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, 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 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$, then:*

1. *$\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}'$ is nonsingular*
2. *$\mathbf{C}\hat{\boldsymbol{\beta}}$ is $N_m[\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}']$*
3. *$SSH/\sigma^2 = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/\sigma^2$ is $\chi^2(m, \lambda)$, where $\lambda = (\mathbf{C}\boldsymbol{\beta})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}']^{-1}\mathbf{C}\boldsymbol{\beta}/2\sigma^2$*
4. *$SSE/\sigma^2 = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}']\mathbf{y}/\sigma^2$ is $\chi^2(n-k)$*
5. *SSH and SSE are independent*

Theorem 9.15 (12.7c). *Let $\mathbf{y}$ be $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, and define the statistic:*

$$F = \frac{SSH/m}{SSE/(n-k)} = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/m}{SSE/(n-k)}$$

Then, if $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is true, F is distributed as $F(m, n-k)$.

Theorem 9.16 (12.7d)**.** *The F test in Theorem 12.7c for* $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ *is a full-reduced-model test.*

Theorem 9.17 (12.7e)**.** *The mean vector and covariance matrix of* $\hat{\boldsymbol{\beta}}_c$ *in (12.30) are:*

1. $E(\hat{\boldsymbol{\beta}}_c) = \boldsymbol{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\boldsymbol{\beta}$
2. $cov(\hat{\boldsymbol{\beta}}_c) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}$

Theorem 9.18 (12.8.1)**.** *In a non-full-rank model, the following properties hold for estimable functions:*

1. *Every row of* $\mathbf{X}\boldsymbol{\beta}$ *is estimable*
2. *Every row of* $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ *is estimable*
3. *All estimable functions can be obtained from linear combinations of rows of* $\mathbf{X}\boldsymbol{\beta}$ *or* $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$

Theorem 9.19 (12.8.2)**.** *For testing* $H_0 : \mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}_3$ *in a balanced two-way model:*

1. *The sum of squares* $SS(\mathbf{a}|\mathbf{m}, \mathbf{b})$ *has 2 degrees of freedom*
2. $SS(\mathbf{a}|\mathbf{m}, \mathbf{b}) = \sum_i \frac{y_{i.}^2}{2} - \frac{y_{..}^2}{6}$
3. *Under* H_0, $\frac{SS(\mathbf{a}|\mathbf{m},\mathbf{b})/2}{SSE/2} \sim F(2, 2)$

Theorem 9.20 (12.8.3)**.** *In a balanced non-full-rank model with orthogonal parameterization:*

1. *The columns of* $\mathbf{X}$ *corresponding to different groups of effects are orthogonal*
2. *The estimates from the full model equal those from reduced models for parameters not involved in the hypothesis*
3. *The sums of squares partition additively:* $SS(\mathbf{m}, \mathbf{a}, \mathbf{b}) = SS(\mathbf{m}) + SS(\mathbf{a}) + SS(\mathbf{b})$

Theorem 9.21 (12.8.4)**.** *For a two-way model with interaction* $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$*, under the side conditions:*

$$\sum_i \alpha_i = 0$$
$$\sum_j \beta_j = 0$$
$$\sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$$

The parameters are uniquely defined and the model becomes full-rank with orthogonal columns.

Theorem 9.22 (12.8.5)**.** *In a balanced complete block design:*

1. *The estimates of treatment effects are uncorrelated with block effects*
2. *The sums of squares for treatments and blocks are independent*
3. *The efficiency relative to a completely randomized design is* $\frac{r}{r-1}$ *where* r *is the number of blocks*

Theorem 9.23 (12.8.6)**.** *For a non-full-rank model, a set of contrasts* $\{\mathbf{l}_1'\boldsymbol{\beta}, \ldots, \mathbf{l}_q'\boldsymbol{\beta}\}$ *is estimable if and only if* $\sum_i c_i\mathbf{l}_i$ *is estimable for all choices of constants* c_i.

Theorem 9.24 (12.9)**.** *If* $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$ *is the hat matrix for a non-full-rank model, then:*

1. $\mathbf{H}$ *is unique and idempotent regardless of the choice of* $(\mathbf{X}'\mathbf{X})^-$
2. $rank(\mathbf{H}) = rank(\mathbf{X}) = k$

3. $tr(\mathbf{H}) = k$

Theorem 9.25 (12.10). *In a balanced non-full-rank model with no missing cells, the following conditions hold:*

1. *The rows of* $\mathbf{X}$ *corresponding to effects in different factors or interactions are orthogonal*
2. *The estimators of parameters in different factors or interactions are uncorrelated*
3. *The sum of squares for different factors or interactions are independent*