MATH 2016 (13177) STATISTICAL MODELLING

Matrix definitions and results

$\mathbf{1}_n$ is a column vector consisting of n ones and n is the number of observations in the random sample. Note that $\mathbf{1}'_n\mathbf{1}_n = n$

$\mathbf{J}_n$ is the square matrix of order n whose elements are all 1 and $\mathbf{J}_n^2 = n\mathbf{J}_n$.

Lemma I.1: The transpose of a matrix displays the following properties:

1. $(\mathbf{A}')' = \mathbf{A}$.
2. $\mathbf{a}$ is a column vector and $\mathbf{a}'$ is the corresponding row vector.
3. $(c\mathbf{A} + d\mathbf{B})' = c\mathbf{A}' + d\mathbf{B}'$.
4. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.
5. $\mathbf{A}' = \mathbf{A}$ provided $\mathbf{A}$ is symmetric.
6. The product of two different symmetric matrices is not symmetric — $(\mathbf{AB})' = \mathbf{BA}$ for $\mathbf{A}$ and $\mathbf{B}$ symmetric.
7. The product of any matrix with its transpose is symmetric — $(\mathbf{A}'\mathbf{A})' = \mathbf{A}'(\mathbf{A}')' = \mathbf{A}'\mathbf{A}$ and $(\mathbf{AA}')' = (\mathbf{A}')'\mathbf{A}' = \mathbf{AA}'$.
8. A column vector premultiplied by a row vector is a scalar and so is symmetric — $\mathbf{a}'\mathbf{b} = \sum_{i=1}^{n} a_i b_j$ and $\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a} = (\mathbf{a}'\mathbf{b})'$.
9. $\mathbf{a}'\mathbf{a} = \sum_{i=1}^{n} a_i^2$, a scalar.
10. A column vector of order n post multiplied by its transpose is a symmetric matrix of order $n \times n$ — from property 7 we have $(\mathbf{aa}')' = \mathbf{aa}'$.

Definition I.6: The **rank** of an $n \times q$ matrix $\mathbf{A}$ with $n \geq q$ is the number of linearly independent columns of the matrix. The matrix is said to be of **full rank**, or rank q, if, none of the columns in the matrix can be written as a linear combination of the other columns. ■

Lemma XI.1: Let $\mathbf{X}$ be an $n \times f$ matrix with $n \geq f$. Then $rank(\mathbf{X}) = rank(\mathbf{X}'\mathbf{X})$.

Definition XII.2: The trace of a square matrix is the sum of its diagonal elements. ■

Lemma XII.3: Let c be a scalar and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be matrices. Then, when the appropriate operations are defined, we have

- $trace(\mathbf{A}') = trace(\mathbf{A})$
- $trace(c\mathbf{A}) = c \times trace(\mathbf{A})$
- $trace(\mathbf{A} + \mathbf{B}) = trace(\mathbf{A}) + trace(\mathbf{B})$

- $trace(\mathbf{AB}) = trace(\mathbf{BA})$
- $trace(\mathbf{ABC}) = trace(\mathbf{CAB}) = trace(\mathbf{BCA})$
- $trace(\mathbf{A} \otimes \mathbf{B}) = trace(\mathbf{A})trace(\mathbf{B})$
- $trace(\mathbf{A}'\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{0}$. ■

For **B** idempotent, $rank(\mathbf{B}) = trace(\mathbf{B})$.

$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{bmatrix}$ then $\mathbf{M}^{-1} = \left(\mathbf{M}^{-1}\right)' = \begin{bmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}' & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{-1} - \mathbf{VB}'\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{BW} \\ -\mathbf{WB}'\mathbf{A}^{-1} & \left(\mathbf{D} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix}$

Definition IX.3: The direct-product operator is denoted $\otimes$ and, if $\mathbf{A}_r$ and $\mathbf{B}_c$ are square matrices of order *r* and *c*, respectively, then the direct product is given by $\mathbf{A}_r \otimes \mathbf{B}_c = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1r}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{r1}\mathbf{B} & \cdots & a_{rr}\mathbf{B} \end{bmatrix}$

Lemma XI.8: Let **A**, **B**, **C** and **D** be square matrices and *a*, *b* and *c* be scalar constants. Then,

$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ provided **A** and **C**, as well as **B** and **D**, are conformable.

$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$$

$$a\mathbf{A} = a \otimes \mathbf{A}$$

$$a\mathbf{A} \otimes b\mathbf{B} = ab\mathbf{A} \otimes \mathbf{B}$$

$$\mathbf{A} \otimes (b\mathbf{B} + c\mathbf{C}) = b\mathbf{A} \otimes \mathbf{B} + c\mathbf{A} \otimes \mathbf{C}$$

$$trace(\mathbf{A} \otimes \mathbf{B}) = trace(\mathbf{A})trace(\mathbf{B})$$

■

Definition V.5: Two columns of a matrix **X**, say $\mathbf{x}_i$ and $\mathbf{x}_j$, are said to be **orthogonal** if $\mathbf{x}_i'\mathbf{x}_j = 0$.

Definition XI.5: Let **A** be an $n \times q$ matrix. A $q \times n$ matrix $\mathbf{A}^-$ such that

$$\mathbf{AA}^-\mathbf{A} = \mathbf{A}$$

is called a **generalised inverse** (g-inverse for short) for **A**. ■

Any matrix **A** has a generalized inverse but it is not unique unless **A** is nonsingular, in which case $\mathbf{A}^- = \mathbf{A}^{-1}$.

To find a generalized inverse $\mathbf{A}^-$ for an $n \times q$ matrix **A** of rank *m*:

1. Find any $m \times m$ minor **H** of **A** i.e. the matrix obtained by selecting any *m* rows and any *m* columns of **A**.
2. Find $\mathbf{H}^{-1}$.

3. Replace **H** in **A** with $\left(\mathbf{H}^{-1}\right)'$.
4. Replace all other entries in **A** with zeros.
5. Transpose the resulting matrix.

Properties of generalized inverses

Let **A** be an $n \times q$ matrix of rank m with $n \geq q \geq m$. Then

1. $\mathbf{A}^{-}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{-}$ are idempotent.
2. $rank\left(\mathbf{A}^{-}\mathbf{A}\right) = rank\left(\mathbf{A}\mathbf{A}^{-}\right) = m$.
3. If $\mathbf{A}^{-}$ is a generalized inverse of **A**, then $\left(\mathbf{A}^{-}\right)'$ is a generalized inverse of $\mathbf{A}'$; that is, $\left(\mathbf{A}^{-}\right)' = \left(\mathbf{A}'\right)^{-}$.
4. $\left(\mathbf{A}'\mathbf{A}\right)^{-}\mathbf{A}'$ is a generalized inverse of **A** so that $\mathbf{A} = \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-}\left(\mathbf{A}'\mathbf{A}\right)$ and $\mathbf{A}' = \left(\mathbf{A}'\mathbf{A}\right)\left(\mathbf{A}'\mathbf{A}\right)^{-}\mathbf{A}'$.
5. $\mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-}\mathbf{A}'$ is unique, symmetric and idempotent as it is invariant to the choice of a generalized inverse. Furthermore $rank\left(\mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-}\mathbf{A}'\right) = m$.

$$\mathbf{H} = \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}' & \mathbf{D} \end{bmatrix} \text{ then } \mathbf{H}^{-1} = \left(\mathbf{H}^{-1}\right)' = \begin{bmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}' & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{-1} - \mathbf{V}\mathbf{C}'\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{C}\mathbf{W} \\ -\mathbf{W}\mathbf{C}'\mathbf{A}^{-1} & \left(\mathbf{D} - \mathbf{C}'\mathbf{A}^{-1}\mathbf{C}\right)^{-1} \end{bmatrix}.$$

Definition I.9: A matrix **E** is idempotent if $\mathbf{E}^2 = \mathbf{E}$. ■

Theorem I.5: Given that the matrix **E** is symmetric and idempotent, then the matrix $\mathbf{R} = \mathbf{I} - \mathbf{E}$ is also symmetric and idempotent. In addition, $\mathbf{ER} = \mathbf{RE} = \mathbf{0}$.

For **B** idempotent, $rank\left(\mathbf{B}\right) = trace\left(\mathbf{B}\right)$.

Lemma XI.4: The inverse of $a\left(\mathbf{I}_m - \tfrac{1}{m}\mathbf{J}_m\right) + b\left(\tfrac{1}{m}\mathbf{J}_m\right)$ is $\dfrac{1}{a}\left(\mathbf{I}_m - \tfrac{1}{m}\mathbf{J}_m\right) + \dfrac{1}{b}\left(\tfrac{1}{m}\mathbf{J}_m\right)$.

Definition XI.10: A set of idempotents $\left\{\mathbf{Q}_{\mathrm{F}}; \mathrm{F} \in \mathcal{F}\right\}$ forms a **complete set of mutually-orthogonal idempotents** (CSMOI) if $\mathbf{Q}_{\mathrm{F}}\mathbf{Q}_{\mathrm{F}'} = \mathbf{Q}_{\mathrm{F}'}\mathbf{Q}_{\mathrm{F}} = \delta_{\mathrm{FF}'}\mathbf{Q}_{\mathrm{F}}$ for all $\mathrm{F}, \mathrm{F}' \in \mathcal{F}$ and $\sum_{\mathrm{F} \in \mathcal{F}} \mathbf{Q}_{\mathrm{F}} = \mathbf{I}_n$ where $\delta_{\mathrm{FF}'} = 1$ if $\mathrm{F} = \mathrm{F}'$ and $\delta_{\mathrm{FF}'} = 0$ otherwise. ■

Lemma XI.7: For a matrix that is the linear combination of a complete set of mutually-orthogonal matrices, such as $\mathbf{V} = \sum_{\mathrm{F} \in \mathcal{F}} \lambda_{\mathrm{F}}\mathbf{Q}_{\mathrm{F}}$, its inverse is $\mathbf{V}^{-1} = \sum_{\mathrm{F} \in \mathcal{F}} \lambda_{\mathrm{F}}^{-1}\mathbf{Q}_{\mathrm{F}}$.

Expectation and variance definitions and results

Theorem I.2: $E\left[a\times v(Y)+b\right]=aE\left[v(Y)\right]+b$

Definition I.2: The **variance**, $\text{var}\left[Y\right]=\sigma_Y^2$, of any random variable Y is defined to be

$$\text{var}\left[Y\right]=E\left[\left(Y-\psi_Y\right)^2\right]=E\left[\left(Y-E\left[Y\right]\right)^2\right]$$

Definition I.3: The covariance of two random variables, X and Y, is defined to be

$$\text{cov}\left[X,Y\right]=E\left[\left(X-E\left[X\right]\right)\left(Y-E\left[Y\right]\right)\right]$$

Lemma XI.6: Let U and V be random variables and a and b be constants. Then

$$\text{var}\left(aU+bV\right)=a^2\,\text{var}\left(U\right)+b^2\,\text{var}\left(V\right)+2ab\,\text{cov}\left(U,V\right).$$

Definition I.4: Let **Y** be a vector of n jointly-distributed random variables with $E\left[Y_i\right]=\psi_i$, $\text{var}\left[Y_i\right]=\sigma_i^2$ and $\text{cov}\left[Y_i,Y_j\right]=\sigma_{ij}\ \left(=\sigma_{ji}\right)$. Then, the **random vector** is

$$\mathbf{Y}=\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix}$$

The **expectation vector** giving the expectation of **Y** is

$$E\left[\mathbf{Y}\right]=\begin{bmatrix}E\left[Y_1\right]\\E\left[Y_2\right]\\\vdots\\E\left[Y_n\right]\end{bmatrix}=\begin{bmatrix}\psi_1\\\psi_2\\\vdots\\\psi_n\end{bmatrix}=\boldsymbol{\psi}$$

The **variance matrix**, **V**, giving the variance of **Y** is

$$\mathbf{V}=E\left[\left(\mathbf{Y}-E\left[\mathbf{Y}\right]\right)\left(\mathbf{Y}-E\left[\mathbf{Y}\right]\right)'\right]=\begin{bmatrix}\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1i} & \cdots & \sigma_{1n}\\\sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2i} & \cdots & \sigma_{2n}\\\vdots & \vdots & & \vdots & & \vdots\\\sigma_{1i} & \sigma_{2i} & \cdots & \sigma_i^2 & \cdots & \sigma_{in}\\\vdots & \vdots & & \vdots & & \vdots\\\sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{in} & \cdots & \sigma_n^2\end{bmatrix}$$

Lemma XI.5: Let

a be a $k \times 1$ vector of constants,
A an $s \times k$ matrix of constants and
Y a $k \times 1$ vector of random variables or random vector.

11. $E[\mathbf{a}] = \mathbf{a}$ and $\text{var}[\mathbf{a}] = \mathbf{0}$;
12. $E[\mathbf{a}'\mathbf{Y}] = \mathbf{a}'E[\mathbf{Y}]$ and $\text{var}[\mathbf{a}'\mathbf{Y}] = \mathbf{a}'\text{var}[\mathbf{Y}]\mathbf{a}$;
13. $E[\mathbf{AY}] = \mathbf{A}E[\mathbf{Y}]$ and $\text{var}[\mathbf{AY}] = \mathbf{A}\text{var}[\mathbf{Y}]\mathbf{A}'$. ■

$E[\mathbf{AYY}'] = \mathbf{A}E[\mathbf{YY}']$

$E[\mathit{trace}(\mathbf{A})] = \mathit{trace}(E[\mathbf{A}])$

Theorem XII.1: Let **Y** be an $n \times 1$ vector of random variables with

$$E[\mathbf{Y}] = \boldsymbol{\psi} \text{ and } \text{var}[\mathbf{Y}] = \mathbf{V},$$

where $\boldsymbol{\psi}$ is a $n \times 1$ vector of expected values and **V** is an $n \times n$ matrix.

Let **A** an $n \times n$ matrix of real numbers.

Then

$$E[\mathbf{Y}'\mathbf{AY}] = \mathit{trace}(\mathbf{AV}) + \boldsymbol{\psi}'\mathbf{A}\boldsymbol{\psi}.$$

Theorem XII.2: Let **Y** be an $n \times 1$ vector of random variables with

$$E[\mathbf{Y}] = \boldsymbol{\psi} \text{ and } \text{var}[\mathbf{Y}] = \mathbf{V},$$

where $\boldsymbol{\psi}$ is a $n \times 1$ vector of expected values and **V** is an $n \times n$ matrix.

Let $\mathbf{Y}'\mathbf{QY}/\nu$ be the mean square where **Q** is an $n \times n$ symmetric, idempotent matrix and $\nu = \mathit{trace}(\mathbf{Q})$ is the degrees of freedom of the sums of squares.

Then

$$E[\mathbf{Y}'\mathbf{QY}/\nu] = \left(\mathit{trace}(\mathbf{QV}) + \boldsymbol{\psi}'\mathbf{Q}\boldsymbol{\psi}\right)/\nu.$$

Haven't updated from here on

Least squares and ANOVA definitions and results

Theorem II.1: Let $\mathbf{Y} = \mathbf{X\theta} + \boldsymbol{\varepsilon}$ where $\mathbf{X}$ is an $n \times q$ matrix of full rank, $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors with mean $\mathbf{0}$ and variance $\sigma^2\mathbf{I}_n$, $q = p+1$ and $n \geq q$. The least squares estimator for $\boldsymbol{\theta}$ is denoted by $\hat{\boldsymbol{\theta}}$ and is given by

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

Theorem II.7: Let $\mathbf{Y}$ be a normally distributed random vector representing a random sample with $E\left[\mathbf{Y}\right] = \mathbf{X\theta}$ and $\text{var}\left[\mathbf{Y}\right] = \mathbf{V}_{\mathbf{Y}} = \sigma^2\mathbf{I}_n$ where $\mathbf{X}$ is an $n \times q$ matrix of full rank, $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters and $n \geq q$. The maximum likelihood estimator for $\boldsymbol{\theta}$ is denoted by $\tilde{\boldsymbol{\theta}}$ and is given by

$$\tilde{\boldsymbol{\theta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

Theorem V.19: Let $\mathbf{Y}$ be a random vector with $\psi = E\left[\mathbf{Y}\right] = \mathbf{X}\theta$ and $\text{var}\left[\mathbf{Y}\right] = \mathbf{V}$ where $\mathbf{X}$ is an $n \times q$ matrix, θ is a $q \times 1$ vector of unknown parameters, $\mathbf{V}$ is an $n \times n$ positive definite matrix and $n \geq q$. Then the generalized least squares estimator for θ is denoted by $\hat{\theta}$ and is given by

$$\hat{\theta} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

Theorem II.2: Let $\mathbf{Y} = \mathbf{X\theta} + \boldsymbol{\varepsilon}$ where $\mathbf{X}$ is an $n \times q$ matrix of full rank, $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors with mean $\mathbf{0}$ and variance $\sigma^2\mathbf{I}_n$, $q = p+1$ and $n \geq q$. The least squares estimator $\hat{\boldsymbol{\theta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$ is an unbiased estimator for $\boldsymbol{\theta}$. Furthermore, $\text{var}\left[\hat{\boldsymbol{\theta}}\right] = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\sigma^2$

Theorem II.5: Let $\mathbf{Y} = \mathbf{X\theta} + \boldsymbol{\varepsilon}$ where $\mathbf{X}$ is an $n \times q$ matrix of full rank, $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors with mean $\mathbf{0}$ and variance $\sigma^2\mathbf{I}_n$, $q = p+1$ and $n \geq q$. Let ℓ' is a $1 \times q$ vector of real numbers. The best linear unbiased estimator for $\ell'\boldsymbol{\theta}$ is $\ell'\hat{\boldsymbol{\theta}}$ where $\hat{\boldsymbol{\theta}}$ is the least squares estimator.

Definition II.10: Let $\hat{\boldsymbol{\theta}}$ be the least squares estimates of $\boldsymbol{\theta}$ so that $\hat{\boldsymbol{\theta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$ where $\mathbf{y}$ is an $n \times 1$ vector of observations from a random sample, $\mathbf{X}$ is an $n \times q$ matrix of full rank, $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters and $n \geq q$. Then define $\mathbf{P}_{\mathbf{X}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ and $\mathbf{R}_{\mathbf{X}} = \mathbf{I} - \mathbf{P}_{\mathbf{X}}$ so that the fitted values are $\mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{P}_{\mathbf{X}}\mathbf{y}$ and the residuals are $\mathbf{e} = \mathbf{R}_{\mathbf{X}}\mathbf{y}$. ■

Theorem II.10: Let $\mathbf{y}'\mathbf{y}$ be the Total sums of squares, $\mathbf{y}'\mathbf{P}_{\mathbf{X}}\mathbf{y}$ be the Regression sums of squares and $\mathbf{y}'\mathbf{R}_{\mathbf{X}}\mathbf{y}$ be the Residual sums of squares where $\mathbf{y}$ is an $n \times 1$ vector of observed values from a random sample and $\mathbf{P_X}$ and $\mathbf{R_X}$ are as given in definition II.10. Then the degrees of freedom of $\mathbf{y'y}$ is *n*, of $\mathbf{y'P_Xy}$ is *q* and of $\mathbf{y'R_Xy}$ is *n*–*q*.

Theorem II.15: Let **A** be an $n \times n$ symmetric matrix and **Y** be an $n \times 1$ normally distributed random vector with $E[\mathbf{AY}] = \mathbf{0}$ and $\text{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$. Then $\left(1/\sigma^2\right)\mathbf{y'Ay}$ follows a chi-squared distribution with *r* degrees of freedom if and only if **A** is idempotent of rank *r*.

Theorem II.16: Let **Y** be an $n \times 1$ normally distributed random vector with $E[\mathbf{Y}] = \mathbf{X\theta}$ and $\text{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$. Let $\mathbf{Y'A}_1\mathbf{Y}, \mathbf{Y'A}_2\mathbf{Y}, \ldots, \mathbf{Y'A}_m\mathbf{Y}$ be a collection of *m* quadratic forms where, for each $i = 1, 2, \ldots, m$, $\mathbf{A}_i$ is symmetric, of rank r_i and $E[\mathbf{A}_i\mathbf{Y}] = \mathbf{0}$. If any two of the following three statements are true, then for each *i*, $\left(1/\sigma^2\right)\mathbf{Y'A}_i\mathbf{Y}$ follows a chi-squared distribution with r_i degrees of freedom. Furthermore, $\mathbf{Y'A}_i\mathbf{Y}$ are independent for $i \neq j$ and $\sum_{i=1}^m r_i = r$ where *r* denotes the rank of $\sum_{i=1}^m \mathbf{A}_i$.

1. All $\mathbf{A}_i$ are idempotent
2. $\sum_{i=1}^m \mathbf{A}_i$ is idempotent
3. $\mathbf{A}_i\mathbf{A}_j = \mathbf{0}, \quad i \neq j$

Theorem II.17: Let U_1 and U_2 be two random variables distributed as chi-squares with r_1 and r_2 degrees of freedom. Then, provided U_1 and U_2, are independent, the random variable $W = \dfrac{U_1/r_1}{U_2/r_2}$ is distributed as Snedecor's *F* with r_1 and r_2 degrees of freedom.

Definition II.11: The estimator of the Residual sums of squares or **deviance** after fitting the model involving parameters $\boldsymbol{\theta}$ is denoted $D(\boldsymbol{\theta})$. The estimator for the Residual sum of squares after fitting the **null model** $E[\mathbf{Y}] = \mathbf{0}$ is $\mathbf{Y'Y}$. The estimator of the **reduction in sum of squares** in going from one model that involves parameters, say $\boldsymbol{\theta}_2$, to a second model that involves more parameters, say $\boldsymbol{\theta}$ made up of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, is denoted $R(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) = D(\boldsymbol{\theta}_2) - D(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and is read 'the reduction in the sums of squares for including the parameters $\boldsymbol{\theta}_1$ in the model given that the parameters $\boldsymbol{\theta}_2$ are already in the model'. The estimator of the reduction in the sum of squares for adding parameters $\boldsymbol{\theta}' = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]'$ to the null model is denoted $R(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{Y'Y} - D(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$; that is, the model parameters **0** are not included.

Definition III.4: For two models, $\boldsymbol{\psi} = \mathbf{X}_1\boldsymbol{\theta}_1$ and $\boldsymbol{\psi} = \mathbf{X}_2\boldsymbol{\theta}_2$, the first model is **marginal** to the second if $\mathcal{C}(\mathbf{X}_1) \subseteq \mathcal{C}(\mathbf{X}_2)$, that is if the columns of $\mathbf{X}_1$ can be written as linear combinations of the columns of $\mathbf{X}_2$.