Andrew J. Spieker, PhD
BIOS 7345 - Advanced Regression Analysis I (Fall 2023)
Document of numbered theorems and lemmas (9/12/2023)

---

1 Review of linear algebra

Lemma 1.1: Property of $\mathbf{A}^T\mathbf{A}$

$\mathbf{A}^T\mathbf{A} = \mathbf{0}$ if and only if $\mathbf{A} = \mathbf{0}$.

Lemma 1.2: Another property involving $\mathbf{A}^T\mathbf{A}$

If $\mathbf{B}\mathbf{A}^T\mathbf{A} = \mathbf{C}\mathbf{A}^T\mathbf{A}$, then $\mathbf{B}\mathbf{A}^T = \mathbf{C}\mathbf{A}^T$.

Theorem 1.1: Symmetric matrices and projection matrices

A symmetric $N \times N$ matrix, $\mathbf{A}$, can be realized as a linear combination of N mutually orthogonal projection matrices.

Theorem 1.2: A condition for $\mathbf{A}^T\mathbf{A}$ to be positive definite

If $\mathbf{A}$ is an $N \times K$ matrix of rank K, then $\mathbf{A}^T\mathbf{A}$ is positive definite.

2 Random vectors and matrices

Theorem 2.1: Expectations of quadratic forms

Let $\mathbf{x}$ denote a random length-N vector (denote $\mathrm{E}[\mathbf{x}] = \boldsymbol{\mu}$ and $\mathrm{Cov}[\mathbf{x}] = \boldsymbol{\Sigma}$). If $\mathbf{A}$ is a constant $N \times N$ matrix, then:

$$\mathrm{E}[\mathbf{x}^T\mathbf{A}\mathbf{x}] = \mathrm{trace}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}.$$

Lemma 2.1: Transforming multivariate normal distributions

Suppose that $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a random length-N vector. Then, for a $K \times N$ constant matrix $\mathbf{C}$, $\mathbf{C}\mathbf{y} \sim \mathcal{N}(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)$.

Lemma 2.2: Univariate-multivariate relationship

$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if and only if $\mathbf{c}^T\mathbf{y}$ is normally distributed for all $\mathbf{c} \neq \mathbf{0}$.

Lemma 2.3: Orthogonal transformations

Let $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, and let $\mathbf{Q}$ denote an orthogonal matrix. Then, $\mathbf{Q}\mathbf{y} \sim \mathcal{N}(\mathbf{Q}\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Theorem 2.2: Distribution of quadratic forms

If $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} > 0$, then $(\mathbf{y} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \sim \chi^2_N$.

Theorem 2.3: Distribution of quadratic forms idempotent case

If $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, and let $\mathbf{P}$ be an $N \times N$ symmetric matrix of rank R. Then $(\mathbf{y} - \boldsymbol{\mu})^T\mathbf{P}(\mathbf{y} - \boldsymbol{\mu})/\sigma^2 \sim \chi^2_R$ if and only if $\mathbf{P}$ is idempotent (and hence a projection matrix).

Theorem 2.4: Independence of $\mathbf{y}^T\mathbf{A}\mathbf{y}$ and $\mathbf{B}\mathbf{y}$

If $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and let $\mathbf{A}$ and $\mathbf{B}$ be constant matrices. Then $\mathbf{y}^T\mathbf{A}\mathbf{y}$ and $\mathbf{B}\mathbf{y}$ are independent if and only if $\mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{0}$.

Theorem 2.5: Independence of $\mathbf{y}^T\mathbf{P}_1\mathbf{y}$ and $\mathbf{y}^T\mathbf{P}_2\mathbf{y}$

If $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and let $\mathbf{P}_1$ and $\mathbf{P}_2$ be projection matrices. Then, $\mathbf{y}^T\mathbf{P}_1\mathbf{y}$ and $\mathbf{y}^T\mathbf{P}_2\mathbf{y}$ are independent if and only if $\mathbf{P}_2\boldsymbol{\Sigma}\mathbf{P}_1 = \mathbf{0}$.

Lemma 2.4: Difference of projection matrices

Suppose that $\mathbf{P}_1$ and $\mathbf{P}_2$ are projection matrices. If $\mathbf{P}_1 - \mathbf{P}_2 \succcurlyeq 0$, then:

- $\mathbf{P}_1\mathbf{P}_2 = \mathbf{P}_2\mathbf{P}_1 = \mathbf{P}_2$.
- $\mathbf{P}_1 - \mathbf{P}_2$ is a projection matrix.

Theorem 2.6: χ^2 distribution and projection matrices

Let $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ be length-N, and $Q_i = (\mathbf{y} - \boldsymbol{\mu})^T\mathbf{P}_i(\mathbf{y} - \boldsymbol{\mu})/\sigma^2$, where $\mathbf{P}_1$ and $\mathbf{P}_2$ are symmetric $N \times N$ matrices. If $Q_i \sim \chi^2_{R_i}$ and $Q_1 - Q_2 \geq 0$, then $Q_1 - Q_2 \sim \chi^2_{R_1 - R_2}$.

Theorem 2.7: The non-central χ^2 and projection matrices

Let $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_N, \sigma^2\mathbf{I}_{N\times N})$ and let $\mathbf{P}$ be a symmetric matrix of rank R. Then, $\mathbf{P}$ is idempotent (and hence an orthogonal projection matrix) if and only if:

$$\frac{1}{\sigma^2}\mathbf{y}^T\mathbf{P}\mathbf{y} \sim \chi^2_R(\boldsymbol{\mu}^T\mathbf{P}\boldsymbol{\mu}/2\sigma^2).$$

3 Ordinary least squares (full-rank)

Lemma 3.1: An important decomposition

$\mathbf{y}$ can be decomposed as $\mathbf{y} = \widehat{\mathbf{y}} + \widehat{\boldsymbol{\epsilon}}$, where $\widehat{\mathbf{y}} \in \mathcal{C}(\mathbf{X})$ and $\widehat{\boldsymbol{\epsilon}} \in \mathcal{N}(\mathbf{X}^T)$; this decomposition is unique.

Lemma 3.2: $\widehat{\mathbf{y}}$ solves the least squares problem in $\mathcal{C}(\mathbf{X})$

The orthogonal projection of $\mathbf{y}$ onto the linear subspace spanned by the column of $\mathbf{X}$ solves the OLS minimization problem. In other words:

$$\widehat{\mathbf{y}} = \underset{\boldsymbol{\theta} \in \mathcal{C}(\mathbf{X})}{\operatorname{argmin}} \, ||\mathbf{y} - \boldsymbol{\theta}||^2.$$

Lemma 3.3: $\widehat{\boldsymbol{\beta}}$ solves the least squares problem in $\mathbb{R}^K$

A least squares estimate of $\boldsymbol{\beta}$, denoted $\widehat{\boldsymbol{\beta}}$, is a solution to the normal equations, so that $\mathbf{X}^T\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y}$.

Lemma 3.4: Some key properties

- Let $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, where $\mathbf{X}$ is of full column rank. Then,
  1. $\mathbf{P}$ is a projection matrix onto $\mathcal{C}(\mathbf{X})$.
  2. $\mathbf{I} - \mathbf{P}$ is a projection matrix and onto $\mathcal{N}(\mathbf{X}^T) = [\mathcal{C}(\mathbf{X})]^\perp$.
  3. $\operatorname{rank}(\mathbf{I} - \mathbf{P}) = \operatorname{trace}(\mathbf{I} - \mathbf{P}) = N - K$.
  4. $\mathbf{P}\mathbf{X} = \mathbf{X}$.

Theorem 3.1: The Gauss-Markov theorem

Suppose that each of the following conditions is satisfied:

- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{0}$.
- $\mathbf{X}$ has full column rank.
- $\mathrm{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{I}$.

Then, $\boldsymbol{a}^T\widehat{\boldsymbol{\beta}} = \mathbf{a}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ is the unique minimum-variance estimator of $\mathbf{a}^T\boldsymbol{\beta}$ among all unbiased linear estimators of $\boldsymbol{a}^T\boldsymbol{\beta}$.

Theorem 3.2: A central limit theorem for OLS

Let $\mathbf{A} := (\mathbf{X}^T\mathbf{X})^{-1/2}\mathbf{X}^T$ (which is $K \times N$), with columns given by $\mathbf{a}_{N1}, \ldots, \mathbf{a}_{NN}$. Let $\mathbf{X}$ be fixed and of full rank, and suppose the following conditions (Lindeberg conditions) hold:

1. $\sum_{i=1}^N ||\mathbf{a}_{Ni}||^k \longrightarrow 0$ for $k > 2$.
2. $\mathrm{E}[|\epsilon|^k] < \infty$ for $k > 2$.

Then, $(\mathbf{X}^T\mathbf{X})^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \longrightarrow_d \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$.

4 Ordinary least squares (rank-deficient)

Lemma 4.1: Non-uniqueness ⇒ infinitely many solutions

Suppose $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ are two different least squares estimators. Then, there are infinitely many least squares estimators of $\boldsymbol{\beta}$.

Lemma 4.2: Two solutions to the normal equations

Suppose $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ are two different least squares estimators. Then, $||\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_1||^2 = ||\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_2||^2$.

Theorem 4.1: Determining the orthogonal projection

The orthogonal projection, $\widehat{\mathbf{y}}$, of $\mathbf{y}$ onto $\mathcal{C}(\mathbf{X})$ is given by $\mathbf{P}\mathbf{y}$, where

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$$

for any generalized inverse, $(\mathbf{X}^T\mathbf{X})^-$.

Lemma 4.3

If $(\mathbf{X}^T\mathbf{X})^-$ is a g-inverse of $\mathbf{X}^T\mathbf{X}$, then $[(\mathbf{X}^T\mathbf{X})^-]^T$ is also a g-inverse of $\mathbf{X}^T\mathbf{X}$.

Lemma 4.4

If $(\mathbf{X}^T\mathbf{X})^-$ is a g-inverse of $\mathbf{X}^T\mathbf{X}$, then $(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}[(\mathbf{X}^T\mathbf{X})^-]^T$ is a symmetric, reflexive g-inverse of $\mathbf{X}^T\mathbf{X}$.

Lemma 4.5

If $(\mathbf{X}^T\mathbf{X})^-$ is a g-inverse of $\mathbf{X}^T\mathbf{X}$, then $\mathbf{X}(\mathbf{X}^T\mathbf{X})^-$ is a g-inverse of $\mathbf{X}^T$ and $(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ is a g-inverse of $\mathbf{X}$.

Lemma 4.6

If $\mathbf{G}$ and $\widetilde{\mathbf{G}}$ are both g-inverses of $\mathbf{X}^T\mathbf{X}$, then $\mathbf{X}\mathbf{G}\mathbf{X}^T = \mathbf{X}\widetilde{\mathbf{G}}\mathbf{X}^T$.

Lemma 4.7

If $(\mathbf{X}^T\mathbf{X})^-$ is a g-inverse of $\mathbf{X}^T\mathbf{X}$, then $\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ is symmetric.

Lemma 4.8: Conditions for uniqueness

There is a unique solution to:

$$\begin{bmatrix} \widehat{\mathbf{y}} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{C} \end{bmatrix} \widehat{\boldsymbol{\beta}} =: \mathbf{D}\widehat{\boldsymbol{\beta}}$$

for any $\widehat{\mathbf{y}} \in \mathcal{C}(\mathbf{X})$ if and only if $\text{rank}(\mathbf{D}) = K$ and the rows of $\mathbf{C}$ are linearly independent of the rows of $\mathbf{X}$.

Lemma 4.9: Which functions are estimable?

The quantity $\mathbf{a}^T\boldsymbol{\beta}$ is estimable if and only if $\mathbf{a} \in \mathcal{R}(\mathbf{X})$.

Lemma 4.10: Uniqueness

If $\mathbf{a}^T\boldsymbol{\beta}$ is estimable, there is a unique $\mathbf{c}_* \in \mathcal{C}(\mathbf{X})$ such that $\mathbf{a} = \mathbf{X}^T\mathbf{c}_*$.

Lemma 4.11: Property of estimable functions

If $\mathbf{a}^T\boldsymbol{\beta}$ is estimable, then $\mathbf{a}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X} = \mathbf{a}^T$ for any g-inverse $(\mathbf{X}^T\mathbf{X})^-$.

Lemma 4.12: Variance of estimable functions

If $\mathbf{a}^T\boldsymbol{\beta}$ is estimable, then $\text{Var}[\mathbf{a}^T\widehat{\boldsymbol{\beta}}] = \sigma^2\mathbf{a}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{a}$.

Theorem 4.2: The Gauss-Markov theorem (version 2)

- Suppose that each of the following conditions is satisfied:
  - $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\text{E}[\boldsymbol{\epsilon}] = \mathbf{0}$.
  - $\mathbf{X}$ may be rank-deficient ($\text{rank}(\mathbf{X}) = R < K$).
  - $\text{Cov}[\mathbf{y}] = \sigma^2\mathbf{I}$.
- Then, $\mathbf{a}^T\widehat{\mathbf{y}}$ is the unique estimate of $\mathbf{a}^T\mathbf{X}\boldsymbol{\beta}$ that achieves minimum variance among all unbiased linear estimators of $\mathbf{a}^T\mathbf{X}\boldsymbol{\beta}$.

Theorem 4.3: The Gauss-Markov theorem (version 3)

- Suppose that each of the following conditions is satisfied:
  - $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\text{E}[\boldsymbol{\epsilon}] = \mathbf{0}$.
  - $\mathbf{X}$ may be rank-deficient ($\text{rank}(\mathbf{X}) = R < K$).
  - $\text{Cov}[\mathbf{y}] = \sigma^2\mathbf{I}$.
- If $\mathbf{a}^T\boldsymbol{\beta}$ is estimable, then $\mathbf{a}^T\widehat{\boldsymbol{\beta}}$ is unique (i.e., the same for all $\widehat{\boldsymbol{\beta}}$ solving the normal equations), and is the BLUE of $\mathbf{a}^T\boldsymbol{\beta}$.

5 Weighted least squares

Theorem 5.1: The Gauss-Markov theorem (version 4)

Let $\mathbf{X}$ be of full rank, and suppose $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{0}$ and $\mathrm{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{V}$. Then, for any constant vector $\mathbf{a}$, $\mathbf{a}^T\widehat{\boldsymbol{\beta}}^*$ is the BLUE of $\mathbf{a}^T\boldsymbol{\beta}$.

6 Hypothesis testing

Lemma 6.1: Difference between $\mathbf{P}_\Omega$ and $\mathbf{P}_\omega$

$\mathbf{P}_\Omega - \mathbf{P}_\omega = \mathbf{P}_{\Omega\cap\omega^\perp}$, with $\omega^\perp \cap \Omega = \mathcal{C}(\mathbf{P}_\Omega \mathbf{D}^T)$.

Lemma 6.2: Difference between $\mathbf{P}_\Omega$ and $\mathbf{P}_\omega$ (more explicitly)

If $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis, then:

$$\mathbf{P}_\Omega - \mathbf{P}_\omega = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{C}^T(\mathbf{C}(\mathbf{X}^T\mathbf{X})^-\mathbf{C}^T)^-\mathbf{C}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T.$$

Theorem 6.1: Expression for $\mathrm{RSS}_H - \mathrm{RSS}$

If $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis, then:

$$\mathrm{RSS}_H - \mathrm{RSS} = (\mathbf{C}\widehat{\boldsymbol{\beta}})^T(\mathbf{C}(\mathbf{X}^T\mathbf{X})^-\mathbf{C}^T)^-\mathbf{C}\widehat{\boldsymbol{\beta}},$$

where $\widehat{\boldsymbol{\beta}}$ is the OLS estimate from the unrestricted model.

Lemma 6.3: Properties involving full-rank C

If $\mathbf{C}$ is of full (row) rank, then:

- $\mathrm{rank}(\mathbf{C}(\mathbf{X}^T\mathbf{X})^-\mathbf{C}^T) = Q$, which is to say that $\mathbf{C}(\mathbf{X}^T\mathbf{X})^-\mathbf{C}^T$ is non-singular.
- $\mathbf{C}(\mathbf{X}^T\mathbf{X})^-\mathbf{C}^T \succ 0$.

Theorem 6.2: Expectation of $\mathrm{RSS}_H - \mathrm{RSS}$

If $\mathrm{rank}(\mathbf{C}) = Q$, then:

$$\mathrm{E}[\mathrm{RSS}_H - \mathrm{RSS}] = \sigma^2 Q + \boldsymbol{\beta}^T\mathbf{C}^T(\mathbf{C}(\mathbf{X}^T\mathbf{X})^-\mathbf{C}^T)^{-1}\mathbf{C}\boldsymbol{\beta}.$$

Theorem 6.3: Distribution of of $\mathrm{RSS}_H - \mathrm{RSS}$ under H_0

Suppose $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\mathrm{rank}(\mathbf{X}) = R \leq K$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. If $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis with $\mathrm{rank}(\mathbf{C}_{Q\times K}) = Q$, then under the null hypothesis:

$$F = \frac{(\mathrm{RSS}_H - \mathrm{RSS})/Q}{\mathrm{RSS}/(N-R)} \sim F_{Q,N-R}$$

Theorem 6.4: Distribution of of $\mathrm{RSS}_H - \mathrm{RSS}$

Suppose $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\mathrm{rank}(\mathbf{X}) = R \leq K$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. If $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis with $\mathrm{rank}(\mathbf{C}_{Q\times K}) = Q$, then:

$$F = \frac{(\mathrm{RSS}_H - \mathrm{RSS})/Q}{\mathrm{RSS}/(N-R)} \sim F_{Q,N-R}(\lambda),$$

where $\lambda = \boldsymbol{\beta}\mathbf{X}^T(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{X}\boldsymbol{\beta}/2\sigma^2$.

7 Analysis of variance (ANOVA)

No numbered theorems or lemmas.

8 Model misspecification

No numbered theorems or lemmas.

9 Confidence regions and prediction

Lemma 9.1: Relationship between the t- and F-distribution

Let $Z \sim \mathcal{N}(0,1)$ and let $U \sim \chi^2_K$, with $U \perp\!\!\!\perp Z$. Further, let

$$T = \frac{Z}{\sqrt{U/K}}.$$

Then, $T \sim t_K$ and $T^2 \sim F(1,K)$.