回归分析

Regression Analysis

武萍 吴贤毅 著
Wu Ping Wu Xianyi

清华大学出版社

回归分析

Regression Analysis

武萍　吴贤毅　著

Wu Ping　Wu Xianyi

清华大学出版社

北　京

内 容 简 介

本书内容分为三部分：(1) 线性回归分析所需要的矩阵理论、多元正态分布；(2) 线性回归的基本理论和方法，包括线性估计的一般小样本理论、关于线性假设的 F-检验方法、基于线性模型的方差分析理论、变量选择问题的讨论、共线性问题、异常值问题以及 Box-Cox 模型等与线性回归相关的内容；(3) 用于分类响应变量的 Logist 回归模型的基本理论和方法。

本书要求读者具有高等代数(或者线性代数)和概率论与数理统计的良好基础。本书的特点之一是在尽可能少的基础知识要求下讲清线性回归分析的理论问题，同时，本书也附带了一些 SAS 代码，这将有助于实际应用中的数据处理。

本书可供统计学专业、数学专业或者其他相关专业作为本科生回归分析课程教材使用，也可作为非统计学专业的研究人员学习回归分析基础理论的参考书使用。

图书在版编目(CIP)数据

回归分析：英文/武萍，吴贤毅著. --北京：清华大学出版社，2016

ISBN 978-7-302-44630-9

Ⅰ.①回…　Ⅱ.①武…　②吴…　Ⅲ.①回归分析–高等学校–教材–英文　Ⅳ.①O212.1

中国版本图书馆 CIP 数据核字(2016)第 179371 号

责任编辑： 汪　操
封面设计： 常雪影
责任校对： 赵丽敏
责任印制： 何　芊

出版发行： 清华大学出版社
网　　址： http://www.tup.com.cn, http://www.wqbook.com
地　　址： 北京清华大学学研大厦 A 座　　**邮　　编：** 100084
社 总 机： 010-62770175　　**邮　　购：** 010-62786544
投稿与读者服务： 010-62776969, c-service@tup.tsinghua.edu.cn
质量反馈： 010-62772015, zhiliang@tup.tsinghua.edu.cn
印 刷 者： 三河市君旺印务有限公司
装 订 者： 三河市新茂装订有限公司
经　　销： 全国新华书店
开　　本： 170mm×230mm　　**印　张：** 14.25　　**字　数：** 269 千字
版　　次： 2016 年 8 月第 1 版　　**印　次：** 2016 年 8 月第 1 次印刷
定　　价： 39.00 元

产品编号：067956-01

Preface

As one of the cornerstones of statistics, regression analysis (especially linear regression analysis) has been playing a significant role in both theoretical research and real-world data analysis for more than one hundred years after Sir Francis Galton invented the term "regression" when he was investigating the relationship between a child and his/her parents. Even in nowadays, this oldest subject still attracts remarkable attentions and interests from active researchers in related areas so that a multitude of relevant academic journals have been publishing countless research papers, not to mention the immense number of efforts made by data analysts using regression analysis techniques from almost all disciplines in the world, including economics, finance, biology, medicine science, healthcare, to name just a few.

This book grows out of the courses we have taught in East China Normal University in the past several years. The most of the students attending those courses majored in statistics and the others are from certain related subjects such as mathematics, economical statistics, financial engineering and insurance and actuarial science.

This book consists of eight Chapters structured as what follows.

(1) The first two Chapters are for preliminaries providing what are believed to form the least required set of preparations on matrix algebra that are generally not taught in the ordinary course under the title "Advanced Algebra" or "Linear Algebra", but are necessary for the deduction of some of linear regression theory, and multivariate normal distributions, which are fundamental for developing the distributional theory and the hypothesis tests in linear regression. We would like to note that, these chapters don't provide sufficient preliminaries in matrices and in order to smoothly go through this book, it is necessary for one to be well trained in linear algebra. Certainly, readers who have got well trained in matrix theory and multivariate normal distributions can simply skip these two chapters.

(2) Chapter 3 is an introduction to the conception of regression models, in which a regression function is interpreted as both conditional expectation given the explanatory variables and linear projection on the space consisting of linear

combinations of the explanatory variables.

(3) The next two chapters are the core parts of this book treating the classical theory on estimation and hypothesis tests in linear regression analysis. Chapter 4 develops the least square estimates, their properties, least square estimates under linear restrictions, generalized LS estimates, and collinearity among explanatory variables. In order to understand Section 4.7, the reader is required to know the basic elements of Bayesian analysis such as loss function, Bayesian risk and Bayesian estimation. Chapter 5 is dedicated to the theory of hypothesis test for linear hypotheses, its parallel theory of confidence regions and a direct application of linear hypothesis test: goodness-of-fit test. Multiple tests for a set of hypotheses are also discussed. Prediction and calibration regions are also discussed in this chapter. Some related quantities such as multiple correlation coefficients and adjusted multiple correlation coefficeints are introduced in this chapter. Moreover, the techniques for analysis of variance are also discussed. Preceding knowledgy on hypothesis test, including significance level, critical value, power, (noncentroal) χ^2, (noncentroal)t- and (noncentroal) F-distribution are required.

(4) Chapter 6 is mainly on variable selection and Chapter 7 provides some miscellaneous topics in linear regression models. In Chapter 6, we introduce a few classical methodologies as well as some newly developed techniques developed for variable selection. While most of the assertions on those classical methodologies are proved, some (especially the proofs involving the theory of asymptotic statistics) are not necessary for understanding the rest of the book. Chapter 7 presents some more interesting topics including collinearity, outliers, testing for heteroskedasticity of errors and two extensions of linear regression models.

(5) Chapter 8 is a separated one that provides a concise and application-oriented introduction on logistic regression models which provide a technique for data analysis when the response is categorical data (binary, multinomial and ordinal data). Because it is application-oriented, the rigor in theory is to some extent not considered very much carefully.

It has long been a consensus that statistics is made of two aspects: science and art. This book puts more focus on the theoretical aspects of regression (the aspect of science) than introduction of a set of data analysis methodologies for regression (the art aspect). The computation involved is recommended to be carried out by means of softwares for statistics. Thus, we provided SAS code for the examples so that the students can learn from it how to use SAS to produce the numerical results of the data under analysis. Besides, a significant feature of this book is that, in theoretical deduction, we also do our best to provide why we deduce as that way, with the attempt to make students know both how and why.

The content of this book may be more than what is needed for a course of 3-4 hours a week. According to our experience in the past years, in a course of 3 hours a week we can taught the first 5 Chapters or 6 Chapters, depending on how the students are trained at the theory and skills of matrix and probability. In addition, there are some contents (sections, subsections or proofs) that are not required for following parts and can be skipped at the first reading. These contents are labelled with the symbol *.

Finally, we are thankful to the Department of Statistics and Actuarial Science of East China Normal University for her kind teaching assignment, so that we have a chance to teach the courses under the name Regression Analysis for many rounds. Many thanks would also be given to the students who have attended the courses in the past years.

Wu Ping and Wu Xianyi

February, 2016

Contents

Chapter 1

Preliminaries: Matrix Algebra and Random Vectors

1.1 Preliminary matrix algebra

This section prepares the necessary preliminaries for matrix algebra for later use in this book. For comprehensive resources in matrix algebra for statisticians, one can be referred to, e.g., the books Graybill (1983) or Seber (2008).

1.1.1 Trace and eigenvalues

Trace of a square matrix is the sum of the diagonal entries, i.e., $\text{tr}(\boldsymbol{A}_{n\times n}) = \sum_{i=1}^{n} a_{ii}$. A number (possibly complex) λ is an eigenvalue of a matrix $\boldsymbol{A}$ if there exists some *nonzero* complex n-vector $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. Throughout the book, for any matrix operation, it is implicitly assumed that the matrices are conformable (i.e., dimensions involved are suitable for defining the operations).

Proposition 1.1 (1) $\text{tr}(\boldsymbol{A}+\boldsymbol{B}) = \text{tr}(\boldsymbol{A}) + \text{tr}(\boldsymbol{B})$.

(2) $\text{tr}(\boldsymbol{A}_{m\times n}\boldsymbol{B}_{n\times m}) = \text{tr}(\boldsymbol{B}\boldsymbol{A})$.

(3) *The nonzero eigenvalues of* $\boldsymbol{A}\boldsymbol{B}$ *are the same as those of* $\boldsymbol{B}\boldsymbol{A}$. *Thus, if* $\boldsymbol{A}$ *and hence* $\boldsymbol{B}$ *are square, then eigenvalues of* $\boldsymbol{A}\boldsymbol{B}$ *and* $\boldsymbol{B}\boldsymbol{A}$ *are identical.*

Proof. (2)

$$\begin{aligned}\text{tr}(\boldsymbol{A}_{m\times n}\boldsymbol{B}_{n\times m}) &= \sum_{i=1}^{m}(\boldsymbol{A}\boldsymbol{B})_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{m} b_{ji}a_{ij} \\ &= \sum_{j=1}^{n}(\boldsymbol{B}\boldsymbol{A})_{jj} = \text{tr}(\boldsymbol{B}\boldsymbol{A}).\end{aligned}$$

(3) Let $\lambda \neq 0$ and $\boldsymbol{ABx} = \lambda \boldsymbol{x}$ for some $\boldsymbol{x} \neq \boldsymbol{0} \implies \boldsymbol{Bx} \neq \boldsymbol{0}$ and $\boldsymbol{BABx} = \lambda \boldsymbol{Bx}$. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices. $\boldsymbol{AB}$ has an eigenvalue 0 means that $\boldsymbol{AB}$ is singular. Hence, so is $\boldsymbol{BA}$, implying that $\boldsymbol{BA}$ has an eigenvalue 0. ■

Remark 1.1 *For Proposition* 1.1, *we note that there exist many couples of matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *such that* $\boldsymbol{AB}$ *has the eigenvalue* 0 *but* $\boldsymbol{BA}$ *has not. For example, if* $\boldsymbol{A}' = \boldsymbol{B} = \begin{pmatrix} \boldsymbol{I}_k & \boldsymbol{0} \end{pmatrix}$, *where* $\boldsymbol{I}_k$ *is the identity matrix of order* k. *Then* $\boldsymbol{AB} = \begin{pmatrix} \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$ *has an eigenvalue* 0 *but* $\boldsymbol{BA} = \boldsymbol{I}_k$ *has not. This holds generally for the case* $\boldsymbol{A}_{m\times n}(n < m)$ *is of full column rank and* $\boldsymbol{B}_{n\times m}$ *is of full row rank* n.

Proposition 1.2 *Let* $\boldsymbol{A}_{n\times n}$ *be a square matrix with* n *eigenvalues* $\lambda_i, i = 1, 2, \cdots, n$. *Then*

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i \quad \textit{and} \quad \det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i,$$

where det *means the determinant.*

Proof. Write $\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise} \end{cases}$ for the indicator of $\{i = j\}$. Check the eigenpolynomial

$$\begin{aligned} f(\lambda) &= \det(\lambda \boldsymbol{I}_n - \boldsymbol{A}) \\ &= \det(\delta_{ij}\lambda - a_{ij})_{n\times n} \\ &= \sum_{j_1 \neq j_2 \neq \cdots \neq j_n} (-1)^* \left(\lambda\delta_{1j_1} - a_{1j_1}\right) \cdots \left(\lambda\delta_{nj_n} - a_{nj_n}\right) \\ &\quad \text{(where * is an appropriately defined positive integer)} \\ &= \prod_{i=1}^{n} (\lambda - \lambda_i) \\ &= \lambda^n - \left(\sum_{i=1}^{n} \lambda_i\right) \lambda^{n-1} + \cdots + (-1)^n \prod_{i=1}^{n} \lambda_i. \end{aligned} \tag{1.1.1}$$

Expanding $\det(\lambda \boldsymbol{I}_n - \boldsymbol{A})$, the term including λ^{n-1} exists only in $(\lambda - a_{11}) \cdots (\lambda - a_{nn})$ and thus $\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii} = \mathrm{tr}(\boldsymbol{A})$.

The second assertion follows from again the identity in (1.1.1) by setting $\lambda = 0$, i.e., the relationship $f(0) = (-1)^n \det(\boldsymbol{A}) = (-1)^n \prod_{i=1}^n \lambda_i$. ■

1.1.2 Symmetric matrices

Proposition 1.3 *Let $\boldsymbol{A}_{n\times n}$ be a symmetric matrix of real numbers. Then:*

(1) The eigenvalues of $\boldsymbol{A}$ are also real numbers and for each eigenvalue we can have an eigenvector of real numbers.

(2) Any two eigenvectors of two different eigenvalues are orthogonal.

(3) The matrix $\boldsymbol{A}$ has n real eigenvalues, writing $\lambda_1, \lambda_2, \cdots, \lambda_n$ and n eigenvectors that can be orthogonalized.

(4) As a result, for any symmetric matrix, there is an orthogonal matrix $\boldsymbol{P}$ such that

$$\boldsymbol{A} = \boldsymbol{P}\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}\boldsymbol{P}' \overset{\text{def}}{=} \boldsymbol{P\Lambda P}', \tag{1.1.2}$$

which is usually referred to as a spectral decomposition of $\boldsymbol{A}$.

(5) For any integer $s > 0$, $\boldsymbol{A}^s = \boldsymbol{P}\begin{pmatrix} \lambda_1^s & & \\ & \ddots & \\ & & \lambda_n^s \end{pmatrix}\boldsymbol{P}' \overset{\text{def}}{=} \boldsymbol{P\Lambda}^s\boldsymbol{P}'$ so that $\text{tr}(\boldsymbol{A}^s) = \sum_{i=1}^n \lambda_i^s$.

(6) $\boldsymbol{A}$ is nonsingular if and only if $\lambda_i \neq 0$ for all i. And the eigenvalues of $\boldsymbol{A}^{-1}$ are $\lambda_i^{-1}, i = 1, 2, \cdots, n$.

(7) The eigenvalues of $(\boldsymbol{I}_n + c\boldsymbol{A})$ are $1 + c\lambda_i, i = 1, 2, \cdots, n$.

(8) The eigenvalues of a positive semidefinite (p.s.d.) matrix are nonnegative; as a straightforward result, $\boldsymbol{A}$ is a p.s.d. matrix then $\text{tr}(\boldsymbol{A}) \geqslant 0$.

Proof. (1) Let λ be an eigenvalue of $\boldsymbol{A}$. Then, there is an n-vector $\boldsymbol{x}$ (maybe complex) such that $\boldsymbol{Ax} = \lambda\boldsymbol{x}$. Taking the conjugate operation, we see that $\bar{\boldsymbol{A}}\bar{\boldsymbol{x}} = \bar{\lambda}\bar{\boldsymbol{x}}$, i.e., $\boldsymbol{A}\bar{\boldsymbol{x}} = \bar{\lambda}\bar{\boldsymbol{x}}$. Hence $\boldsymbol{x}'\boldsymbol{A}\bar{\boldsymbol{x}} = \bar{\lambda}\boldsymbol{x}'\bar{\boldsymbol{x}}$. On the other hand,

$$\boldsymbol{x}'\boldsymbol{A}\bar{\boldsymbol{x}} = \bar{\boldsymbol{x}}'\boldsymbol{A}\boldsymbol{x} = \lambda\bar{\boldsymbol{x}}'\boldsymbol{x}.$$

Thus $\bar{\lambda} = \lambda$.

Let $\boldsymbol{x} = \boldsymbol{a} + \boldsymbol{b}\mathrm{i} \neq \boldsymbol{0}$ be an eigenvector corresponding to an eigenvalue λ. Then

$$\boldsymbol{Ax} = \lambda\boldsymbol{x} = \lambda\boldsymbol{a} + \lambda\boldsymbol{b}\mathrm{i} = \boldsymbol{Aa} + \boldsymbol{Ab}\mathrm{i}.$$

That is,

$$\boldsymbol{Aa} = \lambda\boldsymbol{a} \quad \text{and} \quad \boldsymbol{Ab} = \lambda\boldsymbol{b}.$$

Therefore, the real eigenvector can be taken as $\boldsymbol{a}$ or $\boldsymbol{b}$, whichever is nonzero.

(2) Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two eigenvectors such that $\boldsymbol{Ax} = \lambda\boldsymbol{x}$ and $\boldsymbol{Ay} = \gamma\boldsymbol{y}$ and $\lambda \neq \gamma$. Then

$$\boldsymbol{x}'\boldsymbol{Ay} = \gamma\boldsymbol{x}'\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{Ax} = \lambda\boldsymbol{y}'\boldsymbol{x} = \lambda\boldsymbol{x}'\boldsymbol{y}.$$

That is, $\boldsymbol{x}'\boldsymbol{y} = \boldsymbol{0}$.

(3) The proof of the first part is too long to be presented here. For the second part, using Gram-Schmidt orthogonalization procedure and the assertion in (2) yields the result.

The proof of the rest items are left as exercise. ■

From now on, we will discuss real symmetric matrices only.

We next discuss some more facts on positive semi-definite matrices.

Proposition 1.4 *Let $\boldsymbol{\Sigma}_{m\times m}$ be a symmetric matrix of order m. Then the following statements are equivalent:*

(1) *$\boldsymbol{\Sigma}$ is a positive semi-definite matrix with $r = \mathrm{rank}(\boldsymbol{\Sigma})$.*

(2) *There exist matrices $\boldsymbol{P}_{n\times r}$ and $\boldsymbol{Q}_{n\times n-r}$ such that $(\boldsymbol{P} \vdots \boldsymbol{Q})$ is an orthogonal matrix and*

$$\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Lambda}_r\boldsymbol{P}' \quad \textit{and} \quad \boldsymbol{Q}'\boldsymbol{\Sigma}\boldsymbol{Q} = \boldsymbol{0}. \tag{1.1.3}$$

(3) *There exist matrices $\boldsymbol{P}_{n\times r}$ and $\boldsymbol{Q}_{n\times n-r}$ such that $(\boldsymbol{P} \vdots \boldsymbol{Q})$ is an orthogonal matrix and*

$$\boldsymbol{P}'\boldsymbol{\Sigma}\boldsymbol{P} = \boldsymbol{\Lambda}_r \quad \textit{and} \quad \boldsymbol{Q}'\boldsymbol{\Sigma}\boldsymbol{Q} = \boldsymbol{0}. \tag{1.1.4}$$

Proof. (1)$\Longrightarrow$(2). Let $\boldsymbol{\Sigma}_{m\times m}$ have r positive eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_r$, associated with an $m \times r$ matrix $\boldsymbol{P}$, the columns of which are orthogonal eigenvectors belonging to $\lambda_1, \lambda_2, \cdots, \lambda_r$ and $m - r$ folds zero eigenvalue, associated with $m - r$ orthogonal eigenvectors, by which a matrix $\boldsymbol{Q}$ is composed, such that

$$\begin{pmatrix} \boldsymbol{P}' \\ \boldsymbol{Q}' \end{pmatrix} \boldsymbol{\Sigma} \begin{pmatrix} \boldsymbol{P} & \boldsymbol{Q} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \tag{1.1.5}$$

where $\boldsymbol{\Lambda}_r = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r)$. Hence,

$$\boldsymbol{Q}'\boldsymbol{\Sigma}\boldsymbol{Q} = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{P} & \boldsymbol{Q} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{P}' \\ \boldsymbol{Q}' \end{pmatrix} = \boldsymbol{P}\boldsymbol{\Lambda}_r\boldsymbol{P}', \qquad (1.1.6)$$

(2)$\Longrightarrow$(3). It is obvious.

(3)$\Longrightarrow$(1). Note that $\mathrm{rank}(\boldsymbol{\Sigma}) = r$ is clear. Suppose that

$$\begin{pmatrix} \boldsymbol{P}' \\ \boldsymbol{Q}' \end{pmatrix} \boldsymbol{\Sigma} \begin{pmatrix} \boldsymbol{P} & \boldsymbol{Q} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_r & \boldsymbol{A} \\ \boldsymbol{A}' & \boldsymbol{0} \end{pmatrix}.$$

Then

$$r = \mathrm{rank}(\boldsymbol{\Sigma}) = \mathrm{rank} \begin{pmatrix} \boldsymbol{\Lambda}_r & \boldsymbol{A} \\ \boldsymbol{A}' & \boldsymbol{0} \end{pmatrix} = \mathrm{rank} \begin{pmatrix} \boldsymbol{\Lambda}_r & \boldsymbol{0} \\ \boldsymbol{0}' & -\boldsymbol{A}'\boldsymbol{\Lambda}_r^{-1}\boldsymbol{A} \end{pmatrix}.$$

Hence $\boldsymbol{A} = \boldsymbol{0}$, implying that $\boldsymbol{\Sigma}$ is positive semi-definite. ■

Corollary 1.1 *Write $\boldsymbol{R} = \boldsymbol{P}\boldsymbol{\Lambda}_r^{1/2}$ and $\boldsymbol{S} = \boldsymbol{\Lambda}_r^{-1/2}\boldsymbol{P}$, then $\boldsymbol{R}$ and $\boldsymbol{S}$ are both of full column rank and*

$$\boldsymbol{\Sigma} = \boldsymbol{R}\boldsymbol{R}' \quad \textit{and} \quad \boldsymbol{S}'\boldsymbol{\Lambda}_r\boldsymbol{S} = \boldsymbol{I}_r. \qquad (1.1.7)$$

The following two propositions link the eigenvalues with respect to the quadratic form of matrices.

Proposition 1.5 *Let $\boldsymbol{A}$ be a symmetric matrix. Then*

$$\max_{\boldsymbol{x}'\boldsymbol{x}\neq 0} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \lambda_{\max}(\boldsymbol{A}) \quad \textit{and} \quad \min_{\boldsymbol{x}'\boldsymbol{x}\neq 0} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \lambda_{\min}(\boldsymbol{A}).$$

Proof. By (1.1.2),

$$\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \frac{\boldsymbol{x}'\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{P}\boldsymbol{P}'\boldsymbol{x}} = \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2},$$

where $\boldsymbol{y} = \boldsymbol{P}'\boldsymbol{x} = (y_1, y_2, \cdots, y_n)'$. The assertion thus follows because

$$\lambda_{\min} \leqslant \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2} \leqslant \lambda_{\max}$$

and the both extremes can be achieved by properly selected $\boldsymbol{y}$. ■

Proposition 1.6 *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two positive definite matrices with $\boldsymbol{A} \geqslant \boldsymbol{B}$,* i.e., $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ *for all* $\boldsymbol{x}$*. Then* $\boldsymbol{A}^{-1} \leqslant \boldsymbol{B}^{-1}$.

Proof. Now that $\boldsymbol{A} > \boldsymbol{0}$ and $\boldsymbol{B} > \boldsymbol{0}$, imply that $\boldsymbol{x} \neq \boldsymbol{0} \iff \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} \neq \boldsymbol{0}$, then

$$\begin{aligned}
\boldsymbol{A} \geqslant \boldsymbol{B} &\iff \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} \text{ for all } \boldsymbol{x} \\
&\iff \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}} \geqslant \boldsymbol{1} \text{ for all } \boldsymbol{x} \neq \boldsymbol{0} \\
&\iff \frac{\boldsymbol{y}'\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}} \geqslant \boldsymbol{1} \text{ for all } \boldsymbol{y} \neq \boldsymbol{0} \text{ (by taking } \boldsymbol{y} = \boldsymbol{B}^{\frac{1}{2}}\boldsymbol{x}) \\
&\iff \lambda_{\min}(\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{-\frac{1}{2}}) \geqslant 1 \text{(by Proposition 1.5)} \\
&\iff \lambda_{\max}(\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}^{-1}\boldsymbol{B}^{\frac{1}{2}}) \leqslant 1 \text{(by Proposition 1.3 (6))} \\
&\iff \frac{\boldsymbol{y}'\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}^{-1}\boldsymbol{B}^{\frac{1}{2}}y}{\boldsymbol{y}'\boldsymbol{y}} \leqslant \boldsymbol{1} \text{ for all } \boldsymbol{y} \neq \boldsymbol{0} \text{(by again Proposition 1.5)} \\
&\iff \frac{\boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{B}^{-1}\boldsymbol{x}} \leqslant \boldsymbol{1} \text{ for all } \boldsymbol{x} \neq \boldsymbol{0} \text{ (by taking } \boldsymbol{x} = \boldsymbol{B}^{\frac{1}{2}}\boldsymbol{y}) \\
&\iff \boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x} \leqslant \boldsymbol{x}'\boldsymbol{B}^{-1}\boldsymbol{x} \text{ for all } \boldsymbol{x} \\
&\iff \boldsymbol{A}^{-1} \leqslant \boldsymbol{B}^{-1}.
\end{aligned}$$

This completes the proof. ■

1.1.3 Idempotent matrices and orthogonal projection

Theorem 1.1 *Let $\boldsymbol{A}$ be a symmetric matrix of order n. Then $\boldsymbol{A}^2 = \boldsymbol{A}$ (referred to as idempotent) if and only if it has r eigenvalues 1 and $n-r$ eigenvalues zero for some integer r (under this situation, rank$(\boldsymbol{A}) = r$).*

Proof. Let $\boldsymbol{P}$ be such that $\boldsymbol{A} = \boldsymbol{P}\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}\boldsymbol{P}'$. Then $\boldsymbol{A}^2 =$

$$\boldsymbol{P}\begin{pmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{pmatrix}\boldsymbol{P}' = \boldsymbol{A} = \boldsymbol{P}\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}\boldsymbol{P}' \text{ if and only if } \lambda_i^2 = \lambda_i.$$

This gives the result. Moreover, the rank of $\boldsymbol{A}$ is equal to the number of 1s in the eigenvalues of $\boldsymbol{A}$. ■