Statistical and Application 统计学与应用, 2015, 4, 47-55
Published Online June 2015 in Hans. http://www.hanspub.org/journal/sa
http://dx.doi.org/10.12677/sa.2015.42007

Hans汉斯

The Effects of Different Response Values in Linear Regression Model on Binary Classification

Xiaoying Wang, Yanli Yang, Changlong Chen

School of Mathematics and Physics, North China Electrical Power University, Beijing
Email: yangyanlibang@163.com

Received: Jun. 5th, 2015; accepted: Jun. 20th, 2015; published: Jun. 25th, 2015

Abstract

We use the multiple linear regression model to deal with the classification problem of two populations. Firstly, we assign the response variables and some corresponding values with certain rules, and then construct discriminant function and criterion via least square method. On this basis, we discuss the effects of different response values on classification for balanced and unbalanced data in linear model. In addition, we compare the mentioned discriminant method above with classic discriminant methods including the classical Mahalanobis distance discriminant and Bayes discriminant. At last, we find the inner relation between these methods as well as their advantages and disadvantages.

Keywords

Binary Classification, Response Values, Discriminant Analysis, Linear Regression Model, Least Square

线性回归模型中响应值的选取对二分类问题的影响

王小英，杨岩丽，陈常龙

华北电力大学数理学院，北京

Email: yangyanlibang@163.com

收稿日期：2015年6月5日；录用日期：2015年6月20日；发布日期：2015年6月25日

摘　要

我们利用多元线性回归模型处理两个总体的分类问题，首先对响应变量按一定的规则赋值，并在最小二乘法的基础上构建判别函数及判别准则，进而论证了响应值的选取对平衡及不平衡数据二分类问题的影响。此外，我们将此判别方法与经典判别分析方法如：经典马氏距离判别法、Bayes判别法进行比较，并得到它们之间的内在联系及优缺点。

关键词

二分类问题，响应值选取，判别分析，线性回归模型，最小二乘法

1. 引言

考虑二总体的分类问题，已知有两个总体G_1和G_2，且假定$G_1 \sim N(\mu_1,\Sigma)$，$G_2 \sim N(\mu_2,\Sigma)$。每个个体有p种观测指标，如果进行了n次观测得到的观测矩阵X_β满足：$X_\beta=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，这里$x_i=(x_{i1},x_{i2},\cdots,x_{ip})^{\mathrm{T}}$，$i=1,2,\cdots,n$。其中观测矩阵的前$n_1$行观测值来自第一个总体$G_1$，第$n_1+1$行到第$n$行观测值来自第二个总体$G_2$，且$n=n_1+n_2$。如今给定一个新的样品，判别分析的目的是根据观测矩阵X_β判定此新样品属于两类中的哪一类。

对上述判别分类问题，已有了一些经典的方法，如：距离判别，Bayes 判别等。经典马氏距离判别的思想是：新样品距离哪个总体近就判给哪个总体。而 Bayes 判别的原理是考虑错判损失，依据使总平均损失最小来进行分类判别。其判别准则如下：

$$\text{当}\left(x-\frac{1}{2}(\mu_1+\mu_2)\right)^{\mathrm{T}}\Sigma^{-1}(\mu_1-\mu_2)\geq 2\ln\frac{C(1|2)\pi_2}{C(2|1)\pi_1}\text{时，}\ x\in G_1\text{；否则，}\ x\in G_2\text{。}$$

其中π_1、π_2为总体G_1、G_2的先验概率；$C(i|j)$即把本属于总体G_j的样品错判给G_i时造成的损失。然而通常情况下，总体参数u_1，u_2，Σ未知，需要由样本数据来估计未知参数。这里我们记：

$$\hat{\mu}_1=\frac{1}{n_1}\sum_{i=1}^{n_1}x_i,\ \hat{\mu}_2=\frac{1}{n_2}\sum_{i=n_1+1}^{n}x_i,\ \hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i,\ \hat{\Sigma}=\frac{1}{n-2}\left(\sum_{i=1}^{n_1}(x_i-\hat{\mu}_1)(x_i-\hat{\mu}_1)^{\mathrm{T}}+\sum_{i=n_1+1}^{n}(x_i-\hat{\mu}_2)(x_i-\hat{\mu}_2)^{\mathrm{T}}\right)$$

将求得的参数估计值代入判别函数中，可用相应的判别准则对新样品判别归类。当两正态总体协方差阵相等时，我们可根据距离判别和 Bayes 判别准则导出两个线性判别函数。由判别函数的线性特性及判别函数中指标的多元性，我们考虑多元线性回归模型：

$$Y=X\beta+\varepsilon=\beta_0 1_n+X_\beta b+\varepsilon \tag{1.1}$$

其中$Y=(y_1,y_2,\cdots,y_{n_1},y_{n_1+1}\cdots,y_n)^{\mathrm{T}}$。这里我们按如下规则对响应变量$y$赋值：

$$y=\begin{cases}\xi & \text{当}x\in G_1\text{时}\\ \eta & \text{当}x\in G_2\text{时}\end{cases} \tag{1.2}$$

我们不妨令：$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$，$X=\left(1_n,X_\beta\right)$，$1_n$是一个$n\times1$的列向量，其元素全为1。

$$\beta=\left(\beta_0,\beta_1,\cdots,\beta_p\right)^{\mathrm{T}}=\begin{pmatrix}\beta_0\\ b\end{pmatrix},\quad \varepsilon\sim N\left(0,\sigma^2I_n\right)$$

张尧庭等[1]通过讨论回归分析与判别分析的关系指出：可以用回归分析方法来处理判别分析问题，并指出Fishe线性判别函数与线性回归方程(除常数项α以外)在形式上是一样的；Trevor等[2]提出了分类的线性方法，并认为最小二乘回归系数与LDA (linear discriminant analysis)中的判别系数成比例；Qing和Hui [3]由LDA中的最小二乘公式推导并提出Lassoed判别分析。Jianqing Fan [4]在A ROAD to Classification in High Dimensional Space中提出了Regularized Optimal Affine Discriminant的判别方法。以上方法都是用线性回归的方法做判别，这涉及到对响应变量y的值的选取问题。张尧庭等[1]在多元统计分析引论中提出：为了使各类响应值的均值0，不妨令下(1.2)式中的$\xi=-n_1/n$，$\eta=n_2/n$。Trevor等[2]在文章中令ξ,η分别为−1，1；而Qing和Hui [3]则令ξ,η分别为$-n/n_1$，n/n_2。我们将在前人研究的基础上，进一步讨论回归分析中不同响应值的选取对判别结果的影响。

2. 用线性回归方法做判别

这里我们将用回归分析的方法来处理判别问题。首先我们对观测数据做中心化处理，如邰淑彩等[5]中的：$\tilde{X}=X_\beta-1_n\hat{\mu}^{\mathrm{T}},\tilde{Y}=Y-\bar{y}1_n$，得到(1.1)式的一个新的矩阵表达形式：

$$\tilde{Y}=X_\alpha\tilde{\beta}+\varepsilon=\alpha1_n+\tilde{X}b+\varepsilon \tag{2.1}$$

其中$\tilde{\beta}=\begin{pmatrix}\alpha\\ b\end{pmatrix}$，$X_\alpha=\left(1_n,\tilde{X}\right)$。由多元线性回归模型中系数的最小二乘估计法知：

$$\hat{\alpha}^{ols}=0,\ \hat{b}^{ols}=\left(\tilde{X}^{\mathrm{T}}\tilde{X}\right)^{-1}\tilde{X}^{\mathrm{T}}\tilde{Y},\ \hat{\beta}_0^{ols}=\bar{y}-\hat{\mu}^{\mathrm{T}}\hat{b}^{ols} \tag{2.2}$$

其中$\hat{\alpha}^{ols},\hat{b}^{ols},\hat{\beta}_0^{ols}$分别是$\alpha,b,\beta_0$的最小二乘估计。

Theorem 2.1：在多元线性回归方程(2.1)中，参数b的最小二乘估计$\hat{b}^{ols}$满足式子：

$$\left[(n-2)\hat{\Sigma}+\frac{n_1n_2}{n}\hat{\Sigma}_\beta\right]\hat{b}^{ols}=\frac{n_1n_2(\xi-\eta)(\hat{\mu}_1-\hat{\mu}_2)}{n} \tag{2.3}$$

其中$\hat{\Sigma}_\beta=(\hat{\mu}_2-\hat{\mu}_1)(\hat{\mu}_2-\hat{\mu}_1)^{\mathrm{T}}$，特别的，当$\xi=-\frac{n}{n_1}$，$\eta=\frac{n}{n_2}$时，$\hat{b}^{ols}$满足如下式子：

$$\left[(n-2)\hat{\Sigma}+\frac{n_1n_2}{n}\hat{\Sigma}_\beta\right]\hat{b}^{ols}=n(\hat{\mu}_2-\hat{\mu}_1)$$

由(2.2)式和(2.3)式，我们将系数的最小二乘估计代入判别函数$y=\hat{\beta}_0^{ols}+x^{\mathrm{T}}\hat{b}^{ols}$，并设定判别准则如下(2.4)式，我们用此判别函数及准则对新样本判别归类时能得到下面的定理。

$$x\in\begin{cases}G_1 & \text{当}y\le\bar{y}\text{时}\\ G_2 & \text{当}y>\bar{y}\text{时}\end{cases} \tag{2.4}$$

Theorem 2.2：(1) 若$\hat{\Sigma}$正定，则判别结果只与$\xi-\eta$的符号有关，而与ξ,η的取值无关。即只要$\xi-\eta$的符号$\mathrm{sgn}(\xi-\eta)$相同，用该方法判别得到的结果就相同，且无论n_1和n_2相等与否，该结论都成立。

(2) 当$n_1=n_2$且ξ,η满足$\mathrm{sgn}(\xi-\eta)<0$时，用该判别方法与用距离判别法对新样品分类时得到的判别函数及判别结果相同。

3. 模拟

3.1. 平衡数据模拟

我们通过数据模拟来验证我们的结论。我们随机生成了两类数据G_1， G_2。它们均服从p元正态分布，其中$G_1 \sim N(\mu_1, \Sigma)$，$G_2 \sim N(\mu_2, \Sigma)$。这里我们取$\mu_1 = (1,2,3,2,1,1_{p-5})$，$\mu_2 = (1.3,1.5,2,1,2,1_{p-5})$，$n_1 = n_2 = 100$，$\Sigma$满足：$\Sigma_{ii} = 1$，$i, j = 1,2,\cdots,p$

$$\Sigma_{ij} = \begin{cases} 0.5 & \text{当} i \neq j, |i-j| = 1 \text{且} i, j \leq 6 \\ 0 & \text{其他} \end{cases} \tag{3.1}$$

我们采用五折交叉验证的方法，此方法被 Breiman 等[6]提议并广泛应用于实际，取数据集的 4/5 作训练集，剩下的 1/5 作测试集。然后用训练集拟合判别函数，用测试集评估分类性能。我们分别在$p = 10,30,50,70,90,110,130,150$的八种情况下，用距离判别、判别 I-III 这四种方法对测试集样本判别归类。其中，判别 I-III 即在 **Theorem 2.2** 提出的判别方法中分别取：$\xi = -n_1/n$，$\eta = n_2/n$；$\xi = -n/n_1$，$\eta = n/n_2$；$\xi = -1$，$\eta = 1$得到的三种判别方法。我们重复模拟试验 1000 次，最终得到各自的平均错判率如表 1。

显然，在判别 I-III 中：$\text{sgn}(\xi - \eta) < 0$且$\bar{y} = 0$。由表 1 可知：用判别 I-III 三种方法判别时的模拟错判率相等，这与 **Theorem 2.2 (1)**相符。此外，当$n_1 = n_2$时，用距离判别与用判别 I-III 这三种方法判别时的模拟错判率相等，满足 **Theorem 2.2 (2)**。此外，我们还可从表格中看出：用以上四种方法对平衡数据$(n_1 = n_2)$进行判别时，模拟得到的错判率随着维数p的增加而增加，即判别效果随之降低。

3.2. 不平衡数据模拟

基于上 3.1 中提到的两类数据，我们分别取 p = 10,50,100，并分别在这三种情况下取$n_1 = 100$，$n_2 = 150,300,600$。之后分别用距离判别，Bayes 判别，判别 I-III 这五种方法对前面 9 种情况做判别，我们重复模拟试验 1000 次并取平均值，模拟结果如表 2。

这里我们将用一些特定的评价标准评估不同判别方法的分类性能。当数据不平衡$(n_1 \neq n_2)$时，Weiss [7]指出：为提高分类准确率，分类器往往倾向于将新样品预测为多数类而导致少数类样本的识别率较低，所以错判率不能很好的反映判别方法对不平衡数据集的判别效果。因此我们采用不平衡数据集分类中常用的评价标准：F-value [8]和 G-mean [9]来衡量不同方法判别效果的好坏。这里记：

$$\text{F-value} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}, \ \text{G-mean} = \sqrt{acc^+ * acc^-} \tag{3.2}$$

Table 1. The misclassification rate of four discriminant methods
表 1. 四种判别方法的错判率比较

p	距离判别错判率	判别 I 错判率	判别 II 错判率	判别 III 错判率
10	0.0925	0.0925	0.0925	0.0925
30	0.1153	0.1153	0.1153	0.1153
50	0.1430	0.1430	0.1430	0.1430
70	0.1741	0.1741	0.1741	0.1741
90	0.2095	0.2095	0.2095	0.2095
110	0.2542	0.2542	0.2542	0.2542
130	0.3089	0.3089	0.3089	0.3089
150	0.3910	0.3910	0.3910	0.3910

Table 2. The discriminant outcome comparison of unbalanced data
表 2. 不平衡数据的判别结果比较

(n_2, p)	判别标准	距离判别	Bayes 判别	判别 I	判别 II	判别 III
	错判率	0.0897	0.0938	0.1062	0.1062	0.1062
(150, 10)	F-value	0.8903	0.8738	0.8782	0.8782	0.8782
	G-mean	0.9090	0.8886	0.9004	0.9004	0.9004
	错判率	0.0863	0.0942	0.1919	0.1919	0.1919
(300, 10)	F-value	0.8417	0.7716	0.7212	0.7212	0.7212
	G-mean	0.9117	0.8019	0.8567	0.8567	0.8567
	错判率	0.0840	0.0687	0.3002	0.3002	0.3002
(600, 10)	F-value	0.7586	0.6874	0.4876	0.4876	0.4876
	G-mean	0.9137	0.7328	0.8027	0.8027	0.8027
	错判率	0.1279	0.1292	0.1439	0.1439	0.1439
(150, 50)	F-value	0.8438	0.8259	0.8364	0.8364	0.8364
	G-mean	0.8693	0.8497	0.8627	0.8627	0.8627
	错判率	0.1058	0.1078	0.2124	0.2124	0.2124
(300, 50)	F-value	0.8065	0.7363	0.6972	0.6972	0.6972
	G-mean	0.8871	0.7766	0.8377	0.8377	0.8377
	错判率	0.0921	0.0743	0.3088	0.3088	0.3088
(600, 50)	F-value	0.7346	0.6582	0.4792	0.4792	0.4792
	G-mean	0.8960	0.7118	0.7949	0.7949	0.7949
	错判率	0.1868	0.1833	0.2020	0.2020	0.2020
(150, 100)	F-value	0.7735	0.7559	0.7727	0.7727	0.7727
	G-mean	0.8085	0.7932	0.8039	0.8039	0.8039
	错判率	0.1369	0.1264	0.2406	0.2406	0.2406
(300, 100)	F-value	0.7514	0.6922	0.6642	0.6642	0.6642
	G-mean	0.8482	0.7480	0.8098	0.8098	0.8098
	错判率	0.1047	0.0833	0.3199	0.3199	0.3199
(600, 100)	F-value	0.6985	0.6135	0.4675	0.4675	0.4675
	G-mean	0.8713	0.6815	0.7836	0.7836	0.7836

其中 $acc^+ = \text{recall} = \frac{TP}{TP+FN}$， $acc^- = \frac{TN}{TN+FP}$， $\text{precision} = \frac{TP}{TP+FP}$。$P$：正类(少数类)，$N$：负类(多数类)。$TP$ 与 TN 分别表示被正确分类的正类和负类样本的数目；FN 表示真实类标是正类却被误分为负类的数目，FP 表示真实类标是负类而被误分为正类的数目。acc^+：少数类的查全率，acc^-：多数类的查全率，precision：查准率。陶新民等[10]指出：F-Value 既考虑了查全率又考虑了查准率，只有在查全率和查准率的值都大时，F-value 才会大。同样，只有少数类和多数类样本的查全率同时都大时，G-mean 值才会大。因此，F-value 和 G-mean 能综合考虑少数类和多数类两类样本的分类性能，是对不平衡数据分类性能的两个较好的评测指标。

由表 2 最后三列知：就错判率及不平衡数据的评价标准：F-Value，G-Mean 而言，当数据不平衡 $(n_1 \neq n_2)$ 时，判别 I-III 的判别结果相同且 $\bar{y} \neq 0$，此时用距离判别与用判别 I-III 判别得到的结果不同。

此外，当数据的维数 p 固定不变时，随着数据不平衡程度 n_2/n_1 的增加，距离判别及 Bayes 判别的错判率、F-value 和 G-mean 值变化相对较小，即受不平衡程度的影响较小，判别效果较好；当数据不平衡程度 n_2/n_1 固定不变时，随着维数 p 的增加，距离判别及 Bayes 判别得的错判率较低且 F-value，G-Mean 值较高，判别效果较好；然而我们可以结合线性模型的变量选择方法如：Tibshirani [11]提出的 LASSO (Least Absolute Shrinkage and Selection Operator)、Fan 和 Li [12]提出的 SCAD (Smoothly Clipped Absolute Deviation)、AIC、BIC 及 Fan 和 Lv [13] [14]提出的 SIS (Sure Independence Screening)等方法选择重要变量对数据降维，与此同时，用线性回归的方法对新样品判别归类。用这种方法，我们在降维的同时对数据做判别，可能会使得判别 I-III 的判别效果得到提高，然而这还有待我们以后做进一步研究。

3.3. 实例分析

此外，我们对“Wisconsin Diagnostic Breast Cancer (WDBC)”中的真实数据进行了分析。本文采用的数据来源于 http://www.datatang.com/data/515。

该数据集包含了关于人体细胞核的 30 个相关指标(如：细胞核面积、周长、平滑度等)的大量数据，我们从第一类(未患乳腺癌)和第二类(患乳腺癌)的观测样本中各取 n_1、n_2 个样品，用五折交叉验证的方法，并分别用距离判别、Bayes 判别、判别 I-III 这五种方法对这 $(n_1+n_2)/5$ 个样品“是否患有乳腺癌”进行分析判别，并得到各自的模拟结果如表 3。

Table 3. WDBC discriminant result comparison
表 3. WDBC 判别结果比较

(n_1,n_2)	判别标准	距离判别	Bayes 判别	判别 I	判别 II	判别 III
(200, 200)	错判率	0.0375	0.0375	0.0375	0.0375	0.0375
	F-value	0.9632	0.9632	0.9632	0.9632	0.9632
	G-mean	0.9623	0.9623	0.9623	0.9623	0.9623
(200, 100)	错判率	0.0467	0.0867	0.0700	0.0700	0.0700
	F-value	0.9651	0.9389	0.9445	0.9445	0.9445
	G-mean	0.9417	0.8630	0.9412	0.9412	0.9412
(200, 50)	错判率	0.0320	0.0720	0.1120	0.1120	0.1120
	F-value	0.9805	0.9577	0.9243	0.9243	0.9243
	G-mean	0.9216	0.7803	0.9198	0.9198	0.9198
(200, 40)	错判率	0.0292	0.0583	0.1417	0.1417	0.1417
	F-value	0.9831	0.9670	0.9069	0.9069	0.9069
	G-mean	0.9121	0.7734	0.9010	0.9010	0.9010
(200, 20)	错判率	0.0318	0.0409	0.1818	0.1818	0.1818
	F-value	0.9825	0.9782	0.8884	0.8884	0.8884
	G-mean	0.8817	0.6811	0.8719	0.8719	0.8719
(200, 10)	错判率	0.0143	0.0238	0.2524	0.2524	0.2524
	F-value	0.9925	0.9877	0.8465	0.8465	0.8465
	G-mean	0.8811	0.6811	0.8130	0.8130	0.8130

由表 3 知：由于判别 I-III 中的 $\mathrm{sgn}(\xi-\eta)$ 相等，用判别 I-III 三种方法模拟的判别结果相等。当 $n_1=n_2$ 时，距离判别与判别 I-III 的判别结果相同；当 $n_1\neq n_2$ 时，它们的判别结果不同。而且，随着不平衡程度 n_1/n_2 的增加，距离判别及 Bayes 判别的错判率较低，F-value 值较高，受不平衡程度的影响较小，判别效果较好。

4. 总结

本文主要研究了线性回归模型中响应值的选取对二分类问题的影响。首先，我们对响应变量按一定的规则赋值，然后用最小二乘法拟合，建立判别函数及判别准则，进而得到以下两个结论：1) 该判别方法下的判别结果只与 $\mathrm{sgn}(\xi-\eta)$ 有关，即只要 $\mathrm{sgn}(\xi-\eta)$ 相同，用此判别方法判别的结果就相同。2) 当 $n_1=n_2$ 且 $\mathrm{sgn}(\xi-\eta)<0$ 时，用该判别方法得的判别结果与距离判别结果相同。此外，我们用 r 语言[15]分别对平衡数据、不平衡数据及真实数据 WDBC 进行了模拟，得到了与前面两个结论相符的模拟结果。

基金项目

中央高校基本科研业务费专项资金；北京高等学校青年英才计划项目。

参考文献 (References)

[1] 张尧庭, 方开泰 (1988) 多元统计分析引论. 科学出版社, 北京.

[2] Hastie, T., Tibshirani, R. and Friedman, J. (2009) Elements of statistical learning: data mining, inference and prediction. 2nd Edition, Springer, Berlin.

[3] Mai, Q. and Zou, H. (2012) A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika*, **99**, 29-42.

[4] Fan, J.Q., Feng, Y. and Tong, X. (2012) A road to classification in high dimensional space. *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, **74**, 745-771.

[5] 邰淑彩, 孙韫玉, 何娟娟 (2005) 应用数理统计(第二版). 武汉大学出版社, 武汉.

[6] Breiman, L. and Spector, P. (1992) Submodel selection and evaluation in regression: the x-random case. *International Statistical Review*, **60**, 291-319.

[7] Weiss, G.M. and Provost, F. (2003) Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research*, **19**, 315-354.

[8] Kubat, M., Holte, R. and Matwin, S. (1998) Machine learning for the detection of oil spills in satellite radar images. *Machine Learning*, **30**, 195-215.

[9] Lewis, D. and Gale, W. (1994) Training text classifiers by uncertainty sampling. *Proceedings of ACM-SIGIR Conference on Information Retrieval*, New York, 73-79.

[10] 陶新民, 郝思媛, 张冬雪, 徐鹏 (2013) 不均衡数据分类算法的综述. *重庆邮电大学学报(自然科学版)*, **1**, 106-108.

[11] Tibshirani, R.J. (1996) Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B*, **58**, 267-288.

[12] Fan, J. and Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348-1360.

[13] Fan, J. and Lv, J. (2008) Sure independence screening for ultrahigh dimensional feature space (with discussion). *Journal of the Royal Statistical Society, Series B*, **70**, 849-911.

[14] Fan, J. and Lv, J. (2010) A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, **20**, 101-148.

[15] 薛毅, 陈立萍 (2007) 统计建模与 R 软件. 清华大学出版社, 北京.

附录

Theorem 2.1 的证明：由上(2.2)式知：$\hat{b}^{ols}$ 满足式子 $\left(\tilde{X}^{\mathrm{T}}\tilde{X}\right)\hat{b}^{ols}=\tilde{X}^{\mathrm{T}}\tilde{Y}$，即满足

$$\left[\left(X_{\beta}-1_n\hat{\mu}^{\mathrm{T}}\right)^{\mathrm{T}}\left(X_{\beta}-1_n\hat{\mu}^{\mathrm{T}}\right)\right]\hat{b}^{ols}=\left(X_{\beta}-1_n\hat{\mu}^{\mathrm{T}}\right)^{\mathrm{T}}\tilde{Y} \tag{A.1}$$

(A.1)式左边

$$=\sum_{i=1}^{n}\left(x_i-\hat{\mu}\right)\left(x_i-\hat{\mu}\right)^{\mathrm{T}}\hat{b}^{ols}=\sum_{i=1}^{n}\left(x_i-\frac{\left(n_1\hat{\mu}_1+n_2\hat{\mu}_2\right)}{n}\right)\left(x_i-\frac{\left(n_1\hat{\mu}_1+n_2\hat{\mu}_2\right)}{n}\right)^{\mathrm{T}}\hat{b}^{ols}$$

$$=\left\{\sum_{i=1}^{n_1}\left(x_i-\frac{\left(n_1\hat{\mu}_1+n_2\hat{\mu}_2\right)}{n}\right)\left(x_i-\frac{\left(n_1\hat{\mu}_1+n_2\hat{\mu}_2\right)}{n}\right)^{\mathrm{T}}+\sum_{i=n_1+1}^{n}\left(x_i-\frac{\left(n_1\hat{\mu}_1+n_2\hat{\mu}_2\right)}{n}\right)\left(x_i-\frac{\left(n_1\hat{\mu}_1+n_2\hat{\mu}_2\right)}{n}\right)^{\mathrm{T}}\right\}\hat{b}^{ols}$$

$$=\left[\sum_{i=1}^{n_1}\left(x_i-\hat{\mu}_1+\frac{n_2\left(\hat{\mu}_1-\hat{\mu}_2\right)}{n}\right)\left(x_i-\hat{\mu}_1+\frac{n_2\left(\hat{\mu}_1-\hat{\mu}_2\right)}{n}\right)^{\mathrm{T}}+\sum_{i=n_1+1}^{n}\left(x_i-\hat{\mu}_2+\frac{n_1\left(\hat{\mu}_2-\hat{\mu}_1\right)}{n}\right)\left(x_i-\hat{\mu}_2+\frac{n_1\left(\hat{\mu}_2-\hat{\mu}_1\right)}{n}\right)^{\mathrm{T}}\right]\hat{b}^{ols}$$

$$=\left[\sum_{i=1}^{n_1}\left(x_i-\hat{\mu}_1\right)\left(x_i-\hat{\mu}_1\right)^{\mathrm{T}}+\sum_{i=n_1+1}^{n}\left(x_i-\hat{\mu}_2\right)\left(x_i-\hat{\mu}_2\right)^{\mathrm{T}}+\sum_{i=1}^{n_1}\frac{n_2^{\,2}\left(\hat{\mu}_1-\hat{\mu}_2\right)\left(\hat{\mu}_1-\hat{\mu}_2\right)^{\mathrm{T}}}{n^2}+\sum_{i=n_1+1}^{n}\frac{n_1^{\,2}\left(\hat{\mu}_1-\hat{\mu}_2\right)\left(\hat{\mu}_1-\hat{\mu}_2\right)^{\mathrm{T}}}{n^2}\right]\hat{b}^{ols}$$

$$=\left[\sum_{i=1}^{n_1}\left(x_i-\hat{\mu}_1\right)\left(x_i-\hat{\mu}_1\right)^{\mathrm{T}}+\sum_{i=n_1+1}^{n}\left(x_i-\hat{\mu}_2\right)\left(x_i-\hat{\mu}_2\right)^{\mathrm{T}}+\frac{n_1n_2\left(\hat{\mu}_1-\hat{\mu}_2\right)\left(\hat{\mu}_1-\hat{\mu}_2\right)^{\mathrm{T}}}{n}\right]\hat{b}^{ols}$$

$$=\left[\left(n-2\right)\hat{\Sigma}+\frac{n_1n_2}{n}\hat{\Sigma}_{\beta}\right]\hat{b}^{ols}$$

(A.1)式右边

$$=\sum_{i=1}^{n}\left(x_i-\hat{\mu}\right)\left(y_i-\bar{y}\right)=\sum_{i=1}^{n_1}\left(x_i-\hat{\mu}\right)\left(y_i-\bar{y}\right)+\sum_{i=n_1+1}^{n}\left(x_i-\hat{\mu}\right)\left(y_i-\bar{y}\right)$$

$$=\sum_{i=1}^{n_1}\left(x_i-\frac{\left(n_1\hat{\mu}_1+n_2\hat{\mu}_2\right)}{n}\right)\left(\xi-\frac{\left(n_1\xi+n_2\eta\right)}{n}\right)+\sum_{i=n_1+1}^{n}\left(x_i-\frac{\left(n_1\hat{\mu}_1+n_2\hat{\mu}_2\right)}{n}\right)\left(\eta-\frac{\left(n_1\xi+n_2\eta\right)}{n}\right)$$

$$=\sum_{i=1}^{n_1}\left(x_i-\hat{\mu}_1+\frac{n_2\left(\hat{\mu}_1-\hat{\mu}_2\right)}{n}\right)\frac{n_2\left(\xi-\eta\right)}{n}-\sum_{i=n_1+1}^{n}\left(x_i-\hat{\mu}_2+\frac{n_1\left(\hat{\mu}_2-\hat{\mu}_1\right)}{n}\right)\frac{n_1\left(\xi-\eta\right)}{n}$$

$$=\frac{n_1n_2^2\left(\xi-\eta\right)\left(\hat{\mu}_1-\hat{\mu}_2\right)}{n^2}+\frac{n_2n_1^2\left(\xi-\eta\right)\left(\hat{\mu}_1-\hat{\mu}_2\right)}{n^2}=\frac{n_1n_2\left(\xi-\eta\right)\left(\hat{\mu}_1-\hat{\mu}_2\right)}{n}$$

由(A.1)知：左边 = 右边，所以 $\left[\left(n-2\right)\hat{\Sigma}+\frac{n_1n_2}{n}\hat{\Sigma}_{\beta}\right]\hat{b}^{ols}=\frac{n_1n_2\left(\xi-\eta\right)\left(\hat{\mu}_1-\hat{\mu}_2\right)}{n}$。当 $\xi=-n/n_1$，$\eta=n/n_2$，$\hat{b}^{ols}$ 满足式子 $\left[\left(n-2\right)\hat{\Sigma}+\frac{n_1n_2}{n}\hat{\Sigma}_{\beta}\right]\hat{b}^{ols}=n\left(\hat{\mu}_2-\hat{\mu}_1\right)$。

Theorem 2.2 (1)的证明：由(2.3)式知：

$$\hat{\Sigma}\hat{b}^{ols}=\left[-\frac{n_1n_2}{n\left(n-2\right)}\left(\hat{\mu}_1-\hat{\mu}_2\right)^{\mathrm{T}}\hat{b}^{ols}+\frac{n_1n_2\left(\xi-\eta\right)}{n\left(n-2\right)}\right]\left(\hat{\mu}_1-\hat{\mu}_2\right) \tag{A.2}$$

此处不妨令

$$\lambda=-\frac{n_1n_2}{n\left(n-2\right)}\left(\hat{\mu}_1-\hat{\mu}_2\right)^{\mathrm{T}}\hat{b}^{ols}+\frac{n_1n_2\left(\xi-\eta\right)}{n\left(n-2\right)}=\frac{n_1n_2}{n\left(n-2\right)}\left[\xi-\eta-\left(\hat{\mu}_1-\hat{\mu}_2\right)^{\mathrm{T}}\hat{b}^{ols}\right] \tag{A.3}$$

则 $\hat{\Sigma}\hat{b}^{ols}=\lambda(\hat{\mu}_1-\hat{\mu}_2)$。当 $\hat{\Sigma}$ 正定时：$\hat{b}^{ols}=\lambda\hat{\Sigma}^{-1}(\hat{\mu}_1-\hat{\mu}_2)$。代入(A.3)式得：

$$\lambda=\frac{\xi-\eta}{\frac{n(n-2)}{n_1n_2}+(\hat{\mu}_1-\hat{\mu}_2)^{\mathrm{T}}\hat{\Sigma}^{-1}(\hat{\mu}_1-\hat{\mu}_2)} \tag{A.4}$$

又因 $\hat{\Sigma}^{-1}$ 正定，所以 $(\hat{\mu}_1-\hat{\mu}_2)^{\mathrm{T}}\hat{\Sigma}^{-1}(\hat{\mu}_1-\hat{\mu}_2)>0$，因此 $\mathrm{sgn}(\lambda)=\mathrm{sgn}(\xi-\eta)$。而且由 $\hat{\beta}_0^{ols}=\bar{y}-\hat{\mu}^{\mathrm{T}}\hat{b}^{ols}$，$\hat{b}^{ols}=\lambda\hat{\Sigma}^{-1}(\hat{\mu}_1-\hat{\mu}_2)$ 得：

$$y=\hat{\beta}_0^{ols}+x^{\mathrm{T}}\hat{b}^{ols}=\bar{y}+\lambda(x-\hat{\mu})^{\mathrm{T}}\hat{\Sigma}^{-1}(\hat{\mu}_1-\hat{\mu}_2) \tag{A.5}$$

这里令 $w_1(x)=\lambda(x-\hat{\mu})^{\mathrm{T}}\hat{\Sigma}^{-1}(\hat{\mu}_1-\hat{\mu}_2)$，则(2.4)式等价于：

$$x\in\begin{cases}G_1 & \text{当}w_1(x)\le 0\text{时}\\ G_2 & \text{当}w_1(x)>0\text{时}\end{cases} \tag{A.6}$$

因此，无论 ξ 和 η 取何值，只要 $\mathrm{sgn}(\xi-\eta)$ 相同，则判别函数相同，用上述方法得到的判别结果相同。**Theorem 2.2 (1)**得证。

Theorem 2.2 (2)的证明：距离判别的判别函数为：

$$w_2(x)=\left(x-\frac{(\hat{\mu}_1+\hat{\mu}_2)}{2}\right)^{\mathrm{T}}\hat{\Sigma}^{-1}(\hat{\mu}_1-\hat{\mu}_2) \tag{A.7}$$

距离判别的判别准则：

$$x\in\begin{cases}G_1 & \text{当}w_2(x)\ge 0\text{时}\\ G_2 & \text{当}w_2(x)<0\text{时}\end{cases} \tag{A.8}$$

当 $n_1=n_2$ 时，$\hat{\mu}=\frac{(n_1\hat{\mu}_1+n_2\hat{\mu}_2)}{n}=\frac{\hat{\mu}_1+\hat{\mu}_2}{2}$，即 $w_1(x)=\lambda w_2(x)$，且我们前面已证 $\mathrm{sgn}(\lambda)=\mathrm{sgn}(\xi-\eta)<0$，所以(A.6)式等价于(A.8)式。因此，当 $n_1=n_2$ 时，用距离判别与用线性回归做判别得到的判别结果相同。

当 $n_1\neq n_2$ 时，$\hat{\mu}=\frac{(n_1\hat{\mu}_1+n_2\hat{\mu}_2)}{n}\neq\frac{\hat{\mu}_1+\hat{\mu}_2}{2}$，此时 $w_1(x)\neq\lambda w_2(x)$，用距离判别与用线性回归做判别得到的判别结果不同。**Theorem 2.2 (2)**得证。