



多元统计与矩阵分析

张锋 8125345@qq.com

中国地质大学, 计算机学院, 武汉



均值向量与协方差阵的检验

内容



- (1) 均值向量的检验
- (2) 协方差阵的检验
- (3) 应用案例

内容



- (1) 均值向量的检验
- (2) 协方差阵的检验
- (3) 应用案例



单变量假设检验

- (1)根据问题提出待检验的统计假设 H_0 和 H_1
- (2)选取一个合适的统计量并得出它的抽样分布
- (3)给定显著性水平，通过统计量的抽样分布确定临界值，进而得到拒绝域，建立判别准则
- (4)根据样本观测值计算统计量的值，看是否落在拒绝域中，从而对假设检验做出统计判断，并给出具体的解释



单变量检验

设从总体 $N(\mu, \sigma^2)$ 中抽取一个样本 x_1, x_2, \dots, x_n , 我们要检验假设

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

当 σ^2 已知时, 用统计量 $z = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}$

式中, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ 为样本均值。

当假设成立时, 统计量 z 遵从正态分布, $z \sim N(0,1)$, 从而拒绝域为 $|z| > z_{\alpha/2}$, $z_{\alpha/2}$ 为 $N(0,1)$ 的上 $\alpha/2$ 分位点。



单变量检验

设从总体 $N(\mu, \sigma^2)$ 中抽取一个样本 x_1, x_2, \dots, x_n , 我们要检验假设

$$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$$

当 σ^2 未知时, 用

$$S^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{(n-1)}$$

作为 σ^2 的估计, 然后用统计量

$$t = \sqrt{n} \frac{(\bar{x} - \mu_0)}{S}$$

作为检验统计量。如果原假设成立, 则 t 服从自由度为 $n-1$ 的 t 分布, 拒绝域为 $|t| > t_{n-1}(\alpha/2)$, $t_{n-1}(\alpha/2)$ 为 t_{n-1} 的上 $\alpha/2$ 分位数。类似于 σ^2 已知的情形, 我们也可以选用

$$t^2 = n(\bar{x} - \mu_0)'(S^2)^{-1}(\bar{x} - \mu_0)$$

作为检验统计量。当原假设为真时, t^2 服从第一自由度为 1、第二自由度为 $n-1$ 的 F 分布, 简记为 $t^2 \sim F_{1,n-1}$, 拒绝域为 $t^2 > F_{1,n-1}(\alpha)$, $F_{1,n-1}(\alpha)$ 为 $F_{1,n-1}$ 的上 α 分位数。



多元均值检验

协方差阵已知

设总体 $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 随机样本为 $\mathbf{X}_{(\alpha)} (\alpha = 1, \dots, n)$, 我们要检验这组随机样本的均值是否 否为某一个指定的向量 $\boldsymbol{\mu}_0$, 即

$$H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$$

$$H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

$$\chi_0^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0)$$

我们可以证明当原假设为真时, χ_0^2 服从自由度为 p 的 χ^2 分布。直观上来看 χ_0^2 越大意味着样本均值 $\bar{\mathbf{X}}$ 与 $\boldsymbol{\mu}_0$ 的差距越大, 因此拒绝域应取为 χ_0^2 较大的部分。

拒绝域为 $\{\chi_0^2 > \chi_p^2(\alpha)\}$



在许多统计学软件中，主要是利用P值来给出假设检验的结果。在原假设为真时， $\lambda_0^2 \sim \chi^2(p)$ ，而样本值可以计算出 λ_0^2 的值，记为d,称概率值p为显著性概率值：

$$p = P(\lambda_0^2 \geq d)$$

统计学根据显著性检验方法所得到的p值，一般以 $P < 0.05$ 为有统计学差异， $P < 0.01$ 为有显著统计学差异， $P < 0.001$ 为有极其显著的统计学差异。其含义是样本间的差异由抽样误差所致的概率小于0.05 、0.01、0.001，在统计结果中显示为 $\text{Pr} > F$ 。



多元均值检验

协方差阵未知

由于协方差阵未知，我们需要先估计协方差阵。因为 $\hat{\Sigma} = \frac{1}{n-1}A$ 是 Σ 的无偏估计，所以考虑

$$\begin{aligned} T^2 &= n(\bar{X} - \mu_0)' \hat{\Sigma}^{-1} (\bar{X} - \mu_0) \\ &= n(n-1)(\bar{X} - \mu_0)' A^{-1} (\bar{X} - \mu_0) \end{aligned}$$

的分布。根据 1.4.2 节 Hotelling T^2 分布的性质1

$$T^2 \sim T^2(p, n-1)$$

再利用 T^2 分布与 F 分布的性质（参看 1.4.2 节），我们将检验统计量取为

$$F = \frac{n-p}{(n-1)p} T^2$$

在原假设下， $F \sim F(p, n-p)$ ，拒绝域为 $\{F > F_{p, n-p}(\alpha)\}$ ， $F_{p, n-p}(\alpha)$ 为分布 $F_{p, n-p}$ 的上 α 分位数。

两正态总体均值向量的检验

样本协方差阵相等（但未知）



设 $\mathbf{X}_{(\alpha)}$ ($\alpha = 1, \dots, n_1$) 为来自 p 元正态总体 $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ 的容量为 n_1 的样本,

$\mathbf{Y}_{(\alpha)}$ ($\alpha = 1, \dots, n_2$) 为来自 p 元正态总体 $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ 的容量为 n_2 的样本, 两样本相互独立, $n_1 > p, n_2 > p, \boldsymbol{\Sigma}$ 未知。需要进行的假设检验是

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$$

$$H_1: \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

与单总体均值检验类似, 我们采用的检验统计量形式为:

$$T^2 = \frac{1}{\frac{1}{n_1} + \frac{1}{n_2}} (\bar{\mathbf{X}} - \bar{\mathbf{Y}})' \hat{\boldsymbol{\Sigma}}^{-1} (\bar{\mathbf{X}} - \bar{\mathbf{Y}})$$

其中 $\bar{\mathbf{X}} = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{X}_i, \bar{\mathbf{Y}} = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{Y}_i$ 。协方差阵 $\boldsymbol{\Sigma}$ 的估计采用 $\hat{\boldsymbol{\Sigma}} = \frac{\mathbf{A}_x + \mathbf{A}_y}{n_1 + n_2 - 2}$, 其中 \mathbf{A}_x 与 \mathbf{A}_y 是分别是两个总体的样本离差阵。

当原假设成立时, 利用 1.4.1 节 Wishart 分布的性质 1 与 T^2 统计量的定义可以得到

$$T^2 \sim T_{p, n_1 + n_2 - 2}^2$$

两正态总体均值向量的检验

样本协方差阵相等（但未知）



再利用 T^2 与 F 分布的关系（参看 1.4.2 节）我们可以得到

$$F^* = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2 \sim F_{p, n_1 + n_2 - p - 1}$$

我们取 F^* 作为检验统计量。如果 F^* 的值较大，意味着 T^2 的值较大，也就说明两个总体的距离较远，倾向于拒绝原假设，因此拒绝域取为 F^* 的值较大的区域，即当给定显著性水平 α 后，若 $F^* > F_{p, n_1 + n_2 - p - 1}(\alpha)$ ，则拒绝原假设，否则没有充分的理由拒绝原假设。

两正态总体均值向量的检验

样本协方差阵不等



假设从两个正态总体 $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ 和 $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ 中分别抽取容量为 n_1 和 n_2 的两个样本。当这两个样本协方差不相等时并没有统一的办法处理，下面介绍两种较为简单的情形。如果 $n_1 = n_2$ ，我们可以采取成对数据处理的技巧。令

$$Z_{(i)} = X_{(i)} - Y_{(i)}$$

这样可以将两样本均值检验问题化为单样本均值检验问题，即我们做假设检验：

$$H_0: \boldsymbol{\mu}_Z = \mathbf{0} \quad H_1: \boldsymbol{\mu}_Z \neq \mathbf{0}$$

然后我们对 Z 采用上一节的方法进行假设检验。

如果 $\boldsymbol{\Sigma}_1$ 和 $\boldsymbol{\Sigma}_2$ 相差很大时，我们考虑利用 T^2 统计量的近似分布来构造检验统计量。 T^2 统计量的形式是

$$T^2 = (\bar{X} - \bar{Y})' \left(\frac{\mathbf{A}_x}{n_1(n_1 - 1)} + \frac{\mathbf{A}_y}{n_2(n_2 - 1)} \right)^{-1} (\bar{X} - \bar{Y})$$

其中 $\bar{X}, \bar{Y}, \mathbf{A}_x, \mathbf{A}_y$ 的含义与前文相同。我们记

$$\mathbf{S}_* = \frac{\mathbf{A}_x}{n_1(n_1 - 1)} + \frac{\mathbf{A}_y}{n_2(n_2 - 1)}$$



两正态总体均值向量的检验 样本协方差阵不等

再令

$$f^{-1} = (n_1^3 - n_1^2)^{-1} \left((\bar{X} - \bar{Y})' S_*^{-1} \left(\frac{A_x}{n_1 - 1} \right) S_*^{-1} (\bar{X} - \bar{Y}) \right)^2 T^{-4} \\ + (n_2^3 - n_2^2)^{-1} \left((\bar{X} - \bar{Y})' S_*^{-1} \left(\frac{A_y}{n_2 - 1} \right) S_*^{-1} (\bar{X} - \bar{Y}) \right)^2 T^{-4}$$

可以证明当原假设为真时， $\left(\frac{f-p+1}{fp} \right) T^2$ 近似服从 $F_{p, f-p+1}$ 分布进而可以做假设检验。



多正态总体均值向量的检验

多元方差分析

对于 r 个 p 元正态总体 $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), \dots, N_p(\boldsymbol{\mu}_r, \boldsymbol{\Sigma})$ ，从这 r 个总体中抽取的独立样本为

$$\mathbf{X}_1^{(1)}, \dots, \mathbf{X}_{n_1}^{(1)} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$$

$$\mathbf{X}_1^{(2)}, \dots, \mathbf{X}_{n_2}^{(2)} \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$$

.....

$$\mathbf{X}_1^{(r)}, \dots, \mathbf{X}_{n_r}^{(r)} \sim N(\boldsymbol{\mu}_r, \boldsymbol{\Sigma})$$

总样本数 $n = n_1 + n_2 + \dots + n_r$ 。需要检验的假设是

$$H_0: \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_r \quad H_1: \text{存在 } i \neq j, \text{ 使得 } \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$$

多正态总体均值向量的检验

多元方差分析



为了构造检验统计量，我们先定义以下平方和

$$\text{总偏差平方和 } SST = \sum_{i=1}^r \sum_{j=1}^{n_i} \left(\mathbf{X}_j^{(i)} - \bar{\mathbf{X}} \right)^2$$

$$\text{组内偏差平方和 } SSE = \sum_{i=1}^r \sum_{j=1}^{n_i} \left(\mathbf{X}_j^{(i)} - \bar{\mathbf{X}}_i \right)^2$$

$$\text{组间偏差平方和 } SSA = \sum_{i=1}^r n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^2$$

其中， $\bar{\mathbf{X}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{X}_j^{(i)}$ 是第 i 组的样本均值， $\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} \mathbf{X}_j^{(i)}$ 是总均值，总样本量 $n = n_1 + \cdots + n_r$ 。此时，通过代数运算我们得知如下平方和分解公式成立：

$$SST = SSE + SSA$$

从直观上考察，如果原假设成立，在总偏差平方和 SST 不变的条件下，组间偏差平方和相对于组内偏差平方和应该偏小，因此检验统计量取为

$$F = \frac{SSA/(r-1)}{SSE/(n-r)}$$

拒绝域为 $\{F > F_\alpha\}$ ，其中的 F_α 通过 $P(F > F_\alpha) = \alpha$ 确定。

多正态总体均值向量的检验

多元方差分析



我们将上述方法推广到 r 个 p 元正态总体 $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), \dots, N_p(\boldsymbol{\mu}_r, \boldsymbol{\Sigma})$, 从这 r 个总体中抽取的独立样本为

$$\mathbf{X}_1^{(1)}, \dots, \mathbf{X}_{n_1}^{(1)} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$$

$$\mathbf{X}_1^{(2)}, \dots, \mathbf{X}_{n_2}^{(2)} \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$$

.....

$$\mathbf{X}_1^{(r)}, \dots, \mathbf{X}_{n_r}^{(r)} \sim N(\boldsymbol{\mu}_r, \boldsymbol{\Sigma})$$

总样本数 $n = n_1 + n_2 + \dots + n_r$ 。需要检验的假设是 $H_0: \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_r$ H_1 : 存在 $i \neq j$, 使得 $\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ 前文所叙述的 3 个平方和现在成为了矩阵的形式:

$$\text{总离差阵 } \mathbf{T} = \sum_{i=1}^r \sum_{j=1}^{n_i} \left(\mathbf{X}_j^{(i)} - \bar{\mathbf{X}} \right)^2$$

$$\text{组内离差阵 } \mathbf{A} = \sum_{i=1}^r \sum_{j=1}^{n_i} \left(\mathbf{X}_j^{(i)} - \bar{\mathbf{X}}_i \right)^2$$

$$\text{组间离差阵 } \mathbf{B} = \sum_{i=1}^r n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^2$$

这三者之间仍然有 $\mathbf{T} = \mathbf{A} + \mathbf{B}$ 成立。

多正态总体均值向量的检验

多元方差分析



由于 $\mathbf{T}, \mathbf{A}, \mathbf{B}$ 三者都是矩阵，我们采用1.4.3节所用的广义方差来度量矩阵大小。类似一元情形，我们取检验统计量为

$$\Lambda = \frac{|\mathbf{A}|}{|\mathbf{A} + \mathbf{B}|}$$

可以证明在原假设成立的情况下， Λ 的分布就是 Wilks Λ 分布，即

$$\Lambda \sim \Lambda(p, n - r, r - 1)$$

注意： 此处，分母是组内离差阵的行列式，因此拒绝域为 $\{\Lambda < \lambda_0\}$ ，其中， λ_α 通过 $P(\Lambda < \lambda_\alpha) = \alpha$ 确定。

内容



- (1) 均值向量的检验
- (2) 协方差阵的检验
- (3) 应用案例



单正态总体协方差阵检验

假设 $\mathbf{X}_1, \dots, \mathbf{X}_n$ 是来自 p 元正态总体 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的一个样本, $\boldsymbol{\Sigma}_0$ 是已知的给定矩阵, 且 $\boldsymbol{\Sigma}_0 > 0$ 。考虑假设检验问题

$$\begin{aligned} H_0: \boldsymbol{\Sigma} &= \boldsymbol{\Sigma}_0 \\ H_1: \boldsymbol{\Sigma} &\neq \boldsymbol{\Sigma}_0 \end{aligned}$$

我们用的检验统计量是

$$\lambda = \exp\left(\operatorname{tr}\left(-\frac{1}{2}\mathbf{A}\boldsymbol{\Sigma}_0^{-1}\right)\right) |\mathbf{A}\boldsymbol{\Sigma}_0^{-1}| \left(\frac{e}{n}\right)^{np/2}$$

这个检验统计量的抽样分布很难得到, 通常我们采用 λ 的相关近似分布来得到拒绝域。当样本容量 n 很大时, 如果原假设成立, 那么 $-2\ln \lambda$ 的极限分布是 $\chi^2\left(\frac{p(p+1)}{2}\right)$ 。同时, 如果给定检验水平 α , 当样本容量 n 很大时, 我们可以

由样本值计算出 λ 的值。当 $-2\ln \lambda > \chi_\alpha^2\left(\frac{p(p+1)}{2}\right)$, 即 $\lambda < \exp\left(-\frac{\chi_\alpha^2}{2}\right)$ 时, 拒绝 H_0 。



多总体协方差阵的检验

假定 r 个 p 元正态总体 $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \dots, N_p(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$, 从这 r 个总体中抽取的独立样本为

$$\mathbf{X}_1^{(1)}, \dots, \mathbf{X}_{n_1}^{(1)} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$$

$$\mathbf{X}_1^{(2)}, \dots, \mathbf{X}_{n_2}^{(2)} \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$$

$$\mathbf{X}_1^{(r)}, \dots, \mathbf{X}_{n_r}^{(r)} \sim N(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$$

总样本数 $n = n_1 + n_2 + \dots + n_r$ 。需要检验的假设是

$$H_0: \boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_r$$

H_1 : 存在 $i \neq j$, 使得 $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j$



多总体协方差阵的检验

检验统计量是

$$M = (n - r) \ln \left| \frac{\mathbf{A}}{n - r} \right| - \sum_{t=1}^r (n_t - 1) \ln \left| \frac{\mathbf{A}_t}{n_t - 1} \right|$$

其中 $\mathbf{A}_t = \sum_{i=1}^{n_t} (\mathbf{X}_i^{(t)} - \bar{\mathbf{X}}_t)(\mathbf{X}_i^{(t)} - \bar{\mathbf{X}}_t)'$ 为第 t 个总体的组内样本离差阵, $\bar{\mathbf{X}}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} \mathbf{X}_i^{(t)}$, $t = 1, 2, \dots, r$ 为第 t 个总体的样本均值, $\mathbf{A} = \sum_{t=1}^r \mathbf{A}_t$ 为 r 个总体的样本离差阵的和。当样本容量 n 很大时, 如果原假设 H_0 为真, M 的近似分布为 $(1 - d)M \sim \chi^2(f)$ 。其中 $f = \frac{1}{2}p(p + 1)(r - 1)$

$$d = \begin{cases} \frac{2p^2 + 3p - 1}{6(p + 1)(r - 1)} \left[\sum_{i=1}^r \frac{1}{n_i - 1} - \frac{1}{n - r} \right], & \text{当 } n_i \text{ 不全相等} \\ \frac{(2p^2 + 3p - 1)(r - 1)}{6(p + 1)(n - r)}, & \text{当 } n_i \text{ 全部相等} \end{cases}$$

多个正态总体的均值向量和协方差阵同时检验



假定有 r 个 p 元正态总体 $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \dots, N_p(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$, $\mathbf{X}_i^{(t)}$ 为来自第 t 个总体的随机样本 ($t = 1, \dots, r; i = 1, \dots, n_t$)。检验问题为

$$H_0: \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = \dots = \boldsymbol{\mu}^{(r)}, \text{ 且 } \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \dots = \boldsymbol{\Sigma}_r$$

$$H_1: \boldsymbol{\mu}^{(i)} (i = 1, \dots, r) \text{ 与 } \boldsymbol{\Sigma}_i (i = 1, \dots, r) \text{ 至少有一组不全相等}$$

我们所采用的统计量与上一节略有不同, 形式为

$$\mathbf{M}^* = (n - r) \ln \left| \frac{\mathbf{T}}{(n - r)^p} \right| - \sum_{t=1}^r (n_t - 1) \ln \left| \frac{\mathbf{A}_t}{(n_t - 1)^p} \right|$$

\mathbf{A}_t 与上一节相同, 仍为第 t 个总体的样本离差阵, $\mathbf{T} = \sum_{i=1}^r \sum_{j=1}^{n_i} (\mathbf{X}_j^{(i)} - \bar{\mathbf{X}})^2$ 为总离差阵 (看 2.1.4 节)。当样本容量 n 很大时, 在原假设成立时有这样的近似分布:

$$(1 - b) \mathbf{M}^* \sim \chi^2(f)$$

其中 $f = \frac{1}{2} p(p + 3)(k - 1)$,

$$b = \left(\sum_{i=1}^r \frac{1}{n_i - 1} - \frac{1}{n - r} \right) \left(\frac{2p^2 + 3p - 1}{6(p + 3)(r - 1)} \right) - \frac{p - r + 2}{(n - r)(p + 3)}$$

内容



- (1) 均值向量的检验
- (2) 协方差阵的检验
- (3) 应用案例



应用案例

全面评价上市公司的经营状况有很多方法，本例中采用《国有资本金效绩评价规则》中竞争性工商企业的评价指标体系，即对上市公司考察其八大基本指标：净资产收益率、总资产报酬率、总资产周转率、流动资产周转率、资产负债率、已获利息倍数、销售增长率和资本积累率。表2.1的数据来自三个行业35家上市公司2018年年报数据，均以合并会计报表为依据计算得到。净资产收益率与资产负债率直接取自会计年报，其余各指标计算公式如下：

$$\text{总资产报酬率} = \frac{\text{利润总额} + \text{财务费用}}{(\text{年初总资产} + \text{年末总资产}) / 2} \times 100\%$$

$$\text{总资产周转率} = \frac{\text{主营业务收入}}{(\text{年初总资产} + \text{年末总资产}) / 2}$$

$$\text{流动资产周转率} = \frac{\text{主营业务收入}}{(\text{年初流动资产} + \text{年末流动资产}) / 2}$$

$$\text{已获利息倍数} = \frac{\text{利润总额} + \text{财务费用}}{\text{财务费用}}$$

$$\text{销售增长率} = \frac{\text{本年主营业务收入} - \text{上年主营业务收入}}{\text{上年主营业务收入}} \times 100\%$$

$$\text{资本积累率} = \frac{\text{年末所有者权益} - \text{年初所有者权益}}{\text{年初所有者权益}} \times 100\%$$



如果将不同的行业看做不同的总体，那么以上35家上市公司的数据就可以认为来自3个总体，下面我们尝试对着3个不同行业的上市公司的经营状况进行比较。得到结果：



正态性检验结果

		Shapiro-Wilk	
		统计量	P值
净资产收益率	第一行业	0.931809	0.3598
	第二行业	0.927556	0.3549
	第三行业	0.928107	0.4295
总资产报酬率	第一行业	0.900933	0.1377
	第二行业	0.890505	0.1196
	第三行业	0.812352	0.0205
资产负债率	第一行业	0.963277	0.8033
	第二行业	0.899233	0.1550
	第三行业	0.854927	0.0665
总资产周转率	第一行业	0.600361	<0.0001
	第二行业	0.835126	0.0242
	第三行业	0.942637	0.5827
流动资产周转率	第一行业	0.914545	0.2116
	第二行业	0.907388	0.1975
	第三行业	0.938619	0.5377
已获利息倍数	第一行业	0.689874	0.0004
	第二行业	0.82197	0.0168
	第三行业	0.867204	0.0927
销售增长率	第一行业	0.961605	0.7782
	第二行业	0.966308	0.8686
	第三行业	0.87014	0.1003
资本积累率	第一行业	0.906605	0.1652
	第二行业	0.465201	<0.0001
	第三行业	0.891881	0.1780

方差分析表



源	因变量	平方和	均方	自由度	<i>F</i> 统计量	<i>P</i> 值
模型	净资产收益率	724.885695	362.442847	2	6.32	0.0049
	资产负债率	3862.36971	1931.18485	2	8.52	0.0011
	流动资产周转率	4.00121692	2.00060846	2	4.42	0.0202
	销售增长率	230.400044	115.200022	2	0.54	0.5895
误差	净资产收益率	1834.616150	57.331755	32		
	资产负债率	7257.20396	226.78762	32		
	流动资产周转率	14.48958308	0.45279947	32		
	销售增长率	6861.860271	214.433133	32		
校正合计	净资产收益率	2559.501845		34		
	资产负债率	11119.57367		34		
	流动资产周转率	18.49080000		34		
	销售增长率	7092.260314		34		



两两比较分析

比较		因变量			
		净资产收益率	资产负债率	流动资产周转率	销售增长率
3-2	均值间差值	9.926	7.785	0.6530	3.444
	下限	1.735	-8.218	-0.0749	-12.397
	上限	18.117	23.788	1.3809	19.285
	是否显著	是	否	否	否
3-1	均值间差值	10.204	25.527	-0.0966	6.382
	下限	2.158	9.237	-0.8117	-9.18
	上限	18.25	41.818	0.6185	21.943
	是否显著	是	是	否	否
2-1	均值间差值	0.278	-17.742	-0.7496	2.938
	下限	-7.38	-32.973	-1.4302	-11.872
	上限	7.936	-2.511	-0.0691	17.748
	是否显著	否	是	是	否



Bartlett方差齐性检验

净资产收益率	源	自由度	卡方	Pr > 卡方
	a	2	5.7802	0.0556
资产负债率	源	自由度	卡方	Pr > 卡方
	a	2	0.8541	0.6524
流动资产周转率	源	自由度	卡方	Pr > 卡方
	a	2	15.7135	0.0004
销售增长率	源	自由度	卡方	Pr > 卡方
	a	2	2.7882	0.2481



Levene方差齐性检验

净资产收益率	源	自由度	平方和	均方	F 值	$Pr > F$
	a	2	35487.3	17743.7	4.38	0.0209
	误差	32	129720	4053.8		
资产负债率	源	自由度	平方和	均方	F 值	$Pr > F$
	a	2	75556.4	37778.2	0.53	0.5953
	误差	32	2292983	71655.7		
流动资产周转率	源	自由度	平方和	均方	F 值	$Pr > F$
	a	2	4.8020	2.4010	7.28	0.0025
	误差	32	10.5605	0.3300		
销售增长率	源	自由度	平方和	均方	F 值	$Pr > F$
	a	2	207308	103654	1.55	0.2280
	误差	32	2141412	66919.1		

行业指标统计量



表 2-7 3 个行业各个指标描述统计量的估计

行业	样本量	净资产收益率		资产负债率		流动资产周转率		销售增长率	
		均值	标准差	均值	标准差	均值	标准差	均值	标准差
1	13	9.026	6.890	45.854	15.956	1.455	0.983	7.002	14.520
2	12	9.303	4.858	28.112	16.132	0.705	0.281	9.940	17.647
3	10	19.229	10.569	53.639	12.199	1.358	0.476	13.384	10.030