doi:10.3969/j.issn.1673-9833.2015.03.019

方差成比例时2个正态总体均值的假设检验

冒霜霜，焦肖红，邓锦叶

（中南大学 数学与统计学院，湖南 长沙 410006）

摘　要：在方差未知但相等，或方差未知且大样本的情况下，2 个正态总体均值的假设检验已经给出了解决方法；但是在方差不等且小样本的情况下，2 个正态总体均值的假设检验少有研究。针对这一问题，给出了理论证明的处理方法，并设计了实现流程和用于实现的 MATLAB 程序，最后，以实际案例给出了实现的具体方法。与大样本情况下的方法对比，该方法所需样本数量较小。

关键词：概率论与数理统计；假设检验；方差成比例；MATLAB

中图分类号：O212.1　　**文献标志码：**A　　**文章编号：**1673-9833(2015)03-0101-08

Hypothesis Test of the Means of Two Normal Populations with Proportional Variances

Mao Shuangshuang，Jiao Xiaohong，Deng Jinye

（School of Mathematics and Statistics，Central South University，Changsha 410006，China）

Abstract：For two normal populations with unknown but equal variances or unknown variances but with large sample size, hypothesis test of the means has given a solution. But there is short of research for unequal variances with small sample size. To address the problem, the processing method of theoretical proof is put forward, and the realizing process and MATLAB program for realization are designed. Finally an actual case is used to illustrate the feasibility of the method. Compared with the situation of large sample, this method needs less samples.

Keywords：probability and statistics；hypothesis test；proportional variances；MATLAB

1　研究背景

设总体$X \sim N\left(\mu_1,\sigma_1{}^2\right)$，$Y \sim N\left(\mu_2,\sigma_2{}^2\right)$，$\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$均为未知参数，$(X_1, X_2, \cdots, X_m)$和$(Y_1, Y_2, \cdots, Y_n)$分别是从总体 X，Y 中抽得的相互独立的 2 个样本。记

$$\overline{X}=\frac{1}{m}\sum_{i=1}^{m}X_i，\overline{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i；$$

$$S_1{}^2=\frac{1}{m-1}\sum_{i=1}^{m}\left(X_i-\overline{X}\right)^2，S_2{}^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2。$$

对于假设检验

$$\mathrm{H}_0: \mu_1=\mu_2, \mathrm{H}_1: \mu_1\neq\mu_2,$$

有如下 3 种情形：

情形 1　对于 2 个正态总体方差都已知，要检验 2 个正态总体均值是否相等，有检验统计量[1-2]

$$U=\frac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}}\overset{\mathrm{H}_0成立}{\sim}N(0,1)。$$

情形 2　对于 2 个正态总体方差都未知但相等，要检验 2 个正态总体均值是否相等，有检验统计量

$$U=\frac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}}\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}\overset{\mathrm{H}_0成立}{\sim}t(m+n-2)。$$

情形 3　对于 2 个正态总体方差都未知，要检验 2 个正态总体均值是否相等，当 m, n 很大时($m, n \geqslant 50$)，有检验统计量

收稿日期：2015-04-07

作者简介：冒霜霜（1990-），女，湖南邵阳人，中南大学硕士生，主要研究方向为概率论与数理统计，E-mail: canjingdao@163.com

$$U=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}}\overset{H_0成立}{\sim}N(0,1)。$$

对于方差不等且小样本的情形下，2个正态总体均值的假设检验，目前的研究成果少见。本文给出对于方差不等且小样本的情形下，2个正态总体均值的假设检验的方法。

2 方差成比例时2个正态总体均值的假设检验有关定理

引理1 设总体$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，$\mu_1$，$\mu_2$，$\sigma_1^2$，$\sigma_2^2$均为未知参数，$(X_1, X_2, \cdots, X_m)$和$(Y_1, Y_2, \cdots, Y_m)$分别是从总体$X$，$Y$中抽得的相互独立的2个样本，当$\frac{\sigma_1^2}{\sigma_2^2}=c$为定常数时，统计量$F=\frac{S_1^2}{cS_2^2}\sim F(m-1,n-1)$[3]。

证 记$S_1^2=\frac{1}{m-1}\sum_{i=1}^{m}(X_i-\overline{X})^2$，

$$S_2^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2,$$

则$\frac{m-1}{\sigma_1^2}S_1^2\sim\chi^2(m-1)$，$\frac{n-1}{\sigma_2^2}S_2^2\sim\chi^2(n-1)$。

当$\frac{\sigma_1^2}{\sigma_2^2}=c$为定常数时，$F=\frac{S_1^2}{cS_2^2}\sim F(m-1,n-1)$。

定理1 设总体$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，$\mu_1$，$\mu_2$，$\sigma_1^2$，$\sigma_2^2$均为未知参数，$(X_1, X_2, \cdots, X_m)$和$(Y_1, Y_2, \cdots, Y_n)$分别是从总体$X$，$Y$中抽得的相互独立的2个样本。当$\frac{\sigma_1^2}{\sigma_2^2}=c$为定常数时，则有

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{cn+m}{cmn}}}\sim t(m+n-2),$$

式中$S_\omega=\sqrt{\frac{(m-1)S_1^2+c(n-1)S_2^2}{m+n-2}}$[4-5]。

证 $\overline{X}-\overline{Y}-(\mu_1-\mu_2)\sim N\left(0,\sigma_2^2\left(\frac{c}{m}+\frac{1}{n}\right)\right)$，

$$U\overset{\Delta}{=}\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_2^2\left(\frac{c}{m}+\frac{1}{n}\right)}}\sim N(0,1)。$$

由$S_1^2=\frac{1}{m-1}\sum_{i=1}^{m}(X_i-\overline{X})^2$，$S_2^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2$，

$$\frac{(m-1)S_1^2+c(n-1)S_2^2}{c\sigma_2^2}\sim\chi^2(m+n-2)。$$

记$S_\omega^2=\frac{(m-1)S_1^2+c(n-1)S_2^2}{m+n-2}$，则

$$V\overset{\Delta}{=}\frac{(m+n-2)S_\omega^2}{c\sigma_2^2}\sim\chi^2(m+n-2)。$$

因U，V相互独立，则有

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{cn+m}{cmn}}}=\frac{U}{\sqrt{V/(m+n-2)}}\sim t(m+n-2)。$$

3 方差成比例时2个正态总体均值的假设检验步骤

1）对于2个正态总体方差都未知，要检验2个正态总体均值是否相等，先提出假设检验

$$H_0: \sigma_1^2=\sigma_2^2,\ H_1: \sigma_1^2\neq\sigma_2^2;$$

检验统计量

$$F=\frac{S_1^2}{S_2^2}\sim F(m-1,n-1)。$$

若接受H_0: $\sigma_1^2=\sigma_2^2$，则提出检验

$$H_0: \mu_1=\mu_2,\ H_1: \mu_1\neq\mu_2;$$

检验统计量

$$\frac{(\overline{X}-\overline{Y})}{S_\omega\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(m+n-2),$$

式中$S_\omega^2=\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}$。

2）在1）中，若关于2个正态总体方差是否相等的检验，检验结果为拒绝H_0: $\sigma_1^2=\sigma_2^2$，则作$\frac{\sigma_1^2}{\sigma_2^2}=c$的点估计。从原样本中随机提取（总体$X$中取容量为$m_1$，总体$Y$中取容量为$n_1$）一部分子样，用子样构建$\sigma_1^2$，$\sigma_2^2$的极大似然估计$s_1^{*2}$，$s_2^{*2}$。并设$\overline{x}=\frac{1}{m_1}\sum_{i=1}^{m_1}x_i$，$\overline{y}=\frac{1}{n_1}\sum_{i=1}^{n_1}y_i$；$s_1^{*2}=\frac{1}{m_1}\sum_{i=1}^{m_1}(x_i-\overline{x})^2$，$s_2^{*2}=\frac{1}{n_1}\sum_{i=1}^{n_1}(y_i-\overline{y})^2$。

由极大似然估计的不变性可知，$\hat{C}=\frac{s_1^{*2}}{s_2^{*2}}$也是$c$的极大似然估计，故可用$\hat{C}$代替$c$。再提出检验

$$H_0: \frac{\sigma_1^2}{\sigma_2^2}=c,\ H_1: \frac{\sigma_1^2}{\sigma_2^2}\neq c;$$

检验统计量

$$F=\frac{S_1^2}{cS_2^2}\sim F(m-1,n-1)。\tag{1}$$

此时的拒绝域为

$$W=\left(0,F_{1-\frac{\alpha}{2}}(m-1,n-1)\right)\cup\left(F_{\frac{\alpha}{2}}(m-1,n-1),+\infty\right)。$$

若拒绝H_0，需从原样本中重新随机提取一部分子样来估计c，直至接受H_0为止，并进一步作下述检验：

$$H_0: \mu_1=\mu_2,\ H_1: \mu_1 \neq \mu_2;$$

检验统计量

$$\frac{\left(\bar{X}-\bar{Y}\right)}{S_\omega\sqrt{\frac{cn+m}{cmn}}} \sim t\left(m+n-2\right), \tag{2}$$

式中$S_\omega^2=\dfrac{\left(m-1\right)S_1^{\,2}+c\left(n-1\right)S_2^{\,2}}{m+n-2}$。

此时的拒绝域为

$$W=\left(-\infty,-t_{\frac{\alpha}{2}}\left(m+n-2\right)\right)\cup\left(t_{\frac{\alpha}{2}}\left(m+n-2\right),+\infty\right)。$$

注意：式（1）和式（2）中的S_1^2, S_2^2和$\bar{X}$, $\bar{Y}$是由剩余的样本而求得。

综上所述，在小样本的情形下，对于2个正态总体均值的假设检验可按图1所示流程进行。

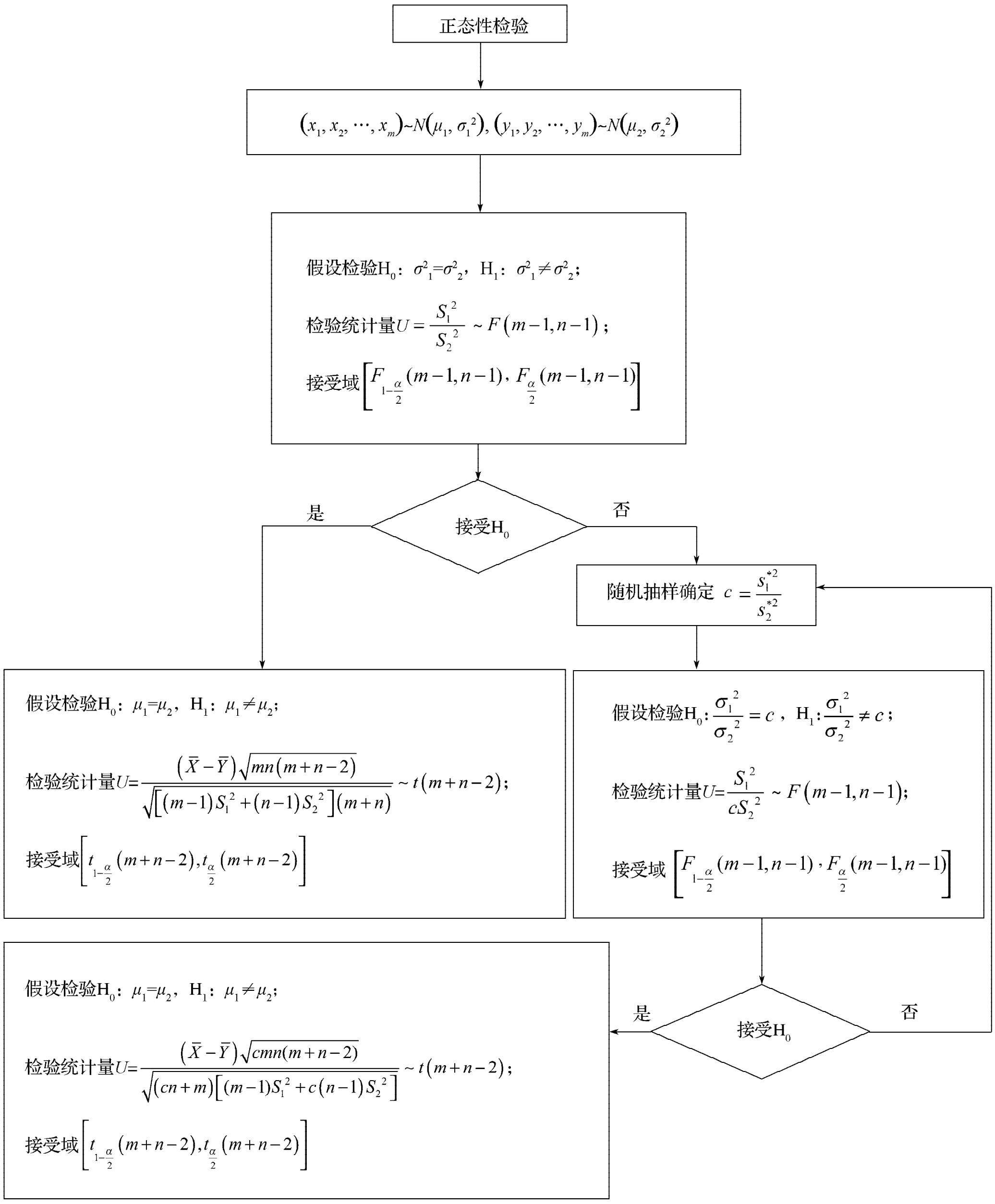

图1　小样本情形下2个正态总体均值的假设检验流程

Fig. 1　Flowchart of hypothesis test of the means of two normal populations in the case of small samples

4　应用案例

根据调查得知男女各250名的月收入[6]（见附表1）。首先，对数据进行正态性检验[7]（相应函数的MATLAB程序见附录1），在MATLAB中输入如下程序：

```
x=[14300 24300 18800 16500 …
27200 26200 29000];
```

y=[3800 22300 12400 6600 …

36100 13200 16800];

hx=fun_1(x)

hy=fun_1(y)

检验结果如下：

hx =

0

hy =

0

正态性检验的结果表明，男性和女性 2 组样本均服从正态分布。

此时，样本容量为 250，属于大样本的情形，可直接利用函数 fun_2[8−9]进行假设检验

$$H_0: \mu_1=\mu_2, H_1: \mu_1 \neq \mu_2。$$

在 MATLAB 中输入如下命令：

[h,ci,stat]=fun_2(x,y,0.05)

运行结果如下：

h =

1

ci =

−1.9600 1.6449

stat =

5.5366

结果表明，拒绝原假设 $H_0: \mu_1=\mu_2$，即认为 2 组数据均值不等。

现在考虑更符合实际的一种情况，由于抽样受限，只得到了其中男女各 40 组数据。

男性: 16 500, 7 800, 1 200, 16 900, 4 900, 7 700, 48 100，44 300，14 500，22 900，14 200，11 900，55 300, 29 400, 30 500, 32 600, 51 000, 1 400, 31 400, 44 400, 19 800, 8 400, 23 700, 18 500, 17 600, 39 400, 37 500, 16 200, 4 400, 40 900, 2 500, 37 200, 17 500, 26 900, 25 400, 15 300, 22 600, 44 500, 34 200, 7 700。

女性：27 400，16 300，11 100，8 200，15 400，11 200, 13 100, 32 000, 29 200, 9 400, 800, 21 500, 18 400, 1 300, 9 000, 13 600, 23 000, 12 400, 11 600, 27 600, 20 500, 200, 13 000, 14 400, 21 800, 40 900, 7 100，8 200，26 900，12 300，8 100，8 400，3 900，9 500, 23 800, 38 800, 16 000, 18 400, 12 900, 2 400。

此时，属于小样本情形，可按图 1 给定的流程进行假设检验。

同样地，首先对数据进行正态性检验，在 MATLAB 中输入以下程序：

x=[16500 7800 1200 …44500 34200 7700];

y=[27400 16300 11100 …18400 12900 2400];

hx=fun_1(x)

hy=fun_1(y)

检验结果如下：

hx =

0

hy =

0

正态性检验的结果表明，男性和女性 2 组样本均服从正态分布。

现为确定男性与女性的工资是否存在显著差异，从男女中各随机抽出 20 组数据。

男性：22 600, 55 300, 7 800, 44 300, 4 900, 16 900, 16 500，39 400，1 400，15 300，48 100，16 200，7 700，32 600，1 200，14 200，44 400，11 900，37 200，17 500。

女性：16 000，18 400，16 300，32 000，15 400，8 200，27 400，40 900，12 400，38 800，1 3100，8 200，2 400，13 600，11 100，800，27 600，21 500，8 400，3 900。

用剩下的样本进行检验，假设男性月收入服从 $N(\mu_1, \mu_1^2)$，女性月收入服从 $N(\mu_2, \mu_2^2)$。假设检验

$$H_0: \sigma_1^2=\sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2。$$

利用函数 fun_3，在 MATLAB 中输入如下命令：

x=[26900 34200 22900 23700 51000 31400 25400 4400 19800 40900 14500 2500 44500 17600 37500 7700 8400 30500 29400 18500];

y=[9500 12900 9400 13000 23000 11600 23800 26900 20500 12300 29200 8100 18400 21800 7100 11200 200 9000 1300 14400];

[h,ci,stat]=fun_3(x,y,0.05)

运行结果如下：

h =

1

ci =

0.3958 2.5265

stat=

2.7858

结果表明，拒绝原假设 $H_0: \sigma_1^2=\sigma_2^2$，即认为 $\sigma_1^2 \neq \sigma_2^2$。

然后，进行假设检验

$$H_0: \frac{\sigma_1^2}{\sigma_2^2}=c, H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq c。$$

此时，用抽中的样本确定常数，得 c=2.139 1。

利用函数 fun_4，在 MATLAB 中输入如下命令：

[h,ci,stat]=fun_4(x,y,0.05,2.1391)

运行结果如下：

h =

0

ci =

0.3958　2.5265

stat=

1.3023

结果表明，接受原假设 H_0: $\frac{\sigma_1^2}{\sigma_2^2}=c$（$c$=2.139 1）。

最后进行假设检验

H_0: $\mu_1=\mu_2$，H_1: $\mu_1\neq\mu_2$。

利用函数 fun_5，在 MATLAB 中输入如下命令：

```
[h,ci,stat]=fun_5(x,y,0.05,2.1391)
```

运行结果如下：

h =

1

ci =

−2.0244　2.0244

stat=

3.0602

结果表明，拒绝原假设 H_0: $\mu_1=\mu_2$，即认为2组数据均值不相等。

综合以上结果可知，男女工资均值不相等，方差成比例，因此男女工资存在显著性差异。这与用大样本的方法直接进行假设检验 H_0: $\mu_1=\mu_2$，H_1: $\mu_1\neq\mu_2$，所得的结果一致。

5　结语

从以上实际应用可以看出，本文在方差成比例时2个正态总体均值的假设检验方法是可行的，而且此方法所需的数据量较少，在实际应用中具有一定的优势。此外，本文编程得到的5个MATLAB函数 fun_1，fun_2，fun_3，fun_4，fun_5，能为解决一些假设检验问题提供便捷的方法。

参考文献：

[1] 邓集贤，杨维权. 概率论及数理统计[M]. 4版. 北京：高等教育出版社，2009：81-102.
Deng Jixian，Yang Weiquan. Probability and Mathematical Statistics[M]. 4th ed. Beijing：Higher Education Press，2009：81-102.

[2] 茆诗松，程依明，濮晓龙. 概率论与数理统计教程[M]. 北京：高等教育出版社，2004：285-304.
Mao Shisong，Cheng Yiming，Pu Xiaolong. Tutorial of Probability and Mathematical Statistics[M]. Beijing：Higher Education Press，2004：285-304.

[3] 姜培华. 两正态总体方差比的最优区间估计和最佳双边检验[J]. 菏泽学院学报，2011，33(2)：1-6.
Jiang Peihua. The Best Interval Estimation and the Best Two-Side Test of Variance Ratio of Two Normal Population [J]. Journal of Heze University，2011，33(2)：1-6.

[4] 王兆军，邹长亮. 数理统计教程[M]. 北京：高等教育出版社，2014：148-152.
Wang Zhaojun，Zou Changliang. Mathematical Statistics Tutorial[M]. Beijing：Higher Education Press，2014：148-152.

[5] Gurland J，McCullough R S. Testing Equality of Means After a Preliminary Test of Equality of Variances[J]. Biometrika，1962，49(3/4)：403-417.

[6] 薛　薇. 统计分析与SPSS[M]. 2版. 北京：中国人民大学出版社，2007：136-145.
Xue Wei. Statistical Analysis and SPSS[M]. 2nd ed. Beijing：China Renmin University Press，2007：136-145.

[7] 易昆南. 基于数学建模的数学实验[M]. 北京：中国铁道出版社，2014：151-155.
Yi Kunnan. Mathematical Experiment Based on Modeling [M]. Beijing：China Railway Press,，2014：151-155.

[8] 傅　鹂，龚　劬. 数学实验[M]. 北京：科学出版社，2000：10-30.
Fu Li，Gong Qu. Mathematical Experiment[M]. Beijing：Science Press，2000：10-30.

[9] 刘二根，王广超. MATLAB与数学实验[M]. 北京：国防工业出版社，2014：116-141.
Liu Ergen，Wang Guangchao. MATLAB and Mathematical Experiment[M]. Beijing：National Defend Industry Press，2014：116-141.

（责任编辑：邓光辉）

附表1　男女各250名月收入表

Appendix 1　Monthly income table of 250 men and 250 women

Ordinal	Sex	Income	Ordinal	Sex	Income	Ordinal	Sex	Income	Ordinal	Sex	Income
1	Female	3 800	8	Male	16 500	15	Male	1 400	22	Male	31 400
2	Male	14 300	9	Female	15 000	16	Male	1 200	23	Female	21 700
3	Female	22 300	10	Male	35 900	17	Female	25 500	24	Male	21 000
4	Male	24 300	11	Female	12 400	18	Female	15 400	25	Male	30 500
5	Female	12 400	12	Female	11 100	19	Male	18 200	26	Male	24 500
6	Female	6 600	13	Female	8 300	20	Female	9 000	27	Male	2 400
7	Male	18 800	14	Female	31 900	21	Male	31 100	28	Female	34 200

续表

Ordinal	Sex	Income	Ordinal	Sex	Income	Ordinal	Sex	Income	Ordinal	Sex	Income
29	Male	27 400	93	Female	2 500	157	Female	26 600	221	Male	19 500
30	Male	28 600	94	Male	4 600	158	Male	48 100	222	Female	26 900
31	Male	31 400	95	Male	11 300	159	Male	17 200	223	Male	34 600
32	Female	1 800	96	Female	7 800	160	Female	15 800	224	Female	20 900
33	Male	23 300	97	Male	14 600	161	Female	4 900	225	Female	30 400
34	Female	7 300	98	Male	21 800	162	Male	8 700	226	Female	19 400
35	Male	1 200	99	Male	16 100	163	Male	56 500	227	Male	10 800
36	Male	14 500	100	Female	18 900	164	Male	25 600	228	Male	43 800
37	Female	11 600	101	Female	2 800	165	Female	15 600	229	Male	14 800
38	Female	23 700	102	Male	34 200	166	Male	32 900	230	Male	7 800
39	Male	9 900	103	Female	28 800	167	Male	52 600	231	Male	17 500
40	Female	9 800	104	Male	17 500	168	Female	12 600	232	Female	37 100
41	Female	19 900	105	Male	27 200	169	Female	10 800	233	Male	40 900
42	Male	20 800	106	Male	19 800	170	Female	18 400	234	Female	13 600
43	Female	13 000	107	Female	28 800	171	Male	17 800	235	Male	14 300
44	Male	26 800	108	Female	9 900	172	Male	6 500	236	Male	14 500
45	Female	16 500	109	Female	9 500	173	Male	9 800	237	Male	16 500
46	Female	11 100	110	Female	31 500	174	Male	55 300	238	Male	18 900
47	Female	26 200	111	Male	19 200	175	Male	13 700	239	Female	26 100
48	Male	18 600	112	Male	13 300	176	Female	24 300	240	Male	11 300
49	Male	8 300	113	Female	15 100	177	Female	3 000	241	Male	16 200
50	Male	16 500	114	Female	6 900	178	Female	8 800	242	Male	6 900
51	Male	9 200	115	Male	23 000	179	Male	19 000	243	Female	16 100
52	Female	2 900	116	Female	11 500	180	Female	23 600	244	Female	5 100
53	Male	45 300	117	Female	32 500	181	Male	13 300	245	Female	2 000
54	Female	27 400	118	Male	30 400	182	Male	6 700	246	Female	19 000
55	Male	12 300	119	Female	21 400	183	Female	23 000	247	Male	7 700
56	Female	5 600	120	Male	11 900	184	Female	29 300	248	Male	32 300
57	Male	10 300	121	Male	29 000	185	Male	9 400	249	Female	16 300
58	Female	36 100	122	Male	26 900	186	Male	42 000	250	Female	5 100
59	Female	23 500	123	Female	21 800	187	Female	12 200	251	Male	20 300
60	Female	4 400	124	Male	15 600	188	Male	21 000	252	Male	23 800
61	Male	23 300	125	Female	23 000	189	Female	6 000	253	Male	27 100
62	Female	20 000	126	Female	18 700	190	Female	4 800	254	Female	12 300
63	Female	8 400	127	Female	10 000	191	Male	35 300	255	Female	15 000
64	Female	24 500	128	Female	9 400	192	Male	34 900	256	Female	29 200
65	Female	33 200	129	Male	14 000	193	Male	28 700	257	Female	27 100
66	Female	4 900	130	Male	8 800	194	Male	28 000	258	Male	27 500
67	Male	12 500	131	Male	600	195	Female	12 000	259	Female	25 600
68	Male	17 700	132	Female	22 600	196	Male	20 300	260	Female	24 600
69	Male	11 400	133	Male	46 300	197	Female	4 700	261	Male	26 500
70	Female	39 600	134	Female	10 200	198	Male	22 500	262	Female	8 200
71	Female	16 000	135	Female	26 000	199	Male	32 600	263	Female	23 600
72	Male	8 000	136	Male	6 800	200	Female	1 400	264	Female	13 300
73	Female	23 800	137	Female	8 400	201	Male	40 900	265	Male	10 900
74	Female	23 300	138	Male	17 600	202	Male	23 800	266	Female	6 800
75	Male	51 000	139	Male	51 000	203	Male	28 400	267	Female	15 900
76	Female	15 400	140	Female	23 100	204	Male	24 300	268	Male	7 000
77	Female	16 500	141	Female	10 200	205	Male	37 100	269	Female	18 400
78	Male	22 600	142	Male	12 400	206	Female	18 600	270	Male	7 900
79	Female	4 900	143	Female	1 100	207	Male	18 800	271	Female	22 000
80	Female	22 100	144	Female	13 100	208	Male	4 400	272	Female	11 800
81	Female	9 300	145	Male	31 400	209	Male	25 700	273	Male	35 100
82	Female	29 700	146	Male	22 900	210	Female	13 600	274	Male	5 300
83	Female	13 800	147	Female	8 900	211	Female	15 000	275	Female	18 200
84	Female	16 800	148	Female	11 700	212	Male	7 000	276	Female	10 800
85	Male	1 700	149	Male	13 000	213	Male	20 900	277	Male	15 200
86	Female	12 900	150	Female	23 100	214	Female	22 700	278	Male	27 900
87	Male	4 200	151	Male	9 500	215	Male	14 300	279	Female	31 200
88	Male	22 800	152	Male	23 900	216	Female	26 100	280	Female	29 500
89	Female	3 900	153	Male	37 200	217	Female	26 600	281	Female	3 200
90	Female	19 000	154	Male	28 100	218	Female	20 100	282	Female	3 400
91	Male	28 800	155	Male	18 200	219	Female	3 700	283	Female	14 500
92	Male	39 400	156	Male	40 200	220	Male	39 400	284	Male	27 500

续表

Ordinal	Sex	Income	Ordinal	Sex	Income	Ordinal	Sex	Income	Ordinal	Sex	Income
285	Male	6 000	339	Female	14 200	393	Female	19 000	447	Male	14 100
286	Female	4 400	340	Female	21 900	394	Male	2 600	448	Male	35 100
287	Female	14 400	341	Female	8 900	395	Female	17 100	449	Female	18 200
288	Female	16 200	342	Male	35 900	396	Female	21 500	450	Male	1 200
289	Female	3 600	343	Male	44 500	397	Male	19 800	451	Male	46 900
290	Male	20 100	344	Female	23 200	398	Male	6 700	452	Female	21 300
291	Male	18 500	345	Female	14 400	399	Male	19 400	453	Female	27 300
292	Female	14 400	346	Male	39 400	400	Male	40 000	454	Male	4 100
293	Male	45 600	347	Male	29 400	401	Male	30 300	455	Male	32 700
294	Male	28 600	348	Male	13 100	402	Male	25 400	456	Female	22 200
295	Female	22 100	349	Male	15 700	403	Female	17 200	457	Male	34 300
296	Female	4 700	350	Male	37 200	404	Female	2 700	458	Female	40 900
297	Female	5 500	351	Female	13 900	405	Male	3 200	459	Female	2 600
298	Female	27 500	352	Female	18 200	406	Male	25 400	460	Female	800
299	Female	7 100	353	Male	36 400	407	Male	20 600	461	Male	1 500
300	Male	41 100	354	Male	14 600	408	Male	11 900	462	Male	18 600
301	Male	12 100	355	Female	5 800	409	Male	21 800	463	Male	22 600
302	Female	14 800	356	Male	24 300	410	Female	27 600	464	Female	19 300
303	Male	23 800	357	Female	11 200	411	Female	1 500	465	Female	8 300
304	Female	15 200	358	Female	15 800	412	Male	4 200	466	Male	21 100
305	Male	33 600	359	Female	7 100	413	Male	44 400	467	Male	29 800
306	Female	29 900	360	Male	19 800	414	Female	13 000	468	Female	3 500
307	Female	31 800	361	Female	9 000	415	Female	21 100	469	Male	39 400
308	Female	23 200	362	Female	15 600	416	Female	8 200	470	Male	4 900
309	Female	15 100	363	Female	16 500	417	Male	16 800	471	Male	14 200
310	Female	28 200	364	Male	27 600	418	Female	8 800	472	Male	40 500
311	Male	19 800	365	Male	58 700	419	Male	23 700	473	Female	200
312	Female	17 100	366	Male	28 300	420	Female	20 900	474	Male	37 700
313	Female	1 200	367	Male	12 800	421	Female	19 000	475	Female	17 700
314	Male	2 900	368	Female	11 800	422	Male	29 600	476	Male	18 000
315	Female	15 700	369	Female	38 800	423	Male	16 900	477	Female	2 800
316	Female	32 400	370	Male	37 300	424	Female	15 400	478	Male	8 400
317	Female	22 500	371	Female	12 800	425	Female	13 000	479	Female	2 400
318	Male	1 700	372	Male	14 400	426	Female	24 700	480	Male	31 200
319	Female	9 300	373	Female	28 800	427	Male	1 000	481	Female	4 900
320	Female	28 800	374	Female	6 600	428	Male	39 800	482	Female	25 800
321	Female	18 400	375	Male	30 600	429	Male	1 400	483	Male	16 700
322	Male	30 400	376	Female	20 500	430	Male	4 900	484	Female	10 500
323	Female	17 000	377	Female	21 100	431	Male	31 000	485	Male	7 700
324	Male	43 300	378	Female	19 300	432	Male	17 400	486	Male	23 300
325	Female	1 300	379	Male	7 700	433	Female	32 700	487	Male	30 200
326	Male	8 500	380	Male	25 200	434	Female	32 000	488	Female	21 100
327	Female	21 300	381	Male	37 500	435	Female	16 500	489	Male	18 500
328	Male	1 100	382	Male	9 400	436	Female	33 500	490	Male	38 500
329	Female	22 000	383	Male	45 100	437	Male	24 500	491	Female	8 100
330	Male	30 700	384	Male	12 800	438	Female	21 600	492	Male	2 500
331	Male	36 700	385	Female	15 300	439	Female	17 300	493	Male	55 700
332	Male	18 600	386	Male	600	440	Female	29 000	494	Male	27 200
333	Male	18 700	387	Male	15 300	441	Male	44 300	495	Female	20 700
334	Female	25 200	388	Female	20 000	442	Female	1 200	496	Female	36 100
335	Female	27 700	389	Male	22 900	443	Female	19 400	497	Female	13 200
336	Male	1 400	390	Female	23 800	444	Female	13 100	498	Female	16 800
337	Female	14 600	391	Male	19 000	445	Female	17 200	499	Male	26 200
338	Female	6 700	392	Female	14 100	446	Male	18 100	500	Male	29 000

附录 1　fun_1~ fun_5 的 MATLAB 程序

1）正态性检验函数 fun_1

```
function [h]=fun_1(x)
[muhat,sigmahat,muci,sigmaci]=normfit(x);
h=kstest((x-muhat)./sigmahat); %返回值h=0，表示样本服从正态分布；返回值 h=1，表示样本不服从正态分布。
```

2)大样本假设检验H_0: $\mu_1=\mu_2$, H_1: $\mu_1\neq\mu_2$函数fun_2

```
function [h,ci,stat]=fun_2(x,y,alpha)
m=length(x);
n=length(y);
s1_2=var(x);
```

```
s2_2=var(y);
xbar=mean(x);
ybar=mean(y);
stat=(xbar-ybar)/sqrt(s1_2/m+s2_2/n); % 检验统计量
c1=norminv(alpha/2,0,1);
c2=norminv(1-alpha/1,0,1);
ci=[c1,c2]; %接受域
if stat>c1 & stat<c2
  h=0; % 接受原假设
else
  h=1; % 拒绝原假设
end
```

3）小样本假设检验 H_0: $\sigma^2{}_1=\sigma^2{}_2$，$H_1$: $\sigma^2{}_1 \neq \sigma^2{}_2$，函数 fun_3

```
function [h,ci,stat]=fun_3(x,y,alpha)
n1=length(x);
n2=length(y);
s1_2=var(x);
s2_2=var(y);
stat=s1_2/s2_2; % 检验统计量
c1=finv(alpha/2,n1-1,n2-1);
c2=finv(1-alpha/2,n1-1,n2-1);
ci=[c1,c2]; %接受域
if stat>c1 &stat<c2
  h=0; % 接受原假设
else
  h=1; % 拒绝原假设
end
```

4）小样本假设检验 H_0: $\frac{\sigma_1{}^2}{\sigma_2{}^2}=c$，$H_1$: $\frac{\sigma_1{}^2}{\sigma_2{}^2}\neq c$函数 fun_4

```
function [h,ci,stat]=fun_4(x,y,alpha,c)
n1=length(x);
n2=length(y);
s1_2=var(x);
s2_2=var(y);
stat=s1_2/(c*s2_2); % 检验统计量
c1=finv(alpha/2,n1-1,n2-1);
c2=finv(1-alpha/2,n1-1,n2-1);
ci=[c1,c2]; %接受域
if stat>c1 & stat<c2
  h=0; % 接受原假设
else
  h=1; % 拒绝原假设
end
```

5）小样本假设检验 H0：$\mu_1=\mu_2$，H_1：$\mu_1 \neq \mu_2$ 函数 fun_5

```
function [h,ci,stat]=fun_5(x,y,alpha,c)
m=length(x);
n=length(y);
s1_2=var(x);
s2_2=var(y);
xbar=mean(x);
ybar=mean(y);
stat=(xbar-ybar)*sqrt(m*n*(m+n-2))/sqrt((c*n+m)*((m-1)*s1_2/c+(n-1)*s2_2)); %检验统计量
c1=tinv(alpha/2,m+n-2);
c2=tinv(1-alpha/2,m+n-2);
ci=[c1,c2];  %接受域
if stat>c1 & stat<c2
  h=0; % 接受原假设
else
  h=1; % 拒绝原假设
end
```