产生服从 F 分布的随机数的例子如下

```
frnd(10,4)          ans = 2.8855
frnd(10,10)         ans = 0.2757
frnd(10,20)         ans = 1.5200
frnd(2,10,5,7)
ans =1.6613   0.1962   1.8418   0.2085   0.4924   0.2682   0.8976
     0.7198   0.3330   0.7767   1.9925   0.5375   1.0791   0.4058
     0.4973   1.0102   3.3659   0.1749   0.2769   0.9471   2.6272
     5.6393   2.8133   1.5102   3.3269   0.2992   0.0420   0.9972
     0.0888   0.0992   0.1880   0.4350   0.3622   0.2180   0.2901
frnd(10,20,5,7)
ans =0.6071   0.8229   0.8108   0.8462   0.9936   1.5513   1.0835
     1.5606   0.6781   0.9967   0.3176   1.5840   0.2994   1.4440
     1.6310   0.1674   0.7198   1.2456   0.5857   0.2091   0.8690
     0.6606   2.5233   2.1569   0.9866   0.3704   0.9820   0.5452
     0.4959   0.3080   2.8579   2.4121   0.6325   0.4102   1.3503
```

求 F 分布数学期望与方差的例子如下

```
[e d]=fstat(10,4)       e = 2          d = NaN
[e d]=fstat(2,10)       e = 1.2500     d = 2.6042
[e d]=fstat(4,15)       e = 1.1538     d = 1.0288
[e d]=fstat(10,20)      e = 1.1111     d = 0.4321
```

第三节 参数估计与假设检验

一、样本的数字特征

统计推断的基础是通过搜集、整理、加工和分析统计数据,使之系统化、条理化,以显示数据资料的趋势、特征和数量关系.

1. 集中趋势(位置)度量

数据样本集中趋势度量的目的在于对数据样本在数据分布线上分布的中心予以定位,即中心位置的度量.集中趋势度量包括几何平均值、调和平均值、算术平均值、中位数和修正的样本均值等.在 MATLAB 中分别用下述函数来求这些度量值,格式如下

geomean(X)　　样本的几何平均值

harmmean(X)　　样本的调和平均值

mean(X)　　样本的算术平均值

median(X)　　样本的中位数

trimmean(X, p)　　修正的样本均数

其中 $\boldsymbol{X}$ 为样本数据.若 $\boldsymbol{X}$ 为向量,则返回该向量的各度量值;若 $\boldsymbol{X}$ 为矩阵,则返回矩阵各列的度量值组成的行向量. P 为计算时剔除样本数据中最高 $P\%$ 和最低 $P\%$ 的数据后的均值.例如分别用指数分布和正态分布产生两个随机矩阵,然后求各种度量值如下

```
x = exprnd(1, 10, 7)
x =  0.0512   0.4854   2.8492   4.1816   0.1766   1.6428   0.7001
     1.4647   0.2333   1.0417   0.2920   3.9302   0.3824   0.1056
     0.4995   0.0814   0.2068   0.8095   0.3838   1.1948   0.1965
     0.7216   0.3035   4.6191   0.0706   0.9690   0.6131   0.4386
     0.1151   1.7358   1.9741   0.7636   0.1842   1.8913   0.2009
     0.2717   0.9021   1.5957   0.8707   0.6875   0.3597   0.4152
     0.7842   0.0667   1.6158   0.1670   0.3432   0.9719   1.0730
     3.9898   0.0868   0.5045   0.6441   0.8465   0.1508   1.2388
     0.1967   0.8909   1.3013   1.5963   1.1887   0.1582   1.0753
     0.8103   0.1124   1.6154   0.3973   1.6626   0.5216   0.6272
ge = geomean(x)
ge = 0.4597   0.2757   1.3137   0.5624   0.6720   0.5726   0.4654
he = harmmean(x)
he = 0.2281   0.1663   0.8708   0.3105   0.4550   0.4016   0.3343
me = mean(x)
me = 0.8905   0.4898   1.7324   0.9793   1.0372   0.7887   0.6071
de = median(x)
de = 0.6105   0.2684   1.6055   0.7038   0.7670   0.5674   0.5329
te = trimmean(x, 10)
te = 0.8905   0.4898   1.7324   0.9793   1.0372   0.7887   0.6071
y = normrnd(5, 2, 10, 7)
y = 4.1349   4.6266   5.5888   4.2002   1.7918   2.9787   5.0001
    1.6688   6.4516   2.3276   6.3800   5.5146   6.2289   4.3643
    5.2507   3.8234   6.4286   6.6312   2.8871   6.0155   7.1900
    5.5754   9.3664   8.2471   6.4238   7.8303   8.3849   1.2520
    2.7071   4.7272   3.6164   7.5805   3.3898   6.1826   5.8564
    7.3818   5.2279   6.7160   6.3372   6.0575   3.7128   6.7913
```

```
    7.3783  7.1335  7.5080  7.3817  5.4386  5.7607  6.4619
    4.9247  5.1186  1.8125  2.5951  3.1562  2.9818  6.1557
    5.6546  4.8087  2.1181  4.9604  0.6587  4.9610  5.0806
    5.3493  3.3353  6.1423  4.6866  4.8816  4.9036  6.3542
gn = geomean(y)
gn = 4.6207  5.2354  4.4383  5.4799  3.4701  4.9580  5.0135
hn = harmmean(y)
hn = 4.1354  5.0362  3.8014  5.1900  2.5658  4.7008  4.2439
mn = mean(y)
mn = 5.0026  5.4619  5.0506  5.7177  4.1606  5.2110  5.4506
dn = median(y)
dn = 5.3000  4.9636  5.8656  6.3586  4.1357  5.3608  6.0060
tn = trimmean(y,10)
tn = 5.0026  5.4619  5.0506  5.7177  4.1606  5.2110  5.4506
```

2. 离中趋势(散布)度量

离中趋势的度量可以理解为样本中的数据偏离中心位置的程度,或称为离差.离中趋势度量一般包括极差、标准差、方差、平均绝对偏差和内四分位数间距等.其中内四分位数间距是指数据的上四分位数 $X_{0.75}$ 与下四分位数 $X_{0.25}$ 之差,是随机变量以 1/2 的概率取值区间的长度,因而只用到了中间 50% 的数据影响其度量值,它对极端异己值的干扰极不敏感.在 MATLAB 中,求这些离中趋势度量的函数如下

iqr(X)	样本的内四分位数间距
mad(X)	样本的平均绝对偏差
range(X)	样本数据的极差
var(X)	样本数据的方差
std(X)	样本数据的标准差
cov(X)或 cov(X, Y)	样本数据的协方差

其中 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 均为样本数据.计算协方差时,若用 cov($\boldsymbol{X}$),$\boldsymbol{X}$ 必须为矩阵;若用 cov($\boldsymbol{X}$, $\boldsymbol{Y}$),$\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 均为长度相等的列向量.若 $\boldsymbol{X}$ 是矩阵,这些函数返回矩阵各列的离中趋势度量值组成的行向量;若 $\boldsymbol{X}$ 为向量,则返回该向量的各种度量值.我们利用产生随机数的办法求各种度量值如下(注意到在计算协方差时是将矩阵 $\boldsymbol{X}$ 的每一列作为一个变量求出其协方差矩阵)

```
x = normrnd(10,3,10,7)
x = 11.7067  11.8697  11.1696   6.4367  6.4423  10.3859
    12.4172   9.2331  12.3971  10.2640  3.3930   6.8323
```

```
    11.9694  10.6949   8.8676  12.8227   8.0936  12.9590
    14.4174   6.4965   7.0307   9.1123   7.0237   8.3213
     8.4441  10.1672   8.6182  14.0188   5.5746  10.6361
    11.3310  10.9821   6.3480   9.2127  10.8685   9.2980
    10.7136   7.1503  10.7022   9.8763   6.3605  14.4368
    10.3553   6.9767  12.3435  10.0644   6.6150   6.0417
    13.4141  10.9444   7.7739  11.7069   6.9882   5.9522
    12.7937   7.9476  14.3305  13.2469   7.5349   7.1586
     9.2167  10.0337   6.1242   8.9471   9.6055   9.2032
     8.8767  12.8604   8.0646   9.7812
in = iqr(x)
in = 1.9974  4.6233  3.2374  3.7140  3.7249  3.8894  5.4665
mn = mad(x)
mn = 1.5978  2.0632  1.6512  2.3084  2.5589  2.0062  2.5293
rn = range(x)
rn = 8.7559  6.2702  5.1933  9.5660  8.4653  6.7520  8.3126
vn = var(x)
vn = 5.1403  5.6484  3.5633  7.5342  8.8353  5.4722  8.6817
sn = std(x)
sn = 2.2672  2.3766  1.8877  2.7449  2.9724  2.3393  2.9465
cn = cov(x)
cn =   5.1403   1.0313  -0.7009  -2.6721  -0.6458   1.3515
      -2.1946   1.0313   5.6484  -1.8270  -0.7081   1.5938
       0.6958  -3.4171  -0.7009  -1.8270   3.5633  -1.1954
      -4.0718   1.4840   0.7338  -2.6721  -0.7081  -1.1954
       7.5342   4.2172  -5.1723   0.3121  -0.6458   1.5938
      -4.0718   4.2172   8.8353  -3.8626  -2.2460   1.3515
       0.6958   1.4840  -5.1723  -3.8626   5.4722  -2.4318
      -2.1946  -3.4171   0.7338   0.3121  -2.2460  -2.4318
       8.6817
y = poissrnd(5, 10, 10)
y = 2  7  3  2   5  4  6  3  9  3
    4  6  4  4   8  8  3  5  8  5
    5  1  2  5  11  4  3  2  6  4
    7  2  3  2   4  8  5  1  2  3
```

```
   3  9  4  6   9  7  4  5  3  6
   5  3  4  5   4  6  4  6  8  3
   3  5  4  5  15  8  4  5  3  6
   7  4  4  6   6  9  4  3  2  6
   3  7  7  8   7  5  7  6  4  5
   4  4  7  5   6  1  0  5  6  4
ip = iqr(y)
ip = 2  4  1  2  4  4  2  2  5  3
mp = mad(y)
mp = 1.3600  2.0000  1.1800  1.4800  3.3200  2.4800  1.3200
     1.4400  2.3000  1.1000
rp = range(y)
rp =  5  8  5  6  11  8  7  5  7  3
vp = var(y)
vp = 2.9000  6.1778  2.6222  3.2889  11.8333  6.2222  3.5556
     2.9889  6.9889  1.6111
sp = std(y)
sp = 1.7029  2.4855  1.6193  1.8135  3.4400  2.4944  1.8856
     1.7288  2.6437  1.2693
cp = cov(y)
cp =   2.9000  -3.0444  -0.7333  -0.3778  -2.0556   1.6667
      -0.5556  -3.0444   6.1778   1.3778   1.1778   0.7778
       0.4444   1.5556  -0.7333   1.3778   2.6222   1.7111
      -0.7778  -1.5556  -0.4444  -0.3778   1.1778   1.7111
       3.2889   1.7778  -0.1111   0.1111  -2.0556   0.7778
      -0.7778   1.7778  11.8333   1.3333  -0.8889   1.6667
       0.4444  -1.5556  -0.1111   1.3333   6.2222   1.6667
      -0.5556   1.5556  -0.4444   0.1111  -0.8889   1.6667
       3.5556  -1.5889   2.2444   1.8667   1.9111   0.8333
      -0.5556  -0.2222  -1.9222   0.2444  -0.3556  -1.5333
      -2.1667  -3.2222  -0.6667  -0.2778   1.4444   0.4444
       1.4444   2.7222   1.4444  -0.1111  -1.5889  -1.9222
      -0.2778
       2.2444   0.2444   1.4444
       1.8667  -0.3556   0.4444
```

```
       1.9111   -1.5333    1.4444
       0.8333   -2.1667    2.7222
      -0.5556   -3.2222    1.4444
      -0.2222   -0.6667   -0.1111
       2.9889    1.1000    0.7222
       1.1000    6.9889   -1.8333
       0.7222   -1.8333    1.6111
x1=unifrnd(1,4,1,10)
x1 =3.9617  3.6616  1.5569  1.7324  1.7406  1.5214  2.0029
    2.2187  2.8713  3.6940
x2=unifrnd(2,5,1,10)
x2 =4.0375  2.3946  3.4187  4.9023  3.4963  3.1912  4.9647
    4.3050  4.9494  2.7178
cov(x1,x2)
ans =   0.9320   -0.2812
       -0.2812    0.8827
```

3. 中心矩

中心矩是关于数学期望的矩. 对于任意的 $k>0$, 称 $\mu_k=E(\boldsymbol{X}-E(\boldsymbol{X}))^k$ 为随机变量 $\boldsymbol{X}$ 的 k 阶中心矩, 一阶中心矩为 0, 二阶中心矩为方差. 在 MATLAB 中, 可用函数 moment() 求中心矩, 格式为

moment(X, k)　　　求样本数据 X 的 k 阶中心矩

其中 $\boldsymbol{X}$ 为随机样本点组成的向量或矩阵, k 为正整数. 如果 $\boldsymbol{X}$ 为向量, 则返回该向量的 k 阶中心矩; 若 $\boldsymbol{X}$ 为矩阵, 则返回由 $\boldsymbol{X}$ 的每一列组成的向量的中心矩构成的行向量. 例如

```
x=normrnd(2.5,1.25,10,7)
x=2.8935  1.5724   3.2112  1.2451  0.8134  3.6640  1.6448
  4.3044  3.8529   1.4729  1.3161  2.1736  2.5141  0.8851
  2.0613  2.3356   2.1680  2.0320  3.6918  1.6936  2.4088
  3.2790  2.9874   1.0153  1.0176  2.6608  3.5072  2.0868
  3.4988  2.6100  -0.2529  1.1801  3.3206  2.7895  1.4455
  3.6761  1.7057   3.7329  4.3406  1.0402  1.2628  3.1222
  1.2599  1.8005   1.8517  2.5697  1.9242  4.1745  4.3606
  2.7650  3.0546   2.9092  0.9784  2.1720  2.8619  1.8169
  2.7974  1.3126   2.7926  2.4485  0.9836  4.3486  1.4416
  1.2403  3.4765   2.5268  1.0896  0.8507  3.9225  2.1921
```

```
y=normrnd(5,2,1,7)
y =4.7007   4.1305   4.8413   8.0703   3.7870   2.3053   5.9388
moment(x,2)
ans =0.9104   0.6750   1.2372   1.0235   0.9778   0.9717   0.8868
var(x)
ans =1.0115   0.7500   1.3747   1.1372   1.0864   1.0797   0.9854
moment(x,3)
ans = -0.2709  0.0966  -0.9486  1.4557  0.3404  -0.4475  0.9098
moment(x,4)
ans =1.7574   0.7855   4.3005   4.2280   1.7492   1.9278   2.8258
moment(y,2)            ans = 2.8138
var(y)                 ans = 3.2828
moment(y,3)            ans = 2.5882
moment(y,4)            ans = 22.0253
```

从上述二阶中心矩与方差结果来看,二阶中心矩并不等于方差.这是什么原因呢? 这是因为两者的定义有所不同

$$\mathrm{var}(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^2,\mathrm{moment}(\boldsymbol{X},2) = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^2,$$

其中 n 为样本容量.如上述 y 的样本容量为 7,则有:$2.8138\times 7\div 6 = 3.2828$,即 $\mathrm{var}(y) = \mathrm{moment}(y,2)\times 7\div 6$.

4. 样本的峰度与偏度

峰度也称峭度,它是单峰分布曲线"峰的平坦程度"的度量.一般峰度的定义为

$$S_4 = \frac{E(\boldsymbol{X}-\mu)^4}{\sigma^4} - 3,$$

这时,正态分布的峰度为 0.在不同文献中,峰度的定义有所不同,通常是不减后面的 3.即正态分布的峰度为 3;比正态分布曲线平坦的分布,其峰度大于 3;反之则小于 3.MATLAB 中使用的峰度定义即为后者,其函数的使用格式如下

kurtosis(X)　　求样本数据 $\boldsymbol{X}$ 的峰度

若 $\boldsymbol{X}$ 为矩阵,则按列求每列的峰度.

偏度是样本数据围绕其均值对称程度的度量.若偏度为负,则数据分布偏向于其均值的左边,反之则偏向其右,定义为

$$S_3 = \frac{E(\boldsymbol{X}-\mu)^3}{\sigma^3},$$

正态分布的偏度为 0.MATLAB 关于偏度的函数使用格式为

skewness(X) 求样本数据 X 的偏度

若 $\boldsymbol{X}$ 为矩阵,则按列求每列的偏度.例如

```
x = normrnd(1.5,2.5,10,7)
x = -0.7589  -2.4510   2.4274   3.2697  -0.0035  -4.3130
     1.0448   1.5897   1.3033   3.3207   6.3935   2.8780
    -1.5791   5.3025  -0.0688  -0.2041   6.7804   2.7614
    -1.2496   4.1391   1.4039   2.8385  -1.0614  -1.8932
     6.1613   1.7150   1.2169   4.5686   2.8822  -1.5859
    -1.0565   0.6505  -3.5114   2.4481  -0.2405   0.9908
     2.2220   4.0946  -1.3494   0.2673   3.8605   1.5188
    -3.6358   0.4267   0.5255   0.9722   2.6551  -3.8011
    -0.4572   1.8314   1.6395  -1.9532   4.4756   0.6975
    -0.1117   2.9673   5.4824   0.5803   2.2889  -1.2905
     4.5914  -0.2608   0.8720   4.0460   0.3376   5.3831
     3.0882  -0.0782  -1.0453   2.7003
y = exprnd(5,10,7)
y =  0.2558  2.4272  14.2459  20.9080   0.8830  8.2142  3.5003
     7.3237  1.1664   5.2083   1.4599  19.6511  1.9120  0.5281
     2.4974  0.4071   1.0341   4.0473   1.9189  5.9740  0.9823
     3.6079  1.5177  23.0957   0.3531   4.8448  3.0655  2.1932
     0.5754  8.6788   9.8703   3.8179   0.9208  9.4566  1.0046
     1.3584  4.5106   7.9785   4.3536   3.4377  1.7984  2.0759
     3.9212  0.3335   8.0792   0.8349   1.7162  4.8594  5.3652
    19.9489  0.4338   2.5226   3.2203   4.2327  0.7540  6.1941
     0.9837  4.4547   6.5063   7.9814   5.9435  0.7911  5.3765
     4.0517  0.5622   8.0769   1.9865   8.3128  2.6081  3.1361
z = frnd(4,9,10,7)
z = 1.9122  1.1848  0.7921  0.6150  2.0475  0.1906  0.4460
    3.8112  2.2065  0.1263  0.4510  0.3146  0.2579  0.6761
    1.6694  0.4877  1.2716  0.4329  0.7716  1.6545  0.0431
    0.7445  2.4418  1.4869  0.0797  1.1820  1.6516  2.2466
    0.1662  1.9040  2.1749  0.5158  0.3436  2.2712  1.0168
    1.2670  2.1547  1.0418  0.8750  1.0062  4.7817  0.3390
    0.4656  3.2058  0.6755  0.4957  2.3920  0.9811  3.0575
    0.4901  1.1466  0.4591  1.7362  1.5714  1.4936  0.3251
```

```
    1.3380   2.0402   1.2353   0.6993   0.8584   0.3702   0.4320
    0.6989   0.7712   1.6238   0.4247   0.6158   0.3420   1.0647
kurtosis(x)
ans = 2.8374   2.1302   1.8027   1.8561   2.6704   1.9227   2.1570
kurtosis(y)
ans = 6.1848   3.7124   3.9252   6.1923   5.5924   2.1023   1.6809
kurtosis(z)
ans = 4.4173   2.1295   2.3971   5.2642   2.2480   4.5196   3.3335
skewness(x)
ans = -0.4709   -0.3339   0.0379   -0.0426   -0.1830   -0.0380   0.4832
skewness(y)
ans = 2.0709   1.3180   1.1632   2.0742   1.8705   0.6844   0.3156
skewness(z)
ans = 1.4251   0.0349   0.1445   1.6084   0.6510   1.4653   1.2738
kurtosis(x(:))         ans = 2.5779
kurtosis(y(:))         ans = 7.3716
kurtosis(z(:))         ans = 5.5041
skewness(x(:))         ans = 0.0346
skewness(y(:))         ans = 2.0837
skewness(z(:))         ans = 1.4338
```

5. 相关系数

若随机变量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的方差 $D(\boldsymbol{X})$,$D(\boldsymbol{Y})$都存在,且均大于零,则称

$$\rho=\frac{\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y})}{\sqrt{D(\boldsymbol{X})}\sqrt{D(\boldsymbol{Y})}}$$

为 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的相关系数.相关系数是随机变量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 间线性相依程度的度量,它满足:$|\rho|\leqslant 1$.当 ρ 为正时,表明 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 间有相互促进作用,称为正相关;当 ρ 为负时,表明 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 间有相互抑制作用,称为负相关.$|\rho|$ 越接近 1,作用越强烈;$\rho=0$ 时,表明 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 间不存在线性依存关系.MATLAB 中用函数corrcoef(X)求矩阵 $\boldsymbol{X}$ 的相关系数矩阵,它把 $\boldsymbol{X}$ 的每一列作为一个随机变量,而把行作为样本点.例如

```
x = 1:10
x = 1   2   3   4   5   6   7   8   9   10
y = 2 * x + normrnd(0,1,1,10)
y = 2.1352   3.8610   4.8366   9.1837   9.9846   12.5362   13.2836
    15.3444   18.3144   20.1068
```

```
corrcoef(x,y)
ans  = 1.0000   0.9938
       0.9938   1.0000
x = exprnd(3,10,7)
x  = 1.5429  0.0014  12.6264  4.0044   1.8508   1.1626  7.7994
     0.1566  4.6537   3.7324  0.8485   1.4177   7.7981  4.9315
     3.7261  2.0890   0.6073  1.2482   1.1311   7.9492  2.9059
     0.3535  3.7086   0.0439  4.6245   1.1687  13.2860  3.8573
     6.8606  1.1891  12.1603  1.5219   0.3943   4.4464  0.7807
     8.1856  0.1288   0.5976  1.5079  13.0538   1.9835  3.4758
     4.3535  0.7976   1.4286  1.2488   3.5097   2.3413  1.3473
     0.2077  1.2188   1.7383  5.0889   0.7489   1.0563  0.0406
     8.2878  6.0989   4.2314  1.3551   3.5398   1.6214  2.0622
     3.9929  7.0486   0.5880  5.3104   0.2284   2.0247  0.1611
corrcoef(x)
ans  =   1.0000    0.0000    0.0911   - 0.5019  0.5558  - 0.4147
       - 0.2971    0.0000    1.0000   - 0.3249  0.1567  - 0.3508
         0.1883  - 0.2574    0.0911   - 0.3249  1.0000  - 0.0951
       - 0.2374  - 0.2722    0.3678   - 0.5019  0.1567  - 0.0951
         1.0000  - 0.3570  - 0.0480   - 0.1098  0.5558  - 0.3508
       - 0.2374  - 0.3570    1.0000   - 0.2693  0.1570  - 0.4147
         0.1883  - 0.2722  - 0.0480   - 0.2693  1.0000    0.2226
       - 0.2971  - 0.2574    0.3678   - 0.1098  0.1570    0.2226
         1.0000
x = normrnd(5,2,10,7)
x  = 9.7453  5.7762  5.5584  5.5370  7.6397  3.3214  3.4654
     5.4586  5.7861  7.7466  6.2500  3.1811  4.5827  4.7856
     4.4668  1.5853  5.3597  2.9053  0.3888  6.5118  3.0459
     6.4033  5.4557  3.9160  8.0713  8.5775  5.7515  3.0720
     4.0248  6.3713  8.2684  5.8689  5.7816  2.3092  0.2417
     8.7250  3.7264  6.6504  1.1657  5.0406  7.9638  3.3236
     7.2137  2.9948  5.4615  5.9399  4.1880  5.0655  5.5147
     2.5449  4.6288  6.3433  7.5487  1.9302  8.7409  4.6323
     3.6602  2.8919  3.9838  6.2771  5.4427  2.5820  4.6648
     7.6819  4.8569  6.7127  7.7616  2.2510  3.4347  4.7660
```

```
corrcoef(x)
ans =    1.0000    0.1586   -0.0358   -0.2682    0.4009   -0.0960
         0.1133    0.1586    1.0000    0.4810    0.4764    0.4984
        -0.3217   -0.3574   -0.0358    0.4810    1.0000   -0.1382
        -0.3160   -0.0590   -0.3193   -0.2682    0.4764   -0.1382
         1.0000    0.1747   -0.3234    0.2380    0.4009    0.4984
        -0.3160    0.1747    1.0000   -0.3525   -0.3183   -0.0960
        -0.3217   -0.0590   -0.3234   -0.3525    1.0000    0.2088
         0.1133   -0.3574   -0.3193    0.2380   -0.3183    0.2088
         1.0000
```

二、参数估计

参数估计是总体分布的数学形式已知,且可以用有限个参数表示的估计问题.参数估计一般可以分为点估计和区间估计.参数估计的方法一般有矩估计法,最小二乘估计法和似然函数估计法三种,而最常用的参数估计法是极大似然函数法.

所谓参数的点估计是指根据样本求出参数的估计值.而区间估计则是给定一可信程度度量 $1-\alpha$ 和一区间,使估计值以概率 $1-\alpha$ 的机会落入给定区间,从而求出该区间端点值的估计.MATLAB 提供了常用分布的参数估计函数,这些函数均采用极大似然函数法计算估计量

binofit(X,N,α)	二项分布的参数估计
expfit(X,α)	指数分布的参数估计
normfit(X,α)	正态分布的参数估计
poissfit(X,α)	泊松分布的参数估计
unifit(X,α)	均匀分布的参数估计

其中 X 为样本数据点,α 为纳伪概率,N 为二项分布的试验次数,这些函数均能同时返回参数的点估计和区间估计.例如对二项分布

```
x = binornd(10,0.6,1,10)
x = 8  9  6  6  7  8  2  5  7  4
[d i] = binofit(x,10)
d = 0.8000  0.9000  0.6000  0.6000  0.7000  0.8000  0.2000
    0.5000  0.7000  0.4000
i = 0.4439  0.5550  0.2624  0.2624  0.3475  0.4439  0.0252
    0.1871  0.3475  0.1216
    0.9748  0.9975  0.8784  0.8784  0.9333  0.9748  0.5561
```

```
    0.8129  0.9333  0.7376
[d i] = binofit(x, 10, 0.01)
d = 0.8000  0.9000  0.6000  0.6000  0.7000  0.8000  0.2000
    0.5000  0.7000  0.4000
i = 0.3518  0.4557  0.1909  0.1909  0.2649  0.3518  0.0109
    0.1283  0.2649  0.0768
    0.9891  0.9995  0.9232  0.9232  0.9630  0.9891  0.6482
    0.8717  0.9630  0.8091
```

其中输入参数中省略 α,则默认 $\alpha = 0.05$;输出参数为两个,上述输出结果中 d 为参数的点估计,i 为参数的区间估计,i 的第一行为区间左边界,第二行为区间右边界.

对泊松分布有

```
x = poissrnd(5, 10, 7)
x = 8  8  3   5  3  7  3
    6  4  9   7  5  1  4
    4  6  7   7  5  4  8
    6  7  7  10  4  4  5
    6  5  4   4  6  6  5
    4  6  8   6  4  5  3
    2  6  3   6  4  5  9
    2  6  7   5  7  7  6
    4  4  6  10  5  8  3
    9  1  2   0  8  3  5
[a b] = poissfit(x)
a = 5.1000  5.3000  5.6000  6.0000  5.1000  5.0000  5.1000
b = 3.8000  3.9000  4.2000  4.5000  3.8000  3.7000  3.8000
    6.5000  6.8000  7.1000  7.6000  6.5000  6.4000  6.5000
[a b] = poissfit(x, 0.01)
a = 5.1000  5.3000  5.6000  6.0000  5.1000  5.0000  5.1000
b = 3.4000  3.5000  3.8000  4.1000  3.4000  3.3000  3.4000
    7.0000  7.3000  7.6000  8.1000  7.0000  6.9000  7.0000
[a b] = poissfit(x(:), 0.05)
a = 5.3143                  b = [ 4.7743,  5.8543 ]
[a b] = poissfit(x(:), 0.01)
a = 5.3143                  b = [ 4.6046,  6.0240 ]
```

对于指数分布

```
x = exprnd(3 , 10 , 7)
x =0.3707   4.1268   1.7089   1.6059   7.0353   0.6755   0.5360
   2.3465   1.6090   1.4736   1.7472   1.3530   2.4970   5.7864
   0.6724   1.9405   6.8198   4.8179   2.4429   0.0571   5.2892
   6.0280   5.4343   5.5295   7.3110   8.1370   7.0315   8.0627
   8.1857   2.1622   2.6482   0.2073   2.9482   1.9189   0.5807
   2.9413   2.1032   1.7373   4.0484   4.1697   1.8174   6.0304
   2.9543   0.5117   3.9427   4.7664   0.2343   3.7727   0.3672
   2.1769   0.6463   0.7291   9.0351   1.3885   2.9763   1.9923
   0.0931   0.4602   2.8412   1.5018   0.3893   8.2143   0.1111
   3.2183   1.4841  10.4229   1.8135   1.3310   1.8220   6.3484
[a b] = expfit(x)
a = 2.8987   2.0478   3.7853   3.6854   2.9429   3.0783   3.5104
b = 1.3900   0.9820   1.8152   1.7673   1.4112   1.4762   1.6834
    4.9524   3.4987   6.4671   6.2965   5.0279   5.2592   5.9975
[a b] = expfit(x , 0.01)
a = 2.8987   2.0478   3.7853   3.6854   2.9429   3.0783   3.5104
b = 1.0774   0.7612   1.4070   1.3699   1.0939   1.1442   1.3048
    5.7970   4.0953   7.5700   7.3703   5.8854   6.1561   7.0203
[a b] = expfit(x(:), 0.05)
a = 3.1356               b = [ 2.4443 , 3.9116 ]
[a b] = expfit(x(:), 0.01)
a = 3.1356               b = [ 2.2544 , 4.1848 ]
```

默认 $\alpha=0.05$，a 为 λ 的点估计，b 为 λ 的区间估计.

对于正态分布

```
x = normrnd(4 , 2.5 , 10 , 7)
x =    2.9186   3.5332   4.7360   3.0003   -0.0102   1.4734    4.0001
      -0.1640   5.8145   0.6595   5.7250    4.6433   5.5362    3.2054
       4.3133   2.5292   5.7858   6.0391    1.3588   5.2694    6.7375
       4.7192   9.4580   8.0589   5.7798    7.5379   8.2311   -0.6850
       1.1338   3.6590   2.2706   7.2256    1.9873   5.4782    5.0705
       6.9773   4.2848   6.1450   5.6715    5.3219   2.3910    6.2391
       6.9729   6.6669   7.1350   6.9771    4.5483   4.9508    5.8274
       3.9059   4.1482   0.0157   0.9939    1.6952   1.4772    5.4446
```

```
     4.8182  3.7609  0.3976  3.9505  -1.4267  3.9512   4.1008
     4.4366  1.9191  5.4279  3.6082   3.8520  3.8794   5.6927
[a b c d] = normfit(x)
a = 4.0032  4.5774  4.0632  4.8971  2.9508  4.2638  4.5633
b = 2.2586  2.2073  2.9746  1.9581  2.7052  2.0982  2.1438
c = 2.3001  2.9130  1.8202  3.4206  0.9109  2.6817  2.9468
    5.7063  6.2418  6.3062  6.3736  4.9906  5.8459  6.1798
d = 1.5535  1.5182  2.0461  1.3468  1.8607  1.4432  1.4746
    4.1232  4.0296  5.4305  3.5747  4.9386  3.8304  3.9137
[a b c d] = normfit(x,0.01)
a = 4.0032  4.5774  4.0632  4.8971  2.9508  4.2638  4.5633
b = 2.2586  2.2073  2.9746  1.9581  2.7052  2.0982  2.1438
c = 1.5565  2.1863  0.8409  2.7759  0.0203  1.9909  2.2410
    6.4498  6.9685  7.2855  7.0183  5.8812  6.5367  6.8856
d = 1.3951  1.3634  1.8374  1.2095  1.6709  1.2960  1.3242
    5.1441  5.0273  6.7750  4.4598  6.1613  4.7788  4.8827
[a b c d] = normfit(x(:),0.05)
a = 4.1884                  b = 2.3299
c = [ 3.6288,4.7479 ]       d = [ 1.9977,2.7956 ]
[a b c d] = normfit(x(:),0.01)
a = 4.1884                  b = 2.3299
c = [ 3.4454,4.9314 ]       d = [ 1.9070,2.9689 ]
```

其中，α 的默认值也是 0.05，a 是 μ 的点估计，b 是 σ 的点估计，c 是 μ 的区间估计，d 是 σ 的区间估计. 另外 MATLAB 还提供了正态分布的对数似然函数等，有兴趣的读者可参阅有关参考书.

其它情况下的参数估计可通过编写估计函数来求得. 以两个正态总体均值差与方差比的估计为例，可编写如下函数

```
function [muhat,sigmahat,muci,sigmaci] = twonormfit(x,y,alpha)
%TWONORMFIT Parameter estimates and confidence intervals for two
%   normal data.
%   NORMFIT(X,Y,ALPHA) Returns estimates of the parameters of the
%   two normal distribution given the data in X-Y.
%
%   [MUHAT,SIGMAHAT,MUCI,SIGMACI] = TWONORMFIT(X,Y,ALPHA)
gives
```

```
%  MLEs and 100(1 - ALPHA) percent confidence intervals given the
%  data.
%  By default the optional paramter ALPHA = 0.05 corresponding
%  to 95% confidence intervals.

if nargin < 3
    alpha = 0.05;
end
[m1,n1] = size(x);[m2,n2] = size(y);
if min(m1,n1) == 1
  x = x(:);
  m1 = max(m1,n1);
  n1 = 1;
end
if min(m2,n2) == 1
  y = y(:);
  m2 = max(m2,n2);
  n2 = 1;
end
if n1~=n2
  error('x and y must be the same colume! ');
end
n=n1;
params1 = zeros(2,n);
params2 = zeros(2,n);
muhat1 = mean(x);
muhat2 = mean(y);
muhat=muhat1-muhat2;
sigmahat1 = std(x);
sigmahat2 = std(y);
sigmahat=sigmahat1./sigmahat2;
if nargout > 2,
  muci = zeros(2,n);
  sigmaci = zeros(2,n);
  tcrit = tinv([alpha/2 1-alpha/2],m1+m2-2);
```

```
    muci = [(muhat + tcrit(1) * sqrt((m1 - 1) * sigmahat1.^2 ...
      + (m2 - 1) * sigmahat2.^2) * sqrt((m1 + m2)/(m1 * m2 * (m1 + m2 -
2))));
        (muhat + tcrit(2) * sqrt((m1 - 1) * sigmahat1.^2 ...
      + (m2 - 1) * sigmahat2.^2) * sqrt((m1 + m2)/(m1 * m2 * (m1 + m2 -
2))))];
    fcrit = finv([alpha/2 1 - alpha/2], m1 - 1, m2 - 1);
    sigmaci = [(sigmahat./sqrt(fcrit(2)));
              (sigmahat./sqrt(fcrit(1)))];
end
```

举例如下

```
x = normrnd(3, 1.5, 10, 7)
x =   3.9529    2.0531   1.4728   3.7202    3.7282   4.4852   2.1882
      2.0979   -0.4878   2.7269   4.0022    2.9925   3.3283   0.9997
      3.8268    1.1525   5.2815   2.8825    2.5857   3.3925   4.6090
      1.3502    4.5835   2.9423   4.3338    4.9147   4.8202   1.9319
      3.1290    2.8302   4.8412   6.4639    5.7951   2.5880   2.9831
     -0.0068    3.5688   1.9557   3.7870    2.2162   2.8003   2.9988
      2.2604    4.4163   3.0113   2.9823    3.1551   1.0942   2.6258
      3.6931   -0.1806   1.8257   4.3697    1.7885   0.5046   3.5949
      2.5185    2.0330   3.8804   3.0839    4.0207   1.9447   2.6040
      4.8548    1.9435   2.6232   1.3394   -0.5469   3.4213   0.5040
y = normrnd(0, 1, 7, 7)
y = -1.0290   -0.4017    0.0714   -0.5803    0.6204   -0.0154
    -0.2751    0.2431    0.1737    0.3165    2.1363    1.2698
     0.5362    2.2126   -1.2566   -0.1161    0.4998   -0.2576
    -0.8960   -0.7164    1.5085   -0.3472    1.0641    1.2781
    -1.4095    0.1352   -0.6556   -1.9451   -0.9414   -0.2454
    -0.5478    1.7701   -0.1390    0.3144   -1.6805   -1.1746
    -1.5175    0.2608    0.3255   -1.1634    0.1068   -0.5735
    -1.0211    0.0097   -0.0132   -1.1190    1.1837    1.8482
    -0.1858
[a b c d] = twonormfit(x, y)
a = 3.5572   2.3388   2.7896   3.5729   2.9206   2.6353   2.6381
b = 2.6342   2.2510   2.2690   0.9615   1.8469   1.5866   0.7785
```

```
c = 2.3411  0.8407  1.6973  2.1620  1.3553  1.3823  1.2314
    4.7733  3.8370  3.8819  4.9837  4.4860  3.8884  4.0448
d = 1.1209  0.9578  0.9654  0.4091  0.7858  0.6751  0.3312
    5.4750  4.6785  4.7158  1.9985  3.8386  3.2977  1.6180
[a b c d] = twonormfit(x, y, 0.01)
a = 3.5572  2.3388  2.7896  3.5729  2.9206  2.6353  2.6381
b = 2.6342  2.2510  2.2690  0.9615  1.8469  1.5866  0.7785
c = 1.8759  0.2676  1.2795  1.6223  0.7565  0.9030  0.6933
    5.2385  4.4101  4.2997  5.5234  5.0847  4.3677  4.5828
d = 0.8172  0.6983  0.7039  0.2983  0.5729  0.4922  0.2415
    7.0359  6.0123  6.0602  2.5682  4.9329  4.2378  2.0793
x = normrnd(2, 1.2, 1, 100)
x = 2.0107   3.0043   1.1333   1.1342  1.7586   1.9754
    2.3347   3.2700   2.7460  -0.1007  2.8368   2.9738
    2.7636   3.5721   2.3925   1.1924  1.8208  -0.9388
    2.5679   2.1403   1.2907   1.2144  0.7032   1.9427
    2.4552   1.6036   1.4001   1.9568  1.7903   0.8513
    3.5511   2.5291   3.5371   1.4027  0.6575   2.9692
    2.0494   1.0925   1.8930  -0.4106  3.3007   0.8226
    1.1738   3.6074   0.9089   1.5046  1.3926   3.9437
    2.0971   0.7027   0.6506   4.0828  4.3250   3.9621
    0.4929   1.7438   1.7613   2.3690  1.3132   0.8268
    1.4638   3.2985   4.8472   2.2751  1.6801   2.8420
    1.4149   4.2350   3.3282   0.5269  1.1961   3.6091
    2.4657   2.4717  -0.0488   2.2734  2.8228   1.2359
    0.7969   1.7773   0.7352   1.9142  2.3350   3.6479
    2.2158   1.3496   3.9610   2.9903  2.2769   2.8060
    1.3903   3.0276   2.3222   2.7500  0.7432   3.8428
    2.5213  -0.3006   2.5639   3.5292
y = normrnd(1, 1.6, 1, 100)
y =   2.0217   3.2093   3.1117  -0.4551  -2.6890   3.8620
      1.6253   1.0325   0.3504  -1.4558   1.3542  -1.1992
     -0.3429   0.6662   2.2095   1.6012  -1.1527   3.3710
      1.0524   3.9927  -0.9344  -0.2522  -0.2277   0.8285
     -0.5633  -0.5424  -2.8067  -0.3411   1.4118   0.7059
```

```
 0.7318   0.8128   1.2696   0.1981  -0.1281   1.8131
 0.3265   1.3666  -0.5352   0.7663   2.1913  -0.4248
 1.2225   0.6222   0.8793   0.4263  -2.3242   0.7703
 3.2293   2.0429   0.3966  -0.0583   1.3983   0.3864
 0.1544   1.0886   3.0060  -3.0320   1.9358  -0.6129
 2.5109  -2.8783   0.6419   1.0929   0.3206   0.6753
-1.4209  -0.8022  -0.3040   1.5866   0.0622   3.4599
 1.2241  -1.9804   0.2733  -0.0433   1.1653   0.6470
 0.5535  -0.1739   0.8967  -1.3104   1.9797  -1.1176
-0.0585   0.7662   1.3969   0.8774   3.7811   3.5952
 2.0023   1.1469  -0.2922   0.2619  -1.2496   0.4008
 0.2465   3.8021   2.2052   1.1040
[a b c d]=twonormfit(x,y)
a = 1.4077                   b = 0.7558
c = [ 1.0286,1.7869 ]        d = [ 0.6200,0.9214 ]
[a b c d]=twonormfit(x,y,0.01)
a = 1.4077                   b = 0.7558
c = [ 0.9077,1.9078 ]        d = [ 0.5822,0.9812 ]
```

其中 α 默认为 0.05，输出结果中，a 为均值差 $\mu_1-\mu_2$ 的估计，b 为均方比 σ_1/σ_2 的估计，c 为 $\mu_1-\mu_2$ 的置信区间估计，d 为 σ_1/σ_2 的置信区间估计．其它各种参数估计的程序可类似编写．

三、假设检验

假设检验是统计推断的基本问题之一，是确定关于样本总体特征的判断是否合理的过程．假设检验就是利用样本提供给我们的信息，对所作的初始假设——零假设或称无效假设，按一定规则判定是否成立，以决定是接受还是拒绝零假设．假设检验的判断和结论是根据样本做出的，故具有“概率性”，所作的结论可能会犯两种错误：弃真和纳伪．一般情况下，伴随着无效假设 H_0 的确定，总能写出备择假设 H_A，备择假设也称对立假设，即拒绝 H_0 一定会接受 H_A，反之亦然．这种拒绝或者接受是在一定概率意义下做出的，因而一般必须预先给定显著性水平 α，α 即为犯错误的概率．假设检验的置信区间是指检验的假设量落入该区间就接受该假设，否则拒绝假设．

1. 正态总体参数的假设检验

对于正态总体来讲，一般可分为方差 σ^2 已知和未知两种情况．当方差已知时，即假设数据来自同一正态分布总体的独立样本，可用 Z 检验；而当方差未知

时,需要根据数据样本对方差作出估计,这时用估计出的方差来代替总体方差,可用 t 检验.这两种检验具有相关的变异度.

$$Z=\frac{\overline{X}-\mu}{\sigma},T=\frac{\overline{X}-\mu}{s},$$

其中 $\overline{X}=\sum_{i=1}^{n}\frac{X_i}{n}$.

若零假设为真,则 Z 服从标准正态分布,T 服从自由度为 $n-1$ 的 t 分布. MATLAB提供了如下检验函数

ttest(X,m,a,t)	单样本均值的 t 检验
ttest2(X,Y,a,t)	双样本均值差异的 t 检验
ztest(X,m,s,a,t)	已知方差的单样本 Z 检验

其中 X 和 Y 为正态总体的样本,m 为欲检验的均值,a 为显著水平,t 为备择假设选项,只有三个值 0,1 和 -1.其中 0 表示"$\neq$",1 表示">",-1 表示"<",缺省时为 $t=0$.

单样本均值的 t 检验主要用于判断在方差未知的情况下,给定的总体样本均值是否为某确定值,即 $H_0:\mu=\mu_0$,上述 m 即为这里的 μ_0.可用随机函数产生一个正态样本,用来模拟这一检验过程,模拟结果如下

```
x = normrnd(0,1,1,100)
x = -0.4326   -1.6656    0.1253    0.2877   -1.1465    1.1909
     1.1892   -0.0376    0.3273    0.1746   -0.1867    0.7258
    -0.5883    2.1832   -0.1364    0.1139    1.0668    0.0593
    -0.0956   -0.8323    0.2944   -1.3362    0.7143    1.6236
    -0.6918    0.8580    1.2540   -1.5937   -1.4410    0.5711
    -0.3999    0.6900    0.8156    0.7119    1.2902    0.6686
     1.1908   -1.2025   -0.0198   -0.1567   -1.6041    0.2573
    -1.0565    1.4151   -0.8051    0.5287    0.2193   -0.9219
    -2.1707   -0.0592   -1.0106    0.6145    0.5077    1.6924
     0.5913   -0.6436    0.3803   -1.0091   -0.0195   -0.0482
     0.0000   -0.3179    1.0950   -1.8740    0.4282    0.8956
     0.7310    0.5779    0.0403    0.6771    0.5689   -0.2556
    -0.3775   -0.2959   -1.4751   -0.2340    0.1184    0.3148
     1.4435   -0.3510    0.6232    0.7990    0.9409   -0.9921
     0.2120    0.2379   -1.0078   -0.7420    1.0823   -0.1315
     0.3899    0.0880   -0.6355   -0.5596    0.4437   -0.9499
     0.7812    0.5690   -0.8217   -0.2656
```

```
mean(x)                  ans = 0.0479
var(x)                   ans = 0.7543
[a b c] = ttest(x, 0, 0.05, 0)
a = 0           b = 0.5823           c = [ -0.1244, 0.2203 ]
[a b c] = ttest(x, 0, 0.01, 0)
a = 0           b = 0.5823           c = [ -0.1802, 0.2760 ]
[a b c] = ttest(x, 0, 0.05, -1)
a = 0           b = 0.7089           c = [ -0.0963, 0.1921 ]
[a b c] = ttest(x, 0, 0.01, -1)
a = 0           b = 0.7089           c = [ -0.1574, 0.2533 ]
[a b c] = ttest(x, 0, 0.05, 1)
a = 0           b = 0.2911           c = [ -0.0963, 0.1921 ]
[a b c] = ttest(x, 0, 0.01, 1)
a = 0           b = 0.2911           c = [ -0.1574, 0.2533 ]
[a b c] = ttest(x, 0.7, 0.05, 0)
a = 1           b = 2.6743e-011      c = [ -0.1244, 0.2203 ]
[a b c] = ttest(x, 0.7, 0.01, 0)
a = 1           b = 2.6743e-011      c = [ -0.1802, 0.2760 ]
[a b c] = ttest(x,0.7, 0.05, -1)
a = 1           b = 1.3372e-011      c = [ -0.0963, 0.1921 ]
[a b c] = ttest(x, 0.7, 0.01, -1)
a = 1           b = 1.3372e-011      c = [ -0.1574, 0.2533 ]
[a b c] = ttest(x, 0.7, 0.05, 1)
a = 0           b = 1.0000           c = [ -0.0963, 0.1921 ]
[a b c] = ttest(x, 0.7, 0.01, 1)
a = 0           b = 1.0000           c = [ -0.1574, 0.2533 ]
```

比较上述理论均值与方差和样本的均值与方差之间的关系,然后看检验结果,即前面分别检验样本均值与理论值是否存在差异,后面检验样本均值与 0.7 是否有显著差异,分别按两个显著水平 $\alpha = 0.05$, $\alpha = 0.01$ 和三种假设方式进行.通过比较可得出什么结论呢?注意上述结果中三个输出项,其中 a 为 0 表示接受 H_0,拒绝 H_A;a 为 1 表示拒绝 H_0,接受 H_A;b 与 T 统计量有关,它表示假设 X 的均值等于 μ 时,T 的观察值较大的概率;c 是样本均值的 $1-\alpha$ 置信区间.

对于两个正态总体样本均值的差异性检验,可用下述过程模拟

```
x = normrnd(2, 2, 1, 100)
```

```
x =   0.2079    2.2704    1.7219   -0.3268    4.3674    1.9691
      3.0724    0.5671    0.6889    2.6287    2.2136    5.6964
      1.4498    6.4251    5.0171   -1.8902   -1.3611    0.8529
      1.6284    2.0179    3.6739    0.5555    0.5570    1.5976
      1.9591    2.5578    4.1166    3.2433   -1.5012    3.3947
      3.6230    3.2727    4.6202    2.6542    0.6540    1.7013
     -2.8980    2.9466    2.2339    0.8178    0.6906   -0.1613
      1.9045    2.7587    1.3393    1.0002    1.9280    1.6505
      0.0855    4.5851    2.8818    4.5619    1.0045   -0.2374
      3.6153    2.0824    0.4876    1.8217   -2.0177    4.1678
      0.0376    0.6230    4.6790    0.1815    1.1743    0.9877
      5.2395    2.1618   -0.1621   -0.2490    5.4714    5.8749
      5.2701   -0.5119    1.5729    1.6021    2.6150    0.8553
      0.0447    1.1064    4.1642    6.7453    2.4586    1.4668
      3.4033    1.0248    5.7250    4.2137   -0.4551    0.6602
      4.6819    2.7762    2.7861   -1.4147    2.4557    3.3713
      0.7264   -0.0052    1.6288   -0.1081
y = normrnd(4, 2, 1, 100)
y = 3.8569  4.5584    6.7466    4.3597    2.9160    7.2684
    5.6504  4.4615    5.3433    2.9838    5.7127    4.5370
    5.2500  1.9053    7.0713    4.8689    0.1657    4.9399
    6.5487  5.2771    6.7616    6.6397    2.1811   -0.6112
    7.5775  4.7816    4.0406    3.1880    0.9302    4.4427
    1.2510  2.3214    3.5827    5.5118    4.7515    1.3092
    6.9638  4.0655    7.7409    1.5820    2.4347    2.4654
    3.7856  2.0459    2.0720   -0.7583    2.3236    4.5147
    3.6323  3.6648    3.7660    4.3370    2.9976    2.5898
    5.0163  3.1582    4.4583    2.0810    3.7079    5.4891
    2.2190  4.2781    3.5277    3.8491    3.2829   -0.1553
    3.7129  6.7867    5.3036    3.2457    2.6771    4.4979
    3.2330  2.9430    4.1108    6.5075   -1.0400    5.1697
    1.9839  5.8886   -0.8479    3.5523    4.1161    3.1508
    3.5942  0.9738    1.7473    2.3700    4.7332    2.8278
    7.0748  4.2801    0.2745    3.0916    2.6959    4.2066
    3.5587  3.4419    2.5327    3.8709
```

```
mean(x)                 ans = 1.9803
mean(y)                 ans = 3.7048
var(x)                  ans = 3.9819
var(y)                  ans = 3.7316
[a b c] = ttest2(x, y, 0.01, 0)
a = 1      b = 3.0657e-009      c = [ -2.4469, -1.0022 ]
[a b c] = ttest2(x, y, 0.05, 0)
a = 1      b = 3.0657e-009      c = [ -2.2723, -1.1769 ]
[a b c] = ttest2(x, y, 0.05, -1)
a = 1      b = 1.5329e-009      c = [ -2.2723, -1.1769 ]
[a b c] = ttest2(x, y, 0.01, -1)
a = 1      b = 1.5329e-009      c = [ -2.4469, -1.0022 ]
[a b c] = ttest2(x, y, 0.05, 1)
a = 0      b = 1.0000           c = [ -2.2723, -1.1769 ]
[a b c] = ttest2(x, y, 0.01, 1)
a = 0      b = 1.0000           c = [ -2.4469, -1.0022 ]
```

上述 a 与 b 的意义同单样本检验，c 表示两样本均值差即 $\overline{X}-\overline{Y}$ 的置信区间，也分两个显著水平 $\alpha=0.05$ 和 $\alpha=0.01$ 及三种假设检验．通过检验结果并结合理论均值与方差和样本均值与方差的比较，请再模拟方差不同质时两个理论均值比较接近的随机样本的检验．

当方差已知时，要检验来自正态总体样本均值与某一数值有无显著差异，模拟如下

```
x = normrnd(4, 2, 1, 100)
x = 1.1120  5.2247  1.3530  2.6768  3.7078   4.4962  3.8467
    7.4763  7.2439  5.2529  4.1836  2.3848   3.0773  1.1881
    3.2509  3.0582  7.5026  5.5064  4.1300   3.4145  4.1656
    5.5324  8.4737  4.6538  5.7266  5.3588   5.1095  6.0033
    6.5187  4.0883  3.3717  4.4534  5.9934   6.4318  2.9146
    5.8245  3.6557  3.3281  5.0830  5.8642   2.8595  1.0028
    3.8993  5.1060  4.1670  7.1550  3.3385   5.5903  2.4304
    1.4738  5.3333  1.2147  1.3989  2.7900   1.0229  5.1171
    3.4453  1.4126  2.2231  2.0270  3.8568  -0.8292  2.6113
    1.2172  4.6593  5.1971  4.2944  3.7971  -1.2700  4.0561
    2.2474  3.4690  3.3448  1.6835  5.1601   4.4795  3.2982
    5.7842  7.1566  1.7837  3.9481  1.7787   5.5017  5.0003
```

```
    2.9655  2.8816  2.4933  5.8516  3.5030   3.7003  1.4832
    4.6252  9.3806  4.5794  1.1544  4.4936   1.1285  4.2971
    0.6139  5.4384
mean(x)              ans = 3.8646
var(x)               ans = 3.8239
[a b c] = ztest(x,4,2,0.05,0)
a = 0          b = 0.4984         c = [ 3.4726,4.2566 ]
[a b c] = ztest(x,4.5,2,0.05,0)
a = 1          b = 0.0015         c = [ 3.4726,4.2566 ]
[a b c] = ztest(x,4,2,0.01,0)
a = 0          b = 0.4984         c = [ 3.3494,4.3798 ]
[a b c] = ztest(x,4.5,2,0.01,0)
a = 1          b = 0.0015         c = [ 3.3494,4.3798 ]
[a b c] = ztest(x,4,2,0.05,-1)
a = 0          b = 0.2492         c = [ 3.4726,4.2566 ]
[a b c] = ztest(x,4.5,2,0.05,-1)
a = 1          b = 7.4413e-004    c = [ 3.4726,4.2566 ]
[a b c] = ztest(x,4,2,0.01,-1)
a = 0          b = 0.2492         c = [ 3.3494,4.3798 ]
[a b c] = ztest(x,4.5,2,0.01,-1)
a = 1          b = 7.4413e-004    c = [ 3.3494,4.3798 ]
[a b c] = ztest(x,4,2,0.05,1)
a = 0          b = 0.7508         c = [ 3.4726,4.2566 ]
[a b c] = ztest(x,4.5,2,0.05,1)
a = 0          b = 0.9993         c = [ 3.4726,4.2566 ]
[a b c] = ztest(x,4,2,0.01,1)
a = 0          b = 0.7508         c = [ 3.3494,4.3798 ]
[a b c] = ztest(x,4.5,2,0.01,1)
a = 0          b = 0.9993         c = [ 3.3494,4.3798 ]
[a b c] = ztest(x,4.5,2,0.01,1)
a = 0          b = 0.9993         c = [ 3.3494,4.3798 ]
```

上述 a,b,c 的意义同单样本 t 检验.以上分别检验样本均值与理论均值是否存在显著差异及样本均值与 4.5 是否存在显著差异.

MATLAB 没有提供检验正态总体的某样本方差是否为某一值的函数.但我们可以充分利用 M 文件功能编写检验函数.下面的函数及其运行结果可供参

考.对于其它要检验的情形,可编写出类似的函数.

```
function [h,sig,ci] = chi2test(x,s,alpha,tail)
% CHI2TEST Hypothesis test: Compares the sample variance
%           to a constant.
%   [H,SIG] = CHI2TEST(X,S,ALPHA,TAIL) performs a
%           CHI2 - test to determine
%   if a sample from a normal distribution (in X) could
%   have variance S.
%   S = 1, ALPHA = 0.05 and TAIL = 0 by default.
%   The Null hypothesis is: "variance is equal to M".
%   For TAIL = 0, alternative: "variance is not M".
%   For TAIL = 1, alternative: "variance is greater than M"
%   For TAIL = -1, alternative: "variance is less than M"
%   TAIL = 0 by default.
%   ALPHA is desired significance level.
%   SIG is the probability of observing the given result
%   by chance given that the null hypothesis is true. Small
%   values of SIG cast doubt on the validity of the null
%   hypothesis.
%   H = 0 => "Do not reject null hypothesis at significance
%   level of alpha."
%   H = 1 => "Reject null hypothesis at significance level
%   of alpha."
if nargin < 1,
    error('Requires at least one input argument.');
end
[m1 n1] = size(x);
if (m1 ~= 1 & n1 ~= 1)
    error('Requires a vector first input argument.');
end
if nargin < 2
    s = 1;
end
  if nargin < 4,
    tail = 0;
```

```
    end
    if nargin < 3,
        alpha = 0.05;
    end
    if (alpha <= 0 | alpha >= 1)
        fprintf('Warning: significance level must be between 0 and 1 \n
');
        h = NaN;
        sig = NaN;
        ci = [NaN NaN];
        return;
    end
    samplesize = length(x);
    ser = var(x);
    chi2val = (samplesize - 1) * ser./s^2;
    sig = chi2cdf(chi2val, samplesize - 1);
    % the significance just found is for the tail = -1 test
    crit1 = sqrt((samplesize-1) * ser./chi2inv((1-alpha), samplesize
- 1));
    crit2 = sqrt((samplesize - 1) * ser./chi2inv(alpha, samplesize
- 1));
    if tail == 1
        sig = 1 - sig;
    elseif tail == 0
        sig = 2 * min(sig, 1 - sig);
        crit1 = sqrt((samplesize-1) * ser./chi2inv((1-alpha/2),...
            samplesize - 1));
        crit2 = sqrt((samplesize-1) * ser./chi2inv(alpha/2, ...
            samplesize - 1));
    end
    ci = [crit1 crit2];
    % Determine if the actual significance exceeds the
    % desired significance
    h = 0;
    if sig <= alpha,
```

```
    h = 1;
end
if isnan(sig),
    h = NaN;
end
```

例如

```
x = normrnd(0,1,1,100)
x = -0.4326  -1.6656   0.1253   0.2877  -1.1465   1.1909
     1.1892  -0.0376   0.3273   0.1746  -0.1867   0.7258
    -0.5883   2.1832  -0.1364   0.1139   1.0668   0.0593
    -0.0956  -0.8323   0.2944  -1.3362   0.7143   1.6236
    -0.6918   0.8580   1.2540  -1.5937  -1.4410   0.5711
    -0.3999   0.6900   0.8156   0.7119   1.2902   0.6686
     1.1908  -1.2025  -0.0198  -0.1567  -1.6041   0.2573
    -1.0565   1.4151  -0.8051   0.5287   0.2193  -0.9219
    -2.1707  -0.0592  -1.0106   0.6145   0.5077   1.6924
     0.5913  -0.6436   0.3803  -1.0091  -0.0195  -0.0482
     0.0000  -0.3179   1.0950  -1.8740   0.4282   0.8956
     0.7310   0.5779   0.0403   0.6771   0.5689  -0.2556
    -0.3775  -0.2959  -1.4751  -0.2340   0.1184   0.3148
     1.4435  -0.3510   0.6232   0.7990   0.9409  -0.9921
     0.2120   0.2379  -1.0078  -0.7420   1.0823  -0.1315
     0.3899   0.0880  -0.6355  -0.5596   0.4437  -0.9499
     0.7812   0.5690  -0.8217  -0.2656
y = normrnd(1.5,2,1,100)
y = -0.8756  -2.9046   3.4727   0.4627   2.1547   1.9681
     1.5429  -0.5079  -0.3943   0.7511  -0.8718  -0.6118
     4.4450   1.6115  -0.9346   1.4175  -0.7567  -1.1986
     0.9778   3.4069   1.7573   2.8129  -0.8356   0.5788
     0.9751  -0.9263  -1.1389   3.3624   1.5225   0.2097
     3.1115   1.9633  -0.4795   4.1792   2.0790   4.4578
     3.7761   0.1317  -1.0839   1.3541   0.8388  -0.1873
     2.4955   4.4770   0.4070  -0.1935   1.0073   2.8260
    -0.2084  -0.9026   1.2603   1.3694   2.4706   0.3090
     1.2007   0.6305   1.3413   4.5703   0.2870  -1.1947
```

```
    2.4388   -0.3071    1.5718    0.2449    2.5708    2.6058
    1.0926   -2.6086    1.7651    4.6859    3.5368   -1.6608
    1.3427    0.1367   -0.5491   -0.9687    2.0776    0.6414
    1.6116    0.7643    0.5701    2.2419    2.9566    5.7243
   -1.2146   -0.5452    3.5757    0.7204   -1.2625    2.1311
    4.6065    2.9158    5.4148    2.5091    5.2291    0.8204
   -0.7796    1.0778    3.8805   -0.7324
[a b c]=chi2test(x,1,0.05,0)
a = 0          b = 0.0648          c = [ 0.7626,1.0089 ]
[a b c]=chi2test(x,1,0.01,0)
a = 0          b = 0.0648          c = [ 0.7330,1.0596 ]
[a b c]=chi2test(x,1,0.05,-1)
a = 1          b = 0.0324          c = [ 0.7785,0.9845 ]
[a b c]=chi2test(x,1,0.01,-1)
a = 0          b = 0.0324          c = [ 0.7447,1.0386 ]
[a b c]=chi2test(x,1,0.05,1)
a = 0          b = 0.9676          c = [ 0.7785,0.9845 ]
[a b c]=chi2test(x,1,0.01,1)
a = 0          b = 0.9676          c = [ 0.7447,1.0386 ]
[a b c]=chi2test(y,2,0.05,0)
a = 0          b = 0.4606          c = [ 1.6589,2.1949 ]
[a b c]=chi2test(y,2,0.01,0)
a = 0          b = 0.4606          c = [ 1.5946,2.3051 ]
[a b c]=chi2test(y,2,0.05,-1)
a = 0          b = 0.2303          c = [ 1.6935,2.1417 ]
[a b c]=chi2test(y,2,0.01,-1)
a = 0          b = 0.2303          c = [ 1.6201,2.2594 ]
[a b c]=chi2test(y,2,0.05,1)
a = 0          b = 0.7697          c = [ 1.6935,2.1417 ]
[a b c]=chi2test(y,2,0.01,1)
a = 0          b = 0.7697          c = [ 1.6201,2.2594 ]
[a b c]=chi2test(y,1.5,0.05,0)
a = 1          b = 3.6643e-004     c = [ 1.6589,2.1949 ]
[a b c]=chi2test(y,1.5,0.05,-1)
a = 0          b = 0.9998          c = [ 1.6935,2.1417 ]
```

```
[a b c] = chi2test(y,1.5,0.05,1)
a = 1            b = 1.8322e-004     c = [ 1.6935, 2.1417 ]
[a b c] = chi2test(x,0.5,0.05,0)
a = 1            b = 0               c = [ 0.7626, 1.0089 ]
[a b c] = chi2test(x,0.5,0.05,-1)
a = 0            b = 1               c = [ 0.7785, 0.9845 ]
[a b c] = chi2test(x,0.5,0.05,1)
a = 1            b = 0               c = [ 0.7785, 0.9845 ]
```

以上输出结果中，a 为 0 表示接受 H_0，拒绝 H_A；a 为 1 表示拒绝 H_0，接受 H_A；b 为接受 H_0 时，统计量较大的概率；c 为 σ 的置信区间.

2. 非参数假设检验

参数检验方法一般用于服从正态分布或近似服从正态分布的总体. 如果对总体服从何种分布知之甚少或完全不知，则需要用到所谓的非参数检验方法. MATLAB提供有三个非参数检验函数，它们是

ranksum(X, Y, α)　　威尔科克秩和检验

signrank(X, Y, α)　　威尔科克符号秩检验

signtest(X, Y, α)　　成对样本的符号检验

其中 X 和 Y 为两个总体样本，α 为指定的显著性概率水平.

秩和检验法的目的是检验两个总体 X 和 Y 的分布是否相同. 调用格式为

$$[p, h] = \text{ranksum}(x, y, \alpha)$$

返回参数中，p 为总体 X 和 Y 分布相同的概率. h 为 0 表示接受分布相同的假设；h 为 1 表示拒绝分布相同的假设，即 X 与 Y 的分布有明显不同. 例如

```
x = normrnd(0,1,1,80)
x = -0.4326  -1.6656   0.1253   0.2877  -1.1465   1.1909
     1.1892  -0.0376   0.3273   0.1746  -0.1867   0.7258
    -0.5883   2.1832  -0.1364   0.1139   1.0668   0.0593
    -0.0956  -0.8323   0.2944  -1.3362   0.7143   1.6236
    -0.6918   0.8580   1.2540  -1.5937  -1.4410   0.5711
    -0.3999   0.6900   0.8156   0.7119   1.2902   0.6686
     1.1908  -1.2025  -0.0198  -0.1567  -1.6041   0.2573
    -1.0565   1.4151  -0.8051   0.5287   0.2193  -0.9219
    -2.1707  -0.0592  -1.0106   0.6145   0.5077   1.6924
     0.5913  -0.6436   0.3803  -1.0091  -0.0195  -0.0482
     0.0000  -0.3179   1.0950  -1.8740   0.4282   0.8956
     0.7310   0.5779   0.0403   0.6771   0.5689  -0.2556
```

```
   -0.3775   -0.2959   -1.4751   -0.2340   0.1184   0.3148
    1.4435   -0.3510
y = normrnd(0.5, 1, 1, 100)
y =  1.1232    1.2990    1.4409   -0.4921    0.7120    0.7379
    -0.5078   -0.2420    1.5823    0.3685    0.8899    0.5880
    -0.1355   -0.0596    0.9437   -0.4499    1.2812    1.0690
    -0.3217    0.2344   -0.6878   -1.7023    1.4863   -0.0186
     0.8274    0.7341    0.5215   -0.5039   -0.4471    0.1256
    -0.6859   -0.5559    1.9725    0.5557   -0.7173    0.4588
    -0.6283   -0.8493    0.2389    1.4535    0.6286    1.1565
    -0.6678    0.0394    0.2376   -0.7132   -0.8194    1.4312
     0.5112   -0.1451    1.3057    0.7316   -0.4898    1.8396
     0.7895    1.9789    1.6380   -0.1841   -0.7919    0.4271
     0.1694   -0.3436    0.9978    1.9885   -0.0465   -0.3468
     0.2537    1.1630   -0.3542   -0.7013    0.3801    0.4347
     0.9853   -0.0955    0.3503    0.0652    0.4207    2.0352
    -0.1065   -0.8474    0.9694   -0.4036    0.5359   -0.1275
     1.0354    1.0529    0.2963   -1.5543    0.6326    2.0929
     1.5184   -1.0804    0.4213   -0.1817   -0.5246   -0.7344
     0.7888    0.0707    0.5558    0.1321
z = normrnd(1, 2, 1, 100)
z =  0.0701    1.7419    2.4566    5.2243   -1.7146   -1.0452
     3.0757    0.2204   -1.7625    1.6311    4.1065    2.4158
     4.9148    2.0091    4.7291    0.3204   -1.2796    0.5778
     3.3805   -1.2324    2.2705   -0.2028    2.1024   -1.1997
     1.1720   -3.0091    0.0138    1.9241    0.3580    3.4731
    -0.2626   -3.6504   -1.4633    3.1113    0.7736    1.7584
     2.8884   -3.2409   -0.2894   -0.4086   -1.0363    0.6358
     4.0420    0.9231    3.4549   -0.3924    1.0150   -0.5658
     2.1739    0.4976    1.9603    2.3363    0.8434    2.7783
     5.6186    2.0493    0.9764    2.8263    1.1119   -1.2141
     1.9710    0.9900    0.4476    3.5529    4.7268   -0.0451
     1.2068   -0.6153    2.3609   -3.7292    2.9802    1.4378
     1.5233    3.4269    0.4507    0.7337   -1.5410   -2.3272
    -0.4071    1.5618   -0.0824   -1.6671    3.1454   -0.4242
```

```
   0.9774  0.9984   0.5011  1.7932   0.4720  -2.3280
  -1.0580  1.4862  -1.5132  0.3056  -0.8827  -1.3491
  -1.0423  0.1967   1.3473  0.7678
[p,h]=ranksum(x,y,0.05)
p = 0.1167                    h = 0
[p,h]=ranksum(x,y,0.01)
p = 0.1167                    h = 0
[p,h]=ranksum(x,z,0.05)
p = 0.0015                    h = 1
[p,h]=ranksum(x,z,0.01)
p = 0.0015                    h = 1
[p,h]=ranksum(y,z,0.05)
p = 0.0207                    h = 1
[p,h]=ranksum(y,z,0.01)
p = 0.0207                    h = 0
```

由上述结论知,X 和 Y 属同分布,X 和 Z 不属同分布,Y 和 Z 在 0.05 的水平上不属同分布,但在 0.01 水平上可认为属同分布.这里必须注意,调用函数 ranksum( )时,X 的样本容量必须小于等于 Y 的样本容量,一般容量应当相等.

符号秩检验法的目的是检验两个配对样本 X 和 Y 的中位数是否相等的假设,这里 X 和 Y 的样本容量必须相同.符号秩检验函数的调用格式为

$$[p,h]=\text{signrank}(X,Y,a)$$

返回参数中,p 为中位数相等的概率,$h=0$ 表示接受中位数相等的假设,$h=1$ 表示拒绝中位数相等的假设.例如

```
x=normrnd(0,1,1,20)
x =1.0641  -0.2454  -1.5175  0.0097   0.0714   0.3165
   0.4998   1.2781  -0.5478  0.2608  -0.0132  -0.5803
   2.1363  -0.2576  -1.4095  1.7701   0.3255  -1.1190
   0.6204   1.2698
y=normrnd(0.5,1,1,20)
y = -0.3960   0.6352   0.3610  -0.6634   1.6837   0.4846
     1.0362  -0.2164  -0.1556   0.8144   0.6068   2.3482
     0.2249   2.7126   2.0085  -1.4451  -1.1805  -0.0735
     0.3142   0.5089
z=normrnd(0,4,1,20)
```

```
z = 3.3478   -2.8891   -2.8860   -0.8047   -0.0819   1.1156
    4.2332    2.4867   -7.0025    2.7894    3.2459   2.5454
    5.2403    1.3084   -2.6920   -0.5973   -9.7961   1.8931
    0.4678   -2.3644
u = normrnd(2,1,1,20)
u = 1.3453   0.9193   1.9523   2.3793   1.6696   1.5001   1.9640
    1.8252   1.0427   3.2925   2.4409   3.2809   1.5023   0.8813
    2.8076   2.0412   1.2438   1.9109   -0.0089  3.0839
[p h] = signrank(x,y,0.05)
p = 0.2568                    h = 0
[p h] = signrank(x,y,0.01)
p = 0.2568                    h = 0
[p h] = signrank(x,z,0.05)
p = 0.1523                    h = 0
[p h] = signrank(x,z,0.01)
p = 0.1523                    h = 0
[p h] = signrank(x,u,0.05)
p = 1.5752e-004               h = 1
[p h] = signrank(x,u,0.01)
p = 1.5752e-004               h = 1
[p h] = signrank(y,z,0.05)
p = 0.1613                    h = 0
[p h] = signrank(y,z,0.01)
p = 0.1613                    h = 0
[p h] = signrank(y,u,0.05)
p = 4.1697e-004               h = 1
[p h] = signrank(y,u,0.01)
p = 4.1697e-004               h = 1
[p h] = signrank(z,u,0.05)
p = 0.0323                    h = 1
[p h] = signrank(z,u,0.01)
p = 0.0323                    h = 0
```

成对样本的符号检验法也是用以检验两个总体样本中位数相等的假设.若 X 和 Y 均为样本,则其样本数必须相等.以上例中的 x,y,z 和 u 为例进行符号检验,结果为

```
[p h] = signtest(x, y, 0.05)
p = 0.5034                    h = 0
[p h] = signtest(x, y, 0.01)
p = 0.5034                    h = 0
[p h] = signtest(x, z, 0.05)
p = 1.1762                    h = 0
[p h] = signtest(x, z, 0.01)
p = 1.1762                    h = 0
[p h] = signtest(x, u, 0.05)
p = 4.0245e-004               h = 1
[p h] = signtest(x, u, 0.01)
p = 4.0245e-004               h = 1
[p h] = signtest(y, z, 0.05)
p = 0.8238                    h = 0
[p h] = signtest(y, z, 0.01)
p = 0.8238                    h = 0
[p h] = signtest(y, u, 0.05)
p = 0.0026                    h = 1
[p h] = signtest(y, u, 0.01)
p = 0.0026                    h = 1
[p h] = signtest(z, u, 0.05)
p = 0.2632                    h = 0
[p h] = signtest(z, u, 0.01)
p = 0.2632                    h = 0
```

比较上述两个结果,只有 z 和 u 在 0.05 水平下检验结果不同,符号秩检验法拒绝接受中位数相等的假设,而成对样本符号检验法接受了中位数相等的假设.事实上 z 样本中位数为 0.791 7,而 u 样本中位数为 1.868 1,可见成对样本符号检验法比威尔科克符号秩检验法接受无效假设 H_0 的条件要宽松些.除此之外,还可用成对样本符号检验法检验总体样本中位数与某一常数是否有差异的假设,只需将上述 y 变为一个常数即可.例如

```
[p h] = signtest(x, 0, 0.05)
p = 0.5034                    h = 0
[p h] = signtest(x, 1, 0.05)
p = 0.0414                    h = 1
[p h] = signtest(x, 0, 0.01)
```

```
p = 0.5034                     h = 0
[p h] = signtest(x , 1 , 0.01)
p = 0.0414                     h = 0
[p h] = signtest(y , 0 , 0.05)
p = 0.2632                     h = 0
[p h] = signtest(y , 0 , 0.01)
p = 0.2632                     h = 0
[p h] = signtest(y , 1 , 0.01)
p = 0.0414                     h = 0
[p h] = signtest(y , 1 , 0.05)
p = 0.0414                     h = 1
[p h] = signtest(z , 0 , 0.05)
p = 0.8238                     h = 0
[p h] = signtest(z , 0 , 0.01)
p = 0.8238                     h = 0
[p h] = signtest(z , 1 , 0.05)
p = 1.1762                     h = 0
[p h] = signtest(z , 1 , 0.01)
p = 1.1762                     h = 0
[p h] = signtest(z , 2 , 0.01)
p = 0.2632                     h = 0
[p h] = signtest(z , 2 , 0.05)
p = 0.2632                     h = 0
[p h] = signtest(z , 3 , 0.05)
p = 0.0118                     h = 1
[p h] = signtest(z , 3 , 0.01)
p = 0.0118                     h = 0
[p h] = signtest(u , 1 , 0.05)
p = 0.0026                     h = 1
[p h] = signtest(u , 1 , 0.01)
p = 0.0026                     h = 1
[p h] = signtest(u , 2 , 0.05)
p = 0.2632                     h = 0
[p h] = signtest(u , 2 , 0.01)
p = 0.2632                     h = 0
```

```
[p h] = signtest(u, 3, 0.05)
p = 0.0026                    h = 1
[p h] = signtest(u, 3, 0.01)
p = 0.0026                    h = 1
```

其它符号检验法如分布的检验,游程检验等可参阅有关文献编写检验函数.

第四节　方差分析与回归分析

方差分析与线性回归分析同属线性模型,是数理统计的重要组成部分.

一、方差分析

方差分析法因其基于统计数据的总变动成分与随机误差成分的比较而得名,是基于不太多的统计数据,定量地分析一个或多个因素对某个(些)响应变量的影响和作用的显著性,这种显著性是基于一定概率条件下而言的,其前提条件是在各因素的作用下,响应变量的分布具有正态性和等方差性.

1. 单因素方差分析

单因素方差分析的基本问题是比较和估计多个等方差正态总体的均值,其基本模型为

$$y_{ij} = \alpha_j + \varepsilon_{ij},$$

其中 y_{ij} 为数据观测值,α_j 是各正态总体的数学期望,ε_{ij} 为各数据的随机误差.在MATLAB中提供了单因素方差分析函数 anoval( ),其使用格式为

$$p = \text{anoval}(X)$$

或

$$p = \text{anoval}(X, g)$$

该函数返回无效假设成立的概率,第一种格式中 $\boldsymbol{X}$ 为一矩阵,函数将矩阵的每一列当作一个总体,矩阵的行数即为样本重复数.若函数返回的概率值接近于零,则无效假设值得怀疑,表明各列的均值事实上是不同的.第二种格式中的 $\boldsymbol{X}$ 为一向量,$\boldsymbol{g}$ 是与 $\boldsymbol{X}$ 同长度的向量,且 $\boldsymbol{g}$ 中的值为整数,最小值为 1,最大值为数据的组数,每一组至少有一个数,但并不要求每组中元素个数相同.因此,第一种格式用于等重复的单因素方差分析,第二种格式用于重复数不等的单因素方差分析.方差分析同时还显示一个图与一张表,表即为方差分析表,它与一般教科书所列方差分析表一般无二;而图则给出了各列数据的 box 图,这种图为一"盒子"形状,因而取名 box 图,其特征为:① 盒子的上底与下底间为内四分位间距;盒子的上、下两条线分别为样本的 25% 和 75% 分位数.② 盒子的中间线为样本的中位数,如果中位数不在盒子中间,表明样本存在一定偏度.③ 虚线贯穿盒子上下,显示了样本其余部分,如果没有奇异值,则样本的最大值为虚线顶点,