Matlab在统计方面的应用

目录

- 随机数的产生
- 随机变量的概率密度计算
- 随机变量的累积概率值(分布函数值)
- 随机变量的逆累积分布函数
- 随机变量的数字特征
- 统计作图
- 参数估计
- 假设检验
- 方差分析

几个常见的分布——二项分布

设随机变量X的分布律为：

$$P\left(x=k\right)=b(k,n,p)=\begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k}, k=1,\cdots,n$$

其中：0< p <1，

n为独立重复试验的总次数，

k为n次重复试验中事件A发生的次数，

p为每次试验事件A发生的概率。

则称X服从二项分布，记为X~B (n, p)。

几个常见的分布——泊松分布

设随机变量X的分布律为：

$$P\{x=k\}=\frac{\lambda^k}{k!}e^{-\lambda} \qquad k=0,1,2,\cdots;\lambda>0$$

则称X服从参数为的Poisson分布，记为

X~P(λ)

几个常见的分布——超几何分布

设一批同类产品共N件，其中有M件次品，从中任取n (n≤N)件，其次品数X恰为k件的概率分布为：

$$P\{X=k\}=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \qquad k=0,1,2,\cdots,\min(n,M)$$

则称次品数X服从参数为 (N, M, n) 的超几何分布。

几个常见的分布——均匀分布

若随机变量X的概率密度为

$$p(x)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & \text{其他}\end{cases}$$

则称X在区间[a, b]上服从均匀分布，记为

X~U (a, b)

几个常见的分布——指数分布

若随机变量X的概率密度为

$$p(x)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0\\ 0 & x<0\end{cases}$$

则称X服从参数为λ的指数分布，记为

$$X\sim E(\lambda)$$

几个常见的分布——正态分布

若随机变量X的概率密度为

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty<x<\infty$$

其中，μ，σ是两个常数，则称X服从参数为μ，σ的正态分布，记为

X~N(μ，σ)

几个常见的分布——Γ分布

若随机变量X的概率密度为

$$p(x)=\begin{cases}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} & x>0\\ 0 & x\le 0\end{cases}$$

记为

$$X\sim\Gamma(\alpha,\beta)$$

几个常见的分布——β分布

若随机变量X的概率密度为

$$p(x)=\begin{cases}\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} & 0<x<1\\ 0 & \text{其他}\end{cases}$$

记为

$$X\sim\beta(\alpha,\beta)$$

几个常见的分布——x²分布

若随机变量X的概率密度为

$$p(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}} & x\geq 0\\ 0 & x<0\end{cases}$$

则称X为服从自由度为n的χ^2分布，记为：

$$X\sim\chi^2(n)$$

几个常见的分布——t分布

若随机变量X的概率密度为

$$p(t)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-\frac{n+1}{2}} \qquad -\infty<t<\infty$$

则称X为服从自由度为n的t分布，记为：

$$X\sim t(n)$$

几个常见的分布——F分布

若随机变量X的概率密度为

$$p(x)=\begin{cases}\dfrac{\Gamma(\frac{n_1+n_2}{2})n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}\cdot\dfrac{x^{\frac{n_1}{2}-1}}{(n_1x+n_2)^{\frac{n_1+n_2}{2}}} & x>0\\ 0 & x\le 0\end{cases}$$

则称X为服从自由度为n_1，n_2的F分布，记为：

$$X\sim F(n_1,n_2)$$

1　随机数的产生

有两种方法可以产生随机数据：

- 通用函数
- 专用函数

(1) 通用函数产生随机数

格式：

y=random('name',A,B,C,m,n)

- name的取值见下页表；
- A，B，C为分布的参数；
- m，n指定随机数的行和列

name 的取值			函数说明
'beta'	或	'Beta'	Beta 分布
'bino'	或	'Binomial'	二项分布
'chi2'	或	'Chisquare'	卡方分布
'exp'	或	'Exponential'	指数分布
'f'	或	'F'	F 分布
'gam'	或	'Gamma'	GAMMA 分布
'geo'	或	'Geometric'	几何分布
'hyge'	或	'Hypergeometric'	超几何分布
'logn'	或	'Lognormal'	对数正态分布
'nbin'	或	'Negative Binomial'	负二项式分布
'ncf'	或	'Noncentral F'	非中心 F 分布
'nct'	或	'Noncentral t'	非中心 t 分布
'ncx2'	或	'Noncentral Chi-square'	非中心卡方分布
'norm'	或	'Normal'	正态分布
'poiss'	或	'Poisson'	泊松分布
'rayl'	或	'Rayleigh'	瑞利分布
't'	或	'T'	T 分布
'unif'	或	'Uniform'	均匀分布
'unid'	或	'Discrete Uniform'	离散均匀分布
'weib'	或	'Weibull'	Weibull 分布

例如:

(1) R = random('Normal',0,1,2,4):

生成期望为 0,标准差为 1 的(2 行 4 列)2× 4 个正态随机数

(2) R = random('Poisson',1:6,1,6):

依次生成参数为 1 到 6 的(1 行 6 列)6 个 Poisson 随机数

(3) 产生12(3x4)个均值为2，标准差为0.3的正态分布的随机数

R = random('Normal',2,0.3,4,4):

(2)专用函数产生随机数

- 1.1 二项分布的随机数据的产生
- 1.2 正态分布的随机数据的产生
- 1.3 常见分布的随机数产生
- 1.4 通用函数求各分布的随机数据

(2.1) 二项分布的随机数据的产生

格式：**R = binornd(N,P,m,n)**

- m,n分别表示R的行数和列数

例如：

```
>>R = binornd(5,0.5,2)
R =
     4     3
     4     2
```

(2.2) 正态分布的随机数据的产生

格式 **R = normrnd(MU,SIGMA,m,n)**

- m,n分别表示R的行数和列数

例如:

```
>>R=normrnd(0,1,2,3)
R =
    1.1892    0.3273   -0.1867
   -0.0376    0.1746    0.7258
```

(2.3) 常见分布的随机数产生

函数名	调用形式	注　　释
Unifrnd	unifrnd (A,B,m,n)	[A,B]上均匀分布(连续) 随机数
Unidrnd	unidrnd(N,m,n)	均匀分布（离散）随机数
Exprnd	exprnd(Lambda,m,n)	参数为 Lambda 的指数分布随机数
Normrnd	normrnd(MU,SIGMA,m,n)	参数为 MU，SIGMA 的正态分布随机数
chi2rnd	chi2rnd(N,m,n)	自由度为 N 的卡方分布随机数
Trnd	trnd(N,m,n)	自由度为 N 的 t 分布随机数
Frnd	frnd(N_1, N_2,m,n)	第一自由度为 N_1,第二自由度为 N_2 的 F 分布随机数
gamrnd	gamrnd(A, B,m,n)	参数为 A, B 的 γ 分布随机数
betarnd	betarnd(A, B,m,n)	参数为 A, B 的 β 分布随机数
lognrnd	lognrnd(MU, SIGMA,m,n)	参数为 MU, SIGMA 的对数正态分布随机数
nbinrnd	nbinrnd(R, P,m,n)	参数为 R，P 的负二项式分布随机数
ncfrnd	ncfrnd(N_1, N_2, delta,m,n)	参数为 N_1，N_2，delta 的非中心 F 分布随机数
nctrnd	nctrnd(N, delta,m,n)	参数为 N，delta 的非中心 t 分布随机数
ncx2rnd	ncx2rnd(N, delta,m,n)	参数为 N，delta 的非中心卡方分布随机数
raylrnd	raylrnd(B,m,n)	参数为 B 的瑞利分布随机数
weibrnd	weibrnd(A, B,m,n)	参数为 A, B 的韦伯分布随机数
binornd	binornd(N,P,m,n)	参数为 N, p 的二项分布随机数
geornd	geornd(P,m,n)	参数为 p 的几何分布随机数
hygernd	hygernd(M,K,N,m,n)	参数为 M，K，N 的超几何分布随机数
Poissrnd	poissrnd(Lambda,m,n)	参数为 Lambda 的泊松分布随机数

2　随机变量的概率密度计算

- 2.1 通用函数计算概率密度函数值
- 2.2 专用函数计算概率密度函数值
- 2.3 常见分布的密度函数作图

2.1 通用函数计算概率密度函数值

格式

- **Y=pdf(name，x，A)**
- **Y=pdf(name，x，A，B)**
- **Y=pdf(name，x，A，B，C)**

说明

返回在X=x处、参数为A、B、C的概率密度值，对于不同的分布，参数个数是不同；

name为分布函数名。

例题:计算正态分布N（0，1）的随机变量X在点0.6578的密度函数值。

```
>>pdf('Normal',0.6578,0,1)
ans =    0.3213
```

例题:自由度为8的卡方分布，在点2.18处的密度函数值。

```
>> pdf('Chisquare',2.18,8)
ans =    0.0363
```

2.2 专用函数计算概率密度函数值

- 二项分布的概率密度函数值

  格式 **binopdf (k, n, p)**

  - 等同于pdf('bino', k, n, p)，
  - p — 每次试验事件发生的概率；
  - K—事件发生k次；
  - n—试验总次数

- 泊松分布的概率值

格式 poisspdf(k, Lambda)

  - 等同于pdf('pois', k, Lambda)

- 正态分布的概率值

格式 **normpdf**(x,mu,sigma)

  - 计算参数为μ=mu，σ=sigma的正态分布密度函数在x处的值

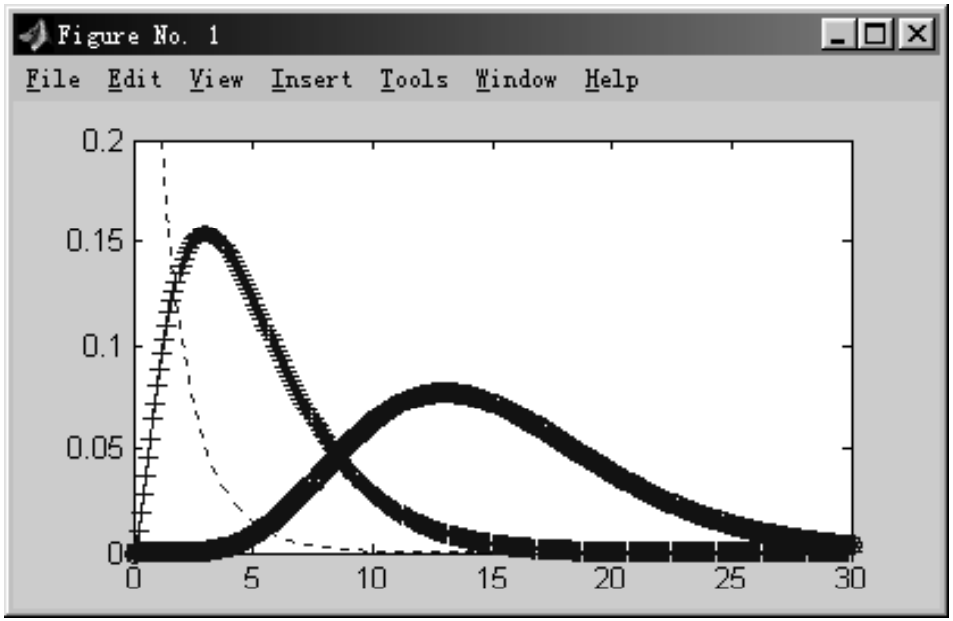
专用函数计算概率密度函数表

函数名	调用形式	注　　释
Unifpdf	unifpdf (x, a, b)	[a,b]上均匀分布(连续)概率密度在 X=x 处的函数值
unidpdf	Unidpdf(x,n)	均匀分布（离散）概率密度函数值
Exppdf	exppdf(x, Lambda)	参数为 Lambda 的指数分布概率密度函数值
normpdf	normpdf(x, mu, sigma)	参数为 mu，sigma 的正态分布概率密度函数值
chi2pdf	chi2pdf(x, n)	自由度为 n 的卡方分布概率密度函数值
Tpdf	tpdf(x, n)	自由度为 n 的 t 分布概率密度函数值
Fpdf	fpdf(x, n_1, n_2)	第一自由度为 n_1,第二自由度为 n_2 的 F 分布概率密度函数值
gampdf	gampdf(x, a, b)	参数为 a, b 的 γ 分布概率密度函数值
betapdf	betapdf(x, a, b)	参数为 a, b 的 β 分布概率密度函数值
lognpdf	lognpdf(x, mu, sigma)	参数为 mu, sigma 的对数正态分布概率密度函数值
nbinpdf	nbinpdf(x, R, P)	参数为 R，P 的负二项式分布概率密度函数值
Ncfpdf	ncfpdf(x, n_1, n_2, delta)	参数为 n_1，n_2，delta 的非中心 F 分布概率密度函数值
Nctpdf	nctpdf(x, n, delta)	参数为 n，delta 的非中心 t 分布概率密度函数值
ncx2pdf	ncx2pdf(x, n, delta)	参数为 n，delta 的非中心卡方分布概率密度函数值
raylpdf	raylpdf(x, b)	参数为 b 的瑞利分布概率密度函数值
weibpdf	weibpdf(x, a, b)	参数为 a, b 的韦伯分布概率密度函数值
binopdf	binopdf(x,n,p)	参数为 n, p 的二项分布的概率密度函数值
geopdf	geopdf(x,p)	参数为 p 的几何分布的概率密度函数值
hygepdf	hygepdf(x,M,K,N)	参数为 M，K，N 的超几何分布的概率密度函数值
poisspdf	poisspdf(x,Lambda)	参数为 Lambda 的泊松分布的概率密度函数值

- 绘制卡方分布密度函数在自由度分别为1、5、15的图形

```
>> x=0:0.1:30;
>> y1=chi2pdf(x,1);
>> plot(x,y1,':')
>> hold on
>> y2=chi2pdf(x,5);
>> plot(x,y2,'+')
>> y3=chi2pdf(x,15);
>> plot(x,y3,'o')
>> axis([0,30,0,0.2])
```

%指定显示的图形区域



3 随机变量的累积概率值(分布函数值)

$$F(x) = P\{X \le x\} = \int_{-\infty}^{x} p(t)dt$$

- 3.1 通用函数计算累积概率值
- 3.2 专用函数计算累积概率值

3.1 通用函数计算累积概率值

格式

- **cdf('name', K, A)**
- **cdf('name', K, A, B)**
- **cdf('name', K, A, B, C)**

说明

返回以name为分布、随机变量X≤K的概率之和的累积概率值

例 求标准正态分布随机变量X落在区间(-∞，0.4)内的概率.

```
>> cdf('norm',0.4, 0,1)
```

例 求自由度为16的卡方分布随机变量落在[0，6.91]内的概率

```
>> cdf('chi2',6.91, 16)
```

3.2 专用函数计算累积概率值

- **二项分布的累积概率值**

格式 binocdf (k, n, p)

- n为试验总次数
- p为每次试验事件A发生的概率
- k为n次试验中事件A发生的次数
- 该命令返回n次试验中事件A恰好发生k次的概率。

- 正态分布的累积概率值

格式 **normcdf(x, mu, sigma)**

- 返回F(x)= $\int_{-\infty}^{x} p(t)dt$ 的值，
- mu、sigma为正态分布的两个参数

设X~N(3, 2^2)，

（1）求 $P\{2<X<5\}$， $P\{-4<X<10\}$， $P\{|X|>2\}$， $P\{X>3\}$

（2）确定c，使得 $P\{X>c\}=\frac{1}{2}P\{X<c\}$

```
>>p1=normcdf(5,3,2)-normcdf(2,3,2)
p1 =
   0.5328
```

p1 = $P\{2<X<5\}$

```
>>p2=normcdf(10,3,2)-normcdf(-4,3,2)
p2 =
   0.9995
```

p2 = $P\{-4<X<10\}$

```
>>p3=1-normcdf(2,3,2)-normcdf(-2,3,2)
p3 =
   0.6853
```

p3 = $P\{|X|>2\}=1-P\{|X|\le 2\}$

```
>>p4=1-normcdf(3,3,2)
p4 =
   0.5000
```

p4 = $P\{X>3\}=1-P\{X\le 3\}$

专用函数的累积概率值函数表

函数名	调用形式	注　　释
unifcdf	unifcdf (x, a, b)	[a,b]上均匀分布(连续)累积分布函数值 F(x)=P{X≤x}
unidcdf	unidcdf(x,n)	均匀分布（离散）累积分布函数值 F(x)=P{X≤x}
expcdf	expcdf(x, Lambda)	参数为 Lambda 的指数分布累积分布函数值 F(x)=P{X≤x}
normcdf	normcdf(x, mu, sigma)	参数为 mu，sigma 的正态分布累积分布函数值 F(x)=P{X≤x}
chi2cdf	chi2cdf(x, n)	自由度为 n 的卡方分布累积分布函数值 F(x)=P{X≤x}
tcdf	tcdf(x, n)	自由度为 n 的 t 分布累积分布函数值 F(x)=P{X≤x}
fcdf	fcdf(x, n_1, n_2)	第一自由度为 n_1,第二自由度为 n_2 的 F 分布累积分布函数值
gamcdf	gamcdf(x, a, b)	参数为 a, b 的 γ 分布累积分布函数值 F(x)=P{X≤x}
betacdf	betacdf(x, a, b)	参数为 a, b 的 β 分布累积分布函数值 F(x)=P{X≤x}
logncdf	logncdf(x, mu, sigma)	参数为 mu, sigma 的对数正态分布累积分布函数值
nbincdf	nbincdf(x, R, P)	参数为 R，P 的负二项式分布概累积分布函数值 F(x)=P{X≤x}
ncfcdf	ncfcdf(x, n_1, n_2, delta)	参数为 n_1，n_2，delta 的非中心 F 分布累积分布函数值
nctcdf	nctcdf(x, n, delta)	参数为 n，delta 的非中心 t 分布累积分布函数值 F(x)=P{X≤x}
ncx2cdf	ncx2cdf(x, n, delta)	参数为 n，delta 的非中心卡方分布累积分布函数值
raylcdf	raylcdf(x, b)	参数为 b 的瑞利分布累积分布函数值 F(x)=P{X≤x}
weibcdf	weibcdf(x, a, b)	参数为 a, b 的韦伯分布累积分布函数值 F(x)=P{X≤x}
binocdf	binocdf(x,n,p)	参数为 n, p 的二项分布的累积分布函数值 F(x)=P{X≤x}
geocdf	geocdf(x,p)	参数为 p 的几何分布的累积分布函数值 F(x)=P{X≤x}
hygecdf	hygecdf(x,M,K,N)	参数为 M，K，N 的超几何分布的累积分布函数值
poisscdf	poisscdf(x,Lambda)	参数为 Lambda 的泊松分布的累积分布函数值 F(x)=P{X≤x}

4 随机变量的逆累积分布函数

已知 $F(x) = P\{X \le x\}$ ，求x。

- 4.1 通用函数计算逆累积分布函数值
- 4.2 专用函数-inv计算逆累积分布函数
- 4.3 随机变量的数字特征

4.1 通用函数计算逆累积分布函数值

格式 icdf('name', P,A,B, C)

- 返回分布为name，参数为A,B,C，累积概率值为P的临界值。
- 如果P= cdf('name', x,A,B,C)，
  则 x = icdf('name', P,A,B,C)

例：在标准正态分布表中，若已知$\Phi(x)=0.975$，求**x**

```
>>icdf('norm',0.975,0,1)
ans =    1.9600
```

例　：在 χ^2分布表中，若自由度为**10**，
α**=0.975**，求临界值**Lambda**。

```
>>icdf('chi2',0.025,10)
ans =   3.2470
```

例：在假设检验中，求临界值问题。已知显著性性水平为**0.05**，查自由度为**10**的双边界检验**t**分布临界值

```
>>icdf('t',0.025,10)
ans = -2.2281
```

4.2 专用函数-inv计算逆累积分布函数

- 正态分布逆累积分布函数

格式 **X=norminv(p,mu,sigma)**

- p为累积概率值，
- mu为均值，sigma为标准差，X为临界值
- 满足：p=P{X≤x}。

例：在标准正态分布表中，若已知$\Phi(x)=0.975$，求**x**

```
>> x=icdf('norm',0.975,0,1)
x =    1.9600
```

例　：在 χ^2分布表中，若自由度为**10**，
α**=0.975**，求临界值**Lambda**。

```
>>Lambda=icdf('chi2',0.025,10)
Lambda =    3.2470
```

例：在假设检验中，求临界值问题。已知显著性性水平为**0.05**，查自由度为**10**的双边界检验**t**分布临界值

```
>>lambda=icdf('t',0.025,10)
lambda =        -2.2281
```

常用临界值函数表

函数名	调用形式	注　　释
unifinv	x=unifinv (p, a, b)	均匀分布(连续)逆累积分布函数（P=P{X≤x}，求 x）
unidinv	x=unidinv (p,n)	均匀分布（离散）逆累积分布函数，x 为临界值
expinv	x=expinv (p, Lambda)	指数分布逆累积分布函数
norminv	x=Norminv(x,mu,sigma)	正态分布逆累积分布函数
chi2inv	x=chi2inv (x, n)	卡方分布逆累积分布函数
tinv	x=tinv (x, n)	t 分布累积分布函数
finv	x=finv (x, n_1, n_2)	F 分布逆累积分布函数
gaminv	x=gaminv (x, a, b)	γ 分布逆累积分布函数
betainv	x=betainv (x, a, b)	β 分布逆累积分布函数
logninv	x=logninv (x, mu, sigma)	对数正态分布逆累积分布函数
nbininv	x=nbininv (x, R, P)	负二项式分布逆累积分布函数
ncfinv	x=ncfinv (x, n_1, n_2, delta)	非中心 F 分布逆累积分布函数
nctinv	x=nctinv (x, n, delta)	非中心 t 分布逆累积分布函数
ncx2inv	x=ncx2inv (x, n, delta)	非中心卡方分布逆累积分布函数
raylinv	x=raylinv (x, b)	瑞利分布逆累积分布函数
weibinv	x=weibinv (x, a, b)	韦伯分布逆累积分布函数
binoinv	x=binoinv (x,n,p)	二项分布的逆累积分布函数
geoinv	x=geoinv (x,p)	几何分布的逆累积分布函数
hygeinv	x=hygeinv (x,M,K,N)	超几何分布的逆累积分布函数
poissinv	x=poissinv (x,Lambda)	泊松分布的逆累积分布函数

设X~N(3, 2^2)，

（1）求 $P\{2<X<5\}$， $P\{-4<X<10\}$， $P\{|X|>2\}$， $P\{X>3\}$

（2）确定c，使得 $P\{X>c\}=\frac{1}{2}P\{X<c\}$

$$P\{X>c\}=\frac{1}{2}P\{X<c\} \longrightarrow 1-P\{X<c\}=\frac{1}{2}P\{X<c\}$$

$$\longrightarrow P\{X<c\}=\frac{2}{3}$$

```
>>X=norminv(2/3, 3, 2)
X=
   3.8615
```

例

公共汽车门的高度是按成年男子与车门顶碰头的机会不超过1%设计的。设男子身高X（单位：cm）服从正态分布N（175，36），求车门的最低高度。

- 设h为车门高度，X为身高，求满足条件的h:

$$P\{X > h\} \le 0.01$$

即P{X<h}>=0.99

```
>> h=norminv(0.99,175,6)
h=          188.9581
```

5 随机变量的数字特征

- 5.1 平均值、中值
- 5.2 数据比较
- 5.3 期望
- 5.4 方差
- 5.5 常见分布的期望和方差
- 5.6 协方差与相关系数

5.1 平均值、中值

- 命令 利用mean求算术平均值
- 格式
  - mean(X)
    - X为向量，返回X中各元素的平均值
  - mean(A)
    - A为矩阵，返回A中各列元素的平均值构成的向量
  - mean(A,dim)
    - 在给出的维数内的平均值
- 说明
  - X为向量时，算术平均值的数学含义是 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ ，即样本均值。

```
>> A=[1 3 4 5;2 3 4 6;1 3 1 5]
A =
     1     3     4     5
     2     3     4     6
     1     3     1     5
>> mean(A)
ans =
    1.3333    3.0000    3.0000    5.3333
>> mean(A,1)
ans =
   1.3333    3.0000    3.0000    5.3333
```

- 命令
  - 忽略NaN计算算术平均值
- 格式
  - nanmean(X)
    - X为向量，返回X中除NaN外元素的算术平均值。
  - nanmean(A)
    - A为矩阵，返回A中各列除NaN外元素的算术平均值向量。

```
>> A=[1 2 3;nan 5 2;3 7 nan]
A =
     1     2     3
   NaN     5     2
     3     7   NaN
>> nanmean(A)
ans =
    2.0000    4.6667    2.5000
```

- 命令 利用median计算中值（中位数）
- 格式
  - median(X)
    - X为向量，返回X中各元素的中位数。
  - median(A)
    - A为矩阵，返回A中各列元素的中位数构成的向量。
  - median(A,dim)
    - 求给出的维数内的中位数

```
>> A=[1 3 4 5;2 3 4 6;1 3 1 5]
A =
     1     3     4     5
     2     3     4     6
     1     3     1     5
>> median(A)
ans =
     1     3     4     5
```

- 命令 忽略NaN计算中位数
- 格式
  - nanmedian(X)
    - X为向量，返回X中除NaN外元素的中位数。
  - nanmedian(A)
    - A为矩阵，返回A中各列除NaN外元素的中位数向量。

```
>> A=[1 2 3;nan 5 2;3 7 nan]
A =
     1     2     3
   NaN     5     2
     3     7   NaN
>> nanmedian(A)
ans =
    2.0000    5.0000    2.5000
```

- 命令 利用geomean计算几何平均数
- 格式
  - M=geomean(X)
    - X为向量，返回X中各元素的几何平均数。
  - M=geomean(A)
    - A为矩阵，返回A中各列元素的几何平均数构成的向量。
- 说明
  - 几何平均数的数学含义是 ，

$$M = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

其中：样本数据非负，主要用于对数正态分布。

```
>> B=[1 3 4 5]
B =     1     3     4     5
>> M=geomean(B)
M =    2.7832
>> A=[1 3 4 5;2 3 4 6;1 3 1 5]
A =
     1     3     4     5
     2     3     4     6
     1     3     1     5
>> M=geomean(A)
M =    1.2599    3.0000    2.5198    5.3133
```

- 命令 利用harmmean求调和平均值
- 格式
  - M=harmmean(X)
    - X为向量，返回X中各元素的调和平均值。
  - M=harmmean(A)
    - A为矩阵，返回A中各列元素的调和平均值构成的向量。
- 说明
  - 调和平均值的数学含义是，其中：样本数据非0，主要用于严重偏斜分布。

$$M = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

```
>> B =     B=[1  3  4  5]
1     3     4     5
>> M=harmmean(B)
M =    2.2430
>> A=[1 3 4 5;2 3 4 6;1 3 1 5]
A =
     1     3     4     5
     2     3     4     6
     1     3     1     5
>> M=harmmean(A)
M =    1.2000    3.0000    2.0000    5.2941
```

5.2 数据比较

- 命令 排序
- 格式
  - Y=sort(X)
    - X为向量，返回X按由小到大排序后的向量。
  - Y=sort(A)
    - A为矩阵，返回A的各列按由小到大排序后的矩阵。
  - [Y,I]=sort(A)
    - Y为排序的结果，I中元素表示Y中对应元素在A中位置。
  - sort(A,dim)
    - 在给定的维数dim内排序
- 说明
  - 若X为复数，则通过|X|排序。

```
>> A=[1 2 3;4 5 2;3 7 0]
A =
     1     2     3
     4     5     2
     3     7     0
>> sort(A)
ans =
     1     2     0
     3     5     2
     4     7     3
>> [Y,I]=sort(A)
Y =
     1     2     0
     3     5     2
     4     7     3
I =
     1     1     3
     3     2     2
     2     3     1
```

- 命令 按行方式排序
- 函数 **sortrows**
- 格式
  - Y=sortrows(A)
    - A为矩阵，返回矩阵Y，Y按A的第1列由小到大，以行方式排序后生成的矩阵。
  - Y=sortrows(A, col)
    - 按指定列col由小到大进行排序
  - [Y,I]=sortrows(A, col)
    - Y为排序的结果，I表示Y中第col列元素在A中位置。
- 说明
  - 若X为复数，则通过|X|的大小排序。

```
>> A=[1 2 3;4 5 2;3 7 0]
>> sortrows(A)
>> sortrows(A,1)
>> sortrows(A,3)
>> sortrows(A,[3 2])
>> [Y,I]=sortrows(A,3)
Y =
     3     7     0
     4     5     2
     1     2     3
I =
     3
     2
     1
```

```
A =
     1     2     3
     4     5     2
     3     7     0

ans =
     1     2     3
     3     7     0
     4     5     2

ans =
     1     2     3
     3     7     0
     4     5     2

ans =
     3     7     0
     4     5     2
     1     2     3

ans =
     3     7     0
     4     5     2
     1     2     3
```

- 命令 求最大值与最小值之差
- 函数 **range**
- 格式
  - Y=range(X)
    - X为向量，返回X中的最大值与最小值之差。
  - Y=range(A)
    - A为矩阵，返回A中各列元素的最大值与最小值之差。

```
>> A=[1 2 3;4 5 2;3 7 0]
A =
     1     2     3
     4     5     2
     3     7     0
>> Y=range(A)
Y =
     3     5     3
```

5.3 期望

- 离散型随机变量X的期望计算

$$EX = \sum_{k=1}^{\infty} x_k p_k$$

(1)sum(x.*p)
(2)样本均值：mean (X)

- 例40 设随机变量X的分布率为:

X	-2	-1	0	1	2
P	0.3	01.	0.2	0.1	0.3

求E(X)和$E(X^2+1)$

```
X=[-2 -1 0 1 2];
p=[0.3 0.1 0.2 0.1 0.3];
EX=sum(X.*p)
Y=X.^2-1
EY=sum(Y.*p)
```

运行后结果如下：

```
EX =
     0
Y =
   3     0    -1     0     3
EY =
   1.6000
```

例 随机抽取6个滚珠测得直径如下

14.70 15.21 14.90 14.91 15.32 15.32

试求样本平均值

解：>>X=[14.70 15.21 14.90 14.91 15.32 15.32]；

>>**mean(X)** %计算样本均值

- 连续型随机变量

$$EX = \int_{-\infty}^{+\infty} xp(x)dx$$

函数： int

格式：inf(f,inf,inf)

已知随机变量X的概率密度

求EX和E (4X-1)。

$$p(x)=\begin{cases}3x^2 & 0<x<1\\ 0 & \text{其它}\end{cases}$$

```
syms x
p_x=3*x^2;
EX=int(x*p_x,0,1)
EY=int((4*x-1)*p_x,0,1)
```

运行结果为：

```
EX =
 3/4
EY =
 2
```

5.4 方差

- 命令 求样本方差
- 函数 **var**
- 格式

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$$

  - D=var(X)
    - var(X)=s^2,若X为向量，则返回向量的样本方差。
  - D=var(A)
    - A为矩阵，则D为A的列向量的样本方差构成的行向量。
  - D=var(X, 1)
    - 返回向量（矩阵）X的简单方差（即置前因子为1/n的方差）
  - D=var(X, w)
    - 返回向量（矩阵）X的以w为权重的方差

- 命令 求标准差
- 函数 **std**
- 格式
  - std(X)
    - 返回向量（矩阵）X的样本标准差（置前因子为1/(n-1)）；
  - std(X,1)
    - 返回向量（矩阵）X的标准差（置前因子为1/n）
  - std(X, 0)
    - 与std (X)相同
  - std(X, flag, dim)
    - 返回向量（矩阵）中维数为dim的标准差值，
    - 其中flag=0时，置前因子为1/(n-1)；否则置前因子为1/n。

$$std = \sqrt{\frac{1}{n-1}\sum_{i}^{n}\left(x_i - \bar{X}\right)^2}$$

- 求下列样本方差和样本标准差，方差与标准差

14.70 15.21 14.90 15.32 15.32

```
>>X=[14.7 15.21 14.9 14.91 15.32 15.32];
>>DX=var(X,1)        %方差
  DX =
      0.0559
>>sigma=std(X,1)     %标准差
sigma =
    0.2364
>>DX1=var(X)         %样本方差
DX1 =
      0.0671
>>sigma1=std(X)      %样本标准差
 sigma1 =
   0.2590
```

- 命令 忽略NaN的标准差
- 函数 **nanstd**
- 格式
  - y = nanstd(X)
    - 若X为含有元素NaN的向量，则返回除NaN外的元素的标准差，
    - 若X为含元素NaN的矩阵，则返回各列除NaN外的标准差构成的向量。

- 命令 样本的偏斜度
- 函数 **skewness**
- 格式
  - y = skewness(X)
    - X为向量，返回X的元素的偏斜度；X为矩阵，返回X各列元素的偏斜度构成的行向量。
  - y = skewness(X,flag)
    - flag=0表示偏斜纠正，flag=1（默认）表示偏斜不纠正。
- 说明
  - 偏斜度样本数据关于均值不对称的一个测度，如果偏斜度为负，说明均值左边的数据比均值右边的数据更散；如果偏斜度为正，说明均值右边的数据比均值左边的数据更散，因而正态分布的偏斜度为0；
  - 偏斜度的定义：

$$y = \frac{E(x-\mu)^3}{\sigma^3}$$

其中：μ为x的均值，σ为x的标准差，E(.)为期望值算子

5.5 常见分布的期望和方差

- 命令 均匀分布（连续）的期望和方差
- 函数 **unifstat**
- 格式
  - [M,V] = unifstat(A,B)
    - A、B为标量时，就是区间上均匀分布的期望和方差
    - A、B也可为向量或矩阵，则M、V也是向量或矩阵。

```
>>a = 1:6; b = 2.*a;
>>[M,V] = unifstat(a,b)
M =
    1.5000    3.0000    4.5000    6.0000    7.5000    9.0000
V =
    0.0833    0.3333    0.7500    1.3333    2.0833    3.0000
```

- 命令 正态分布的期望和方差
- 函数 **normstat**
- 格式
  - [M,V] = normstat(MU,SIGMA)
    - MU、SIGMA可为标量也可为向量或矩阵，则M=MU，V=SIGMA2。

```
>> n=1:4;
>> [M,V]=normstat(n'*n,n'*n)
M =
     1     2     3     4
     2     4     6     8
     3     6     9    12
     4     8    12    16
V =
     1     4     9    16
     4    16    36    64
     9    36    81   144
    16    64   144   256
```

- 命令 二项分布的均值和方差
- 函数 **binostat**
- 格式
  - [M,V] = binostat(N,P)
    - N，P为二项分布的两个参数，可为标量也可为向量或矩阵。

```
>>n = logspace(1,5,5)
n =
        10        100        1000        10000        100000
>>[M,V] = binostat(n,1./n)
M =
     1     1     1     1     1
V =
    0.9000    0.9900    0.9990    0.9999    1.0000
>>[m,v] = binostat(n,1/2)
m =
        5        50        500        5000        50000
v =    1.0e+04 *
    0.0003    0.0025    0.0250    0.2500    2.5000
```

常见分布的均值和方差

函数名	调用形式	注　释
unifstat	[M,V]=unifstat (a, b)	均匀分布(连续)的期望和方差，M 为期望，V 为方差
unidstat	[M,V]=unidstat (n)	均匀分布（离散）的期望和方差
expstat	[M,V]=expstat (p, Lambda)	指数分布的期望和方差
normstat	[M,V]=normstat(mu,sigma)	正态分布的期望和方差
chi2stat	[M,V]=chi2stat (x, n)	卡方分布的期望和方差
tstat	[M,V]=tstat (n)	t 分布的期望和方差
fstat	[M,V]=fstat (n_1, n_2)	F 分布的期望和方差
gamstat	[M,V]=gamstat (a, b)	γ 分布的期望和方差
betastat	[M,V]=betastat (a, b)	β 分布的期望和方差
lognstat	[M,V]=lognstat (mu, sigma)	对数正态分布的期望和方差
nbinstat	[M,V]=nbinstat (R, P)	负二项式分布的期望和方差
ncfstat	[M,V]=ncfstat (n_1, n_2, delta)	非中心 F 分布的期望和方差
nctstat	[M,V]=nctstat (n, delta)	非中心 t 分布的期望和方差
ncx2stat	[M,V]=ncx2stat (n, delta)	非中心卡方分布的期望和方差
raylstat	[M,V]=raylstat (b)	瑞利分布的期望和方差
Weibstat	[M,V]=weibstat (a, b)	韦伯分布的期望和方差
Binostat	[M,V]=binostat (n,p)	二项分布的期望和方差
Geostat	[M,V]=geostat (p)	几何分布的期望和方差
hygestat	[M,V]=hygestat (M,K,N)	超几何分布的期望和方差
Poisstat	[M,V]=poisstat (Lambda)	泊松分布的期望和方差

5.6 协方差与相关系数

- 命令 协方差
- 函数 **cov**
- 格式
  - cov(X)
    - 求向量X的协方差
  - cov(A)
    - 求矩阵A的协方差矩阵，该协方差矩阵的对角线元素是A的各列的方差，即：var(A)=diag(cov(A))。
  - cov(X,Y)
    - X,Y为等长列向量，等同于cov([X Y])。

```
>> X=[0 -1 1]';Y=[1 2 2]';
>> C1=cov(X)     %X的协方差
C1 =     1
>> C2=cov(X,Y)
C2 =
    1.0000         0
         0    0.3333
>> A=[1 2 3;4 0 -1;1 7 3];
>> C1=cov(A)
C1 =
    3.0000   -4.5000   -4.0000
   -4.5000   13.0000    6.0000
   -4.0000    6.0000    5.3333
>> C2=var(A(:,1))
C2 =     3
>> C3=var(A(:,2))
C3 =    13
>> C4=var(A(:,3))
```

- 命令 相关系数
- 函数 **corrcoef**
- 格式
  - corrcoef(X,Y)
    - 返回列向量X,Y的相关系数，等同于corrcoef([X Y])。
  - corrcoef (A)
    - 返回矩阵A的列向量的相关系数矩阵

```
>> A=[1 2 3;4 0 -1;1 3 9]
A =
     1     2     3
     4     0    -1
     1     3     9
>> C1=corrcoef(A)
C1 =
    1.0000   -0.9449   -0.8030
   -0.9449    1.0000    0.9538
   -0.8030    0.9538    1.0000
>> C1=corrcoef(A(:,2),A(:,3))
C1 =
    1.0000    0.9538
    0.9538    1.0000
```

6 统计作图

- 6.1 正整数的频率表
- 6.2 经验累积分布函数图形
- 6.3 最小二乘拟合直线
- 6.4 绘制正态分布概率图形
- 6.5 绘制威布尔(Weibull)概率图形
- 6.6 样本数据的盒图
- 6.7 给当前图形加一条参考线
- 6.8 在当前图形中加入一条多项式曲线
- 6.9 样本的概率图形
- 6.10 附加有正态密度曲线的直方图
- 6.11 在指定的界线之间画正态密度曲线

6.1 正整数的频率表

- 命令 正整数的频率表
- 函数 tabulate
- 格式
  - table = tabulate(X)
    - X为正整数构成的向量，返回3列：
      第1列中包含X的值第2列为这些值的个数，
      第3列为这些值的频率。

例49

6.2 经验累积分布函数图形

- 函数 cdfplot
- 格式
  - cdfplot(X)
    - 作样本X（向量）的累积分布函数图形
  - h = cdfplot(X)
    - h表示曲线的环柄
  - [h,stats] = cdfplot(X)
    - stats表示样本的一些特征

例50

6.3 最小二乘拟合直线

- 函数 lsline
- 格式
  - lsline
    - 最小二乘拟合直线
  - h = lsline
    - h为直线的句柄

例51

6.4 绘制正态分布概率图形

- 函数 normplot
- 格式
  - normplot(X)
    - 若X为向量，则显示正态分布概率图形，
    - 若X为矩阵，则显示每一列的正态分布概率图形。
  - h = normplot(X)
  - 返回绘图直线的句柄
- 说明
  - 样本数据在图中用“+”显示；
  - 如果数据来自正态分布，则图形显示为直线，而其它分布可能在图中产生弯曲。

例53

6.5 绘制威布尔(Weibull)概率图形

- 函数 weibplot
- 格式
  - weibplot(X)
    - 若X为向量，则显示威布尔(Weibull)概率图形，
    - 若X为矩阵，则显示每一列的威布尔概率图形。
  - h = weibplot(X)
    - 返回绘图直线的柄
- 说明
  - 绘制威布尔(Weibull)概率图形的目的是用图解法估计来自威布尔分布的数据X
  - 如果X是威布尔分布数据，其图形是直线的，否则图形中可能产生弯曲。

例54

6.6 样本数据的盒图

- 函数 boxplot
- 格式
  - boxplot(X)
    - 产生矩阵X的每一列的盒图和“须”图，“须”是从盒的尾部延伸出来，并表示盒外数据长度的线，如果“须”的外面没有数据，则在“须”的底部有一个点。
  - boxplot(X,notch)
    - 当notch=1时，产生一凹盒图，notch=0时产生一矩箱图。
  - boxplot(X,notch,'sym')
    - sym表示图形符号，默认值为“+”。
  - boxplot(X,notch,'sym',vert)
    - 当vert=0时，生成水平盒图，vert=1时，生成竖直盒图（默认值vert=1）。
  - boxplot(X,notch,'sym',vert,whis)
    - whis定义“须”图的长度，默认值为1.5，若whis=0则boxplot函数通过绘制sym符号图来显示盒外的所有数据值。

例55

6.7 给当前图形加一条参考线

- 函数 refline
- 格式
  - refline(slope,intercept)
    - slope表示直线斜率，intercept表示截距
  - refline(slope)
    - slope=[a b]，图中加一条直线：y=b+ax。

例56

6.8 在当前图形中加入一条多项式曲线

- 函数 refcurve
- 格式
  - h = refcurve(p)
    - 在图中加入一条多项式曲线，
    - h为曲线的环柄，
    - p为多项式系数向量，p=[p1,p2, p3,…,pn]，其中p1为最高幂项系数。

例57

6.9 样本的概率图形

- 函数 capaplot
- 格式
  - p = capaplot(data,specs)
    - data为所给样本数据，
    - specs指定范围，
    - p表示在指定范围内的概率。
- 说明
  - 该函数返回来自于估计分布的随机变量落在指定范围内的概率

例58

6.10 附加有正态密度曲线的直方图

- 函数 histfit
- 格式
  - histfit(data)
    - data为向量，返回直方图和正态曲线。
  - histfit(data,nbins)
    - nbins指定bar的个数，缺省时为data中数据个数的平方根。

例59

6.11 在指定的界线之间画正态密度曲线

- 函数 normspec
- 格式
  - p = normspec(specs,mu,sigma)
    - specs指定界线，
    - mu,sigma为正态分布的参数
    - p为样本落在上、下界之间的概率。

例60

7 参数估计

- 7.1 常见分布的参数估计
- 7.2 非线性模型置信区间预测
- 7.3 对数似然函数

7.1 常见分布的参数估计

- 命令β分布的参数a和b的最大似然估计值和置信区间
- 函数 betafit
- 格式
  - PHAT=betafit(X)
  - [PHAT,PCI]=betafit(X,ALPHA)
    - PHAT为样本X的β分布的参数a和b的估计量
    - PCI为样本X的β分布参数a和b的置信区间，是一个2×2矩阵，其第1例为参数a的置信下界和上界，第2例为b的置信下界和上界，ALPHA为显著水平，（1-α）×100%为置信度。

例61

例61

- 随机产生100个β分布数据，相应的分布参数真值为4和3。求该样本的最大似然估计值和置信度为99%的置信区间。

```
>>X = betarnd (4,3,100,1);
>>[PHAT,PCI] = betafit(X,0.01)
PHAT =
      3.9010    2.6193
PCI =
      2.5244    1.7488
      5.2776    3.4898
```

- 命令 正态分布的参数估计
- 函数 normfit
- 格式
  - [muhat,sigmahat,muci,sigmaci] = normfit(X)
  - [muhat,sigmahat,muci,sigmaci] = normfit(X,alpha)
    - muhat, sigmahat分别为正态分布的参数μ和σ的估计值，
    - muci, sigmaci分别为置信区间，其置信度为(1-α)x100%；
    - alpha给出显著水平α，缺省时默认为0.05，即置信度为95%。

例62

- 有两组(每组100个元素)正态随机数据，其均值为10，均方差为2，求95%的置信区间和参数估计值。

```
>>r = normrnd (10,2,100,2);
>>[mu,sigma,muci,sigmaci] = normfit(r)
mu =
   10.1455   10.0527        %各列的均值的估计值
sigma =
    1.9072    2.1256        %各列的均方差的估计值
muci =                      %各列均值的置信区间
    9.7652    9.6288
   10.5258   10.4766
sigmaci = =                 %各列均方差的置信区间
    1.6745    1.8663
    2.2155    2.4693
```

例63

- 分别使用金球和铂球测定引力常数
  - （1）用金球测定观察值为：6.683 6.681 6.676 6.678 6.679 6.672
  - （2）用铂球测定观察值为：6.661 6.661 6.667 6.667 6.664
- 设测定值总体为，μ和σ为未知。对(1)、(2)两种情况分别求μ和σ的置信度为0.9的置信区间。

```
X=[6.683  6.681  6.676  6.678  6.679  6.672];
Y=[6.661  6.661  6.667  6.667  6.664];
[mu,sigma,muci,sigmaci]=normfit(X,0.1)  %金球测定的估计
[MU,SIGMA,MUCI,SIGMACI]=normfit(Y,0.1)  %铂球测定的估计
```

- 命令 利用mle函数进行参数估计
- 函数 mle
- 格式
  - phat=mle('dist', X)
    - dist为分布函数名，如：beta(分布)、bino（二项分布）等，
    - X为数据样本
    - 返回用dist指定分布的最大似然估计值
  - [phat, pci]=mle ('dist', X)
    - 置信度为95%
  - [phat, pci]=mle ('dist', X, alpha)
    - 置信度由alpha确定，alpha为显著水平α，(1-α)x100%为置信度
  - [phat, pci]=mle ('dist', X, alpha, pl)
    - 仅用于二项分布，pl为试验次数。

例64

参数估计函数表

函数名	调 用 形 式	函 数 说 明
binofit	PHAT= binofit(X, N) [PHAT, PCI] = binofit(X,N) [PHAT, PCI]= binofit (X, N, ALPHA)	二项分布的概率的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的参数估计和置信区间
poissfit	Lambdahat=poissfit(X) [Lambdahat, Lambdaci] = poissfit(X) [Lambdahat, Lambdaci]= poissfit (X, ALPHA)	泊松分布的参数的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的 λ 参数和置信区间
normfit	[muhat,sigmahat,muci,sigmaci] = normfit(X) [muhat,sigmahat,muci,sigmaci] = normfit(X, ALPHA)	正态分布的最大似然估计，置信度为 95% 返回水平 α 的期望、方差值和置信区间
betafit	PHAT =betafit (X) [PHAT, PCI]= betafit (X, ALPHA)	返回 β 分布参数 a 和 b 的最大似然估计 返回最大似然估计值和水平 α 的置信区间
unifit	[ahat,bhat] = unifit(X) [ahat,bhat,ACI,BCI] = unifit(X) [ahat,bhat,ACI,BCI]=unifit(X, ALPHA)	均匀分布参数的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的参数估计和置信区间
expfit	muhat =expfit(X) [muhat,muci] = expfit(X) [muhat,muci] = expfit(X,alpha)	指数分布参数的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的参数估计和置信区间
gamfit	phat =gamfit(X) [phat,pci] = gamfit(X) [phat,pci] = gamfit(X,alpha)	γ 分布参数的最大似然估计 置信度为 95%的参数估计和置信区间 返回最大似然估计值和水平 α 的置信区间
weibfit	phat = weibfit(X) [phat,pci] = weibfit(X) [phat,pci] = weibfit(X,alpha)	韦伯分布参数的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的参数估计及其区间估计
Mle	phat = mle('dist',data) [phat,pci] = mle('dist',data) [phat,pci] = mle('dist',data,alpha) [phat,pci] = mle('dist',data,alpha,p1)	分布函数名为 dist 的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的最大似然估计值和置信区间 仅用于二项分布，pl 为试验总次数

7.2 非线性模型置信区间预测

- 命令 高斯—牛顿法的非线性最小二乘数据拟合
- 函数 nlinfit
- 格式
  - beta = nlinfit(X,y,FUN,beta0)
    - 返回在FUN中描述的非线性函数的系数。
    - FUN为用户提供形如 $\hat{y}=f(\beta,X)$ 的函数，该函数返回已给初始参数估计值β和自变量X的y的预测值。
    - 若X为矩阵，则X的每一列为自变量的取值，y是一个相应的列向量。
    - 如果FUN中使用了@，则表示函数的柄。
  - [beta,r,J] = nlinfit(X,y,FUN,beta0)
    - beta为拟合系数，r为残差，J为Jacobi矩阵，beta0为初始预测值。

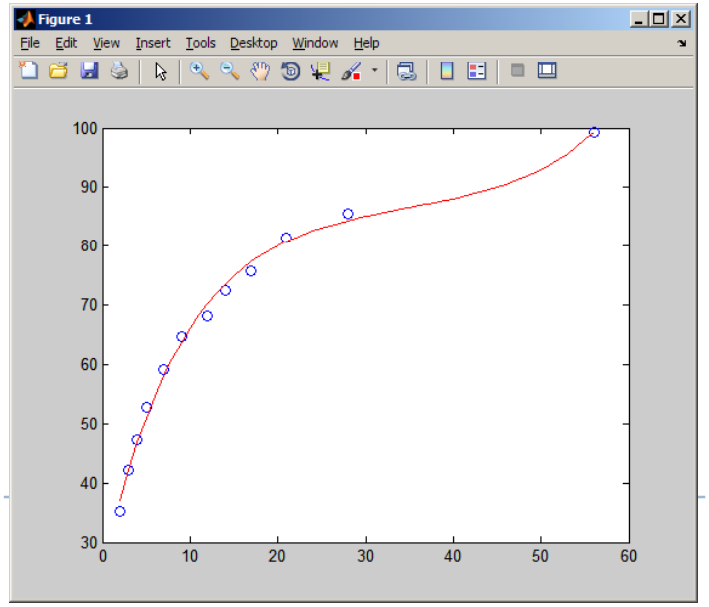
例65

- 混凝土的抗压强度随养护时间的延长而增加，现将一批混凝土作成12个试块，记录了养护日期x（日）及抗压强度y（kg/cm2）的数据：
  养护时间x=[2 3 4 5 7 9 12 14 17 21 28 56]
  抗压强度y=[35 42 47 53 59 65 68 73 76 82 86 99]+r
  r为 一个[-0.5,0.5]之间的随机向量
  建立非线性回归模型，对得到的模型和系数进行检验。
  %模型为：y=a+k1*exp(m*x)+k2*exp(-m*x);

```
clc;clear;
x=[2 3 4 5 7 9 12 14 17 21 28 56];
y1=[35 42 47 53 59 65 68 73 76 82 86 99];
r=rand(1,12)-0.5;
y=y1+r
myfunc=inline('beta(1)+beta(2)*exp(beta(4)*x)+beta(3)*exp(-beta(4)*x)','beta','x');
beta=nlinfit(x,y,myfunc,[0.5 0.5 0.5 0.5]);
a=beta(1),k1=beta(2),k2=beta(3),m=beta(4)
%test the model
xx=min(x):max(x);
yy=a+k1*exp(m*xx)+k2*exp(-m*xx);
plot(x,y,'o',xx,yy,'r')
```

结果:

```
a = 87.5244
k1 = 0.0269
k2 = -63.4591
m = 0.1083
```



- 命令 非线性模型的参数估计的置信区间
- 函数 nlparci
- 格式
  - ci = nlparci(beta,r,J)
    - 返回置信度为95%的置信区间，
    - beta为非线性最小二乘法估计的参数值，
    - r为残差，
    - J为Jacobian矩阵。
    - nlparci可以用nlinfit函数的输出作为其输入。

例66

```
clc;clear;
x=[2 3 4 5 7 9 12 14 17 21 28 56];
y1=[35 42 47 53 59 65 68 73 76 82 86 99];
r=rand(1,12)-0.5;
y=y1+r
myfunc=inline('beta(1)+beta(2)*exp(beta(4)*x)+beta(3)*exp(-beta(4)*x)','beta','x');
[beta, resids, J]=nlinfit(x,y,myfunc,[0.5 0.5 0.5 0.5]);
ci = nlparci(beta, resids, J)
```

结果:

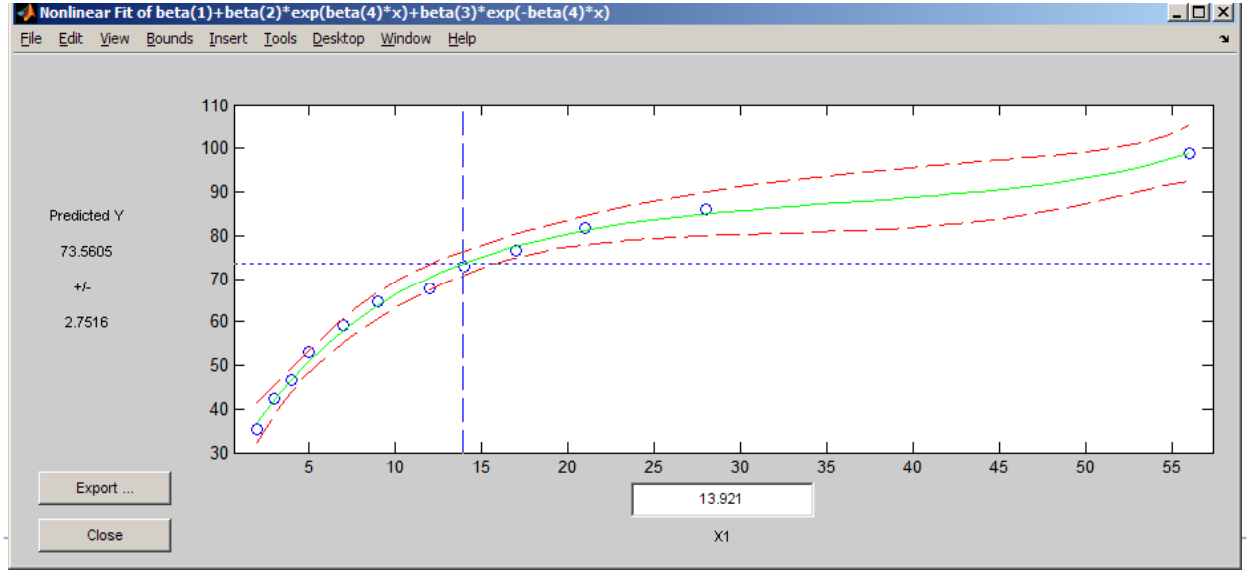
```
ci =
   82.6664   92.4917
   -0.0051    0.0618
  -67.3282  -58.1145
    0.0804    0.1336
```

```
a = 87.5244
k1 = 0.0269
k2 = -63.4591
m = 0.1083
```

- 命令 非线性拟合和显示交互图形
- 函数 nlintool
- 格式
  - nlintool(x,y,FUN,beta0)
    - 返回数据(x,y)的非线性曲线的预测图形，它用2条红色曲线预测全局置信区间。
    - beta0为参数的初始预测值，置信度为95%。
  - nlintool(x,y,FUN,beta0,alpha)
    - 置信度为(1-alpha)×100%

例67

```
clc;clear;
x=[2 3 4 5 7 9 12 14 17 21 28 56];
y1=[35 42 47 53 59 65 68 73 76 82 86 99];
r=rand(1,12)-0.5;
y=y1+r
myfunc=inline('beta(1)+beta(2)*exp(beta(4)*x)+beta(3)*exp(-beta(4)*x)','beta','x');
nlintool(x,y,myfunc,[0.5 0.5 0.5 0.5]);
%nlinfit(x,y,myfunc,[0.5 0.5 0.5 0.5]);
```



- 命令 非线性模型置信区间预测
- 函数 nlpredci
- 格式
  - ypred = nlpredci(FUN,inputs,beta,r,J)
    - ypred 为预测值，FUN与前面相同，beta为给出的适当参数，r为残差，J为Jacobian矩阵，inputs为非线性函数中的独立变量的矩阵值。
  - [ypred,delta] = nlpredci(FUN,inputs,beta,r,J)
    - delta为非线性最小二乘法估计的置信区间长度的一半，当r长度超过beta的长度并且J的列满秩时，置信区间的计算是有效的。[ypred-delta,ypred+delta]为置信度为95%的不同步置信区间。
  - ypred = nlpredci(FUN,inputs,beta,r,J,alpha,'simopt','predopt')
    - 控制置信区间的类型，置信度为100(1-alpha)%。
    - ‘simopt’ = ‘on’ 或 ‘off’ (默认值)分别表示同步或不同步置信区间。
    - ‘predopt’=‘curve’ (默认值) 表示输入函数值的置信区间，
    - 'predopt'='observation' 表示新响应值的置信区间。nlpredci可以用nlinfit函数的输出作为其输入。

例68

续前例，在[15 25 45]处的预测函数值和置信区间一半宽度

```
clc;clear;
x=[2 3 4 5 7 9 12 14 17 21 28 56];
y1=[35 42 47 53 59 65 68 73 76 82 86 99];
r=rand(1,12)-0.5;
y=y1+r
myfunc=inline('beta(1)+beta(2)*exp(beta(4)*x)+beta(3)*exp(-beta(4)*x)','beta','x');
[beta, resids, J]=nlinfit(x,y,myfunc,[0.5 0.5 0.5 0.5]);
[ypred, delta]=nlpredci(myfunc, [15 15 45], beta, resids, J)
```

```
ypred =
   75.1812   75.1812   90.2229
delta =
    1.7299
    1.7299
    4.3058
```

- 命令 有非负限制的最小二乘
- 函数 lsqnonneg
- 格式
  - x = lsqnonneg(C,d)
    - 返回在x≥0的条件下使得 ||C*x-d|| 最小的向量x
    - C和d必须为实矩阵或向量。
  - x = lsqnonneg(C,d,x0)
    - x0为初始点，x0≥0
  - x = lsqnonneg(C,d,x0,options)
    - options为指定的优化参数，参见options函数。
  - [x,resnorm] = lsqnonneg(…)
    - resnorm表示norm(C*x-d).^2的残差
  - [x,resnorm,residual] = lsqnonneg(…)
    - residual表示C*x-d的残差

例70

```
>> A =[0.0372 0.2869;
0.6861 0.7071;
0.6233 0.6245;
0.6344 0.6170];
>> b=[0.8587 0.1781 0.0747 0.8405]';
>> [x,resnorm,residual] = lsqnonneg(A,b)
x =
         0
    0.6929
resnorm =
    0.8315
residual =
    0.6599
   -0.3119
   -0.3580
    0.4130
```

7.3 对数似然函数

- 命令 负分布的对数似然函数
- 函数 Betalike
- 格式
  - logL=betalike(params,data)
    - 返回负 分布的对数似然函数，params为向量[a, b]，是 分布的参数，data为样本数据。
  - [logL,info]=betalike(params,data)
    - 返回Fisher逆信息矩阵info。如果params 中输入的参数是极大似然估计值，那么info的对角元素为相应参数的渐近方差。
- 说明
  - betalike是 分布最大似然估计的实用函数。似然函数假设数据样本中，所有的元素相互独立。因为betalike返回负 对数似然函数，用fmins函数最小化betalike与最大似然估计的功能是相同的。

例71

- 命令 负分布的对数似然估计
- 函数 Gamlike
- 格式
  - logL=gamlike(params,data)
    - 返回由给定样本数据data确定的分布的参数为params（即[a，b]）的负对数似然函数值
  - [logL,info]=gamlike(params,data)
    - 返回Fisher逆信息矩阵info。如果params中输入的参数是极大似然估计值，那么info的对角元素为相应参数的渐近方差。
- 说明
  - gamlike是分布的最大似然估计函数。因为gamlike返回对数似然函数值，故用fmins函数将gamlike最小化后，其结果与最大似然估计是相同的。

例72

- 命令 负正态分布的对数似然函数
- 函数 normlike
- 格式
  - logL=normlike(params,data)
    - 返回由给定样本数据data确定的、负正态分布的、参数为params(即[mu，sigma])的对数似然函数值。
  - [logL,info]=normlike(params,data)
    - 返回Fisher逆信息矩阵info。如果params中输入的参数是极大似然估计值，那么info的对角元素为相应参数的渐近方差。

- 命令 威布尔分布的对数似然函数
- 函数 Weiblike
- 格式

例73

  - logL = weiblike(params,data)
    - 返回由给定样本数据data确定的、威布尔分布的、参数为params(即[a，b])的对数似然函数值。
  - [logL,info]=weiblike(params,data)
    - 返回Fisher逆信息矩阵info。如果params中输入的参数是极大似然估计值，那么info的对角元素为相应参数的渐近方差。
- 威布尔分布的负对数似然函数定义为

$$-\log L = -\log \prod_{i=1}^{n} f(a,b \mid x_i) = -\sum_{i=1}^{n} \log f(a,b \mid x_i)$$

8 假设检验

- 8.1 已知，单个正态总体的均值μ的假设检验（U检验法）
- 8.2 未知，单个正态总体的均值μ的假设检验(t检验法)
- 8.3 两个正态总体均值差的检验（t检验）
- 8.4 两个总体一致性的检验——秩和检验
- 8.5 两个总体中位数相等的假设检验——符号秩检验
- 8.6 两个总体中位数相等的假设检验——符号检验
- 8.7 正态分布的拟合优度测试
- 8.8 正态分布的拟合优度测试
- 8.9 单个样本分布的 Kolmogorov-Smirnov 测试
- 8.10 两个样本具有相同的连续分布的假设检验

8.1 σ^2 已知，单个正态总体的均值μ的假设检验（U检验法）

- 函数 ztest
  - h = ztest(x,m,sigma)
    - x为正态总体的样本，m为均值，sigma为标准差，显著性水平为0.05(默认值)
  - h = ztest(x,m,sigma,alpha)
    - 显著性水平为alpha
  - [h,sig,ci,zval] = ztest(x,m,sigma,alpha,tail)
    - sig为观察值的概率，当sig为小概率时则对原假设提出质疑，
    - ci为真正均值μ的1-alpha置信区间，zval为统计量的值。
    - 若h=0，表示在显著性水平alpha下，不能拒绝原假设；
    - 若h=1，表示在显著性水平alpha下，可以拒绝原假设。

  - 原假设
    - 若tail=0，表示备择假设：$H_0: \mu_1 \neq \mu_0 =$（默认，双边检验）；
    - tail=1，表示备择假设：$H_1: \mu_1 > \mu_0 =$（单边检验）；
    - tail=-1，表示备择假设：$H_2: \mu_1 < \mu_0 = m$（单边检验）。

例74

- 某车间用一台包装机包装葡萄糖，包得的袋装糖重是一个随机变量，它服从正态分布。当机器正常时，其均值为0.5公斤，标准差为0.015。某日开工后检验包装机是否正常，随机地抽取所包装的糖9袋，称得净重为（公斤）0.497, 0.506, 0.518, 0.524, 0.498, 0.511, 0.52, 0.515, 0.512；问机器是否正常？
- 总体μ和σ已知，该问题是当σ^2为已知时，在水平 $\alpha = 0.05$ 下，根据样本值判断μ=0.5还是μ!=0.5 。
- 原假设：$H_0: \mu = \mu_0 = 0.5$ 备择假设：$H_1: \mu \neq 0.5$

```
X=[0.497,0.506,0.518,0.524,0.498,0.511,0.52,0.515,0.512];
[h,sig,ci,zval]=ztest(X,0.5,0.015,0.05,0)
```

8.2 σ^2未知，单个正态总体的均值μ的假设检验(t检验法)

- 函数 ttest
- 格式
  - h = ttest(x,m)
    - x为正态总体的样本，m为均值μ_0，显著性水平为0.05
    - 若h=0，表示在显著性水平alpha下，不能拒绝原假设；
    - 若h=1，表示在显著性水平alpha下，可以拒绝原假设。
  - h = ttest(x,m,alpha)
    - alpha为给定显著性水平
  - [h,sig,ci] = ttest(x,m,alpha,tail)
    - sig为观察值的概率，当sig为小概率时则对原假设提出质疑，
    - ci为真正均值μ的1-alpha置信区间。
- 原假设
  - tail=0表示备择假设：$H_0: \mu = \mu_0 = m$ （默认，双边检验）；
  - tail=1，表示备择假设：$H_0: \mu_1 \neq \mu_0 = m$ （单边检验）；
  - tail=-1，表示备择假设：$H_1: \mu_1 > \mu_0 = m$ （单边检验）。

$$H_2: \mu_1 < \mu_0 = m$$

- 某种电子元件的寿命X（以小时计）服从正态分布，u, σ^2均未知。现测得16只元件的寿命如下
  159 280 101 212 224 379 179 264 222 362 168 250 149 260 485 170
  问是否有理由认为元件的平均寿命大于225（小时）？
- σ^2未知，在 水平下检验假设:

$$\alpha = 0.05$$

$$H_0 : \mu < \mu_0 = 225 \qquad H_1 : \mu > 225$$

```
>> X=[159 280 101 212 224 379 179 264 222 362 168 250 149 260 485 170];
>> [h,sig,ci]=ttest(X,225,0.05,1)
```

8.3 两个正态总体均值差的检验（t检验）

- 两个正态总体方差未知但等方差时，比较两正态总体样本均值的假设检验
- 函数 ttest2
  - [h,sig,ci]=ttest2(X,Y)
    - X，Y为两个正态总体的样本，显著性水平为0.05
    - 若h=0，表示在显著性水平alpha下，不能拒绝原假设；
    - 若h=1，表示在显著性水平alpha下，可以拒绝原假设。
  - [h,sig,ci]=ttest2(X,Y,alpha)
  - [h,sig,ci]=ttest2(X,Y,alpha,tail)
    - sig为当原假设为真时得到观察值的概率，当sig为小概率时则对原假设提出质疑，ci为真正均值μ的1-alpha置信区间。
- 原假设：　　，　(为X为期望值，为Y的期望值)
  - tail=0，表示备择假设：　　（默认，双边检验）；
  - tail=1，表示备择假设：　　（单边检验）；
  - tail=-1，表示备择假设：　　（单边检验）。

$$H_0: \mu_1 = \mu_2$$

$$H_0: \mu_1 = \mu_2$$

$$H_0: \mu_1 > \mu_2$$

$$H_1: \mu_1 < \mu_2$$

例76

- 在平炉上进行一项试验以确定改变操作方法的建议是否会增加钢的产率，试验是在同一只平炉上进行的。每炼一炉钢时除操作方法外，其他条件都尽可能做到相同。先用标准方法炼一炉，然后用建议的新方法炼一炉，以后交替进行，各炼10炉，其产率分别为

（1）标准方法：78.1 72.4 76.2 74.3 77.4 78.4 76.0 75.5 76.7 77.3

（2）新方法： 79.1 81.0 77.3 79.1 80.0 79.1 79.1 77.3 80.2 82.1

设这两个样本相互独立，且分别来自正态总体 $N(\mu_1,\sigma^2)$ 和 $N(\mu_2,\sigma^2)$，μ_1,μ_2,σ^2 均未知。问建议的新操作方法能否提高产率？（取α=0.05）

- 两个总体方差不变时，在α=0.05水平下检验假设：

$$H_0:\mu_1=\mu_2 \qquad H_1:\mu_1<\mu_2$$

```
>> X=[78.1  72.4  76.2  74.3  77.4  78.4  76.0  75.5  76.7  77.3];
>>Y=[79.1  81.0  77.3  79.1  80.0  79.1  79.1  77.3  80.2  82.1];
>> [h,sig,ci]=ttest2(X,Y,0.05,-1)
```

8.4 两个总体一致性的检验——秩和检验

- 函数 ranksum
- 格式
  - p = ranksum(x,y,alpha)
    - x、y为两个总体的样本，可以不等长，
    - alpha为显著性水平
    - P为两个总体样本X和Y为一致的显著性概率，若P接近于0，则不一致较明显。
  - [p,h] = ranksum(x,y,alpha)
    - h为检验结果，h=0表示X与Y的总体差别不显著；
    - h=1表示X与Y的总体差别显著
  - [p,h,stats] = ranksum(x,y,alpha)
    - stats中包括：ranksum为秩和统计量的值以及zval为过去计算p的正态统计量的值

例77

- 某商店为了确定向公司A或公司B购买某种商品，将A和B公司以往的各次进货的次品率进行比较，数据如下：
  A：7.0 3.5 9.6 8.1 6.2 5.1 10.4 4.0 2.0 10.5
  B：5.7 3.2 4.1 11.0 9.7 6.9 3.6 4.8 5.6 8.4 10.1 5.5 12.3
  设两样本独立。问两公司的商品的质量有无显著差异，取α=0.05。
- 设 ， 别为A、B两个公司的商品次品率总体的均值。在水平α=0.05下检验假设:

μ_A μ_B

$$H_0 : \mu_A = \mu_B \quad H_1 : \mu_A \neq \mu_B$$

```
>> A=[7.0 3.5 9.6 8.1 6.2 5.1 10.4 4.0 2.0 10.5];
>> B=[5.7 3.2 4.1 11.0 9.7 6.9 3.6 4.8 5.6 8.4 10.1 5.5 12.3];
>> [p,h,stats]=ranksum(A,B,0.05)
```

8.5 两个总体中位数相等的假设检验——符号秩检验

- 函数 signrank
- 格式
  - p = signrank(X,Y,alpha)
    - X、Y为两个总体的样本，长度必须相同，
    - alpha为显著性水平，
    - P两个样本X和Y的中位数相等的概率，p接近于0则可对原假设质疑。
  - [p,h] = signrank(X,Y,alpha)
    - h为检验结果：h=0表示X与Y的中位数之差不显著，h=1表示X与Y的中位数之差显著。
  - [p,h,stats] = signrank(x,y,alpha)
    - stats中包括：signrank为符号秩统计量的值以及zval为过去计算p的正态统计量的值。

例78

8.6 两个总体中位数相等的假设检验——符号检验

- 函数 signtest
- 格式
  - p=signtest(X,Y, alpha)
    - X、Y为两个总体的样本，长度必须相同，
    - alpha为显著性水平，
    - P两个样本X和Y的中位数相等的概率，p接近于0则可对原假设质疑。
  - [p, h]=signtest(X,Y, alpha)
    - h为检验结果：h=0表示X与Y的中位数之差不显著，h=1表示X与Y的中位数之差显著。
  - [p,h,stats] = signtest(X,Y,alpha)
    - stats中sign为符号统计量的值

例79

8.7 正态分布的拟合优度测试

- 函数 jbtest
- 格式
  - H = jbtest(X)
    - 对输入向量X进行Jarque-Bera测试，显著性水平为0.05。
    - H为测试结果，若H=0，则可以认为X是服从正态分布的；若H=1，则可以否定X服从正态分布。
    - X为大样本，对于小样本用lillietest函数
  - H = jbtest(X,alpha)
    - 在水平alpha下施行 Jarque-Bera 测试，alpha在0和1之间。
  - [H,P,JBSTAT,CV] = jbtest(X,alpha)
    - P为接受假设的概率值，P越接近于0，则可以拒绝是正态分布的原假设；
    - JBSTAT为测试统计量的值，CV为是否拒绝原假设的临界值。

例80

8.8 正态分布的拟合优度测试

- 函数 lillietest
- 格式
  - H = lillietest(X)
    - 对输入向量X进行Lilliefors测试，显著性水平为0.05。例81
    - H为测试结果，若H=0，则可以认为X是服从正态分布的；若X=1，则可以否定X服从正态分布。
  - H = lillietest(X,alpha)
    - 在水平alpha而非5%下施行Lilliefors测试，alpha在0.01和0.2之间。
  - [H,P,LSTAT,CV] = lillietest(X,alpha)
    - P为接受假设的概率值，P越接近于0，则可以拒绝是正态分布的原假设；
    - LSTAT为测试统计量的值，CV为是否拒绝原假设的临界值。

8.9 单个样本分布的 Kolmogorov-Smirnov 测试

- 函数 kstest
- 格式
  - H = kstest(X)
    - 测试向量X是否服从标准正态分布，测试水平为5%。
    - 原假设为X服从标准正态分布。若H=0则不能拒绝原假设，H=1则可以拒绝原假设。
  - H = kstest(X,cdf)
    - 指定累积分布函数为cdf的测试(cdf=[ ]时表示标准正态分布)，测试水平为5%
  - H = kstest(X,cdf,alpha)
    - alpha为指定测试水平
  - [H,P,KSSTAT,CV] = kstest(X,cdf,alpha)
    - P为原假设成立的概率，KSSTAT为测试统计量的值，CV为是否接受假设的临界值。

例82

8.10 两个样本具有相同的连续分布的假设检验

- 函数 kstest2
- 格式
  - H = kstest2(X1,X2)
    - 测试向量X1与X2是具有相同的连续分布，测试水平为5%。
    - 原假设为具有相同连续分布。测试结果为H，若H=0，表示应接受原假设；若H=1，表示可以拒绝原假设。这是Kolmogorov-Smirnov测试方法。
  - H = kstest2(X1,X2,alpha)
    - alpha为测试水平
  - [H,P,KSSTAT] = kstest(X,cdf,alpha)
    - 与指定累积分布cdf相同的连续分布，
    - P为假设成立的概率，
    - KSSTAT为测试统计量的值。

例8.3

9 方差分析

- 9.1 单因素方差分析
- 9.2 双因素方差分析

9.1 单因素方差分析

- 单因素方差分析是比较两组或多组数据的均值，它返回原假设——均值相等的概率
- 函数 anova1
  - p = anova1(X)
    - X的各列为彼此独立的样本观察值，其元素个数相同，p为各列均值相等的概率值，若p值接近于0，则原假设受到怀疑，说明至少有一列均值与其余列均值有明显不同。
    - anova1函数产生两个图：标准的方差分析表图和盒图。
  - p = anova1(X,group)
    - X和group为向量且group要与X对应
  - p = anova1(X,group,'displayopt')
    - displayopt=on/off表示显示与隐藏方差分析表图和盒图
  - [p,table] = anova1(…)　% table为方差分析表
  - [p,table,stats] = anova1(…)　% stats为分析结果的构造

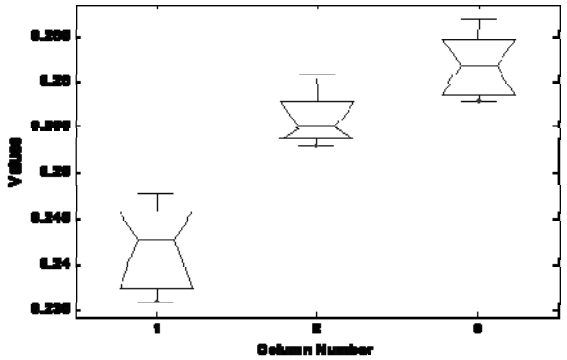
例84

- 设有3台机器，用来生产规格相同的铝合金薄板。取样测量薄板的厚度，精确至‰厘米。得结果如下：
  机器1：0.236 0.238 0.248 0.245 0.243
  机器2：0.257 0.253 0.255 0.254 0.261
  机器3：0.258 0.264 0.259 0.267 0.262
  检验各台机器所生产的薄板的厚度有无显著的差异？

```
>> X=[0.236  0.238  0.248  0.245  0.243; 0.257  0.253  0.255  0.254  0.261;
0.258  0.264  0.259  0.267  0.262];
>> P=anova1(X')
```

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	0.00105	2	0.00053	32.92	1.34305e-005
Error	0.00019	12	0.00002		
Total	0.00125	14			



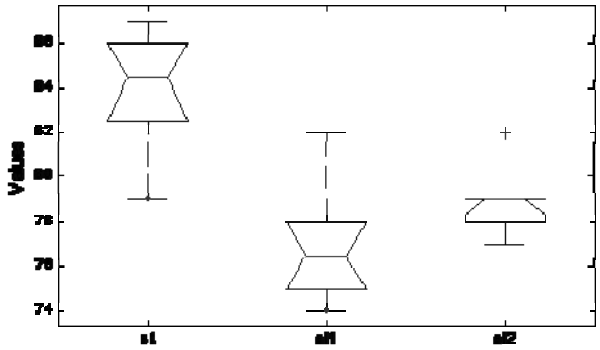
例85

- 建筑横梁强度的研究：3000磅力量作用在一英寸的横梁上来测量横梁的挠度，三种材料的测试强度是
  钢筋：82 86 79 83 84 85 86 87；
  合金1：74 82 78 75 76 77；
  合金2：79 79 77 78 82 79
  检验这些合金强度有无明显差异？

```
>> strength = [82 86 79 83 84 85 86 87 74 82 78 75 76 77 79 79 77 78 82 79];
>>alloy = {'st','st','st','st','st','st','st','st', 'al1','al1','al1','al1','al1','al1',
                'al2','al2','al2','al2','al2','al2'};
>> [p,table,stats] = anova1(strength,alloy,'on')
```

ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	184.8	2	92.4	15.4	0.0002
Error	102	17	6		
Total	286.8	19			



9.2 双因素方差分析

- 函数 anova2
- 格式
  - p = anova2(X,reps)
  - p = anova2(X,reps,'displayopt')
  - [p,table] = anova2(…)
  - [p,table,stats] = anova2(…)

$$\begin{array}{cc} A=1 & A=2 \\ \end{array}$$
$$\begin{bmatrix} x_{111} & x_{112} \\ x_{121} & x_{122} \\ x_{211} & x_{212} \\ x_{221} & x_{222} \\ x_{311} & x_{312} \\ x_{321} & x_{322} \end{bmatrix} \begin{array}{l} \}B=1 \\ \\ \}B=2 \\ \\ \}B=3 \\ \\ \end{array}$$

- 说明
  - 执行平衡的双因素试验的方差分析来比较X中两个或多个列（行）的均值，不同列的数据表示因素A的差异，不同行的数据表示另一因素B的差异。如果行列对有多于一个的观察点，则变量reps指出每一单元观察点的数目，每一单元包含reps行，如右上图，则reps=2
  - 其余参数与单因素方差分析参数相似。

例86

- 一火箭使用了4种燃料，3种推进器作射程试验，每种燃料与每种推进器的组合各发射火箭2次，得到结果如下：

推进器（B）		B1	B2	B3
燃料A	A1	58.2000	56.2000	65.3000
		52.6000	41.2000	60.8000
	A2	49.1000	54.1000	51.6000
		42.8000	50.5000	48.4000
	A3	60.1000	70.9000	39.2000
		58.3000	73.2000	40.7000
	A4	75.8000	58.2000	48.7000
		71.5000	51.0000	41.4000

- 考察推进器和燃料这两个因素对射程是否有显著的影响？