



石家莊鐵道大學  
SHIJIAZHUANG TIEDAO UNIVERSITY

在线开放课程

MATLAB数据分析

数据统计

主讲：卞建鹏

# 1、最大值和最小值

## (1) 求向量的最大值和最小值

(1)  $y=\max(x)$ : 返回向量 $x$ 的最大值存入 $y$ , 如果 $x$ 中包含复数元素, 则按模取最大值。

(2)  $[y,I]=\max(x)$ : 返回向量 $x$ 的最大值存入 $y$ , 最大值的序号存入 $I$ , 如果 $x$ 中包含复数元素, 则按模取最大值。

```
命令行窗口
>> x=[1 3 2 5]
x =
     1     3     2     5
>> [y,I]=max(x)
y =
     5
I =
     4
```

# 1、最大值和最小值

## (2) 求矩阵的最大值和最小值

(1)  $\max(A)$ : 返回一个行向量，向量的第  $i$  个元素是矩阵  $A$  的第  $i$  列上的最大值。

(2)  $[Y,U]=\max(A)$ : 返回行向量  $Y$  和  $U$ ， $Y$  向量记录  $A$  的每列的最大值， $U$  向量记录每列最大值的行号。

(3)  $\max(A,[],dim)$ :  $dim$  取 1 或 2。 **$dim$  取 1 时，该函数和  $\max(A)$  完全相同**； $dim$  取 2 时，该函数返回一个列向量，其第  $i$  个元素是  $A$  矩阵的第  $i$  行上的最大值。

# 1、最大值和最小值

```
A=[1 2 3;4 2 4;5 2 6];
```

```
[Y,U]=max(A)
```

```
max(A,[],2)
```

```
max(max(A))
```

```
A =  
  
     1     2     3  
     4     2     4  
     5     2     6
```

```
>> max(A, [], 2)
```

```
ans =
```

```
     3
```

```
     4
```

```
     6
```

```
>> max(max(A))
```

```
ans =
```

```
     6
```

```
>> [Y,U]=max(A)
```

```
Y =
```

```
     5     2     6
```

```
U =
```

```
     3     1     3
```

# 1、最大值和最小值

## (3) 两个向量或矩阵对应元素的比较

(1)  $U=\max(A,B)$ : A,B是两个同型的向量或矩阵, 结果U是与A, B同型的向量或矩阵, U的每个元素等于A, B对应元素的较大者。

(2)  $U=\max(A,n)$ : n是一个标量, 结果U是与A同型的向量或矩阵, U的每个元素等于A对应元素和n中的较大者。

```
命令行窗口
>> x=[1 3 2 5];y=[2 1 4 5];
>> U=max(x, y)
U =
     2     3     4     5
```

## 2、和与积

设 $x$ 是一个向量， $A$ 是一个矩阵：

`sum(x)`：返回向量 $x$ 各元素的和。

`prod(x)`：返回向量 $x$ 各元素的乘积。

`sum(A)`：返回一个行向量，其第 $i$ 个元素是 $A$ 的第 $i$ 列的元素和。

`prod(A)`：返回一个行向量，其第 $i$ 个元素是 $A$ 的第 $i$ 列的元素乘积

```
>> x=[1 2 3;0 1 4]
```

```
x =
```

```
    1    2    3
```

```
    0    1    4
```

```
>> sum(x)
```

```
ans =
```

```
    1    3    7
```

```
>> prod(x)
```

```
ans =
```

```
    0    2   12
```

# 3、平均值和中值

mean(x): 返回向量x的算术平均值。

median(x): 返回向量x的中值。

mean(A): 返回一个行向量，其第i个元素是A的第i列的算术平均值

median(A): 返回一个行向量，其第i个元素是A的第i列的中值。

```
命令行窗口
>> x=[1 2 3;0 1 4]
x =
     1     2     3
     0     1     4
>> mean(x)
ans =
    0.5000    1.5000    3.5000
>> median(x)
ans =
    0.5000    1.5000    3.5000
```

## 4、累加和与累乘积

`cumsum(x)`: 返回向量 $x$ 累加和向量。

`cumprod(x)`: 返回向量 $x$ 累乘积向量。

`cumsum(A)`: 返回一个矩阵，其第 $i$ 列是 $A$ 的第 $i$ 列的累加和向量。

`cumprod(A)`: 返回一个矩阵，其第 $i$ 列是 $A$ 的第 $i$ 列的累乘积向量。

```
命令窗口
>> x=[1 2 3;0 1 4]
x =
     1     2     3
     0     1     4
>> cumsum(x)
ans =
     1     2     3
     1     3     7
>> cumprod(x)
ans =
     1     2     3
     0     2    12
```



# 5、排序

对向量x中的元素按升序排列。也可以对矩阵A的各列或各行重新排序，其调用格式为：

**[Y,I]=sort(A,dim)**

其中dim指明对A的列还是行进行排序。若dim=1，则按列排；若dim=2，则按行排。Y是排序后的矩阵，而I记录Y中的元素在A中位置。

A =

1	2	3
4	2	4
5	2	6

```
>> [Y,I]=sort(A,2)
```

Y =

1	2	3
2	4	4
2	5	6

I =

1	2	3
2	1	3
2	1	3

## 6、标准方差

对于向量x，std(x)返回一个标准方差。对于矩阵A，std(A)返回一个行向量，它的各个元素便是矩阵A各列或各行的标准方差。

**Y=std(A,flag,dim)**

其中dim取1或2。当dim=1时，求各列元素的标准方差；当dim=2时，则求各行元素的标准方差。flag取0或1，当flag=0时，按 $\sigma_1$ 所列公式计算标准方差，当flag=1时，按 $\sigma_2$ 所列公式计算标准方差。缺省flag=0，dim=1。

$$\sigma_1 = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\sigma_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

```
x =  
    1    2    3  
    0    1    4  
>> std(x)  
ans =  
    0.7071    0.7071    0.7071
```

# 7、相关系数

`corrcoef(x,y)`: x,y是向量。

`corrcoef(A)`: 返回从矩阵A形成的一个相关系数矩阵。此相关系数矩阵的大小与矩阵A一样。它把矩阵A的每列作为一个变量，然后求它们的相关系数。

```
A = 1.0000  0.2000  0.5000  
     1.2000  0.2500  0.5100  
     1.0200  0.2100  0.5100
```

```
>> corrcoef(A)
```

```
ans = 1.0000  0.9951  0.5766  
      0.9951  1.0000  0.6547  
      0.5766  0.6547  1.0000
```

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

例：生成满足正态分布的 $10000 \times 5$ 随机矩阵，然后求各列元素的均值和标准方差，再求这5列随机数据的相关系数矩阵。



在线开放课程

```
X=randn(10000,5); M=mean(X); D=std(X);
```

```
R=corrcoef(X)
```

```
M = 0.0011    0.0066    0.0009    0.0264    0.0101
```

```
D = 1.0011    1.0036    1.0049    1.0058    1.0061
```

```
R = 1.0000    0.0119    0.0051   -0.0114   -0.0011
```

```
    0.0119    1.0000    0.0093   -0.0012    0.0071
```

```
    0.0051    0.0093    1.0000    0.0048    0.0095
```

```
   -0.0114   -0.0012    0.0048    1.0000   -0.0017
```

```
   -0.0011    0.0071    0.0095   -0.0017    1.0000
```

# 8、常用描述性统计量函数

函数名	说 明	函数名	说 明
corr	线性（或秩）相关系数	moment	中心矩
corrcoef	线性相关系数	prctile	百分位数
cov	协方差矩阵	quantile	分位数
iqr	内4分位极差	range	极差
kurtosis	峰度	skewness	偏度
mean	均值	std	标准差
median	中位数	tabulate	频率分布表
mode	众数	var	方差

# 9、常用统计绘图函数

函数名	说 明	函数名	说 明
boxplot	箱线图	ksdensity	核密度估计图
cdfplot	分布函数图	lsline	添加最小二乘拟合线
ecdf	分布函数图	normplot	正态概率图
ecdfhist	频率直方图	parallelcoords	多元数据的平行坐标图
gline	添加参考线	probplot	概率图
glyphplot	星图或脸谱图	qqplot	Q-Q图（分位数图）
gplotmatrix	散点图矩阵	refcurve	添加参考多项式曲线
gscatter	分组散点图	refline	添加参考线
hist	二维直方图	scatterhist	二维散点图和边缘直方图
hist3	三维直方图	wblplot	威布尔分布概率图

# 9、常用统计绘图函数

用normrnd函数产生1000个标准正态分布随机数，并做出频数直方图和经验分布函数图。

```
>> x = normrnd(0, 1, 1000, 1);
```

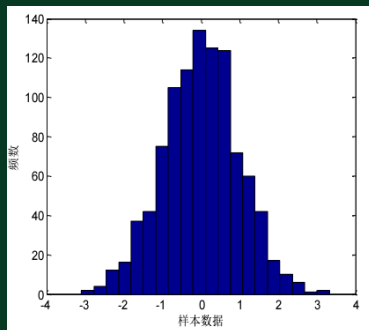
```
>> hist(x, 20);
```

```
>> xlabel('样本数据');
```

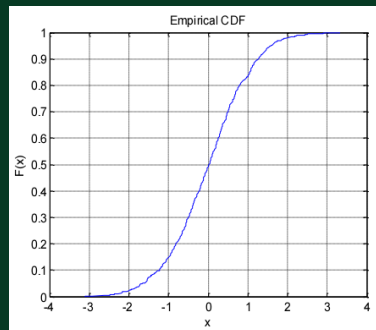
```
>> ylabel('频数');
```

```
>> figure;
```

```
>> cdfplot(x);
```



频数直方图



经验分布函数图

# 10、统计函数的应用

```
>> VC =load ('data.txt') ;  
>> figure; % 新建图形窗口  
>> [f, xc] = ecdf(VC); % 调用ecdf函数计算  
%xc处的经验分布函数值f  
>> ecdfhist(f, xc); % 绘制频率直方图  
  
>> x = linspace(1000,8000,30); % 产生一个新的横坐标向量x  
>> y = normpdf(x,mean(VC),std(VC));  
% 计算正态分布在向量x处的密度函数值
```

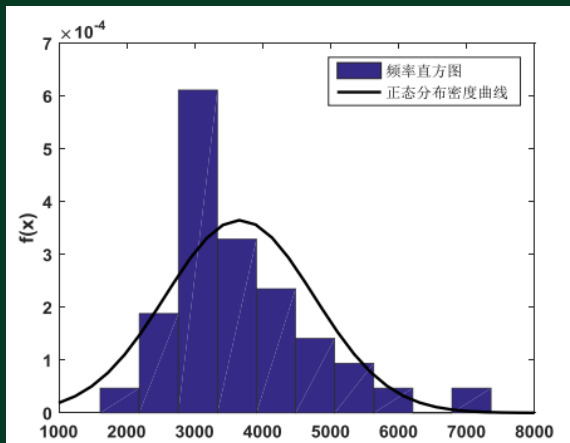


# 10、统计函数的应用

```
>> hold on
```

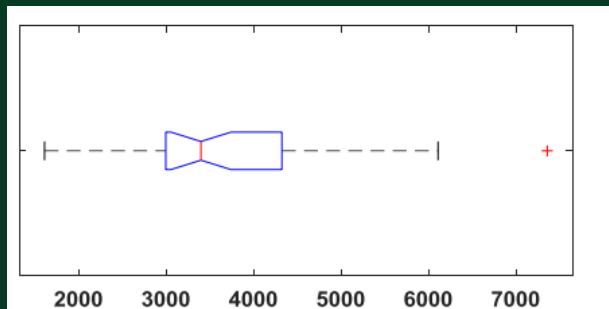
```
>> plot(x,y,'k','LineWidth',2) % 绘制正态分布的密度函数曲线，  
                                并设置线条为黑色实线，线宽为2
```

```
>> legend('频率直方图','正态分布密度曲线','Location','NorthEast');
```



# 10、统计函数的应用

```
>> figure; % 新建图形窗口  
% 绘制带有刻槽的水平箱线图  
>> boxplot(VC,boxlabel,'notch','on','orientation','horizontal')
```

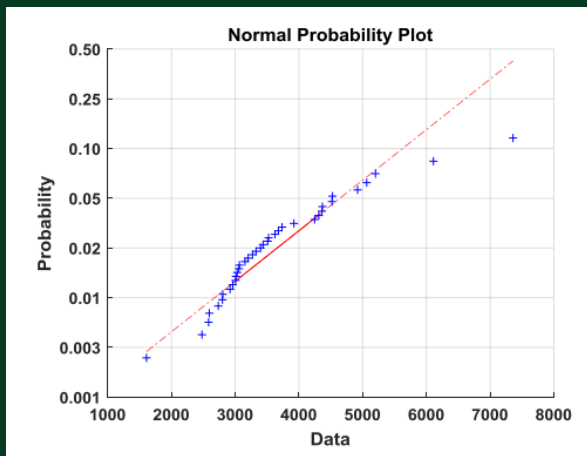


# 10、统计函数的应用

## 绘制正态概率图

```
>> figure;
```

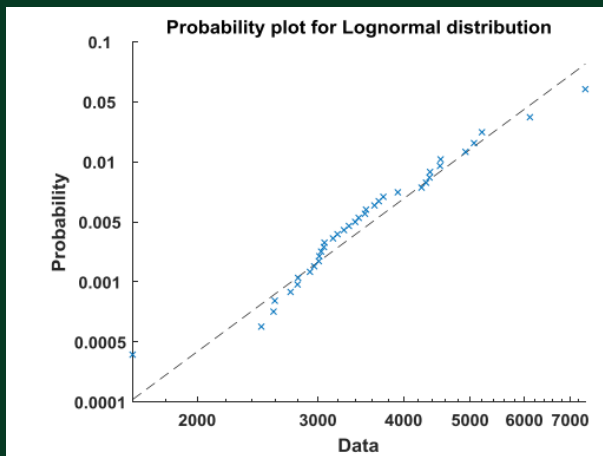
```
>> normplot(VC);
```



## 绘制对数正态概率图

```
>> figure;
```

```
>> probplot('lognormal',VC)
```



# 10、统计函数的应用

## 频数和频率分布表

```
>> x = [2 2 6 5 2 3 2 4 3 4 3 4 4 4 4 2 2  
6 0 4 7 2 5 8 3 1 3 2 5 3 6 2 3 5  
4 3 1 4 2 2 2 3 1 5 2 6 3 4 1 2 5];  
>> tabulate(x(:))
```

Value	Count	Percent
0	1	1.96%
1	4	7.84%
2	14	27.45%
3	10	19.61%
4	10	19.61%
5	6	11.76%
6	4	7.84%
7	1	1.96%
8	1	1.96%

# 10、统计函数的应用

统计字符串中各字符出现的频数、频率

```
>> x = ['If x is a numeric array, TABLE is a numeric matrix.'];
```

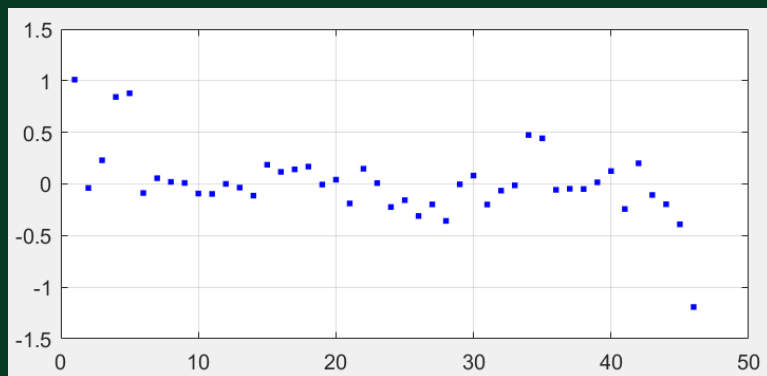
```
>> tabulate(x)
```

Value	Count	Percent
I	1	2.44%
f	1	2.44%
x	2	4.88%
i	5	12.20%
s	2	4.88%
a	5	12.20%
n	2	4.88%
u	2	4.88%
...		

# 10、统计函数的应用

```
a=[1.0100 -0.0393 0.2301 0.8424 0.8792 -0.0879 0.0552 0.0204  
0.0101 -0.0922 -0.0971 0.0009 -0.0356 -0.1138 0.1867 0.1169 0.1410  
0.1687 -0.0061 0.0422 -0.1884 0.1478 0.0077 -0.2228 -0.1573 -  
0.3110 -0.1979 -0.3573 -0.0028 0.0815 -0.1985 -0.0650 -0.0129 0.4747  
0.4424 -0.0576 -0.0463 -0.0499 0.0159 0.1248 -0.2422 0.2009 -  
0.1074 -0.1960 -0.3907 -1.1918];
```

```
n=length(a)  
plot(1:n,a,'b.')
```



# 10、统计函数的应用

`figure(1); hist(a);` %作频数直方图

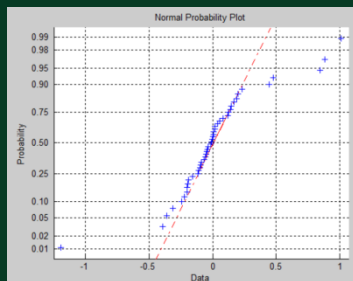
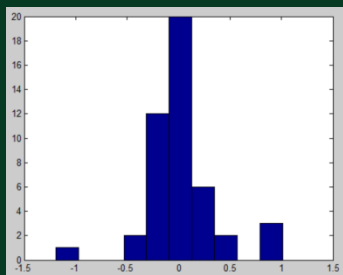
`figure(2);`

`normplot(a);` %分布的正态性检验

`[muhat,sigmahat,muci,sigmaci]= normfit(a)`

%参数估计 均值、方差、均值的0.95置信区间，方差的0.95置信区间

`[h,sig,ci]= ttest(a,muhat)` %假设检验



# 10、统计函数的应用

**muhat = 0.0159**

**sigmahat = 0.3432**

**muci =**

**-0.0860**

**0.1178**

**sigmaci =**

**0.2847**

**0.4323**

**h = 0**

**sig = 1**

**ci = -0.0860 0.1178**

①布尔变量 $h=0$ ，表示不拒绝零假设，说明提出的假设“均值 0.0159”是合理的；

②95%的置信区间为  $[-0.0860 \quad 0.1178]$ ，它完全包括 0.0159，且精度很高；

③sig 的值为1，远超过0.5，不能拒绝零假设。



# 小结

1. 最大值和最小值、和与积、累加和与累乘积、排序、标准方差
2. 相关系数
3. 常用描述性统计量函数
4. 常用统计绘图函数
5. 统计函数的应用