常用统计量

实验目的

学习常用统计量MatLab软件计算的基本过程与方法

- 平均值、中位数(mean,median)
- 方差，标准差，极差(var, std, range)
- 偏度(skewness)
- 峰度(kurtosis)
- 直方图(hist, histfit, cdfplot)

常用统计量

在对数据进行深入分析之前，首先需要将已有数据以一定的格式读取，并利用常用的统计量进行概括性分析.

这些常用统计量包括: 平均值、中位数、方差、标准差、极差、偏度、峰度等.

下面介绍如何计算上述统计量.

常用统计量

平均值与中位数

平均值又称均值、数学期望. 设 $X_1, X_2, \cdots, X_n$ 为一串数据

平均值即: $\frac{1}{n}\sum_{i=1}^{n} X_i$ 计算命令为 **mean(x).**

中位数则是将一串数据从小到大排列后，位于中间位置的数据.若数据有偶数个，中位数是指位于中间两个数字的平均值. 计算命令为 **median(x).**

常用统计量

若x为向量，无论行向量还是列向量，mean(x)和median(x)输出的都是该组数字的平均值和中位数.

若x为矩阵，mean(x)和median(x)输出的是行向量，每个元素均为x中对应列的平均值、中位数.

接下来要介绍的各个统计量， 其输出规则与此相同.

例. 产生服从N(3,9)分布的100个随机数，计算其平均值和中位数.

```
r = normrnd(3,3,1,100);   mean(r)     median(r)
```

常用统计量

方差、标准差与极差

描述数据分散程度的统计量, 主要有方差、标准差和极差.

设一组数据 $X_1, X_2, \cdots, X_n$ 的平均值为 $\bar{X}$

方差的定义为: $\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ 计算命令为 **var(x).**

标准差的定义为方差的平方根, 计算命令为 **std(x).**

极差为最大值与最小值之差, 计算命令为 **range(x).**

常用统计量

方差、标准差与极差

例. 产生服从E(5)分布的1000个随机数，计算其平均值、方差、标准差和极差.

```
x=exprnd(5,1000,1);
mean(x),  var(x),   std(x),  range(x)
```

常用统计量

例. 现有一组学生的期末考试成绩. 分别计算其平均值、中位数、方差、标准差、极差.

```
>> x=[89,90;78,98;64,80;90,85;
      55,65;73,89;21,60;97,80]
```

>> mean(x)	平均值	70.8	80.8
>> median(x)	中位数	75.5	82.5
>> var(x)	方差	602.7	164.1
>> std(x)	标准差	24.5	12.8
>> range(x)	极差	76	38

学生编号	高等数学	大学物理
1	89	90
2	78	98
3	64	80
4	90	85
5	55	65
6	73	89
7	21	60
8	97	80

常用统计量

偏度

偏度是描述数据分布非对称程度的数字特征，反映的是数据分布偏斜方向和程度。

设一组数据 $X_1, X_2, \cdots, X_n$ 的平均值为 $\bar{X}$ ，其偏度计算公式为：

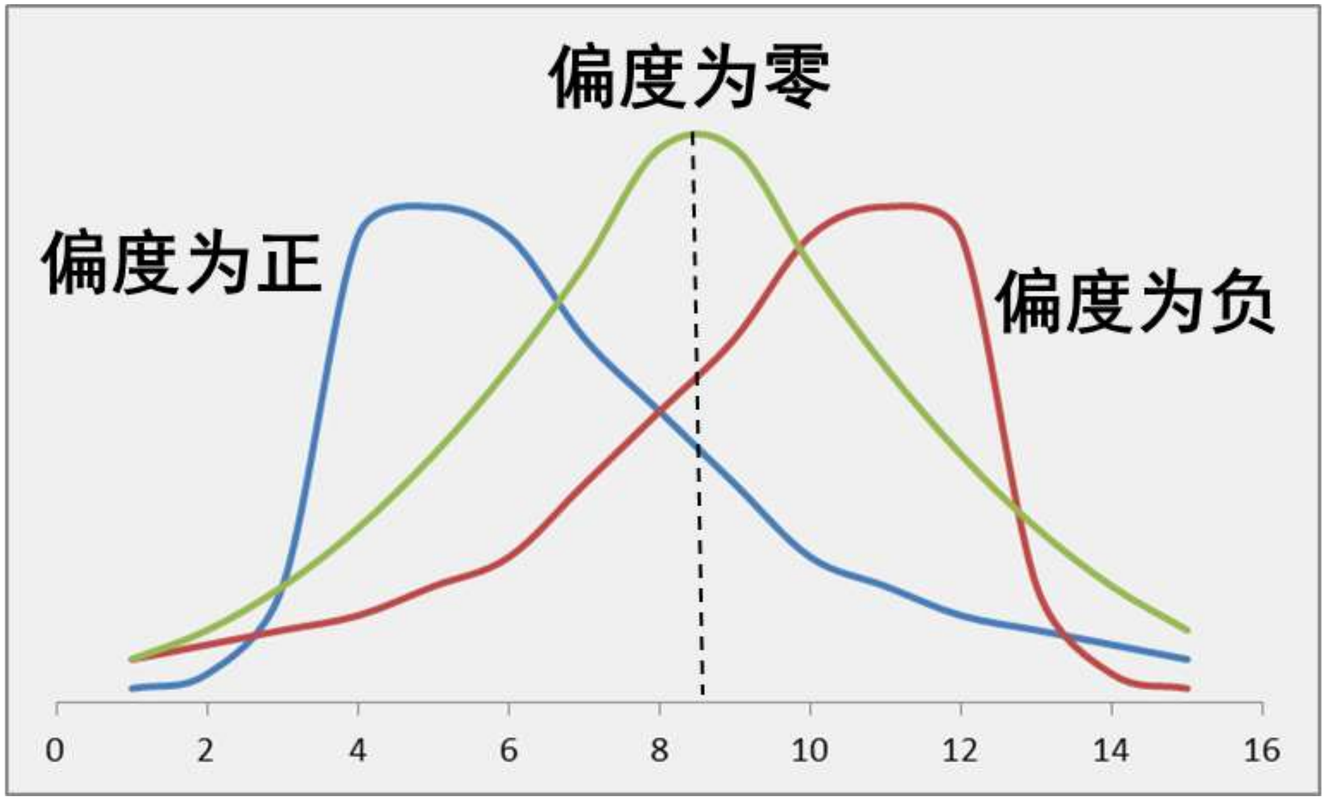
$$G1 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right)^3}$$

偏度的计算命令为skewness(x).

常用统计量

偏度若为零, 即数据相对均匀地分布在平均值两侧; 偏度为正, 表明数据在右侧有长尾现象,为负则表明数据向左有长尾现象. 参见右图.

例. 分别产生服从E(0.1)、E(9)、E(100)的1000个随机数，计算其各自偏度.



常用统计量

例. 随机生成四组随机数, 分别计算其偏度.

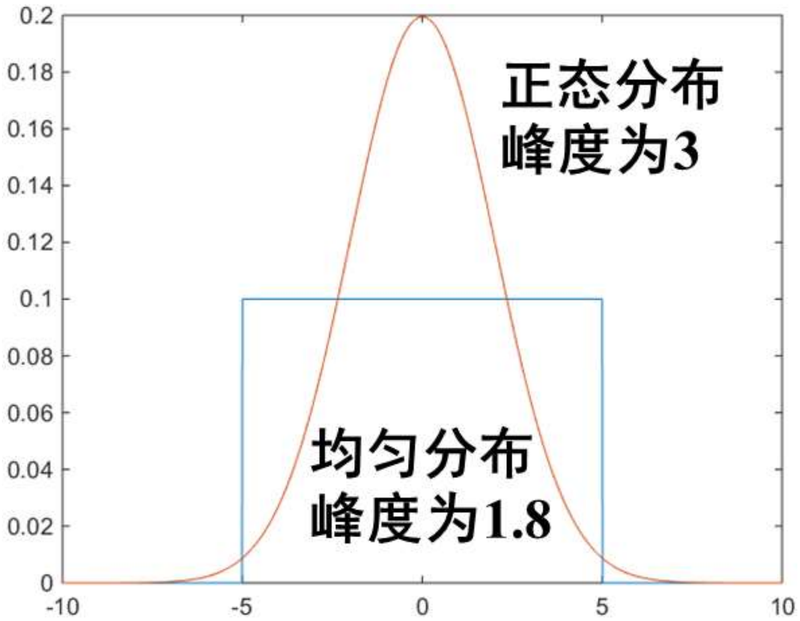
```
>> x=normrnd(0,1,1000,1);y=exprnd(5,1000,1);
>> z=-1*exprnd(5,1000,1);w=unifrnd(-4,4,1000,1);
>> skewness(x)      0.095
>> skewness(y)      2.189(右长尾)
>> skewness(z)     -2.065(左长尾)
>> skewness(w)     -0.033
```

常用统计量

峰度

峰度是描述数据分布形态陡缓程度的指标, 计算公式为:

$$G2=\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^4}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right)^2}$$



峰度越大, 数据分布的顶峰会显得越尖、越陡峭.

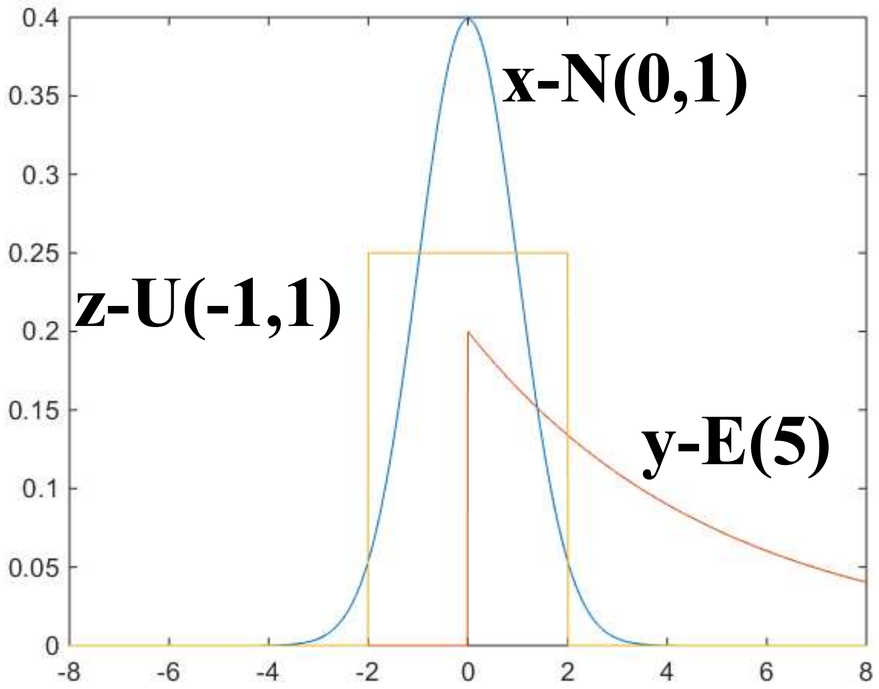
峰度的计算命令为kurtosis(x).

常用统计量

例. 随机生成三组随机数, 分别服从N(0,1)分布、E(5)分布和U(-1,1)分布. 分别计算其峰度.

```
>> x=normrnd(0,1,1000,1); y=exprnd(5,1000,1);
>> z=unifrnd(-1,1,1000,1);
```

```
>> kurtosis(x)    3.281
>> kurtosis(y)    10.225
>> kurtosis(z)    1.817
```



常用统计量

频率直方图

频率直方图是对概率密度函数的一种近似表达, 做法是选择一个覆盖所有数据的区间, 然后将此区间n等分, 计算每个小区间内落入的数据频次, 并据此作图.

常用统计量

频率直方图

具体做法如下：

假设共获得m个观测数据，第i个小区间内落入了v_i个数据，则该区间的频率记为v_i/m，到该区间为止的累计频率为$\sum_{k=1}^{i}\frac{v_k}{m}$

设第i个小区间为$[t_{i-1}, t_i]$，现以$[t_{i-1}, t_i]$为底，以v_i/m为高作长方形，得到频率直方图；若以$\sum_{k=1}^{i}\frac{v_k}{m}$为高作长方形得到累计频率直方图(经验分布函数图)

常用统计量

Matlab中绘制直方图有三个命令.

绘制频率直方图： hist(data,k)

绘制带有密度曲线的频率直方图： histfit(data,k)

绘制累计频率直方图(经验分布函数图)： cdfplot(data)

其中data表示需要分析的观测数据,k表示划分的小区间的个数,即准备将数据分成k组来计数.

常用统计量

例. 现有一组学生的期末考试成绩.

459 362 624 542 509 584 433 748 815 505 612 452 434

982 640 742 565 706 593 680 926 653 164 487 734 608

428 1153 593 844

分别绘制其频率直方图、带有密度曲线的频率直方图以及累计频率直方图.

频率直方图： hist(data,k)

带密度曲线的频率直方图： histfit(data,k)

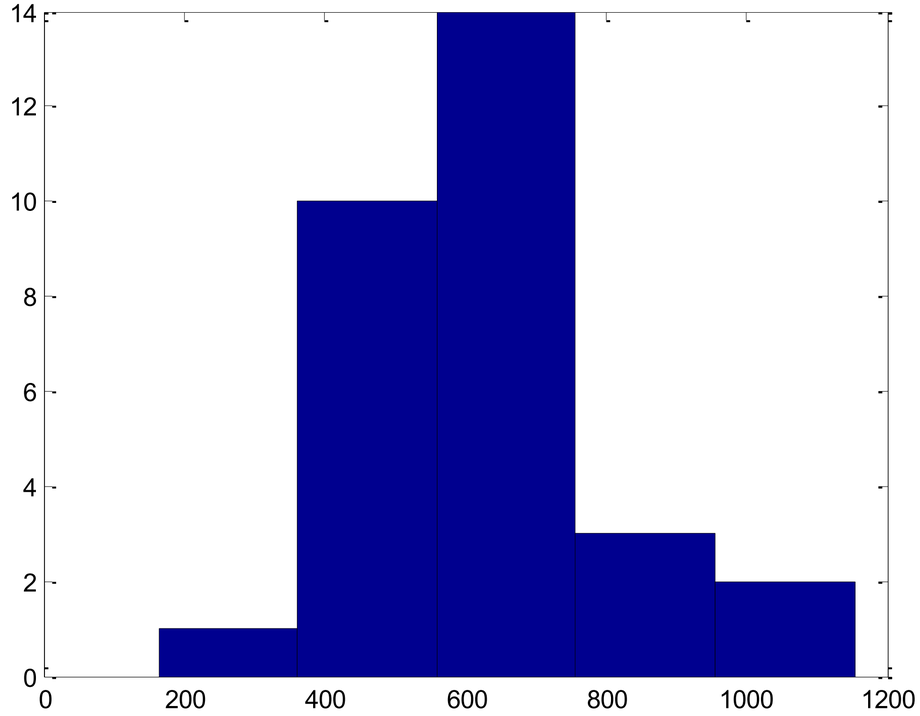
累计频率直方图(经验分布函数图)： cdfplot(data)

常用统计量

首先定义向量：

data=[459 362 624 542 509 584 433 748 815 505 612
452 434 982 640 742 565 706 593 680 926 653 164
487 734 608 428 1153 593 844]

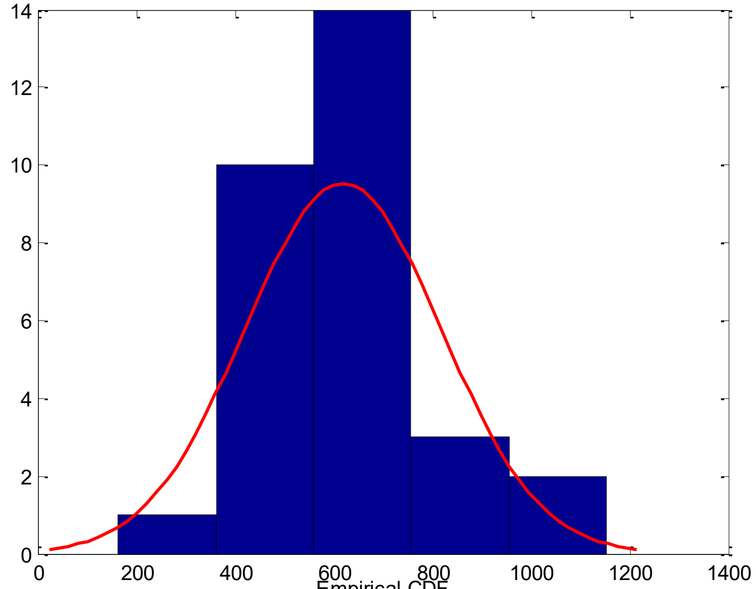
频率直方图： >> hist(data,5)



常用统计量

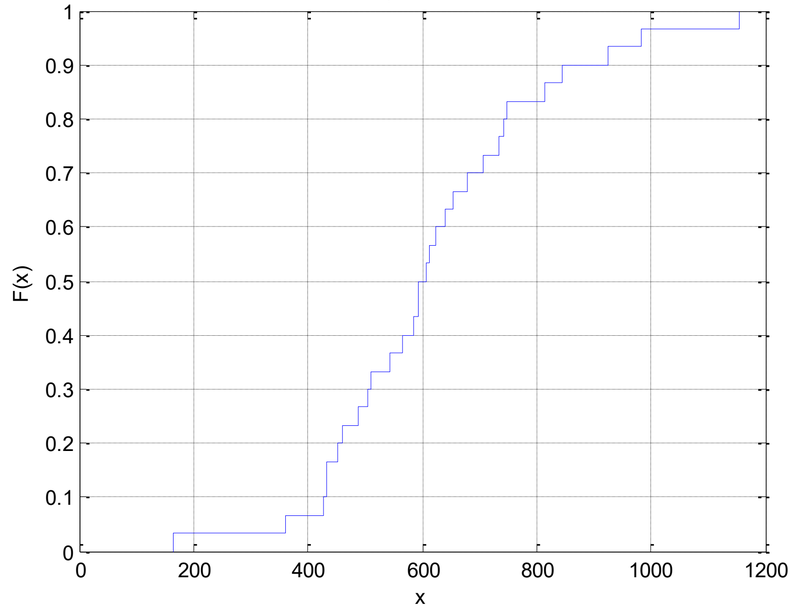
带有密度曲线的频率直方图：

```
>> histfit(data,5)
```



累计频率直方图(经验分布函数图)：

```
>> cdfplot(data)
```



常用统计量

例. 产生参数为10的指数分布随机数500个，并画出直方图和经验分布函数图

绘制频率直方图： hist(data,k)

绘制累计频率直方图(经验分布函数图)： cdfplot(data)

产生随机数 x=exprnd(10,500,1);

作直方图 hist(x,9)

作经验分布函数图 cdfplot(x)

常用统计量

常用的统计量还有很多，这里就不再一一介绍。在对数据进行深入统计分析之前可以将常用统计量都计算、整理出来，帮助我们初步了解数据的整体特征.

课堂练习：随机生成服从泊松分布π(10)的1000个均匀分布随机数，并计算其平均值、中位数、方差、标准差、极差、偏度、峰度，绘制其频率直方图、带有密度曲线的频率直方图和累计频率直方图.

作业

生成10000个服从如下分布的随机数，并计算其平均值、中位数、方差、标准差、极差、偏度、峰度，绘制其频率直方图、带有密度曲线的频率直方图和累计频率直方图.

$$1.F_X(x)=\begin{cases}\dfrac{1}{3}e^{x}, & x\le 0\\ 1-\dfrac{2}{3}e^{-x}, & x>0\end{cases}$$

2. Y~π(5).