参数估计与假设检验

实验内容

1. 参数估计

2. 假设检验

一　参数估计

参数估计问题的一般提法

设有一个统计总体，总体分布函数为$F(x,\theta)$，其中θ是未知参数，现从该总体抽样，得样本

$$X_1, X_2, \ldots, X_n$$

要依据该样本对参数θ作出估计，或估计θ的某个已知函数$g(\theta)$.

例如：对学生体重抽样检测得{wi}，猜测服从N(μ,σ^2)分布，那么μ和σ？？

答：求计算数据均值和方差，EX，EX^2 即可

一　参数估计

参数估计 $\begin{cases} \text{点估计} \\ \text{区间估计} \end{cases}$

点估计 —— 估计未知参数的值。

区间估计 —— 根据样本构造出适当的区间，使它以一定的概率包含未知参数或未知参数的已知函数的真值。

一　参数估计

（一）点估计的求法

1. 矩估计法（用样本矩估计总体矩）

设总体分布含有个k未知参数 $\theta_1, \ldots, \theta_k$,构造$l$阶矩$\mu_l$：

$$\mu_l = E(X^l) = \int_{-\infty}^{+\infty} x^l f(x;\theta_1,\theta_2,\ldots,\theta_k)\mathrm{d}x \quad \text{（}X\text{ 连续型）}$$

或 $$\mu_l = E(X^l) = \sum_x x^l p(x;\theta_1,\theta_2,\ldots,\theta_k) \quad \text{（}X\text{ 离散型）}$$

$l=1,\ldots, k$

一般说，它们是$\theta_1, \theta_2\ldots, \theta_k$的函数。

一　参数估计

由于样本的l阶矩 $A_l=\frac{1}{n}\sum_{i=1}^{n}X_i^l$

依概率收敛到总体的l阶矩μ_l。所以令

$$\mu_l(\theta_1,\ldots,\theta_k)=A_l,\ l=1,\ldots,k$$

解此方程组得其根为

$$\hat{\theta}_i(X_1,\ldots,X_n),\ i=1,\ldots,k$$

分别估计参数θ_i, $i=1,\ldots,k$, 并称其为θ_i的矩估计。

（构造线性方程组）

2. 极大似然估计法

设总体X有概率密度$f(x;\theta)$（或分布律$p(x;\theta)$，$\theta=(\theta_1,\ldots,\theta_k)$。设$X_1,\ldots,X_n$是来自总体的简单随机样本，$x_1,\ldots, x_n$是样本观测值。

极大似然估计是寻找参数θ_i，使样本 $X_1,\ldots,X_n$在样本值 $x_1,\ldots,x_n$附近取值的概率达到最大。

对学生体重抽样检测得{wi},假设服从N(μ,σ^2)分布.既然一次检测出现wi这个结果，那么它的概率就应该非常大.所以用N(μ,σ^2)计算 {wi} 发生的概率，令其最大化，求得μ和σ.

一　参数估计

构造似然函数：$L(\theta_1,\theta_2\ldots,\theta_k)=\prod_{i=1}^{n} f(x_i,\theta_1,\ldots,\theta_k)$

$$L(\theta_1,\theta_2\ldots,\theta_k)=\prod_{i=1}^{n} p(x_i,\theta_1,\ldots,\theta_k)$$

若有参数$\theta=(\theta_1,\ldots,\theta_k)$的取值，

$$\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$$

使得似然函数$L(\theta_1,\ldots,\theta_k)$达到最大，则称它为参数$\theta_1,\ldots,\theta_k$的极大似然估计。

（二）区间估计

设总体X的分布中含有未知参数θ，若对于给定的概率$1-\alpha(0<\alpha<1)$，存在两个统计量$\hat{\theta}_1(X_1,X_2,\ldots,X_n)$和$\hat{\theta}_2(X_1,X_2,\ldots,X_n)$使得$P(\hat{\theta}_1<\theta<\hat{\theta}_2)=1-\alpha$

若已知概率分布，可以直接计算概率；不知概率分布，只能根据样本估计：随机变量落在某个范围内的概率是多少？

则称随机区间$(\hat{\theta}_1,\ \hat{\theta}_2)$为参数$\theta$的置信水平为$1-a$的置信区间，$\hat{\theta}_1$称为置信下限，$\hat{\theta}_2$称为置信上限.

置信区间的意义：

反复抽取容量为n的样本，可以据此估计θ的取值范围（区间）.
参数θ位于该区间内的概率为$1-a$。

一　参数估计

1. 数学期望的置信区间

设样本$(X_1,X_2,\ldots,X_n)$来自正态母体$X\sim N(\mu,\sigma^2)$

（1）方差σ^2已知，μ的置信区间

$$\left(\overline{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \overline{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

%期望落在此范围内的概率为1-α。

（2）方差σ^2未知，μ的置信区间

$$\left(\overline{X}-t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}},\ \overline{X}+t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right)$$

一　参数估计

2. 方差的区间估计

μ未知时，方差σ^2的置信区间为

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \right)$$

%方差落在此范围内的概率为1-α。

S^2是样本方差.

（三）参数估计的命令

1. 正态总体的参数估计

点估计和区间估计可同时由以下命令获得：

[muhat,sigmahat,muci,sigmaci] = normfit(X, alpha)

alpha为显著性水平, 1-alpha为置信水平. alpha缺省时设定为0.05.

muhat 输出正态分布均值的点估计值

sigmahat 输出标准差的点估计值

muci输出均值的区间估计

sigmaci输出标准差的区间估计

X为数据矩阵（列为变量）时，输出行变量。

一　参数估计

例1．分别给出容量为50、100、1000的正态分布$N(10, 2^2)$的随机数，并以此为样本值，给出μ和σ的点估计和区间估计.

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

命令：

```
r=normrnd(10,2,50,1);
[mu,sigm,muci,sigmci]=normfit(r)
r=normrnd(10,2,100,1);
[mu,sigm,muci,sigmci]=normfit(r)
r=normrnd(10,2,1000,1);
[mu,sigm,muci,sigmci]=normfit(r)
```

例2. 一组身高数据，假设服从正态分布，计算学生身高的均值和标准差的点估计和置信水平为0.95的区间估计。

170.1,179,171.5,173.1,174.1,177.2,170.3,176.2,175.4,
163.3,179.0,176.5,178.4,165.1,179.4,176.3,179.0,173.9,
173.7,173.2,172.3,169.3,172.8,176.4,163.7,177.0,165.9,
166.6,167.4,174.0,174.3,184.5,171.9,181.4,164.6,176.4,
172.4,180.3,160.5,166.2,173.5,171.7,167.9,168.7,175.6,
179.6,171.6,168.1,172.2

```
r=[......]
[mu,sigm,muci,sigmci]=normfit(r)
```

一　参数估计

练习1：生成服从N(2,9)分布的100个随机数，给出其均值和标准差的点估计，和置信水平99%的区间估计。

```
r=normrnd(mu,sigma,i,j);
[mu,sigm,muci,sigmci]=normfit(r,alpha)
```

一 参数估计

2. 其它分布的参数估计

（1）取容量充分大的样本(n>50)，按中心极限定理，它近似地服从正态分布；

（2）使用特定分布总体的估计命令.

1^0[muhat, muci] = expfit(X,alpha)-----在显著性水平alpha下，求指数分布数据X的均值的点估计、区间估计.

2^0[lambdahat, lambdaci] = poissfit(X,alpha)-----在显著性水平alpha下，求泊松分布数据X的参数的点估计、区间估计.

一　参数估计

函数名	参数估计对应的参数	调用格式
mle	极大似然估计	phat=mle('dist',data) [phat,pci]=mle('dist',data) [phat,pci]=mle('dist',data,alpha) [phat,pci]=mle('dist',data,alpha,pl)
normlike	对数正态似然函数	L=normlike(params,data)
normfit	正态分布	[muhat,sigmahat,muci,sigmaci]=normfit(X,alpha)

一　参数估计

函数名	参数估计对应的参数	调用格式
poissfit	泊松分布	lambdahat=poissfit(X) [lambdahat,lambdaci]=poissfit(X) [lambdahat,lambdaci]=poissfit(X, alpha)
unifit	均匀分布	[ahat,bhat]=unifit(X) [ahat,bhat,ACI,BCI]=unifit(X) [ahat,bhat,ACI,BCI]=unifit(X, alpha)

一　参数估计

函数名	参数估计对应的参数	调用格式
weibfit	威布尔分布	phat=weibfit(X) [phat,pci]=weibfit(X) [phat,pci]=weibfit(X, alpha)
weiblike	威布尔对数似然函数	logL=weiblike(params,data) [logL,info]=weiblike(params,data)

一　参数估计

练习2：生成参数为0.5的指数分布、泊松分布随机数100个，计算其参数的点估计和置信水平97%的区间估计.

```
[lamta,lamtaci]=expfit(r,alpha)
[lamta,lamtaci]=poissfit(r,alpha)
```

```
r=exprnd(0.5,100,1);
[lamta,lamtaci]=expfit(r)
[lamta,lamtaci]=expfit(r,0.01)
```

二　假设检验

统计推断的另一类重要问题是假设检验问题。

在总体的分布函数完全未知或只知其形式，但不知其参数的情况，为了推断总体的某些未知特性，提出某些关于总体的假设。

对总体X的分布律或分布参数作某种假设，根据抽取的样本观察值，运用数理统计的分析方法，检验这种假设是否正确，从而决定接受假设或拒绝假设.

比如拿到一组数据，主观判断其服从泊松分布，那么这种判断究竟对不对？---------大胆假设，小心求证。

二　假设检验

1. 参数检验：如果总体的分布函数类型已知，这时构造出的统计量依赖于总体的分布函数，这种检验称为参数检验.

参数检验的目的往往是对总体参数及其有关性质给出明确的判断.

2. 非参数检验：如果所检验的假设并非是对某个分布的参数作出明确的判断，检验统计量的分布函数不依赖于总体的分布类型，这种检验叫非参数检验.

如判断总体分布类型的检验就是非参数检验.

二　假设检验

假设检验的一般步骤是：

① 根据实际问题提出原假设H_0与备择假设H_1，即说明需要检验的假设的具体内容。

② 选择适当的统计量，构造恰当的拒绝域.

③ 根据样本观测值计算统计量的观测值，看其是否落入拒绝域中，从而在检验水平条件下对拒绝或接受原假设H_0作出判断.

二　假设检验

（一）参数检验

1. 单个正态总体$X\sim N(\mu, \sigma^2)$均值检验

------- 方差σ^2已知时采用z检验

------- 方差σ^2未知，采用t检验

二 假设检验

	H_0	H_1	总体方差 σ^2 已知 统计量 $z=\dfrac{\overline{X}-\mu_0}{\dfrac{\sigma}{\sqrt{n}}}$	总体方差 σ^2 未知 统计量 $t=\dfrac{\overline{X}-\mu_0}{\dfrac{s}{\sqrt{n}}}$
			在显著水平 α 下拒绝 H_0，若	
I	$\mu=\mu_0$	$\mu\neq\mu_0$	$\lvert z\rvert > u_{1-\frac{\alpha}{2}}$	$\lvert t\rvert > t_{1-\frac{\alpha}{2}}(n-1)$
II	$\mu\leq\mu_0$	$\mu>\mu_0$	$z>u_{1-\alpha}$	$t>t_{1-\alpha}(n-1)$
III	$\mu\geq\mu_0$	$\mu<\mu_0$	$z<-u_{1-\alpha}$	$t<-t_{1-\alpha}(n-1)$

2. 单个正态总体方差检验———— χ^2

	H_0	H_1	均值 μ 已知 统计量 $\chi^2=\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i^2-\mu)^2$	均值 μ 未知 统计量 $\chi^2=\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i^2-\overline{X})^2$
			在显著水平 α 下拒绝 H_0，若	
I	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2<\chi^2_{\frac{\alpha}{2}}(n)$ 或 $\chi^2>\chi^2_{1-\frac{\alpha}{2}}(n)$	$\chi^2<\chi^2_{\frac{\alpha}{2}}(n-1)$ 或 $\chi^2>\chi^2_{1-\frac{\alpha}{2}}(n-1)$
II	$\sigma^2=\sigma_0^2$	$\sigma^2>\sigma_0^2$	$\chi^2>\chi^2_{1-\alpha}(n)$	$\chi^2>\chi^2_{1-\alpha}(n-1)$
III	$\sigma^2=\sigma_0^2$	$\sigma^2<\sigma_0^2$	$\chi^2<\chi^2_{\alpha}(n)$	$\chi^2<\chi^2_{\alpha}(n-1)$

3. 两个正态总体$N(\mu_1, \sigma_1^2)$和$N(\mu_2, \sigma_2^2)$均值检验

（1）已知$\sigma_1^2=\sigma_2^2$，选取统计量 $z=\dfrac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$

（2）方差未知 $\sigma_1^2=\sigma_2^2$ ，检验H_0: $\mu_1=\mu_2$

选取统计量

$$t=\frac{\bar{X}-\bar{Y}}{\sqrt{(n_1-1)s_1^2+(n_2-1)s_2^2}}\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}$$

（2）方差未知$\sigma_1^2=\sigma_2^2$，检验H_0: $\mu_1=\mu_2$

选取统计量

$$t=\frac{\overline{X}-\overline{Y}}{\sqrt{(n_1-1)s_1^2+(n_2-1)s_2^2}}\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}$$

当$|t|>t_{1-\alpha/2}(n_1+n_2-2)$时拒绝原假设，否则接受原假设。

------$t_{1-\alpha/2}(n_1+n_2-2)$为$t$分布$t(n_1+n_2-2)$的下测1-$\alpha$/2分位数，$n_1$为来自总体$N(\mu_1, \sigma_1^2)$的样本的容量，$n_2$是来自总体$N(\mu_2, \sigma_2^2)$的样本的容量。

4. 两个正态总体方差检验

<table>
<tr><td rowspan="2"></td><td rowspan="2">H$_0$</td><td rowspan="2">H$_1$</td><td>均值 μ_1,μ_2 已知
统计量 F_0</td><td>均值 μ_1,μ_2 未知
统计量 F</td></tr>
<tr><td colspan="2">在显著水平 α 下拒绝 H$_0$，若</td></tr>
<tr><td>I</td><td>$\sigma_1^2=\sigma_2^2$</td><td>$\sigma_1^2\neq\sigma_2^2$</td><td>$F_0>F_{1-\frac{\alpha}{2}}(n_1,n_2)$ 或
$F_0<\dfrac{1}{F_{1-\frac{\alpha}{2}}(n_2,n_1)}$</td><td>$F>F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)$ 或
$F<\dfrac{1}{F_{1-\frac{\alpha}{2}}(n_2-1,n_1-1)}$</td></tr>
<tr><td>II</td><td>$\sigma_1^2\leq\sigma_2^2$</td><td>$\sigma_1^2>\sigma_2^2$</td><td>$F_0>F_{1-\alpha}(n_1,n_2)$</td><td>$F>F_{1-\alpha}(n_1-1,n_2-1)$</td></tr>
<tr><td>III</td><td>$\sigma_1^2\geq\sigma_2^2$</td><td>$\sigma_1^2<\sigma_2^2$</td><td>$F_0<\dfrac{1}{F_{1-\alpha}(n_2,n_1)}$</td><td>$F<\dfrac{1}{F_{1-\alpha}(n_2-1,n_1-1)}$</td></tr>
</table>

$$F_0=\frac{\frac{1}{n_1}\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\frac{1}{n_2}\sum_{i=1}^{n_2}(Y_i-\mu_2)^2},$$

$$\boldsymbol{F}=\frac{s_1^2}{s_2^2}$$

二　假设检验

5. 参数检验的计算机命令

1^0 z检验

（1）命令ztest函数

（2）功能：给定方差条件下进行正态总体均值的检验

（3）语法：[h, sig, ci]=ztest(x, m, sigm, alpha, tail);

h=1，则拒绝原假设，h=0，则接受原假设

ztest(x，m，sigm)在alpha水平下进行Z检验，以确定服从正态分布的样本均值是否为m，sigm为给定的标准差。Alpha缺省表示置信水平0.05.

参数tail指定单侧检验/双侧检验

二　假设检验

[h, sig, ci]=ztest(x, m, sigm, alpha, tail)

- tail=0（为默认设置）指定备择假设 $\mu \neq \mu_0$
- tail=1指定备择假设 $\mu > \mu_0$
- tail=-1指定备择假设 $\mu < \mu_0$

sig为与Z统计量相关的p值。

ci为均值真值的1-alpha置信区间。

二 假设检验

例6. 生成100个标准正态分布的随机数，假设均值和标准差的观测值与真值之间没有差异，进行检验。

[h, sig, ci]=ztest(x, m, sigm, alpha, tail)

- tail=0（为默认设置）指定备择假设 $\mu \neq \mu_0$
- tail=1指定备择假设 $\mu > \mu_0$
- tail=-1指定备择假设 $\mu < \mu_0$

x=normrnd(0, 1, 1, 100); [h, sig, ci]=ztest(x, 0, 1)

结果：h=0, sig=0.6317, ci=[-0.1481   0.2439]

h=1，则拒绝原假设，h=0，则接受原假设

二 假设检验

例7. 某批矿砂的5个样品中的镍含量，经测定为（%）

3.25 3.27 3.24 3.26 3.24

设测定值总体服从正态分布，标准差为0.04，问在0.01水平上能否接受假设：这批镍含量的均值为3.25。

[h, sig, ci]=ztest(x, m, sigm, alpha, tail)

- tail=0（为默认设置）指定备择假设 $\mu \neq \mu_0$
- tail=1指定备择假设 $\mu > \mu_0$
- tail=-1指定备择假设 $\mu < \mu_0$

h=1，则拒绝原假设，h=0，则接受原假设

```
x=[3.25 3.27 3.24  3.26  3.24];
[h, sig, ci]=ztest(x, 3.25, 0.04, 0.01)
```

结果：

h=0, sig=0.9110, ci=[3.2059 3.298]

例8. 下面列出的是某工厂随机选取的20只部件的装配时间

9.8 10.4 10.6 9.6 9.7 9.9 10.9 11.1 9.6 10.2

10.3 9.6 9.9 11.2 10.6 9.8 10.5 10.1 10.5 9.7

设总体服从正态分布，标准差为0.4，问在0.05水平上能否认为装配时间的均值显著的大于10。需检验H$_0$: $\mu \le 10$, H$_1$: $\mu > 10$

```
[h, sig, ci]=ztest(x, m, sigm, alpha, tail)
```

- tail=0（为默认设置）指定备择假设 $\mu \neq \mu_0$
- tail=1指定备择假设 $\mu > \mu_0$
- tail=-1指定备择假设 $\mu < \mu_0$

h=1，则拒绝原假设，h=0，则接受原假设

[h, sig, ci]=ztest(x, 10, 0.4, 0.05, 1) 结果：h=1, sig=0.0127, ci=[10.0529 inf]

二 假设检验

2^0单个样本的t检验

（1）命令ttest函数

（2）功能：未知方差条件下进行正态总体均值的检验

（3）语法：[h, sig, ci]=ttest(x, m, alpha, tail)；

h=1，则拒绝原假设，h=0，则接收原假设

二　假设检验

例9. 测得一批刚件20个样品的屈服点（单位：T/mm^2）为：

4.98 5.11 5.20 5.11 5.00 5.61 4.88 5.27 5.38 5.20

5.46 5.27 5.23 4.96 5.35 5.15 5.35 4.77 5.33 5.54

设屈服点服从正态分布，在0. 05水平上，检验该样本的均值是否为5. 20的假设检验。需检验H_0: $\mu = 5.20$, H_1: $\mu \neq 5.20$

过程如下：

```
x=[4.98  5.11  5.20  5.11  5.00  5.61  4.88
   5.27  5.38  5.20  5.46  5.27  5.23  4.96
   5.35  5.15  5.35  4.77  5.33  5.54];
m=mean(x)
[h, sig, ci]=ttest(x, 5.20, 0.05)
```

结果：

```
m=5.2075
h=0
sig=0.8796
ci=[5.1052   5.3098]
```

3^0两个样本的t检验

（1）命令ttest2函数

（2）功能：两个样本均值差异的t检验

（3）语法：[h, significance, ci]=ttest2(x, y, alpha, tail);

h=1，则拒绝原假设，h=0，则接收原假设

例10. 对两种不同的水稻品种A, B分别统计了8个地区的单位面积产量（单位：kg）

品种A：86 87 56 93 84 93 75 79

品种B：80 79 58 91 77 82 76 66

要求检验两个水稻品种的单位面积产量之间是否有显著差异?

过程如下：

```
x=[86 87 56 93 84 93 75 79];
y=[80 79 58 91 77 82 76 66 ];
[h, significance, ci]=ttest2(x, y);
```

结果：

```
h=0
significance=0.3393
ci=-6.4236 17.4236
```

二　假设检验

练习3：生成服从N(4.1, 1)的100个随机数向量x，服从π(4)的100个随机数向量y，检验其均值是否有显著差异？

```
x=normrnd(4.1, 1, 100, 1);
y=poissrnd(4, 100, 1);
[h, significance, ci]=ttest2(x, y)
h = 0
m = 0.3139
n = -0.6171    0.1992
```

二　假设检验

（二）非参数检验

1. Jarque-Bera检验

（1） 数学原理：Jarque-Bera检验是评价X服从正态分布的假设是否成立。该检验基于样本偏度和峰度，样本偏度接近于0，样本峰度接近于3。基于此构造一个包含χ^2统计量：

JB=n(g1^2+(g2-3)^2/4)/6（n为样本容量）

Jarque和Bera证明了在正态性假定下，JB渐进的服从自由度为2的χ^2分布，若JB超过了$\chi_{1-\alpha}^2$（2），即χ^2（2）的下测1-α分位数，则拒绝正态分布零假设，反之，接受零假设。

二　假设检验

（2）函数名称：jbtest

（3）语法：H=jbtest(x);

H=jbtest(x, alpha);

[H, p, jbstat, cv]=jbtest(x, alpha);

H=1，则拒绝服从正态，H=0，则接收服从正态

（4）alpha为显著水平，p为p值，jbstat为检验统计量的值，cv为确定是否拒绝原假设的的临界值。

（5）应用实例

例11. 对下列数据确定其是否服从正态分布。

```
x = [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
     527 552 513 781 474 388 824 538 862 659 775 859 755 49 697
     515 628 954 771 609 402 960 885 610 292 837 473 677 358 638
     699 634 555 570 84 416 606 1062 484 120 447 654 564 339 280
     246 687 539 790 581 621 724 531 512 577 496 468 499 544 645
     764 558 378 765 666 763 217 715 310 851];
```

```
[H, p, jbstat, cv]=jbtest(x, 0.05);
```

结果：H=0　p= 0.6913　Jbstat= 0.7384　cv=5.9915

三　线性回归

一元线性回归

回归模型和参数确定

一元线性回归研究因变量与一个自变量之间的线性关系。模型为：$y = b_0 + b_1 x$

y是因变量，x是自变量，b_0，b_1为待估参数。

2. 函数名称：regress

（1）语法：b=regress(y, x);

（2）说明：b返回参数的估计值。y为列向量，　x为自变量取值矩阵。

注意：在回归分析时，可先对数据划出散点图，看是否有线性关系，再进行回归分析，散点图的命令为：scatter

例13. 为研究某一化学反应过程中，温度x(^{0}C)对产品得率Y（%）的影响，测得数据如下：

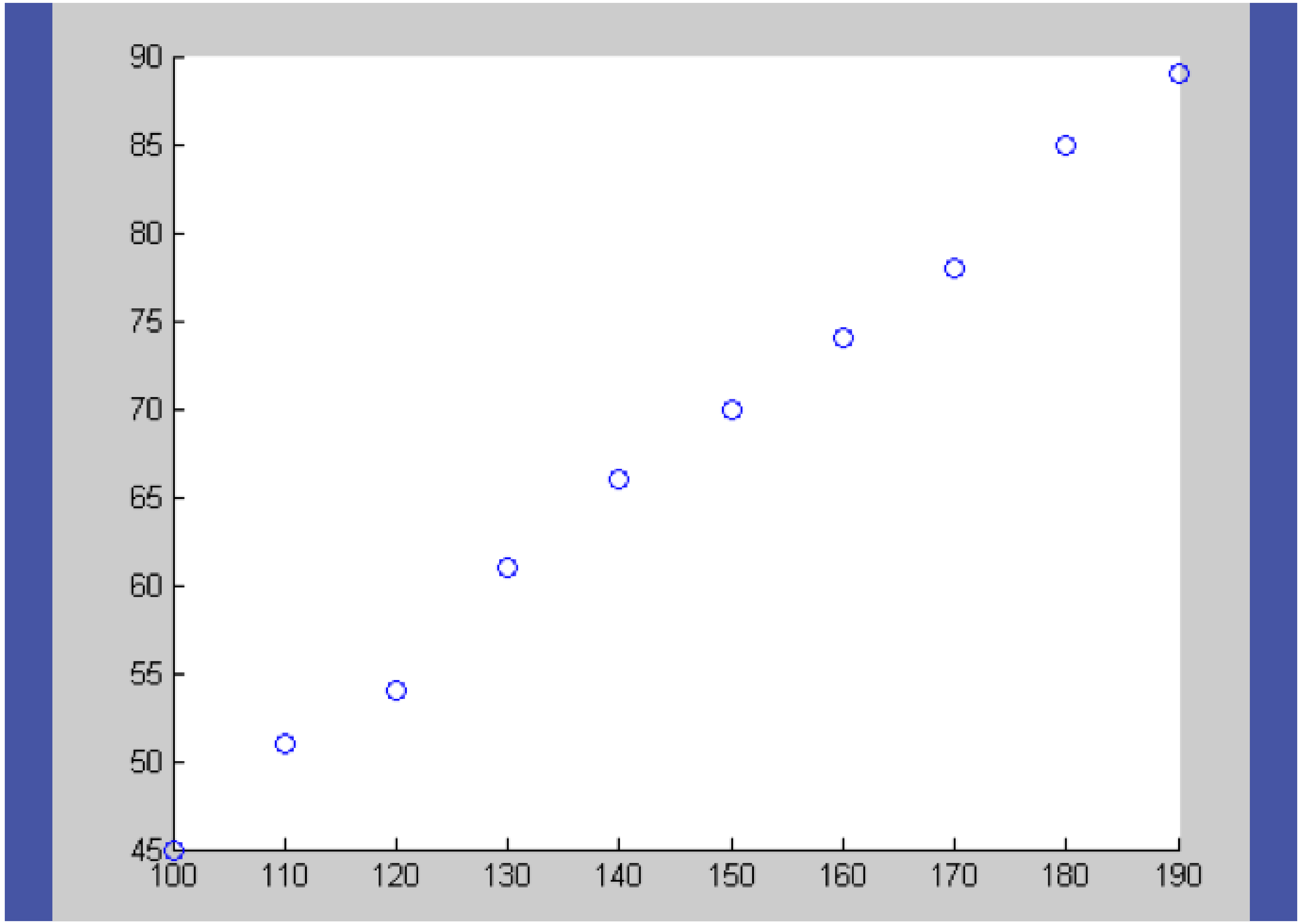
温度x	100	110	120	130	140	150	160	170	180	190
得率y	45	51	54	61	66	70	74	78	85	89

求y关于x的线性回归方程。

过程如下：先画散点图

```
x=[100; 110; 120; 130; 140; 150; 160; 170; 180; 190] ;
y=[45; 51; 54; 61; 66; 70; 74; 78; 85; 89] ;
scatter(x, y)
```

三　线性回归



90
85
80
75
70
65
60
55
50
45
100
110
120
130
140
150
160
170
180
190

三　线性回归

```
a=ones(length(x), 1);     %生成一个与x同型的1向量
z=[a, x];                 %1向量与x合成一个2列的矩阵
b=regress(y, z)           %调用线性回归命令
```

b=-2.7394　0.4830

表明y=-2.7394+0.4830*x