- 随机试验
- 数据的统计分析
- 参数估计
- 假设检验

随机试验

古典概率： 事件A发生的概率

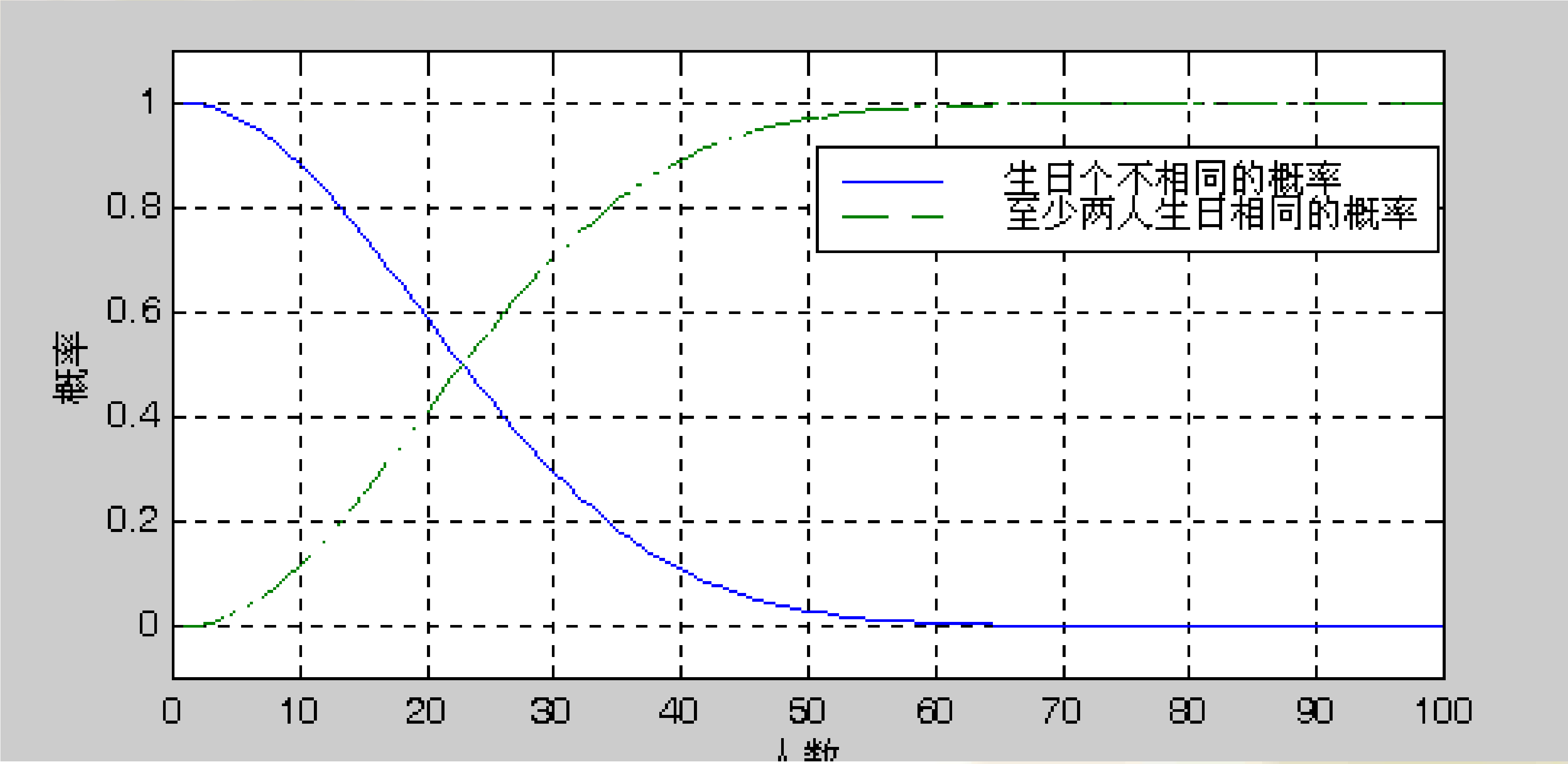
$$P(A)=\frac{m}{n}=\frac{A\text{包含的样本点个数}}{\text{样本点总数}}$$

在100人的团体中，如果不考虑年龄的差异，研究是否有两个以上的人生日相同。假设每人的生日在一年365天中的任意一天是等可能的，那么随机找 n 个人（<365）

问这些人生日各不相同的概率是多少？

至少有两个人生日相同的概率为多少？

```
for n=1:100
   p0(n)=prod(365:-1:365-n+1)/365^n;
   p1(n)=1-p0(n);
end
n=1:100;
plot(n,p0,n,p1,'--')
xlabel(‘人数’),ylabel(‘概率’)
legend(‘生日各不相同的概率’,‘至少两人相同的概率’)
axis([0 100 -0.1 1.1]),grid on
```


生日个不相同的概率
至少两人生日相同的概率
概率
人数
0
0.2
0.4
0.6
0.8
1
0
10
20
30
40
50
60
70
80
90
100

重要的概率分布

表8-2 概率分布的命令字符

分布	离散型随机变量				连续型随机变量					
	均匀分布	二项分布	泊松分布	几何分布	均匀分布	指数分布	正态分布	χ^2分布	t分布	F分布
字符	unid	bino	poiss	geo	unif	exp	norm	chi2	t	f

表8-3 运算功能的命令字符

功能	概率密度	分布函数	逆概率分布	均值与方差	随机数生成
字符	pdf	cdf	inv	stat	rnd

正态分布

$X \sim N(\mu, \sigma^2)$ 概率密度为 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty < x < +\infty$

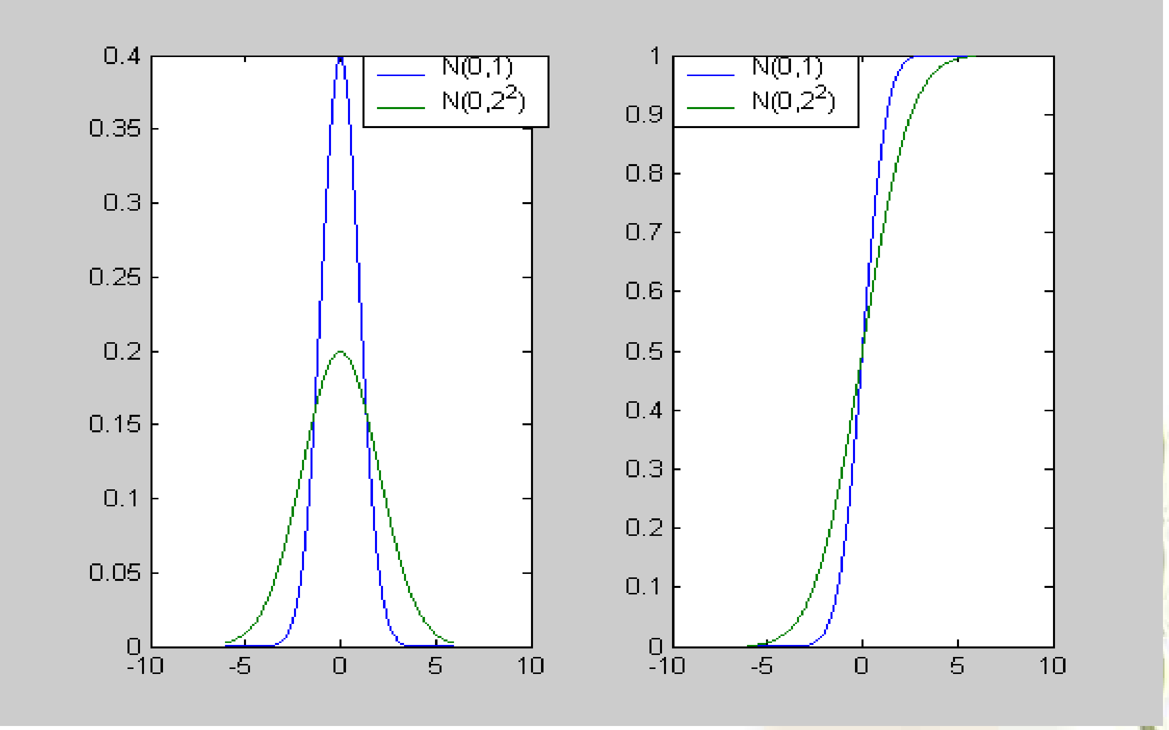
y=normpdf(x, mu, sigma) $\mu = mu$, $\sigma = sigma$ 的正态分布的密度函数

y=normpdf(x) 标准正态分布的密度函数

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

y=normcdf(x, mu, sigma) $\mu = mu$, $\sigma = sigma$ 的正态分布的分布函数

y=normcdf(x) 标准正态分布的分布函数




```
x=-6:0.01:6;
y1=normpdf(x);
z1=normcdf(x);
y2=normpdf(x, 0, 2);
z2=normcdf(x, 0, 2);
subplot(1, 2, 1),plot(x, y1, x, y2);
 legend('N(0,1)','N(0,2^2)')
 subplot(1, 2, 2),plot(x, z1, x, z2);
 legend('N(0,1)','N(0,2^2)')
```

例：某厂生产一种设备，其平均寿命为**10**年，标准差为**2**年.如该设备的寿命服从正态分布，求寿命不低于**9**年的设备占整批设备的比例？

解： 设随机变量 ξ 为设备寿命，由题意 $\xi \sim N(10,2^2)$

$$P(\xi \geq 9) = 1 - P(\xi < 9)$$

```
>>clear
>> p1=normcdf(9,10,2)
p1 =0.3085
>>1-p1
  ans = 0.6915
```

P180

统计作图

在数据较小、较少的情况下输入 - - - **Matlab**交互环境下输入

数据量较大，且不以计算机可读形式存在 **load *.M** - - - **M**文件的形式输入数据

load *.txt - - - 读数据文件的命令读入

基本统计函数表

函数名称	功能简介
Max(x)	求最大值
Min(x)	求最小值
Median(x)	求中值
Range(x)	求极差
Mean(x)	求算术平均值
Std(x)	求样本标准差
Var(x)	求样本方差
Cov(x)	求协方差矩阵

例8-19 某班（共有120名学生）的高等数学成绩如下：

74	63	78	76	89	56	70	97	89	94	76	88
65	83	72	41	39	72	73	68	14	76	45	70
90	46	54	61	75	76	49	57	78	66	64	74
78	87	86	73	47	67	21	66	79	67	68	65
56	84	66	73	68	72	76	65	70	94	53	65
77	78	53	74	59	50	98	67	89	78	63	92
54	87	84	80	63	64	85	66	69	69	60	54
75	33	30	62	74	65	84	73	55	85	75	76
81	71	83	72	56	84	76	75	67	65	35	94
59	47	45	67	75	36	78	82	94	70	84	75

根据以上数据作出该门课程成绩的频数表和直方图。

解：（1）数据输入：

- 方法1：在Matlab的交互环境下直接输入;
- 方法2：将以上数据以一列的形式存为A.txt文件,用 load A.txt 命令读入数据。

(2) 用hist命令作频数表和直方图：(区间个数为5，可省略)

- [N,X]=hist(A,5) 120名学生高数成绩的频数表；
- hist(A,5) 120名学生高数成绩的直方图；

matleb命令：

load A.txt↙

disp('高数成绩的频数表'),[N,X]=hist(A,5)%N为频数

频数是如何计算的？

```
[N,X]=hist(A,5)
```

N =

3   10   22   60   25

X =

22.4000   39.2000   56.0000   72.8000   89.6000

```
a1=min(A);a2=max(A);
disp(['成绩最小值',blanks(4),'最大值'])
disp([a1,a2])
```

成绩最小值   最大值

14   98

—

84 ÷5=16.8   14+16.8=30.8

22.4从哪儿来的？

所以，成绩在14~30.8之间有3人。

hist(A,5)%直方图

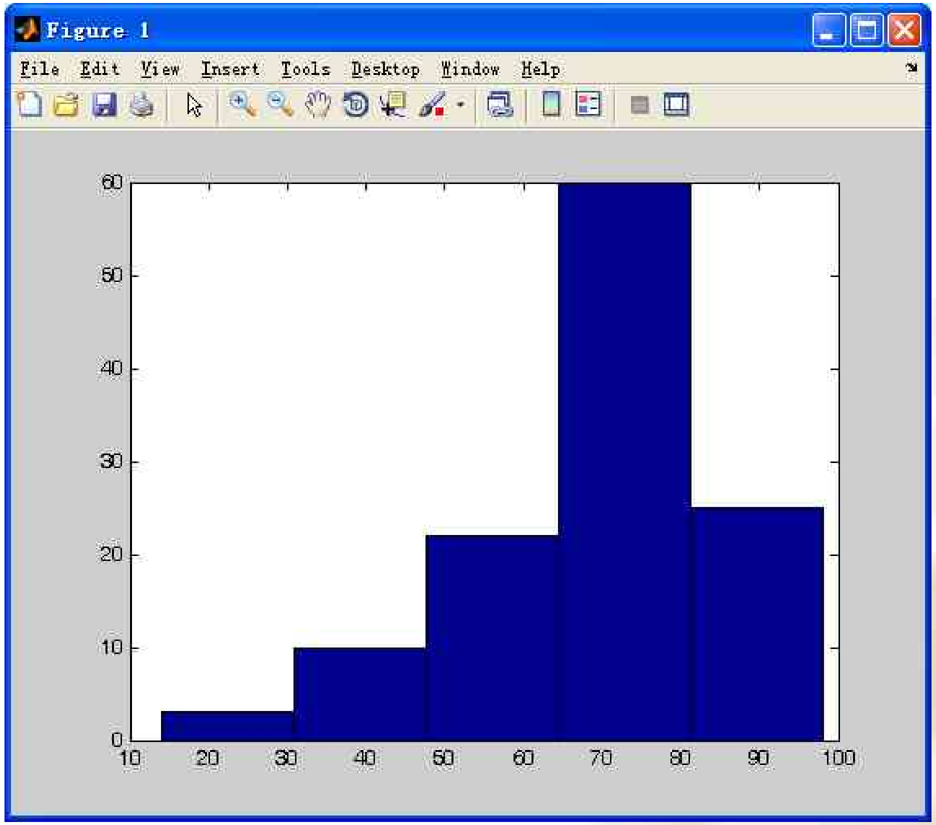


Figure 1
File Edit View Insert Tools Desktop Window Help
0
10
20
30
40
50
60
10 20 30 40 50 60 70 80 90 100

例**8-20**　求例**8-19**中**A**的均值、中位数、极差、方差和标准差。

解：在命令窗口输入：

```
M=[mean(A),median(A),range(A),var(A),std(A)]
```

```
M = 68.9583   71.5000   84.0000  249.5697   15.7978
```

参数估计

参数估计的Matlab函数

函数	功能
[mu,sigma,muci,sigmaci]=normfit(x,alpha)	正态总体的均值、标准差的极大似然估计mu和sigma,返回在显著性水平alpha下的均值、标准差的置信区间muci和sigmaci，x是样本（数组或矩阵），当alpha缺省时设定为0.05.
[mu,muci]=expfit(x,alpha)	指数分布的极大似然估计，返回显著性水平alpha下的置信区间muci，x是样本（数组或矩阵），当alpha缺省时设定为0.05.
[a,b,aci,bci]=unifit(x,alpha)	均匀分布的极大似然估计，返回显著性水平alpha下的置信区间aci,bci ，x是样本（数组或矩阵），当alpha缺省时设定为0.05.
[p,pci]=binofit(x,n,alpha)	二项分布的极大似然估计, 返回在显著性水平alpha下的置信区间pci，x是样本（数组或矩阵），当alpha缺省时设定为0.05.
[lambda, lambdaci]=poissfit(x,alpha)	泊松分布的极大似然估计，返回显著性水平alpha下的置信区间lambdaci，x是样本（数组或矩阵），当alpha缺省时设定为0.05.

例8-22　某厂生产的瓶装运动饮料的体积假定服从正态分布，抽取10瓶，测得体积（毫升）为

595，602，610，585，618，615，605，620，600，606，

求均值 μ 、标准差 σ 的极大似然估计值及置信水平为0.90的置信区间。

```
x=[595 602 610 585 618 615 605 620 600 606];
[mu,sigma,muci,sigmaci]=normfit(x,0.90)
```

```
mu =
  605.6000
sigma =
   10.8033
muci =
  605.1584
  606.0416
sigmaci =
   10.8864
   11.5724
```

假设检验

正态总体的均值和方差假设检验

原假设	检验统计量及其分布	备择假设 H_1	否定域 W
$\mu=\mu_0$（σ^2已知）	$Z=\dfrac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)$	$\mu>\mu_0$	$Z\geq z_\alpha$
		$\mu<\mu_0$	$Z\leq -z_\alpha$
		$\mu\neq\mu_0$	$\lvert Z\rvert\geq z_{\alpha/2}$
$\mu=\mu_0$（σ^2未知）	$t=\dfrac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}\sim t(n-1)$	$\mu>\mu_0$	$t\geq t_\alpha(n-1)$
		$\mu<\mu_0$	$t\leq -t_\alpha(n-1)$
		$\mu\neq\mu_0$	$\lvert t\rvert\geq t_{\alpha/2}(n-1)$
$\sigma^2=\sigma_0^2$（μ未知）	$\chi^2=\dfrac{(n-1)s^2}{\sigma_0^2}\sim\chi^2(n-1)$	$\sigma^2>\sigma_0^2$	$\chi^2\geq\chi_\alpha^2(n-1)$
		$\sigma^2<\sigma_0^2$	$\chi^2\leq\chi_{1-\alpha}^2(n-1)$
		$\sigma^2\neq\sigma_0^2$	$\chi^2\geq\chi_{\alpha/2}^2(n-1)$ 或 $\chi^2\leq\chi_{1-\alpha/2}^2(n-1)$

单个正态总体 $N(\mu,\sigma^2)$ 均值 μ 的假设检验

1. σ^2 已知，关于 μ 的检验（U检验法）

[h, p, ci]=ztest(x, mu, sigma, alpha, tail)

x	--------------	样本（数组或矩阵）
mu	--------------	原假设 H_0 中的 μ_0
sigma	--------------	总体标准差 σ
alpha	--------------	显著性水平 α，缺省时**0.05**
tail	--------------	对备择假设 H_1 的选择

tail=0时，备择假设 $H_1: \mu \neq \mu_0$，（默认）

tail=1时，备择假设 $H_1: \mu > \mu_0$

tail=-1时，备择假设 $H_1: \mu < \mu_0$

[h, p, ci]=ztest(x, mu, sigma, alpha, tail)

h=0表示“在显著性水平**alpha**的情况下，接受 H_0 ”

h=1表示“在显著性水平**alpha**的情况下，拒绝 H_0 ”

p -------------- 假设 H_0 下样本均值出现的概率

ci -------------- 均值的置信区间

例8-23　在某粮店的一批大米中，随机地抽测6袋，其重量为26.1，23.6，25.1，25.4，23.7，24.5（单位：千克）。设每袋大米的重量 $X \sim N(\mu, 0.1)$，问能否认为这批大米的袋重是25千克（ $\alpha = 0.01$）？

解：按题意：H_0 ：$\mu = 25$，H_1 ：$\mu \neq 25$

已知 $\sigma = 0.316, \alpha = 0.01$ ，程序如下：

```
x=[26.1 23.6 25.1 25.4 23.7 24.5];
[h,p,ci]=ztest(x,25,0.316,0.01,0)
```

结果：

```
h =     0        ---       接受H0
p =    0.0387
ci =   24.4010   25.0656
```

当 $\alpha = 0.1$ 时，命令为：

```
x=[26.1 23.6 25.1 25.4 23.7 24.5];
[h,p,ci]=ztest(x,25,0.316,0.1,0)
```

结果：

```
h =
     1
p =
    0.0387
ci =
   24.5211   24.9455
```

--- 拒绝H_0

单个正态总体 $N(\mu,\sigma^2)$ 均值 μ 的假设检验

2. σ^2 未知，关于 μ 的检验（t检验法）

```
[h, p, ci]=ttest(x, mu, alpha, tail)
```

例　某种原件的寿命X（以小时计）服从正态分布 $N(\mu,\sigma^2)$ 其中 μ , 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）？

$$H_0: \mu \le \mu_0 = 225, \quad H_1: \mu > 225$$

解：按题意需检验：H_0：$\mu \le \mu_0 = 225$，H_1：$\mu > 225$

取 $\alpha = 0.05$ ，程序如下：

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

```
h =    0
p =   0.2570
ci =  198.2321      Inf
```

h=0,p=0.2570,说明在显著水平为**0.05**的情况下，不能拒绝原假设，认为元件的平均寿命不大于**225**小时。

单个正态总体 $N(\mu,\sigma^2)$ 方差 σ^2 的假设检验

总体 $X \sim N(\mu,$ … 自总体的样本。

$x_1, x_2, \cdots x_n$ 是样 …

chi2inv(p,n)是求 $X \sim \chi^2(n)$ 时，$P(X,x) = P(X \le x) = p$ 中的 x，即 χ^2 分布的逆概率函数

检验假设 $H_0: \sigma^2 = \sigma_0^2$，$H_1: \sigma^2 \ne \sigma_0^2$

```
x=[x1,x2,…,xn];
chi2=(n-1)*var(x)/sigma^2;
u1=chi2inv(alpha/2,n-1)
u2=chi2inv(1-alpha/2,n-1)
if  chi2<u1
   h=1
elseif  chi2>u2
   h=1
else
  h=0
end
```

$$P(X \ge \chi_{\alpha/2}{}^2(n-1)) = \frac{\alpha}{2}$$

$$\Rightarrow P(X \le \chi_{\alpha/2}{}^2(n-1)) = 1 - \frac{\alpha}{2}$$

$\sigma^2 = \sigma_0{}^2$ (μ未知)	$\chi^2 = \frac{(n-1)s^2}{\sigma_0{}^2} \sim \chi^2(n-1)$	$\sigma^2 > \sigma_0{}^2$	$\chi^2 \ge \chi_\alpha{}^2(n-1)$
		$\sigma^2 < \sigma_0{}^2$	$\chi^2 \le \chi_{1-\alpha}{}^2(n-1)$
		$\sigma^2 \ne \sigma_0{}^2$	$\chi^2 \ge \chi_{\alpha/2}{}^2(n-1)$ 或 $\chi^2 \le \chi_{1-\alpha/2}{}^2(n-1)$

例8-24　例8-23中能否认为每袋大米质量的标准差 $\sigma = 0.316(kg)$ ？

```
x=[26.1 23.6 25.1 25.4 23.7 24.5];
chi2=5*var(x)/0.1
u1=chi2inv(0.01/2,5)
u2=chi2inv(1-0.01/2,5)
if  chi2<u1
   h=1
elseif  chi2>u2
   h=1
```

```
chi2 =
  48.5333
u1 =
```

例8-23　在某粮店的一批大米中，随机地抽测6袋，其重量为26.1，23.6，25.1，25.4，23.7，24.5（单位：千克）。设每袋大米的重量 $X \sim N(\mu, 0.1)$ ，问能否认为这批大米的袋重是25千克（ $\alpha = 0.01$ ）？

P190

15 16 18 22 24

考试安排

- 考试时间：**40**分钟。每人**2**题。
- 开卷考试，只能带书本，不可拷贝.

注意事项：

- 1、在E盘建立word文档（程序以及运行结果），文件名为“考试+姓名+学号”；
- 2、随时保存；
- 3、提交以后到教室机确认是否提交成功