第一章　数理统计的基本概念

1、总体与样本
2、统计量
3、经验分布函数
4、抽样分布
5、正态总体的抽样分布定理

1、 总体、个体、样本

总体 — 研究对象全体的集合,记为X

个体 —组成母体的每个元素,记为$X\ i$

样本 —从总体中抽取的部分个体

$$(X_1, X_2, \boxed{?}, X_n)$$

为总体X的一个容量为n的样本,

$(x_1, x_2, \boxed{?}, x_n)$ 为样本的观察值.

随机抽样

重复抽样 ⟶ 独立同分布

非重复抽样 ⟶ 近似独立同分布

简单随机样本（iid）

若母体 X 的子样 $(X_1, X_2, \cdots, X_n)$ 满足：

$X_1, X_2, \cdots, X_n$ 有相同的分布

$X_1, X_2, \cdots, X_n$ 相互独立

则称 $(X_1, X_2, \cdots, X_n)$ 为简单随机样本.

设总体X的分布函数为$F(x)$,则iid样本$(X_1, X_2, \boxed{?}, X_n)$ 的联合概率分布函数为

$$F_n(x_1, x_2, \boxed{?}, x_n) = \prod_{i=1}^{n} F(x_i)$$

若总体X的密度函数为$f(x)$,则iid样本的概率密度函数为

$$f_n(x_1, x_2, \boxed{?}, x_n) = \prod_{i=1}^{n} f(x_i)$$

若母体X为离散情形,则iid 样本

$$(X_1, X_2, \boxed{?}, X_n)$$

的联合概率分布为

$$P_n(x_1, x_2, \boxed{?}, x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

例1 设某批产品共有N个,其中的次品数为M, 其次品率为 $p = M / N$

若 p 是未知的,则可用抽样方法估计.

从这批产品中任取一个产品,用随机变量 X 来描述它是否是次品:

$$X = \begin{cases} 1, & \text{所取的产品是次品} \\ 0, & \text{所取的产品不是次品} \end{cases}$$

其概率分布

$$f(x, p) = p^x (1-p)^{1-x}, \qquad x = 0, 1$$

设有放回地抽取一个容量为n的样本

$$(X_1, X_2, \boxed{?}, X_n)$$

是iid,其样本观察值为 $(x_1, x_2, \boxed{?}, x_n)$

样本空间为

$$\Omega = \{(x_1, x_2, \boxed{?}, x_n) \mid x_i = 0,1,\ \ i = 1,2, \boxed{?}, n\}$$

$(X_1, X_2, \boxed{?}, X_n)$的联合分布为

$$f_{总}(x_1, x_2, \boxed{?}, x_n) = \prod_{i=1}^{n} f(x_i)$$

$$= p^{\sum\limits_{i=1}^{n} x_i} (1-p)^{n-\sum\limits_{i=1}^{n} x_i}$$

若抽样是无放回的,则前次抽取的结果会影响后面抽取的结果.例如

$$P(X_2=1|X_1=1)=\frac{M-1}{N-1}=\frac{p-1/N}{1-1/N}\xrightarrow{N\to\infty}p$$

$$P(X_2=1|X_1=0)=\frac{M}{N-1}=\frac{p}{1-1/N}\xrightarrow{N\to\infty}p$$

当子样容量 n 与母体中个体数目 N 相比很小时, 可将无放回抽样近似地看作放回抽样.

2、统计量

定义1.2.1 样本$(X_1, X_2, \boxed{?}, X_n)$的函数

$$g(X_1, X_2, \boxed{?}, X_n)$$

且不含任何未知参数,则称 $g(X_1, X_2, \boxed{?}, X_n)$ 为**统计量**.

例 $X \sim N(\mu, \sigma^2)$, μ, σ^2 是未知参数,则

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \text{是统计量,}$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \longrightarrow \text{不是统计量.}$$

2.1 常用统计量(Statistic) —— 样本矩

样本均值（Sample Mean）：

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

样本方差（Sample Variance）：

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 .$$

样本均值具有下述性质：

设总体的数学期望和方差分别为 μ 和 σ^2

则样本均值的数学期望和方差为

$$E(\overline{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\mu$$

$$D(\overline{X})=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{\sigma^2}{n}$$

样本均值的数学期望不变，方差缩小n倍。

样本方差具有下述性质：

设总体的数学期望和方差分别为 μ 和 σ^2

则样本方差的数学期望为

$$E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) =$$

$$= \frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}E(X_i^2) - nE(\overline{X}^2)\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)\right) = \sigma^2$$

样本标准差（**Sample Standard Deviation**）：

$$S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}.$$

样本矩（**Sample Moment**）

k[illegible]
$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\ k=1,2\boxed{?}.$$

k[illegible]
$$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k,\ k=1,2\boxed{?}.$$

由大数定律可知：

若总体的r阶矩存在，则样本的直到r阶的样本矩（样本中心矩和样本原点矩）当样本容量n趋于无穷时，必依概率收敛（或几乎处处收敛）于相应的总体矩。

这为我们“利用样本矩去估计总体矩”提供了理论上的保证。

2.2 顺序统计量

将子样值 $x_1, x_2, \boxed{?}, x_n$ 由小到大排列

$$x_1^* \le x_2^* \le \boxed{?} \le x_n^*$$

定义 $X_{(k)} = x_k^*, k = 1, 2, \boxed{?}, n$

则称统计量 $X_{(1)}, X_{(2)}, \boxed{?}, X_{(n)}$

为顺序统计量.

极小统计量 $X_{(1)}$

极大统计量 $X_{(n)}$

中位数

$$m_e = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & n\text{为奇数} \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & n\text{为偶数} \end{cases}$$

极差

$$R = X_{(n)} - X_{(1)}$$
$$= \max\{X_1, \boxed{?}, X_n\} - \min\{X_1, \boxed{?}, X_n\}$$

样本p分位数

$$m_p = \begin{cases} X_{[np+1]}, & np\text{不是整数} \\ \frac{1}{2}(X_{(np)} + X_{(np+1)}), & np\text{是整数} \end{cases}$$

样本的四分位数与箱线图

$$m_0 = X_{(1)}, m_{0.25}, m_{0.5}, m_{0.75}, \quad m_1 = X_{(n)}$$

2.3 秩统计量

将子样值 $x_1, x_2, \boxed{?}, x_n$ 由小到大排列

$$x_1^* \le x_2^* \le \boxed{?} \le x_n^*$$

R_k表示x_k在顺序统计量中的位次称为X_k的秩。

$$(R_1, R_2, \boxed{?}, R_n)$$

称为**秩统计量**。

例3 从一批机器零件毛坯中随机地抽取10件，测得其重量为(单位: 公斤):

210, 243, 185, 240, 215,

228, 196, 235, 200, 199

求这组样本的均值、方差、中位数与极差.

解 将样本由小到大重排

$$(x_{(1)}, x_{(2)}, \boxed{?}, x_{(10)})$$

$$= (185, 196, 199, 200, 210,$$
$$215, 228, 235, 240, 243)$$

秩统计量 R= (5,10,1,9,6,7,2,8,4,3)

$$\bar{x} = \frac{1}{10}(230 + 243 + 185 + 240 + 215$$

$$+ 228 + 196 + 235 + 200 + 199) = 217.1$$

$$a_2 = \frac{1}{10}\sum_{i=1}^{10} x_i^{\ 2} = 47522.5$$

$$s^2 = a_2 - \bar{x}^2 = 47522.5 - (217.1)^2 = 390\ .09$$

$$\boldsymbol{me = (x_{(5)} + x_{(6)})/2 = (210 + 215)/2 = 212.5}$$

$$R = x_{(10)} - x_{(1)} = 58$$

2.4 顺序统计量的分布

$X_{(n)}$的分布函数

$X_{(1)}$的分布函数

$\left(X_{(1)}, X_{(n)}\right)$ 的联合分布

作为一个研究课题，请同学们自行推导。

3、经验分布函数

设样本为 $(x_1, x_2, \boxed{?}, x_n)$，对任意实数 x

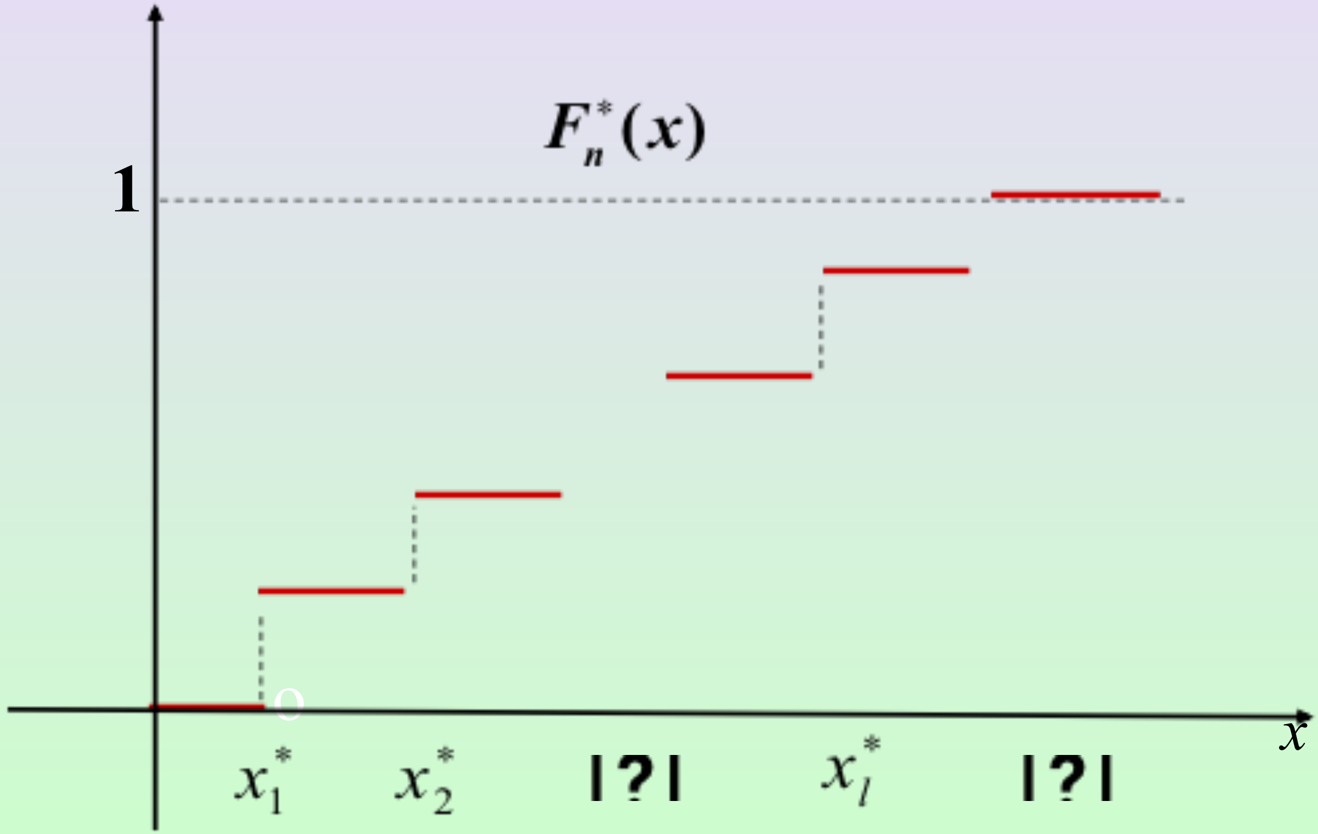
样本中小于或等于x的个数为 $m(x)$，则称

$$F_n^*(x) = \frac{m(x)}{n}$$

为样本的经验分布函数.

$$F_n^*(x) = \begin{cases} 0, & x < x_1^* \\ \dfrac{k}{n}, & x_k^* \le x < x_{k+1}^* \\ 1, & x \ge x_n^* \end{cases}$$

经验分布函数图形



由Chebyshev 大数定律, 对任意 $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|F_n^*(x) - F(x)\right| < \varepsilon\right) = 1$$

对每个固定x $F_n^*(x)$ 依概率收敛于 $F(x)$.

格利汶科(W.Glivenko)定理

$$P\left(\lim_{n\to\infty} \sup_{x\in R}\left| F_n^*(x) - F(x)\right| = 0\right) = 1$$

当 $n\to\infty$时, 经验分布函数

几乎处处一致收敛于总体分布函数$F(x)$。

例4 从织布车间抽取12匹布，检查每匹布的疵点数，样本

(1,0,0,2,1,3,2,0,1,1,2,1)

求样本频数分布和经验分布函数.

将12个数从小到大排列,得顺序统计量

0,0,0,1,1,1,1,1,2,2,2,3

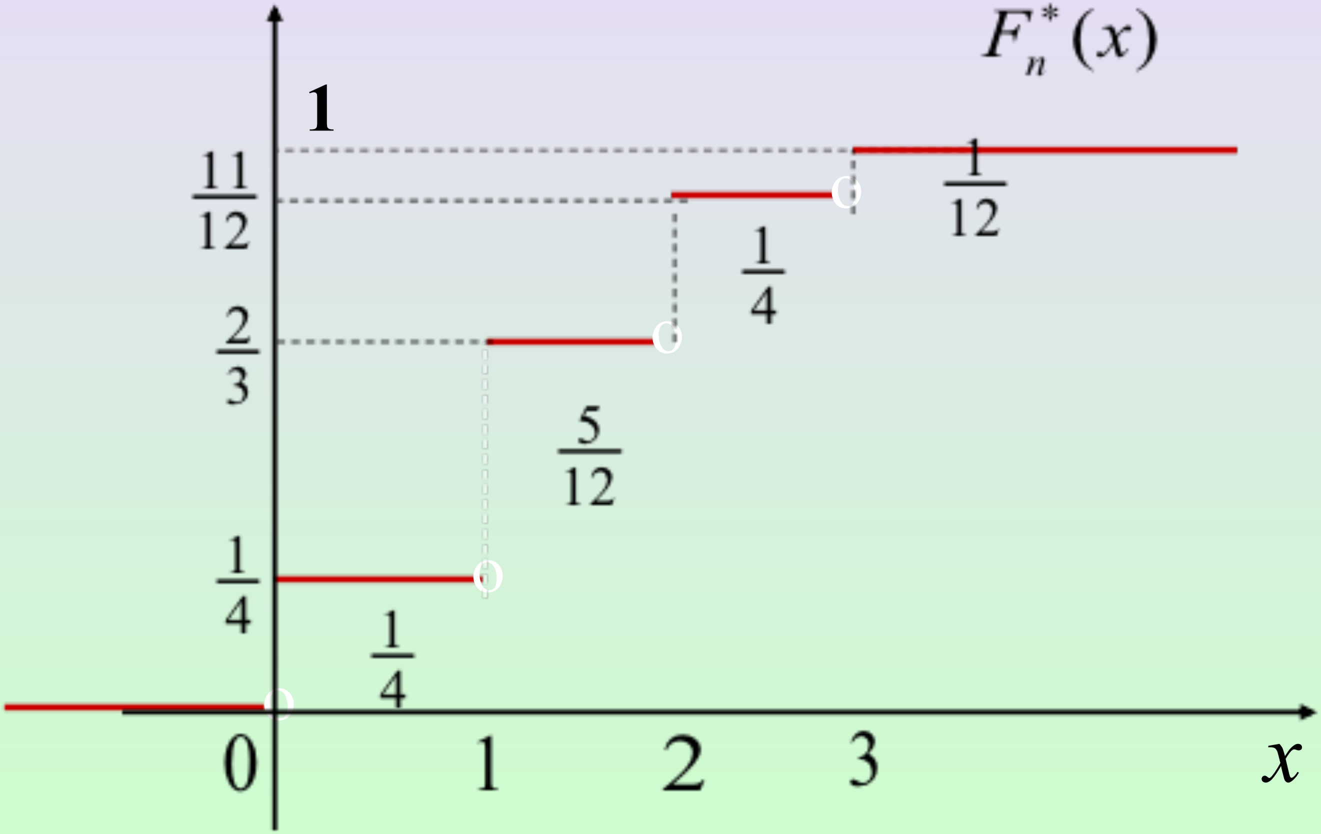
频数表

X	0	1	2	3
频数	3	5	3	1

样本经验分布函数：

$$F_n^*(x)=\begin{cases}0, & x<0\\ 1/4, & 0\le x<1\\ 2/3, & 1\le x<2\\ 11/12, & 2\le x<3\\ 1, & x\ge 3\end{cases}$$

经验分布函数图形



$F_n^*(x)$
1
$\frac{11}{12}$
$\frac{2}{3}$
$\frac{1}{4}$
$\frac{1}{4}$
$\frac{5}{12}$
$\frac{1}{4}$
$\frac{1}{12}$
0
1
2
3
x

频率（频数）直方图

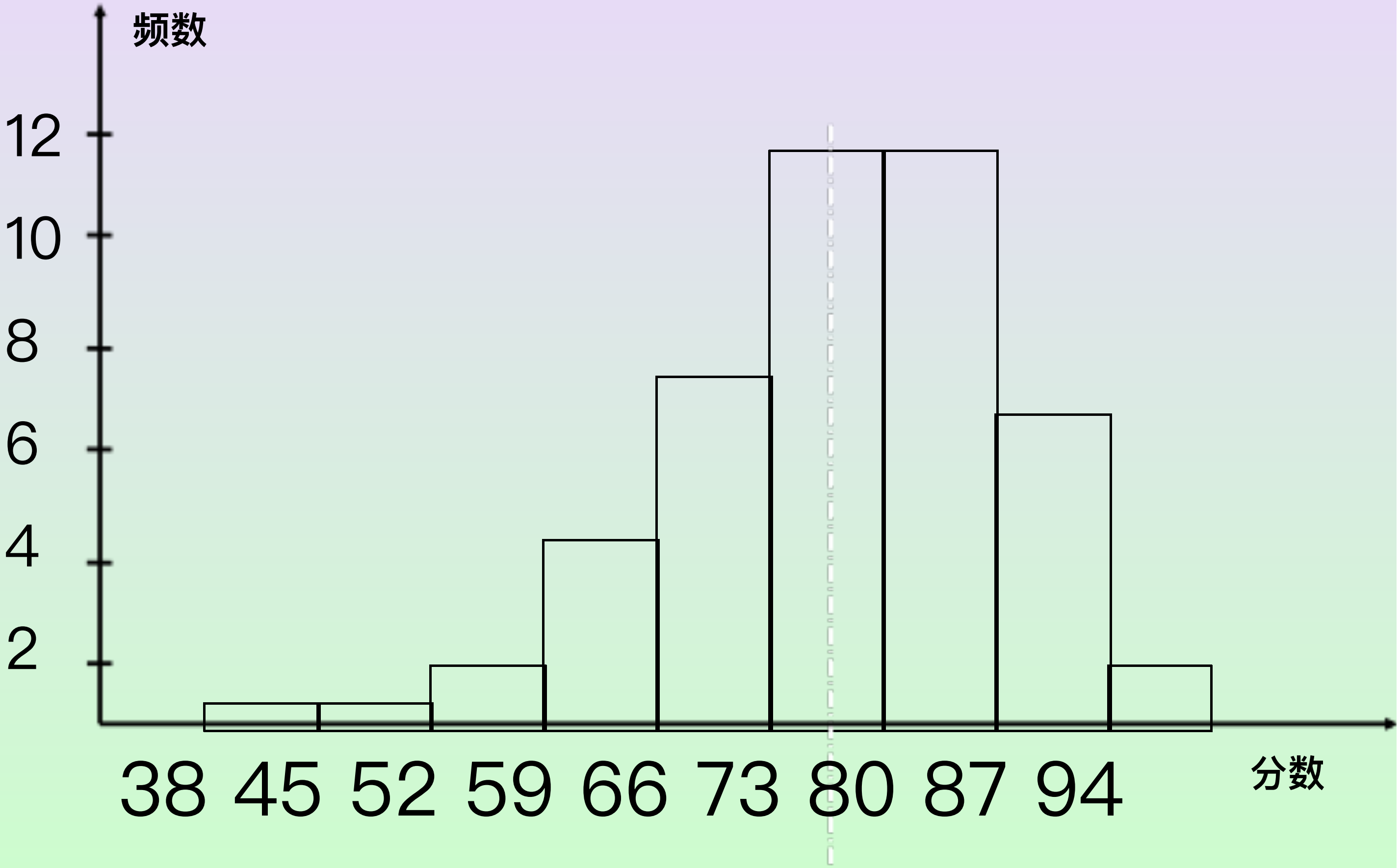
例5 **某班50名学生概率考试成绩如下：**

75	65	80	81	92	63	77	79	54	98
85	72	66	84	83	60	82	78	64	90
81	78	76	86	68	76	73	71	88	87
65	57	46	89	78	66	87	79	84	78
96	88	67	38	67	75	83	82	68	85

频数分布表

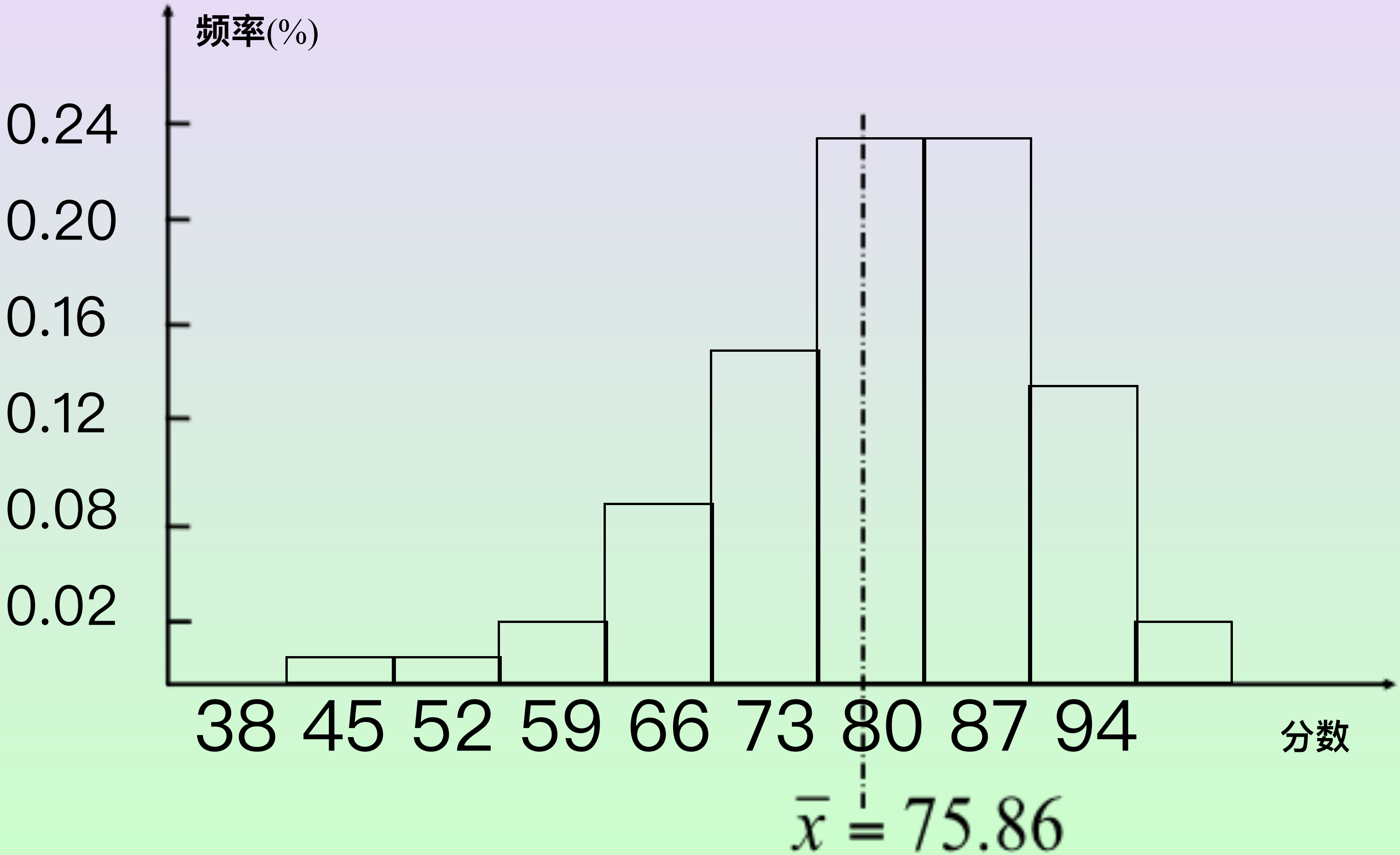
组号	组限	组中值	组频数	组总分
1	38~44	41	1	41
2	45~51	48	1	48
3	52~58	55	2	110
4	59~65	62	5	310
5	66~72	69	8	552
6	73~79	76	12	912
7	80~86	83	12	996
8	87~93	90	7	630
9	94~99	97	2	194
合		$\bar{x} = 75.86$	50	3793

频数直方图

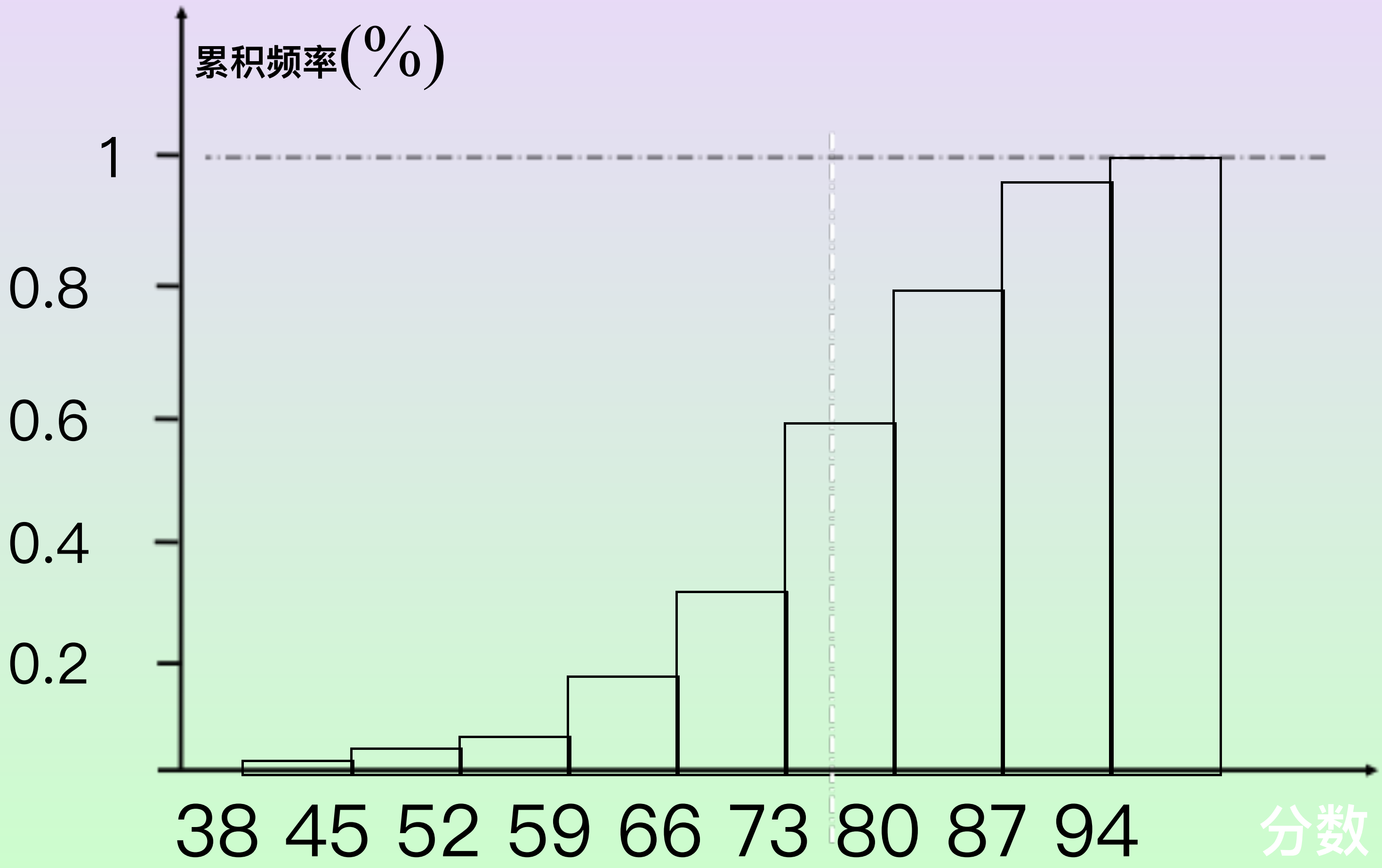


组号	组限	组中值	频数	频率	累积频率
1	38~44	41	1	0.02	0.02
2	45~51	48	1	0.02	0.04
3	52~58	55	2	0.04	0.08
4	59~65	62	5	0.10	0.18
5	66~72	69	8	0.16	0.34
6	73~79	76	12	0.24	0.58
7	80~86	83	12	0.24	0.82
8	87~93	90	7	0.14	0.96
9	94~99	97	2	0.04	1
合			50		

频率直方图



累积频率直方图



例6 从某地区随机抽取50户农民，调查其人均年收入情况，得到数据（单位:元）如下：

924	800	916	704	870	1040	824	690	574	490
972	988	1266	684	764	940	408	804	610	852
602	754	788	962	704	712	854	888	768	848
882	1192	820	878	614	846	746	828	792	872
696	644	926	808	1010	728	742	850	864	738

试对该地区农民收入的水平和贫富悬殊程度做个大致分析。

解：显然，如果不进行加工，面对这一大堆大小参差不齐的数据，很难得出什么印象．但是可以对这些数据稍事加工，如记各农户的人均年收入分别为 $x1$，$x2$，...，$x50$，计算得到

$$\bar{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = 809.52,$$

$$s = \sqrt{\frac{1}{50-1}\sum_{i=1}^{50}(x_i - \bar{x})^2} = 155.85$$

这样，就可以了解到该地区农民的平均收入和该地区农民贫富悬殊的大致情况：农民的年人均平均收入大约为809.52元，标准差约为155.85元，贫富悬殊不算很大.

【实验】用Excel对 例6 中的数据计算统计量样本均值、样本方差和样本标准差的观测值.

实验准备:

- (1) 函数AVERAGE的使用格式:

  AVERAGE(number1, number2, ...)

  功能：计算给定样本的算术平均值.
- (2) 函数VAR的使用格式:

  VAR(number1,number2,...)

  功能：计算给定样本的方差.
- (3) 函数STDEV的使用格式:

  STDEV(number1,number2,...)

  功能：计算给定样本的标准差.

- 实验方法一：
- (1) 输入数据及统计量名，
- (2) 计算样本均值，在单元格H2中输入公式：
- = AVERAGE(A2:E11)
- (3) 计算样本方差*s*2，在单元格H3中输入公式：
- = VAR(A2:E11)
- (4) 计算样本标准差*s*，在单元格H4中输入公式：
- = STDEV(A2:E11)

	A	B	C	D	E	F	G	H
1	年收入						统计量	
2	924	972	602	882	696		样本均值 $\bar{x}$	809.52
3	800	988	754	1192	644		样本方差 s^2	24288.908
4	916	1266	788	820	926		样本标准差 s	155.84899
5	704	684	962	878	808			
6	870	764	704	614	1010			
7	1040	940	712	846	728			
8	824	408	854	746	742			
9	690	804	888	828	850			
10	574	610	768	792	864			
11	490	852	848	872	738			

AVERAGE(A2:E11)

VAR(A2:E11)

STDEV(A2:E11)

实验二：

- (1) 输入整理数据，如图左所示.
- (2) 在Excel主菜单中选择

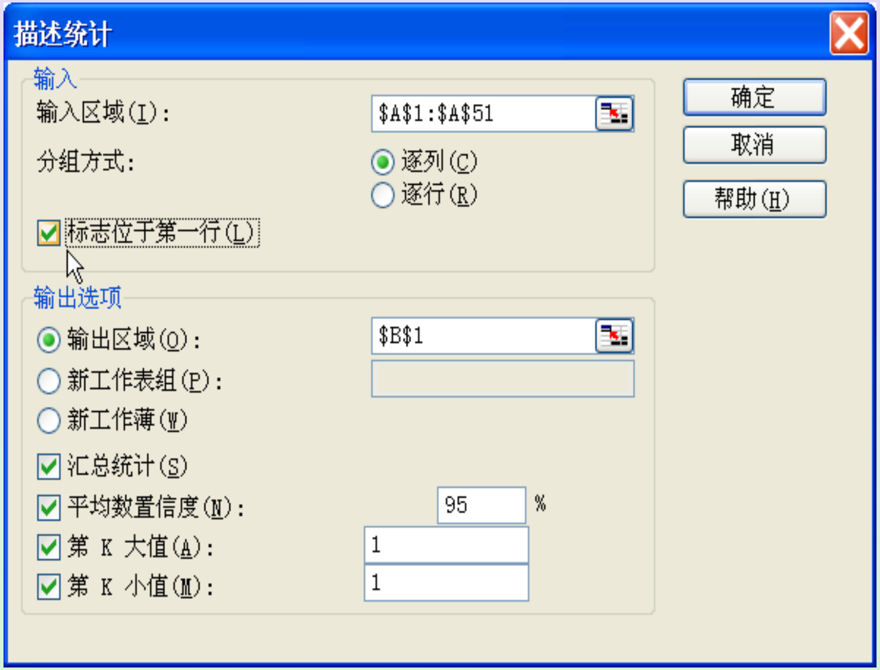
  “工具” □ “数据分析”,

  打开“数据分析”对话框，在“分析工具”列表中

  选择“描述统计”选项，单击“确定”按钮.
- (3) 打开的“描述统计”对话框，依次输入“输入区域”

  和“输出区域”,

  选中“标志位于第一行”复选框，单击“确定”按钮.
- 得到描述统计的结果如图右所示.


描述统计
输入
输入区域(I):
$A$1:$A$51
分组方式:
逐列(C)
逐行(R)
标志位于第一行(L)
输出选项
输出区域(O):
$B$1
新工作表组(P):
新工作薄(W)
汇总统计(S)
平均数置信度(N):
95
%
第 K 大值(A):
1
第 K 小值(M):
1
确定
取消
帮助(H)

	A	B
1	年收入	
2	924	
3	800	
4	916	
5	704	
6	870	
7	1040	
8	824	
9	690	
10	574	
11	490	
12	972	
13	988	
14	1266	
15	684	
16	764	
17	940	
18	408	

B	C
年收入	
平均	809.52
标准误差	22.04037557
中位数	814
众数	704
标准差	155.8489902
方差	24288.90776
峰度	1.446415185
偏度	0.247132211
区域	858
最小值	408
最大值	1266
求和	40476
观测数	50
最大(1)	1266
最小(1)	408
置信度(95.0%)	44.29179212

习题一： 1 - 10

4、充分统计量

定义1.2.2

设总体分布族为 $\{P_\theta : \theta \in \Theta\}$

$X_1, X_2, \boxed{?}, X_n$ 是该总体的iid样本

$T(X_1, X_2, \boxed{?}, X_n)$ 是一个样本统计量。

如果在给定 $T(X_1, X_2, \boxed{?}, X_n) = t$ 条件下

样本 $X_1, X_2, \boxed{?}, X_n$ 的条件分布函数

$F_\theta(x_1, x_2, \boxed{?}, x_n \mid t)$ 与参数 θ 无关,

$T(X_1, X_2, \boxed{?}, X_n)$ 称为 θ 的**充分统计量**。

例1 在贝努力试验中，总体服从两点分布

$$P_\theta\{X=x\}=\theta^{x}(1-\theta)^{1-x}, x=0,1$$

$X_1, X_2, \boxed{?}, X_n$ 是该总体的iid样本

$$P_\theta(X_1=x_1, \boxed{?}, X_n=x_n)=\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

取 $T=\sum_{i=1}^{n} X_i$，

在观察值 $T=\sum_{i=1}^{n} x_i=t$ 条件下

样本的条件分布为

$$P_\theta(X_1 = x_1, \boxed{?}, X_n = x_n \mid \sum_{i=1}^{n} X_i = t)$$

$$= \frac{P_\theta(X_1 = x_1, \boxed{?}, X_n = x_n, \sum_{i=1}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)}$$

$$= \frac{\theta^t (1-\theta)^{n-t}}{C_n^t \theta^t (1-\theta)^{n-t}} = \frac{1}{C_n^t} \text{与}\theta\text{ 无关。}$$

所以，$T = \sum_{i=1}^{n} X_i$是θ的充分统计量.

因子分解定理（定理1.2.1）

设总体分布族为$\{P_\theta, \theta \in \Theta\}$，$T(X)$是充分统计量当且仅当:

样本$X = (X_1, \boxed{?}, X_n)$的联合分布可分解为

$$P(x;\theta) = g(T(x), \theta)h(x)$$

在上例的贝努利试验场合：

$$P_\theta(X_1 = x_1, \boxed{?}, X_n = x_n) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

可见，$T = \sum_{i=1}^{n} X_i$是θ的充分统计量.

例2 *Poison*分布族

例3 正态分布族

例4 均匀分布族

注：充分统计量不唯一

“最小充分统计量”

习题一：11-15

第三节 抽样分布

3.1 特征函数

一维分布： $\varphi_X(t)=E(e^{itX})$

n 维分布： $\varphi_X(t)=E(e^{it'X})$

特征函数是分布密度（分布列）的傅立叶变换

分布函数与特征函数互相唯一确定。

二项分布$B(n,p)$: $\varphi(t)=(pe^{it}+q)^n$

泊松分布$P(\lambda)$: $\varphi(t)=\exp(\lambda(e^{it}-1))$

正态分布$N(\mu;\sigma^2)$: $\varphi(t)=\exp(i\mu t-\frac{1}{2}\sigma^2t^2)$

特征函数的性质

（1）有界性：

（2）线　性：

（3）独立性：

（4）矩关系：

（5）唯一性：

3.2 统计学三大分布

χ^2-分布

设随机变量 $X_1, X_2, \boxed{?}, X_n$ 相互独立且同服从标准正态分布$N(0,1)$，称随机变量

$$\chi^2 = X_1^2 + X_2^2 + \boxed{?} + X_n^2$$

所服从的分布是自由度为 n 的χ^2-分布，

记为 $\chi^2 \sim \chi^2(n)$

定理1.3.3

设$X \sim \chi^2(n)$，则

(1) $\varphi(t) = (1-2it)^{-\frac{n}{2}}$

(2) $E(X) = n; Var(X) = 2n$

定理1.3.4（可加性）

定理1.3.15

$x=(X_1,X_2,\boxed{?},X_n)'$ [illegible] N(0,1)的 iid 样[illegible] ,

则 $Y=x'Ax\sim\chi^2(p)$ [illegible]

$A^2=A$ (幂[illegible])且 $rank(A)=p$

科赫伦（Cochran）分解定理

设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立且同服从标准正态分布$N(0,1)$，又设

$$Q_1 + Q_2 + \cdots + Q_k = \sum_{k=1}^{n} X_i^2$$

其中$Q_j(j=1,2,\cdots,k)$是秩为 n_j的$(X_1, X_2, \cdots, X_n)$非负定二次型，则

$$Q_j(j=1,2,\cdots,k)\text{相互独立且 } Q_j \sim \chi^2(n_j)$$

当且仅当 $\sum_{j=1}^{k} n_j = n.$

定理1.3.2

假设简单样本 $X_1, X_2, \cdots, X_n$ 来自正态总体 $N(\mu, \sigma^2)$， 则有

$$\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2(n)$$

$t-$分布

设随机变量 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且 X 与 Y 相互独立，则称随机变量

$$T = \frac{X}{\sqrt{Y/n}}$$

所服从的分布是自由度为 n 的 $t-$分布，

记为 $T \sim t(n)$.

F -分布

设随机变量 $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, 且 X 与 Y 相互独立，则称随机变量

$$F = \frac{X / n_1}{Y / n_2}$$

所服从的分布是自由度为 n_1, n_2 的 F -分布，

记为 $F \sim F(n_1, n_2)$.

3.3 正态总体的抽样分布

定理1.3.9

设随机变量 $X_1, X_2, \boxed{?}, X_n$ 相互独立且同服从

正态分布 $N(\mu, \sigma^2)$，则其线性组合

$$U = a_1 X_1 + \boxed{?} + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

也服从正态分布，期望方差分别为

$$EU = \mu \sum_{i=1}^{n} a_i, \quad Var(U) = \sigma^2 \sum_{i=1}^{n} a_i^2$$

定理1.3.10

设 $X_1, X_2, \boxed{?}, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的 iid 样本，　　A 是 $p \times n$ 阶矩阵，则

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \boxed{?} \\ Y_p \end{pmatrix} = A \begin{pmatrix} X_1 \\ X_2 \\ \boxed{?} \\ X_n \end{pmatrix} + b \sim N_p(\mu A \mathbf{1}_n + b, \sigma^2 A A^T)$$

定理1.3.11

设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的 iid 样本，则

(1) $\bar{X} \sim N(\mu, \dfrac{\sigma^2}{n})$,

(2) $\bar{X}$ 与 S^2 相互独立;

(3) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$,

定理1.3.12

设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的 iid 样本，则

$$(1) \quad T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

$$(2) \quad \frac{(\overline{X} - \mu)^2}{S^2/n} \sim F(1, n-1)$$

定理

$$(X_1, X_2, \boxed{?}, X_n) \sim N(\mu_1, \sigma_1^{\ 2})$$

$$(Y_1, Y_2, \boxed{?}, Y_m) \sim N(\mu_2, \sigma_2^{\ 2})$$

是来自正态总体的两组异方差独立样本

且方差已知，则

$$U = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \sim N(0,1)$$

定理1.3.13

$$(X_1, X_2, \boxed{?}, X_n) \sim N(\mu_1, \sigma^2)$$

$$(Y_1, Y_2, \boxed{?}, Y_m) \sim N(\mu_2, \sigma^2)$$

是来自正态总体的两组同方差独立样本

且方差未知，则

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \sqrt{\frac{mn(m+n-2)}{m+n}}$$

$$\sim t(m+n-2)$$

定理1.3.14

$$(X_1, X_2, \cdots, X_n) \sim N(\mu_1, \sigma_1^2)$$

$$(Y_1, Y_2, \cdots, Y_m) \sim N(\mu_2, \sigma_2^2)$$

是来自正态总体的两组异方差独立样本

且方差已知，则

$$F = \frac{S_X^2 / \sigma_1^2}{S_Y^2 / \sigma_2^2} \sim F(n-1, m-1)$$

习题一：16-39