第6章

数理统计的基本概念

数理统计的分类：

描述统计学

研究对随机现象进行观测、试验，以取得有代表性的观测值

推断统计学

对已取得的观测值进行整理、分析，作出推断、决策，推断出所研究的对象的规律性

分析

推断统计学

- 参数估计 √√
- 假设检验 √√
- 方差分析
- 回归分析

×

案例1 随着技术的快速更新，社会的迅猛发展，职业的选择受到从业者的高度关注。比如职业发展研究人员做特定职业满意度调查时，把满意度分为四类：“特别不满意”、“不满意”、“基本满意”、“满意”，同时为给出量化评级体系，把这四类满意度对应了得分0、1、3、5。我们知道，全国可能有数十万、甚至数百、数千万的从业者，所以只能采取抽样调查，比如全国随机抽查了100个从业者，得分数据如下：

0	3	3	5	0	5	3	0	5	0	3	3	3	0	1	0	3	1	5	1
3	1	0	3	5	0	5	5	3	1	0	1	1	5	1	3	1	1	3	3
1	3	1	5	0	1	1	3	1	0	5	5	1	5	3	3	3	3	3	3
3	3	5	0	3	1	1	3	3	1	5	1	3	3	0	3	1	3	3	3
0	3	0	3	3	3	0	0	1	1	3	0	0	0	0	3	3	0	3	3

问题是能从100个数据给出该职业满意度得分的概率分布吗？用什么方法，用理论依据是什么？

案例2 考查某品牌、某型号的智能手机寿命，选了18个手机去做（疲劳）寿命实验，数据如下：（单位：小时）

3655 3510 3649 3519 3469 3506 3484 3470

3768 3390 3471 3462 3506 3425 3418 3510

3436 3180

我们关心如下问题：

- ❑ 该型号的手机的疲劳寿命是不是正态分布?
- ❑ 如果是正态分布，参数是多少?

一　总体和样本

总体—— 所研究的对象的全体

所研究的对象的某个(或某些)数量指标的全体

个体—— 组成总体的每一个元素称为一个个体,

即总体的每个数量指标.

抽样—— 从总体中抽取部分个体

样本—— 从总体中抽样出来n的个体，称为**容量为n的样本**

任务：由部份推断总体

案例2（续） 某品牌、某型号智能手机寿命，选了18个手机做（疲劳）寿命实验，数据如下：（单位：小时）

3655	3510	3649	3519	3469	3506	3484	3470	3768
3390	3471	3462	3506	3425	3418	3510	3436	3180

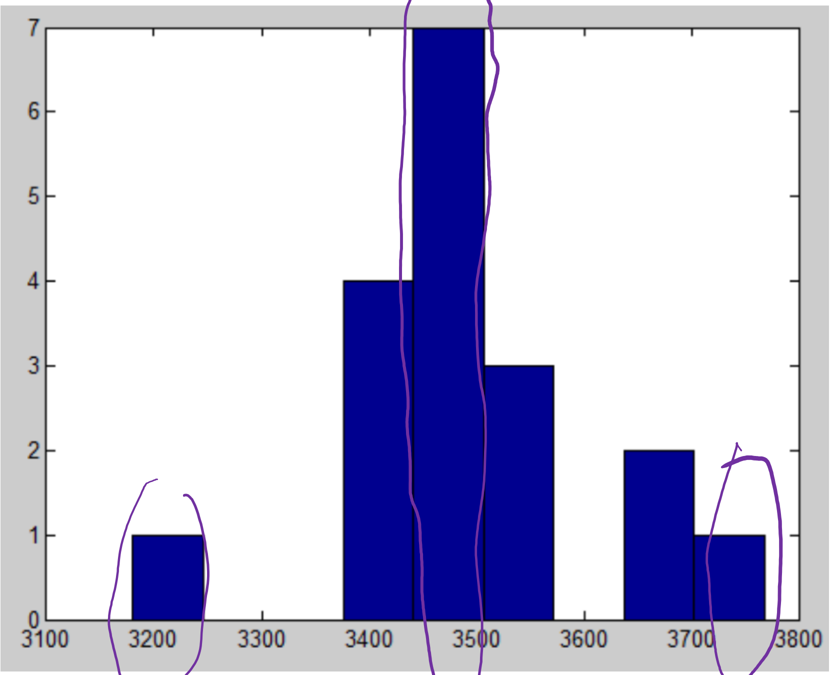
。。。

在这个案例中

- 总体是什么？ ⟵ 这种型号的所有手机的寿命
- 个体是什么？ ⟵
- 样本是什么？ ⟵

3655	3510	3649	3519
3469	3506	3484	3470
3768	3390	3471	3462
3506	3425	3418	3510
3436	3180		

❒ 总体中不同个体的数量指标是不同的，背后有一定的分布，这个分布就是总体分布



总体看成是随机变量(或多维随机变量)，记为 X .

总体分布就是随机变量 X 的分布

设X为总体，设$(X_1, X_2, \dots, X_n)$为容量为n的样本

- 依次对样本的每个个体进行观测得到n个数据：$(x_1, x_2, \dots, x_n)$，称其为**样本观测值**，简称**样本值**.

Ω

- 样本$(X_1, \dots, X_n)$的所有可能取值的集合称为**样本空间**，记为$\mathscr{X}$

$\mathscr{X} = \{(x_1, \cdots, x_n);\ x_i \in [a_i, b_i]\}$

简单随机样本

设$(X_1, \dots, X_n)$是来自总体X的一个样本，若满足：

（1）$X_i, i = 1, \dots, n$与X同分布； ← 代表性

（2）独立性：$X_1, \dots, X_n$相互独立； ← 独立性

则称 $(X_1, \dots, X_n)$ 为**简单随机样本**

$X_1, \dots, X_n$ i.i.d.

设总体 X 的分布函数为$F(x)$，$(X_1, X_2, \dots, X_n)$
为总体 X 的简单随机样本，

则$(X_1, X_2, \dots, X_n)$的联合分布函数为

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

若总体X的概率密度函数为$f(x)$，则
$(X_1, X_2, \dots, X_n)$的联合概率密度函数为

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

二 统计量

$X_1, X_2, \ldots, X_n$

$x_1, x_2, \ldots, x_n$

定义 设$(X_1, X_2, \ldots, X_n)$是取自总体X的一个样本，

设 $g(r_1, r_2, \ldots, r_n)$ 为一实值连续函数，

且不含有未知参数，

则称随机变量 $g(X_1, X_2, \ldots, X_n)$ 为**统计量**.

设$(x_1, x_2, \ldots, x_n)$是一个样本值，

称 $g(x_1, x_2, \quad , x_n)$

为统计量 $g(X_1, X_2, \quad , X_n)$ 的一个**样本值**

常用的统计量

设$(X_1,\dots,X_n)$是来自总体 X 的样本

(1) 样本均值

总体均值

（2）样本方差

总体方差

样本标准差

总体标准差

（3）样本k阶原点矩

总体k 阶原点矩

(4)样本k阶中心矩

总体k 阶中心矩

(5) 顺序统计量与极差

设 $(X_1, X_2, \cdots, X_n)$ 为样本 $(x_1, x_2, \quad, x_n)$ 为样本值,且 $x_1^* \le x_2^* \le \quad \le x_n^*$

定义 随机变量 $X_{(k)} := x_k^*$, $k = 1, 2, \ldots, n$

则称统计量 $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ 为**顺序统计量**.

其中, $X_{(1)} = \min_k \{X_k\}$, $X_{(n)} = \max_k \{X_k\}$

$D_n = X_{(n)} - X_{(1)}$ 为**极差**

案例2（续） 某品牌、某型号智能手机寿命，选了18个手机做（疲劳）寿命实验，数据如下：（单位：小时）

3655 3510 3649 3519 3469 3506 3484 3470 3768

3390 3471 3462 3506 3425 3418 3510 3436 3180

上述部分统计量的样本值为

（1） $\bar{x}=\frac{1}{18}\sum_{i=1}^{18}x_i=3490.4$ ； （2） $s^2=\frac{1}{17}\sum_{i=1}^{18}\left(x_i-\bar{x}\right)^2=14846$ ；

（3） $cm_2=\frac{1}{18}\sum_{i=1}^{18}(x_i-\bar{x})^2=14021$； （4） $cm_3=\frac{1}{18}\sum_{i=1}^{18}\left(x_i-\bar{x}\right)^3=-107430$

案例4 设总体X的概率密度函数为

$$f(x)=\begin{cases}|x| & |x|<1\\ 0 & |x|\geq 1\end{cases}$$

$(X_1, X_2, \dots, X_{50})$为总体的样本，求 (1)$\bar{X}$的数学期望与方差

案例4（续）

$$f(x)=\begin{cases}|x| & |x|<1\\ 0 & |x|\ge 1\end{cases}$$

(2) 下表数据是用matlab根据上述分布生成50个数据计算的样本均值

$$Y=\frac{1}{50}\sum_{i=1}^{50}X_i$$

$$EY=0,$$
$$DY=0.01$$

共100次的样本数据

数据经过四舍五入处理

0.02	-0.07	0.08	-0.02	0.09	-0.13	0.18	0	0	-0.03
-0.12	0.06	-0.02	-0.14	0.22	0.11	0.17	-0.07	-0.06	0.06
0	-0.1	-0.04	0	-0.08	0.03	0.05	0.05	0.07	0.04
0.01	-0.07	0.08	0.03	0.03	-0.06	0.14	0.13	0.11	-0.15
-0.02	-0.03	-0.08	-0.03	-0.17	-0.13	0.01	-0.01	-0.14	0.14
0.08	0.11	0.15	0.16	0.14	0.1	0.04	0	0.06	-0.17
-0.01	0.1	0.11	-0.06	0	0.25	0.02	-0.18	-0.03	0.1
-0.08	0.04	-0.25	0.03	0.07	0.13	-0.07	0.04	0	-0.2
0.08	-0.01	-0.03	-0.07	0.13	0.08	-0.2	-0.06	-0.25	-0.13
0.07	-0.03	0.18	-0.07	0.02	-0.07	-0.15	0.18	0.02	-0.06

$$\bar{y}=0.004484$$
$$s_Y^2=0.0110$$

案例4（续） (3) 计算 $P(|\overline{X}|>0.02)$

$EY=0,$
$DY=0.01$

根据表格中的数据计算随机事件$\{|\bar{X}|>0.02\}$的频率

0.02	-0.07	0.08	-0.02	0.09	-0.13	0.18	0	0	-0.03
-0.12	0.06	-0.02	-0.14	0.22	0.11	0.17	-0.07	-0.06	0.06
0	-0.1	-0.04	0	-0.08	0.03	0.05	0.05	0.07	0.04
0.01	-0.07	0.08	0.03	0.03	-0.06	0.14	0.13	0.11	-0.15
-0.02	-0.03	-0.08	-0.03	-0.17	-0.13	0.01	-0.01	-0.14	0.14
0.08	0.11	0.15	0.16	0.14	0.1	0.04	0	0.06	-0.17
-0.01	0.1	0.11	-0.06	0	0.25	0.02	-0.18	-0.03	0.1
-0.08	0.04	-0.25	0.03	0.07	0.13	-0.07	0.04	0	-0.2
0.08	-0.01	-0.03	-0.07	0.13	0.08	-0.2	-0.06	-0.25	-0.13
0.07	-0.03	0.18	-0.07	0.02	-0.07	-0.15	0.18	0.02	-0.06

例 设 总体X的分布列如下:

X	0	1	2
p	1/3	1/3	1/3

X_1, X_2, X_3是来自于该总体的样本, 求$X_{(1)}$, $X_{(3)}$ 的分布

作业 习题六

- 2，5，6，7
- 开放式案例分析题

补充题

设总体X的分布列如下:

X	0	1	2
p	1/3	1/3	1/3

X_1, X_2, X_3是来自于该总体的样本,

(1) 求$(X_{(1)}\ , X_{(3)})$的联合分布律;

(2) 求 $\mathrm{Cov}(X_{(1)}, X_{(3)})$和$\rho_{X_{(1)}, X_{(3)}}$

$X_{(3)}$ \ $X_{(1)}$	0	1	2
0			
1			
2			