概率统计学习辅导

徐玉华　陆永余　等编

中央广播电视大学出版社

概率统计学习辅导

徐玉华　陆永余　等编

中央广播电视大学出版社

（京）新登字163号

概率统计学习辅导
徐玉华　陆永余　等编
中央广播电视大学出版社出版
新华书店总店科技发行所发行
国防科工委印刷厂印装
开本787×1092　1/32　印张6.125　千字130
1986年12月第1版　1992年2月第5次印刷
印数　71501～73700
定价：2.40元
ISBN 7-304 00184 4 TB·6

重版说明

本书已经过两届电视大学学生使用。此次重版除对错误之处作了更正外，还根据教学基本要求，对偏难的典型例题和自我检查题加了“*”号，以使本书更好地配合教学，发挥其作用。

中央电大

数学教研室

1989. 12

序　言

这本《概率统计学习辅导》（试用本）是与张尧庭教授主编的《概率统计》一书相配合的辅导教材。本书按照教材的章节次序编写了九章辅导材料，目的是指导学员阅读、理解主讲教师编写的基本教材，为此每章首先说明主讲教师在这一章中所要达到的目的；其次把内容提要（尽可能简明扼要）与对学员提出的要求并列，使每一条要求都有明确的针对性；然后，针对各项要求选择一些典型例题以便帮助学员切实掌握相应的教学内容。最后提供一批自我检查题，以便帮助学员自己检查一下是否达到本章的各项要求。

但是由于基本教材较前稿改动较大，修改稿直到一九八六年八月底才脱稿，此后由徐玉华、夏杏菊、王可宪、仲崇彬、陆永余、金乃正、唐承谨、张伟涤和谭英仕等老师根据基本教材的修改稿分头编写，并由徐玉华老师负责统稿，廖祖伟老师审校。这项多人合作的任务，仅仅在短短两个月内完成，加之我们水平有限，欠妥之处在所难免，所以这九章书面辅导材料还需要通过将来的多次教学实践，作进一步的改进。

显然，上述情况决定了本辅导材料只能是一个试用本。因此诚恳希望各地老师和广大学员多多提出宝贵的批评建议，使将来的改编工作搞得更好！

编　者

一九八六年十月

目　　录

第一章　数据的简单分析

一、教学目的

1. 介绍有关求和号 Σ 的一些知识。
2. 说明均值、方差、中位数、频数…等概念。
3. 介绍数据分析的一些简单的和直观的方法。

二、内容提要与要求

基本内容	要　求
1. 求和号 Σ (1) $\sum_{i=1}^{n} a_i \triangleq a_1 + a_2 + \cdots + a_n$ (2) 性质。 (3) 双重和号 $\sum_i \sum_j a_{ij}$ 及性质。	了解求和号 Σ 的含义，熟悉 Σ 的有关性质；弄清 Σ 在公式变形中的作用。
2. 均值、方差 (1) 设已知一组数据 $x_1, x_2, \cdots, x_n$，称 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 为数据的均值，称 $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 为数据方差。	深入理解均值、方差的概念；熟练掌握 $\bar{x}, S^2$ 的计算方法。

(2) 均值、方差性质：

1° $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

2° $\sum_{i=1}^{n}(x_i - c)^2 \geqslant \sum_{i=1}^{n}(x_i - \bar{x})^2$ 等号仅在 $c = \bar{x}$ 时成立。

3° $S^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2$

4° 若 $y_i = bx_i - a$ 则

$\bar{y} = b\bar{x} - a$

$S_y^2 = b^2 S_x^2$ 其中 S_x^2, S_y^2 分别是 $\{x_i\}$ 与 $\{y_i\}$ 的方差。

5° $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

$= \sum_{i=1}^{n}x_i y_i - n\bar{x}\bar{y}$

3. 中位数、标准差、极差、变异系数的概念

了解中位数、标准差 … 等概念。

4. 加权平均

给定一组数 $x_1, x_2, \cdots, x_n$，及 $p_1, p_2, \cdots, p_n$，且 (1) $p_i > 0$，$(i = 1, 2 \cdots, n)$，(2) $\sum_{i=1}^{n}p_i = 1$，称

$\bar{x} = \sum_{i=1}^{n}p_i x_i$ 是 $x_1, x_2, \cdots, x_n$ 的加权平均数，p_i 称为 x_i 的权。

着重理解加权平均数的意义。

5. 直方图

掌握绘制直方图的方法步骤；

(1) 直方图概念。 (2) 绘制直方图步骤。 (3) 频数直方图、频率直方图。	能区分频数直方图和频率直方图。
6. 平方和分解公式 给定一组数据 $x_1,x_2,\cdots,x_n$，及常数 c，称 $\sum_{i=1}^{n}(x_i-c)^2$ $=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-c)^2$ 为平方和分解公式。	初步了解平方和分解公式的意义。

三、典型例题

例 1.1　任给一个常数 c，总有

$$\sum_{i=1}^{n}(x_i-c)^2 \geqslant \sum_{i=1}^{n}(x_i-\bar{x})^2$$

等号仅在 $c=\bar{x}$ 时成立。试证明之。

证　本例在教材第一章 § 3 性质 2 已有证明。在此再给出另一方法：

对于给定的数据 $x_1,x_2,\cdots,x_n$ 函数 $f(c)=\sum_{i=1}^{n}(x_i-c)^2$ 是 c 的二次函数。今考虑其极值，利用微分法。

$$f'(c)=-2\sum_{i=1}^{n}(x_i-c)=0$$

解得
$$c=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x} \qquad (驻点)$$

由于 $f(c)$ 是 c 的二次函数，且二次项系数大于零，故 $f(c)$ 只有最小值 $f(\bar{x})$。于是

$$f(c) \geqslant f(\bar{x})$$

即
$$\sum_{i=1}^{n}(x_i - c)^2 \geqslant \sum_{i=1}^{n}(x_i - \bar{x})^2$$

且当 $c = \bar{x}$ 时取等号。

例 1.2 设一组数据99.3，98.7，100.5，101.2，98.4，99.7，99.3，100.5，求（1）中位数，（2）均值，（3）方差，（4）变异系数。

解 将八个数字依大小顺序排列，设为 $x_1, x_2 \cdots, x_8$，并令 $y_i = 100(x_i - 100)(i = 1,2\cdots,8)$ 列表计算：

序号	1	2	3	4	5	6	7	8	$\sum$
x_i	98.4	98.7	99.3	99.3	99.7	100.5	100.5	101.2	
y_i	−160	−130	−70	−70	−30	50	50	120	−240
y_i^2	25600	16900	4900	4900	900	2500	2500	14400	72600

$$中位数M = \frac{1}{2}(99.3 + 99.7) = 99.5$$

$$\bar{y} = \frac{1}{8} \times (-240) = -30$$

$$\bar{x} = \frac{1}{100}\bar{y} + 100 = 99.7$$

$$S_y^2 = \frac{1}{8}\sum_{i=1}^{8} y_i^2 - \bar{y}^2 = \frac{1}{8} \times 72600 - 900 = 8175$$

$$S_x^2 = \frac{1}{100^2}S_y^2 = 0.8175$$

$$S_x \approx 0.904$$

故
$$\frac{S_x}{|\bar{x}|} = 0.00907$$

按：解此题用到了求均值方差的简化公式，其中 $y_i = 100(x_i - 100)$，在 x_i 中减去100，目的是简化，即使之变为绝对值较小的数进行计算，100是中位数附近的数；乘以100，是去掉小数，“化整”，这在数字计算中是经常用到的一种方法。

例 1.3 举例说明加权平均数的概念。

解 以线段 AB 的内分点 P 为例。设 A,P,B 的坐标分别为 a,x,b。P 分 AB 为 $AP : PB = \lambda$(大于0的常数)，则

$$\frac{x-a}{b-x} = \lambda \quad x = \frac{a+b\lambda}{1+\lambda} = \frac{1}{1+\lambda}a + \frac{\lambda}{1+\lambda}b$$

由于 $\frac{1}{1+\lambda}, \frac{\lambda}{1+\lambda}$ 大于零且 $\frac{1}{1+\lambda} + \frac{\lambda}{1+\lambda} = 1$，故可视为数 a,b 的权。x 则是 a,b 的加权平均。λ 越大表明 P 离 A 越远。

再如，函数 $f(x)$ 在 $[a,b]$ 上的平均数 $\frac{1}{b-a}\int_a^b f(x)dx$。也可看成加权平均。事实上

$$\frac{1}{b-a}\int_a^b f(x)dx = \frac{1}{b-a}\lim_{\substack{n\to\infty \\ \Delta x\to 0}} \sum_{i=1}^{n} f(x_i)\Delta x_i$$

$$= \lim_{\substack{n\to\infty \\ \Delta x\to 0}} \sum_{i=1}^{n} f(x_i)\cdot\frac{\Delta x_i}{b-a}$$

其中 $\frac{\Delta x_i}{b-a}$ 满足(1) $\frac{\Delta x_i}{b-a} > 0$，(2) $\sum_{i=1}^{n}\frac{\Delta x_i}{b-a} = 1$. 故 $\frac{\Delta x_i}{b-a}$ 就是“权”，而 $\sum_{i=1}^{n} f(x_i)\cdot\frac{\Delta x_i}{b-a}$ 正是 $f(x_1),f(x_2),\cdots,f(x_n)$ 的加权平均。取极限的过程，只是说明对 $[a,b]$ 的无穷分割而已。

例 1.4 某食品厂为加强质量管理，对某天生产的罐头

抽查 100 个，测得各个罐头净重如下：

342,340,348,346,343,342,346,341,344,348,346,346,340,344,342,
344,345,340,344,344,343,344,342,343,345,339,350,337,345,349,
336,348,348,344,345,332,342,342,340,350,343,347,340,344,353,
340,340,356,346,345,346,340,339,342,352,342,350,348,344,350,
335,340,338,345,345,349,336,342,338,343,343,341,347,341,347,
344,339,347,348,343,374,346,344,345,350,341,338,343,339,343,
336,342,339,343,350,341,340,341,345,344,342。

（单位：克）试绘出罐头净重的频数、频率直方图。

解 （1）求极差

由表中 $\max\limits_{1\leqslant i\leqslant 100} x_i = 356$，$\min\limits_{1\leqslant i\leqslant 100} x_i = 332$。所以 $R = 24$ 。

（2）分组定组距

由于 100 个数据，拟分 10 组。组距 $d = 2.5$。

（3）定分点，定区间

取区间为[331.05,356.05]。分点为：$x_0, x_1, \cdots x_{10}$，即

331.05,333.55,336.05,338.55,341.05,343.55,
346.05,348.55,351.05,353.55,356.05

这里小数点后多取一位，目的是数据和分点不重合。当然不如此亦可，彼时分成的十个小区间可采用半开半闭 $(x_i, x_{i+1}]$ 式，$(i = 1,2,\cdots,10)$ 。

（4）列频数分布表（用唱票法）

罐头重量(克)	频　　数		频率
331.05～333.55	一	1	0.01
333.55～336.05	下	3	0.03
336.05～338.55	正	4	0.04
338.55～341.05	正正正正一	21	0.21
341.05～343.55	正正正正一	21	0.21
343.55～346.05	正正正正正正	29	0.29
346.05～348.55	正正	10	0.10
348.55～351.05	正下	8	0.08
351.05～353.55	丅	2	0.02
353.55～356.05	一	1	0.01

(5) 画频数直方图

画频率直方图（请读者自行画出）只须将纵坐标易为 $\frac{v}{n \cdot d}$，其图形与频数直方图同。换言之，频率直方图上各个小矩形的面积恰为所观察到的值落在该小区间的概率。这一点对今后学习是十分重要的。

例 1.5　设 $y=\theta+\ln x+\varepsilon$. 试用最小二乘法，利用数据 $(x_1,y_1),(x_2,y_2),\cdots(x_n,y_n)$ 对参数 θ 作估计。

解　实测值 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 应满足

$$y_i=\theta+\ln x_i+\varepsilon_i \qquad (i=1,2,\cdots n)$$

其中 θ 为真值(未知)，ε_i 为误差。

问题是怎样由实测值 (x_i,y_i)，$i=1,2,\cdots n$ 来估计 θ。为此，只须用最小二乘法，求得估值 $\hat{\theta}$ 使

$$\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}(y_i-\theta-\ln x_i)^2$$

为最小。

对上式求导并令为 0：

$$-2\sum_{i=1}^{n}(y_i-\theta-\ln x_i)=0$$

解得 $$\hat{\theta}=\frac{1}{n}\left(\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}\ln x_i\right)=\bar{y}-\ln\left(\prod_{i=1}^{n}x_i\right)^{\frac{1}{n}}$$

这里 $$\left(\prod_{i=1}^{n}x_i\right)^{\frac{1}{n}}=\sqrt[n]{x_1x_2\cdots x_n}$$

叫做 $x_1,x_2,\cdots,x_n$ 的几何平均值。

四、自我检查题

1. 给定数据 11.20,11.12,11.28,11.20,11.40,求(1)均值;(2)方差;(3)标准差;(4)中位数;(5)变异系数。

2. 对 100 个靶进行射击,各打 10 发子弹,只记录击中与否,射击结果如下:

每靶击中数 x_i	0	1	2	3	4	5	6	7	8	9	10
频数 m_i	0	0	1	5	10	23	24	20	11	4	2

求每靶击中数的均值与方差。

3. 设两组数据 $x_1,x_2,\cdots,x_n;y_1,y_2,\cdots y_m$;今将之并为一组:$z_1,z_2,\cdots,z_{m+n}$,证明

$$\bar{z}=\frac{n\bar{x}+m\bar{y}}{m+n}$$

$$S_z^2=\frac{nS_x^2+mS_y^2}{m+n}+\frac{mn}{(m+n)^2}(\bar{x}-\bar{y})^2$$

其中 $\bar{x},\bar{y},\bar{z},S_x^2,S_y^2,S_z^2$ 分别为 $\{x_i\},\{y_j\},\{z_t\}$ 的均值与方差。($i=1,2,\cdots,n;j=1,2,\cdots,m;t=1,2,\cdots,m+n$)。

4. 给定 50 个数据:

13. 39,13. 42,13. 38,13. 53,13. 51,13. 30,13. 40,13. 28,13. 43,13. 46,13. 53,
13. 55,13. 29,13. 24,13. 34,13. 50,13. 65,13. 43,13. 42,13. 38,13. 34,13. 57,
13. 26,13. 33,13. 40,13. 43,13. 50,13. 44,13. 53,13. 48,13. 48,13. 34,13. 59,
13. 35,13. 44,13. 34,13. 33,13. 25,13 28,13. 49,13. 33,13. 26,13. 26,13. 55,
13. 54,13. 37,13. 31,13. 37,13. 33,13 36。

试画出频数直方图。

第二章　事件与概率

一、教学目的

1. 讲述事件与概率等基本概念。

2. 着重介绍概率计算的加法公式和乘法公式，并由此导出一些常见的概率计算公式。

3. 利用条件概率和独立性概念，推导出全概率公式和二项分布。

二、内容提要与要求

基本内容	要求
1. 事件与概率	正确理解事件与概率概念。
(1) 概念。	
(2) 事件间的关系与运算： 事件和、积、差、包含、相等、互逆、互斥。 事件的运算律：交换律、结合律、分配律、德·摩根律。	对照集合论知识，弄清事件运算关系以及与概率运算的联系。
(3) 概率性质 1° $P(\Omega)=1$ 2° $P(\phi)=0$	熟悉概率的基本性质。