



全国高等农业院校教材

渔业生物统计学

● 陈兆祥 主编

● 渔业资源、淡水渔业、海水养殖、水生生物专业用

中国农业出版社

全国高等农业院校教材

渔业生物学

陈兆祥 主编

渔业资源、淡水渔业
海水养殖、水生生物 专业用

中国农业出版社

全国高等农业院校教材

渔业生物统计学

陈兆祥 主编

* * *

责任编辑 白洪信

中国农业出版社出版 (北京市朝阳区农展馆北路2号)
新华书店北京发行所发行 农业出版社印刷厂印刷

850×1168mm32开本 8.625印张 215千字

1995年10月第1版 1995年10月北京第1次印刷

印数 1-3,200册 定价 8.25元

ISBN 7-109-03714-2/S·2361

主编 陈兆祥（上海水产大学）
参编 王绪昌（大连水产学院）
主审 林德光（华南热带作物学院）

前 言

本书是编者在历年各自教学讲义的基础上，按全国高等农业院校教材指导委员会审定的《渔业生物统计学》编写大纲编写而成。内容着重于统计方法在渔业生物学中的应用，而不强调统计公式的推演和证明。本书特点是把实用统计方法与数理统计原理建立了有机联系，既无数理统计的痕迹，亦非单纯操作。抽象的概念均附以直观内容，在弄清统计原理的来龙去脉后再行应用。这样做，期望有助于读者对统计原理的理解及正确使用。

本书紧密联系渔业生产和渔业生物学研究，所选例子多为渔业生物学资料，针对性强，更易为本专业的学生及科技工作者接受。

鉴于研究多指标问题的重要性及电子计算机的广泛使用，介绍“多元分析”的时机已经成熟。故本书安排了“多元分析初步”这一章。这在本科教材中是一个开端，今后可望不断充实。

作为本教材的配套书，编者另外编写了《渔业生物统计习题与计算实习》。在计算实习中使用通用统计软件，以助于计算机使用及提高科研水平。

限于编者水平，书中谬误与不妥之处在所难免，深望使用本书的师生和科技工作者批评指正。

编 者

1993年10月

目 录

第一章 绪论	1
第一节 统计数据	1
第二节 总体与样本	2
第三节 随机样本与统计量	3
第四节 渔业生物统计的任务	4
复习思考题	5
第二章 数据整理	6
第一节 分布表与分布曲线	6
第二节 特征数字	11
复习思考题	18
第三章 概率基础	19
第一节 随机事件和事件的概率	19
第二节 概率计算法则	23
第三节 概率分布	25
第四节 正态分布	28
第五节 二项分布	36
第六节 普哇松分布	39
复习思考题	42
第四章 总体平均数的检验与估计	43
第一节 样本平均数的分布	43
第二节 显著性检验的基本原理	46
第三节 总体平均数的检验	51
第四节 总体平均数的估计	59
复习思考题	64

第五章 总体百分率的估计与检验.....	63
第一节 单个百分率	63
第二节 两个百分率之差	65
复习思考题	68
第六章 非参数检验.....	69
第一节 符号检验法	69
第二节 符号秩和检验	71
第三节 非配对两样本秩和检验	73
第四节 多个样本秩和检验	75
复习思考题	77
第七章 χ^2 检验	78
第一节 χ^2 检验的原理	78
第二节 适合性检验	80
第三节 独立性检验	82
复习思考题	87
第八章 试验设计	88
第一节 试验设计的原则和要求	88
第二节 单因素随机试验	90
第三节 两因素随机试验	98
第四节 随机区组试验	109
第五节 拉丁方试验	114
第六节 正交试验设计	116
第七节 多重比较	127
第八节 缺失数据的弥补	130
第九节 数据转换	132
复习思考题	136
第九章 回归分析	138
第一节 相关与回归的概念	138
第二节 一元线性回归方程	140
第三节 回归显著性检验	145
第四节 估计与预测	149

第五节	相关系数	153
第六节	曲线回归	157
第七节	多元线性回归	167
第八节	逐步回归	176
第九节	协方差分析	185
复习思考题		189
第十章	多元分析基础	190
第一节	多变量的样本	190
第二节	总体均值向量的检验	194
第三节	总体均值各分量的估计	196
第四节	两总体均值向量之差的检验与估计	197
第五节	多元方差分析	200
复习思考题		212
附表		213
参考书目		264

第一章 绪 论

第一节 统计数据

一、变异的数据与随机变量

在渔业生产或科学研究中为了定量地进行分析，我们按预定计划搜集有关的数据资料，例如：一艘渔轮在一个航次中各网的渔获量；一次网获中200尾小黄鱼的体长；某海区内20个水样的盐度；10个样本鱼肉中蛋白质的含量，等等。这样的数据称为统计数据。同一网获中200尾小黄鱼的体长不尽相同，对此我们不会感到惊异，然而这正是统计数据的本质特征。

统计数据毫无例外地参差不齐，是事物变异性的反映。引起变异有诸多因素，其中有的我们可以认识或控制，但也存在许多影响微小的因素，尚不能认识。我们把那些尚不能认识的因素称作随机因素。通常正是种种随机因素的干扰，造成个体的变异。变异普遍存在，程度大小不一。故统计数据就是具有变异性的数据。

变异性的存在，使群体中的个体在同一标志上的量值不能保持一个恒定的值。为了表示全部个体（某标志上的度量值），就需采用变量。例如，太湖中二龄青鱼的体长因个体而异，是一个变量。这个变量取什么值将随取到的个体而定，即具有偶然性。但是偶然中又蕴含着必然，全部个体大小呈规律性散布（称为分布），因而变量取值又有规律依循。我们称这样的变量为随机变量，常以大写字母 X 、 Y 等记之。

二、数据分类

生物性状一般可分为数量性状和质量性状。数量性状如：鱼的年龄、体长、体重、鳍条数、怀卵量；鱼体上的寄生虫数；水体中的含氧量等。对数量性状观测所得到的数据称为定量数据——量数。质量性状如：鱼体的颜色；有鳞与无鳞；存活与死亡；抱卵与不抱卵等等。质量性状一般无法度量。但通过对个体的观察不难确定它属于哪个类别。从而可统计出各个类别中个体的数目，称为定性数据——计数。为了进一步处理，按数据结构把量数与计数又分为连续型与离散型两类。

(一) 连续型数据 在取定的单位下，长度、重量、温度这些变量的量值并不限于整数，整数后面可以带有小数。而且所带小数的位数可随测量仪器精密度的提高而增加。设想仪器可精确到任意程度，则小数的位数将会无限增加。也就是说这类变量可以在某个区间内连续取值，故称连续型。

(二) 离散型数据 年龄、鳍条数、鱼塘中死鱼尾数、礁石上附着的蛹数，这些变量只能取某些间断的数值（一般为整数），在数轴上是一些孤立的点，故称离散型。

第二节 总体与样本

一、总 体

直观地说，总体就是符合指定条件的研究对象的全体。例如，研究对象是舟山渔场越冬带鱼群，那么该鱼群中的全部是我们讨论的总体。其中每条鱼是一个观测单元，称为个体。研究某港湾中礁石上附着的蛹数，那么总体就是全部礁石。每一块礁石是一个观测单元，即一个个体。

如果总体中包含的个体为有限个，则称有限总体。例如，研究的对象是某养殖场的草鱼，那么该场的全部草鱼是讨论的总体，为

有限总体,如果研究的对象是在某些条件下饲养的草鱼,而无时间、空间的限制,那么这样的草鱼数量将是无限,构成一个无限总体。

· 总体就是研究对象的全体,这样说虽然直观,但在统计意义上并不贴切。统计学中所言总体不是指研究对象的实体或某些方面,而是指全部个体在某一标志上的度量值组成的数据集合。也就是说总体可表示为一个随机变量。

二、样 本

为研究舟山渔场越冬带鱼群,若能将该鱼群全部捕获进行观测,从而作出关于该鱼群的结论,当然准确无误,但显然这是无法实现的。即使对捕获到的这一部分,皆作观测也还是难以办到,因为数量仍十分庞大。可行的方法只能是少之又少,只从渔获中抽取少量个体,观测只在取出的个体中进行,即所谓统计抽样。抽出的这一部分称为样本。样本中所包含的个体数称为样本容量。与总体的概念类似,撇开实体,样本也是数据集合,是总体的子集。例如,研究的总体是舟山渔场越冬带鱼群的体长,从渔获中随机地抽取100尾,测得100个体长数据,这100个数据便是取到的一个样本。

一般观测只能施行于样本,而期望了解的却是总体。因此样本必须充分反映总体信息,这是统计抽样的根本原则。只有这样,我们才能由样本提供的信息作出关于总体的正确结论。为了达到这一要求,除样本中包含的个体要有一定数量外,通常还要求随机抽样,即总体中的任何个体都有同等的机会被抽中。随机抽样可采用抽签、拈阄及使用随机数字表等方法。

第三节 随机样本与统计量

一、随机样本

由总体中抽取部分个体组成样本。哪些个体被抽中,具有偶

然性。样本将随抽到的个体而异，即样本具有可变性。我们把可变的样本称作随机样本。通过调查或试验可取得一个具体的样本，即一组数据，它们是随机样本在这次试验中的取值。在今后的讨论中，不论是指随机样本或是指样本值，皆称作样本。在具体问题中只要稍加留心将不致混淆。

随机样本记为 $\{x_1, x_2, \dots, x_n\}$ ， x_i ($i=1, 2, \dots, n$) 为第 i 次从总体中取出的个体。显然它可以是原总体中的任何一个个体，即 x_i 可看作表示原总体的随机变量。

二、统计量

凡是由样本构成的各种量（其中不含有未知的总体参数）称为统计量。简单的统计量如：样本平均数、样本百分率。统计量将随样本而变化，故统计量是随机变量。统计量的可变性正是我们推测总体的着眼点，对统计量的研究是统计学的重要内容。

第四节 渔业生物统计的任务

在渔业生物科学研究中如何进行定量分析，根本问题是处理变异的数据。由于存在变异，一条鱼的怀卵量不足以说明整个群体的怀卵量，一口鱼塘的产量不能代表相同条件下所有鱼塘的产量。通过试验或观测取得的数据一般总是样本，样本的变异有可能使我们迷惑不解。我们必须去区分处理变异与随机变异。安排两组鱼作两种饵料效果的对比试验，试验结果两组数据出现的差异可能是饵料效果（处理）的差异与随机误差的混合，但也可能两种饵料效果并无差异，两组间的差异纯属随机误差。

处理变异数据，从变异中寻求规律，是数理统计学研究的内容。数理统计学为处理变异数据提供了普遍适用的理论和方法。如今在各个科学领域及工、农业生产中数理统计的方法已被广泛采用，成果卓著。数理统计与生物学研究相结合形成生物统计。

在渔业生产与科学研究中，若能正确地运用生物统计这一工具，将使我们的计划与判断都有明确的数量标准，更具科学性，更有说服力。从而使我们能更有效地组织生产与提高科研水平。

统计分析虽然是试验结束取得数据以后的事，但为了使数据能符合统计分析的要求，既无短缺亦无多余，在进行试验之前首先需拟定和设计试验方案，即所谓“试验设计”。试验设计得好，就能以较少的人力、物力和时间取得充分反映总体的信息，达到事半功倍的效果。一切统计推断皆基于试验，数据资料的好坏对推断的正确性起着首要作用。统计分析不能代替试验。试验设计虽然已成为统计学的一个分支，但两者紧密联系，不可分割。一般在《生物统计》教材中，“试验设计”总占一章。

渔业生物统计是处理渔业生物学数据资料的得力工具，工具只有通过使用它的人才能发挥其作用。渔业科技工作者掌握了此项工具，凭借他们的专业知识与实践经验，就能合理地解释数据，并作出试验中研讨问题的正确结论。

复习思考题

1. 何谓统计数据？其本质特征是什么？
2. 何谓“量数”？何谓“计数”？举例说明。
3. 统计学中的“总体”与通常所言“群体”差别何在？为什么可用随机变量表示总体？
4. 统计抽样应遵循的基本原则是什么？试说明样本容量大、小之得失。
5. 何谓统计量？为什么说统计量是随机变量？
6. 简述“渔业生物统计”的任务。

第二章 数据整理

第一节 分布表与分布曲线

通过试验或调查收集到一堆带有变异的数据，乍一看，令人眼花缭乱，不得要领。为了清楚地显示其变动规律，就必须进行统计整理，通常整理成频数（次数）分布表。在着手整理前，首先应检查、核对原始数据，以免遗缺或差错。对异常数据（特大或特小）的舍弃不能凭主观好恶，应遵循一定的统计原则。离散型与连续型两类数据的整理方法有所差别，下面分别予以讨论。

一、离散型数据的整理

离散型数据的整理十分简便，首先找出不相同的数据，按大小依次列出，然后读数据，逐个归类，记下各类中数据的个数。例如，吕泗渔场小黄鱼生殖鱼群，从产卵场外的一次渔获中任意取50尾，测定性腺成熟度，得1—4期的数据50个，分类整理成表2—1，称**样本频数分布表**，或称**样本的频率分布**。

表 2—1 小黄鱼性腺成熟度分布

性腺成熟度X	1	2	3	4	合 计
频 数 f	2	6	14	28	50
频 率 ω	0.04	0.12	0.28	0.56	1

表中性腺“成熟度”X为变量，鱼数f称为**频数**或**次数**。

频数的总和等于观测的个体总数n，即 $\sum f = n$ （ $\sum$ 为求和号）。

f 与 n 的比值为相对频数，称为频率，记作 ω ，即 $\omega = f/n$
显然有 $0 \leq \omega \leq 1$

为了更直观地给出频数分布表所示规律，可将其绘成频数分布图。横轴置变量 X ，纵轴置频数。纵轴若置频率，则称频率分布图。分布图的作法是在变量 X 的各取值点作竖直线段，高度等于相应的频数（或频率）。频率分布图与频数分布图仅是纵轴刻度不同，而图形一样。图2-1是按表2-1绘成的频率分布图。

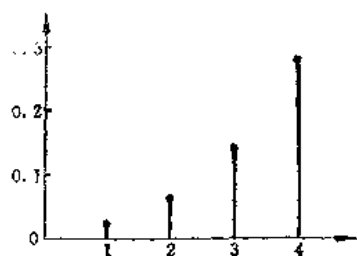


图 2-1 表2-1资料的分布图

由频数分布表或分布图不仅可以看出变量取值的范围，而且还可看到它取各个不同值的次数多寡，故采用“分布”这一形象化的名词。

二、连续型数据的分组整理

(一) 分组步骤 对连续型数据需采用分组的方法进行整理，下面结合例2.1介绍分组步骤。

例2.1 将下列100尾小黄鱼体长数据整理成次数分布表。

241	192	264	191	222	282	274	221	263	221
272	186	220	256	218	220	260	190	224	203
198	216	223	218	240	270	226	208	216	248
224	182	215	214	184	180	215	257	212	235
255	214	205	194	198	189	255	234	250	234
254	210	223	206	207	232	209	248	220	210
229	186	211	190	204	192	230	209	213	233
210	220	167	228	190	212	238	232	226	230
209	188	179	203	193	234	236	240	244	244
242	234	180	205	231	201	219	241	201	196

1. 求极差 数据中最大值与最小值之差称极差，记为 R 。例 2.1 最大值为 282，最小值为 167，于是极差为

$$R = \text{最大值} - \text{最小值} = 282 - 167 = 115$$

2. 确定组数与组距 分多少组视数据多寡而定，数据多，则可多分一些组。分组过多或过少都将影响数据分布规律的显示。一般可分 10—25 组，表 2—2 可作参考。例 2.1 分为 12 组。

组数初步确定后，接着定组距。一个组就是一个小区间。区间的左端点称组下限；右端点称组上限；组段的中点称组中值；区间的长称组距，即

$$\text{组距} = \text{组上限} - \text{组下限}。$$

一般采用等距分组。确定组距可用极差除以组数的商作参考值。例 2.1 中

$$\frac{\text{极差}}{\text{组数}} = \frac{115}{12} = 9.625,$$

若以 9.625 为组距，计算将很不方便。为方便计算，组距可取与该参考值接近的整数或位数较少的小数，例 2.1 组距定为 10 (cm)。

为了使组距为整数，先前初步确定的组数可稍作增减，第一组的下限可取得小于资料的最小值，最后一组的上限可取得大于资料的最大值。

数据归组后一律以所在组的组中值计算，而不管原来是何数值。因此组中值应尽量取整。第一组的组中值以接近资料中的最小值为好。

3. 读数据，划列归组 一般可用记号“+”或“正”划列计数，将数据逐个归入各组。统计各组次数，即整理成频数分布表。分布表上亦可列出频率。表 2—3 是例 2.1 的资料整理成的频数分布表。

若某一数据恰好为两组的界值，该数据如何归组呢？划归前一组或划归后一组二者皆可，但对所有这样的数据归组应按统一

表 2-2 组数与样本容量的关系

样 本 容 量	组 数
50—100	9—12
100—200	12—15
200—400	15—20
400—600	20—25
600以上	25—30

表 2-3 100尾小黄鱼体长频数分布

1	2	3	4	5	6
组 段	组中值X	f	Σf	fX	fX ²
165—175	170	1	1	170	28900
175—185	180	6	6	900	162000
185—195	190	11	17	2090	397100
195—205	200	13	30	2600	520000
205—215	210	16	46	3360	705600
215—225	220	17	63	3740	822800
225—235	230	15	78	3450	793500
235—245	240	10	88	2400	576000
245—255	250	6	94	1500	375000
255—265	260	3	97	780	202800
265—275	270	2	99	540	145800
275—285	280	1	100	280	78400
合 计		100		21810	897900

方法处理。我们采用归上限组的办法。

(二) 绘制分布曲线图

1. 频率曲线 数据整理成频数分布表后进而可绘制分布图。在坐标图中横轴置变量，按各组组限划分小区间，各小区间的长等于组距；纵轴置频数（或频率）。以各小区间为底，对应的频数