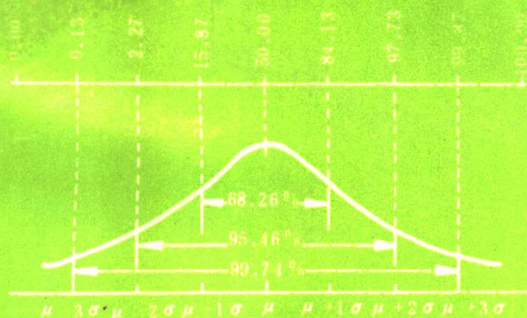


农业技术干部统计学自学丛书

畜牧统计学

杨永年等 编著



东北林业大学出版社

农业技术干部统计学自学丛书

畜牧统计学

杨永年 主编

江占文 副主编
杨尚勇

东北林业大学出版社

内 容 简 介

本书结合畜牧业教学、科研和生产实践,比较系统地、深入浅出地阐述了统计原理和方法,以及畜牧试验设计技术。主要内容有数量资料的整理,平均数和标准差,概率和常见的概率分布,差异显著性检验,卡平方检验,方差分析的基本原理和方法,直线相关和回归,复回归和复相关,协方差分析,畜牧试验设计等。全书各章均有练习题,书后附有常用统计用表和电子计算器使用方法。

本书为畜牧技术人员的自学用书,亦可作为农林院校畜牧专科、中专的教材和教学参考书。

畜 牧 统 计 学

Xumu—Tongjixue

杨永年 等编著

东北林业大学出版社出版发行

(哈尔滨市和兴路8号)

黑龙江省教委印刷厂印刷

开本787×1092毫米 1/32 印张7.75 字数175千字

1990年8月第1版 1990年8月第1次印刷

印数 1—5000册

ISBN 7—81008—155—1/S·40 定价: 2.55元

前 言

畜牧科学的发展,已进入定量地研究各种畜、禽生命现象的阶段,需要用数理统计原理和方法,分析和解释畜、禽经济性状的量变化和量之间的关系,从而获得符合客观规律的结论,用于指导畜牧生产。因此,广大畜牧技术人员和农林院校师生迫切希望有一本适合畜牧专业大、中专学生用的教科书。鉴于这种实际需要,从1989年8月便开始了畜牧统计学的编著工作。

本书是基于二十多年农业教学和科研积累的资料和统计学教学经验,并同畜牧生产第一线 and 数理统计教学第一线的同志们合作,编著而成的。内容侧重于常用统计方法和试验设计技术在畜牧科学上的应用,而不强调各种公式数学推导的严密性。但又尽可能地将书中公式的来龙去脉交待清楚,使读者对统计学原理有较系统地了解,克服过去只会运算不明道理的弊端。为了使读者易晓易学,笔法力图深入浅出,通俗易懂。全书共十章,前九章系统地论述统计学原理和方法,最后一章介绍试验设计技术和相应的试验结果分析方法。

本书承蒙北京农业大学张沅教授进行审稿,并为书写了序。编写过程中,曾得到佳木斯师专高校图书代办站、集贤县科协、佳木斯市民盟、佳木斯市郊区畜牧局等单位的大力协助。在此一并致谢。

由于我们水平所限,书中定有疏漏或谬误,敬请读者批评指正,以便再版修订。

杨永年

1990年8月于哈尔滨

本书编著人员

主 编 杨永年 (佳木斯师范专科学校)

副主编 江占文 (佳木斯师范专科学校)

杨尚勇 (佳木斯市郊区畜牧局)

编 委 吴进发 (集贤县科学技术协会)

董锡文 (佳木斯师范专科学校)

目 录

序	(1)
第一章 数量资料的整理	(3)
第一节 常用统计名词的概念	(3)
第二节 资料的整理	(5)
第二章 平均数和标准差	(13)
第一节 平均数	(13)
第二节 标准差	(16)
第三章 概率和常见的概率分布	(29)
第一节 概率的概念和运算	(29)
第二节 二项分布	(34)
第三节 正态分布曲线	(37)
第四节 抽样分布	(42)
第四章 差异显著性检验	(53)
第一节 差异显著性检验的步骤和方法	(53)
第二节 t 分布和 t 检验	(56)
第三节 总体平均数的置信区间	(60)
第四节 两个样本平均数间的差异显著性检验	(62)
第五节 两个百分数资料差异显著性检验	(69)
第五章 卡平方检验	(75)
第一节 卡平方的概念和分布	(75)
第二节 适合性检验	(79)
第三节 独立性检验	(81)
第六章 方差分析的基本原理和方法	(87)
第一节 平方和及自由度的分割	(87)

第二节	F 分布和 F 检验	(93)
第三节	多重比较	(95)
第七章	直线相关和回归	(107)
第一节	直线相关	(107)
第二节	直线回归	(125)
第三节	相关系数与回归系数的关系	(140)
第八章	复回归和复相关	(144)
第一节	复回归	(144)
第二节	复相关	(152)
第九章	协方差分析	(160)
第一节	协方差分析的意义	(160)
第二节	协方差分析的原理	(161)
第三节	协方差分析的步骤和方法	(169)
第十章	试验设计	(176)
第一节	畜牧试验的种类	(177)
第二节	试验设计的基本原则	(179)
第三节	试验设计方法	(181)
附表		
附表 1	正态曲线下的面积表	(203)
附表 2	随机数字表	(204)
附表 3	t 值表	(208)
附表 4	χ^2 (卡平方) 表	(209)
附表 5	F 值表	(210)
附表 6	SSR 值表	(216)
附表 7	q 值表	(220)
附表 8	r 及 R 的显著数值表	(224)
附录	计算器的使用方法	(225)

序

当今，定量性的研究方法已经成为现代科学技术发展的重要标志之一。生物科学的发展也不例外，进入了定量地研究生命现象的阶段。畜牧科学属于应用生物学的一个重要领域，它涉及到许多量的问题，需要分析和解释量的变化与量之间的关系。只有这样，才能对这类事物的本质获得规律性的认识。畜牧调查与各类畜牧科学实验提供了大量数据资料。这些数据往往是千差万别和杂乱无章的。只有通过适当的统计分析，才能从中推导出客观的规律性。应用这些规律就能够进一步指导畜牧生产的发展。

畜牧生物统计是应用于畜牧领域生物现象的统计方法。它是应用数理统计原理分析和推断畜牧数量资料的一门学科。因此，它是进行畜牧调查和实验必须掌握的研究工具。

多年来，广大初级畜牧工作者和中等畜牧专业师生迫切需要有一本适合于畜牧专科或专科以下学校使用的生物统计学教科书。本书作者多年从事生物统计教学与科研工作，积累了一些经验，又参阅了大量有关论著和文献，编著了此书。这必将对我国畜牧中专、专科生物统计教学工作是一大促进。

本书内容侧重于各种统计方法在畜牧科学上的应用，而不强调各种公式的严密推导。但又尽可能地将大部分公式的来源交待清楚，使读者能够对统计学的原理有一概括了解。畜牧试验设计在畜牧科学中占有重要地位，而试验设计又与生物统计有着密切的关系。因此，本书兼述了生物统计理论和方法，以及试验设计与试验结果分析。书中列举了畜牧科

学中经常遇到的大量实例，便于读者学习、理解和应用生物统计方法解决实际问题。

本书内容系统性强，深入浅出，紧密结合畜牧科研、生产，不仅可作为中等畜牧专科学学生的教材，对于具有高中文化程度的畜牧工作者也是一本很好的自学参考书。

張 沅

1990年元月于北京

第一章 数量资料的整理

在进行家畜、家禽的科学研究和育种工作时,经过调查和试验,常常收集到一些重要的数量资料,为我们研究问题提供了许多有用的信息。然而,这些原始数据资料,还只是一堆杂乱无章的数字,显示不出任何规律性来。畜牧工作者只有运用生物统计的方法,对这些数量资料进行分组、归类、整理,使资料系统化,以便进一步据此分析研究,才能透过现象看本质,掌握其发生和发展的规律。这种整理统计分析的第一步工作,就是数量资料的整理。

因此,在讲述统计原理和方法前,首先讲述数量资料的整理方法。

第一节 常用统计名词的概念

学习生物统计,需要弄清下面一些常遇到的统计名词的概念。

一、总体和样本

在进行家畜、家禽的科学试验时,要对动物的一些性状进行调查和记载。例如,我们调查来杭鸡 500 日龄产蛋数,其目的是了解来杭鸡这个品种的产蛋率。为此,要对不同地区 and 不同年份能够产蛋的来杭鸡所有个体进行调查,才能得到真正可靠的结果。像这样根据研究目的确定的研究对象的

全体称为总体，它是由性质相同的所有个体组成。个体是组成总体的最小研究单位。

总体中的个体数可以是有限多个，也常常是无穷多个。因此，对总体的所有个体都进行调查记载是不可能办到的，只能从总体中抽取一定数量的个体进行研究。这些个体就组成了样本。样本是总体的代表。为使样本具有代表性而避免人为的偏差，样本要从总体中随机地抽取，即随意地、不加选择地任意抽取。这样得到的样本称为随机样本。在家畜的科学试验和育种工作中调查所得的数据，一般都属于样本。

二、数量资料和变数

为了了解家畜的某一性状，经过调查和记载，得到一些观测值，称为数量资料，简称资料。例如，调查 100 只来杭鸡 500 日龄产蛋数，得到 100 个观测值便是来杭鸡 500 日龄产蛋数的样本资料。因为生物体的遗传基础和遇到的环境条件不尽相同，因而同一品种不同个体的同一性状便有不同的表现，出现不同程度的变异，从而得到的观测值就是变化的数量，简称变数或变量。因为，这些观测值，人们事先不能确定，带有偶然性、随机性，故又称为随机变数或随机变量。

变数可分为连续性变数和非连续性变数。例如家畜的体长、体重、产乳量和产毛量等就都是变化的数量，一般用米、公斤等单位表示。这些变数并不限于整数，根据精确度的要求，在相邻的整数中间，可以有很多中间数。例如乳牛的每日产乳量在 20kg 与 21kg 之间，可以有 20.1kg、20.11kg 等许许多多中间数。这类变数称为连续性变数。家畜的产仔数、成活数、产蛋数等变数，其观测值就只能取整数，在相

邻的整数之间没有中间数出现。这类变数称为非连续性变数。

三、参数和统计量

调查家畜的某一性状，如果对总体的所有个体都进行了调查和记载，那么由总体的全部个体观测值计算所得说明总体资料的特征值称为参数，如总体平均数、标准差等。假设调查来杭鸡500日龄产蛋数总体中的所有个体，得到了总体的数量资料，并计算出总体资料的平均数是235个，这一特征值便称为参数。但总体资料实际上是不易获得的，一般只能用样本的观测值进行计算。计算所得说明样本资料的特征值称为统计量。例如调查100只来杭鸡500日龄产蛋数，从这一样本资料计算出的平均数是205个，这一特征值称为统计量。样本的统计量是用来估计总体参数的，所以说统计量是参数的估计值。

第二节 资料的整理

对家畜某一性状进行调查和记载，得到了一份数量资料，如果资料所含的观测值不多，如几个或十几个，那么这样的资料不进行整理就能看出变化情况，因此无需进行整理。观测值较多的资料，如几十个到几百个，不经过整理就难以看出分布和变化的情况。例如表1.1是100头大白猪成年母猪产仔一月的窝重资料，不加整理就不易看出资料分布和变化的情况。同时，经过整理的资料，进一步统计分析时，还可以简化计算过程，所以对观测值较多的资料的整理是统计分析的第一步。

表1.1 100头大白猪成年母猪产仔一月的窝重 单位: kg

12.6	40.5	49.6	55.0	58.2	63.8	68.7	20.1	41.0	49.9
96.7	39.8	45.6	51.8	56.0	94.4	85.8	69.5	40.2	69.8
55.2	58.7	64.0	67.8	113.0	25.5	41.5	50.4	55.5	104.7
71.7	79.5	46.0	52.5	71.8	47.2	54.9	60.5	67.5	72.5
101.1	58.9	68.6	74.5	83.0	86.7	91.8	80.5	85.0	31.5
53.0	67.3	61.2	72.3	54.3	76.9	56.4	67.1	64.0	74.5
89.5	81.0	70.8	72.0	42.5	53.0	66.5	67.5	90.5	75.5
56.7	63.7	74.0	87.5	62.2	56.8	86.8	76.8	66.5	64.5
34.4	43.9	51.6	36.0	57.2	54.1	65.0	88.5	73.0	70.5
84.0	62.5	75.9	65.0	76.5	62.8	65.5	81.5	63.3	79.5

一、频数分布表

资料整理最常用的方法是编制频数分布表。对于收集到的数据资料,首先划分出等距离的组区间,然后将数据列入各相分的组内,即可出现一个有规律的表格,我们称它为频数分布表。每组数据出现的次数称为频数,一般用 f 表示。频数分布表的编制方法因变数种类的不同而异,现将连续性变数和非连续性变数频数分布表的编制方法分别介绍于下:

(一) 连续性变数资料的整理

现以表 1.1 的资料为例说明连续性变数资料整理方法。

1. 确定组数 编制频数分布表首先是确定分几个组。分组过少,不能充分反映观测值的分布情况,同时用它计算出的统计量精确度也较低。但分组过多,不容易看出观测值的集中情况,利用它计算统计量时也比较费事。因此,分组不宜过多或多少,一般都是根据资料所含观测值的多少来确定,可参考表 1.2 的分组数。

大白猪成年母猪产仔一月窝重资料的观测值为 $n=100$, 适宜的分组数为 $k=8\sim 16$, 现暂定分为 10 组,在分组过程中

表1.2

不同数量观测值的样本适宜分组数

样本观测值的个数 (n)	50	100	200	300	500	1000
适宜分组数 (k)	5—10	8—16	10—20	12—24	15—30	20—40

还可以根据情况适当增加或减少。

2. 计算组距 组的大小或组与组的距离称为组距，一般用 I 或 (i) 表示，而用 R 表示最大观测值（记 X_{max} ）和最小观测值（记 X_{min} ）之差，称为极差，又称为全距。计算组距公式为

$$\text{组距 } (i) = \frac{R}{K}$$

本例极差 $R = 113.0 - 12.6 = 100.4\text{kg}$ ，则组距

$$(i) = \frac{100.4}{10} = 10.04(\text{kg}),$$

一般都以近似的整数作为组距，本例的组距可定为 10kg 。

3. 分组 每个组都应有明确的界线，才能使每个观测值划分到相应的组内，组的界限称为组限，位于每组下端数值小的称为下限，位于上端数值大的称为上限。排列组限时首先从第一组开始，通常是把最小的观测值安置在第一组的中心位置上，这样可以使第一组一般只有一两个观测值，比较容易反映出最低和最高的组所含观测值最少的规律。本例最小观测值 $X_{min} = 12.6$ ，组距 $(i) = 10$ ，则按下式计算

$$\text{第一组下限} = X_{min} - \frac{1}{2}(i) = 12.6 - \frac{1}{2} \times 10 = 7.6 \approx 8.$$

第一组下限值加上一个组距值作为第一组的上限和第二组的下限值，以此类推，直至最大数据包括在最后一组内为止。一般采用只写下限，而不写上限的办法，恰好在各组上限的

值应归到后一组去。

组限确定后，可采用记选票的方法，把各个观测值按其数值的大小，划记在相应的格子内，最后统计出各组观测值的频数（即次数）。

4. 求组中值 组中值是上下组限间的中点值，又称组值，是各组观测值的代表值。计算方法是

$$\text{组中值 } (x) = \frac{\text{下组限} + \text{上组限}}{2}$$

本例第一组的组中值为 $(8+18)/2=13$ 。因各组的组距相同，第一组的组中值加组距即为第二组的组中值，即 $13+10=23$ ，以此类推，求出所有的组中值。

各组的组限、频数和组中值确定以后，便可列出频数分布表。

表1.3 100头大白猪成年母猪产仔一月窝重频数分布表

组区间 (kg)	组中值 (x_i)	划 记	频数 (f_i)
8—	13	—	1
18—	23	丁	2
28—	33	下	3
38—	43	正正	10
48—	53	正正正正	19
58—	63	正正正正正	24
68—	73	正正正正	20
78—	83	正正丁	12
88—	93	正—	6
98—	103	丁	2
108—	113	—	1
总和(Σ)			100

(二) 非连续性变数资料的整理

现以表 1.4 资料为例,说明非连续性资料频数分布表的制作方法。

表 1.4 70头成年母猪的繁殖力资料 单位: 头

7	8	11	14	10	12	11	10	10	9
10	12	11	10	10	11	9	12	8	10
12	10	10	11	8	10	8	10	11	13
10	9	11	12	10	12	9	9	11	10
11	11	13	11	14	13	10	11	13	11
13	10	10	9	11	11	8	9	9	11
10	7	10	13	12	12	13	10	11	9

这一资料的繁殖力在 7~14 头范围内变化,如以 1 头作为组距,分为 8 组,对 70 个观测值的资料是比较适宜的分组数。将各个观测值进行分组,统计出各组的频数后,便可列出表 1.5 的频数分布表。

表 1.5 70头成年母猪繁殖力频数分布表

产仔头数 (x)	7	8	9	10	11	12	13	14	总频数(n)
频 数 (f)	2	5	9	20	17	8	7	2	70

有些资料非连续性变数变化范围比较小,如母羊的繁殖力等可采用上例的方法,以 1 作为组距进行分组,组数可少于一般的适宜分组数的范围。对于变化幅度较大的资料,如母鸡的年产蛋数等,若以 1 作为组距,则分组过多,因此可大于 1,使分组数控制在适宜分组数的范围之内。

二、频数分布图

试验和调查的数量资料除用频数分布表来表示分布情况外,还可用频数分布图表示。频数分布图能够更形象地表明频数分布情况,因而在农业科技书刊和试验报告中应用比较

广泛。常用的频数分布图有柱形图和多边形图。

(一) 柱形图

现以 100 头大白猪成年母猪产仔一月窝重的频数分布表为例，说明柱形图的绘制方法。先在坐标纸上画一直角坐标，横轴表示各组的组限。本例共分 11 个组，在横轴上划出 11 个等分，每一等分代表一个组，标明各组的组限，先写第一组的下限，第二组的下限兼作第一组的上限，这样一直写到最后一组的上限。在组限的下方要写明研究的内容和数量单位，纵轴表示各组的频数。本例频数最多的组为 24，可将纵轴分成 5 个等分，每个等分表示频数为 5，并在纵轴的左侧写明频数。然后，在横轴各组的位置上，按本组观测值的频数画出相应高度的柱形。这样各组便以柱形的高度或面积表示观测值出现的频数。最后，在图的下方写明图的名称（图 1.1）。

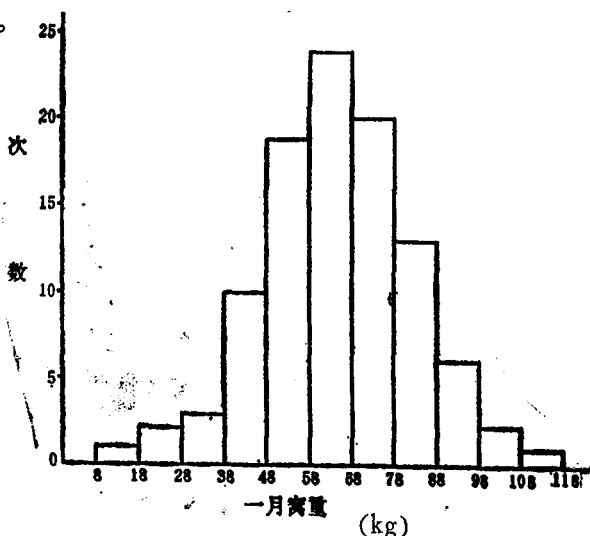


图 1.1 100 头大白猪成年母猪产仔一月窝重频数分布柱形图