




全国高等农业院校教材

全国高等农业院校教材指导委员会审定



# 生物统计

• 吴仲贤 主编

• 畜牧 动物遗传育种专业用



北京农业大学出版社

## 前 言

本书适用于畜牧专业或与生物科学有关的专业。主要内容包括：描述性统计学（1~5章）；推断性统计学（6~9章）；方差分析及试验设计（10~15章）；相关与回归分析（16~20章）；以及21章的协方差分析和22章的非参数统计学。

本书侧重于统计学的应用，对其理论尽可能简略一些。其中个别章节的内容有一点初等微积分和矩阵代数的知识，如果略去不读也不妨碍对结论的应用。

本书在教学安排上可较为灵活，基本上可供50，70，90三种学时使用。编者对不同学时应学的章节建议如下：

50学时的学习章节：①→②→③→④→⑤→⑧→⑨→⑩→⑫→⑬→⑭→⑯→⑰→⑱。

70学时可在50学时的基础上，可再增加如下章节：⑥→⑦→⑮→⑯→⑰。

90学时除了删节个别章节外，基本可以学完此书。

各章的习题放在本书的正文之后，供教师和学生选择。

本书的第11章是张勤博士编写的，第20章是杨子恒博士编写的。

为了保证本教材的质量，专门聘请四川农业大学的明道绪教授主审，提出许多宝贵意见。张勤博士又审阅了一些主要章节，对此深表谢忱。

在教材出版过程中，曾请吴常信教授通读全书；王爱国博士和张勤博士在一校时又付出了辛勤的劳动；副编审沈永华先生对本书的文字也做了一些润色工作；于凡工程师绘制了本书的全部插图，谨在此对以上人员致以衷心的感谢。

编者

1993年1月

# 目 录

<b>第一章 概论</b> .....	1
第一节 生物统计学发展简史 .....	1
第二节 生物统计学的地位 .....	2
第三节 统计学的几个基本概念 .....	4
一、试验与事件 .....	4
二、频率与概率 .....	4
三、总体与个体 .....	5
四、抽样与样本 .....	5
五、参数与统计量 .....	6
六、准确度和精确度 .....	6
第四节 学习生物统计 .....	7
<b>第二章 资料整理</b> .....	9
第一节 统计表 .....	9
一、统计表的概 念 .....	9
二、统计表的种类 .....	10
第二节 资料的分类 .....	11
一、计量资料 .....	11
二、计数资料 .....	11
第三节 数据的分组 .....	12
一、间断性变数资料的整理与分组 .....	12
二、连续性变数资料的整理和分组 .....	13
第四节 统计图 .....	16
一、统计图的作用 .....	16
二、统计图的种类 .....	17
<b>第三章 集中趋势的度量</b> .....	22
第一节 算术平均数 .....	22
一、算术平均数的定义 .....	22

二、用频数分布表求平均数 .....	22
三、变数线性变换对平均数的影响 .....	24
四、用假定平均数等级法求平均数 .....	25
五、平均数的性质 .....	26
第二节 中位数、众数、几何平均数和调和平均数 .....	27
一、中位数 .....	27
二、众数 .....	29
三、几何平均数 .....	31
四、调和平均数 .....	33
第三节 5种平均数关系及评价 .....	33
一、三种平均数之间的关系 .....	33
二、对5种度量集中趋势的指标的评价 .....	34
<b>第四章 离散性的度量</b> .....	<b>35</b>
第一节 标准差 .....	35
一、标准差的定义 .....	35
二、分组资料的计算方法 .....	37
三、变数的线性变换对标准差的影响 .....	38
四、联合均方 .....	39
第二节 范围 .....	39
一、范围 .....	39
二、内百分位范围 .....	41
三、中四分位范围 .....	42
第三节 平均绝对离差 .....	42
第四节 相对离散度 .....	43
<b>第五章 正态分布</b> .....	<b>45</b>
第一节 随机变量及分布 .....	45
一、分布 .....	45
二、随机变量 .....	45
三、离散随机变量和连续随机变量 .....	46
第二节 正态分布 .....	46
一、正态分布的概念 .....	47
二、正态分布的定义 .....	48

三、正态分布的性质 .....	48
四、与正态分布有关的表格 .....	50
<b>第六章 抽样分布 .....</b>	<b>58</b>
<b>第一节 统计量 .....</b>	<b>58</b>
一、什么叫统计量 .....	58
二、简单随机样本 .....	58
三、概率抽样 .....	58
<b>第二节 正态总体样本平均数和均方的抽样分布 .....</b>	<b>59</b>
一、样本平均数的分布 .....	59
二、样本方差的分布 .....	61
三、从一个正态总体的抽样实验 .....	62
<b>第三节 <math>t</math>分布 .....</b>	<b>68</b>
一、 $t$ 分布 .....	68
二、 $t$ 分布的性质 .....	68
三、 $t$ 分布表和 $t$ 分布的双侧分位数表 .....	68
<b>第四节 <math>\chi^2</math>分布 .....</b>	<b>69</b>
一、 $\chi^2$ 分布的定义 .....	69
二、 $\chi^2$ 分布的性质 .....	70
三、分布的适合性检验 .....	70
四、 $\chi^2$ 分布的有关表格 .....	71
<b>第五节 <math>F</math>分布 .....</b>	<b>72</b>
一、 $F$ 分布的定义 .....	72
二、 $F$ 分布的性质 .....	72
<b>第七章 统计推断 .....</b>	<b>74</b>
<b>第一节 参数估计 .....</b>	<b>74</b>
一、参数的点估计 .....	74
二、参数的区间估计 .....	75
<b>第二节 假设检验 .....</b>	<b>81</b>
一、假设检验的步骤 .....	82
二、两类错误 .....	85
<b>第三节 单侧和双侧检验 .....</b>	<b>89</b>
一、双侧检验和单侧检验的定义 .....	89

二、双侧和单侧检验实例 .....	90
<b>第八章 两个样本平均数的比较</b> .....	93
第一节 两个独立样本的平均数的比较 .....	93
一、抽样分布 .....	93
二、 $Z$ 检验 .....	94
三、 $t$ 检验 .....	95
四、等样本含量 .....	98
五、单侧与双侧检验 .....	98
六、未知 $\sigma_1^2, \sigma_2^2$ , 且 $\sigma_1^2 \neq \sigma_2^2$ , 假设检验 $H_0: \mu_1 = \mu_2$ .....	99
第二节 样本含量的确定 .....	103
第三节 配对样本平均数间的比较 .....	105
一、配对比较的计算 .....	105
二、配对比较能提高试验的精确度 .....	107
<b>第九章 率的显著性检验</b> .....	108
第一节 二项分布 .....	108
一、定义 .....	108
二、二项分布的性质 .....	109
三、连续性矫正 .....	110
第二节 率的区间估计 .....	112
第三节 样本率对总体率的检验 .....	113
第四节 两个率的比较 .....	114
一、两组独立样本率的比较 .....	114
二、配对样本率的比较 .....	116
三、两个率比较时样本含量的估计 .....	117
<b>第十章 方差分析及试验设计</b> .....	120
第一节 试验设计的一些概念 .....	120
第二节 试验设计三原则 .....	122
一、随机化 .....	122
二、局部控制 .....	123
三、重复 .....	123
四、试验设计三原则关系图 .....	124
第三节 方差分析的基本原理 .....	125

一、平方和与自由度的剖分 .....	125
二、 $F$ 分布与 $F$ 检验 .....	129
三、多重比较 .....	130
<b>第十一章 方差分析的数学模型</b> .....	<b>132</b>
第一节 模型 .....	132
一、线性模型的概念 .....	132
二、线性模型的分类 .....	133
三、按参数的性质分类的模型的基本含义 .....	133
第二节 单因子模型的平方和、均方和期望均方 .....	136
一、模型 .....	136
二、平方和、自由度和均方 .....	137
三、期望均方 .....	138
第三节 双因子交叉分组模型的平方和、均方和期望均方 .....	145
一、双因子无交互作用 .....	145
二、双因子有交互作用 .....	151
第四节 双因子套分组模型 .....	160
一、模型 .....	160
二、平方和、自由度和均方 .....	161
三、期望均方 .....	161
<b>第十二章 完全随机化设计</b> .....	<b>169</b>
第一节 随机化分组方法 .....	169
第二节 等重复的完全随机化设计 .....	170
一、计算各项平方和与自由度 .....	170
二、建立方差分析表 .....	172
三、多重比较 .....	172
第三节 不等重复的完全随机化设计 .....	177
第四节 $t$ 检验和 $F$ 检验的关系 .....	179
一、 $t$ 检验 .....	180
二、 $F$ 检验 .....	180
第五节 完全随机化试验设计的优缺点 .....	181
一、优点 .....	181
二、缺点 .....	181

<b>第十三章 多向分类的方差分析</b> .....	182
<b>第一节 随机化完全区组设计</b> .....	182
一、随机区组的动物试验 .....	182
二、随机化完全区组平方和与自由度的剖分 .....	183
三、随机化完全区组设计实例 .....	185
四、配对设计与随机化完全区组设计 .....	187
五、漏失数据的处理 .....	189
六、相对效益的测度 .....	190
七、随机化完全区组设计的特点 .....	192
<b>第二节 双向区组——拉丁方设计</b> .....	193
一、拉丁方实例 .....	195
二、拉丁方中漏失数据的处理 .....	197
三、效益上得益的估计 .....	198
四、拉丁方设计的优缺点 .....	199
<b>第三节 平衡不完全区组设计</b> .....	200
<b>第四节 裂区设计</b> .....	203
一、裂区设计的应用 .....	204
二、几种不同设计的裂区安排 .....	205
三、裂区设计实例 .....	208
<b>第十四章 析因试验</b> .....	214
<b>第一节 一些基本概念</b> .....	214
一、符号与定义 .....	214
二、三种效应 .....	215
三、正交比较 .....	217
<b>第二节 两因子的析因试验</b> .....	218
一、 $2^2$ 析因试验 .....	218
二、 $3 \times 4$ 析因试验 .....	220
<b>第三节 三因子的析因试验</b> .....	223
<b>第四节 系统分组</b> .....	227
一、次级样本含量相等的方差分析 .....	228
二、次级样本含量不等的方差分析 .....	230
三、试验单元与抽样单元 .....	233



<b>第十五章 数据的变换</b> .....	234
<b>第一节 非加性检验</b> .....	234
一、非加性测验法 .....	234
二、变换函数的确定 .....	236
<b>第二节 方差同质性检验</b> .....	237
一、 $k$ 个等含量样本的方差同质性检验 .....	237
二、 $k$ 个方差同质性的Cochran检验 .....	237
三、 $k$ 个方差同质性的Bartlett检验 .....	242
<b>第三节 方差稳定性变换</b> .....	244
一、平方根变换 .....	245
二、对数变换 .....	246
三、反正弦变换 .....	248
四、倒数变换 .....	249
<b>第十六章 一元线性回归</b> .....	252
<b>第一节 概说</b> .....	252
一、什么是回归分析 .....	252
二、一元线性回归的数学模型 .....	252
三、回归分析要解决的问题 .....	253
<b>第二节 回归分析</b> .....	253
一、参数 $\beta_0$ 和 $\beta_1$ 的最小二乘估计 .....	253
二、实例 .....	255
三、回归方程的显著性检验 .....	256
四、 $b_0$ 和 $b_1$ 的显著性检验 .....	258
五、两条回归直线的比较 .....	261
<b>第三节 利用回归方程进行估计和预测</b> .....	264
一、对 $\beta_0 + \beta_1 X$ 的区间估计 .....	264
二、利用回归方程进行预测 .....	265
<b>第四节 可化为直线的曲线回归</b> .....	268
一、计算基本统计量 .....	269
二、建立估计的回归方程 .....	270
三、对回归方程进行显著性检验 .....	270
<b>第十七章 多元线性回归</b> .....	271

第一节 多元线性回归的数学模型及矩阵表示 .....	271
第二节 参数 $\beta$ 的最小二乘估计 .....	272
一、参数 $\beta$ 的最小二乘估计 .....	272
二、求解求逆紧凑变换法 .....	274
三、 $b$ 的期望和方差 .....	278
四、实例 .....	279
第三节 二元线性回归分析 .....	280
一、回归面 .....	280
二、变量的中心化 .....	281
三、实例 .....	283
第四节 多元线性回归的检验问题 .....	285
一、回归关系的显著性检验 .....	285
二、回归系数的显著性检验 .....	286
三、回归平方和的剖分 .....	288
<b>第十八章 相关分析</b> .....	293
第一节 简单相关 .....	293
一、相关系数 .....	293
二、双变量总体 .....	294
三、样本相关系数的计算 .....	296
四、相关系数的显著性检验 .....	297
五、相关关系的显著性检验 .....	298
六、 $\rho$ 的置信区间估计 .....	299
七、相关系数的齐性检验 .....	300
八、相关与回归 .....	302
第二节 偏相关系数 .....	304
一、偏相关系数的定义 .....	304
二、变量的“标准化”变换 .....	304
三、偏相关系数的计算 .....	306
四、实例 .....	307
五、复相关系数 .....	310
<b>第十九章 逐步回归</b> .....	313
第一节 自变量的选择 .....	313

第二节	相关系数矩阵 .....	315
第三节	实例 .....	318
<b>第二十章</b>	<b>非线性模型基础 .....</b>	<b>329</b>
第一节	一般线性模型与变量转换 .....	329
一、	正交多项式 .....	330
二、	对因变量 $y$ 的一个转换族 .....	333
三、	方差稳定化转换 .....	335
四、	百分比观察值的转换 .....	338
五、	对变量转换的评价 .....	339
第二节	非线性模型概述 .....	340
第三节	非线性最小二乘原理 .....	342
第四节	非线性模型的参数估计方法 .....	345
一、	迭代算法的一般步骤及可能出现的问题 .....	346
二、	Gauss-Newton法 .....	347
三、	最速下降法 .....	349
四、	阻尼最小二乘法 .....	350
五、	两种直观的评价方法：方格表与等值线图 .....	351
六、	非线性模型参数估计初值的选取 .....	353
第五节	模型配合的适合性与参数估计的误差 .....	354
一、	模型配合效果的评价 .....	354
二、	参数估计的误差 .....	355
第六节	应用于生物、农业中的非线性模型 .....	356
一、	生长模型 .....	356
二、	其它模型 .....	359
第七节	软件与程序 .....	359
一、	SAS的NLIN过程 .....	360
二、	Fortran子程序库SSL II的子程序NOLF1和NOLG1 .....	361
三、	C语言函数nls1和nls2 .....	363
习题	.....	365
<b>第二十一章</b>	<b>协方差分析 .....</b>	<b>368</b>
第一节	协方差分析的用途 .....	368
第二节	协方差的模型与假定 .....	370

第三节	单因素的协方差分析 .....	371
第四节	两因素的协方差分析 .....	380
第五节	两个辅助变量的协方差分析 .....	383
	对 $\bar{Y}$ 进行修正 .....	385
<b>第二十二章</b>	<b>非参数统计</b> .....	387
第一节	$\chi^2$ 检验 .....	387
一、	$\chi^2$ 分布 .....	387
二、	适合性检验 .....	389
三、	独立性检验 .....	390
四、	费歇的 $2 \times 2$ 的精确概率计算 .....	393
五、	4格表的配对检验 .....	395
六、	自由度大于一的列联表 .....	397
第二节	分布的假设检验 .....	399
一、	卜瓦松分布的好适度检验 .....	399
二、	二项分布的好适度检验 .....	400
三、	正态分布的好适度检验 .....	401
第三节	符号和符号的等级检验法 .....	401
一、	符号检验 .....	401
二、	成对资料的等级检验法 .....	402
三、	两个组群资料的符号等级检验法 .....	403
四、	多个独立样本的等级检验 .....	405
五、	随机化区组设计的等级检验 .....	406
第四节	非参数的相关分析 .....	408
一、	列联系数 $C$ .....	408
二、	Spearman秩相关系数 $r_s$ .....	409
三、	Kendall秩相关系数 $\tau$ .....	412
附一	习题集 .....	420
附二	数表 .....	449
	参考文献 .....	510

# 第一章 概 论

## 第一节 生物统计学发展简史

19世纪末叶，生物学家们为了解决达尔文进化论中的复杂问题，经常需要借助统计学手段，而在运用这个手段过程中，原有统计方法的不足与局限性逐步地暴露出来。后来，许多学者在改善统计手段方面做了许多工作。

初创生物统计学的是高尔登(Francis Galton, 1822~1911)。他在极其广泛地搜集资料的同时，为了能使他的遗传理论建立在比较精确的基础上，出色地引入中位数 (Median)，百分位数 (Percentile)，四分位数 (Quartile)，四分位差 (Quviation) 以及分布，相关，回归等重要的统计学概念和方法。从统计学角度来看《自然遗传》(Natural Inheritance) 是高尔登最重要的著作。是作者首先提出“生物统计学” (Biometry) 一词，指出：“所谓生物统计学，是应用于生物学科中的现代统计方法。”

对生物统计学倾注心血，并把它上升到通用方法论的高度，是高尔登的得意门生皮尔逊 (Karl Pearson, 1857~1936)。

皮尔逊花了近50年的时间和精力，才将概率论和统计学熔为一炉。他的贡献是多方面的，主要有：一、变异数据的处理；二、分布曲线的分配；三、卡方检验的提出；四、回归相关的发展。皮尔逊在“对进化论的数理贡献”等论文中，对回归相关的理论和应用贡献的进展，使这个始于生物统计领域的概念被推论到一般统计的方法论上。以皮尔逊为主将的生物统计学派在研究进化、遗传等生物现象时所发展的统计方法，也适用于社会、经

济和人口等方面的统计研究。

与研究达尔文理论并行的还有生前默默无闻的孟德尔(Gregor Johann Mendel, 1822~1884), 人们对他的评价是“爱好统计学和园艺学, 这两者的结合使他在遗传学作出了卓越的贡献。”

与皮尔逊齐名的是英国遗传学的所谓剑桥学派的首要人物贝特森(William Bateson), 他是遗传实验学派。他与皮尔逊关于进化论的争论, 不但推进了遗传学的发展, 也发展了生物统计学。

接下来是哥塞特(William Seely Gosset, 1876~1937)的“小样本理论”, 人们认为哥塞特的研究成果的战略意义远比其战术意义要大, 它打开了人们的思路, 启发了后人发展出许多统计方法。由于哥塞特开创的“t分布”理论使统计学开始由大样本向小样本, 由描述向推断发展。R·费歇称他为“统计学中的法拉第”。

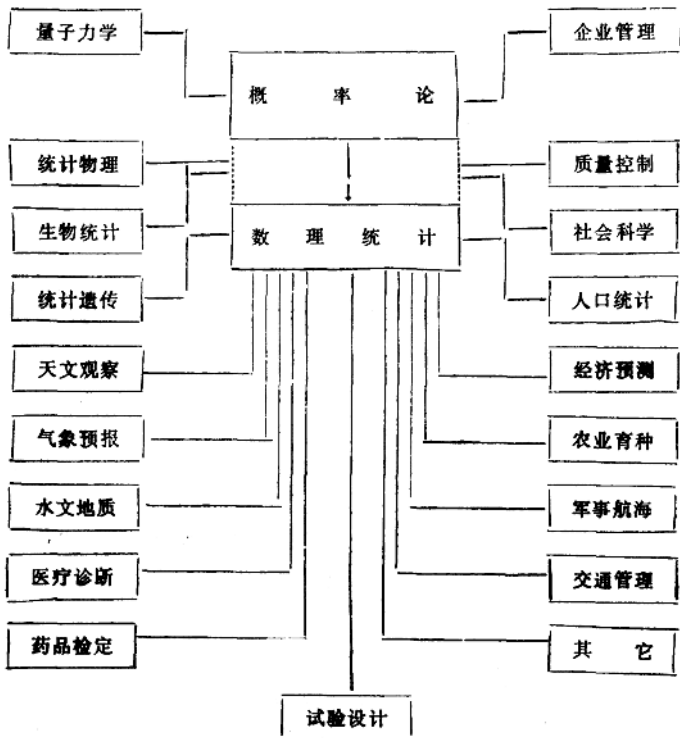
最后应提到的是统计发展史上地位显赫的R·费歇(Ronald Aylmer Fisher, 1890~1962)。他非常强调统计学是一门“通用方法论”, 并且指出:“资料的简化(Reduction of data)就是统计方法的目标”, 他的“假设无限总体”(Hypothetical infinite population)概念的提出有利于达到推断总体。他在“抽样理论”, “方差分析”, “试验设计”及“遗传育种理论”诸方面都有杰出的贡献。

(此节是根据陈善林、张浙先生的《统计发展史》一书编写的)。

## 第二节 生物统计学的地位

数理统计学(Mathematical Statistics)是数学的一个分支学科。它研究的是怎样有效地收集、整理和分析带有随机性的数据,

以对所考察的问题作出推断或预测，直至为采取一定的决策和行动提出依据和建议。数理统计学可用于种种专门领域(物理、化学、工程、生物、经济和社会等)，它应用于生物领域，就是我们 要介绍的生物统计学，下面用表格给出生物统计学、统计遗传和试验设计所处的地位。



此表摘自韩冠堂编著的《现代生活中的统计学和概率论》。

### 第三节 统计学的几个基本概念

#### 一、试验与事件 (Experiment and event)

试验是一个综合条件的实现，假定这种综合条件可以任意地重复实现很多次。

投掷一枚分币，看是“正面朝上”，还是“正面朝下”是一次试验。研究几种不同的饲料配方，看一看哪一种饲料配方使育肥猪长得更快也是一次试验。

事件是试验的结果和现象。

事件按其结果可以分三大类：在一定试验条件下必然发生的事件为必然事件；在一定试验条件下不可能发生的事件为不可能事件。

例如有10个同类产品，其中有8个是正品，2个是次品。从中任意抽取三个。其中至少有一个是正品，则是必然事件。三个都是次品的是不可能事件。

在一定的试验条件下，现象 $A$ 可能发生也可能不发生的事件，称为随机事件。

比如投掷一枚分币，这种试验是在一定条件下作的：我们规定：“分币是匀称的，放在手心上，用一定动作垂直上抛，让分币自由的落在有弹性的桌面上，等等。”“正面朝上”这件事，可能发生，也可能不发生，我们就说事件 $A$ （“正面朝上”）是一个随机事件。

#### 二、频率与概率 (Frequency and probability)

上面投掷分币例子的条件我们称为条件组 $S$ ，随机事件 $A$ 在 $n$ 次试验中发生了 $m$ 次，这时又称发生的次数 $m$ 为频数，我们定义：

$$A \text{ 发生的频率} = \frac{\text{频数}}{\text{试验次数}} = \frac{m}{n}$$



在人们的心目中，由于长期的经验积累所得，认为在大量投掷分币之后，正面向上的频率应接近 $1/2$ 。而投掷的次数越多，越接近 $1/2$ 。

于是我们下简单的定义：“频率的稳定值叫做该随机事件的概率。”

### 三、总体与个体 (Population and individual)

总体，又称母体。一个统计问题所研究对象的全体。

总体更准确一点是指“统计总体”，R·费歇说：“如果我们已有关于一万名新兵身长的资料，统计研究的对象不是新兵的整体，而是各种身长尺寸的总体”。所以我们对构成总体各因素的某些标志感兴趣，而不是各因素的本身。

组成总体的每一个元素（单元）称为个体。总体的性质取决于个体的性质。一般构成总体的所有个体数目用 $N$ 表示，称为总体容量。

若构成总体中的个体为有限个，则称为有限总体。若构成总体的个体有无限多个，则称为无限总体。

R·费歇提出一个“假设无限总体”的概念，他认为“建立假设无限总体就可以达到使问题简化的目的。所谓假设的无限总体，即现有的资料是它的随机样本。”

一个饲养试验，虽然只是20头猪吃了某种饲料，但我们假设它是从一个吃这种饲料的无限总体抽取的一个样本。

### 四、抽样与样本 (Sampling and sample)

在研究总体特征时，因为总体的无限性或测量方法具有破坏性，不允许对总体中的每一个个体一一进行度量，而采用随机抽样的方法，从总体中抽取一定数量的个体进行研究。这部分个体叫做样本。

随机抽样是指总体中的每个个体都有同等的机会进入样本。