

猪育种统计方法

华北农大畜牧系编

华北地区猪育种科研协作组印

一九七五年九月

毛主席语录

思想上政治上的路线正确与否是决定一切的。

必须在以农业为基础、工业为主导的发展国民经济总方针的指导下，逐步实现工业、农业、科学技术和国防的现代化。

我们不能走世界各国技术发展的老路，跟在别人后面一步一步地爬行。我们必须打破常规，尽量采用先进技术，在一个不太长的历史时期内，把我国建设成为一个社会主义的现代化强国。

对情况和问题一定要注意到它们的数量方面，要有基本的数量分析。

不注意懂得事物的数量方面不懂得注意基本的统计、主要的百分比，不懂得注意决定事物质量的数量界限，一切都是胸中无“数”，结果就不能不犯错误。

以养猪为中心，全面发展畜牧业。

改良畜种。

养猪是关系肥料、肉食和出口换取外汇的大问题，一切合作社都要将养猪一事放

在自己的计划内,当然省、专、县、区都应有自己的计划。

看来,养猪业必须有一个大发展。

目 录

绪 论	1
-----	---

§ 1、猪改良育种方法的发展趋向	1
------------------	---

§ 2、统计在猪育种工作中的意义	2
------------------	---

第一章 数据资料的整理和计算	4
----------------	---

§ 1、基本概念	4
----------	---

§ 2、频率分布	5
----------	---

§ 3、 $\bar{X}$ 和S的计算	13
----------------------	----

§ 4、变异系数	20
----------	----

§ 5、平均数标准差（标准误）	22
-----------------	----

§ 6、小结	23
--------	----

第二章 统计测验	24
----------	----

§ 1、统计测验概说	24
------------	----

§ 2、U测验	26
---------	----

§ 3、t测验	30
---------	----

§ 4、 χ^2 测验	36
------------------	----

§ 5、F测验和变异分析	40
--------------	----

§6、小结	51
-------	----

第三章 回归相关 52

§1、回归相关的意义	52
------------	----

§2、直线回归相关	54
-----------	----

§3、曲线回归相关	64
-----------	----

§4、二元回归相关	67
-----------	----

§5、小结	73
-------	----

第四章 数量性状遗传的基础知识 74

§1、性状的遗传和变异	74
-------------	----

§2、数量性状的构成和剖析	76
---------------	----

§3、通径系数和通径分析	78
--------------	----

第五章 重复力 80

§1、重复力的意义	80
-----------	----

§2、重复力的计算	82
-----------	----

§3、重复力的性质	88
-----------	----

§4、重复力的应用	89
-----------	----

第六章 遗传力 91

§1、遗传力的意义	91
-----------	----

§2、遗传力的计算	92
§3、遗传力的取值范围	101
§4、不同方法估计遗传力的精确性的比较	101
§5、影响估计遗传力正确性的因素	102
§6、遗传力的应用	104
§7、猪主要经济性状的遗传力	105

第七章 遗传相关

107

§1、遗传相关的意义	107
§2、遗传相关的计算	108
§3、遗传相关的应用	114
§4、猪主要经济性状间的遗传相关	116

第八章 育种值和选种

117

§1、育种值在选种中的意义	117
§2、个体成绩测验	117
§3、祖先谱系测验	120
§4、旁系同胞测验	127
§5、子女后裔测验	136
§6、各种测验的相对效率	140

第九章 亲缘关系和选配

146

§1、选配的意义	146
§2、亲缘系数(血缘相关系数)	147
§3、近交系数	150
§4、近亲繁殖的应用	156
§5、杂种优势的估计方法	157
§6、杂交繁殖的应用	159

第十章 试验设计

§1、试验设计的目的和原则	160
§2、试验设计的方法	161
§3、抽样调查	181
§4、试验单位(猪)数目的确定	183

緒 論

§1. 猪改良育种方法的发展趋向

为了发展我国社会主义农业，实现“养猪业必须有一个大发展”的要求，培育适应我国各地条件的优良猪种是一个重要方面。目前猪的改良育种方法，已从过去的以个别优秀个体为中心的品系繁育，转向以具有优良性状的基础猪群的群体系统选育。进行群体系统选育，要在开始前调查基础猪群的性能，根据调查结果确定重点改良选育项目。在群体系统选育过程中仍要重视个体选择，特别更要搞好个体选择记录，常年积累数据资料，依据被选择个体的本身记录（如日增重、背膘厚度等）和同胞测验记录（如饲料要求率、眼肌面积、后腿比例等）评定成绩，选出优良个体组成下一代猪群来代替原有的猪群。运用这种改良育种方法，就可以逐代提高猪的素质，能够加快猪的改良育种速度。但是这样的群体系统选育比过去的个体品系繁育还要引入更多“数”的概念，这也要求畜育种科学走向现代化的一种发展趋势。我们从事猪改良育种工作的同志，有必要进一步学习和掌握有关育种统计的方法及其应用。

2.2. 统计在猪育种工作中的意义

统计是处理大量偶然现象中的规律性的方法，能使我们从浩繁的事实中抽取意义深远的结论，使估计上的误差和人力物力上浪费减少至最低限度。我们在调查研究进行试验特别是在用现代育种方法以繁殖的育种工作时，必须妥善记载有关性状的的大量观测数据。如何正确地收集、整理和分析这些观测数据，从大量数据记录中总结经验，找出规律，预测发展趋势和指导改进育种工作都离不开统计方法。所以育种统计是育种工作中的一个重要部分，在育种工作中具有重要意义。

“对情况和问题一定要注意到它们的数量方面，要有基本的数量分析”，这是搞好育种工作的关键。因为在以提高猪的经济性状（如日增重、饲料要求率等）为主要目的育种工作中，不能离开它们的数量方面和数量分析。就是判断猪的经济性状的优劣好坏也是以生产过程中所需要的时间、成本和产量等的数量多少为标准的，这就反映了“任何质都表现为一定的数量，没有数量也就没有质”的辩证关系。所以在育种工作中忽视统计是错误的。

由于应用有关公式，计算数据的育种统计，仅是育种工作中的一个重要组成部分，只搞公式数据不能代替全部的育种工

作。所以只搞数据和公式的育种工作也是片面的。只有在实际育种工作中，正确运用统计方法，使统计和育种实践结合起来，才能多快好省地使错误的育种工作。

第一章 数据资料的整理和计算

§ 1. 基本概念

育种实践中得到的观测数据提供了很有用的信息（情报），可以帮助我们发现存在的问题，认识事物的内在规律，是采取措施改进育种工作的依据。可是蕴藏在大量数据之中的信息并非一目了然，只有对数据作科学的整理和分析，才能去粗取精，去伪存真，提取出正确的信息来。

整理数据之前，首先要了解数据的代表性质，判断一批数据是总体的度量观测值，还是样本的度量观测值。所谓总体（或称集团）就是我们所研究对象的全体，其中的一个单位则叫做个体。譬如，我们研究母猪的繁殖力（产仔数），总体就是产仔母猪的全体，即所有的产仔母猪，而其中的一头产仔母猪则是一个个体。所谓样本（或称样品）就是研究对象的总体的一部分，样本中所含个体的数目叫做样本的大小（或称容量）。譬如，每次抽取十头、百头或千头产仔母猪进行观测，不论每次头数多少都是产仔母猪的一部分，都是样本。我们将总体或样本中每一个个体的性状的观测数据叫做变数，通常用 X 代表变数。一般将变数在30个以下的样本称为小样本，变数在30个以上的样本称为大样本。区分了总体和样本后，用总体测得的数据所计算出来的统计量叫做常轨数，用样本测得的数据所计算出来的统计量叫做估计常轨数。常轨数是总体的真正数值，是个常数。估计常轨数由于样本大小的不同和取样时个体的变动而有所变化，是个变数。由于无限总体所含个体的数目可以很多，甚至无穷，以致我们不可能一一加以考察；有时尽管有

限总体所含个体的数目不是很多，因条件限制也不允许我们加以全部考察。这样就不易得到或无法得到总体的常轨数，更多地只能是通过样本来了解总体，用样本性状的估计常轨数来估计代表总体性状的常轨数。如何用样本来研究总体，通过样本来了解总体，这就是统计所要解决的问题之一。

整理数据时，还要了解数据的特性。我们都知道在同样饲养管理条件下，同品种同胎次的产仔母猪的产仔数并不完全一样，表现出一定的差异（波动）。这种差异不是杂乱无章的，而是表现出一定的规律。一般产仔母猪的产仔数多在8至12头之间，不论抽取多少样本对母猪产仔数进行观测，也多在8至12头之间。经过多次观测，发现各次观测的产仔数变异的情况是十分相似的，而且产仔数落在任一范围的数目在多次数据中的比例是比较稳定的，这种稳定的比例叫做“概率”。这就是说产仔数散布的情况是有规律性的。观测数据既有差异又有规律的特性，正是统计所要研究的问题。我们从有差异的数据中，找出其规律性，也是统计所要解决的问题之一。

§ 2. 频率分布

频率分布根据使用需要，有时也叫频数分布，有时也叫次数分布。

2.1. 小样本（不分组）资料的整理。

表1——1 是10头仔猪的初生重（公斤）。

表1——1

1.5	1.2	1.3	1.4	0.8	0.9	1.0	1.1	1.6	1.2
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

整理方法是先将原始数据按数值的大小，由小到大或由大到小依次排成顺序表，如表1——2。

表 1—2

0.8	0.9	1.0	1.1	1.2	1.2	1.3	1.4	1.5	1.6
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

作出依次顺序表后，可以一目了然地看出数据资料的集中和变异情况。对于任何一组观测数据，排列在当中的数值叫做中数（或称中位数）。在偶数个观测数据的情况下，我们取当中两个数值的平均数做为中数。一组数据的下限叫最小值，上限叫最大值。最大值与最小值之间叫做全距（或称极差）。表 1—2 的中数 $= 1.2 + 1.2 / 2 = 1.2$ ，最大值 $= 1.6$ ，最小值 $= 0.8$ ，全距 $= 1.6 - 0.8 = 0.8$ 。

2.2 大样本（分组）资料的整理

当变数增多时，编制依次顺序表既费时间又不能使我们从庞大的数据行列中看出明显的规律性来，要搞清数据变异更完整的规律，有必要将庞大的依次顺序表压缩成简单的频率分布表（或称次数分配表）。

表 1—3 是 200 头母猪 30 天泌乳力（公斤）的原始数据。这叫度量数据资料，因为变数是连续性的。我们指性状的男是用长度、重量或容积来表示的，取值不一定为整数，在两个整数之间有无限个数，即在变异范围内可以取得任何数值的变数称为连续性变数。

对表 1—3 的整理步骤：

1. 求出资料的全距。资料的最大值为 118.5，最小值为 11.6，全距 $= 118.5 - 11.6 = 106.9$ 。

2. 确定分组的组数。在样本的个体比较多时，通常分成 10—20 个组；样本的个体少于 50 时，可分成 5—6 个组。对表 1—3 的数据可以分成 10—15 个组，经考虑为计算方便合理起见，可分成 11 个组。

3. 决定组距。一般用字母 h 代表组距。组距等于全距除以

表 1—3

47.2	67.67	78.5	90.5	53.9	31.5	60.7	84.0	62.2	62.2
57.1	51.6	74.5	97.8	58.2	78.0	64.6	84.8	50.5	68.7
40.5	56.8	58.0	41.5	85.8	58.8	51.8	60.5	51.8	63.7
78.2	29.3	23.5	67.8	62.5	96.5	84.0	36.0	74.5	79.6
58.2	41.5	63.8	67.3	80.7	11.6	64.0	34.1	28.0	65.0
75.9	59.8	79.5	57.2	76.5	34.5	49.6	67.5	18.4	53.0
61.2	63.3	71.7	65.8	103.5	12.0	58.5	69.8	40.2	55.0
85.0	66.5	39.8	75.5	55.5	104.7	85.0	55.2	104.2	57.3
60.0	96.7	20.5	15.7	20.0	26.3	79.3	77.4	102.4	95.1
63.0	56.0	70.3	76.8	74.0	118.5	42.4	26.6	86.8	53.0
43.8	83.0	71.8	36.0	67.8	72.5	96.5	101.0	41.7	76.1
69.0	65.0	100.0	31.4	71.8	94.5	70.1	89.7	53.4	27.9
81.5	36.7	88.6	79.4	69.5	79.5	72.1	65.5	23.2	40.2
49.9	81.0	41.5	76.9	72.0	82.1	64.4	69.8	54.1	68.6
40.5	86.7	59.5	25.0	62.8	70.0	50.6	67.1	56.7	73.8
62.3	72.3	113.0	90.3	57.8	86.1	34.0	39.4	49.7	41.0
50.4	67.5	79.5	68.3	38.3	63.6	34.0	14.6	25.5	54.3
94.3	88.4	85.0	60.8	81.0	52.5	117.0	78.0	23.3	71.7
85.1	78.0	81.0	74.5	72.9	64.0	80.5	108.6	72.5	87.5
56.0	90.8	89.4	70.6	55.5	102.2	93.5	94.4	67.3	47.3

组数。对表 1—3 的组距， $i = 106.9/11 = 9.7 \approx 10$ 。

4. 确定分点。表 1—3 的资料既然可分为 10—20，20—30，30—40……11 个组，那么泌乳力恰好是 20、30、40……的样本应分到上下哪一组呢？为了避免这种麻烦，通常要使分点比尾精确度提高一位，即分成 10—19.9，20—29.9，30—39.9……。

5. 求出各組中值（或稱組中介）。組中值等於組上限加組下限的二分之一。組中值 = 組上限 + 組下限 / 2。

6. 做出頻數。將全部變數按數值大小劃歸到相應各組。

經過上述步驟，編制出表1—3所擬的頻率分布表，形式如表1—4。

表1—4

分 組	組中值(X)	頻 數 計 標	頻數(f)	相對頻數
10~19.9	15	正	5	0.025
20~29.9	25	正正T	11	0.055
30~39.9	35	正正T	12	0.060
40~49.9	45	正正正一	16	0.080
50~59.9	55	正正正正正正	30	0.150
60~69.9	65	正正正正正正正T	39	0.195
70~79.9	75	正正正正正正正一	36	0.180
80~89.9	85	正正正正正	25	0.125
90~99.9	95	正正T	13	0.065
100~109.9	105	正正	9	0.045
110~119.9	115	T	3	0.015
Σ			200	1.000

一般習慣用X代表變數，在分組資料里也可用X代表組中值，此時組中值相當於變數。f在分組資料里是代表頻數（或稱次數）的符號。Σ讀Sigma，是代表同類數求和的符號。

表1—5是70頭因豬的產仔數（頭）的原始數據。這是計數資料，變數是非連續性的。我們指性狀的變只能用整數而不能用小數表示，即在變異範圍內只能數個數的變數，稱為非連續性變數。例如每胎仔豬數只能數出8、9、10……，不能數出8.5頭

仔猪这类的数据。

表1—5

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

对表1—5的整理步骤：

1. 求出资料的全距。70头母猪产仔猪的变异范围是7—14头。
2. 变数按每窝所产仔猪头数分组，求出频数，编制出频数分布表。表1—6就是10头母猪产仔猪的频数分布表。

表1—6

每窝仔猪数(X)	频数计称(母猪头数)	频数(f)	相对频数
7	T	2	0.029
8	正	5	0.071
9	正正	9	0.128
10	正正正正	20	0.286
11	正正正T	17	0.243
12	正下	8	0.114
13	正T	7	0.100
14	T	2	0.029
Σ		70	1.000

2.3. 用图形表示数据资料

频数分布情况除用频率分布表表示外，还可绘成图表示，图形可使研究的事物或现象更加一目了然。