

《科技写作》教学参考资料

科学论文的整理和撰写方法

(节 译)

富田军二 著

小泉贞明 修订

石馆基

肇永和 译

哈尔滨科学技术大学高教研究会

一九八四年五月

科学论文的整理和撰写方法(节译)

目 录

第一篇 绪 论

1. 科学论文的种类和内容..... 1
2. 科学论文的基本要求..... 2

第二篇 撰写论文前的准备工作

- (A) 有关整理研究结果的一般注意 事项 3
3. 整理研究结果的第一步..... 3
4. 假 说..... 4
5. 各种事件间的关系、秩序..... 5
- (B) 数据资料的处理和表现..... 6
6. 数据的数学处理..... 6
7. 把数据编列成表..... 7
8. 画成图表..... 7
9. 频数分布表、频数分布图..... 8
10. 代表值..... 9
11. 加权平均..... 9
12. 误差、散射度、置信界限..... 9
13. 测定误差及其种类..... 10
14. 以生物为材料的实验或统计中的变差的特点..... 10
15. 相关系数..... 11
- (C) 在数据处理中容易出现的错误..... 11
16. 一般注意事项..... 11
17. 比较事件(群)间的数值的场合..... 12
18. 阐述相关关系的场合..... 14

第三篇 论文的组成

19. 内容的划分方法和排列顺序..... 15
20. 论文的标题..... 17
21. 作者名..... 20
22. 单位名..... 21
23. 同论文第 1 页有关的其它事项..... 23
24. 目 录..... 25
25. 绪 论..... 27
26. 谢 词..... 27

27. 材料及方法.....	29
28. 成 果.....	30
29. 讨 论.....	30
30. 结 论.....	31
31. 提 要.....	32
32. 注 脚.....	33
33. 追加、附录.....	34

第六篇 图 表

(A) 有关图表的一般注 意 事 项.....	35
72. 图表的重要性.....	35
73. 有关图表的一般注意事项.....	35
74. 图表的序号和说明.....	35
75. 在文章中引用图、表、图片时的注意事项.....	36
76. 正文中图表的位置.....	37
(B) 表.....	37
77. 表的改进.....	37
78. 表的印刷形式.....	43
(C) 图画、照片.....	44
79. 凸版和照相铜版.....	44
80. 照 片.....	44
81. 绘图时的注意事项.....	45
82. 图中的文字、数字.....	47
83. 图的放大率、缩小率.....	48
(D) 曲线的画 法.....	55
84. 坐标轴的选择.....	55
85. 方格纸的选择.....	56
86. 刻度的选择.....	56
87. 曲线的绘制方法及其平滑化.....	57
88. 制图上的各种注意事项.....	58

第八篇 原稿的加工和校正

101. 原 稿 的 誊 写.....	60
102. 原稿的检查、订正、完善.....	60
103. 校 样.....	62
104. 校对符号.....	63
译后记.....	66

第1篇 绪 论

1. 科学论文的种类和内容 自然科学的作者发表的科研成果(广义的论文),由于专业领域不同,种类或名称甚多。大体上可以分为下列4种:(1) 科学论文,(2) 资料 (3) 综述 (4) 单行本。

(1) 科学论文:科学论文在学术成果中最为重要,而且同学者、工程技术人员的关系极为密切。一篇科学论文应该包括哪些内容,因专业领域或研究部类而异,不是简单用几句话就能说清楚的。具体说来,至少要具有下列项目中的某一项内容。

a) 属于某一特定领域的新见解或创见。它可以只是实验结果或新的研究方法,也可以是对自己或别人的实验成果提出的新的解释。

b) 对于在逻辑上认为根据还不够充分,但已经为学者们所接受或承认了的学说或推论进一步进行试验所得的结果。

c) 对于意见或看法不一致的推论或假说进一步进行确定性研究所取得的成果。

d) 对于过去为人们所忽视或没有引起人们注意的一些问题,经过充分研究所取得的成果。

e) 问题本身即或不大,但它有可能进一步发展,甚至有希望引出其它有意义的重要突破的内容。

科学论文的正文一般包括绪论、原始资料、方法、结果、评论、结论、摘要等部分,依次加以论述。但从对各种杂志所进行的调查研究的结果可知,论文的目的、种类、长短不同,以及杂志的经营方针不同,论文所包括的内容也各不相同。

(2) 资料:资料包括调查报告、观测成果、视察报告等内容。由于在这些材料中出现的资料或数据本身具有实用价值,因而一个非常重要的问题是事实的记载要准确,而且还必须取材范围广,亲眼目睹的资料和传闻的资料要严格区分开来,不能以个人的私见或爱憎作为取舍材料的标准或根据。根据不同的使用目的,资料的绪论或评论部分可以简化,或者全部省略。

(3) 综述:综述是就某一特定问题,从综合的公正无私的立场出发,对已经发表过的重要成果进行批判、评论,使读者对过去的发展过程或现状一目了然。根据情况,有时主要只以最近10年或近1年来发表的成果为主进行综合论述。

(4) 单行本:单行本多半是以上述3者,即科学论文、资料、综述的某一种或者两种,或者全部为内容,范围极其广泛。

预 报:除上述4种外还有1种,这就是预报(*preliminary note*)。预报主要是例如虽然在细节上还有必须进行研究的课题,但大致的方向已经明确了场合,或者实验实例较少,但由于材料的时间性或经费等原因,要想充实实验实例还需要时间的场合,为了取得优先权,暂且做一简要报道之类的文章。

必须注意,预报同把科学论文分成若干部分进行发表时的第1次报道,在意义上有着本质的不同。就是说,科学论文的第1次报道、第2次报道、……,是把范围广泛的研究课题中的某一部分或某一系统经过整理分别发表的;相反,预报则是要涉及到科研课题的全部,对其结果或结论作简要的报道。因此,预报是以报道新的见地为主要着眼点,哪些冗长的绪论、评论、谢词等,则不加叙述。

在写完预报之后,由于就某些细节做了进一步的研究,或者进一步积累了大量的实验例,结果在后来发表正式论文时,有时出现对预报的某些内容必须加以订正的情况,这是不得已的。但是,如果出于为了过分贪图功名,以致必须对预报中的重要事项进行订正的做法,不能被认为是体面的。

2、科学论文的基本要求 科学论文不只是研究人员本身的备忘录,而是要广泛地向同行们转达自己所取得的研究成果或意见,除供参考外,还有征求意见的目的。因此,精心注意论文的表述是很重要的。每个具体问题将在后面详细叙述,下面只列举一些基本要求。

(1) 态度: 在撰写科学论文时,最根本、最重要的问题是忠于真理,在学术问题上要老老实实。因此,在任何情况下,都必须采取冷静的、公正无私的、不为先入观念所左右的态度。不得为了顺乎自己的主观愿望而歪曲事实或妄加议论;不能从大量具有同等可能性的事项中只挑选那些于己方便的内容;不能隐讳于己不利的内容;不能对别人的学说使用为攻击而攻击之类的不恰当的过激言词;不能为了迁就老师或前辈,对那些应坚持己见的问题也不加坚持。

(2) 记述: 论文中的记述必须绝对准确。从数字、数式到正文中的所有记述,乃至引用文献的记述都不得有丝毫错误。

(3) 逻辑: 在论文中,无论是论点的阐述、评论的展开、下结论、用词和文章的表现方式等,都必须在逻辑上没有矛盾或缺陷。

一般最容易犯的逻辑上的错误,大致有下列各点。

- ① 用词、造句表现方式上的错误
- ② 假设在立意方法、使用方法方面的错误
- ③ 评论在态度、开展方法方面的错误(例如:先入观念,事实和意见的混同,脱离目标的评论;以假定为前提的无休止的评论;不恰当的比较或类似的例子等)
- ④ 在下结论方面的错误(例如,结论过大或过小;因果关系方面的错误判断等)

第2篇 撰写论文前的准备工作

(A) 有关整理研究结果的一般注意事项

3. 整理研究结果的第一步 (1) 利用表或图表表示研究成果：在归纳整理研究结果的最初阶段，就是利用表或图表来表示数据。这是明确数据之间的关系、解释和说明研究结果的最佳手段。正因如此，不管这些表或图表是否能载入论文之中，也还是以这样做为好。当然，必须注意表或图表的表示方法是有许多种的。

(2) 研究结果的解释、说明：认真研究图或图表以及经过整理了的笔记，找出数据之间的内在联系，对结果进行解释或说明，然后引出结论。在这种情况下，必须以冷静、公正无私的态度，对在逻辑上存在可能性的所有场合给予同等的考虑，不得漫不经心地只强调其中的某一方面，而要对其它方面的事实或证据等，也要客观地加以综合考虑，从中找出可能性更大的问题。此时，哪怕是在最初阶段浮现出来的可能性，也都要毫无遗漏地当场记载下来。有人也可能想要在事后经过整理再写。但是采取这种做法，往往会出现遗忘或漏记的现象。特别是那些第1次想到的问题，思路最清晰，对它的印象或注意力也特别深，所以必须及时记载下来。

(3) 准确程度的研究：在做如上所述结论时，在任何时刻都不得忘记注意研究数据的准确程度或可靠限度。否则，有时会导致所得结论失真，甚至错误。

(4) 追加实验：如果时间充裕，为了进一步证实按上述方法引出的结论，可以进行追加实验。此时，如果可能，最好努力从性质完全不同的另外的角度再取得证据。

(5) 结论的修改：如果通过追加实验以及后来的考察，认识到原来所下的结论有必须订正的部分时，要加以妥善处理。此时，既不能草率地变更或撤销原来结论，也不能过分地拘泥于旧说。

(6) 例外的事实：对于同自己所提出的结论不一致的例外事实或特殊事例要更加予以注意。那种对于推论自己所提出的结论不利就予以忽视，或者不作准确记录之类的做法，不是科学的态度。不仅要经常想起许多重要的发现往往产生于乍一看来似乎属于例外或者特例之中的这一事实，而且还要认真研究为什么出现了这样的例外，并试图加以说明；还要考虑由于存在着这些例外，有没有改变原来所下结论的必要，进而能否引出包括这些例外在内而且有可能作出解释的结论。

(7) 所有的内容都要记录下来：前面曾经提过，凡是在自己的脑海中出现过的想法都要立即全部记录下来。常常有这样的情况，由于使用的记录用纸大小不一，纸张的质量也不好，以致对于那些好容易记录下来的东西，却采取漫不经心的态度，随便乱扔，实在是令人惋惜的。因此，记录用纸最好裁成一定的大小（例如裁成标准规格的图书卡片或者2倍于卡片大小的卡片），放在实验室的一些地方或自己的家里（同时准备削好了尖的铅笔），也可以放在文件袋里，以便随时可以用来记录在自己脑海中出现的想法。另

外,在1张卡片中只能记载1项内容,卡片的上方要写上适当的标题和编排序号,并有序地放入卡片箱内。也可以使用相当于中学笔记本大小的带格的纸制成的代用卡片(把它夹在纸夹里,或者装在信封里备用。但这种代用卡片同标准规格的卡片相比,用起来不方便,也不那么经济)。

另一个比较好的方法是,当同一或同类项目的记载达到一定数量后要进一步加工整理,经过挑选决定取舍。必要时还可经过进一步补充,形成更加完善的资料,再用原稿纸或打字腊纸誊写清楚,最后,逐渐达到论文原稿水平。

4. 假说 为了整理、归纳自己所得的实验结果,并且对它进行解释,假说起着很大的作用。常常有这样的情况,那些看起来好象彼此互无联系或没有任何意义的未经整理的原始资料,只有建立了适当的假说,通过整理分类,然后才能弄清它们之间的意义。假说还可用来作为进行新的实验、观察的出发点。如果最初进行的实验、观察的结果达到预期目的,为了进一步提高对自己的假说是正确的可靠程度,一般要进一步通过实验取得充分的证据。就是说,假说可以启发人们如何进一步进行实验、观察,并以此为出发点,构成研究向更加深广的领域扩展的基础。

假说在任何时候都是一种推测。有时由于后来发现了反证这一假说的新事实而不得不予以放弃。但是,如果说明这一假说的新的事实不断出现,而且确认这一假说在任何情况下都能适用时,那么,该假说可以升格进入学说(*theory*)的部类,再提高一步,就进入法则(*law*)的部类。

如上所述,假说乃是研究人员的最重要的判断手段之一。万一假说的建立方法或应用方法不当,反而会妨碍研究的正常开展。因此,必须注意下列各点。

(1) 假说的建立方法:为了建立有用的好的假说,需要有审慎的思想准备、丰富的经验和预备知识,而不要轻率地把在脑海中偶然出现的想法原封不动地拿来做为假说。这是因为自己的某些想法不可能只是一时的假设,如果一旦定为假说,就很难再用别的什么东西去代替它。所以,应从一开始就要在充分酝酿的基础上建立假说。从阅读一些前辈的传记中可知,的确也有某些以在脑海中偶然出现的想法为基础提出某种假想,并据以取得重大研究成果的实例。但总地说来,这种例子毕竟是极少数。大多数的重大发现,还是必须以经过周密的考虑建立起来的假说为基础而取得的。因此,绝对不是漫不经心的毫无根据的情况下就能建立假说的。

(2) 假说的内容:假说必须具备有能够解释过去不能解释的事实或关系的内容,或者有助于预见新的事实或关系的内容。假说本身不得有矛盾,而且对既定的事实或原理不得有忽视之处。另外,所提假说如果不能进行试验或证明,将是没有什么大用处的。

(3) 假说的选择:在有可能建立几种假说的情况下,不得毫无理由地只强调其中某一个假说。当不是用复杂的假说而是用简单的假说也同样能够说明问题时,那就要选用简单的假说,而不选用复杂的假说。

(4) 同假说相比更要尊重事实:人们在建立了一个假说之后,往往容易进行“带主观愿望的观测(*wishful thinking*)”,并由此而使观察、解释、判断受到影响。为了不致陷入这种弊端,要从平时就养成尊重事实的思想方法和习惯。在任何情况下都要记住:假说毕竟不过是一种推测,无论什么时候,都要以客观证据为主,以假说为从。

(5) 假说的变更和坚持：如果已经发现假说同事实不一致，就应该修订或者全部废弃。人们往往容易坚持最初建立起来的假说，对于那些于假说不利的事实，通过不加深究或者托词支吾的解释等手段力图维护自己的假说，或者对于自己的假说的反证采取转移视线做法。然而，这样做，假说反而会对研究起负作用。不过必须指出，这是指的已经出现了明确的反证，以致原假说无法继续维护下去的场合而言。因而只要没有直接了当的反证，假说是不能轻易放弃的。这是因为，有时假说虽然完全正确，但也可能在实验或观察的方法上有缺点。这样，我们就应该做两种思想准备，一方面，在“事实”面前，随时果断地修改自己的假说；另一方面，也要坚持不能轻易放弃。此外，对于每一个判断，都必须持以不为感情所左右的公证无私的态度。

(6) 不能把别人已经提出的假说作为新的假说加以发表。但是，如果前人的假说有缺陷，经过加工修改后而作为新的假说发表，当然是可以的。

5. 各种事件间的关系、秩序 从实验、观察、观测得来的未经过加工的数据，既或积累得再多，如果原封不动，那也只不过是各自孤立的事实，用途甚微。科学的目的在于找出各种事件间的关系或秩序。这时，如能从数量上，特别是从统计学的角度，对于上述原始数据加以分析处理，那就可以很容易地发现和证明这些数据之间的关系。

(1) 非常初步的而且是基本的秩序：我们把许多物质分别称呼为水、树、铜，这些称呼本身就已经是一种秩序（当然不是因果性的秩序），探讨这种非常初步的秩序这一工作本身，不仅是在开辟科学的尚未开垦的领域的原始时期，就是在今天，也仍是一项基本的重要的工作。

(2) 因果关系：同科学关系最密切的秩序是因果性。但除了因果关系以外，还有各种关系，把这些所有的关系都作为因果性来看待，或者不找出因果性就不能认为是科学的观点，是错误的。

有关因果关系的常识性的观念和严正的观念：例如，人们一听说“张三把窗户玻璃打坏了”，一般就常识性地认为打破玻璃的原因是张三。但这种认识会立即给人们以一种含糊不清的印象。如果把打破玻璃的原因不是单纯地归结于张三本身，而是归结于张三的某种行为，就显得更加严密。如果我们对于张三的这一行为再作进一步的分析，就会注意到其中存在着物理的重要原因。又如我们说“铁在潮湿的空气中是要生锈的”。但铁生锈并不一定只限于在潮湿的空气中。相反，也有铁在潮湿空气中偏偏不生锈的情况。因而促使人想到铁生锈的原因并不是单一的，而是同许多因素有关联的。从上述两例可知，人们是从最初的比较粗糙的近似的说法逐渐发展到更加严密的思考方法，进而沿着找出经过更加充分分析了的确不移的因果关系的方向发展的。

另外，有时有这样的情况。严格说来，本来不能作为原因和结果的关系看待的问题，却硬要把它同原因和结果之间的关系扯在一起。例如，我们常常听到“力是产生加速度的原因”这样的说法。事实上，加速度的出现同力的作用经常处于同时发生、同时消失的状态之中，它们之间不是在时间上表现为先后发生的因果关系，只不过是用力概念来表述产生加速度的这一事实而已。恐怕这是把原因这个词同理由这个词在概念上混淆在一起了。

(3) 数量的关系式：各种不同因素间的关系，常常需要用数学关系式来表达。例

如, 杠杆原理、波义耳—查理定律、欧姆定律等, 就是如此。这种关系不是因果性的关系。

(4) 理论: 在相对理论或气体的运动学理论之类的理论 (*theory*) 中, 绝不是所有的相互之间有一定关系的因素都是能够直接观测得到的。而且, 人们也不是主张在因素之间适用的关系都要直接通过实验才能加以确认的。但是, 我们知道, 借助于这样的理论, 通过实验能够确认的许多数量方面或性质方面的法则, 也不是互无关系、各自孤立的。例如, 气体的温度、比热之间的数量关系, 固体的熔点、压力、容积之间的关系等, 都可以是通过物质的运动学理论的假定, 用逻辑的方法推导出来的。

(B) 数据、资料的处理和表现

6. 数据的数学处理 从实验、观察、观测得来的“原始”的数据, 有时由于某种目的也可以原封不动地加以记载和发表。但在想要不仅了解数值本身, 而且还要弄清事件间的关系或包孕于其中的原理、原则时, 一般要对原始的数据进行数学处理和综合整理。但是, 如果数学处理方法不当, 尽管原始数据的数值再准、数量再多, 也很难讲清其中存在的关系或原理。

为了从所得数据中大体估计事件之间有没有关系, 以及如果有, 它将是一个什么样关系的问题, 根据情况, 最好先求出算术平均, 或者制成表, 或者画成图表。如果用某一种方法不能找出关系时, 还要接连不断地试用其它的方法。这样的实例是很多的, 当学生或年轻研究人员向导师做科学研究的中间汇报时说“没有发现A和B之间的关系”然而导师却当场亲自作出了高度准确的相关图。由此可见, 不能由于只用一两个方法遭到失败而就灰心失望地做出结论。

另外, 在开始阶段根本无法估计某些因素之间到底存在什么样关系的时候, 应该先用简单的方法一步一步地探讨, 而不要一下子就使用过于复杂的方法。

有关数学处理的技巧问题, 由于许多国家已出版了很多好书, 因此不在这里作说明。正如在“前言”里所讲的那样, 只就同“精密计算以前, 统计学以前、推测统计学以前”的基本问题有关的注意事项, 介绍如下。

数值计算上的注意: 为了整理实验、观测等结果, 随之而来的一项工作是数值计算。为了进行数值计算, 需要注意下列各点: (1) 计算方法的选择, (2) 计算的记录, (3) 有效数字的确定, (4) 验算。

(1) 计算方法的选择: 当求某一数值时, 它的计算方法常常有许多种。另外, 在计算用的器械或表格中, 也有各种计算器、计算尺、算盘、对数表及其它特殊的表格等。用哪种计算方法, 使用哪种计算手段, 这要根据计算的种类或要求的精度来决定。如果计算方法不当, 将空徒浪费时间和劳动, 甚至产生不准确的结果。因而计算方法的选择, 应当是很慎重的。

(2) 计算的记录: 数字要清楚和不要太小, 书写方式要统一。注意0和6、1和7、5和8等数字的写法不要弄混。多位数字的定位方法要一致。数字写错时, 不要用橡皮擦, 要在必须擦的数字的上面, 清楚地划1条斜线或横线, 后然后在旁边写上订正数字, 不要把用纸擦得皱皱褶褶的, 以免原来的字迹看不清楚。这是因为常常有这样的例子,

由于一念之差，原来的数字反而是准确的。

确定计算的程序也是很重要的。要在考虑计算的性质、方法的基础上，把计算分成几个阶段，把每个阶段的计算结果恰当地一一列记在同一表上。在计算中，例如，哪怕是某数值乘以2这样的简单运算也好，都要整整齐齐地写在该数值的近旁。这样做，对于事后检查有无错误是非常方便的。

这样，由于所有各个阶段的计算都详细写出，有时需要很大篇幅的纸面。因此，必要时也可以把几张纸接在一起。但是为了缩小纸面，而用小字写得密密麻麻的，或者把辅助计算写在其它小纸片上之类的做法，乃是产生错误的根源（不得已必须用另纸计算时，一定要把另纸粘在主要计算用纸上）。例如，计算 $m = (2a + 3b^2)^2$ 的 m 的数值时，要象下表那样，中间计算也要毫不简略地明确记载下来（使用对数表时，要设计另外的样式）：

a	$2a$	b	b^2	$3b^2$	$2a + 3b^2$	$(2a + 3b^2)^2$

计算的程序、阶段的划分方法、记录的顺序等，如能充分注意按具体情况加以确定的话，那么，后面的计算就不大需要再花费多大的脑筋，可以机械地进行下去，而且效率也高。

(3) 有效数字的确定：计算数值时必须时刻确定计算出来的数字取几位才算准确的问题。不考虑有效数字，只是一味地罗列位数多的数字，是没有意义的。为了确定有效数字，必须随时牢记关系误差（或百分比误差）同有效数字之间的关系。

(4) 验算：进行数值计算时，至少要验算1次。验算的方法不是固定不变的，要根据问题的性质加以选定。例如，使用同原来不同的计算顺序或方法进行验算，或者请别人给计算，或者用计算器或对数表计算原来用计算尺计算的结果，等等。验算不能参照原来的计算，必须在取得最后的结果之前完全独立地在另外的纸上进行。

7. 把数据编列成表 作为数据分析的第1步，把数据列成表是很重要的。但是，要想制成一种非常有利于进行分析的并且是很有效率的表，需要下一番工夫加以仔细琢磨。数据以怎样排列为最好的问题，常常在不同的情况下，形式是多种多样的。首先要确定数据是横着排还是竖着排，其次是确定按着读取的顺序排，还是按大小的顺序排，或者是分成几个部分排。

表的用纸必须留有宽裕的空白（最好是方眼纸或格纸），数字不要过小，书写要清楚，并与附近的数字之间留有空隙，无论是横的方向或是竖的方向都要写直（看起来似乎不难做到，但实际上做起来却是相当难以办到的重要问题）。从这些原始数值导出其它数值时，不管是纳入正式印刷的表中，那些计算过程中的计算值也要整整齐齐地一一写在各自的部位上，以绝对不在另外的纸上进行计算而只把结果（即最后阶段）转记到表上为好。

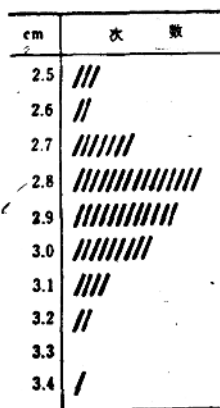
此外，有关纳入论文中的表的制法，请参阅77、78。

8. 画成图表 大多数的实验结果，如能制成图表，则有关数据之间的关系能够最清楚地表现出来。因此，不管数据是否纳入论文之中，都应当制成图表，进行探讨。

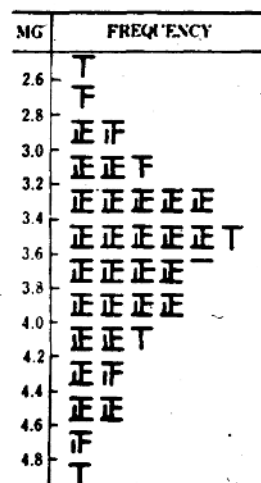
用图表表示的方法很多。首先在纵横两轴上分别画上等距离刻度。但是必须注意，如果刻度的大小不合适，既或把数据画在图表之中，也不能清楚地表现出它们之间的关系。这时，就要反复变更刻度的大小，有时需要把一个轴或两个轴改成对数刻度（使用半对数方格纸或对数方格纸），才能把数据的相互关系或一般的规律表示清楚。

有关出现在论文中的图表的画法，请参阅84—88。

9. 频数分布表、频数分布图 简单地整理数据的最基本的手段是把数据按照适当的项目进行分类。分类的方法，依研究的目的而异。首先是制作频数分布表，这对于综合观察数据来说是很有效的。例如，当测量了从某试验所采集来的许多蝗虫的体长时，假定这些材料中的最大、最小个体分别是5.5、3.2cm，那么，就在3.2cm和5.5cm之间划分成适当的间隔，然后看体长在4.5cm和4.6cm之间的蝗虫有多少；体长在4.6cm以上4.7cm以下的蝗虫有多少……，分别加以整理，并以此为基础制作频数分布表。如有必要，还可以以频数分布表为基础制作频数分布图。根据统计资料直接制作简单的频数分布图时，可按下述方法进行（第1图）：



第1图 根据观测或测定直接制作的频数分布图的
1个例子（件数比较少的场合），1根斜线表示1件



第2图 根据观测或测定直接制作频数分布图的
1个例子（件数较多的场合），1个正字相当于5件

先画1条竖线，分成适当的等间隔，画上刻度。其次，对试样——进行检查，分清哪个试样属于哪个范围，然后在刻度的相当位置的右侧分别画1条斜线（即1条斜线相当于1个数据）。把斜线画成相同的长度、间隔，则可制成如第1图所示的频数分布图。数据数量多时，也可象第2图那样反复画正字（1个正字相当于5个数据）。当通过一个频数分布表制作的分布图得不到相当清楚的关系时，不要马上就灰心失意，这是因为如果重新划分间隔或改变划分方法，进行重新分类，往往还可以把数据之间的关系弄清楚的。

10. **代表值** 一般说来,常常用算术平均值来代表某一事项的某一群的数值。但这并不是说以这种算术平均值为代表值的做法在任何情况下都适用。在不少的情况下,以中位数(*median*)、众数(*mode*)、几何平均值、调和平均值中的某一个值做为代表可能更合适。究竟选定其中的哪一个做为代表值,这要在慎重地考虑了各种情况或因素的基础上合理地确定下来。在难于做出合理的选择的场合,要多取几个值分别做为代表值加以试用,再从中寻找最能弄清关系的代表值。

11. **加权平均** 例如,某月生活费的变动,以上个月为标准(100)进行比较,主食费为102,副食费为97,衣着费为95,供热照明费为100,其它为96。如果把上列数字原封不动地平均一下得98,由此而就得出该月份的生活费用比上个月宽裕2%的结论是错误的。必须把各种费用分别占实际生活费的百分比,换句话说,即把各项费用的权(*weight*)考虑进去进行计算。例如,在进行某一测定时,A测定了3次,得平均值25.3g,其次,B测定了12次,得平均值25.7g的结果,把这两个平均值加起来用2除得25.5g的答案是错误的。应该把A、B两者测定结果的权考虑进去,做

$$\frac{25.3 \times 3 + 25.7 \times 12}{3 + 12} = 25.6g$$

的计算。

12. **误差、散射度、置信界限** 例如“29次的实验”或“用于检查的个体数共计183个”等场合的29或183这些数字本身,都具有绝对的意义,没有必要考虑准确度或置信界限。相反,“测定温度为25.5℃”或“平均生存期间为21.8小时”等场合的数字,则并不是绝对没有偏差的而是带有某种程度变差的数字。严格说来,在记载这些数字的同时还必须明确记载它的变差范围。又如,写做“就1个个体的含量来说,A群为25.5mg,B群为24.3mg,两群之间没有差异”的场合,如果不同时写明双方数字的置信界限,则读者就不了解做出“没有差异……”这一判断的根据。这样,在需要记载置信界限的场合,如果不写上误差范围、散射度、置信界限等,则读者把作者所得的数值用来做参考、比较,或是以此为根据进行其它计算,或是对其评论或结论进行批评或探讨等问题,不是非常困难,就是不可能的。

但是,误差或散射度的表示方法很多。例如,为了表示变差的范围,既有只记其最大值和最小值,或者附记两者之差就足以满足要求的场合(或者,不能做更高的处理时,例如,实验中的温度只在实验的最后通过最高最低温度计读取1次的场合),也有必须取平均偏差、标准差、四分位偏差、概率偏差、变差中的适当的项目表示的情况。究竟应当如何表示,要分别根据每个具体情况慎重确定,而不是无理地从中随意选定一种就可以了事的。

13. **测定误差及其种类** 测定误差的原因,大体可以分为下列5个方面:

1. 仪器、工具不完备
2. 实验、观测所必需的条件、状态不能严格实现
3. 实验材料在实验中不是固定不变或不能严格符合标准
4. 人们在感性判断上的差异或错误

5: 其它各种错误

除此之外, 还有各种情况。当整理实验结果时, 要充分考虑上述各点, 对于结果应如何修正, 误差应该如何表示的问题, 要根据具体情况适当确定。此时, 如果使用不恰当的方法修改结果, 有时反而会更加降低结果的准确度, 需要特别注意。

从数量的特性来看, 误差可以分为下列两种。

a. 固定的或系列性误差: 归根到底, 指的是可以归结为若干固定原因的误差, 它对于所有的测定产生同样的或同一方向的影响, 例如, 尺的长度由于热膨胀而变化, 机器、器械结构上的不完备(刻度不准确或指针回转轴的位置不正等), 个人感觉上的某种缺点(如刻度读数经常过大, 或者确定某一事件的发生瞬间经常超前等), 都属于这种误差。

b. 偶然误差: 是指事先无法预计的完全偶然发生的误差。例如, 器械的可动部分的“空转(play)”或灵敏度不稳定、视觉、触觉等。就同一个人而言, 灵敏度的界限也不一定是固定不变的。虽然企图在温度不变的情况下进行实验, 但严格说来这是很难实现的, 常常是在一瞬间过高、而在另一瞬间又过低等。上述这些情况都是产生偶然误差的原因。此外, 原因不明的偶然误差或变差是非常之多的。

固定的或系列性误差比较容易追究原因并加以避免, 而且既或不能避免, 也可以通过理论的或其它适当的方法进行计算加以订正。但是, 偶然误差则无论如何也是消除不了的。这就要通过误差论(theory of errors)中的方法, 从包括多数偶然误差的测定值的一群中, 确定“似乎最准确的值”。

此外, 还有例如把0.95写成0.59的这种由于过失或疏忽而产生的错误(mistake), 这同上述误差、变差有着本质的不同, 是用任何理论也无法订正的。但是, 其中非常大的误差或变差, 可以通过若干次重复同样测定的方法发现; 非常小的误差或变差可以做偶然误差处理。

如上所述, 千万不要忘记能够合理适用误差论的处理方法, 只是对偶然误差而言的。

通过对于向标的射击的类例的研究, 就可以弄清固定误差和偶然误差的区别。如第3图所示, 弹痕的平均位置即黑点的群落中心偏离在标的的右下方, 这是固定误差。但观察一下每一次射击, 弹痕却是在中心的四周大体上同等程度地分布着, 这相当于偶然误差。在实验、观测中重复同一测定时, 各个测定相当于该图的各种



第3图 向标的射击时的弹痕, 表示测定中的误差同该种情况的相似性

弹痕。从伴同偶然误差的各个中弹位置, 用误差论中的方法可以求出黑点的群落中心, 但不能从这一计算中求出标的位置。

14. 以生物为材料的实验或统计中的变差的特征 例如, 在同一条条件下集体饲养的同龄鱼(不是通过短小形和长大形交配的具有特殊的遗传学交配而饲养的鱼)的体

长或体重的变差等那种在特定条件下的自然物的偶然变差中，与关于测定的偶然误差的理论同样的理论恰好适用，这一点是由我们的经验所证明了的。但是，一般说来，象物理、化学那样，以无生物为材料进行的测定和在生物学、医学中以生物为材料进行的测定，在本质上有着很大的差别。就是说，前者的实验材料一般在实验中没有自身变化，或者自身变化很小。但后者所用的材料，有时在实验过程中变化很大。其中，常常由于材料自身的心理方面的原因起作用，使实验结果发生显著变差。因此，一般说来，把用于物理、化学等方面的偶然误差理论原封不动地用来解决以生物为材料做的实验或统计的结果中的不齐或变差的问题是危险的。有关生物的实验、观察、统计中的结果的变差，同一般的测定误差在本质上是意义完全不同的问题。因此，在变差本身也往往具有重要意义的场合，通过对这种变差的探讨，有时也发现关于某种性质的生物的季节性变化的重要周期。

15. 相关系数 当我们要研究两个事件的关系时，通过比较双方的代表值（当然，要把误差、置信界限等考虑进去）而达到目的场合是很多的。然而，根据问题的性质，使用这样的方法，也有不能令人十分满意的情况。例如，现在想通过对几百枚树叶的长度和宽度的测定，想要弄清长度和宽度之间有没有什么关系。此时，作为一种模糊的全体性的印象，例如说不定叶的长度和宽度成正比例也未可知，但既或算出宽度的平均值和长度的平均值，并用来分别作普通的频数分布图（表），也不能具体地明确它们之间有没有什么关系。不过，把树叶按着长度的顺序规整地排列起来，通过研究在这种情况下在宽度方面有没有某种顺序的方法，就可以大体弄清它们之间的关系，此时，相关图表就可以使这种方法更加精密可靠。于是，在了解了有某种关系之后，下一步则是要弄清是什么样的关系，以及这种关系的程度如何的问题，这是相关系数的任务。

（C）在数据处理中容易出现的错误

对数据进行统计处理是非常有价值的。但如果搞错用法或解释，反而适得其反。下面叙述那些初步的、然而却是经常容易发生的错误。

16. 一般注意事项 (1) 核对：对于数值进行统计处理之前，最重要的问题是最大限度地核对这些数字是否准确。根据数据的性质或复杂程度，有各种核对方法。例如，在调查恰好是1000人的性别时，假定得出男550人、女500人的结果，那就一定能立即发现该调查在什么地方出现了错误。但是例如下述的场合，只是大致看一下的话，是不可能判断它的正确与否的。

在某男女合校高中，要想对恰好1000名学生了解准备升学的人数和不准备升学的人数、男女性别以及家长职业（分为农业户、非农业户两种）之间的关系而作了调查，结果如下表。

不希望升学的学生	661	非农业户	565
不希望升学的农业户	366	希望升学的男生	310
男 生	586	希望升学的男生非农业户	205
男生的非农业户	295		

该表只是看一下，是不能发现这些数字有无错误的。但以希望升学的非农业户男生数205为突破口，就可以依次算出下列的详细数字。

希望升学的男生非农业户	205
不希望升学的男生非农业户	90
希望升学的男生农业户	105
不希望升学的男生农业户	186
不希望升学的女生非农业户	205
不希望升学的女生农业户	180
希望升学的女生非农业户	65

但是，最后剩下的只是希望升学的女生农业户的人数未能从上述计算中直接求出来。不过，它可以把上述7种人的人数加在一起，然后从1000中减掉，是应该能够求出来的。然而，这7种人的总数是1036，所以希望升学的女生农业户的人数为-36，这显然是不合理的。由此才判明上表中的某处有错误。

(2) 对于某一群统计数字，不是经过上述认真分析就可以立即认定该群统计数字具有严密的不变的关系的。例如，某国的自杀人数既或每年几乎不变，也不能得出结论每年必须发生如此这般数量的自杀者。这是因为，导致自杀的确切原因不清楚，而且这种确切的原因是否每年都起作用也不清楚。

(3) 统计数字只是说明有关某一特定群的特点的信息，对于该群中的每个场合并没有任何的具体说明。例如，不能从男的出生数约占总出生数的一半这一常识出发，就具体地推论下一个将要出生的婴儿究竟是男是女。

(4) 在整理统计材料时，有时在选择方法上有意无意地出现偏颇。如果以这些材料为基础进行分析研究，常常会导出错误的见解。特别是材料数量少的场合、目前感到有问题的题材或与此有关的原因的知识不足的情况，就更要特加注意。

(5) 内插法、外插法的危险：使用内插法(interpolation)或外插法(extrapolation)时要特加注意。特别是外插法既或在表示每个实测值的点都在非常整齐的1条很规则的曲线上的场合使用，也不大可能求出足以置信的数值。另外，在实测范围内时和范围外时可以予想情况或条件有不同的场合，既或值的变动在实测范围内是有规律的，使用外插法也几乎是没有什么意义的。例如，某微生物的氧消耗量，既或在15℃到17℃之间表现为随着温度的上升而非常有规律地增加的这个结果，但如果在另一方面已经弄清在40℃附近该微生物的繁殖在减退这一事实的话，那么就可以预想到在40℃附近，代谢当然也要出现异常。所以，根据在未满37℃时测定的结果，用外插法去预见把温度提高到37℃以上时的氧消耗量，是困难的。

17. 比较事件(群)间的数值的场合 将某一事项的数字，在两个以上事件或群之间比较的场合，必须注意下列问题。

(1) 要弄清各群数值的单位是否具有等价性。例如，为了研究日本的一年贸易额，只从数字上看是在第2次世界大战后有明显的增长，但同战前相比，日元的价值是下降了，而且不是说战后的下降处于只是一次性的急剧下降，而在那以后就稳定了下来的状态，因而不能只从金额就轻易地引出某种结论。

(2) 要研究在划分部类时的判定标准是否各群一致。例如，在比较各国发生的强盗

罪的件数或种类问题上,由于不同的国家在犯罪的分类法或法院在处理方法上的不同,要想作出结论也并不是很简单的事。又如,把自己进行的抗结核剂对肺结核的疗效是因病情的轻重而异的研究结果同别人的同一内容的研究报告相比较时,既或两者划分病情轻重的方法相同(例如,极轻、轻、中等、稍重、重),还要注意两者确定轻重程度的标准是否一致。再如,癌症患者人数虽然出现逐年增加的事实,但这一数字仅仅是由于诊断水平日益提高,因而越来越准确的结果。然而,如果是过去的话,由于诊断不准确,常常把那些病因不明或其它病症也都包括进来,结果,也许会出现这种情况,本来实际上癌症患者人数在减少,但表面看来却是在增加。

(3) 要仔细研究收集数据的方法是否同一。例如,在比较A、B两地结核患者死亡率的场合,如果A地的统计是以死亡诊断书为依据、B地的统计是以对各户进行实际情况调查为依据的话,那就不可能由此而得出严密的结论。这是因为,例如,从前,如果谁在诊断书上写的死亡原因是结核的话,深恐被别人发现而讨厌,所以就托医生给改写成其它病名,或者不请医生诊断而自行死去等。再有,某一个人究竟是否因患结核病而死亡的问题,有时外行的判断同医生的判断也不同。此外,调查中所写的死亡原因也可能是伪造的。

(4) 表示事件的尺度或单位不同时,对于它的解释要加以充分注意。例如,为了测定某组织的成长尺度,根据组织的种类、时间、场合,需要使用组织的特定部位的重量或面积、细胞数、分裂中的细胞数、DNA量、氧消耗量等。由于这些值的变化不一定是平行进行的,有时反而会向相反方向变化。

(5) 当各群材料的素质、性质不同时,也需要注意。例如,不能由于A大学或B大学的入学考试竞争率分别为每19.3人、18.5人取1人,就说B大学可以录取更多的优秀考生。如果B大学的入学考试是在A大学以及其它许多大学之后举行的,那么就会有許多考其它大学的落榜生来参加B大学的入学考试,这样,A、B两大学的报考学生素质不同,单纯地从竞争率来判断新生入学质量的优劣,是错误的。

此外,下述事实是我本人的亲身经验,这就是从某个池子采集了复盖在水底上的名为 *Asellus* 的小动物用于某生理学的实验时的整个过程。大约到了7月的时候,*Aselles*的抗药能力突然下降。开始时,我认定这种动物到了夏天,活动力要下降。可是,由于一到夏天就有许多孩子进到池子里来回走,于是我怀疑是否在所采集的*Asellus*中有许多是被孩子们踩过的,由于不了解这个情况用于实验,以致抗药能力下降。于是我就用从附近私人住宅的处于同一条件的池子(当然没有受过孩子们的骚扰)中采集的标本进行实验,结果没发现抗药能力因季节不同而变动的现象,这才弄清我的想象是对的。

(6) 含于双方的群中的个体数大体上不相同,不能用所得数的绝对值比较,应该用含于那个群的总个体数的百分比进行比较。

例如,在pH7.8的海水和pH8.2的液体中分别进行海胆卵的受精,其结果如下。

在pH7.8时,800个卵中的受精卵为656个

在pH8.2时,710个卵中的受精卵为602个,

如果用受精卵的绝对数比较上述结果,则得出pH7.8时的受精率高的结论。但这是不正确的。如果用百分比进行比较。

pH7.8.....受精率 82%

假设这种场合的差异有意义的话，就得出了pH8.2时的受精率高的相反结论。

18. 阐述相关关系的场合 (1) 对于本来互相完全无关的两个事件，既或求出它们之间的关系也是没有意义的。例如，既或日本癌症患者发生例数和咖啡的进口量之间的相关系数非常高，也不能因此就说两者是密切相关的。

(2) 同上述场合不同，A事件和B事件确有关系时，既或A和B之间的相关系数高，也不能只根据这一情况就作为A只同B有密切关联的证据。如果A同B以外的B'、B''、...也好象有关系的话，那就有必要同时就A和B'、A和B''、...之间的相关关系进行研究。例如享受生活困难补助者数目的变动，同生活保护方针上的变化以及工资、物价、就业等变化等都存在相关关系的可能性。在这些可能性中哪一种相关优先的问题，还要在综合考虑其它更多情况之后才能确定。

(3) 由于A事件往往同B事件相关而进行观察，因而就认定A同B是有某种意义的相关，还兼过早。既或在B100例中有90例被确认同A事件相关，但如果不弄清在A事件是以什么样的比例同B以外的事件也相关的话，那也是不能下结论的。例如，人们常常认为在望日的时候天气晴朗，但如果不清楚望月以外时的晴朗天气的比例如何的话，是不能下结论的。

(4) 这里有与B有关的2组不同数据B₁、B₂，虽然A和B₁或A和B₂之间没有大的相关关系，但如果把B₁和B₂的数据混在一起后再研究它同A之间的相关关系时，有时却显示出非常高的系数，然而却不能因此就说A和B是密切相关的。

(5) 把就只有有限的少数事例的群进行研究而计算出的相关系数轻率地适用于含有比之更多的事例的大群之中是危险的。相关系数同其它统计数字一样，一般是依采取标本的范围的大小而显示的值也不同。例如，1对骰子掷100次，如果研究一下一个骰子出现的点儿同另一个骰子的点儿之间的相关关系，将得比0大得多的值。但是，只根据这一点就连置信界限也不加以考虑地说在双方的骰子点儿在显示方法上有关连性，是不可以的（如果再增加试行次数，将发现系数逐渐向0靠近的趋势。因为从理论上讲，双方的骰子点儿是完全没有关系的，当然相关系数应该是0）。