第 1 章　待分析的数据是否值得分析

经典的数理统计是分析数据的一门学科。对于给定的一些数据，数理统计会专注于研究下述内容：如何描述其概率分布规律；如何估计总体中某些参数的数值和研究估计方法的性质；如何基于样本对特定的总体参数进行假设检验和研究假设检验方法的特性；如何揭示结果变量依赖多个原因变量(即影响因素或自变量)变化而变化的规律性；如何基于多个变量的取值情况反映多个个体或样品之间的亲疏关系；如何透过现象探索多个变量之间的本质联系；如何揭示多个变量与多个样品之间的内在关联性。然而，数理统计学家在开始实施前述的任何一类数据分析之前，都是有一个最基本的假定的，这就是“待分析的数据是值得分析的”！在实践中，人们在进行数据分析之前，无论分析者是否为数理统计学家，都应首先弄清这个“假定”是否成立。因为任何拟分析的数据不会因为处理数据的人是“数理统计学家”而变得“俯首帖耳”或“值得分析”。

1.1　哪些情形下获得的数据是不值得分析的

1.1.1　人为编造的数据是不值得分析的

当你试图运用统计分析方法处理数据时，首先要问的问题不是选择什么统计分析方法更合适，而是此数据是否值得分析。

【例 1-1】　对于表 1-1 所示资料，有人曾采用三种分析策略(参见后文中张飞给出的处理方法)得出了两个相互矛盾的结论，这一命题被称为“Symposon 悖论”。事实上，对于该命题，具有聪明头脑的人(简称张飞)与具有正常心态和思维的人(简称百姓)会给出截然不同的解读方法和答案。

表 1-1　同时按年龄和性别分层后吸烟与否与是否患肺癌的调查结果

年龄(岁)	吸烟与否	例　数					
		男性：	患	未患	女性：	患	未患
≤40	吸烟		5	5		40	50
	不吸烟		60	55		5	5
>40	吸烟		30	10		5	55
	不吸烟		30	5		5	35

【分析与解答】　张飞这样处理：分别按三种策略分析此资料，策略一，只考察吸烟与不吸烟患肺癌的概率之间的差别；策略二，在按性别分层的基础上，再采用策略一分析；策略三，先按性别分层，后按年龄分层，再采用策略一分析。得到两种自相矛盾的结论：结论一，吸烟有利于健康(策略一与策略三)；结论二，吸烟有害于健康(策略二)。

百姓这样处理：首先提出一个问题，此资料是否值得分析？经过审读，发现它不值得分析，因为它是造假的产物！何以见得？此资料中有两个“疑点”，由此可推测其纯属人为编

造的“调查资料”。疑点之一，调查了≤40 岁男性吸烟者 10 人，居然有一半人患了肺癌；调查了≤40 岁女性不吸烟者 10 人，居然也有一半人患了肺癌。这样的调查结果只有在肺癌病房中才有可能获得，即肺癌病房中有 5 位患者，另 5 位是照顾病人的家属或护工。疑点之二，随机调查 400 位对象，再按“年龄”、“性别”、“吸烟与否”和“是否患肺癌”将其划分成 16 个小组，每个小组中人数的个位数不是 0 就是 5(由基本常识可知，一个正整数的个位数有 0、1、2、…、9 共 10 种可能性)，出现这种奇特现象的概率是极低的，大约为 $P = \left(\frac{2}{10}\right)^{16} = (0.2)^{16} = 6.5536 \times 10^{-12}$。以如此低的概率出现的奇特现象，在实际调查资料中几乎是不可能出现的！

另外，即便表 1-1 中未出现上述两个疑点，而且，调查资料是在科学完善调查设计方案指导下获得的，张飞的三个分析策略都是不正确的，它们均属于“用单因素分析方法取代多因素分析方法”的错误。正确的分析方法是直接分析三个原因(“年龄”、“性别”、“吸烟与否”)对“是否患肺癌”这个“二值结果变量”的影响是否具有统计学意义。具体地说，可以选用对数线性模型分析，也可以选用多重 logistic 回归模型分析。若分析目的仅仅是希望考察吸烟与不吸烟患肺癌的概率之间的差别是否具有统计学意义，还可以选用 CMH 校正的χ^2检验，此法是把性别和年龄视为两个分层因素，对每一层(研究层内吸烟与不吸烟患肺癌概率之间的差别)做一些基本计算，然后再进行合并计算，最后得出汇总分析的结果和结论(即在消除分层因素影响的前提下，分析吸烟与不吸烟患肺癌概率之间的差别是否具有统计学意义)。

事实上，基于表 1-1 的调查设计，即使在样本含量、样本的代表性和资料的准确性等方面都做得很好，统计分析方法选用得也正确，其结论也不一定就很科学。因为导致一个人是否患肺癌，还有很多可能的影响因素未加以考察，如是否有肺癌的家族史、生活环境(如空气污染情况、饮水的质量、饮食构成、饮食习惯、生活习惯)、锻炼身体情况、职业、劳动强度、家庭经济状况、家庭生活是否幸福美满、人际关系是否良好、心态是否平稳、睡眠质量和每晚平均深睡眠时间长短等。这一切也都与一个人“是否患肺癌”有很大关系，它们被统称为“重要的非试验因素”。在任何一项调查或试验研究中，很关键的一个问题是，是否“找准找全”了对研究目的和主要评价指标有影响的“重要非试验因素”。不仅如此，还必须在尽可能大的样本含量且对总体有极好代表性的受试对象身上准确地获得试验因素、全部重要非试验因素和主要评价指标的数值。所有这一切，都与“数据是否值得分析”密切相关。

由此可知，统计学与纯数学(这里特指数理统计)是有本质区别的！纯数学所研究的数据是有很多隐含的假定的，在这些隐含假定(即所给定的数据是值得分析的)成立的条件下，直接研究数据之间内在的规律性。而统计学则不然，在使用者分析数据之前，必须对那些“隐含假定”逐一进行考证。例如，这些数据是否来自于研究目的相符合的总体(即数据是否具有同质性、样本对于总体的代表性如何)？它们是一个指标的不同取值还是多个指标的不同取值(即是一元资料还是多元资料)？它们是定量的还是定性的(即资料的性质是什么)？它们受到多少个因素的影响(即数据是受单因素影响还是受多因素影响的结果、是什么设计类型下收集的资料)？它们是怎样被收集到的(即是随意或人为选取的还是随机获得的)？所提供的数据对要揭示的问题是否足够多(即样本含量是否充足)？不同组数据之间是否具有可比性(即是否有合理的对照组、组间所受影响是否均衡)？在实施研究或进行试验或调查过程中的质量控制是否严格(即难以控制的众多非试验因素的影响是否被控制在尽可能低的水平上；至少应有

可靠的方法确保所有非试验因素在试验因素的不同水平组中处于基本平衡状态)？只有弄清了上述诸多疑问之后，当答案是“此资料值得分析”时，才能选择具体的统计分析方法(首先是进行探索性统计分析，以了解资料的基本情况，然后再进行正式的统计分析，以便达到合理选择统计分析方法，正确实现研究目的)对资料予以处理。

然而，上述看似烦琐的“循证”过程并非是正确应用统计学的全部“法宝”，它仅仅适用于有了数据之后。在更多情况下，是在仅仅有了研究课题和确定了研究目标之后，就如何能够“多快好省而又科学严谨”地实现既定的研究目标提供“强有力的保证和具有可操作性的做法”，才是统计学的精华之所在！也就是说，统计学的精华是如何制订出科学完善、严谨高效、经济可靠的设计方案。在此方案的指导下，努力做好试验或调查过程中的“质量控制”，才有望获得值得进行统计分析的研究数据。

应当清楚，这里所说的“数据”是广义的，正确的表述应当是“科研资料”。有时，科研资料是纯文本的，更多情况下，科研资料包括原因变量、结果变量、变量的具体取值和变量的专业含义。也就是说，变量、变量的取值(含单位)和变量的专业含义构成了科研资料的三要素。

怎样才能确保所制订出来的设计方案质量高、可操作性强呢？这离不开正确的统计思想的指导，离不开“透过现象看本质”的具体措施保驾护航、离不开“基本常识”的鼎力相助，离不开“各科专业知识”这个坚如磐石的后盾。

【例 1-2】 下面有两组试验记录，试审查此类资料是否值得分析。第一组试验记录见表 1-2，第二组试验记录见表 1-3。

表 1-2　某项试验的试验记录

某药浓度(?)	?	?	?
0	72	97.042	6987.024
25	72	93.431	6727.032
50	72	98.056	7060.032
100	72	96.056	6916.032
200	72	130.694	9409.968
300	72	139.403	10037.020

注：表中“?”处为空白，读者不知道其下方的数据代表什么含义，也不知道它们的单位分别是什么。

表 1-3　某项试验的试验记录

动物编号	性　别	体重(g)	*X*	*Y*	*Z*
1	1	230	3.73	300	16
4	1	225	3.73	600	24～36
5	1	231	3.73	510	16～24
6	0	211	3.73	540	16
12	0	220	3.73	570	24～36

注：*X* 代表气溶胶浓度(mg/L)；*Y* 代表翻正反射消失持续时间(min)；*Z* 代表恢复时间(h)；*Z* 的取值为“16”的还有 7 只动物，即共有 9 只动物都在 16h 恢复。

【分析与解答】 在表 1-2 中，第 3 列上有 2 个数据的小数部分都是“056”，而第 4 列上有 3 个数据的小数部分都是“032”，这种巧合令人难以置信！在表 1-3 中，有 9 只动物的恢复时间都是 16h，还有 3 只动物的恢复时间的跨度很大，前者过于精确，后者过于不准，都说明结果不真实！如此不真实的试验结果是不值得分析的！

1.1.2 产生于质量控制不严的试验的数据是不值得分析的

有些大型科研课题，需要划分成大约 10 个分课题，每个分课题下面可能还需要再分成 5 个子课题。这样一来，子课题的总数目有 30 ~ 50 个。其中，有相当多的观测指标是相同的或在专业上是有联系的，为了探讨这些指标之间的相互关系和依赖关系，人们常常需要把它们合并在一起进行统计分析。很显然，完成这些子课题研究的研究者们的责任心、技术水平等可能相差很多，若事先没有统一的操作规程(SOP)和严格的技术培训，所收集到的试验数据必然是参差不齐甚至杂乱无章的。

【例 1-3】 笔者曾参与评审过一种新药的临床试验项目，该项目由 5 个临床试验中心(通常是 5 所规模相当的医院)共同完成。其中，有一个主要安全性评价指标是某种不良事件发生率(%)，一所医院的试验结果为 22.3%，另一所医院的试验结果为 78.2%，其他医院的试验结果介于这两个数据之间。同一个重要的评价指标在不同医院之间相差如此之大，表明此项临床试验研究的质量控制很差，其整个临床试验数据的准确性非常值得怀疑。因此，质量控制做得不好的任何科研课题(最常见的是多中心临床试验研究课题)所产生的数据都是不值得分析的。

1.1.3 经过错误的方法加工整理后的数据是不值得分析的

【例 1-4】 有些医生常需要采用两种或多种方法对疑似患有某种疾病的患者或患者的标本同时进行检测，以便诊断患者是否确实患了预先推断的某种疾病。某医生收集了同时用 B 超和 CT 检查的 94 例某病患者的检测资料，整理成表 1-4 所示的形式。收集和分析此资料的目的是希望回答可否用 B 超取代 CT。基于表 1-4 的资料，这位医生选择什么统计分析方法可以实现其分析目的呢？

表 1-4 B 超及 CT 检查结果

检查方法	例数			
	轻度	中度	重度	正常
B 超	18	3	3	70
CT	38	7	3	46

【分析与解答】 与这位医生的分析目的对应的统计分析方法属于“一致性检验”或称“Kappa 检验”。然而，该医生将两种仪器检测 94 名患者的结果按表 1-4 的形式整理后，就无法实施一致性检验了。因为表 1-4 属于“结果变量为多值有序变量的 2 ×4 列联表资料”，本质上是“单因素两水平设计一元定性资料”，就是人们习惯表述的“两个独立样本资料”。针对此资料，只能采用秩和检验，其分析结果只能回答“两种仪器检测结果之间的差异是否具有统计学意义”。事实上，实际收集资料是这样操作的：用两种仪器检测每位患者，每种仪器检测结果都可能出现 4 种结果(正常、轻度异常、中度异常、重度异常)之一，故对每位患者而言，两种仪器就有 16 种可能的组合结果之一出现，即应按“配对扩大形式”来列表，见表 1-5。

基于表 1-5(注意：表中的 f_{ij} 处必须是具体的频数)采用 Kappa 检验，就可回答两种仪器检测结果之间是否具有一致性。

表 1-5　B 超及 CT 检查结果比较(例数)(表 1-4 的正确表达形式)

B 超检查结果	例　数				
	CT 检查结果:正常	轻　度	中　度	重　度	合　计
正常	f_{11}	f_{12}	f_{13}	f_{14}	70
轻度	f_{21}	f_{22}	f_{23}	f_{24}	18
中度	f_{31}	f_{32}	f_{33}	f_{34}	3
重度	f_{41}	f_{42}	f_{43}	f_{44}	3
合计	46	38	7	3	94

注：表中的 f_{ij} 为第 i 行第 j 列上的例数，由于原始资料整理错误而无法获知。

1.1.4　不符合特定统计分析方法要求的数据是不值得分析的

【例 1-5】　表 1-6 是某市工业部门 13 个行业 8 项指标的数据，拟采用主成分分析方法确定 8 项指标的样本主成分(综合变量)，若要求损失信息不超过 15%。现分别基于相关系数矩阵、协方差矩阵进行计算，所用程序及结果如下。

表 1-6　某市工业部门 13 个行业 8 项指标的数据

编　号	行　业	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	冶金	90342	52455	101091	19272	82.00	16.100	197435	0.172
2	电力	4903	1973	2035	10313	34.20	7.100	592077	0.003
3	煤炭	6735	21139	3767	1780	36.10	8.200	726396	0.003
4	化学	49454	36241	81557	22504	98.10	25.900	348226	0.985
5	机械	139190	203505	215898	10609	93.20	12.600	139572	0.628
6	建材	12215	16219	10351	6382	62.50	8.700	145818	0.066
7	森工	2372	6572	8103	12329	184.40	22.200	20921	0.152
8	食品	11062	23078	54935	23804	370.40	41.000	65486	0.263
9	纺织	17111	23907	52108	21796	221.50	21.500	63806	0.276
10	缝纫	1206	3930	6126	15586	330.40	29.500	1840	0.437
11	皮革	2150	5704	6200	10870	184.20	12.000	8913	0.274
12	造纸	5251	6155	10383	16875	146.40	27.500	78796	0.151
13	文教艺术用品	14341	13203	19396	14691	94.60	17.800	6354	1.574

注：X_1 表示年末固定资产净值(万元)；X_2 表示职工人数(人)；X_3 表示工业总产值(万元)；X_4 表示全员劳动生产率(元/人·年)；X_5 表示百元固定原资产值实现产值(元)；X_6 表示资产利税率；X_7 表示标准燃料消费量；X_8 表示能源利用效果。

以下是算法 1(基于相关系数矩阵)进行主成分分析的第一部分计算结果：

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	3.10491252	0.20747090	0.3881	0.3881
2	2.89744162	1.96722608	0.3622	0.7503
3	0.93021555	0.28809329	0.1163	0.8666
4	0.64212226	0.33803813	0.0803	0.9468
5	0.30408413	0.21748637	0.0380	0.9848
6	0.08659776	0.05441338	0.0108	0.9957
7	0.03218438	0.02974261	0.0040	0.9997
8	0.00244178	·	0.0003	1.0000

从这部分分析结果可以看出，取前 3 个主成分就可以解释全部原始数据 86.66% 的信息。

以下是算法 2(基于协方差阵)进行主成分分析的第一部分计算结果：

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	5.44387E10	4.63867E10	0.8668	0.8668
2	8052070898	7827548141	0.1282	0.9950
3	224522758	136007774	0.0036	0.9986
4	88514983.7	86405779.1	0.0014	1.0000
5	2109204.6	2105915.97	0.0000	1.0000
6	3288.62949	3274.59278	0.0000	1.0000
7	14.0367172	13.9062234	0.0000	1.0000
8	0.13049382	.	0.0000	1.0000

从这部分分析结果可以看出，只需取 1 个主成分，就可以解释原始数据 86.68% 的信息。算法 2 的分析结果看上去似乎更为合理，更好理解。

【分析与解答】 一般而言，根据上述算法 2 中的这个第 1 主成分的值，可以对 13 个行业进行排序和分类。而且，由于这个第 1 主成分主要包含的是 X_7 的信息，也就是说，本资料可以根据 X_7 的信息对 13 个行业进行排序和分类。

表面上，基于算法 2 分析本资料后得到了一个令人兴奋的结果！其实不然，在上面的分析和解释过程中，已经犯了一个不可饶恕的错误！

显然，由算法 2 所得的结果，与算法 1 所得的结果很不一样。算法 1 是基于相关系数矩阵进行的主成分分析；算法 2 是基于方差协方差矩阵进行的主成分分析。按照统计学上的基本常识，这两种算法所得结果不应相差很大，而对本资料而言，其结果却相差十分悬殊。问题出在资料不符合主成分分析的要求上！

严格地说，主成分分析适合用在单组设计多元定量资料上。所谓单组设计，即受试对象或样品来自于同一个总体，他们(或它们)对于研究目的和全部评价指标而言应具有相同的性质。此时，可以考察不同定量指标之间内在的相互和依赖关系，甚至包括它们与隐变量(即主成分变量)之间的关系。而在本例中，13 个行业之间存在着本质区别，X_7 一个指标就可明显区分出哪些行业是以“燃料消费多”为标志的行业。用这一个指标就可对这 13 个行业进行排序，这个结果似乎令人感到十分欣慰，但这是十分片面的，甚至是相当错误的，因为它仅能反映哪些行业燃料消耗量大、哪些行业燃料消耗量小，而并不能反映其他任何方面！

所以，进行主成分分析时，务必要强调受试对象或样品应具有同质性，不要盲目地滥用主成分分析。若分析本资料的目的是为了基于给定的 8 项定量指标进行综合考虑，给 13 个行业进行排序，建议采用“综合评价”方法，如秩和比法。

综上所述，什么样的数据值得进行统计分析呢？数据必须满足以下四个前提条件：其一，数据是真实的，不是造假的，也不是随意添加或删除某些数据而形成的；其二，所取数据的试验(或调查)设计无严重错误；其三，数据在收集和整理过程中没有出现任何严重的失误或偏倚；其四，数据应满足拟选用的统计分析方法的要求(或前提条件)。

以上四点都是十分重要的，对事先确定的研究目的来说：

第一点关系到研究者或数据分析者的科研道德和科研作风问题。研究者或数据分析者无

论出于什么目的，造假及随意添加或删除数据都是绝对不允许的！

第二点关系到对总体的定义是否清楚，所抽取的样本对于总体的代表性是否很好，样本中个体的同质性是否好，样本是否足够大，所考察的影响因素是否“找准找全”，所观测的指标是否涵盖了与研究目的对应的全部主要指标，指标的测定时间是否合适和测定结果是否准确，课题设计的质量成为所收集的数据是否值得分析的第一道防线。

第三点关系到试验（或调查）过程中可能会受到来自环境、试验条件（仪器设备、试剂等）、研究者和受试者的心理等因素或条件的改变对观测结果造成的歪曲实际的影响，“质量控制”的严格程度成为所收集的数据是否值得分析的第二道防线。

第四点应当注意的是，应根据设计类型、比较类型（如新药临床试验中，还有非劣效性检验、等效性检验、优效性检验，通常的假设检验被称为差异性检验）和统计分析目的，选择最合适的统计分析方法，而不应当盲目套用。例如，有些人针对某病患者与正常人的多项定量指标的数值，进行变量聚类分析和样品聚类分析，这是很不合适的。若研究目的是考察全部定量指标的平均值（即均值向量）在两组之间的差别是否具有统计学意义，就应当选用成组设计定量资料多元方差分析处理资料；若研究目的是希望构造出函数式，以便对新个体究竟属于某病患者还是属于正常人进行判定，就适合选用判别分析。还有人面对多项二值变量及其取值时，采用主成分分析和探索性因子分析，完全忽视了这些多元统计分析方法要求拟分析的资料应当是多元定量资料的前提条件。考察拟分析的科研数据与拟选用的统计分析方法是否吻合成为所收集的数据是否值得分析的第三道防线。

1.1.5　盲目解释基于误用统计分析方法所得到的分析结果是不可取的

有些研究者习惯于脱离实际或无中生有地解释多元统计分析的计算结果。有些多元资料并不适合所选用的某些多元统计分析方法，得出的计算结果是解释不清的。然而，分析者却凭自己对多元统计分析方法的一知半解，盲目解释计算结果，不仅牵强附会，更严重的是对未来的研究工作或政策制订会产生歪曲的指导。

【例 1-6】　原文题目为《过敏性紫癜患儿血清白三烯 B4 白介素 -5 的测定及其临床意义》，测得患儿 31 例，正常儿 27 例的血清 IL-5、LTB4、CRP 的结果如表 1-7 所示。

表 1-7　两组血清 IL-5、LTB_4、CRP 的检验结果（$\bar{x} \pm s$）

组　别	例　数	IL-5（pg/ml）	LTB_4	CRP
正常组	27	12.7 ± 3.2	17.6 ± 5.7	4.75 ± 2.85
患儿组急性期	31	53.8 ± 4.2	95.3 ± 12.0	36.10 ± 11.78
恢复早期	31	37.8 ± 3.9	45.7 ± 10.1	18.35 ± 6.43

原作者对此资料做了多种统计分析，其中包括直线相关分析，给出了相关系数的计算结果，见表 1-8，该计算结果可信吗？

表 1-8　血清 IL-5、LTB_4、CRP 的相关系数 r（n = 89）

	IL-5	LTB_4	CRP
IL-5	—	0.772	0.715
LTB_4	0.772	—	0.735
CRP	0.715	0.735	—

【分析与解答】 原作者给出的表1-8中的计算结果是不可信的！因为所做的线性相关分析是错误的。道理很简单，由表1-7可知，任何两个定量变量的89对数据实际上来自58名儿童，“89”是“例次”而不是“例数”，因为有31名儿童被重复测量了两次。即使都是不同的个体，把他们的数据放在一起进行线性相关分析也是错误的，因为他们是不同质的。正常儿童与处于不同时期的患病儿童，在与所研究疾病有关的指标上的取值大小是不同的，而进行线性相关与回归分析有两个隐含的前提条件：其一，数据所测自的受试者必须具有同质性；其二，所研究的变量或指标在专业上应有联系(例如，若研究人的身高与转氨酶之间的相关关系或回归关系就缺乏专业依据了)。

1.1.6 缺失值过多的数据是不值得分析的

有些课题在实施过程中出现了不可控的严重问题，导致所收集的科研数据严重缺失。例如，在新药临床试验研究中，可能由于未采用盲法分配受试对象或盲法执行得不够严格，导致有相当比例的受试者在试验过程中要求更换治疗方案或干脆脱离此项临床试验研究，这些受试者的试验数据就不完整，较多变量或较多时间点上的变量表现为缺失值。当具有缺失值的受试者数目的比例占总例数的20%以上时，应宣布此临床试验失败了。又例如，在一个调查研究课题中，若发出的问卷调查表的数目为10万份，而收回的调查问卷为1万份，其中各项内容填写有效的问卷为3千份。若仅依据这3千份有效调查问卷，即使采用各种正确的统计分析方法予以处理，得出的结论很可能是无用的，甚至有可能是有严重错误的。因为有效问卷的回收率仅为3%，这些被调查者很可能在全部10万名被调查者中对调查目的或其中某项项目具有很强的倾向性。换句话说，有效问卷回收率低的调查数据是不值得进行统计分析的。

1.2 怎样保证数据是值得分析的

1.2.1 必须制订出科学完善的科研设计方案

由1.1节介绍的内容可知，数据是否值得分析的关键有两条，其一，考察科研设计是否正确，只有在无严重错误的科研设计方案指导下收集的数据才具备值得分析的首要前提；其二，在科研工作实施(试验或调查或从文献中提取信息)过程中，考察质量控制是否严格，是否在可能出现问题的所有环节上都事先想到了，重在防范一切可能影响结果准确可靠性的非试验因素的干扰和影响，一旦出现也有切实可行的对策使其影响降到最低程度。

什么是科研设计呢？在开始一项科学研究之前，需要提出研究目标，为了很好地实现事先制订的研究目标所做的一切考虑和安排，称为该研究项目或研究课题的科研设计方案，包括课题框架设计方案和课题技术设计方案两大类。对于任何一个科研课题而言，要想圆满地完成事先确定的研究目标，要考虑的问题或方面往往是很多的，稍有不慎，一旦落下了某些重要方面或在某些方面考虑不周，就可能导致整个课题研究前功尽弃。因此，对于任何一个科研课题来说，事先制订出科学完善的科研设计方案显得尤为重要。

科研设计的涵盖面非常广，其基本内容概括起来如图1-1和图1-2所示。图1-1是从结构上来划分的；而图1-2是从功能上来划分的，它实际上就是图1-1中“课题技术设计方案”的一种变形。

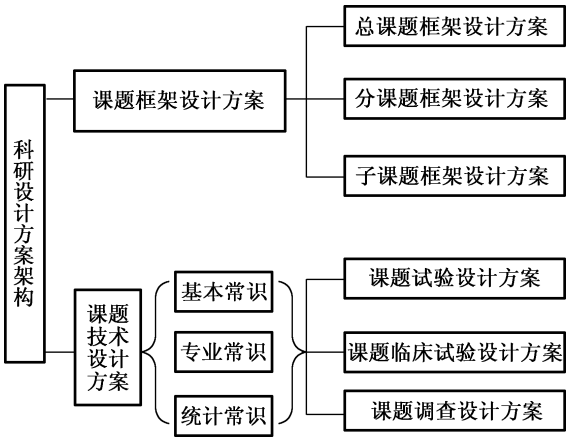

图 1-1　科研设计方案架构图

在图 1-2 中，专业设计包括基本常识与专业知识；统计设计包括试验性研究设计和观察性研究设计，而试验性研究设计又根据受试对象的不同，可进一步划分为试验设计(以动物或样品为受试对象)和临床试验设计(以健康人或某病患者为受试对象)。

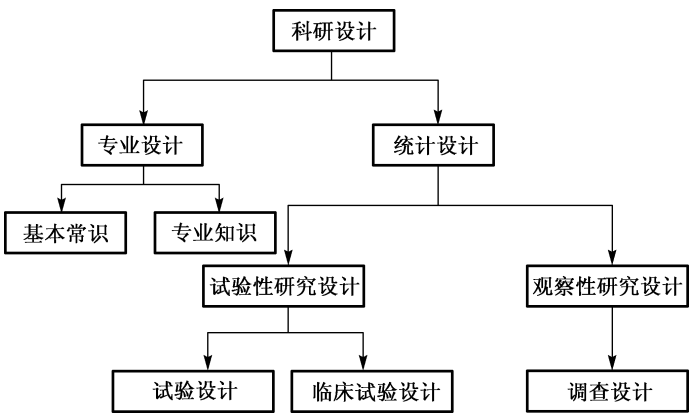

图 1-2　科研设计框图

在图 1-2 中，似乎仅涉及统计学知识，其实不然，对于任何一个科研课题，只具备统计学知识而不具备与拟研究问题有关的专业知识的制订者，制订不出有用的统计设计方案。事实上，制订统计设计方案的人，不仅需要具备丰富的专业知识，更需要具备正常人的思维和心态，要用够用足基本常识，并且要会将基本常识、专业知识和统计学知识相互融合，才有可能制订出具有极高使用价值的统计设计方案，如图 1-3 所示。

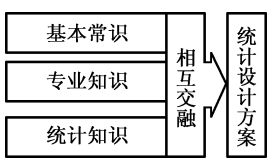

图 1-3　课题统计设计方案的内核

在科研课题获得批准之后且课题正式实施之前，应制订出科学完善的科研设计方案，其标志的具体内容如下：

(1)人力、物力、财力和时间满足设计要求。也就是说，申请书上计划的科研条件在实际科研环境中均能一一兑现。例如，检测某种物质需要一种昂贵的仪器设备，而本课题组已经拥有或可以借到，或可以请他人代为检测，若国内没有一个单位拥有，则属于“空头支票”，不具备完成科研课题的基本条件。又如，完成此科研课

题需要 10 个专业领域的专家参与，而现有研究人员仅涉及 6 个专业领域，另外 4 个专业领域的研究人员目前几乎不可能得到。笔者在检查重大和重点科研课题时经常发现，一些涉及多因素、多指标的复杂科研课题，按照所制订的研究目标，必须有高水平的统计学专家和计算机专家花较多时间自始至终参与，并花大力气研究和攻关，方可达到事先制订的研究目的。然而，在那些课题组的全部参与人员中，却找不到具备前述提及条件的那两类专家，这属于人力或技术水平上达不到设计要求。再比如，要想完成某项难度较大的科研课题，至少需要 5 年时间，而申请时考虑不周，计划 3 年完成，这属于时间上不符合设计要求，事先夸下了海口，而无力实现。

(2)试验设计的“三要素”(受试对象、影响因素、试验效应)、“四原则”(随机原则、对照原则、重复原则、均衡原则)、“设计类型”(安排因素及其水平的一种架构)和“质量控制”(确保所获得的科研数据准确可靠的措施)均符合专业和统计学要求。这方面的内容很多，请参见有关“医学统计学”专著。这里仅举一例，拟使用某种动物制造一种用于试验的动物模型，但这种动物很特殊，根据目前的条件尚无法获得，这就属于三要素中的受试对象不符合设计要求。

必须强调指出的是，对于三要素中的受试对象，应把握质量标准和数量要求。质量标准中包括受试对象种类的确定和五类标准(纳入标准、排除标准、中止标准、终止标准、剔除标准)的制订；数量标准指有根据地估算出每个小组中需要的合适样本含量。影响因素中应特别重视找全找准重要非试验因素，并尽可能按重要非试验因素对受试对象进行分层随机化。试验效应方面，注意三类指标(诊断性指标、疗效评价指标和安全性评价指标)都应有主次之分，还应注意指标的测定时间和测定方法。

在四原则方面，应巧妙地使用分层随机化，确定合理的对照组，有根据地估计各小组所需要的样本含量和确保组间具有高度的均衡性。

(3)重要的影响因素(包括试验因素和非试验因素)和观测指标没有遗漏，并作了合理安排。影响因素和观测指标属于三要素中的两个重要要素，再一次特别提出来，是为了强调其重要性。因为在很多科研课题的设计中，研究者常将某些重要的试验因素和观测指标遗漏，导致科研结果回答不了需要回答的问题，甚至很容易得出错误的结论。

(4)必须找准找全重要的非试验因素并使之都得到很有效的控制。很多科研课题的结论经不起时间和实践的检验，问题的症结就在于重要的非试验因素未得到有效的控制，它们“污染”了试验结果，从而得出了不够正确、甚至错误的结论。

(5)研究过程中可能出现的各种情况都已考虑在内，并有相应的对策。这一点也是至关重要的。比如，在新药临床试验研究中，若中途退出的受试者人数较多，可能会导致整个临床试验前功尽弃，一旦发现可能出现较高比例的受试者退出临床试验，应预先想好对策，尽可能降低退出比例。再比如，某些受试者的某些测定指标的数值明显偏高或偏低，应事先对此种可能的异常现象有所警惕，必要时应对所有受试者进行多次重复观测，用每位受试者该指标多次测定值的平均值取代其测定指标的数值，以降低个别受试者在个别时间点上出现难以处理的异常值。有时，极少数异常值可能会颠覆整个研究结论。

(6)对操作方法，试验数据的收集、整理、分析等均有一套明确的规定和方法。尤其是有多个单位多位研究者合作的研究课题，统一规定试验操作规程，统一培训，仪器和试剂统一质量指标和校准，有规范化且符合试验设计要求的收集和整理资料的表格，对获得准确可靠试验资料是极其重要的。有了数据后，将由谁来进行数据处理，将采用哪些统计分析方法分析，事

先都应做到胸中有数，不要等到试验做完再来考虑请谁做数据分析。到那时，很可能这个科研课题的数据已无分析的价值了。因为若科研设计有严重错误或质量控制有问题或收集资料的方法不正确，都可能得到包含很多错误信息的资料，无论用何种高级的统计分析方法去分析，其结论都是荒谬的。也就是说，做统计分析的人应参与一开始的科研设计，应在科研过程中参与质量控制，应协助研究人员确保科研资料准确、可靠、完整。

1.2.2　必须严格控制课题实施过程中的质量

对于科研课题实施过程中的质量控制，事先考虑的内容和拟采取的对策应该包含在科研设计方案之中。若在实施过程中出现一些影响结果正确性的问题，但的确未包括在事先制订的科研设计方案之中的，应临时补充和修改与之有关的科研设计方案的内容，尽可能把突发事件所造成的坏影响降到最低程度。

在科研工作中，较多的属于试验性研究。下面就针对试验研究过程中如何做好质量控制作一些详细论述。首先，通过一幅图(见图 1-4)粗略地了解质量控制的大致内容；然后，再对质量控制作一些详细的论述。

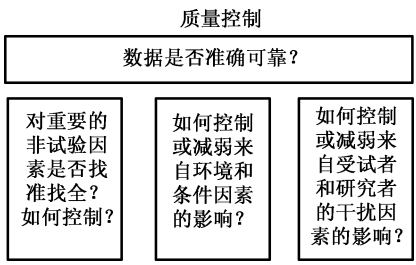

图 1-4　课题实施过程中质量控制的内容

1. 进行质量控制的必要性

由图 1-4 可知，无论试验设计方案制订得多么科学严谨，它毕竟是纸面上的东西，要将其落到实处，需要通过实施者所持有的态度和采取的具体行动，需要受试对象(特别是人)的全力配合，需要环境和条件处于设计者所希望的状态。然而，在较长时间的试验过程中，前述提及的诸多方面都或多或少会发生一些变化，一旦那些变化超出了允许的范围，试验结果就会偏离其真实值较远，结论的可信度就必然会大打折扣。因此，对试验过程中出现的可能会影响结果准确度的任何问题，都应能及时发现，并能在尽可能短的时间内消除其影响。因此，在试验过程中，进行质量控制是非常必要的。

2. 进行质量控制的环节与措施

需要进行质量控制的环节包括：控制研究者的责任心、心理状态和技术水平对试验结果可能造成的不利影响；控制受试者的心理状态和不依从性对试验结果可能造成的不利影响；控制环境的变化(如温度、湿度、光线、清洁度等)对试验结果可能造成的不利影响；控制试验条件的改变(如仪器的状态发生改变、试剂过期、观测指标的单位不同、测定的方法不同等)对试验结果可能造成的不利影响。

对试验过程中的质量进行控制的措施有：对不同单位的多名研究者进行统一的培训，内容包括增强责任心教育、调适好心理状态(在言行上不应有个人的倾向性或偏好)、接受正规的技术培训，并要求每位研究者自觉遵守统一的操作规程。对于动物或样品类的受试对象而言，选用与研究目的和试验因素相对应(同质)的动物或样品；对于某病患者的受试对象而言，制订严格的纳入与排除标准。尽可能采取双盲(受试对象和研究者都不知道具体的分组结果)或三盲法(受试对象、研究者和数据统计分析者都不知道具体的分组结果)分配受试对象，尽可能减少或消除来自研究者，特别是受试者心理的影响。对于环境因素的影响，尽可能创造条件，使试验研究在特定的温度、湿度、光线和清洁度条件下进行；对于仪器设备、试剂尽可能

校准，并经常对其进行检查和调试；对于指标的测定单位、测定方法应完全统一，实在不能统一的，应采取合适的方法进行换算；对于主要评价指标，若测定方法无法统一，应确定一个有资质的测定部门来完成测试，称为“中心试验室”，这是消除多个试验研究单位测定结果参差不齐的重要措施之一。

1.2.3 必须有实时记录科研数据的规格化表格

1. 实时记录科研数据的规格化表格的种类

(1)调查表

在调查研究中，通常都会事先编制一个与调查目的吻合的调查表。通常每位被调查对象的调查结果都记录在一张调查表上。

(2)病例报告表

在临床试验研究中，通常都会事先编制出一个与临床试验目的吻合的病例报告表(简称CRF，Case Report Form)。通常每位临床受试对象的试验结果都记录在一张病例报告表上。

(3)各种设计类型对应的表

在以动物或样品为受试对象的试验研究中，不存在与前述类似的调查表或病例报告表，只有与试验设计类型对应的表。在试验研究中，与各种试验设计类型对应的表格形式是相对固定的，只要在试验开始之前，把那些规格化的表格编制出来，实时记录试验数据的问题就迎刃而解了。

2. 与常见试验设计类型对应的规格化表

(1)单组设计

【实际问题1】 从一组受试对象身上观测某一个定量指标的数值，研究者关心的是，如何推断该单组设计一元定量资料所代表的总体平均值与一个给定的标准值或理论值之间的差别是否具有统计学意义，属于什么统计学问题？

【例1-7】 已知玉米单交种群105的平均穗重为300g。喷药后，随机抽取9个果穗，其穗重(g)分别为308、305、311、298、315、300、321、294、320。问喷药后与喷药前的果穗平均重量之间的差别是否具有统计学意义？

【分析与解答】 这是一个单组设计一元定量资料的记录与呈现的问题，只需要写出一个定量指标的名称(含单位)及其在所有受试对象上的取值，不需要用表格形式呈现。

【实际问题2】 研究者希望对一组受试者同时观测两项或多项在专业上有一定联系的定量指标的数值，希望同时考察这些定量指标的平均值与多个给定的标准值之间的差别是否具有统计学意义，属于什么统计学问题？

【例1-8】 为探讨结核病患者的营养状况，某研究者对2000年收治的45例患者的营养状况进行了调查，调查的指标包括白蛋白(ALB)、白蛋白/球蛋白(A/G)比值，数据见表1-9。已知这些指标的下限分别为：白蛋白35g/L，白蛋白/球蛋白1.5。试评价结核病患者的营养状况。

【分析与解答】 这是一个单组设计二元定量资料的记录与呈现的问题，可用表1-9的形式呈现单组设计二元定量资料。

表 1-9　45 例结核病患者的营养状况

ID	ALB	A/G	ID	ALB	A/G	ID	ALB	A/G
1	30.2527	0.9332	16	29.4046	0.4953	31	37.1407	0.4976
2	37.2461	1.1546	17	31.7185	1.3290	32	24.2056	1.0685
3	31.9332	1.1154	18	26.3070	0.9545	33	34.1015	0.6992
4	38.2970	0.8270	19	42.2041	1.0580	34	36.4848	0.8745
5	29.7258	1.4642	20	28.2290	0.8142	35	41.3322	1.0341
6	40.2492	1.1738	21	32.0035	0.9595	36	28.3449	1.1210
7	29.6518	1.7670	22	33.4976	0.8141	37	41.3502	1.1723
8	33.9290	0.7915	23	28.7575	0.9331	38	47.1740	0.5704
9	35.8982	1.4726	24	23.6329	1.3659	39	25.1091	0.9331
10	33.2792	1.2376	25	32.3452	0.7535	40	30.5990	0.9196
11	33.0777	0.3949	26	35.2587	0.7430	41	33.3743	0.6501
12	23.7641	0.8388	27	33.1799	1.3128	42	28.5072	0.6796
13	41.5968	1.1160	28	34.1699	0.8658	43	34.8188	0.9514
14	39.0836	1.2769	29	31.8015	1.2527	44	29.3374	0.6549
15	33.1353	1.2509	30	31.5403	0.9285	45	33.9863	1.0733

(2)配对设计

【实际问题】 怎样才能尽可能消除来自受试对象的重要非试验因素对结果的影响，使两种处理所产生的效应之差能更真实地表现出来？

【例 1-9】 对血小板活化模型大鼠以 ASA 进行试验性治疗，以血浆 TXB_2(ng/L)为指标，其结果见表 1-10。

表 1-10　大鼠血小板活化模型 ASA 治疗前后血浆 TXB_2 的变化

大鼠号	血浆 TXB_2(ng/L)	
	给药前	给药后
1	250	184
2	226	205
⋮	⋮	⋮
10	176	176

【例 1-10】 从 8 窝仔猪中每窝选出性别相同、体重接近的两头仔猪进行饲料营养价值对比试验，将每窝两头仔猪随机分配到两个饲料组中，试验期 30 天，结果见表 1-11。

表 1-11　喂饲不同饲料的仔猪体重

窝号	仔猪体重(kg)	
	甲饲料	乙饲料
1	10.0	9.8
2	11.2	10.6
⋮	⋮	⋮
8	10.8	9.8

【例 1-11】 一位社会学者为了判定聪明人选择的配偶是否也聪明，随机选取 32 对夫妻，并测得智商得分(IQ)，见表 1-12。问配偶之间的差别是否有统计学意义？

表 1-12　32 对夫妻 IQ 得分测定结果

配偶号	IQ 得分	
	丈夫	妻子
1	95	95
2	103	98
⋮	⋮	⋮
32	96	102

【分析与解答】 上述 3 个问题看似相差甚远，但它们在试验设计类型上都叫做“配对设计”，其资料可称为“配对设计一元定量资料”。其微小的区别是各对数据的“配对条件”不同而已。表 1-10 被称为“自身配对设计”；表 1-11 被称为“同源配对设计”；表 1-12 被称为“条件相近者配对设计(本例为‘家庭’)”。

若在上述各种情况下，同时测定两项或多项在专业上有一定联系的定量指标的数值，列表格式类似，只是纵向上的定量指标的数目增多了而已。其资料被称为配对设计多元定量资料。

(3)单因素 K 水平设计($K \geqslant 2$)

【实际问题】 在找不到恰当的配对或配伍条件时，如何能真实地显露出两种或多种处理的效应之间的差别是否具有统计学意义？

【例 1-12】 一个小麦新品种经过 6 代选育，从第 5 代(A 组)中抽出 10 株，株高(cm)为 66、65、66、68、62、65、63、66、68、62；又从第 6 代(B 组)中抽出 10 株，株高(cm)为 64、61、57、65、65、63、62、63、64、60。问株高性状是否已经达到稳定？

【例 1-13】 某研究者欲考察单抗、利巴韦林两种药物对乙型脑炎病人的治疗效果，将 16 名病情等方面相近的患者随机等分到两个处理组中去，测量退热时间(日)，具体数据如下。

单抗组：1，2，1，2，5，9，6，4；

利巴韦林组：32，13，6，7，10 ，8，5，7。

请问两组治疗所得的平均效果之间的差别是否具有统计学意义？

【例 1-14】 从津丰小麦 4 个品系中分别随机抽取 10 株，测量其株高(cm)，数据如下。

品系 0－3－1：63，65，64，65，61，68，65，65，63，64；

品系 0－3－2：56，54，58，57，57，57，60，59，63，62；

品系 0－3－3：61，61，67，62，62，60，67，66，63，65；

品系 0－3－4：53，58，60，56，55，60，59，61，60，59。

问不同品系津丰小麦的平均株高之间的差别是否具有统计学意义？

【例 1-15】 试就表 1-13 所示资料说明大白鼠感染脊髓灰质炎病毒后，再作伤寒或百日咳预防接种是否影响生存日数。

表 1-13　感染脊髓灰质炎病毒后的三组大鼠接受不同处理后生存日数的记录结果

编号	生存日数		
	处理：接种伤寒疫苗	接种百日咳疫苗	空白对照
1	5	6	8
2	7	6	9
3	8	7	10
4	11	10	12
5	11	16	14
6	12	11	16

【分析与解答】 上述 4 个问题看似相差甚远，但它们在试验设计类型上都叫作单因素 K 水平设计($K \geqslant 2$)问题。当试验因素的水平数为 2 时，还可简称为成组设计。其中，前 3 个问题未用表格形式呈现。其实，完全可以用类似表 1-13 的形式呈现设计和定量资料。

若在上述各种情况下，同时测定两项或多项在专业上有一定联系的定量指标的数值，列表

格式类似，只是纵向上的定量指标的数目增多了而已。其资料称为单因素 K 水平设计（$K \geqslant 2$）多元定量资料。

（4）无重复试验随机区组设计

【实际问题】 怎样才能尽可能消除来自受试对象的重要非试验因素对结果的影响，使两种以上的处理所产生的效应能更真实地表现出来？换句话说，如何才能把三种配对形式拓展或镶嵌到单因素多水平设计中去？

①如何才能把条件相近者配对设计拓展或镶嵌到单因素多水平设计中去？

【例 1-16】 某研究者欲比较 3 种抗癌药物对小白鼠肉瘤的抑瘤疗效，先将 15 只染有肉瘤的小白鼠按体重大小配成 5 个区组，使每个区组内的 3 只小白鼠体重最接近，然后随机决定每个区组中的 3 只小白鼠接受 3 种药物的治疗。以肉瘤的重量为指标，试验结果见表 1-14。试比较不同抗癌药物对小白鼠肉瘤的抑瘤效果之间的差别是否有统计学意义。

表 1-14　不同药物作用后小白鼠肉瘤重量测定结果

区　组	肉瘤重量(g)		
	药物种类：	A	B	C
1		0.82	0.65	0.51
2		0.73	0.54	0.23
3		0.43	0.34	0.28
4		0.41	0.21	0.31
5		0.68	0.43	0.24

②如何才能把同源配对设计拓展或镶嵌到单因素多水平设计中去？

【例 1-17】 为了比较准确地评价分别含 5 种不同蛋白质饲料的营养价值，有人设计了如下的试验：用 8 窝大白鼠，每窝选取性别相同、体重接近的 5 只，用随机的方法将每窝中的 5 只分到 5 个饲料组中去，收集的试验数据见表 1-15。

表 1-15　8 窝大白鼠分别用 5 种饲料喂养 9 周后所增体重测定结果

窝　别	体重增加量(g)				
饲料：	甲	乙	丙	丁	戊
1	24	15	37	57	82
2	42	28	37	51	66
⋮	⋮	⋮	⋮	⋮	⋮
7	47	21	32	54	73
8	53	37	42	59	90

【分析与解答】 上述两个实际问题看上去很不同，但在试验设计类型上它们都是无重复试验的随机区组设计（一个为试验因素，另一个为区组因素。通常，区组因素为与受试对象有直接关系的“属性因素”，如窝别、性别、血型、体重分组、病情等）。不同点仅仅在于区组因素不同，前者为“体重分组”，后者为“窝别”。对定量的观测指标而言，若遗传因素对其影响大于非遗传因素，则选择“窝别”作为区组因素效果比较好，反之亦然。

若在上述各种情况下，同时测定两项或多项在专业上有一定联系的定量指标的数值，列表格式类似，只是纵向上的定量指标的数目增多了而已。其资料称为无重复试验随机区组设计多元定量资料。

(5)有重复试验随机区组设计

【实际问题】 如何在随机区组设计中体现重复试验?

【例 1-18】 现有 3 种降低转氨酶的药物 A、B、C,为了考察它们对甲型肝炎和乙型肝炎患者转氨酶降低的程度之间的差别是否有统计学意义,收集试验数据,见表 1-16(即从两型患者的总体中各随机抽取 9 例,然后分别随机均分入 3 个药物组中去),此为具有 3 次独立重复试验的随机区组设计(以前称为 3×2 析因设计,但不够严格,因为"型别"是一个属性因素)。其中,药物是试验因素;型别是属性因素。

表 1-16 3 种药物对两型肝炎患者转氨酶降低效果的试验结果

肝炎型别	转氨酶降低值(kPa)								
药物种类:	A 药			B 药			C 药		
甲型	100	85	90	120	90	110	50	60	40
乙型	65	75	100	45	30	50	50	60	45

【例 1-19】 某研究者希望研究 3 种减肥方案的效果,分别从办公室与车间选取女工,资料见表 1-17,这是具有 5 次重复试验的随机区组设计。其中,减肥方案为试验因素;工作地点为区组因素(也是一种特殊的属性因素)。

表 1-17 3 种减肥方案中办公室和车间女工的体重减少量

工作地点	体重减少量(磅)														
减肥方案:	食疗				运动				食疗+运动						
办公室	6	2	10	−1	8	3	4	−2	6	−2	8	12	7	10	5
车间	3	15	4	8	6	−4	6	8	−2	3	15	8	10	16	3

【分析与解答】 无重复试验与有重复试验的随机区组设计之间的区别在于以下两点:其一,有重复试验时,能更好地反映受试对象的个体差异的大小;其二,有重复试验时,可以分析试验因素与区组因素之间是否存在不可忽视的交互作用。

若在上述各种情况下,同时测定两项或多项在专业上有一定联系的定量指标的数值,列表格式类似,只是纵向上的定量指标的数目增多了而已。其资料被称为有重复试验随机区组设计多元定量资料。关于多元定量资料的表达,可如法炮制,后面就不赘述了。

(6)具有一个重复测量的单因素设计

【实际问题】 如何才能把自身配对设计拓展或镶嵌到单因素多水平设计中去?

【例 1-20】 某外科医生在每个病人的手术过程中的 3 个不同阶段测得病人门静脉压力的资料见表 1-18,研究者希望考察 3 个阶段门静脉压力之间的差别是否具有统计学意义。

表 1-18 10 名病人在手术过程中的 3 个不同阶段门静脉压力的测定结果

病例号	门静脉压力(cmH_2O)		
A(阶段):	A_1	A_2	A_3
1	40	36	36
2	19	17	18
⋮	⋮	⋮	⋮
10	36	37	38

【分析与解答】 该资料所代表的试验属于具有一个重复测量的单因素设计,若只有 2 个阶段,就属于自身配对设计了。

若在上述情况下，同时测定两项或多项在专业上有一定联系的定量指标的数值，列表格式类似，只是纵向上的定量指标的数目增多了而已。其资料称为具有一个重复测量的单因素设计多元定量资料。

(7)平衡不完全随机区组设计

【实际问题】 在某试验研究中，若试验因素的水平数大于每个区组内可以容纳的受试对象个数，应该如何处置呢？

【例 1-21】 用 4 种方法治疗脚气，受试者为两只脚都患脚气的患者，每只脚接受 1 种处理，观察指标为治疗效果评分。设计格式和资料见表 1-19。

表 1-19　4 种方法治疗脚气的疗效评分

受试者编号	疗效评分			
	治疗方法：	甲	乙	丙	丁
1		5	3	…	…
2		…	…	4	7
3		2	…	6	…
4		…	4	…	8
5		3	…	…	7
6		…	2	5	…

【分析与解答】 这是一个平衡不完全随机区组设计。此设计比较烦琐，单个体型区组因素(每行多个数据来自一个受试对象)身上能接受处理的“单位数”小于处理因素的水平数。需要选用多少个个体，必须通过计算才能确定下来。

本例中，$v=4$(药物有 4 种)，$a=2$(每位患者形成一个区组，每人只有 2 只脚)，$r=3$(每种药物被重复使用的次数)，$b=6$(区组个数，即本例中的患者人数)，满足 $rv=ab$，即 $3\times4=2\times6$；同时满足 $\lambda=r(a-1)/(v-1)=3(2-1)/(4-1)=1$ 为整数。故本例为平衡不完全区组设计。受试对象是 6 名患脚气病的患者，试验因素是“治疗方法”，区组因素为每位患者(即个体差异)，定量的观测指标为“疗效平分”。

(8)无重复试验的双因素设计

【实际问题】 如何看待两个试验因素各水平组合条件下只做一次试验的结果？

【例 1-22】 将落叶松苗木栽在 4 块不同的苗床(B)上，每块苗床的苗木又分别使用 3 种不同的肥料(A)以观察肥效差异，2 年后于每个苗床的各施肥小区内用随机方法抽取一株，测量其高度，资料见表 1-20。

表 1-20　不同肥料 A 和不同苗床 B 对落叶松苗高影响的测定结果　　(单位：cm)

肥　料　A	苗高(cm)				
	苗床 B：B_1	B_2	B_3	B_4	合　　计
A_1	55	47	47	53	202
A_2	63	54	57	58	232
A_3	52	42	41	48	183
合计	170	143	145	159	617

【分析与解答】 在上述试验中，涉及两个试验因素，但因素各水平组合条件下未做独立重复试验。

【特别提示】 因为当试验中同时涉及两个试验因素时，通常都需要先考察它们之间的交

互作用是否具有统计学意义。而当未做独立重复试验时，对其定量指标的试验结果进行统计分析是无法考察交互作用项的效应大小的。因此，基于此设计所得出的结论是否正确取决于以下两点：其一，两试验因素之间的交互作用可忽略不计；其二，各试验条件下的试验结果十分稳定，其误差在专业上允许的范围之内。

(9)不可拆分的单个体型无重复取样的拉丁方设计

【实际问题】 如何才能使用最少的样本含量获得比较准确的试验结果？

【例 1-23】 试评价 4 台秤测定结果是否一致。样品重量为观测指标，设计和资料见表 1-21。

表 1-21 4 台秤称样品重量的一致性试验

样品代号	秤的代号(样品重量，mg)			
	测定顺序： 1	2	3	4
1	C(　)	B(　)	A(　)	D(　)
2	B(　)	A(　)	D(　)	C(　)
3	A(　)	D(　)	C(　)	B(　)
4	D(　)	C(　)	B(　)	A(　)

【分析与解答】 此为单个体型无重复取样(一个体反复被观测)的 4×4 拉丁方设计。上述安排称为不可拆分的单个体型无重复取样的 4×4 拉丁方设计，此设计涉及一个试验因素、两个区组因素。一般认为，试验因素与区组因素之间的交互作用项的效应很微弱，可以忽略不计。

【特别提示】 表中括号内填写样品重量；所选取的 4 个样品的重量最好不等，且包含实际使用时重量的最小值和最大值，这样有利于评价结果在一定重量范围内的准确性。

(10)可以拆分的单个体型有重复取样的拉丁方设计

【实际问题】 当样品可以被分成若干份时，如何才能使用最少的样本含量获得比较准确的试验结果？

【例 1-24】 观察 A、B、C 3 种中药的促凝作用，以生理盐水作对照(D)。每次试验时，从动物身上抽取血样，往血样中加入不同药物(包括生理盐水)，观测指标为“血凝时间”。现取品种相同、体重接近的雄性白兔 4 只，每只白兔的血样被均分成 4 份，以不同的顺序接受不同的药物处理，以血浆复钙凝固时间为指标，设计和资料见表 1-22。

表 1-22 不同中药组测得的血凝时间

白兔编号	中药代号[血凝时间(s)]			
	试验顺序号：1	2	3	4
1	C(85)	D(102)	A(90)	B(100)
2	D(108)	A(88)	B(110)	C(80)
3	A(95)	B(98)	C(82)	D(110)
4	B(105)	C(78)	D(104)	A(92)

【分析与解答】 此为单个体型有重复取样(抽自一个体的血样被均分成 4 份)的 4×4 拉丁方设计。上述安排称为可拆分的单个体型有重复取样的 4×4 拉丁方设计。应特别注意的是，若将药物给动物灌服后再测量指标的数值，则上述拉丁方设计所得结果是最不准确的，在实际科研中，应尽可能避免这样的设计。

(11)多个体型无重复试验的拉丁方设计

【实际问题】 何时必须选用多个体型区组因素？