

该表是两个样本率比较的基本数据,其中 14, 86, 30, 90 是两个样本率比较的基本数据,其他数据是从这四个数据推算出来的,因此这样的资料称为四格表资料。其一般形式见表 3.4。

表 3.4 四格表资料

组别	阳性数	阴性数	合计
1	$a(E_{11})$	$a(E_{12})$	R_1
2	$a(E_{21})$	$a(E_{22})$	R_2
合计	C_1	C_2	N

表格中括号内的值代表各自格子的理论频数,任一格的理论频数均可由公式 $E_{ij} = \frac{R_i C_j}{N}$ 计算出,其中 N 为总例数。

用以四格表资料检验的统计量为 $\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, O_{ij} 表示实际观察的频数, C 为行数, R 为列数。显然, O_{ij} 与 E_{ij} 相差越小, $\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ 就越小, χ^2 值也就越小,认为 O_{ij} 与 E_{ij} 之间吻合程度越高,它们来自同一总体的可能性就比较大;反之, χ^2 值越大, O_{ij} 与 E_{ij} 之间吻合程度越差,可认为两率来自同一总体时,抽到这样大 χ^2 值的可能性比较小。 χ^2 值的大小除与 $O_{ij} - E_{ij}$ 有关外,亦随格子数(即自由度)的增大而加大。自由度 $\nu = (R - 1)(C - 1)$ 。因此四格表的自由度 $\nu = (R - 1)(C - 1) = (2 - 1)(2 - 1) = 1$ 。

四格表资料的 χ^2 检验主要用于两个样本率的比较。

1. 提出假设

$H_0: \pi_1 = \pi_2$ 两个样本率所代表的两个总体是同一总体,即两样本来自同一总体;

$H_1: \pi_1 \neq \pi_2$ 两样本来自不同的总体;

给定显著性水平 $\alpha = 0.05$ 。

2. 计算检验统计量

首先计算每一格子的理论频数,再计算每一格子的实际观察

频数与期望频数差的平方并除以相应的期望频数。然后用公式

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \text{ 计算 } \chi^2 \text{ 统计量。}$$

3. 确定概率 P 值

计算 χ^2 的自由度 $\nu = (R - 1)(C - 1)$, 并根据给定的显著性水平 α , 查 χ^2 临界值表, 得到 $\chi_{\alpha, \nu}^2$ 。对于四格表资料, 由于自由度为 1, 因此 $\chi_{\alpha, \nu}^2 = \chi_{0.05, 1}^2 = 3.84$ 。

4. 判断结果

依据以上的统计量及得到的概率值和假设中的检验水平比较, 进行统计推断。结果如表 3.5 所示。

表 3.5 临界值与检验假设的关系

χ^2 值	P 值	假设	判断
$< \chi_{0.05, \nu}^2$	> 0.05	不拒绝 H_0	差异无显著性
$\geq \chi_{0.05, \nu}^2$	≤ 0.05	拒绝 H_0	差异有显著性
$\geq \chi_{0.01, \nu}^2$	≤ 0.01	拒绝 H_0	差异有极显著性

实际应用时, 为了简化计算, 由 χ^2 检验的基本公式可推导出四格表专用公式 $\chi^2 = \left(\frac{ad - bc}{R_1 R_2 C_1 C_2} \right)^2 N$ 。其中 a, b, c, d 分别是四个格子的实际观察频数, N 为总例数。应用该公式时, 只需将四格表内的观察频数一次代入, 便可计算出 χ^2 统计量。

由于数学上的 χ^2 分布是连续性的分布, 用 χ^2 统计量检验列联表独立性假设时, 是通过用连续性 χ^2 分布作为 χ^2 检验统计量分布的近似方法寻求检验的概率, 而 χ^2 检验统计量往往超度, 特别是在小自由度的情况下。因此人们提出了矫正的方法。考虑到 χ^2 检验统计量由样本含量、自由度和实际观察频数与期望频数的离差三部分构成, 在四格表中, 当样本含量 N 较大 (大于 40), 各格子的期望频数均大于 5 时, 虽然其自由度只有 1, 但由于此时仍有较好的近似性, 为方便计算, 可以不予矫正。但某一格的期望频数很

小时,常会使 χ^2 分量值不合理地变大。例如,当实际观察频数 $O = 30$,期望频数 $E = 20$ 时,此时 $O - E = 10$, χ^2 分量为 $(30 - 20)^2/20 = 5$,但当 $O = 14, E = 4$ 时,尽管仍有 $O - E = 10$,但 χ^2 分量为 $(14 - 4)^2/4 = 25$ 。为此,在四格表中,当样本含量大于 40,如果某一格子的理论频数小于 5 时,就必须采用连续校正,校正公式为

$$\chi_c^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}, \text{ 或在专用公式基础上进行:}$$

$$\chi_c^2 = \frac{\left(|ad - bc| - \frac{N}{2}\right)^2 N}{R_1 R_2 C_1 C_2}。$$

例 5 已知甲、乙两种药物治疗某病的结果见表 3.6。问两种药物有效率的差异是否有统计学显著性?^[1]

表 3.6 甲、乙两种药物治疗某病的效果比较

组别	有效例数	无效例数	合计
甲药	3	31	34
乙药	7	11	18
合计	10	42	52

解:观察例数 $N = 52$,计算各格子的理论频数,有 $E_{21} = 3.46 < 5$,应作连续性校正。设甲、乙两种药物的治疗有效率分别为 π_1 和 π_2 ,并假定二者相同,两药物样本的有效率不同是由于抽样引起的。于是

$$H_0: \pi_1 = \pi_2; H_1: \pi_1 \neq \pi_2; \alpha = 0.05。$$

$$\text{由于 } \chi_c^2 = \frac{\left(|ad - bc| - \frac{N}{2}\right)^2 N}{R_1 R_2 C_1 C_2} = \frac{\left(|3 \times 11 - 31 \times 7| - \frac{52}{2}\right)^2 52}{34 \times 18 \times 10 \times 42} = 5.05,$$

自由度 $\nu = (R - 1)(C - 1) = (2 - 1)(2 - 1) = 1$,按 $\alpha = 0.05$ 水平,查 χ^2 临界值表,得 $\chi^2 > \chi_{0.05,1}^2 = 3.84, \chi^2 < \chi_{0.01,1}^2 = 6.63$,得到 $0.01 < P < 0.05$ 。

于是,本例在 0.05 概率水平下拒绝 H_0 ,认为两种药物疗效差

异在统计上有显著性意义。

在四格表的 χ^2 检验中,若有理论频数小于1,或者样本含量小于40,尤其是用其他检验方法计算的概率接近检验水平时,较理想的检验方法是四格表的精确概率法。四格表的精确概率法的基本思想是:在四格表周边合计不变的条件下,由公式 $P = \frac{R_1!R_2!C_1!C_2!}{a!b!c!d!N!}$ 可直接计算出表内四个数据的各种组合下的概率。

因四格表的自由度为1,在计算各种组合的时候,只需依次增减四格表中任何一个格子的数据,便可得到周边合计不变条件下的各种组合的四格表。计算时,如果是双侧检验,需要分别计算各种组合中所有 $|O - E|$ 等于和大于样本 $|O - E|$ 值的各个四格表的按上式算出的 P 值,然后将这些 P 值相加得到双侧检验的 P 值;若是单侧检验,按研究目的计算一侧的所有 $|O - E|$ 等于和大于样本 $|O - E|$ 值的各个四格表的按上式算出的 P 值,然后将这些 P 值相加得到单侧检验的 P 值。

例6 用新旧两种药物治疗某种疾病的结果如表3.7所示,问新药的效果是否比旧药更好?

表 3.7 新旧两种药物治疗某种疾病的效果比较

组别	治愈例数	未愈例数	合计
新药	5	0	5
旧药	2	4	6
合计	7	4	11

解:由于样本例数 $N = 11 < 40$,本例需用四格表的精确概率法计算。

$H_0: \pi_1 = \pi_2; H_1: \pi_1 \neq \pi_2; \alpha = 0.05$ 。样本的 $|O - E|$ 值 $E_{11} = \frac{5 \times 7}{11} = 3.2$ 。根据研究目的,判定为两样本含量不等的单侧检验,

算出周边合计不变的各种组合,选择一侧组合中 $|O - E| \geq 1.8$

者,分别按公式 $P = \frac{R_1!R_2!C_1!C_2!}{a!b!c!d!N!}$ 计算 P 值。

1	2	3	4	5
1 4 6 0	2 3 5 1	3 2 4 2	4 1 3 3	5 0 2 4
5 6 7 4 11	5 6 7 4 11	5 6 7 4 11	5 6 7 4 11	5 6 7 4 11

$|O-E|$ 2.2 1.2 0.2 0.8 1.8

依题意,只需计算5号四格表的 P 值:

$$P = \frac{R_1!R_2!C_1!C_2!}{a!b!c!d!N!} = \frac{5!6!7!4!}{5!0!2!4!11!} = 0.0455$$

按 $\alpha = 0.05$ 水平拒绝 H_0 ,认为新药的治疗效果比旧药好。

(二) 行 \times 列表资料的 χ^2 检验

四格表的基本数据只有两行两列,对于多于两行两列的情况,统称为行 \times 列表或列联表,简记为 $R \times C$ 表。行 \times 列表资料的 χ^2 检验主要用于解决多个样本率的比较、样本构成比的比较以及计数资料的关联性分析,其检验的基本原理、计算 χ^2 的基本公式与四格表 χ^2 检验相同,不过在行 \times 列表中计算各格子的理论频数十分烦琐。其统计量计算公式为

$$\chi^2 = \left(\sum_{i=1}^R \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}} - 1 \right) N$$

式中各符号的意义同前,其自由度为 $\nu = (R - 1)(C - 1)$ 。

1. 多个样本率的比较

例7 研究复方哌唑嗪对高血压病治疗效果的临床试验并与复方降压片和安慰剂作对照,结果如表3.8所示,问三种药物效果有无差别?

表 3.8 三种药物治疗高血压病的有效率比较

组别	有效	无效	合计	有效率/%
复方哌唑嗪	35	35	40	87.50
复方降压片	20	10	30	66.67
安慰剂	7	25	32	21.88
合 计	32	40	102	60.78

解: $H_0: \pi_1 = \pi_2 = \pi_3$, 即三种药物的总体有效率相同; $H_1: \pi_1, \pi_2, \pi_3$ 不全相同。

$$\chi^2 = 102 \times \left(\frac{35^2}{40 \times 62} + \frac{5^2}{40 \times 40} + \frac{20^2}{30 \times 62} + \frac{10^2}{40 \times 30} + \frac{7^2}{62 \times 32} + \frac{25^2}{40 \times 32} - 1 \right) = 32.74$$

自由度 $\nu = (R - 1)(C - 1) = 2$, 给定 $\alpha = 0.05$, 查 χ^2 临界值表得 $P \leq 0.05$, 于是拒绝 H_0 , 认为三种药物的治疗效果不完全相同。

2. 样本构成比的比较

例 8 1979 年某地爆发松毛虫病, 333 例患者按年龄以 14 岁为界分成 2 组, 资料如表 3.9, 试考察两组病人病变类型的构成比有无差别?

表 3.9 某地两组松毛病患者型别构成比较

年龄分组	皮炎型	骨关节炎型	软组织炎型	混合型	合计
儿童组	50	48	18	72	188
成人组	105	10	7	23	145
合计	155	58	25	95	333

解: H_0 : 两组患者中患各型松毛虫病的构成比相同;

H_1 : 两组患者中患各型松毛虫病的构成比不同;

$\alpha = 0.05$ 。

按公式计算统计量的值为 $\chi^2 = 70.14$, $\nu = (R - 1)(C - 1) = 3$ 。

查 χ^2 临界值表得 $P \leq 0.05$, 于是拒绝 H_0 , 认为两组患者中患各型松毛虫病的构成比不同。

第四节 方差分析

方差分析是检验多个总体均值是否相等的统计方法, 被广泛

应用于分析心理学、生物学、工程和医药的试验数据检验。它通过检验各总体的均值是否相等来判断分类型自变量对数值型因变量是否有显著影响。我们把被检验的对象称为“因素”或“因子”，而因素的具体表现称为“水平”。在只有一个因素的方差分析（称为单因素方差分析）中，涉及两个变量：一个是分类型自变量，一个是数值型自变量。当研究分类型自变量对数值型因变量的影响时，所用的方法就是方差分析。之所以叫方差分析，是因为在判断均值之间是否有差异时要用到方差。具体地说就是通过对数据误差来源的分析来判断不同总体的均值是否相等，进而分析自变量对因变量是否有影响。因此，进行方差分析时，需要考虑误差的来源。一般情况下，来自同一总体中的样本，其观察值是各不相同的，它们的差异可能是由于抽样的随机性造成的，称之为随机误差。而来自不同总体的样本，其观测值也各不相同，这种差异除抽样的因素之外，还可能由于系统因素造成的，称之为系统误差。衡量因素的不同水平（同一总体）下样本数据的误差，称为组内误差，衡量因素的不同水平（不同总体）下各样本之间的误差，称为组间误差。显然，组内误差只包含随机误差，而组间误差既包含随机误差，也包含系统误差。

方差分析的前提是基于下面的三个假定：

(1) 每个总体都服从正态分布，即对于因素的每一个水平，其观测值是来自正态分布的简单随机样本；

(2) 各个总体的方差 σ^2 必须相同。或者说，各组观测数据是从具有相同方差的正态总体中抽取的；

(3) 观测值是相互独立的。

在上述假设成立的前提下，要分析自变量对因变量是否有影响，实际上就是要检验自变量的每个水平（总体）的均值是否相等。下面对问题的一般提法进行归纳。

设因素有 k 个水平，每个水平的均值分别用 $\mu_1, \mu_2, \dots, \mu_k$ 来表示，要检验 k 个水平（总体）的均值是否相等，提出如下假设：

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ 自变量对因变量没有显著影响

$H_1: \mu_1, \mu_2, \cdots, \mu_k$ 不全相等 自变量对因变量有显著影响

对于上面的问题,如果用一般的假设检验方法,如 t 检验,由于一次只能检验两个样本,要检验来自 k 个总体的样本,需要做 $C_k^2 = \frac{k(k-1)}{2}$ 次不同的检验,不仅十分繁琐,而且置信水平也会降低,犯第一类错误的概率也相应地增加。而方差分析不仅可以提高检验的效率,而且由于它是将所有的样本信息结合在一起,因此也增加了分析的可靠性。

一、单因素方差分析

为叙述方便,在单因素方差分析中,我们用 A 表示因素,因素的 k 个水平(总体)分别用 A_1, A_2, \cdots, A_k 表示,每个观测值用 x_{ij} ($i = 1, 2, \cdots, k; j = 1, 2, \cdots, n$) 表示,即 x_{ij} 表示第 i 个水平(总体)的第 j 个观测值。其中,从不同水平中所抽取的样本容量可以相等,也可以不等。应用时,为了清楚直观,常常列出单因素方差分析的数据结构,见表 3.10。

表 3.10 单因素方差分析的数据结构

样本单位(编号)	观测值			
	A_1	A_2	...	A_k
1	x_{11}	x_{21}	...	x_{k1}
2	x_{12}	x_{22}	...	x_{k2}
⋮	⋮	⋮		⋮
n	x_{1n}	x_{2n}	...	x_{kn}

为检验 H_0 是否成立,需要确定检验统计量。其过程可以概括为以下步骤:

(1) 计算因素各水平的均值。假定从第 i 个总体中抽取一个容量为 n_i 的简单随机样本,令 \bar{x}_i 表示第 i 个总体的样本均值,则有

$$\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i} \quad (i = 1, 2, \cdots, k)$$

其中 n_i 为第 i 个总体的样本观测值的个数, x_{ij} 为第 i 个总体的第 j 个观测值。

(2) 计算全部观测值的总均值。它是全部观测值总和除以观测值的总个数。令总均值为 \bar{x} , 则

$$\bar{x} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}}{n} = \frac{\sum_{i=1}^k n_i \bar{x}_i}{n}$$

式中, $n = n_1 + n_2 + \cdots + n_k$ 。

(3) 计算误差平方和。为构造检验统计量, 在方差分析中, 需要计算三个误差平方和:

① 总误差平方和。它是全部观测值与总平均值的误差平方和, 反映了全部观测值的离散情况。计算公式为

$$\text{总误差平方和} = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$$

② 水平项误差平方和。它是各组平均值与总平均值的误差平方和, 反映各总体的样本均值之间的差异程度, 又称为组间平方和。计算公式为

$$\text{水平项误差平方和} = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$$

③ 误差项平方和。它是每个水平或组的各样本数据与其组平均值误差的平方和, 反映了每个样本各观测值的离散程度, 又称为组内平方和或残差平方和。该平方和实际上反映的是随机误差的大小。其计算公式为

$$\text{误差项平方和} = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

上述三个平方和之间的关系为:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$$

(4) 计算统计量。如果 $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ 为真, 则表明没有系统误差, 组间平方和除以其自由度后的均方与组内平方和除以其自由度后的均方差异就不会太大; 如果组间均方显著大于组内均方, 说明各水平之间的差异不仅有随机误差, 还有系统误差。要检验这种差异, 就要构造一个用于检验的统计量。我们知道, 各误差平方和的大小与观测值的多少有关, 为了消除观测值多少对误差平方和大小的影响, 需要将其平均, 也就是用各平方除以它们所对应的自由度, 这一结果称为均方。设全部观测值的个数为 n , 则总误差平方和的自由度为 $n - 1$; 水平项误差平方和的自由度为 $k - 1$, 其中 k 为因素水平(总体)的个数; 误差项平方和的自由度为 $n - k$ 。由于我们主要是比较组间均方和组内均方之间的差异, 所以通常只计算水平项误差平方和的均方和误差项平方和的均方。二者的比就是所需要的检验统计量 F , 当 H_0 为真时, 二者的比服从第一自由度为 $k - 1$ 、第二自由度为 $n - k$ 的 F 分布。计算出检验统计后, 将统计量的值 F 与给定的显著性水平 α 的临界值 F_α 进行比较, 从而做出对原假设 H_0 的决策。

对于给定的显著性水平 α , 在 F 分布表中查找与第一自由度 $k - 1$ 、第二自由度 $n - k$ 相应的临界值 $F_\alpha(k - 1, n - k)$, 若 $F > F_\alpha$, 则拒绝 H_0 , 即 $\mu_1 = \mu_2 = \cdots = \mu_k$ 不成立, 表明 $\mu_i (i = 1, 2, \cdots, k)$ 之间的差异是显著的, 即所检验的因素对观测值有显著的影响; 若 $F < F_\alpha$, 则不拒绝 H_0 , 没有数据表明 $\mu_i (i = 1, 2, \cdots, k)$ 之间有显著的差异, 或者说, 还不能认为所检验的因素对观测值有显著的影响。

为了使计算过程更加清楚, 通常将上述过程的内容列成方差分析表, 见表 3. 11。

表 3.11 方差分析表的一般形式

误差来源	平方和	自由度	均方	F 值	P 值	F_{α}
组间(因素影响)	水平项误差平方和	$k - 1$	组间均方	组间均方 / 组内均方		
组内(误差)	误差项平方和	$n - k$	组内均方			
总和	总误差平方和	$n - 1$				

二、双因素方差分析

单因素方差分析只是考虑一个分类型自变量对数值型因变量的影响。实际中,有时需要考虑几个因素对试验结果的影响。当方差分析涉及两个分类型自变量时,称为双因素方差分析。在双因素方差分析中,如果两个影响因素对试验结果的影响是相互独立的,这时的双因素方差分析称为无交互作用的双因素方差分析,否则称为有交互作用的双因素方差分析。由于后者理论上相对复杂且计算较烦琐,这里只介绍无交互作用的双因素方差分析。

在无交互作用的双因素方差分析中,由于有两个因素,在获取数据时,需要将一个因素安排在“行”的位置,称为行因素;另一个因素安排在“列”的位置,称为列因素。设行因素有 k 个水平:行 1, 行 2, \dots , 行 k ; 列因素有 r 个水平:列 1, 列 2, \dots , 列 r 。行因素和列因素都可以搭配成一组,观察它们对试验指标的影响,共抽取 $k \times r$ 个观察数据,其数据结构如表 3.12 所示。

表 3.12 中,行因素共有 k 个水平,列因素共有 r 个水平。每一个观测值 x_{ij} 看作是由行因素的 k 个水平和列因素的 r 个水平所组合成的 $k \times r$ 个总体中抽取的容量为 1 的独立随机样本。这里 $k \times r$ 个总体中的每一个总体都服从正态分布,且有相同的方差。

表 3.12 双因素方差分析的数据结构

	观测值				平均值
	列 1	列 2	...	列 r	\bar{x}_i
行 1	x_{11}	x_{12}	...	x_{1r}	\bar{x}_1
行 2	x_{21}	x_{22}	...	x_{2r}	\bar{x}_2
⋮	⋮	⋮		⋮	⋮
行 k	x_{k1}	x_{k2}	...	x_{kr}	\bar{x}_k
平均值 \bar{x}_j	$\bar{x}_{\cdot 1}$	$\bar{x}_{\cdot 2}$...	$\bar{x}_{\cdot r}$	$\bar{\bar{x}}$

$\bar{x}_i = \frac{\sum_{j=1}^r x_{ij}}{r}$ ($i = 1, 2, \dots, k$) 是行因素的第 i 个水平下各观测

值的平均值; $\bar{x}_j = \frac{\sum_{i=1}^k x_{ij}}{r}$ ($j = 1, 2, \dots, r$) 是第 j 个水平下各观测值

的平均值; 而 $\bar{\bar{x}} = \frac{\sum_{i=1}^k \sum_{j=1}^r x_{ij}}{kr}$ 是全部 $k \times r$ 个样本数据的总平均值。

双因素方差分析的步骤如下:

(1) 提出假设

对行因素提出的假设为: $H_0: \mu_1 = \mu_2 = \dots = \mu_k; H_1: \mu_1, \mu_2, \dots, \mu_k$ 不完全相等。

式中 μ_i 为行因素的第 i 个水平的均值。

对列因素提出的假设为: $H_0: \mu_1 = \mu_2 = \dots = \mu_r; H_1: \mu_1, \mu_2, \dots, \mu_r$ 不完全相等。

式中 μ_j 为列因素的第 j 个水平的均值。

(2) 构造检验统计量

为检验 H_0 是否成立, 需要分别确定检验行因素和列因素的统

计量。总误差平方和是全部样本观测值与全部的样本平均值的误差平方和,记为 SST,即

$$\begin{aligned} SST = \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - \bar{x})^2 &= \sum_{i=1}^k \sum_{j=1}^r (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^k \sum_{j=1}^r (\bar{x}_j - \bar{x})^2 \\ &+ \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2 \end{aligned}$$

上式右端的第一项是行因素所产生的误差平方和,记为 SSR,第二项是列因素所产生的误差平方和,记为 SSC;第三项是除行因素和列因素影响之外的剩余因素影响产生的误差平方和,称为随机误差项平方和,记为 SSE。

在上述误差平方和的基础上,换算成均方,即将各平方和除以各自的自由度。SST 的自由度为 $kr - 1$;SSR 的自由度为 $k - 1$;SSC 的自由度为 $r - 1$;SSE 的自由度为 $(k - 1)(r - 1)$ 。因此,行因素的均方(记为 MSR)为 $MSR = \frac{SSR}{k - 1}$;列因素的均方(记为 MSC)为

$$MSC = \frac{SSC}{r - 1}; \text{ 随机误差项的均方(记为 MSE)为 } MSE = \frac{SSE}{(k - 1)(r - 1)}。$$

为检验行因素对因变量的影响是否显著,采用下面的统计量:

$$F_R = \frac{MSR}{MSE} \sim F[k - 1, (k - 1)(r - 1)]$$

为检验列因素对因变量的影响是否显著,采用下面的统计量:

$$F_C = \frac{MSC}{MSE} \sim F[r - 1, (k - 1)(r - 1)]$$

(3) 统计决策

根据给定的显著性水平 α 和两个自由度,查 F 分布表得到相应的临界值 F_α ,然后将 F_R 和 F_C 与 F_α 进行比较。

若 $F_R > F_\alpha$,则拒绝原假设 H_0 ,认为 $\mu_1 = \mu_2 = \dots = \mu_k$ 不成

立, 即有 $1 - \alpha$ 的把握认为所检验的行因素对观测值有显著影响。

若 $F_C > F_\alpha$, 则拒绝原假设 H_0 , 认为 $\mu_1 = \mu_2 = \dots = \mu_k$ 不成立, 即有 $1 - \alpha$ 的把握认为所检验的列因素对观测值有显著影响。

为使计算过程更加清楚, 对上面的计算内容列成方差分析表, 见表 3.13。

表 3.13 双因素方差分析表

误差来源	误差平方和	自由度	均方	F 值	P 值	F_α
行因素	SSR	$k - 1$	MSR	F_R		
列因素	SSC	$r - 1$	MSC	F_C		
误差	SSE	$(k - 1)(r - 1)$	MSE			
总和	SST	$kr - 1$				

第五节 统计推断中的常见错误辨析

在科技论文中所出现的统计错误, 绝大部分都是统计推断方面的错误。由于统计推断在理论上有一定的深度, 许多作者在实际应用时并未真正理解统计推断的相关原理, 照本宣科, 结果导致各种错误的发生。现将统计推断中的常见错误按统计推断的应用类型进行分类分析。

一、常见的抽样推断错误分析

由于总体中的每个个体的观察指标具有个体的差异, 因而一般情况下, 样本均数不会恰好等于总体均数, 从而引起抽样误差。但根据数理统计学中的中心极限定理, 样本均数的分布近似服从正态分布, 即使从一个偏态总体中抽样, 当样本数足够大时, 样本的均数也近似服从正态分布。中心极限定理还表明, 从总体中随机抽取一个样本, 当这个样本的含量逐渐增大时, 样本的均数就会逐

渐接近总体的均数。或者当抽取的样本数量很多时,样本的均数也接近总体的均数,即样本均数就是总体均数。因此,抽样误差和样本含量是统计推断最重要的两个因素。抽样误差越小,表示样本均数与总体均数越接近,用样本均数估计总体均数的可靠性就越大,反之,用样本均数估计总体均数的可靠性就越小。但许多医学刊物中所刊载的抽样研究的文章的作者未提供抽样误差,使编者及读者无法估计总体参数的可信度及可靠程度。有些作者对样本统计值之差未进行显著性检验,仅凭资料表面的差值比较来加以判断^[6]。针对上述的两个因素,结合统计量的抽样分布规律,可以将实际应用中的一些常见错误列举如下:

1. 样本含量太小

这是一些科技论文中最常见的错误,作者往往只是利用了较小的样本含量就盲目地给出分析的结论。这固然是实验条件或缺乏信息等因素导致的结果,但样本含量太小所得资料不稳定,结论缺乏充分的根据。有些论文中样本含量 ≤ 3 ,当指标的变异度较大时,很难呈现出资料之间的内在规律性^[6]。

2. 忽视统计分布的条件

这类错误的发生与作者的数学基础有关。由于这类错误不是与统计数据及统计方法等“表面”因素相联系,因而不容易被发现。例如,如果总体 $X \sim N(0, \sigma^2)$, (X_1, X_2) 是取自总体的一个样本,考察统计量 $\frac{(X_1 + X_2)^2}{(X_1 - X_2)^2}$ 的分布规律。有的解法如下:

因为 $X_1 \sim N(0, \sigma^2)$, $X_2 \sim N(0, \sigma^2)$, 所以 $X_1 + X_2 \sim N(0, 2\sigma^2)$, $\frac{X_1 + X_2}{\sqrt{2}\sigma} \sim N(0, 1)$, $X_1 - X_2 \sim N(0, 2\sigma^2)$, $\frac{X_1 - X_2}{\sqrt{2}\sigma} \sim N(0, 1)$ 。因此 $\left(\frac{X_1 + X_2}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1)$, $\left(\frac{X_1 - X_2}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1)$, 根据 F 分布的定义,

$$\text{有: } \frac{\frac{(X_1 + X_2)^2}{2\sigma^2} / 1}{\frac{(X_1 - X_2)^2}{2\sigma^2} / 1} \sim F(1,1), \text{ 即 } Y \sim F(1,1).$$

上述解法的错误在于没有考虑 F 分布的定义中的要求,即分子随机变量与分母随机变量要相互独立。本例是要求 $\left(\frac{X_1 + X_2}{\sqrt{2}\sigma}\right)^2$ 与 $\left(\frac{X_1 - X_2}{\sqrt{2}\sigma}\right)^2$ 相互独立。尽管求解的结果是正确的,但该解法由于缺乏对定义条件的说明,因而是不完整的,结论的获得带有一定的偶然性。正确的做法应该补充说明 $\left(\frac{X_1 + X_2}{\sqrt{2}\sigma}\right)^2$ 与 $\left(\frac{X_1 - X_2}{\sqrt{2}\sigma}\right)^2$ 相互独立。正确解法为:

$$\text{设 } \xi = \frac{X_1 + X_2}{\sqrt{2}\sigma}, \eta = \frac{X_1 - X_2}{\sqrt{2}\sigma}, \text{ 由于 } E(\xi) = 0, E(\eta) = 0 \text{ 所以}$$

$$\text{有 } \text{Cov}(\xi, \eta) = E[\xi - E(\xi)][\eta - E(\eta)] = E(\xi\eta) = \frac{1}{2\sigma^2}E(X_1 + X_2)(X_1 - X_2) = \frac{1}{2\sigma^2}[E(X_1^2) + E(X_2X_1) - E(X_1X_2) - E(X_2^2)]$$

因为 $X_1 \sim N(0, \sigma^2), X_2 \sim N(0, \sigma^2)$ 且 X_1, X_2 相互独立, 所以 $E(X_1^2) = E(X_2^2), E(X_1X_2) = E(X_1)E(X_2)$, 因此 $\text{Cov}(\xi, \eta) = 0$, 即 ξ, η 不相关, 又因为 $\xi \sim N(0, 1), \eta \sim N(0, 1)$, 由正态分布随机变量不相关即是相互独立知 ξ 与 η 独立。于是 ξ^2, η^2 也相互独立, 且 $\xi^2 \sim \chi^2(1), \eta^2 \sim \chi^2(1)$ 。于是有 $Y = \frac{\xi^2}{\eta^2} \sim F(1, 1)$ 。

3. 错用、乱用抽样分布

抽样分布是进行统计推断的基础和依据,也是进行科学研究和撰写某些科技论文的工具。但抽样分布涉及较多的数学知识,将

其准确的应用到实际问题中并不容易。常常出现的错误是：

(1) 忽视抽样分布中对样本条件的要求

抽样分布定理一般都对样本提出了一定的条件，样本是否满足这些条件是决定抽样分布应用正确与否的关键。但由于对这些条件的理解需要一定的高等数学基础，使得对抽样分布规律的应用常常出现混乱。例如，有些样本并非简单随机样本，或者所取自的总体根本不服从正态分布规律，但有些论文作者仍将此样本应用于具有上述属性要求的抽样推断，这样导出的结论显然没有任何说服力。也有的作者忽视样本个体间的独立性，盲目将其处理为简单随机样本。我们经常在一些论文中看到这样的字样：“设总体 X 服从正态分布”、“设总体服从指数分布”等等。其实在实际中，这样的假设很多情况下仅是作者的“一厢情愿”，除了一些理论上已经证明了的随机现象或者由统计规律解释过的结论外，要说明总体服从哪一种分布，必须提供一定的依据。

(2) 对定义的理解含混不清

有些自以为是的做法表现为对定义未能真正理解。现举一例说明。

例 已知正态总体 $X \sim N(\mu, \sigma^2)$ ， μ, σ 未知，在 X 中抽取简单随机样本 X_1, X_2, \dots, X_n ，求样本均值的分布。

典型的错误解法：根据 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 且 X_i 独立同分布于 $N(\mu, \sigma^2)$ 的条件，直接得出统计量 $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ 。事实上， $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ 含有未知参数 μ, σ ，所以不是统计量。统计量是指不含任何参数的样本的函数，由于 \bar{X} 正是 μ 的一个估计量，且统计量 σ^2 未知，因此在求 μ 的估计量 \bar{X} 的分布时，应该用样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 来代替 σ^2 。

正确解法: 因为 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, \frac{\sigma^2}{n})$, 所以 $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,$

1)。而 $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} (n-1) \sim \chi^2(n-1)$ 且

$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ 与 $\frac{(n-1)S^2}{\sigma^2}$ 相互独立。

(3) 对理论的掌握不足, 导致对各种抽样分布的使用不准确

根据统计学原理, 在进行统计推断时, 各种抽样分布的应用都必须根据问题的类型和条件的不同而有所区别。一般情况下, 在 μ 和 σ^2 均未知的条件下正态总体方差的区间估计用 χ^2 分布; 两个正态总体方差比的区间估计用 F 分布(因为方差比可以类比于两个 χ^2 统计量的比, 故用 F 分布)。而在生物学和医学的相关领域, 所涉及的统计推断的抽样分布的应用, 情况更加复杂, 抽样分布使用不准确的现象, 经常见诸文端, 应该引起足够的重视。

二、参数估计的典型错误分析

参数估计中的常见错误主要是方法上的。由于涉及较复杂的数学运算, 所以有时对此类错误的判断并不容易。下面举一例说明。

例 假设 X_1, X_2, \dots, X_n 是独立同分布的, 其概率密度为

$$f(x_i; \theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_2} e^{-\frac{x_i - \theta_1}{\theta_2}}, & \theta_1 \leq x \leq +\infty, -\infty \leq \theta_1 \leq +\infty, \\ 0 \leq \theta_2 \leq +\infty \\ 0, & \text{其他} \end{cases}$$

求 θ_1 与 θ_2 的极大似然估量。

错解: $L(x_1, x_2, \dots, x_n; \theta_1, \theta_2) = \prod_{i=1}^n \frac{1}{\theta_2} e^{-\frac{x_i - \theta_1}{\theta_2}} = \frac{1}{\theta_2^n} e^{-\sum_{i=1}^n \frac{x_i - \theta_1}{\theta_2}}$, 两

边取对数, 有

$$\ln L(x_1, x_2, \dots, x_n; \theta_1, \theta_2) = -n \ln \theta_2 - \sum_{i=1}^n \frac{x_i - \theta_1}{\theta_2}$$

$$\text{分别对 } \theta_1, \theta_2 \text{ 求导数, 得} \begin{cases} \frac{\partial \ln L}{\partial \theta_1} = \frac{n}{\theta_2} \\ \frac{\partial \ln L}{\partial \theta_2} = -\frac{n}{\theta_2} + \frac{\sum_{i=1}^n (x_i - \theta_1)}{\theta_2^2} \end{cases}$$

由 $\frac{\partial \ln L}{\partial \theta_2} = 0$ 得, $\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \theta_1)$ 。由于 $\frac{\partial \ln L}{\partial \theta_1} = \frac{n}{\theta_2} > 0$, 所以 θ_2 的极大似然估计为 $\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \theta_1)$ 。

分析: 首先 $\frac{\partial \ln L}{\partial \theta_1} = \frac{n}{\theta_2} > 0$ 不能说明 θ_1 的极大似然估计不存在,

其次, 若 θ_1 的估计不存在, 则 θ_2 的极大似然估计 $\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \theta_1)$ 的 θ_1 又从何而来?

正确解法: 由于 $\frac{\partial \ln L}{\partial \theta_1} = \frac{n}{\theta_2} > 0$, 所以 $L(x_1, x_2, \dots, x_n; \theta_1, \theta_2)$ 是关于 θ_1 的单调上升函数, 当 θ_1 最大时, $L(x_1, x_2, \dots, x_n; \theta_1, \theta_2)$ 达到最大, 又 $\theta_1 \leq x$, 因此有 $\theta_1 \leq \min_{1 \leq i \leq n} x_i$ 。所以当 $\theta_1 = \min_{1 \leq i \leq n} x_i$ 时, 似然函数达到最大值。所以 θ_1 的极大似然估计为: $\hat{\theta}_1 = \min_{1 \leq i \leq n} x_i$ 。

再令 $\frac{\partial \ln L}{\partial \theta_2} = 0$, 得 θ_2 的极大似然估计为: $\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n x_i - \hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n x_i - \min_{1 \leq i \leq n} x_i$ 。

三、假设检验中的常见错误分析

统计推断中, 假设检验方面的错误是最普遍和最常见的。而假

设检验本身较强的理论性是造成对其理解和应用产生困难的直接原因。许多研究者由于在原理上没有把握假设检验的本质,在统计推断中,照搬假设检验的程序,结果导致了各种错误的发生。下面根据假设检验在各领域的应用中出现的错误形式进行分类分析。

(一) 未能正确区分使用“双侧检验”还是“单侧检验”

错误地使用“双侧检验”或“单侧检验”,会影响结论的可靠性,甚至导致不正确的结论。对于其严重后果,我们通过具体的例子来说明。

例1 下面列出的是某工厂随机选取20只部件的装配时间(单位:min)

9.8, 10.4, 10.6, 9.6, 9.7, 9.9, 10.9, 11.1, 9.6, 10.2,
10.3, 9.6, 9.9, 11.2, 10.6, 9.8, 10.5, 10.1, 10.5, 9.7

设装配时间总体服从正态分布,能否认为装配时间均值显著大于10? ($\alpha = 0.05$)^[6]

错解:依题意提出检验问题: $H_0: \mu = 10; H_1: \mu > 10$ 。

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = 10.2, s = \sqrt{\frac{1}{19} \sum_{i=1}^{20} (x_i - \bar{x})^2} = 0.51,$$

$$t = \frac{\bar{x} - \mu_0}{s} \sqrt{n} = \frac{10.2 - 10}{0.51} \sqrt{20} = 1.75,$$

因 $\alpha = 0.05$,查表得 $t_{\frac{\alpha}{2}}(n-1) = t_{0.025}(19) = 2.093$,因为 $t = 1.75 < 2.093 = t_{0.025}(19)$,所以接受 H_0 ,认为装配时间的均值与10无显著差别。

分析:本题中备择假设 $H_1: \mu > 10$,故该问题实际上是右侧检验问题,其拒绝域的形式为 $t \geq t_{\alpha}(n-1)$ 。上述解法的错误在于把这个检验问题处理为双侧检验问题,从而认为拒绝域的形式为 $|t| \geq t_{\frac{\alpha}{2}}(n-1)$ 。

正确解法:依题意提出检验问题: $H_0: \mu = 10; H_1: \mu > 10$,计算得

$$\bar{x} = 10.2, s = 0.51, t = \frac{\bar{x} - \mu_0}{s} \sqrt{n} = 1.75$$

由 $\alpha = 0.05$, 查表得 $t_{\alpha}(n-1) = t_{0.05}(19) = 1.729$ 。因为 $t = 1.75 > 1.729 = t_{0.05}(19)$, 所以拒绝 H_0 , 认为装配时间的均值明显大于 10。

类似的错误在医学领域尤其多见。下面再举一例。

例2 某医生比较一新药与一旧药的疗效, 取随机样本 16 例, 随机分组设立新药组、旧药组各 8 例, 研究某指标结果见表 3.14, 问二药疗效有无差异?^[7]

表 3.14 新药组、旧药组某指标测定结果

新药组	5.2	9.3	7.8	6.9	4.2	8.1	10.6	8.1
旧药组	5.3	4.1	3.2	5.1	6.1	8.9	4.4	6.9

这是一项计算均数及进行显著性检验的临床试验。计算可得, 二组均数、标准差分别为 7.5, 2.1; 5.5, 1.8。统计量 $t = 2.05$, 自由度为 $8 + 8 - 2 = 14$ 。取单侧显著性临界值 $t_{0.05, 14} = 1.761$, 得单侧 $P < 0.05$, 可以认为二组差异明显, 新药疗效优于旧药。假如采用双侧检验, 则双侧显著性临界值 $t_{0.05, 14} = 2.145$, 得双侧 $P > 0.05$, 二组未见显著差异, 即二药疗效无差异。事实的结果显然只有一种, 之所以会有上面的两种不同结果, 是因为不同的检验方式所导致的。就是说, 在两种假设检验中, 必有一种是不恰当的。那么如何判断或确定究竟应该采用哪一种检验方式呢? 一般情况下, 选择单侧检验还是双侧检验首先应根据专业知识来确定, 同时也应考虑所要解决的问题本质或研究目的^[7]。如果是比较新法是否优于旧法, 则用单侧检验; 如果比较两种方法是否相同, 则需用双侧检验。因此本例宜用双侧检验。

(二) 未检验统计量使用的条件

统计量使用的盲目性源于对各种检验条件的理解不足或理解错误。

例3 某农业试验站为了研究某种新化肥对农作物产量的效力,在若干小区进行试验,测得产量见表3.15。

表 3.15 施肥与未施肥的产量结果(单位:kg)

施肥	34	35	32	33	34	30	
未施肥	29	27	32	31	28	32	31

设农作物产量服从正态分布,检验该种化肥对提高产量的效力是否显著($\alpha = 0.10$)。

错解:显然这是一个单侧检验问题,依题意提出检验假设:

$$H_0: \mu_1 \leq \mu_2; H_1: \mu_1 > \mu_2.$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} =$$

$$\frac{33 - 30}{\sqrt{\frac{5 \times 3.2 + 6 \times 4}{11}}} \sqrt{\frac{42}{13}} = 2.828$$

因 $\alpha = 0.10$,查表得 $t_{\alpha}(n_1 + n_2 - 2) = t_{0.10}(11) = 1.3634$ 。由 $t = 2.828 > t_{0.10}(11)$,故拒绝 H_0 ,认为该种化肥对提高产量的效力是显著的。

分析:在两个正态总体均值差的检验问题中,当两个总体方差(未知)相等时,才能使用统计量 $t = \frac{\bar{X} - \bar{Y} - \sigma}{s_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$,尽管上面解

法的结论正确,但由于对统计量的使用带有一定的盲目性,使得结论的获得带有很大的偶然性。本题首先要提出检验问题 $H_0: \sigma_1^2 = \sigma_2^2$ 。忽视“方差相等”是初学者常见的错误。

正确解法:设 X 为施肥后的产量, Y 为未施肥的产量。据题意, $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$,由于总体方差 σ_1^2, σ_2^2 未知,故先对方差进行检验。

$$H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2$$

$$\bar{x} = \frac{1}{6} \sum_{i=1}^6 x_i = 33, \bar{y} = \frac{1}{7} \sum_{i=1}^7 y_i = 30$$

$$s_1^2 = \frac{1}{5} \sum_{i=1}^6 (x_i - \bar{x})^2 = 3.2, s_2^2 = \frac{1}{6} \sum_{i=1}^7 (y_i - \bar{y})^2 = 4$$

$$F = \frac{s_1^2}{s_2^2} = \frac{3.2}{4} = 0.8$$

$\alpha = 0.10, n_1 = 6, n_2 = 7$ 。查表得 $F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0.05}(5, 6) = 4.95$ 。因为 $F = 0.8 < F_{0.05}(5, 6)$ ，所以接受 H_0 ，即认为 $\sigma_1^2 = \sigma_2^2$ 。然后提出检验问题

$$H'_0: \mu_1 \leq \mu_2; H'_1: \mu_1 > \mu_2 \text{ (以下解法与上面相同, 不再赘述)。}$$

对统计量的盲目使用所导致的错误往往与各种分布的误用交织在一起。上面例子的典型错误在医学杂志中比较常见。文献[8, 9]对《解放军医学杂志》2002年科研论文中科研资料假设检验的误用, 给出了比较系统的分析和释疑, 读者可以参照分析。

(三) t 检验的误用

t 检验是统计推断中的一种重要的假设检验方法。要实现对 t 检验的正确运用, 前提就是充分理解 t 检验的适用范围和应用条件。 t 检验的结果提示现有差别不仅仅是抽样误差所致, 且提示犯第一类错误的可能性大小, 它仅适用于正态或近似正态分布(包括偏态转换)和其方差是齐性资料的检验, 这里的方差齐性是指各组资料所取自的总体方差相等。在对均数进行假设检验时, t 检验仅用在单因素两水平设计(包括配对设计和成组设计)和单组设计(给出一组数据和一个标准值的资料)的定量资料的检验场合。在计量资料的统计分析中, 误用 t 检验代替方差分析是较为严重的错误。研究工作者过多地考虑实验设计, 统计分析过程中选用方法较简单的 t 检验, 常常混淆因素与水平之间的关系, 拆散多因素的整体设计, 使得资料的利用率低, 误差项的自由度变小, 结果的可靠性降低^[8]。下面列举几个 t 检验误用的实例, 这几个例子来

源于文献[10]。

例4 某研究对膀胱癌组与正常对照组 mRNA 定量 PCR 检测,结果差别有非常显著意义(t 检验, $t = 4.17, P < 0.01$ 检验), 见表 3.16。

表 3.16 两组 mRNA 定量 PCR 检测结果

组别	例数	mRNA 定量 PCR 读数($\bar{x} \pm s$)
膀胱癌组	12	0.40 \pm 0.31
对照组	24	0.05 \pm 0.16

分析: 本资料为完全随机设计的单因素两水平的资料, 表中对照组数据标准差比平均数大 2 倍多, 提示数据严重偏离了正态分布, 且两总体方差亦不相等($F = 3.75; F_{0.05(11, 23)} = 2.62, F > F_{0.05(11, 23)}, P < 0.05$), 不满足方差齐性的要求, 因此, 不能使用完全随机设计 t 检验。

解决办法: 一是对资料进行相应的变量变换, 能满足 t 检验的前提条件后, 再使用 t 检验; 二是选用适合分析偏态分布的非参数检验进行分析。

例5 某研究观察茶色素对慢性苯中毒的疗效。中毒者 2 例, 采用用药前后自身对照方法, 观察血液流变学情况; 统计学处理作 t 检验, 结果服药前后对比差异有显著性($P < 0.01$), 见表 3.17。

表 3.17 服茶色素前后血液流变学比较

检查项目	用药前	用药后	t 值
高切值	2.3 \pm 0.54	4.3 \pm 0.35	17.54
中切值	3.4 \pm 0.23	5.0 \pm 0.41	19.28
低切值	5.2 \pm 0.37	8.0 \pm 0.46	26.73
红细胞聚集指数	1.5 \pm 0.02	2.1 \pm 0.10	13.42

分析: 本例用完全随机 t 检验(根据 t 值结果)与原实验设计不符, 不宜使用。原设计为自身对照配对设计, 是最简单的重复测量设计, 这种比较应采用配对 t 检验。自身配对 t 检验与完全随机 t

检验的不同之处在于:①两者设计类型不同;②观察的内容也不同,完全随机 t 检验是比较两样本均数有无差异,而自身配对 t 检验是在同一个体上进行实验,以观察每个实验单位处理前后(或两种处理方法)有无差异;③配对 t 检验能最大限度地减少个体变异对处理间变异的可能影响,提高检验效能。

例6 某研究探究汞作业工人慢性过量汞接触的早期敏感指标。检测汞作业工人的血清、尿 N-乙酰- β -D 氨基葡萄糖苷酶(NAG)的总活性及尿 β_2 -微球蛋白(β_2 -MG)、血肌酐(Cr)等项目;汞作业工人工龄不足4年者为 Hg-1 组($n = 61$),工龄满4年者为 Hg-2 组($n = 104$),选择无汞接触史的机关工作人员 98 人为对照组,年龄、性别与 Hg 组相匹配。统计方法为两组间比较采用 t 检验,并对 3 组作了两两比较,结果见表 3.18。

表 3.18 各组血清、尿 NAG 总活性、尿 -MG 和血肌酐水平

组别	血清 NAG /(U · L ⁻¹)	尿 NAG /(U · L ⁻¹)	尿 NMG /(U · L ⁻¹)	Cr /(μ mol · L ⁻¹)
Hg-1 组	6.15 ± 0.18*	9.91 ± 1.26	62.66 ± 26.45	36.2 ± 19.4
Hg-2 组	6.75 ± 0.54*	13.08 ± 2.07*	95.12 ± 37.23**	37.1 ± 16.8
对照组	5.10 ± 0.27	11.05 ± 1.12	69.46 ± 22.85	37.1 ± 13.3

同对照组比较有差异: * $P < 0.05$; ** $P < 0.01$ 。

分析: 本例原设计为条件相近者的配对 t 检验,但配对设计是要求各组间的实验对象要一一配对,用对子差值计算结果,而本例各组人数均不相等,无法配对求差值,不能用配对 t 检验。本例也不宜用完全随机 t 检验。因为完全随机 t 检验只能分析单因素 2 水平的资料,本例的设计为单因素 3 水平,对于单因素 $k(k \geq 3)$ 水平设计的定量资料,如果满足正态性和方差齐性两个条件,应使用单因素 3 水平设计资料的方差分析方法,否则可使用非参数检验方法。用 t 检验代替方差分析,会增大犯第一类错误的概率,本例有 3 个均数,即有 3 个对比组,如果每次比较的显著性水平 $\alpha = 0.05$,

则每次不犯第一类错误的概率为 $1 - 0.05$, 3次都不犯第一类错误的概率为 $(1 - 0.05)^3$, 这时犯第一类错误的概率为 $1 - (1 - 0.05)^3 = 0.14$, 远大于 0.05 。

(四) χ^2 检验的误用

χ^2 检验是对样本的频数分布所来自的总体分布是否服从某种理论分布或某种假设分布所作的假设检验。前已述及 χ^2 检验的原理, χ^2 检验也是理论性相对较强的假设检验方法。类似于 t 检验, 在统计推断中, χ^2 检验在应用中出现的错误也相对多见。为了方便读者理解和辨识科技期刊论文中应用 χ^2 检验出现的各种错误, 现将 χ^2 检验的使用范围和应用中的注意事项进行归纳。

(1) 使用范围: χ^2 检验决定一组实际观察得到的频数与有关总体的理论频数是否一致的问题; 或者决定实际观察得到的频数与假设总体分布是真实时的期望频数是否一致的问题。所以, χ^2 检验属于自由分布的非参数检验。 χ^2 检验可以处理一个因素分为多种类别, 或多种因素各有多种类别的各种资料的非参数检验。各种比率检验也可用 χ^2 检验。

(2) 注意事项: 计算 χ^2 值时, 必须用绝对数, 而不能用相对数, 因为 χ^2 值的大小与频数大小有关。当自由度为 1 时, 需考虑进行连续性校正, 当求出的 χ^2 值略大于 3.84 时, 校正更为必要。因为往往会改变原来的结论。多组资料进行 χ^2 检验时, 如理论频数小于 5 的个数占全部理论频数个数的五分之一以上时, 要进行适当的并组, 但应注意, 要并得合理, 若比较几个等级资料相差是否显著时, 用 Ridit 分析法或秩和检验为好。 χ^2 检验的结论, 只提示两样本从同一总体中抽得的概率。在医学方面的应用中, 当两种疗法的病死率相差显著时, 并不证明某疗法疗效差, 只是说两样本从同一总体抽得的概率小, 因此可能两样本来自不同的总体, 差别有意义。当相差不显著时, 并不证明两疗法疗效相同, 只是说两样本来自同一总体的可能性大, 因此, 做出有判别的结论, 数据不足, 至于

有无差别的理由,要从专业上加以探讨:差别的大小,要根据医学上的实际意义来评论。再就是进行 χ^2 检验的资料应合乎随机抽样原则,且各组除处理因素不同外,其他条件应尽量控制其相同。下结论时,不能绝对化,因为显著性界限是人为规定的、相对的。即使是根据 $P < 0.01$ 做出的结论,仍有 1% 错误的可能。当检验结果与显著性界限很接近时,可加大样本继续观察,不要急于下结论。

下面根据 χ^2 检验误用的不同表现形式,给出几个实例。这些实例及分析来源于文献[11]。

例 7 某研究对海上石油运输船中船员的健康状况调查,症状与体征体查结果见表 3.19。

表 3.19 症状与体征体查结果

症状或体征	调查组($n = 219$)		对照组($n = 114$)	
	检出人数	患病率/%	检出人数	患病率/%
神经症候群	46	21.00	29	25.44
胃肠道症	33	15.07**	5	4.39
四肢关节酸痛	14	6.39	4	3.51
高血压	12	5.48*	1	0.88

χ^2 检验: * $P < 0.05$, ** $P < 0.01$ 。

分析:表中为 4 个项目的 2 组患病率比较,均直接使用四格表资料的 χ^2 检验(适用于:当样本含量 $n \geq 40$,且理论频数均 ≥ 5 时),但对于高血压项目不适用,因为对照组高血压项的理论频数为 4.45,不能直接用 χ^2 检验。应该用连续校正的 χ^2 检验。为便于比较,现把“高血压”项目体检结果列成四格表形式分析(见表 3.20)。

表 3.20 两组高血压患病率比较

组别	高血压	血压正常	合计	患病率/%
调查组	12(8.55)	207(210.45)	219	5.48
对照组	1(4.45)	113(109.55)	114	0.88
合计	13	5.48*	333	3.09

括号内数据为理论频数。

由表3.20,用连续校正 χ^2 检验,结果: $\chi^2 = 3.10, P > 0.05$,与原结论正好相反。可见, χ^2 检验不当应用会得出错误的结论。

例8 某研究对膀胱癌组与正常对照组的人巨细胞病毒(HCMV)阳性检出率进行比较,结果见表3.21。

表3.21 两组的HCMV阳性检出结果

组别	例数	HCMV 阳性人数 / %
膀胱癌组	12	8(66.7) *
对照组	24	0(0)

* χ^2 检验: $\chi^2 = 19.29, P < 0.01$ 。

分析:例8为对2个样本率的比较,原文用四格表资料的 χ^2 检验,但本例的总样本含量为36,不满足四格表资料 χ^2 检验的条件,而适用的是Fisher精确概率检验(当总样本含量 $N < 40$,或有理论频数 < 5 时)。Fisher检验结果为 $P = 0.000016$,这与原结论一致,但原来所用的假设检验方法不当,而且 χ^2 值也应为20.57,或校正后的 χ^2 值为16.90。

例9 某研究对60例胃癌患者不同区域癌基因蛋白强度比较,结果见表3.22。

表3.22 不同区域C-met的癌基因蛋白强度级别分布

胃癌区域	-	+	++	+++	χ^2	P值
肿瘤区	18	12	14	16		
移行区	42	10	5	3	4.15	> 0.05
癌旁正常黏膜区	50	6	4	0	6.22	< 0.05

分析:本例中每位患者同时有3个检测值,不能用表3.22的形式表示。表3.22适用于单纯随机分组设计,即观察相互独立的180例患者的数据,这与研究的内容不符,而且原文中所用的两两比较的 χ^2 检验也不合适。本组资料属于配伍设计,只有同一观察对象的3个区域才有可比性;因此,本资料适合用随机组设计资料的秩和检验,当结果差别有统计意义时再进行组间的两两比较。

以上列举的几个误用 χ^2 检验的实例,只在医学研究领域的一个侧面反映了 χ^2 检验方法在应用时出现的问题,事实上 χ^2 检验误用的表现形式十分复杂,在不同学科或不同的研究领域中, χ^2 检验在应用时都与所研究的问题的形式有直接的联系,由于涉及过多的专业知识和较深的数学要求,在此不再一一列举,有兴趣的读者可以参阅相关的书籍。