生物统计基础知识(II)

刘来福

(北京师范大学数学系)

二、两个样本平均数的比较

在一次饲料实验中，测定了饲以不同等级蛋白质饲料的两组大白鼠八周后的增重量。一组白鼠饲以高蛋白，平均增重量为 120 克;另一组白鼠饲以低蛋白，平均增重量为 101 克。这

是否就可以断言饲料的不同会影响到白鼠体重的增加呢？马上下这个结论还为时过早。因为影响白鼠体重增长的原因，除了饲料不同之外，还有许多其它的因素。如白鼠个体的差异，同种饲料内的差异，进食量的不同，对环境反应的不同等等，这些都是随机因素。在这些因素的影响下，也会使这些白鼠的增重量各不相同。因此上述两组白鼠平均增重之间的差异(19克)，肯定是有随机因素的作用。这就产生一个问题：对饲以不同饲料的两组白鼠来说，19克的差异中所反映的不同的饲料所引起的那部分差异是否明显地高于由随机因素所引起的那部分差异？从理论上讲，饲料的不同当然会影响到白鼠的增重。问题是在我们所考虑的试验中由随机因素产生的误差有多大？如果随机误差很大，以致于饲料不同所产生的差异还没有随机误差大，那么我们就没有理由认为这个试验所得到的差异是由不同饲料所产生的结果。如果在试验中随机误差很小，不同饲料所引起的增重量的差异要比随机误差大得多，当然我们就有充分的把握断言所得到的差异主要是由不同的饲料所引起的。这也就是比较两个样本平均数所要解决的问题。

(一)假设检验及差异的显著性

前一讲最后提到的X射线照射雄鼠的例子实际上就是要检验这14只雄鼠在照射前后平均体重的差异是否显著。

如果鼠的体重服从正态分布，那么照射前后体重的变化仍然服从正态分布，14个观测值x_i就是这个总体中的一个样本。

假定照射前后雄鼠的体重没有发生变化，也就是说上述正态总体的平均值应是零。这时$t=\bar{x}/s_{\bar{x}}$将服从自由度为$n-1$的t分布。由前面的讨论我们知道，对于数$\alpha>0$，我们能从t值表上找到这样一个t_α值，使得$|t|>t_\alpha$的概率是α。也就是说使得$|\bar{x}|>t_\alpha s_{\bar{x}}$的概率为$\alpha$。如果我们把$\alpha$取得足够小(如取$\alpha=0.05$或$\alpha=0.01$)。一般来说，差异的平均数$|\bar{x}|$应该不超过$t_\alpha s_{\bar{x}}$。如果$|\bar{x}|$比较大，而且$|\bar{x}|>t_\alpha s_{\bar{x}}$。那么这个平均差异很难由随机因素的影响来解释，只能怀疑前面关于照射前后体重没有变化的假设的正确性。应该认为照射前后体重确实发生了变化。这就是对前面的假设进行检验的过程。

在这个例子中，已经算得$\bar{x}=1.15$克，$s_{\bar{x}}=0.2451$克，自由度$d.f.=14-1=13$，由t值表查得$t_{0.01}=3.0123$。有$t_{0.01}s_{\bar{x}}=0.74$，$|\bar{x}|=1.15>0.74=t_{0.01}s_{\bar{x}}$。因此我们就拒绝关于照射前后体重没有变化的假设而判断：照射前后体重确实有变化。

当然，在拒绝接受这个假设时也可能做出错误的判断，即原来的假设是正确的但被错误地判断为照射前后体重有明显的变化。这种错误的可能性不会超过α。因此α就可做为判断差异是否存在的一个标准，称之为显著性水准。显然，α越小，拒绝接受原来假设的把握就越大，因此差异的显著性水准就越高。

在生物学上通常定为

当$|t|\leqslant t_{0.05}$时称为差异不显著；

当$t_{0.05}<|t|\leqslant t_{0.01}$时称为差异显著；

当$t_{0.01}<|t|$时称为差异极显著。

为计算方便，有时也用$t_\alpha s_{\bar{x}}$做为区别差异显著性的标准，即

当差异$\leqslant t_{0.05}s_{\bar{x}}$时称为差异不显著；

当$t_{0.05}s_{\bar{x}}<$差异$\leqslant t_{0.01}s_{\bar{x}}$时称为差异显著；

当$t_{0.01}s_{\bar{x}}<$差异时称为差异极显著。

数$t_\alpha s_{\bar{x}}$以后将称为水准为α的最小显著差数，记为$(\mathrm{L.S.D.})_\alpha$。(Least significant difference)。

(二)不同情形下平均数差异显著性的检验

1. 两组含量相等的样本平均数的比较

例如　对雄小鸡做性激素的效应试验。把初生的鸡随机地分为两组，每组11只，饲养在同一笼内，做上标记以示区别。使一组接受睾丸激素(A)处理，另一组接受雄甾烯醇酮(C)处理。到第15天取它们的鸡冠分别称重。结果见表4。要检验这两种激素的效应是否不同，即检验$\bar{x}_1-\bar{x}_2=41$与0是否有明显的不同。这时样本平均数之差的标准差(即两组样本之差的标准误差)由下式计算

$$s_{\bar{x}_1-\bar{x}_2}=s\sqrt{\frac{2}{n}},$$

表 4 雏鸡(♂)性激素效应试验结果

激素	鸡数	鸡冠重量(毫克)	平均
$A(x_1)$	11	57 120 101 137 119 117 104 73 53 68 118	97($\bar{x}_1$)
$C(x_2)$	11	89 30 82 50 39 22 57 32 96 31 88	56($\bar{x}_2$)

其中，n 为每组样本的含量，

$$s=\sqrt{\frac{\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2+\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2\right]}{2(n-1)}}$$

$$=\{[(57-97)^2+\cdots+(118-97)^2+(89-56)^2+\cdots+(88-56)^2]/2\times10\}^{\frac{1}{2}}$$

$$=\sqrt{811}$$

代入上式得

$$s_{\bar{x}_1-\bar{x}_2}=\sqrt{811\times\frac{2}{11}}=\sqrt{147.5}$$
$$=12.14\ (\text{毫克})$$

自由度 $d.f.=2\times(n-1)=2\times(11-1)=20$。由 t 值表查得 $t_{0.05}=2.0860$，$t_{0.01}=2.8453$。由此得 $L.S.D.$ 值为

$$(L.S.D.)_{0.05}=t_{0.05}s_{\bar{x}_1-\bar{x}_2}=2.0860\times12.14=25.32,$$

$$(L.S.D.)_{0.01}=t_{0.01}s_{\bar{x}_1-\bar{x}_2}=2.8453\times12.14=34.54。$$

由于

$$|\bar{x}_1-\bar{x}_2|=41>34.54=(L.S.D.)_{0.01},$$

因此可以判断两种激素的差异极显著。

2. 两组含量不相等的样本平均数的比较

在实验中各组所用的动物数相等计算起来比较简单。可是在不少实验中要求两组动物的数目相等是困难的。例如从两批不同处理过的蛋孵化出的雏鸡在数目上总不会相等，又如在实验过程中也可能有动物发生意外的死亡。如果这两个样本含量相差不太悬殊，并且各自的方差相差也不太大(一般，当方差较小的样本在10个上下时，另一个样本的方差应低于前一样本的方差的三倍即可)。这时，在比较这样的两个样本平均数时，计算方法略加改变就可进行检验。举例如下：

对两组雌鼠分别给以高蛋白或低蛋白饲料，八周后观察各鼠所增的体重(见表5)。

表 5 雌鼠饲料试验体重的增重

饲　料	高蛋白(x_1)	低蛋白(x_2)	差值(d)
鼠　数	12(n_1)	7(n_2)	
各鼠所增的体重(克)	83	70	13
	146	118	28
	119	101	18
	104	85	19
	120	107	13
	161	132	29
	107	94	13
	134		
	115		
	129		
	99		
	123		
平　均	120($\bar{x}_1$)	101($\bar{x}_2$)	19($\bar{d}$)

可以算出两个样本的方差分别为

$$s_1^2=\frac{1}{n_1-1}\sum(x_1-\bar{x}_1)^2=\frac{1}{11}[(83-120)^2+\cdots+(123-120)^2]$$
$$=445.82,$$

$$s_2^2=\frac{1}{n_2-1}\sum(x_2-\bar{x}_2)^2=\frac{1}{6}[(70-101)^2+\cdots+(94-101)^2]$$
$$=425.33。$$

平均数差异的标准差为

$$s_{\bar{x}_1-\bar{x}_2}=s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

其中

$$s=\sqrt{[(n_1-1)s_1^2+(n_2-1)s_2^2]/(n_1+n_2-2)}$$
$$=\sqrt{[\sum(x_1-\bar{x}_1)^2+\sum(x_2-\bar{x}_2)^2]/(n_1+n_2-2)}$$
$$=\sqrt{[11\times445.82+6\times425.33]/17}$$
$$=\sqrt{438.59}=20.94$$

代入上式

$$s_{\bar{x}_1-\bar{x}_2}=20.94\times\sqrt{\frac{1}{12}+\frac{1}{7}}=9.96$$

这时，自由度 $d.f.=n_1+n_2-2=17$，由 t 值表查得 $t_{0.05}=2.1098$。由此得到 $L.S.D.$ 值为

$$(L.S.D.)_{0.05} = t_{0.05}s_{\bar{x}_1-\bar{x}_2} = 2.1098 \times 9.96 = 20.91。$$

由于

$$|\bar{x}_1 - \bar{x}_2| = 19 < 20.91 = (L.S.D.)_{0.05}$$

故此两种饲料对八周雌鼠体重的变化没有显著的影响。

3. 成对数据的比较

实验时，所处的环境条件愈相象，愈能表明不同处理间的真实效果。在上述的饲料实验中，虽然我们可以使得这两组鼠在体重等各方面的条件力求相似。但是仍然不能避免血统上以及起始体重等方面的差异。为了改进上面的试验设计，可以取同胎生而且体重相近的两只分别喂以不同的饲料。这样所得到的试验结果将有效地缩小血统及体重方面的差异。例如，在表2中所列的前七对数据如果是按上面所说的设计方式得到的数据。即可以用它们的差数 d 来进行分析现在要检验 $\bar{d}=19$ 与零是否有显著的不同。与本段开始时一样，差值 d 的标准差应是

$$s_d = \sqrt{\frac{1}{n-1}\sum(d-\bar{d})^2}$$

$$= \left\{\frac{1}{7-1}[(13-19)^2+(28-19)^2 + \cdots + (13-19)^2]\right\}^{\frac{1}{2}}$$

$$= \sqrt{\frac{1}{6}\times 290} = \sqrt{48.33} = 6.95$$

平均差值的标准差为

$$s_{\bar{d}} = s_d/\sqrt{n} = 6.95 \times \frac{1}{\sqrt{7}} = 2.63$$

自由度 $d.f. = n - 1 = 7 - 1 = 6$，由 t 值表查得 $t_{0.05} = 2.4469$，$t_{0.01} = 3.7074$。由此得 $L.S.D.$ 值为

$$(L.S.D.)_{0.05} = t_{0.05}s_{\bar{d}} = 2.4469 \times 2.63 = 6.44$$

$$(L.S.D.)_{0.01} = t_{0.01}s_{\bar{d}} = 3.7074 \times 2.63 = 9.75$$

由于

$$|\bar{d}| = 19 > 9.75 = (L.S.D.)_{0.01}$$

故可以断言，饲以高蛋白的雌鼠的增重，极显著地高于饲以低蛋白的雌鼠的增重。

由此可见，如果能观测到成对的数据，就可以大大提高试验及其结果的鉴别能力。因此在对实验材料的了解逐步深入的基础上，逐步提高试验的精度，把成组比较的试验安排成个体成对比较的试验，将会大大提高试验的效率。但是要注意的是数据是否配对，要依靠生物学的知识及试验设计的方案来判断，统计分析方法的选择必须依试验设计的方式而定。千万不能人为地随意将两组数据配成对进行分析。

4. 大样本平均数的比较

如果两个样本的含量都比较大(一般要求不少于30)，这时可以用

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

来计算平均数差异的标准差，其中 s_1，s_2 分别是两个样本的标准差，n_1，n_2 为两个样本的含量。$L.S.D.$ 值的计算与前面一样。

例如，测得100头某种牛的体高，得到 $\bar{x} = 133$ 厘米，$s_1 = 4.07$ 厘米，而120头另一种牛的体高为 $\bar{x}_2 = 131$ 厘米，$s_2 = 2.92$ 厘米。这两种牛的体高的差异是否显著？

按上面的 $s_{\bar{x}_1-\bar{x}_2}$ 的计算公式有

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{4.07^2}{100}+\frac{2.92^2}{120}} = 0.4865$$

由 t 值表(可将自由度取做∞)查得 $t_{0.05} = 1.9600$，$t_{0.01} = 2.5758$，于是有

$$(L.S.D.)_{0.05} = t_{0.05}s_{\bar{x}_1-\bar{x}_2} = 1.9600 \times 0.4865 = 0.95,$$

$$(L.S.D.)_{0.01} = t_{0.01}s_{\bar{x}_1-\bar{x}_2} = 2.5758 \times 0.4865 = 1.25。$$

由于

$$|\bar{x}_1 - \bar{x}_2| = 2 > 1.25 = (L.S.D.)_{0.01}。$$

因此这两种牛的体高差异极显著。

三、计数资料的检验——χ^2 检验

生物学上的数字资料大体可分做两大类。一类是由测量所得的记录，如体高，体重，耗氧量，时间等。观测的结果一般是连续变量，我们称为计量资料。另一类是由清点数目所得的记录，例如某个动物总体中雌、雄的只数，对某些

试验有无反应的动物的只数或者在实验结果中动物生存、死亡的只数等等。这类资料通常与离散变量有关，所计数的是定性的而非定量的性状。它们包含着分在各个分明的组类中的个体数目。

对于计数资料，前述的检验方法一般不适用。最常见的假设检验，是检验在各个组类中实际观测的频数与按照假设所计算出来的各个组类中应出现的理论频数是否是符合。为此需要引入一个新的统计量 χ^2。进行检验时，主要是要根据检验假设来计算理论频数，并且依据 χ^2 统计量进行推断。如果资料是单向分类的，这时检验的目的一般是需要检验观测频数与在某个假设下的理论频数是否符合。这时我们称为适合性检验。如果资料为两向分类，这时往往是要来判断这两个方向的因子之间的作用是否独立。因此称之为独立性检验。举例说明如下：

（一）适合性检验

在某试验中，对由若干母豚鼠所产的幼豚鼠的性别做了调察，查得 94 只 (a_1) 为雌性，80 只 (a_2) 为雄性。而豚鼠在一般情况下，性比为 1:1，问这个试验是否影响到所产的幼豚鼠的性比。

我们假设试验没有影响到所产的幼豚鼠的性比。也就是说，这些母豚鼠所产的幼豚鼠的性比应为 1:1。在这 94+80=174 只幼豚鼠中应该有 87 只 (t_1) 雌性和 87 只 (t_2) 雄性。这就是这两组的理论频数。

为反映观测频数与理论频数的差异，我们用 $(a_i-t_i)^2/t_i$ 表示第 i 组中的观测频数 a_i 与理论频数 t_i 的相对误差。那么在 n 个组中观测频数与理论频数总的相对误差可以用

$$\chi^2=\sum_{i=1}^{n}(a_i-t_i)^2/t_i。$$

显然它是一个统计量，而且 χ^2 值的大小可以反映观测频数与理论频数相差的程度。理论上的分析表明，它将服从自由度为 $n-1$ 的 χ^2 分布。对于给定的 α，在不同的自由度下 χ^2 的临界值 χ^2_α 将由 χ^2 表给出（表 6）

表 6 χ^2 值表

自由度 \ α	0.50	0.30	0.20	0.10	0.05	0.02	0.01
1	0.455	1.074	1.642	2.706	3.841	5.412	6.635
2	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	2.366	3.665	4.642	6.251	7.815	9.837	11.345
4	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	4.351	6.064	7.289	9.236	11.010	13.388	15.086
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

如果由观测资料算得的 χ^2 越小，就表明观测的数值与理论频数越接近。这时就可以接受原来的假设。如果它比较大，而且大于显著性水准为 α 时的临界值 χ^2_α，那么我们将在这个水准上否定原来的假设。

在我们的问题中有

$$\chi^2=\frac{(a_1-t_1)^2}{t_1}+\frac{(a_2-t_2)^2}{t_2}$$
$$=\frac{(94-87)^2}{87}+\frac{(80-87)^2}{87}$$
$$=0.5632+0.5632=1.1264$$

自由度 $d.f.=n-1=2-1=1$。由 χ^2 表查得 $\chi^2_{0.05}=3.841$。然而

$$\chi^2=1.1264<3.841=\chi^2_{0.05}。$$

因此我们接受原来的假设，认为雌豚鼠所产的幼豚鼠的性比仍为 1:1。

（二）独立性检验

有人统计了小家鼠到两种诱捕器（以同样食物为诱饵）去的频数。一号诱捕器沾了以往小家鼠的尿，二号则是清洁的（见表 7）。

表 7 小家鼠到两种诱捕器的频数

结果 \ 诱捕器	一　号	二　号	总　计
进入诱捕器	17(a_{11})	4(a_{21})	21(C_1)
不进诱捕器	9(a_{12})	24(a_{22})	33(C_2)
总　计	26(R_1)	28(R_2)	54(G)

这两种不同的诱捕器的诱捕能力是否相同？

这里所得到的资料就是以诱捕器的不同以及是否进入诱捕器这两个因素的不同分类而得到的 $2\times2=4$（组）计数资料。问题实际上是

要我们分析不同的诱捕器会不会影响到小家鼠进入诱捕器的可能性。也就是说要检验诱捕器的清洁与否是不是与小家鼠进入诱捕器是独立的。因此它是独立性检验的问题。

假设诱捕器的不同并不影响小家鼠进入诱捕器。这时小家鼠进入诱捕器的频率应该是：$p = C_1/G = 21/54 = 0.3889$，不进入诱捕器的频率为：$q = C_2/G = 33/54 = 0.6111$。于是走到一号诱捕器前的 $R_1 = 26$ 只鼠中，应该有 $t_{11} = R_1C_1/G = 26 \times 0.3889 = 10.1$ 只进入了诱捕器，有 $t_{12} = R_1C_2/G = 26 \times 0.6111 = 15.9$ 只不进去。在走到二号诱捕器前的 $R_2 = 28$ 只鼠中，应该有 $t_{21} = R_2C_1/G = 10.89$ 只进入了诱捕器，而有 $t_{22} = R_2C_2/G = 28 \times 0.6111 = 17.11$ 只不进去。

这时，在前面关于独立性的假设成立的前提下，计算出来的理论频数为 t_{ij}，而观测频数是 a_{ij}。于是 χ^2 统计量

$$\chi^2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{(a_{ij} - t_{ij})^2}{t_{ij}}$$

将服从自由度为 $d.f. = (n_1 - 1)(n_2 - 1)$ 的 χ^2 分布，其临界值仍由表 3 给出。

在这个问题中有

$$\chi^2 = \frac{(17 - 10.1)^2}{10.1} + \frac{(9 - 15.9)^2}{15.9} + \frac{(4 - 10.89)^2}{10.89} + \frac{(24 - 17.11)^2}{17.11}$$
$$= 4.71 + 2.99 + 4.36 + 2.78$$
$$= 14.84$$

自由度 $d.f. = (n_1 - 1)(n_2 - 1) = (2 - 1) \times (2 - 1) = 1$，由表 4 查出 $\chi^2_{0.05} = 3.841$，$\chi^2_{0.01} = 6.635$。有

$$\chi^2 = 14.84 > 6.635 = \chi^2_{0.01}$$

因此否定原来的假设，有 99% 的把握认为小家鼠进入到不同诱捕器的概率是不一样的。