生物统计基础知识（IV）

刘 来 福

（北京师范大学数学系）

五、方 差 分 析

第 II 讲介绍了两组样本平均数比较时进行显著性检验的方法。如果相互比较的不是两个组，而是三个，四个或更多个组的时候，就需要把前面所用的方法加以推广。适用于比较多组平均数之间差异的检验方法称之为方差分析。所谓方差是指离均差的平方和（简称平方和）与自由度之商。分析时主要是把平方和与自由度按不同的变异起因分解为若干个部分。从而构成来自不同起因的方差，利用它来检验各组平均数之间差异的显著性。

方差分析是分析和处理观测数据的主要方法之一，应用范围甚广。对于从不同试验设计中得出的观测资料，进行方差分析时将有不同的计算方法。但是，其基本思想却大同小异。这里仅结合单因素试验中的几个试验设计加以介绍，以便对这个方法的基本思想有所了解。至于适用于不同试验设计的方差分析方法可参考有关的生物统计书籍。

所谓单因素试验是指在试验中所考察的因素只有一个。例如对动物所进行的饲料试验，药效试验等等。至于其它的影响因素，如动物在性别，年龄（或日龄），体重等方面的差异，季节的差异，环境的差异等等都应尽量控制在可能最小的变异范围之内。专门考察一个因素变化时所产生的影响。因素在试验中所分的等级叫做水平，而同一个试验条件下的一次试验叫做一个处理。

下面进行方差分析时，我们将假定观测数据总是来自若干个正态总体，各数据所受随机误差影响的方差是相同的。

（一）完全随机设计 所谓完全随机设计是指把一批试验对象完全混合，然后分成若干组。再把各种不同的处理随机地安置在这些组上。每一种处理的各个试验结果就可以理解为这种处理的一个随机样本。

例如 把四种不同的饲料分别喂给 4 组小鸡，每组 5 只，它们的增重量列表于12。试对所得的结果进行分析。

我们把试验中所考察的因素（饲料）称为因素 A。用 x_{ij} 表示试验中所得到的各个观测值。

表 12 饲料试验中小鸡的增重

组别	增重 (x_{ij})					$X_{i\cdot}$	$\bar{x}_{i\cdot}$
1	55	49	42	21	52	219	43.8
2	61	112	30	89	63	355	71.0
3	42	97	81	95	92	407	81.4
4	169	137	168	85	154	713	142.6

第一个指标 i 表示因素A取第 i 个水平（相当于例题中的第 i 组），而第二个指标 j 表示每组中的第 j 个观测值。 例如 $x_{12}=49$，$x_{35}=92$ 等。 $p=4$ 为因素A的水平数，$r=5$ 为每组内的个体数。用 $X_{i\cdot}$ 表示第 i 组观测数据的总和 $X_{i\cdot}=\sum_{j=1}^{r}x_{ij}$（如 $X_{1\cdot}=55+49+\cdots+52=219$）。用 $\bar{x}_{i\cdot}$ 表示第 i 组观测数据的平均数 $\bar{x}_i=\frac{1}{r}X_{i\cdot}=\frac{1}{r}\sum_{j=1}^{r}x_{ij}$（如 $\bar{x}_{1\cdot}=219/5=43.8$）。 $X_{\cdot\cdot}$ 表示所有 x_{ij} 的总和 $X_{\cdot\cdot}=\sum_{i=1}^{p}\sum_{j=1}^{r}x_{ij}=55+49+\cdots+154=1694$。

$\bar{x}_{\cdot\cdot}=\frac{1}{pr}X_{\cdot\cdot}=\frac{1}{4\times5}1694=84.7$ 是总平均。

$X_{i\cdot}$，$\bar{x}_{i\cdot}$ 的计算结果列于表 12 的最后两列。

1. 平方和的分解

整个试验结果的差异可以用全部观测数据的平方和来刻划，这个平方和我们称之为总平方和，记作 S_T。 于是有

$$S_T^*=\sum_{i=1}^{p}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{\cdot\cdot})^2=(55-84.7)^2+(49-84.7)^2+\cdots+(154-84.7)^2=37626.2。$$

这个平方和还可以用另外的公式来计算，即

$$S_T=\sum_{i=1}^{p}\sum_{j=1}^{r}x_{ij}^2-\frac{1}{pr}X_{\cdot\cdot}^2=(55^2+49^2+\cdots+154^2)-\frac{1}{4\times5}1694^2=37626.2。$$

不难看出，这个试验中产生上述差异的原因不外乎因素A的不同水平及随机误差两种。

$\sum_{i=1}^{p}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2$ 就表示各处理间平均每只小鸡所产生的差异的大小。然而在我们的试验中每个处理内都有 5 只小鸡。因此A的不同水平所引起的差异的大小应是

$$S_A=r\sum_{i=1}^{p}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2=5\times[(43.8-84.7)^2+(71-84.7)^2+\cdots+(142.6-84.7)^2]=26119。$$

还可以写成

$$S_A=\frac{1}{r}\sum_{i=1}^{p}X_{i\cdot}^2-\frac{1}{rp}X_{\cdot\cdot}^2=\frac{1}{5}(219^2+355^2+407^2+713^2)-\frac{1}{4\times5}1694^2=26119。$$

在每个组内部的数据间的差异 $\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i\cdot})^2$ 反映了每一组内随机误差的大小，因此整个试验中随机误差的平方和 S_e 就应该等于

$$S_e=\sum_{i=1}^{p}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i\cdot})^2=(55-43.8)^2+(49-43.8)^2+\cdots+(154-142.6)^2=11507.2。$$

还可以写成

$$S_e=\sum_{i=1}^{p}\sum_{j=1}^{r}x_{ij}^2-\frac{1}{r}\sum_{i=1}^{p}X_{i\cdot}^2=(55^2+49^2+\cdots+154^2)-\frac{1}{5}(219^2+\cdots+713^2)=11507.2。$$

为了计算方便，我们引入如下的记法

$$W=\sum_{i=1}^{p}\sum_{j=1}^{r}x_{ij},\quad P=\frac{1}{r}\sum_{i=1}^{p}X_{i\cdot}^2,$$

$$C=\frac{1}{pr}X_{\cdot\cdot}^2。$$

由前面的推导可知，有下面的平方和计算公式

$$S_T = W - C, \ S_A = P - C,$$

$$S_e = W - P = S_T - S_A。$$

这样，计算时只要把 W，P，C 算出来，就可以由上面的式子计算 S_T，S_A，S_e 的值。

2. 自由度的分解

由于方差等于平方和与自由度之商。在对平方和按不同变异原因进行分解之后，还要对自由度进行分解。

总的自由度应该是所有观测数据的个数减1，即 $d.f_T = pr - 1$。

因素A的自由度等于 $d.f_A = p - 1$。

在计算每个组内的误差时，可知每一组内误差的自由度为 $r - 1$。整个试验随机误差的自由度应等于每组内误差自由度的总和，$d.f_e = p(r-1)$。它恰好等于总自由度与因素A的自由度之差，即 $d.f_e = d.f_T - d.f_A$。

在我们的例子中有

$$d.f_T = 4 \times 5 - 1 = 19, \ d.f_A = 4 - 1 =$$

$$d.f_e = 19 - 3 = 16$$

3. 计算方差

因为参与计算平方和的数据的个数不同，直接用平方和进行比较不合理，因此要用方差来进行比较。按照平方和除以自由度的公式，可以计算出不同变异原因的方差

$$V_A = S_A/d.f_A = 26119/3 = 8706.3,$$

$$V_e = S_e/d.f_e = 11507/16 = 719.2。$$

把上面的计算结果列成方差分析表(见表13)。

表 13 小鸡饲料试验方差分析表

方差来源	平方和	自由度	均方	F	临界值
饲料 (A)	26119	3	8706.3	12.1**	$F_{0.05} = 3.2$
机误 (E)	11507.2	16	719.2		$F_{0.01} = 5.3$
总和	37626.2	19			

** 特别显著。

4. 显著性检验

方差是反映差异大小的一个指标。由上面的分析可知，V_A 中除包含有随机误差之外，还可能有饲料不同所引起的差异；而 V_e 全部都是随机误差。要检验在这个试验中饲料的不同对小鸡的增重是否有影响，实际上就是要检验 V_A 是否明显地高于 V_e。如果饲料不同对小鸡的增重不发生影响，那么 V_A 所反映的也应该是随机误差。这样 V_A 与 V_e 应该没有明显的不同。如果不同的饲料对小鸡的增重有影响，那么，V_A 就应该明显地要比 V_e 大。

为了检验饲料不同所引起的小鸡增重的差异的显著性，我们构成一个新的统计量 $F = V_A/V_e$。这样，检验饲料间方差的显著性的问题就变成检验 F 是否明显地大于 1 的问题了。

统计分析的理论表明，统计量 F 服从自由度为 $n_1 = d.f_A$，$n_2 = d.f_e$ 的 F 分布。当显著性水准为 0.05 和 0.01 时不同自由度下 F 的临界值由表 14 给出。

我们用 $F_\alpha(n_1, n_2)$ 表示当显著性水准为 α，相应的两个自由度为 n_1，n_2 时 F 的临界值。于是就可以查得 $F_{0.05}(3, 16) = 3.2$，$F_{0.01}(3, 16) = 5.3$。利用这个临界值就可以判断因素的影响是否是显著的。

如果 $F > F_{0.01}$，则认为因素A的影响特别显著，记为“＊＊”；如果 $F_{0.01} \geq F > F_{0.05}$，则认为因素A的影响显著，记为“＊”；如果 $F_{0.05} \geq F$，则不能认为因素A有显著的影响。

本例中的 $F = 12.1 > 5.3 = F_{0.01}$。故饲料不同对小鸡的增重影响极显著。

5. 平均数间的比较

前面求出的四种饲料所喂的小鸡平均增重试验，经方差分析可以看出，有非常显著的差异。其中最高的第四组与最低的第一组当然有明显的差别。但这四种饲料之间是否都有显著的差别呢？例如第一组与第二组，第一组与第三组平均数的差别怎样呢？这要做进一步的分析：对所得到的四个平均数逐对地进行比较。

表 14 *F* 表

n_2 \ n_1	1	2	3	4	5	6	12	24	∞
⋮									
11	4.8	4.0	3.6	3.4	3.2	3.1	2.8	2.6	2.4
	9.7	7.2	6.2	5.7	5.3	5.1	4.4	4.0	3.6
12	4.8	3.9	3.5	3.3	3.1	3.0	2.7	2.5	2.3
	9.3	6.9	6.0	5.4	5.1	4.8	4.2	3.8	3.4
13	4.7	3.8	3.4	3.2	3.0	2.9	2.6	2.4	2.2
	9.1	6.7	5.7	5.2	4.9	4.6	4.0	3.6	3.2
14	4.6	3.7	3.3	3.1	3.0	2.9	2.5	2.3	2.1
	8.9	6.5	5.6	5.0	4.7	4.5	3.8	3.4	3.0
15	4.5	3.7	3.3	3.1	2.9	2.8	2.5	2.3	2.1
	8.7	6.4	5.4	4.9	4.6	4.3	3.7	3.3	2.9
16	4.5	3.6	3.2	3.0	2.9	2.7	2.4	2.2	2.0
	8.5	6.2	5.3	4.8	4.4	4.2	3.6	3.2	2.8
⋮									

注：表中每一格内上面的数字为 $\alpha=0.05$ 的临界值，下面一个为 $\alpha=0.01$ 的临界值。

为了比较平均数之间差异的显著性，需要把各平均数俩俩之间的差异全部计算出来。为方便起见，我们可以把各平均数依大小次序排列，然后在表15的平均数差异表中计算出来。表15中差异的第一列数分别是 $\bar{x}_4$ 与其它三个平均数之差，$\bar{x}_4-\bar{x}_3=61.2$，$\bar{x}_4-\bar{x}_2=71.6$，$\bar{x}_4-\bar{x}_1=98.8$。而第二列为 $\bar{x}_3$ 与 $\bar{x}_2$，$\bar{x}_1$ 之差，第三列为 $\bar{x}_2-\bar{x}_1$。

表 15 平均数差异表

组别	平均数	差		异
第四组	142.6			
第三组	81.4	61.2*		
第二组	71	71.6**	10.4	
第一组	43.8	98.8**	37.6	27.2
		第四组	第三组	第二组

* 显著。 ** 特别显著。

多个平均数之间差异的显著性检验方法很多，这里只介绍一个比较简单的方法，即图基(Tukey)的方法(作了一些修正)。

类似于两个平均数差异的显著性检验，首先要计算这个试验中的标准误差（平均数的标准差)。它等于机误方差 V_e 与组内动物数 r 之商的平方根，即 $\sqrt{V_e/r}=\sqrt{719.2/5}=11.99$。这时，在水准为 α 时任两个平均数具有显著差异的临界值将由标准误差的若干倍给出，即

$$D_\alpha=q_\alpha(r,\ d.f_e)\sqrt{V_e/r},$$

式中 $q_\alpha(r,\ d.f_e)$ 是显著性水准为 α 时，对应于每组有 r 个个体且机误自由度为 $d.f_e$ 的一个统计量的临界值。它由 q 值表给出（见表 16)。例子中 $D_{0.05}=q_{0.05}(5,16)\sqrt{V_e/r}=4.33\times11.99=51.92$，$D_{0.01}=q_{0.01}(5,16)\sqrt{V_e/r}=5.49\times11.99=65.83$。与表 15 所列的平均数之差异比较可见 $\bar{x}_4$ 与 $\bar{x}_1$，$\bar{x}_2$ 差异极显著，$\bar{x}_4$ 与 $\bar{x}_3$ 差异显著，而其余的差异不显著。也就是说，试验中第四组饲料的增重显著地优于其它三组。

（二）样本含量不等的完全随机试验 前面介绍的单因素试验中，每一个处理内样本的含量都是相等的。但在动物试验中有时候很难使试验中每个处理的样本含量相等。

例如，有 8 头母猪产了八窝仔猪，每头仔猪出生重量如表 17 所示。

各窝出生仔猪的平均体重是否有显著的差异？

含量不等的单因素试验的分析与等含量的分析方法类似，仅仅是计算时略有不同。

由于各窝内动物的个数 r_i 不相同，因此进行平方和分解时计算公式稍有变化。如果记

表 16 q 值表

$d.f_e$	2	3	4	5	6	7	8	9	10	15
⋮										
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.88
	4.32	5.04	5.50	5.84	6.10	6.32	6.51	6.67	6.81	7.36
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.79
	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	7.19
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.71
	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	7.05
15	3.01	3.67	4.08	4.37	5.59	4.78	4.94	5.08	5.20	5.65
	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.93
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.59
⋮	4.13	4.78	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.82

注：每一格上下两个数字分别表示 $\alpha=0.05$，$\alpha=0.01$ 时的 q 值。

表 17 八窝猪的出生重量

窝号 i	数目 r_i	出生重量 x_{ij}										$X_{i\cdot}$
1	10	2.0	2.8	3.3	3.2	4.4	3.6	1.9	3.3	2.8	1.1	28.4
2	8	3.5	2.8	3.2	3.5	2.3	2.4	2.0	1.6			21.3
3	10	3.3	3.6	2.6	3.1	3.2	3.3	2.9	3.4	3.2	3.2	31.8
4	8	3.2	3.3	3.2	2.9	3.3	2.5	2.6	2.8			23.8
5	6	2.6	2.6	2.9	2.0	2.0	2.1					14.2
6	4	3.1	2.9	3.1	2.5							11.6
7	6	2.6	2.2	2.2	2.5	1.2	1.2					11.9
8	4	2.5	2.4	3.0	1.5							9.4
Σ	56											152.4

$$X_{\cdot\cdot}=\sum_{i=1}^{p}\sum_{j=1}^{r_i}x_{ij},\quad X_{i\cdot}=\sum_{j=1}^{r_i}x_{ij},$$

$$r=\sum_{i=1}^{p}r_i;$$

$$W=\sum_{i=1}^{p}\sum_{j=1}^{r_i}x_{ij}^2,\quad P=\sum_{i=1}^{p}\frac{1}{r_i}X_{i\cdot}^2,$$

$$C=\frac{1}{r}X_{\cdot\cdot}^2。$$

则，各平方和的计算公式仍为

$$S_T=W-C,\quad S_A=P-C,$$

$$S_e=S_T-S_A。$$

相应的自由度为

$$d.f_T=r-1,\quad d.f_A=p-1,$$

$$d.f_e=d.f_T-d.f_A=r-p。$$

在例子中有

$$r=56,\quad p=8,\quad X_{\cdot\cdot}=152.4$$

$$W=\sum_{i=1}^{8}\sum_{j=1}^{r_i}x_{ij}^2=2.0^2+2.8^2+\cdots+1.5^2=439.40,$$

$$P=\sum_{i=1}^{8}\frac{1}{r_i}X_{i\cdot}^2=\frac{1}{10}\times28.4^2+\frac{1}{8}\times21.3^2+\cdots+\frac{1}{4}\times9.4^2=422.23,$$

$$C=\frac{1}{r}X_{\cdot\cdot}^2=\frac{1}{56}\times152.4^2=414.75。$$

由此得平方和

$$S_T=W-C=439.40-414.75=24.65,$$

$$S_A=P-C=422.23-414.75=7.48,$$

$$S_e=S_T-S_A=24.65-7.48=17.17。$$

自由度为 $d.f_T=56-1=55$，$d.f_A=8-1=7$，$d.f_e=55-7=48$。这样就可列出这个问题的方差分析表（见表 18）。

表 18　仔猪出生重量方差分析表

方差来源	平方和	自由度	方　差	F	临界值
窝　间	7.48	7	1.07	2.97*	$F_{0.05}=2.21$
窝内(机误)	17.17	48	0.36		$F_{0.01}=3.04$
总　和	24.65	55			

* 显著。

由于算得的 $F=2.97>2.21=F_{0.05}(7,48)$，因此可以认为窝之间有显著的差异。

各窝平均出生重量的比较过程与前面也类似。只是在计算标准误差时，前面用的是机误方差与每个处理内个体数之商的平方根。这里，由于不同处理内的个体数不同了，因此需要给出这些个体数 r_i 的一个平均估值。统计理论的分析表明，正确的估值应由公式

$$r_0=\frac{1}{p-1}\left(r-\frac{1}{r}\sum_{i=1}^{p}r_i^2\right)$$

给出。在我们的例子中应该是

$$r_0=\frac{1}{8-1}\left[56-\frac{1}{56}(10^2+8^2+\cdots+4^2)\right]$$

$$=\frac{1}{7}\left[56-\frac{1}{56}\times 432\right]=6.90。$$

由此就可以依前面那样来对平均数间的差异进行显著性检验了。

要指出，含量不等的单因素试验虽然也可以进行分析，但是计算起来要比等含量的情形麻烦。不仅如此，在对数据的利用及分析的精度上也不如含量相等的试验。因此不在毫无办法的情况下，不必人为地去安排这样的试验。

（三）完全随机区组设计　前述的完全随机设计之主要缺点在于其效率往往比较低。因为在进行随机化时不受限制，致使各个处理内部的试验误差一般偏高。如果采用局部控制的思想，在安排处理时先把试验对象有意识地分成若干组群，使得每一个组群内再对影响试验的某些条件加以控制。譬如按动物的家系，初重，品种，性别，年龄或日龄的不同来分组群。要注意，这时要求每个组群内的组数要等于处理数或处理数的某个倍数。我们把不同的组群称为区组。然后，用随机化的方法把不同的处理安排在每个区组内的各组中去。这种试验设计的方法称为完全随机区组设计。由于试验的各处理是随机地在各个区组内进行的，每个区组内部又进行了局部控制，使得试验对象的差异尽可能地小。因此将会提高试验的效率。

例如，对每窝 3 只的三窝猪中每只分别饲以 B_{12} 含量不同的三种饲料，令 A_1: 每磅饲料中含 B_{12} 5 微克，A_2: 含 10 微克，A_3: 含 15 微克。分别观测其日增重，得结果如表 19。

表 19　含 B_{12} 饲料试验日增重

窝 / 饲料	B_1	B_2	B_3	$X_{i\cdot}$	$\bar{x}_{i\cdot}$
A_1	1.26(x_{11})	1.21(x_{12})	1.19(x_{13})	3.66	1.22
A_2	1.29(x_{21})	1.23(x_{22})	1.22(x_{23})	3.74	1.25
A_3	1.38(x_{31})	1.27(x_{32})	1.23(x_{33})	3.88	1.30
$X_{\cdot j}$	3.93	3.71	3.64	11.28	$\bar{x}_{\cdot\cdot}=1.25$

试分析 B_{12} 的不同含量对猪的增重是否有影响？

这种情况的方差分析过程与前面基本类似。令 $p=3$ 为处理的水平数，$q=3$ 为区组(窝)数。仍采用前面的记法，有

$$W=\sum_{i=1}^{p}\sum_{j=1}^{q}x_{ij}^2=1.26^2+1.21^2+\cdots$$

$$+1.23^2=14.1634,$$

$$P=\frac{1}{q}\sum_{i=1}^{p}X_{i\cdot}^2$$

$$=\frac{1}{3}(3.66^2+3.74^2+3.88^2)$$

$$=\frac{1}{3}\times 42.4376=14.1459,$$

表 20　B_{12} 含量试验方差分析表

方差来源	平方和	自由度	方差	F	临界值
A	0.0083	2	0.00415	7.44*	$F_{0.05}=6.94$
B	0.0153	2	0.00765	13.63*	$F_{0.01}=18$
机误（E）	0.0022	4	0.00056		
总　和	0.0258	8			

* 显著。

$$Q=\frac{1}{p}\sum_{j=1}^{q}X^2_{\cdot j}$$

$$=\frac{1}{3}(3.93^2+3.71^2+3.64^2)$$

$$=\frac{1}{3}\times 42.4586=14.1529,$$

$$C=\frac{1}{pq}X^2_{\cdot\cdot}=\frac{1}{3\times 3}\times 11.28^2$$

$$=\frac{1}{9}\times 127.2384=14.1376。$$

于是平方和可以分解为

总平方和

$$S_T=W-C=14.1634-14.1376=0.0258,$$

因素A平方和

$$S_A=P-C=14.1459-14.1376=0.0083,$$

因素B平方和

$$S_B=Q-C=14.1529-14.1376=0.0153,$$

机误平方和

$$S_e=S_T-S_A-S_B=0.0258-0.0083-0.0153=0.0022。$$

自由度可分解为

$$d.f_T=pq-1=8,\ d.f_A=p-1=2,$$

$$d.f_B=q-1=2,$$

$$d.f_e=d.f_T-d.f_A-d.f_B=(p-1)(q-1)=4.$$

可以得出表 20 的方差分析表。

F 检验表明因素 A（B_{12} 含量）各水平间和区组（窝）间都有显著差异。如果这个试验不考虑区组（窝）间的不同，只看成是每个处理内各有三头猪，按前面第一个例子的方法来分析。那么区组间的差异将全部变成了机误，这时显然，F 检验就不会呈显著差异了。

至于平均数间的比较方法与前面所述完全相同，就不再介绍了。

表 21　统计符号一览表

符　号	其它记法	含　意
α		统计检的显著性水准
μ		总体平均值
σ		总体标准差
σ^2		总体方差
$\bar{x}$		样本平均数
s	SD	样本标准差
s^2	V	样本方差
$s_{\bar{x}}$	SE	样本平均数的标准差即标准误差
C.V.	CVC	变异系数
S_d		成对观测值的差异标准差
$S_{\bar{d}}$	SE_d	成对观测值的差异标准误差
$S_{\bar{x}_1-\bar{x}_2}$	$SE_D\ s_D$	两个样本平均数差异的标准差
t		统计量
t_α		显著性水准为 α 时 t 的临界值
$d.f.$	$DF\ df\ \nu$	自由度
$(L.S.D.)_\alpha$	$LSD_\alpha\ lsd(\alpha)$	水准为 α 时两平均数的最小显著差数
χ^2		统计量
χ^2_α	$\chi^2_{\alpha,\nu}$, $\chi^2_\alpha(\nu)$	(自由度为 ν)水准为 α 的 χ^2 的临界值
t	$E\ T$	计数资料χ^2统计量中的理论频数
a	$O\ A$	计数资料χ^2统计量中的实际观测频数
b		回归系数
a		回归常数
l_{xx}	Σ_{x^2}	变量 x 的样本平方和
l_{yy}	Σ_{y^2}	变量 y 的样本平方和
l_{xy}	Σ_{xy}	变量 x 与变量 y 的样本乘积和
r		样本相关系数
S_A	SS_A	因素 A 的平方和
S_e	SS_e	机误平方和
C	CT	矫正数
F		统计量
$F_\alpha(n_1, n_2)$	F_α	自由度为 n_1，n_2，水准为 α 时 F 的临界值

生物统计基本知识就简单介绍到这里。毫无疑问，生物统计是分析和处理生物科学研究中观测资料的一个重要的数学工具。因此生物

统计的知识应该是生物科学工作者必须掌握的知识之一。有了它就可以使我们能更充分地利用观测资料所提供的信息找出各种生命现象内在的统计规律性。另一方面，在应用生物统计方法时也要注意到统计知识的基本原理，各种分析方法的前提条件及应用范围。如果不顾条件地套用生物统计的分析方法，其结果并不一定是生命现象的内在规律，甚至是错误的结论。至于生物统计的原理和方法的全面而深入的知识，许多中外书籍都有介绍，想深入了解和学习的读者可以参阅有关的书籍。

另外，由于生物统计中所用的符号不统一，为阅读方便，这里特将本文中所用的主要符号与其它的记法及它们的含义列表21供读者参考。