目　录

第一章　完全随机化单因子试验设计

试验中只考虑一个因子(假定为A)，选择它的 a 个水平 A_1、A_2、……、A_a，比较这 a 个水平的试验称为单因子试验。更一般地，不把 A_1、A_2、……、A_a看成是因子 A 的水平，而看成是给定的 a 个方法或条件。当然，在各水平要重复多次试验，假定重复数为 n，则总共要进行an次试验。各水平的重复数也可以不相同。

§1.1 试 验 方 法

这里，an次试验全部按随机顺序进行，称为完全随机化试验法。这种试验方法除所提出的因子外，其他因子的试验环境变动完全随机化，都归并到试验误差上去。

〔例1.1〕为提高某化工产品的强度，只提出反应温度（A）作为试验因子。选定120℃(A_1)、140℃(A_2)、160℃

试验顺序表　　　　表 1.1

顺序号	1	2	3	4	5	6	7	8	9	10	11	12
水　平	A_2	A_3	A_4	A_4	A_4	A_2	A_3	A_3	A_1	A_1	A_4	A_3
日　期	第一天				第二天				第三天			

顺序号	13	14	15	16	17	18	19	20
水　平	A_3	A_2	A_1	A_1	A_1	A_4	A_2	A_2
日　期	第四天				第五大			

(A_3)、180℃(A_4)4个水平，各水平重复5次。找出最适宜的操作条件。

本试验a＝4，n＝5。总计20次。完全按随机顺序进行，如表1.1所示。

所得试验数据整理如表1.2所示。

产 品 强 度　　　　表 1.2

水　　平	A_1	A_2	A_3	A_4
各水平重复	7.9	8.0	8.3	8.3
	7.5	8.6	8.9	7.8
	7.9	8.1	8.5	7.8
	7.6	8.4	8.4	7.9
	7.7	8.1	8.4	8.1

假定强度值越大越好，需要分析这些数据，作出结论。一般把试验数据整理成表1.3那样。

完全随机化的试验数据　　　　表 1.3

水　　平	A_1	A_2	……	A_i	……	A_a
各水平重复	x_{11}	x_{21}	……	x_{i1}	……	x_{a1}
	x_{12}	x_{22}	……	x_{i2}	……	x_{a2}
	⋮	⋮		⋮		⋮
	x_{1n}	x_{2n}	……	x_{in}	……	x_{an}
和	$x_{1\cdot}$	$x_{2\cdot}$	……	$x_{i\cdot}$	……	$x_{a\cdot}$
平　　均	$\bar{x}_{1\cdot}$	$\bar{x}_{2\cdot}$	……	$\bar{x}_{i\cdot}$	……	$\bar{x}_{a\cdot}$

表中数据x_{ij}表示A因子在i水平，第j次重复的数据。其他记号意义如下：

$$x_{i\cdot} = \sum_{j=1}^{n} x_{ij} = A_i \text{ 水平的数据和;}$$

$$\overline{x}_{i\cdot} = \frac{1}{n} x_{i\cdot} = \frac{1}{n} \sum_{j=1}^{n} x_{ij} = A_i \text{ 水平的数据平均值;}$$

$$x_{\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{n} x_{ij} = \text{全部数据之和;}$$

$$\overline{x}_{\cdot\cdot} = \frac{1}{an} x_{\cdot\cdot} = \frac{1}{an} \sum_{i=1}^{a} \sum_{j=1}^{n} x_{ij} = \text{全部数据之平均值。}$$

下标处若有黑点(·)，表示对此下标求和；字母上若有一横(—)，表示对有黑点的那个下标取和，然后平均。本书原则上按此规定使用记号。

§1.2 数据构造模型

数据构造模型是在分析数据时提出的一种假定。由于有了这个假定才可能应用数理统计方法。

数据构造模型需根据试验方法的本质进行假定。在试验中，由于存在试验误差，在 A_i 水平上进行 n 次试验所得的 n 个数据不一定是相等的，所以试验数据 x_{ij} 可以表达成：

$$x_{ij} = \mu_i + e_{ij} \tag{1.1}$$

式中 μ_i——A_i 水平真值；

e_{ij}——数据中包含的误差值。

这里 μ_i 和 e_{ij} 都是未知的。而真值 μ_i 可以表达成：

$$\mu_i = \overline{\mu}_{\cdot} + (\mu_i - \overline{\mu}_{\cdot}) = \mu + \alpha_i \tag{1.2}$$

式中 $\mu = \overline{\mu}_{\cdot} = \frac{1}{a} \sum_{i=1}^{a} \mu_i$

$\alpha_i = \mu_i - \overline{\mu}_{\cdot}$

这里 μ 称为一般平均。α_i 是 μ_i 对于 $\overline{\mu}_{\cdot}$ 的偏移，称为 A_i

水平效应或A_i的主效应(定义$\sum_{i=1}^{a}\alpha_i=0$)。所以把μ_i理解为(一般平均)+(A_i水平效应)。

误差值e_{ij}，由于试验顺序完全随机化，所以它是一个随机变量。可以证明误差$\{e_{ij}\}$服从互相独立，平均值为零，方差为σ^2的正态分布，用记号N.I.D.$(0,\ \sigma^2)$或$N(0,\ \sigma^2)$表示。因此我们可以把混入各试验数据的误差e_{11}，e_{12}，……e_{an}看成是从$N(0,\ \sigma^2)$抽出的容量为an的一个随机抽样。随之，σ^2将成为衡量试验误差的一个尺度，称为误差方差。

所采用的数据构造模型重新整理如下：

$$x_{ij}=\mu+\alpha_i+e_{ij} \tag{1.3}$$

这里μ和α_i都是未知的。但一般平均值μ和A_i水平效应α_i是常量，即参数。而e_{ij}是变量(随机变量，或变量的实现值)。本书中，参数将使用希腊字母μ、α、……变量将使用拉丁字母e、f、……

§1.3 方差分析

我们希望知道的是：水平A_1，A_2，……A_a之间是否有差异；如果有差异，哪一个好。现在，A_i的水平效应设为α_i。问题变成，α_1，α_2，……α_a之间是否有差异；如果有差异，哪一个最好。对此我们运用数理统计的方法，一是检验A水平间没有差异的假设；二是估计A的水平效应。而这种估计和检验均以方差分析为基础。

(一) 方差分析

所谓方差分析，就是按给出离散度的各种因素将总平方和进行分解。对于给定的数据x_1，x_2，……x_n，我们用各

数据对于全体数据平均值的偏差平方和$\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)$作为它们的离散度。在这里，整个试验数据的离散度为总平方和S_T：

$$S_T=\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij}-\overline{x}_{..})^2 \qquad (1.4)$$

全体an个试验数据的值可能不一样。而造成这种差异（离散度）的原因是取了a个不同的水平和存在试验误差。因此将整个离散度按这两个因素分解，则上述方差分解为：

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij}-\overline{x}_{..})^2=n\sum_{i=1}^{n}(\overline{x}_{i.}-\overline{x}_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij}-\overline{x}_{i.})^2$$

即：总平方和(S_T)＝A间平方和(S_A)

＋误差平方和(S_e)

上式右边第一项，A间平方和项，若除以乘数n，则是$\overline{x}_{1.}$，$\overline{x}_{2.}$，……$\overline{x}_{a.}$与总平均值$\overline{x}_{..}$的偏差平方和，表示$\overline{x}_{1.}$，$\overline{x}_{2.}$，……$\overline{x}_{a.}$之间的离散度，因此是因子在A水平差异的平方和。每个水平都进行n次试验，因此乘n，最终为总的A间平方和S_A。

上式右边第二项是x_{i1}，x_{i2}，……，x_{in}与其平均值$\overline{x}_{i.}$的偏差平方和$\sum_{j=1}^{n}(x_{ij}-\overline{x}_{i.})^2$在a个水平的试验误差的总计。称为误差平方和或残差平方和，用S_e表示。

现验证（1.4）式的左边等于右边，首先分解

$$x_{ij}-\overline{x}_{..}=(\overline{x}_{i.}-\overline{x}_{..})+(x_{ij}-\overline{x}_{i.})$$

将两边平方得：

$$(x_{ij}-\overline{x}_{..})^2=(\overline{x}_{i.}-\overline{x}_{..})^2+(x_{ij}-\overline{x}_{i.})^2$$
$$+2(\overline{x}_{i.}-\overline{x}_{..})(x_{ij}-\overline{x}_{i.})$$

两边对 i 、 j 求和：$\sum\limits_{i}^{a}\ \sum\limits_{j}^{n}$。其中右边第三项

$$2\sum_{i=1}^{a}\ \sum_{j=1}^{n}(\overline{x}_{i.}-\overline{x}_{..})(x_{ij}-\overline{x}_{i.})$$

$$=2\sum_{i=1}^{a}(\overline{x}_{i.}-\overline{x}_{..})\ \sum_{j=1}^{n}\ (x_{ij}-\overline{x}_{i.})$$

$$=2\sum_{i=1}^{a}(\overline{x}_{i.}-\overline{x}_{..})\left(\sum_{j=1}^{n}x_{ij}-\ n\overline{x}_{i.}\right)=0$$

(因为按定义可知 $\sum\limits_{j=1}^{n}x_{ij}-\ n\overline{x}_{i.}=0$)

这样左边 $\sum\limits_{i=1}^{a}\sum\limits_{j=1}^{n}\ (x_{ij}-\overline{x}_{..})^2=S_T$

右边 $\sum\limits_{i=1}^{a}\sum\limits_{j=1}^{n}(\overline{x}_{i.}-\overline{x}_{..})^2+\sum\limits_{i=1}^{a}\sum\limits_{j=1}^{n}\ (x_{ij}-\overline{x}_{i.})^2+0$

$$=n\sum_{i=1}^{a}(\overline{x}_{i.}-x_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}\ (x_{ij}-\overline{x}_{i})^2+0$$
$$=S_A+S_e$$

即 $S_T=S_A+S_e$，验证了（1.4）式成立。

所以，这种试验方法中总平方和可以分解为A间平方和与误差平方和。

（二）自由度

数理统计中 χ^2分布的模数是自由度。自由度不同，χ^2 分布的形状也不同。可以证明， 平方和乘以某数将服从 χ^2 分

布，或者在某条件下服从 χ^2 分布。方差分析中的自由度这时就是 χ^2 分布的自由度。

而实际上自由度与独立成分的个数相一致，自由度 n 说明平方和由 n 个独立成分构成。首先看一下

$$S_A = n\sum_{i=1}^{a}(\overline{x}_{i.}-\overline{x}_{..})^2,$$

它是 $(\overline{x}_{1.}-\overline{x}_{..}), (\overline{x}_{2.}-\overline{x}_{..}), \cdots\cdots, (\overline{x}_{a.}-\overline{x}_{..})^2$ 这 a 个成分的平方和。这 a 个成分都是对总平均值的偏差，而全部相加为零，所以 a 个成分约束的个数为 1，独立成分的个数则为 (a−1)。可知 A 间平方和 S_A 的自由度是 a−1。

其次看一下

$$S_e = \sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij}-\overline{x}_{i.})^2,$$

它是

$$\begin{array}{lllll}
(x_{11}-\overline{x}_{1.}), & (x_{21}-\overline{x}_{2.}), & (x_{31}-\overline{x}_{3.}), & \cdots\cdots, & (x_{a1}-\overline{x}_{a.})\\
(x_{12}-\overline{x}_{1.}), & (x_{22}-\overline{x}_{2.}), & (x_{32}-\overline{x}_{3.}), & \cdots\cdots, & (x_{a2}-\overline{x}_{a.})\\
\vdots & \vdots & \vdots & \vdots & \vdots\\
(x_{1n}-\overline{x}_{1.}), & (x_{2n}-\overline{x}_{2.}), & (x_{3n}-\overline{x}_{3.}), & \cdots\cdots, & (x_{an}-\overline{x}_{a.})
\end{array}$$

这 an 个成分的平方和。由于各列 n 个成分之和为 0，所以独立成分是 an−a=a(n−1) 个。则误差平方和的自由度是 a(n−1)。

最后看一下

$$S_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij}-\overline{x}_{..})^2$$

它是

$$\begin{array}{llll}
(x_{11}-\overline{x}_{..}), & (x_{21}-\overline{x}_{..}), & \cdots\cdots, & (x_{a1}-\overline{x}_{..})\\
(x_{12}-\overline{x}_{..}), & (x_{22}-\overline{x}_{..}), & \cdots\cdots, & (x_{a2}-\overline{x}_{..})\\
\vdots & \vdots & \vdots & \vdots\\
(x_{1n}-\overline{x}_{..}), & (x_{2n}-\overline{x}_{..}), & \cdots\cdots, & (x_{an}-\overline{x}_{..})
\end{array}$$

这an个成分的平方和。而所有这些成分之和为0，独立成分个数是an－1。所以总平方和S_T的自由度是an－1。

用记号ϕ表示自由度，ϕ_T、ϕ_A、ϕ_e分别表示S_T、S_A、S_e的自由度。对应于方差分解，等式

$$\phi_T=\phi_A+\phi_e \tag{1.5}$$

成立。

实际求自由度的方法是：

总平方和自由度ϕ_T＝(所有数据的个数)－1

A间平方和自由度ϕ_A＝(因子A的水平数)－1

这样，误差平方和的自由度$\phi_e=\phi_T-\phi_A$。

（三）平均平方与平均平方的期望值

平方和除以自由度称为平均平方，用记号V表示。则V_A表示A间平均平方，V_e表示误差平均平方。即：

$$V_A=\frac{S_A}{a-1}$$
$$V_e=\frac{S_e}{a(n-1)} \tag{1.6}$$

平均平方是从数据$\{x_{ij}\}$计算出来的，但由于存在试验误差，即使是做相同的试验，数据$\{x_{ij}\}$的值也要变化。因此平均平方V_A、V_e的值每次试验时也都要变。按数理统计的概念，V_A、V_e是样本（数据）的函数，是统计量。作为统计量的一个重要的数字特征就是它的数学期望。V_A、V_e的期望值用$E\{V_A\}$，$E\{V_e\}$表示。通过运算可以知道这些期望值是：

$$E\{V_A\}=\sigma^2+n\sigma_A^2 \tag{1.7}$$
$$E\{V_e\}=\sigma^2$$

其中：$\sigma^2=V_{ar}\{e_{ij}\}$是试验误差的方差。

$$\sigma_A^2=\frac{1}{a-1}\sum_{i=1}^{a}\alpha_i^2 \text{是因子A主效应方差。}$$

将以上这些结果整理成方差分析表如下：

方差分析表　　　　表 1.4

变差来源	平方和	自由度	平均平方	F_0	平均平方期望值
A间	S_A	$a-1$	$V_A=\frac{S_A}{a-1}$	V_A/V_e	$\sigma^2+n\sigma_A^2$
误差	S_e	$a(n-1)$	$V_e=\frac{S_e}{a(n-1)}$		σ^2
总计	S_T	$an-1$			

（四）假设检验

在数理统计中假设检验的思想方法是：提出一个假设，把它与数据进行对照，判断是否舍弃它。其判断步骤如下：

（Ⅰ）设假设 H_0 正确，可导出一个理论结论，设此结论为R_0；

（Ⅱ）再根据试验得出一个试验结论，与理论结论相对应，设为R_1；

（Ⅲ）比较R_0与R_1，若R_0与R_1没有大的差异，则没有理由怀疑H_0，从而判定为："不舍弃H_0"（采用H_0）；若 R_0 与 R_1有较大差异，则可以怀疑 H_0，此时判定为："舍弃H_0"。

比如，我们希望检验的假设是：A水平间没有差异。可以表示为 H_0：$\alpha_1=\alpha_2=\cdots\cdots=\alpha_a=0$；也可以表示为 H_0：$\sigma_A^2=0$。

由方差分析表1.4可知，平均平方期望值 $E\{V_e\}=\sigma^2$，$E\{V_A\}=\sigma^2+n\sigma_A^2$；前者表示$V_e$平均，后者表示$V_A$平均。一般来说$V_A$大于$V_e$。然而，若假设$H_0$成立，$\sigma_A^2=0$，则$V_A$与$V_e$平均说来相同，$V_A$与$V_e$之比接近1；若$H_0$不真，则$V_A/V_e$将大于1。所以，直观上的检验方法是

V_A/V_e接近1 ⇒不舍弃H_0；

V_A/V_e比1大得多⇒舍弃H_0。

但是，V_A/V_e比1大多少才能舍弃H_0呢？为确定这个量的界限，需要利用数理统计中关于F分布的理论。

若y_1服从自由度为ϕ_1的χ^2分布，y_2服从自由度为ϕ_2的χ^2分布，并且y_1、y_2相互独立，则$\dfrac{y_1}{\phi_1}\Big/\dfrac{y_2}{\phi_2}$服从自由度为（$\phi_1$，$\phi_2$）的F分布。F分布是连续分布，分布模数是两个自由度（ϕ_1，ϕ_2）。称ϕ_1为分子自由度，称ϕ_2为分母自由度。在自由度为（ϕ_1，ϕ_2）的F分布中，某点右侧面积为p，也就是F比此值大的概率为p，把这个值写为$F_{\phi_2}^{\phi_1}(p)$。若检验的显著性水平（或危险率）给定为α时，则可以把$F_{\phi_2}^{\phi_1}(\alpha)$作为临界值来检验假设。

这里，S_e/σ^2服从自由度为ϕ_e的χ^2分布；当H_0成立，$\sigma_A^2=0$时，S_A/σ^2也服从自由度为ϕ_A的χ^2分布；又S_A与S_e相互独立，所以$\dfrac{S_A}{\phi_A\sigma^2}\Big/\dfrac{S_e}{\phi_e\sigma^2}=V_A/V_e$服从自由度为（$\phi_A$，$\phi_e$）的F分布。这就是假定$H_0$正确时的理论结论$R_0$。而试验结论$R_1$要与理论结论$R_0$相比较。由给定的显著性水平，通常是$\alpha=0.05$；分子自由度$\phi_1=\phi_A=a-1$，分母自由度$\phi_2=\phi_e=a(n-1)$；查F分布表得出$F_{a(n-1)}^{a-1}(\alpha)$。所以$H_0$：$\alpha_1=\alpha_2=\cdots\cdots=\alpha_a=0$（$\sigma_A^2=0$）的检验是：（显著性水平$\alpha$）

$$
\begin{aligned}
&V_A/V_e > F^{a-1}_{a(n-1)}(\alpha) \Rightarrow \text{舍弃}H_0; \\
&V_A/V_e \leq F^{a-1}_{a(n-1)}(\alpha) \Rightarrow \text{不舍弃}H_0。
\end{aligned}
\tag{1.8}
$$

通常，$F^{a-1}_{a(n-1)}(\alpha)$一般性地表示成$F^{\phi_A}_{\phi_B}(\alpha)$或$F_\alpha(\phi_A, \phi_B)$。

（五）具体做法

为了检验必须作出方差分析表。作方差分析表的关键在于平方和的计算。为简便起见，通常使用线性变换

$$u_{ij}=(x_{ij}-a)\times b \tag{1.9}$$

把原数据 $\{x_{ij}\}$ 变换成简单的数据 $\{u_{ij}\}$。先计算出它的平方和：S'_A、S'_e、S'_T。在计算平方和时使用其定义式的变形公式

$$
\begin{aligned}
S'_T &= \sum_{i=1}^{a}\sum_{j=1}^{n} u_{ij}^2 - \frac{u_{..}^2}{an} \\
S'_A &= \frac{1}{n}\sum_{i=1}^{a} u_{i.}^2 - \frac{u_{..}^2}{an}
\end{aligned}
\tag{1.10}
$$

$$S'_e = S'_T - S'_A$$

其中 $\dfrac{u_{..}^2}{an} = \dfrac{(\text{所有数据之和})^2}{\text{所有数据的个数}}$，称为修正项CT′。

再通过变换，得：

$$S_A = \frac{1}{b^2}S'_A,\quad S_e = \frac{1}{b^2}S'_e,\quad S_T = \frac{1}{b^2}S'_T \tag{1.11}$$

算得原数据的平方和。如此，即可列表检验。

〔例1.1〕对表1.2的试验数据进行检验。

首先对原数据$\{x_{ij}\}$实行变换

$$u_{ij}=(x_{ij}-8.0)\times 10$$

变为简单数据$\{u_{ij}\}$，把它们记入表1.5。

变换后的试验数据表 表 1.5

水　平	A_1	A_2	A_3	A_4	
u_{ij}	−1	0	3	3	
	−5	6	9	−2	
	−1	1	5	−2	
	−4	4	4	−1	
	−3	1	4	1	
$u_{i\cdot}$	−14	12	25	−1	$u_{\cdot\cdot}=22$

对应表1.5做出表1.6。

变换后数据的平方 表 1.6

水　平	A_1	A_2	A_3	A_4	
u_{ij}^2	1	0	9	9	
	25	36	81	4	
	1	1	25	4	
	16	16	16	1	
	9	1	16	1	
$u_{i\cdot}^2$	196	144	625	1	$u_{\cdot\cdot}^2=484$

$$\left(\sum_{i=1}^{a}\sum_{j=1}^{n}u_{ij}=272,\quad \sum_{i=1}^{a}u_{i\cdot}^2=966\right)$$

利用表1.6，根据式（1.10)，计算各项平方和。修正项

$$CT'=\frac{u_{\cdot\cdot}^2}{an}=\frac{484}{4\times 5}=24.2$$

$$S_T'=\sum_{i}^{a}\sum_{j}^{n}u_{ij}^2-CT'=272-24.2=247.8$$

$$S_A' = \frac{1}{n}\sum_{i}^{a} u_{i.}^2 - CT' = \frac{1}{5} \times 966 - 24.2 = 169.0$$

$$S_e' = S_T' - S_A' = 247.8 - 169.0 = 78.8$$

与原数据对应的平方和按式（1.11）算出，

$$S_T = \frac{1}{b^2} S_T' = \frac{1}{10^2} \times 247.8 = 2.478$$

$$S_A = \frac{1}{b^2} S_A' = \frac{1}{10^2} \times 169.0 = 1.690$$

$$S_e = \frac{1}{b^2} S_e' = \frac{1}{10^2} \times 78.8 = 0.788$$

这样就可以列出方差分析表。

例 1.1 的方差分析表　　　　表 1.7

变差来源	平方和	自由度	平均平方	F_0
A　间	1.690	3	0.563	11.5
误　差	0.788	16	0.049	
总　计	2.478	19		

如果给定显著性水平 $\alpha=0.05$，查F分布表，可知$F_{16}^{3}(0.05)=3.24$，因为$F_0=11.5>3.24$，所以按5％显著水平检验，舍弃A间没有差异的假设。如果按1％显著水平检验，查F分布表，得$F_{16}^{3}(0.01)=5.29$，也要舍弃假设。这说明A的各水平间有显著差异，选择A的不同水平 A_i 确实能使产品的强度不同。

§1.4 最适水平的决定

经检验，判定A的水平间有显著差异时，那么这个差异是多少？最适水平是哪个？我们设A_i的水平效应为α_i，则问题就是估计α_i。

估计未知参数有点估计与区间估计两种方法。所谓的点估计，是对未知参数取的一个估计值(点)。问题是怎样选择估计所用的统计量，即估计量。这里有所谓"无偏估计量"的概念。当$\overline{x_{i.}}$的期望值是$\mu+\alpha_i$时，称$\overline{x_{i.}}$是$\mu+\alpha_i$的无偏估计量，$\overline{x_{i.}}$平均地为$\mu+\alpha_i$，可以认为$\overline{x_{i.}}$是$\mu+\alpha_i$的没有偏倚的好估计量。在参数的点估计中，一般都使用无偏估计量。

所谓区间估计，是给出未知参数的某个区间，选择两个估计量，构成一个随机区间，在一定的概率保证下包含这个被估计的未知参数。要想做出100%正确的区间估计是不可能的，所以一般都以95%的概率作出此区间。我们把它称作置信度95%的预测区间。

上已谈到，找出最适水平的关键是估计A_i的水平效应α_i。为分析之便，加上一个常数，把$\mu+\alpha_i$看作是A_i的水平效应。

首先考虑$\mu+\alpha_i$的点估计。由数据构造模型：$x_{ij}=\mu+\alpha_i+e_{ij}$ (1.3) 计算

$$\overline{x_{i.}}=\frac{1}{n}\sum_{j=1}^{n}x_{ij}=\frac{1}{n}\sum_{j=1}^{n}(\mu+\alpha_i+e_{ij})=\mu+\alpha_i+\overline{e_{i.}}$$

求出期望值和方差

$$E\{\overline{x_{i.}}\}=E\{\mu+\alpha_i+\overline{e_{i.}}\}=\mu+\alpha_i \qquad (1.12)$$

$$V_{ar}\{\overline{x_{i.}}\}=V_{ar}\{\overline{e_{i.}}\}=\frac{1}{n}\sigma^2 \qquad (1.13)$$

可见$\overline{x_{i\cdot}}$是$\mu+\alpha_i$的无偏估计量。A_i的水平效应的估计可以用A_i水平的试验数据平均值来代替。

再考虑$\mu+\alpha_i$的区间估计。$\mu+\alpha_i$的95%置信度的预测区间是：

$$\text{点估计量} \pm t(\phi, 0.05)\sqrt{\text{点估计量方差的估计值}}$$

其中ϕ是方差估计的自由度，由于σ^2的无偏估计量是V_e，所以σ^2的自由度就是V_e的自由度，即$a(n-1)$，请参见表1.3。

由于点估计量$\overline{x_{i\cdot}}$的方差是σ^2/n；V_e又是σ^2的无偏估计量，所以用V_e来估计σ^2，则点估计量方差的估计值是V_e/n。

这样，$\mu+\alpha_i$的95%置信度的预测区间由

$$\overline{x_{i\cdot}} \pm t(a(n-1),\ 0.05)\sqrt{\frac{V_e}{n}} \qquad (1.14)$$

给出。这里$t(a(n-1),\ 0.05)$表示 t 分布临界点。

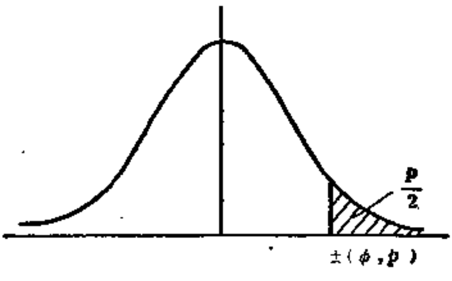

图 1.1 t分布的临界点

在 t 分布中，如果某点为$t(\phi, p)$，则此点的右侧，也就是取值比这一点大的概率是$\frac{p}{2}$。如图1.1所示。