第九章　方差分析

Analysis of Variance （ANOVA）

在生产实践、科学研究和经济活动中，影响事物的因素往往很多，其中有些因素影响较大，有些因素影响较小，我们需要了解哪些因素对事物的影响是显著的。

例如影响农作物单产的因素有：作物品种，施肥的数量和质量，土地的位置和肥沃程度，水质以及灌溉技术等。弄清哪些因素对作物单产影响特别显著，对提高作物产量意义很大。

通常先用科学的方法(试验设计)来获取有用的试验数据，然后对所得数据进行分析。方差分析就是对试验结果的数据作分析的一种常用的统计方法。

方差分析分为单因素方差分析和多因素方差分析。

第一节　单因素方差分析

研究问题:

一、研究问题的背景

二、单因素方差分析的模型

三、单因素方差分析的方法

四、单因素方差分析的应用

本节重点:

一、离差平方和的分解

二、单因素方差分析的检验

一、研究问题的背景

1、一些术语：

1）因素：试验中，那些需要考察的可控条件称为因素或因子，用 A、B、C、⋯ 表示。

2）水平：因素变化的各个等级称为水平，因子 A 的 p 个水平用 A_1、A_2、⋯、A_p 表示。

3）单因素试验：试验时只有一个因素在变化，而其它条件控制不变的试验。

4）多因素试验：试验中的变化因素多于一个。

5）单因素方差分析：只考虑一个因素，各水平相互独立。

6）多因素方差分析：考虑多个因素，各水平独立或相关。

2、研究案例

农科所研究人员引进五种高产油菜品种，想知道哪一种适合在泰顺种植，现把每一品种在四块试验田试种，每块亩亩产量见下表：

品种 \ 亩产量 \ 地块	1	2	3	4	$\bar{x}_i$
A_1 $\xi_1 \sim N(\mu_1,\sigma^2)$	256	222	280	298	264
A_2 $\xi_2 \sim N(\mu_2,\sigma^2)$	244	300	290	275	277
A_3 $\xi_3 \sim N(\mu_3,\sigma^2)$	250	277	230	322	270
A_4 $\xi_4 \sim N(\mu_4,\sigma^2)$	288	280	315	259	286
A_5 $\xi_5 \sim N(\mu_5,\sigma^2)$	206	212	220	212	213

这是单因素方差分析；研究的问题是：各品种的均值是否有差异。即检验 $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

利用单因素方差分析模型可以解决这种问题！

二、单因素方差分析的模型

表9.1.1　单因素分析所得的试验数据

因素水平 \ 试验号		1	2	$\cdots$	n_i	$\bar{x}_{i\cdot}$
A_1	$N(\mu_1,\sigma^2)$	x_{11}	x_{12}	$\cdots$	x_{1n_1}	$\bar{x}_{1\cdot}$
A_2	$N(\mu_2,\sigma^2)$	x_{21}	x_{22}	$\cdots$	x_{2n_2}	$\bar{x}_{2\cdot}$
$\vdots$	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
A_p	$N(\mu_p,\sigma^2)$	x_{p1}	x_{p2}	$\cdots$	x_{pn_p}	$\bar{x}_{p\cdot}$

这p个正态总体$\xi_1,\xi_2,\cdots,\xi_p$ 相互独立，同方差。$n_1,n_2,\cdots,n_p$是相互独立的自然数，可以相等也可以不等，$x_{i1},x_{i2},\cdots,x_{in_i}$ 是取自正态总体 $\xi_i\sim N(\mu_i,\sigma^2)$ 的一个样本，$\bar{x}_{i\cdot}$是第 i 个样本的平均值，$i=1,2,\cdots,p$, 令 $\varepsilon_{ij}=x_{ij}-\mu_i$, 则 $\varepsilon_{ij}\sim N(0,\sigma^2)$。

ε_{ij}, $i=1,2,\cdots,p$, $j=1,2,\cdots,n_i$ 是相互独立的随机误差。

于是得单因素方差分析的数学模型是：

$$\begin{cases} x_{ij}=\mu_i+\varepsilon_{ij} \\ \varepsilon_{ij}\sim N(0,\sigma^2) \\ \varepsilon_{ij},\ i=1,\cdots,p,\ j=1,\cdots,n_i \text{ 相互独立} \end{cases}$$

我们的任务（要检验的问题）是：

$$H_0:\mu_1=\mu_2=\cdots=\mu_p$$

应用方差分析法解决上述问题的关键是：

利用数据中隐含的信息，把引起数据变化的随机误差和系统误差区分开来。数据变化的波动、随机误差和系统误差应该用什么统计量来表示？

为了强调因素各水平的作用，下面把模型进一步细化：

记 $\mu=\frac{1}{n}\sum_{i=1}^{p}n_i\mu_i$，称为总体平均值。 令 $\alpha_i=\mu_i-\mu$，这是第i个总体的均值与总平均值之差，称为第i水平的效应，$i=1,2,\cdots,p$。显然有 $\sum_{i=1}^{p}n_i\alpha_i=0$

当假设 $H_0:\mu_1=\mu_2=\cdots=\mu_p$ 成立时，有 $\mu_i=\mu$ ，$i=1,2,\cdots,p$ 。这时 $\alpha_i=0$，因此，单因素方差分析模型可改写为：

$$\begin{cases} x_{ij}=\mu+\alpha_i+\varepsilon_{ij} \\ \varepsilon_{ij}\sim N(0,\sigma^2) \\ \varepsilon_{ij},\ i=1,\cdots,p,\ j=1,\cdots,n_i \text{ 相互独立} \end{cases}$$

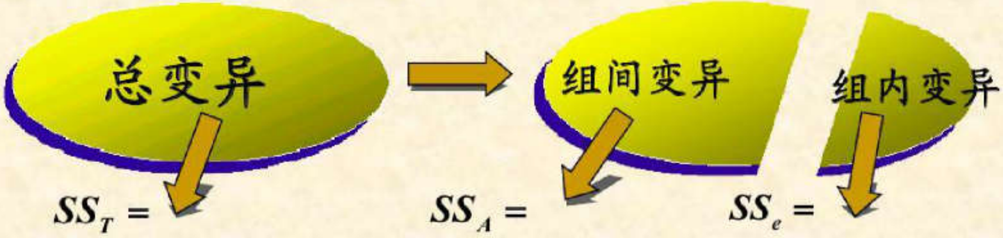
要求检验假设是：

$$H_0:\alpha_1=\alpha_2=\cdots=\alpha_p=0$$

下面讨论总离差平方和、组内平方和与组间平方和的关系。

方差分析的思想及其示意图

方差分析实际上是把总离差平方和进行分解，使每一部分分解可追溯到一种来源可辨别的变异。因此方差分析实际上应称为变异分析或变差分析。



$SS_T=$ $\qquad$ $SS_A=$ $\qquad$ $SS_e=$

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2 \overset{?}{=} \sum_{i=1}^{p}n_i(\bar{x}_{i\cdot}-\bar{x})^2+\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2$$

三、单因素方差分析的方法

1、离差平方和的分解

应用方差分析法解决上述问题的关键是：把总离差平方和分解成组内平方和与组间平方和之和。

$$SS_T=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2$$ 表示总离差平方和；

$$SS_e=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2$$ 表示组内离差平方和；

$$SS_A=\sum_{i=1}^{p}n_i(\bar{x}_{i\cdot}-\bar{x})^2$$ 表示组间离差平方和。

下面的任务是证明：

$$SS_T=SS_e+SS_A$$

记：$n=\sum_{i=1}^{p}n_i$ 为样本总容量； $\bar{x}_{i\cdot}=\frac{1}{n_i}\sum_{j=1}^{n_i}x_{ij}$，$i=1,2,\cdots,p$ 称为第i组平均值；$\bar{x}=\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}x_{ij}=\frac{1}{n}\sum_{i=1}^{p}n_i\bar{x}_i$ 为样本总平均值。

这时总离差平方和：

$$\begin{aligned}SS_T&=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot}+\bar{x}_{i\cdot}-\bar{x})^2\\&=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2+\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\bar{x}_{i\cdot}-\bar{x})^2+\boxed{2\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})(\bar{x}_{i\cdot}-\bar{x})}\\&=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2+\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\bar{x}_{i\cdot}-\bar{x})^2=SS_e+SS_A\end{aligned}$$

式中 $SS_A=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\bar{x}_{i\cdot}-\bar{x})^2=\sum_{i=1}^{p}n_i(\bar{x}_{i\cdot}-\bar{x})^2$ 是组间平方和；

$SS_e=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2$ 是组内平方和，或称误差平方和。

由于 $\bar{x}_{i.}=\frac{1}{n_i}\sum_{j=1}^{n_i}x_{ij}=\frac{1}{n_i}\sum_{j=1}^{n_i}(\mu+\alpha_i+\varepsilon_{ij})=\mu+\alpha_i+\bar{\varepsilon}_i$，$\bar{\varepsilon}_{i.}=\frac{1}{n_i}\sum_{j=1}^{n_i}\varepsilon_{ij}$

$$\bar{x}=\frac{1}{n}\sum_{j=1}^{p_i}n_i\bar{x}_{i.}=\frac{1}{n}\sum_{j=1}^{p_i}n_i(\mu+\alpha_i+\bar{\varepsilon}_{i.})=\mu+\bar{\varepsilon},\quad \bar{\varepsilon}=\frac{1}{n}\sum_{j=1}^{p_i}n_i\bar{\varepsilon}_{i.}$$

故有 $SS_e=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\mu+\alpha_i+\varepsilon_{ij}-\mu-\alpha_i-\bar{\varepsilon}_{i.})^2$

$=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\varepsilon_{ij}-\bar{\varepsilon}_{i.})^2$ 是服从正态分布变量的平方和

SS_e 仅依赖于随机误差 ε_{ij}，故也称之为误差平方和。

$$SS_A=\sum_{i=1}^{p}n_i(\bar{x}_{i.}-\bar{x})^2=\sum_{i=1}^{p}n_i(\mu+\alpha_i+\bar{\varepsilon}_{i.}-\mu-\bar{\varepsilon})^2$$

$=\sum_{i=1}^{p}n_i(\alpha_i+\bar{\varepsilon}_{i.}-\bar{\varepsilon})^2$ 是服从正态分布变量的平方和

SS_A 既与随机误差有关，还与各水平间的效应 α_i 有关，是由各水平效应所决定的。

2、构造假设检验统计量

由于 $E[\frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2]=\sigma^2$，故组内与组内平方和的期望

分别是：$E(SS_e)=E[\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2]=\sum_{i=1}^{p}E[\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2]$

$$=\sum_{i=1}^{p}E[\frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2](n_i-1)=\sum_{i=1}^{p}(n_i-1)\sigma^2=(n-p)\sigma^2$$

$$E(SS_A)=E[\sum_{i=1}^{p}n_i(\alpha_i+\bar{\varepsilon}_{i.}-\bar{\varepsilon})^2]=\sum_{i=1}^{p}n_i[E\alpha_i^2+E(\bar{\varepsilon}_{i.}-\bar{\varepsilon})^2]$$

$$=\sum_{i=1}^{p}n_i\alpha_i^2+\sum_{i=1}^{p}n_iE\bar{\varepsilon}_{i.}^2-nE\bar{\varepsilon}^2=(p-1)\sigma^2+\sum_{i=1}^{p}n_i\alpha_i^2$$

记 $\boxed{Q_E=\frac{1}{n-p}SS_e,\ Q_A=\frac{1}{p-1}SS_A}$ 分别称为组内和组间均方，

则 $E(Q_E)=\sigma^2$，$E(Q_A)=\sigma^2+\frac{1}{p-1}\sum_{i=1}^{p}n_i\alpha_i^2\ge E(Q_E)$

当 $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$ 成立时，$\sum_{i=1}^{p} n_i \alpha_i^2 = 0$，因此

$$E(Q_A) = \sigma^2 + \frac{1}{p-1}\sum_{i=1}^{p} n_i(\mu_i - \mu)^2 = \sigma^2 = E(Q_E)$$

取统计量 $$F = \frac{Q_A}{Q_E} = \frac{SS_A/(p-1)}{SS_e/(n-p)}$$

当 $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$ 成立时，统计量F的值应该不大。如果能求得统计量的分布，就可得到检验统计量。

$x_{ij} \sim N(\mu_i, \sigma^2)$，$j = 1, \cdots, n_i$ 可以看作取自同一正态总体

$$\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{i.})^2 / \sigma^2 \sim \chi^2(n_i - 1),\quad i = 1, 2, \cdots, p$$

$$SSe/\sigma^2 = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{i.})^2/\sigma^2 \sim \chi^2\left(\sum_{i=1}^{p}(n_i - 1)\right) = \chi^2(n-p)$$

$$SS_T/\sigma^2 = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij} - \overline{x})^2/\sigma^2 \sim \chi^2(n-1)$$

由于 $SS_T/\sigma^2 = SS_A/\sigma^2 + SS_e/\sigma^2$，$n-1 = (n-p) + (p-1)$，

由Cochran定理知：$SS_A/\sigma^2 \sim \chi^2(p-1)$，

且统计量 SS_A/σ^2 与 SSe/σ^2 相互独立。

于是在 H_0 为真时，检验统计量是：

$$F = \frac{SS_A/(p-1)}{SS_e/(n-p)} \sim F(p-1, n-p)$$

由前面讨论知，在 H_0 真时，SSA与SSE相比较不应太大。因此上式中的F值也不应太大。因此用F统计量检验 H_0 可采用右单边检验。即对给定的检验水平 α，查表求分位数 $F_{1-\alpha}$，使 $P(F \ge F_{1-\alpha}) = \alpha$，然后根据样本计算 $F_{值}$，进行比较即可。

3、确定拒绝域

检验统计量 $F=\dfrac{SS_A/(p-1)}{SS_e/(n-p)}\sim F(p-1,n-p)$

由前面讨论知，在 H_0 为真时，SSA与SSE之比不应太大因此上式中的F值也不应太大。因此用F统计量检验 H_0 可采用右单边检验。即对给定的检验水平 α，查表求分位数 $F_{1-\alpha}$，使

$$P(F\geq F_{1-\alpha})=\alpha,$$

然后根据样本计算 $F_{值}$，

若 $F_{值}\geq F_{1-\alpha}$，则拒绝原假设 H_0；

若 $F_{值}<F_{1-\alpha}$，则接受原假设 H_0。

以上过程可列成方差分析表如下：

4、方差分析表

表9.1.2 单因素方差分析表

方差来源	平方和	自由度	均方	F 值
因素A(组间)	$SS_A=\sum_{i=1}^{p}n_i(\bar{x}_{i.}-\bar{x})^2$	$p-1$	$Q_A=SS_A/(p-1)$	$F_{值}=Q_A/Q_E$
误差E(组内)	$SS_e=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2$	$n-p$	$Q_E=SS_e/(n-p)$	若 $F_{值}\geq F_{1-\alpha}(p-1,n-p)$ 则拒绝 H_0，否则接受 H_0
总离差	$SS_T=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2$	$n-1$		

计算 SS_T,SS_A,SS_e 简化公式如下：

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

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

$$SS_A=\sum_{i=1}^{p}n_i(\bar{x}_{i.}-\bar{x})^2=\sum_{i=1}^{p}n_i(\bar{x}_{i.}-2\bar{x}_{i.}\bar{x}+\bar{x}^2)$$

$$=\sum_{i=1}^{p}n_i\bar{x}_{i.}^2-2\bar{x}\sum_{i=1}^{p}n_i\bar{x}_{i.}+\bar{x}^2\sum_{i=1}^{p}n_i=\sum_{i=1}^{p}n_i\bar{x}_i^2-2n\bar{x}^2+n\bar{x}^2$$

$$=\sum_{i=1}^{p}n_i\bar{x}_{i.}^2-n\bar{x}^2=\sum_{i=1}^{p}\frac{1}{n_i}(\sum_{j=1}^{n_i}x_{ij})^2-\frac{1}{n}(\sum_{i=1}^{p}\sum_{j=1}^{n_i}x_{ij})^2$$

$$SSe=SS_T-SS_A$$

要计算 SS_T,SS_A,SS_e，只要计算 SS_T,SS_A 中的量：

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}x_{ij}^2\ ,\ T=\sum_{i=1}^{p}\sum_{j=1}^{n_i}x_{ij},\ T^2,\quad T_{i.}=\sum_{j=1}^{n_i}x_{ij},\ T_{i.}^2=(\sum_{j=1}^{n_i}x_{ij})^2$$

5、表上作业法

序号 / 水平	1	2	…	n_i	$T_{i.}$	$T_{i.}^2$	$T_{i.}^2/n_i$	$\sum_{j=1}^{n_i}x_{ij}^2$
A_1	x_{11}	x_{12}	…	x_{1n_1}	$T_{1.}$	$T_{1.}^2$	$T_{1.}^2/n_1$	$\sum_{j=1}^{n_i}x_{1j}^2$
A_2	x_{21}	x_{22}	…	x_{2n_2}	$T_{2.}$	$T_{2.}^2$	$T_{2.}^2/n_2$	⋮
⋮	…	…	…	…	…	…	…	
A_p	x_{p1}	x_{p2}	…	x_{pn_p}	$T_{p.}$	$T_{p.}^2$	$T_{p.}^2/n_p$	$\sum_{j=1}^{n_i}x_{pj}^2$
$\sum$					T		$\sum T_{i.}^2/n_i$	$\sum_{i=1}^{p}\sum_{j=1}^{n_i}x_{ij}^2$

$$SS_T=\sum_{i=1}^{p}\sum_{j=1}^{n_i}x_{ij}^2-T^2/n,\quad SS_A=\sum_{i=1}^{p}T_{i.}^2/n_i-T^2/n,$$

$$SS_e=SS_T-SS_A$$

四、单因素方差分析的应用

1、对本节案例进行的检验

品种＼地块	1	2	3	4	$T_{i.}$	$T_{i.}^2$	$T_{i.}^2/n_i$	$\sum_{j=1}^{n_i} x_{ij}^2$
A_1	256	222	280	298				
A_2	244	300	290	275				
A_3	250	277	230	322				
A_4	288	280	315	259				
A_5	206	212	220	212				
				$\sum$				

$$SS_T = \sum_{i=1}^{p}\sum_{j=1}^{n_i} x_{ij}^2 - T^2/n = 1395472 - 1370784.8 = 24687.2$$

$$SS_A = \sum_{i=1}^{p} T_{i.}^2/n_i - T^2/n = 1383980 - 1370784.8 = 13195.2$$

$$SS_e = SS_T - SS_A = 24687.2 - 13195.2 = 11492$$

把上面的分析结果整理成如下方差分析表。

表9.1.3 方差分析表

方差来源	平方和	自由度	均方	$F_{值}$	临界值
因素(A)	13195.2	5−1=4	Qa=3298.8	4.31	$F_{0.95}(4,15)=3.06$
误差(E)	11492.0	20 − 5 = 15	Qe = 766.1		
总离差	24687.2	20 − 1 = 19			

由于 $F_{值}=4.31>3.06=F_{0.95}(4,15)$，故在 $\alpha=0.05$ 显著性水平上拒绝 H_0，即不同品种的亩产量在0.05水平上有显著差异。

例9.1.2　从某校初中二年级的四个平行班各随机抽取一个学生先后参加五次年段数学竞赛，其结果如下表。

成绩 实验号 水平	1	2	3	4	5
A_1	81	80	88	85	95
A_2	83	89	85	91	88
A_3	76	92	83	90	95
A_4	70	99	82	80	78

问这四个学生成绩是否存在差异（$\alpha=0.025$）？

解：用 ξ_1,ξ_2,ξ_3,ξ_4 表示A_1,A_2,A_3,A_4四个学生数学竞赛的成绩，这时有 $\xi_i \sim N(\mu_i,\sigma^2)$, $i=1,2,3,4$, 假定这四个学生学习成绩不存在差异，即需检验原假设 $H_0:\mu_1=\mu_2=\mu_3=\mu_4$

下面对此进行检验。先对表中数据计算如下表：

实验号 水平	1	2	3	4	5	$\sum_{j=1}^{5}x_{ij}$	$(\sum_{j=1}^{5}x_{ij})^2$	$\frac{1}{n_i}(\sum_{j=1}^{5}x_{ij})^2$	$\sum_{j=1}^{5}x_{ij}^2$
A_1	81	80	88	85	95	429	184041	36808.2	36955
A_2	83	89	85	91	88	436	190096	38019.2	38060
A_3	76	92	83	90	95	436	190096	38019.2	38254
A_4	76	99	92	80	78	409	167281	33456.2	33909
$\sum$						1710		146303	147178

$$SS_T=\sum_{i=1}^{4}\sum_{j=1}^{5}x_{ij}^2-(\sum_{i=1}^{4}\sum_{j=1}^{5}x_{ij})^2/(4\times5)=147178-146205=973,$$

$$SS_A=\sum_{i=1}^{4}\frac{1}{n_i}(\sum_{j=1}^{5}x_{ij})^2-\frac{(\sum\sum x_{ij})^2}{n}=146303-146205=98,$$

$$SS_e=SS_T-SS_A=973-98=875$$

把上面的分析结果整理成如下方差分析表。

表9.1.4　方差分析表

方差来源	平方和	自由度	均方	$F_{值}$	临界值
因素(A)	98	4 − 1 = 3	Qa = 32.7	0.596	$F_{0.975}=4.08$
误差(E)	875	20 − 4 = 16	Qe = 54.7		
总离差	973	20 − 1 = 19			

由于 $F_{0.975}=4.08>F_{值}=0.596$，故接受 H_0，即在 $\alpha=0.025$ 下，四个考生成绩无差异。

可以对数据作线性变换：$x_{ij}'=(x_{ij}-c)/d$，$d\neq 0$，

则有　$SS_T'=SS_T/d^2$，$SS_A'=SS_A/d^2$，$SS_e'=SS_e/d^2$，

于是有　$$F'=\frac{SS_A'/f_A}{SS_e'/f_e}=\frac{SS_A/f_A}{SS_e/f_e}=F,$$

即用简化后数据作方差分析不会影响结论。

为处理方便起见，可先对数据作简化处理再进行计算。令 $x_{ij}'=x_{ij}-80$，$i=1,2,3,4$，$j=1,2,3,4,5$，于是得下表：

x_{ij}' 水平	1	2	3	4	5	$\sum x_{ij}'$	$(\sum x_{ij}')^2/5$	$\sum x_{ij}'^2$
A_1	1	0	8	5	15	29	168.2	315
A_2	3	9	5	11	8	36	259.2	300
A_3	−4	12	3	10	15	36	259.2	494
A_4	−10	19	2	0	−2	9	16.2	469
$\sum$						110	703	1578

$$SS_T=\sum\sum x_{ij}'^2-(\sum\sum x_{ij}')^2/n=1578-(110)^2/20=973$$

$$SS_A=\sum\frac{1}{n_i}(\sum x_{ij}')^2=705-(110)^2/20=98$$

$$SS_e=SS_T-SS_A=973-98=875$$

经计算得方差分析表如下:

方差来源	平方和	自由度	均方	
因素A	$SS_A=98$	3	$Q_A=32.7$	$F_{值}=\frac{Q_A}{Q_E}=0.596$
误差e	$SS_e=875$	16	$Q_e=54.7$	
总离差	$SS_T=973$	19		$F_{0.975}(3,16)=4.08$

$\because F_{0.975}(3,16)=4.08>0.596=F_{值}$,

$\therefore$ $F_{值}$ 落入接受域，故接受 H_0，即认为四个学生的学业成绩在 $\alpha=0.025$ 水平上无显著差异。

小结:

单因素方差分析的计算步骤如下:

①、在必要时简化数据，这不影响 F 值的计算;

②、在数据表上计算SST，SSA和SSE;

③、计算均方 Q_A，Q_E 和 $F_{值}$

④、根据 α 值查表求 $F_{1-\alpha}(p-1,n-p)$，

比较$F_{值}$ 和 $F_{1-\alpha}(p-1,n-p)$ 的大小，

若 $F_{值}\geq F_{1-\alpha}$，则拒绝原假设 H_0;

若 $F_{值}<F_{1-\alpha}$，则接受原假设 H_0。

定理8.1.1（Cochran定理） 设 $\xi_1,\xi_2,\cdots,\xi_n$ 为n个相互独立且服从$N(0,1)$的变量，$Q=\sum_{i=1}^{n}\xi_i^2$ 为 $\chi^2(n)$ 变量。若有

$$Q=Q_1+Q_2+\cdots+Q_k$$

其中 Q_i 为某些正态变量的平方和，这些正态变量分别是 ξ_1，$\xi_2,\cdots,\xi_n$ 的线性组合，其自由度为 f_i，则各 Q_i 相互独立，且为 $\chi^2(f_i)$ 变量的充要条件是：$\sum_{i=1}^{n}f_i=n$。

几个结论：

1、设 $\xi_1,\xi_2,\cdots,\xi_n$ 是正态总体 $N(\mu,\sigma^2)$的一个样本，其方差与修正方差是：$S_n^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2, S^2=\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2$

则 $E[\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2]=\sigma^2$，$\frac{nS_n^2}{\sigma^2}=\frac{\sum(\xi_i-\bar{\xi})^2}{\sigma^2}\sim\chi^2(n-1)$

2、设 $\xi_1\sim\chi^2(n_1)$，$\xi_2\sim\chi^2(n_2),\cdots$，$\xi_p\sim\chi^2(n_p)$ 且相互独立，则 $\xi_1+\xi_2+\cdots+\xi_p\sim\chi^2(\sum_{i=1}^{p}n_i)$

3、设 $\xi\sim\chi^2(n)$，$\eta\sim\chi^2(m)$，且相互独立，则

$$F=\frac{\xi/n}{\eta/m}\sim F(n,m)$$