高等学校教材

工业技术应用数理統計学

下册

周华章 編

高等教育出版社

高等学校教材

工业技术应用数理統計学

下册

周华章编

高等教育出版社

本书对数理統計学在工业技术上的应用作了比較詳細的闡述，分上下两册出版。此次出版的下册是第一版。下册的主要內容有方差分析，迴归分析与相关分析，抽样理論，质量控制，計件与計点质量控制等部分。为了照顾高等工业学校学生及工程人員的学习特点，本书在叙述时幷不过份追求数学的严格性，而尽量举例說明。本书可供高等工业学校作教学参考用书，亦可供工业部門研究人員，工厂技术檢驗人員作参考。

本书原由人民教育出版社出版。現經上級决定，自 1965 年 1 月 1 日起，另行成立“高等教育出版社”；本书今后改用高等教育出版社名义继續印行。

工业技术应用数理統計学

下　册

周　华　章　編

北京市书刊出版业營业許可証出字第 119 号

高等教育出版社出版（北京景山东街）

人　民　教　育　印　刷　厂　印　装

新　华　书　店　北　京　发　行　所　发　行

各　地　新　华　书　店　經　售

統一书号K13010·1131　开本 850×1168 1/32　印张 11 7/8　插页 2

字数 270,000　印数 9,301—13,300　定价（5）¥1.20

196[illegible]年[illegible]月第1版　1965 年 3 月北京第 3 次印刷

說　明

(I)本书下册是第一版发行。在此以前，上册已分别于1960年及1963年发行第一版及第二版。上册第二版，即修訂版的內容比之第一版已有多处經过修改，所用符号亦已有較多的改变，这在上册第二版卷首已有說明。現在下册的符号是沿用上册第二版的，为了讀过上册第一版但未看到第二版的讀者方便起見，特将上册第二版卷首对书中符号所作的說明照录如下：

(1)凡遇只有一个随机变量的場合，概用ξ表示随机变量。ξ的观測值用x表示。譬如$x_1, x_2, \cdots$分别表示ξ的第1, 2, …个观測值。

(2)ξ的期望值$E(\xi)$用$\bar{u}$表示；方根差用σ_ξ(或簡写作σ)表示，即ξ的方差$\mathrm{Var}(\xi)=\sigma_\xi^2$。

(3)ξ的子样平均数作为一个随机变量，用$\bar{\xi}$表示。$\bar{\xi}$的观測值用$\bar{x}$表示。譬如$\bar{x}_1, \bar{x}_2, \cdots$分别表示第1, 2, …个子样观測值的平均数。

(4)$\bar{\xi}$的期望值$E(\bar{\xi})$因等于$E(\xi)$，故仍用$\bar{u}$表示。$\bar{\xi}$的方根差用$\sigma_{\bar{\xi}}$，即$\bar{\xi}$的方差$\mathrm{Var}(\bar{\xi})=\sigma_{\bar{\xi}}^2$。

(5)ξ的子样方根差作为一个随机变量，用σ_x表示。σ_x的观測值，亦即由ξ的子样观測值$x_1, x_2, \cdots, x_n$算出的方根差，用s表示。譬如$s_1, s_2, \cdots$分别表示第1, 2, …个子样观測值的方根差。

(6)σ_x的期望值为$E(\sigma_x)$。σ_x的方根差用σ_{σ_x}表示，即σ_x的方差$\mathrm{Var}(\sigma_x)=\sigma_{\sigma_x}^2$。

(7)遇有二个以上随机变量出現的場合，用ξ, η, ζ, …等表示随机变量。其观測值分别用$x, y, z, \cdots$表示。

(8)$\xi, \eta, \zeta, \cdots$的期望值$E(\xi), E(\eta), E(\zeta), \cdots$分别用$\bar{u}, \bar{v}, \bar{w}, \cdots$表示。$\xi, \eta, \zeta, \cdots$的方根差分别以$\sigma_\xi, \sigma_\eta, \sigma_\zeta, \cdots$表示。

(Ⅱ)按照上面所录8条关于符号的說明，可以看出一条規則，即凡随机变量都用希腊字母，而凡随机变量的观測值則都用拉丁字母来表示。但在下册中这条規則在某些場合不便应用。譬如在迴归分析中，子样迴归系数的观測值用b来表示。根据上述規則，子样迴归系数作为一个随机变量，最好用β来表示。但在迴归分析中，β已被用来表示母体迴归系数。为此，只得用b^*来表示“子样迴归系数”这一个随机变量，而b則为b^*的观測值，譬如以b_1，$b_2, \cdots$分别表b^*的第1，2，…个观測值。其他还有数处有类似情况。遇到这种情况，下册中总是以带星号的拉丁字母表示随机变量，而以不带星号的同一拉丁字母表示該随机变量的观測值。

(Ⅲ)下册中凡遇引述上册中某些公式或定理时，如注明“見上册（修訂版）某章某节”，表示該公式或定理仅在修訂版中可以找到，而在第一版中沒有該公式或定理。如注明“見上册某章某节”，則表示該公式或定理在上册第一版或第二版的同一章节中都能找到。

(Ⅳ)本书虽然列为高等学校教材，但也可供工业部門研究人員及工厂质量檢驗人員作参考之用。为此，本书內容較注重于数理統計方法在工业技术中之实际应用，尽量多举实例加以說明，对于数学推导則不作过份严格的要求，希讀者鉴諒。

内容目录

第十三章　方差分析

§13.1　方差分析总説

对同一品种工业产品的某一性质，如重量、口徑等，进行多次的量度，結果往往不能一致，而有所差异。进一步探求这差异发生的原因，就可知道，在任何一种試驗中，試驗的結果常受一些环境的因素或条件的影响。这就是說，如果試驗的环境或条件改变，各次試驗的結果就表現出一定的差异。例如将同一种布剪成每段一丈，在不同溫度的水中洗濯后，在同溫度的溫室內晾干，結果发現各段布的收縮率有相当明显的差异。这差异有可能是水的溫度不同所造成。又如在进行木材抗压强度試驗时，需要加荷于品种相同而比重不同的各块木材。結果也发現比重不同的各块木材的抗压强度有相当明显的差异。这差异有可能是木材比重不同所造成的。凡是由于試驗条件不同而引起的試驗結果的差异，称为“条件誤差”。

在实际的試驗工作中，虽然我們尽量把試驗条件与环境固定不变，而各次的試驗結果仍表現出一定程度的差异。譬如将同一种布五段，每段一丈，同时在一盆水中洗濯，然后同时在同一溫室內晾干，但每一段布的收縮率仍不完全相同。引起差异的原因，一方面是試驗条件虽然尽量固定不变，但仍不能避免这些条件的微細的变异，另一方面还存在着其他使試驗結果发生差异的因素，如称衡时的誤差，仪器的不准确，原料的不均一，人的因素，以及环境的其他性质对試驗結果的影响等。由于这一类因素（試驗条件除外）所产生的試驗結果的差异，称为“試驗誤差”。

条件誤差与試驗誤差的区別在于：前者是系統性的，后者是随

机性的。譬如說，我們能够断定水的溫度愈高，布的收縮率愈大，那么水的溫度就是造成系統性誤差的一种因素。但当水的溫度以及其他凡人力能使其固定(例如測量收縮率的仪器为同一台仪器)的因素都固定不变时，如果布的收縮率仍有差异，那么这种差异就不是某一特定因素所造成的，而是許多个偶然性因素，或称随机性因素，所造成的。正如以前在第七章最后一节（上册修訂版）中已提到过，如果造成各段布的收縮率不均一的原因幷非一种或数种系統性因素，而是不能确定种数的众多的相互独立的偶然性因素，那么各段布的收縮率通常要形成正态分布。反之，如果有系統性的因素在起作用，它們就不能形成正态分布。方差分析(或称二級动差分析，亦称离差分析)的目的，就是要确定一項試驗的結果中，有沒有条件誤差存在。換言之，就是要确定这一項試驗中有沒有系統性因素在起作用。

方差分析是数理統計方法中很有效的方法之一。它是小子样統計推断理論的推广应用。它的方法和理論根据将在以下一节中詳細說明。

§13.2 一个变异因素的方差分析的問題

如上节所述，一項試驗在其进行时，可能受到某种試驗条件的影响，而使試驗結果表現出系統性的条件誤差。这里所說的試驗条件有时是一种，有时是二种，或甚至二种以上。現在我們将先談一种条件的方差分析。換言之，先解决一种試驗条件对試驗結果有无显著影响的問題。

一种試驗条件的方差分析的进行方法，首先是把这种条件分为若干級来进行試驗，記录其試驗結果。譬如拿水的溫度对布的收縮率的影响問題来說，我們用四种溫度不同的水，如各为 20°C，40°C，60°C 和 80°C，在每一种溫度的水中洗濯五段布，每段一丈，

所有的布都屬于同一品种；将各段布的收縮率列成下表，进行分析。

表 13.1

試件分批	温	度		
	1	2	3	4
1	x_{11}	x_{12}	x_{13}	x_{14}
2	x_{21}	x_{22}	x_{23}	x_{24}
3	x_{31}	x_{32}	x_{33}	x_{34}
4	x_{41}	x_{42}	x_{43}	x_{44}
5	x_{51}	x_{52}	x_{53}	x_{54}

表中 x_{ij} 表示第 i 批試件在第 j 級温度的水中洗濯后的收縮率。普遍言之，用 a 表示試件批数，b 表示温度的級数。

这里有两点应該注意：

(1)第一点要注意的是，所謂变异条件的分級，并不一定要像本例那样为数量性的。在上例中，水的温度可分为 20°C, 40°C 等，是按攝氏温度計上的度数来分級；但另举一例就可以看出，分級就不是数量性的。譬如三名工人在同样温度的水中洗濯同一品种的布，洗完后在同一温度的温室內晾干，問这三名工人的洗濯技术对布的收縮率有无显著影响，亦即三名工人的洗濯技术有无显著差异。这时变异条件，或称变异因素，是工人的洗濯技术，但技术的优劣并不能像温度那样用严格的数字来衡量。再看一个例子：用四架輪廓仪对机械零件的表面光洁度进行量测，其結果有无显著差异这一問題，也同样屬于这种性质。

(2)第二点要注意的是，如果变异因素是数量性的，那么分級进行試驗时最好是分得較为均匀，但不必完全均匀。譬如水的温度可分为 20°C、37°C、58°C、83°C 四級；或分三級，五級亦可。譬如

說木材抗压强度試驗中之变异因素“木材比重”也是一种数量性的因素，可分为三級、四級、五級等。分級的多少完全根据实驗者的需要。

用方差分析来解决試驗条件对試驗結果有无显著影响的問題，其理論根据可以說明如下。如果这个变异因素对試驗結果并无显著影响，那么試驗結果 $a\times b=n$ 个数值 x_{ij} 应当为来自同一个正态分布母体的一个随机性子样。如果这个变异因素对試驗結果有显著影响，譬如說水的溫度对布的收縮率，木材比重对木材抗压强度，确实有影响，那么在表 13.1 中每一列的 x_{ij} 应当是来自一个正态母体的一个随机性子样，但不同的两列中的 x_{ij} 并不来自同一母体。这是因为在每一列中的五段布都是在同一溫度的水中洗濯的，譬如第一列中五个数字 $x_{11}, x_{21}, x_{31}, x_{41}, x_{51}$ 是在溫度为20°C的水中洗濯的五段布的收縮率，这五个数字应当是来自一个正态母体的一个随机性子样。同理，第二列的五个数字，也是来自一个正态母体的一个随机性子样。可是，这两个母体却不是同一个正态母体。方差分析的实质是：提出一項統計假設，即假設所有 $a\times b$ 个 x_{ij} 都来自同一正态母体，然后用 F 检驗法来推断这种假設是否可信。說得更精确些，方差分析假定在每一种溫度的水中洗濯的布的收縮率，其均方根差是相同的(尽管是未知的)，分析的目的是要檢驗收縮率的母体平均值，是否因水的溫度不同而有所不同。推断时所用信度通常为5%。如果檢驗結果表示統計假設不可信，那么我們的結論就是：就5%的信度而論，水的溫度对布的收縮率的母体平均数有显著影响，亦即条件誤差是显著存在的。反之，如檢驗結果表示所提統計假設为可信，則水的溫度对布的收縮率的母体平均值无显著影响。同样，显著二字是相对于5%的信度而言。

§13.3 一个变异因素的方差分析的方法

根据上节所說的理論，为了檢驗所提統計假設是否可信，先将試驗結果写成下列表格。

表 13.2

試件分批	因素分級				
	1	2	………	b	
1	x_{11}	x_{12}	………	x_{1b}	
2	x_{21}	x_{22}	………	x_{2b}	
⋮	⋮	⋮	⋮	⋮	
a	x_{a1}	x_{a2}	………	x_{ab}	
总計	T_1	T_2	………	T_b	T
平均数	$\bar{x}_1$	$\bar{x}_2$		$\bar{x}_b$	$\bar{x}$

表中 $T_1, T_2, \cdots, T_b$ 表示各列数字相加之和，$\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_b$ 表示各列数字的平均数，而 T 及 $\bar{x}$ 分别表示全体 $a\times b$ 个 x_{ij} 的总和及总平均数。

根据上表計算:

$$\bar{x}_j=\frac{\sum_{i=1}^{a}x_{ij}}{a}=\frac{T_j}{a},\quad (j=1, 2, \cdots, b);$$

$$\bar{x}=\frac{\sum_{j=1}^{b}\sum_{i=1}^{a}x_{ij}}{ab}=\frac{T}{ab}.$$

現在我們将一个变异因素的方差分析的假定条件及有待檢驗的統計假設 H_0 列述如下，然后接着討論檢驗 H_0 的方法。(有待

檢驗的統計假設通常用H_0表示，称为“待驗假設”，或“解消假設”）。

假定条件如下：

(1)在变异因素的某一个等級，譬如第j个等級进行試驗所得的观测数据，形成正态分布$N(\bar{u}_j;\sigma_{\varepsilon_j})$。这里$\bar{u}_j$和$\sigma_{\varepsilon_j}$都是未知的参数。这条假定也可换一个説法，即在第$j$个等級进行的第$i$个試件所得的观测数据$\xi_{ij}$，不論$i$是1, 2, …, a中那一个值，都是$N(\bar{u}_j;\ \sigma_{\varepsilon_j})$变量。如表13.2所列，对于每一个$\xi_{ij}$只有一观测值$x_{ij}$。

(2)σ_{ε_j}与j无关，或

$$\sigma_{\varepsilon_1}=\sigma_{\varepsilon_2}=\cdots=\sigma_{\varepsilon_b}=\text{常数}\ \sigma_{\varepsilon},$$

尽管这个常数σ_ε是未知的。

(3)每一个子样$(x_{1j},\ x_{2j},\cdots,x_{aj})$ $(j=1,2,\cdots,b)$都是随机性子样。

(4)各个子样是相互独立的。

統計假設H_0: $\bar{u}_1=\bar{u}_2=\cdots=\bar{u}_b=$常数$\bar{u}$。

檢驗方法:

将表13.2中$a\times b$个观察值x_{ij}的离差平方和分解为两个离差平方和如下:

$$\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x})^2=\sum_{j=1}^{b}\sum_{i=1}^{a}\{(x_{ij}-\bar{x}_j)+(\bar{x}_j-\bar{x})\}^2=$$

$$=\sum_{j=1}^{b}\sum_{i=1}^{a}\{(x_{ij}-\bar{x}_j)^2+2(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x})+$$

$$+(\bar{x}_j-\bar{x})^2\}=\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)^2+$$

$$+2\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x})+$$

$$+a\sum_{j=1}^{t}(\bar{x}_j-\bar{x})^2,$$

因为

$$\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x})=\sum_{j=1}^{b}\{(\bar{x}_j-\bar{x})\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)\}=$$

$$=\sum_{j=1}^{b}(\bar{x}_j-\bar{x})(T_j-a\bar{x}_j)=0,$$

所以有

$$\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x})^2=\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)^2+a\sum_{j=1}^{b}(\bar{x}_j-\bar{x})^2. \tag{13.1}$$

上式等号右方第一項为各組內部离差平方和，第二項为組与組間离差平方和。所謂“各組內部”的“組”是这样理解：例如，同一种溫度的水中洗濯的 a 段布的收縮率构成一組，即表內同一列中的数字。

根据所提統計假設，如果試驗条件对試驗結果无显著影响，試驗結果应形成正态分布 $N(\bar{u};\sigma_t)$。事实上，我們只有从 $a\times b$ 个試驗得到的 $a\times b$ 个观測值，并不知道 $\bar{u}$ 和 σ_t 等于多少，只不过用这两个符号分别表示此正态母体的平均数和均方根差而已。

今以 σ_t^2 除(13.1)式左右两方，得：

$$\frac{1}{\sigma_t^2}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x})^2=\frac{1}{\sigma_t^2}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)^2+$$

$$+\frac{a}{\sigma_t^2}\sum_{j=1}^{b}(\bar{x}_j-\bar{x})^2. \tag{13.2}$$