徐中儒　编著

农业试验
最优回归设计

$Y = X\beta + \varepsilon$

黑龙江科学技术出版社

农业试验最优回归设计

徐中儒　编著

黑龙江科学技术出版社

1988年·哈尔滨

责 任 编 辑：常濂莲
封 面 设 计：刘连生

农业试验最优回归设计

徐中儒 编著

黑龙江科学技术出版社出版
（哈尔滨市南岗区建设街35号）
黑龙江新华印刷厂附属厂印刷

787×1092毫米32开本 14.875印张4插页 300千字
1988年10月第1版·1988年10月第1次印刷
印数：1—4000册　　定价：5.35元
ISBN 7-5388-0340-8/S·27

前　　言

最优回归设计是数理统计中的一个分支，在农业生产和科学研究中有着广泛的应用。它是解决农业科学试验中多变量问题的数据处理，建立数学模型，制定最佳农艺措施，获得优化决策的极其有效的统计数学方法。

本书是在我多年运用回归设计方法所取得的一些应用成果和收集国内外有关资料的基础上，经过系统整理而写成的。编写中着重以实际应用为主，数学理论推导则尽量从简。全书包含古典回归分析、回归正交设计、回归旋转设计、回归最优设计、均匀设计、回归模型的选择准则及回归方程的优化。通过30多个实例，详细说明如何运用各种最优设计方法和它们的分析与应用。这些实例涉及作物栽培、耕作、农业气象、温室花卉生产、土壤肥料、农业生态、农业机械设计等方面的内容。

本书可供从事农业科技和教育工作的同志阅读，亦可作为有关专业的大学生和研究生学习的参考书，对于从事科学试验的科研人员尤有重要参考价值。

在编写本书过程中，得到侯中田、张瑞忠教授和葛家麒、赵妮珊等同志多方面的支持和帮助，在此表示感谢。

由于编著者水平所限，难免有许多缺点和不妥之处，敬

请读者批评指正

徐中[illegible]

于东北农学院　1988年4月

目　录

第一章　古典回归分析

§1　一元线性回归

§2　多元线性回归

§3　非线性回归

第二章 回归正交设计

第三章　回归旋转设计

第四章 回归的最优设计

§1 回归问题的最优设计

§2 饱和最优设计

§3 最优设计的统计检验

第五章 均匀设计

§1 一次均匀设计

第六章　回归模型的选择准则

第七章　回归方程的优化

第一章　古典回归分析

在自然现象中，同时出现的多个变量，通常不是各自孤立的在变化，而是相互依赖、相互联系的。变量之间相互依赖的关系大体上可分为两大类：一类是确定性关系。例如，电路中的欧姆定律就是确定性关系，即电压V、电阻R给定时，电流I就可由$I=V/R$完全确定。再如，自由落体运动中，物体下落的距离S与时间t之间，就有关系$S=\frac{1}{2}gt^2$，如果取定了t的值，S的值就完全确定了。变量之间的这种确定性关系，称为函数关系。另一类是非确定性关系。例如，降水量与农作物产量之间的关系是非确定性的，即使在同样的雨水条件下，作物产量也不是唯一确定的。再如，犁的耕深与牵引阻力的关系。这一类变量之间的关系，只有通过大量的观测试验才能发现它们之间也具有一定的规律性，称为统计规律性。这一类变量之间的关系，称为非确定性关系，亦称相关关系。比如人的身高与体重间的关系，虽然一个人的身高并不能确定体重，但是，通过大量观测，平均说来，身高者，体也重，即身高与体重这两个变量具有相关关系。

一般把讨论随机变量与非随机变量之间的关系的问题，

称为回归分析；把讨论随机变量之间关系的问题，称为相关分析；为了简单起见，我们不区分回归和相关，而统称为相关。若从变量的个数来区分，把两个变量之间的相关，称为简单相关；把3个变量以上的相关，称为复相关；把在复相关的条件下，仅研究两个变量的相关，称为偏相关。

§1 一元线性回归

回归分析是研究变量间相关关系的有力工具。它不仅提供了建立变量间关系的数学表达式（通常称为经验公式）的一般方法，而且进行分析讨论判明所建立的经验公式的有效性，以及如何利用所得到的经验公式去达到预测、控制等目的。回归分析要解决的问题可归纳为：

（1）确定二组或二组以上相对应的变量之间的相关关系，找出这些变量之间的定量关系式。

（2）对这个定量关系式的可靠性进行统计检验。

（3）根据变量之间的定量关系式作出预报和控制，进行优化分析。

（4）对多因素问题进行因素分析，确定各因素之间的主次关系。

（5）应用回归分析的原理，作出试验处理少，统计性质好的试验设计。

由以上各点可见，回归分析在工农业生产及科学技术中有广泛的应用。

一、一元线性回归方程

1．回归方程

假定有两个随机变量 X 和 Y，它们之间相互依赖存在相关关系，由于 Y 的随机性，即使在随机变量 X 取某定值的条件下，随机变量也没有确定的值与之对应，而只有确定的概率分布与之对应。为了要寻求 Y 和 X 之间的关系，一个自然的想法是取 $X=x_0$ 时，所对应的 Y 值的数学期望值作为 Y 的代表值，亦即取

$$Y|_{X=x_0}=E(Y|X=x_0)$$

其中 $E(Y|X=x_0)$ 表示在 $X=x_0$ 条件下 Y 的条件数学期望，它的值随 X 取确定值而确定，当 X 取不同的 x 值时，Y 的数学期望值 $E(Y|X=x)$ 也将取不同的值，即 $E(Y|X=x)$ 的值随 x 的变化而变化，也就是说，$Y|_{X=x}=y=E(Y|X=x)$ 是 x 的函数，可记为

$$y=\mu(x)=E(Y|X=x) \tag{1.1}$$

用上式来描述变量 X 与 Y 之间的变化规律，称公式（1.1）为 Y 对 X 的回归方程，它所对应的曲线，称为回归曲线。

可以证明，对于任意函数 $\varphi(x)$，恒有

$$E[Y-\varphi(x)]^2\geqslant E[Y-\mu(x)]^2 *$$

---

* 在 $X=x$ 时有

$$\begin{aligned}E[Y-\mu(x)]^2&=E[Y-E(Y)]^2=E[(Y-\varphi(x))-(E(Y)-\varphi(x))]^2\\&=E[Y-\varphi(x)]^2-2E[(Y-\varphi(x))(E(Y)-\varphi(x))]\\&\quad+E[E(Y)-\varphi(x)]^2\\&=E[Y-\varphi(x)]^2-[E(Y)-\varphi(x)]^2\end{aligned}$$

$\because$ $[E(Y)-\varphi(x)]^2\geqslant 0$，

$\therefore$ $E[Y-\mu(x)]^2\leqslant E[Y-\varphi(x)]^2$

这说明，在一切 x 的函数中，只有回归函数 $\varphi(x)$ 作为 Y 的估计，才能使估计的偏差平方和达到最小，因此，回归方程(1.1)成为研究变量间相关关系的一个重要工具。

一般说来，从任意函数 $\varphi(x)$ 中，找一个使估计偏差最小的回归函数 $\mu(x)$ 来是比较困难的。另外，对条件分布 $E(Y|X=x)=\mu(x)$ 一般不知道，因而也无法确定条件数学期望 $\mu(x)$ 的值，故不从寻求 $\mu(x)$ 入手，而转向借助样本资料来判断回归函数 $\mu(x)$ 属于哪一类型函数，并来估计回归函数，这样得出的方程叫做经验方程。回归分析的基本内容就是怎样来估计 $\mu(x)$，然后利用这个估计结果作预测或控制。估计 $\mu(x)$ 的问题又称为求 y 对 x 的回归问题。

例 1. 大豆栽培试验中，测得株龄（周）x 与株高（厘米）y 的数据如下：

x_i： 1，2，3，4，5

y_i： 5,17,24,33,41

这里，x 是一般变量，y 是随机变量，求 y 对 x 的回归。

将每对观测值 (x_i, y_i) 在直角坐标系中描出，得图1～1，称为散点图。通过散点图可以粗略地了解用什么形式的函数估计随机变量 y 的数学期望来得好些，由图1～1看出它大致成一直线，

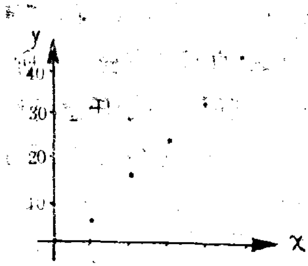

图1～1 散点图

故在本例中用 x 的线性函数 $\alpha+\beta x$ 来估计 y 的数学期望是适宜的。

用线性函数$\alpha+\beta x$来估计y的数学期望的问题，称为一元线性回归问题，这种估计问题，相当于对x的每一个值，假设$E(y)=\alpha+\beta x$，进一步我们还假设对于x的每一个值，$y\sim N(\alpha+\beta x,\sigma^2)$，其中$\alpha,\beta$及$\sigma^2$都是未知参数，并且都不依赖于$x$，对$y$作这样的正态假设，相当于设

$$y=\alpha+\beta x+\varepsilon \quad \varepsilon\sim N(o,\sigma^2) \tag{1.2}$$

上式就是所谓一元线性回归模型。

回归分析的主要问题是根据x，y的观测值

$$(x_i,\ y_i) \quad (i=1,2,\cdots,n)$$

给出α,β的估计值a,b，同时对a,b作统计检验，以便确定这些估计值的可靠程度。$\alpha+\beta x$的估计为$a+bx$，记作$\hat{y}$，而方程

$$\hat{y}=a+bx$$

称为y对x的线性回归方程或回归方程，其图形称为回归直线。

2. α,β的最小二乘估计

对于已知具有线性相关关系的两个变量X和Y，求它们之间的定量表达式，实际上就是对(1,2)中的α,β进行估计。α,β的任一对估计值a,b决定了一条回归直线

$$\hat{y}=a+bx$$

问题是在这些直线中哪一直线更能反映X,Y的线性相关关系呢？即怎样取α,β的估计值最合适？可采用最小二乘法来估计α,β。

设给定n个点(x_1,y_1)，$(x_2,y_2)\cdots$，(x_n,y_n)，那么，对于平面上任意一条直线l：$y=a+bx$，我们用数量