

施肥方案对作物、蔬菜的影响

喻梅 金青松 唐福明

教练: 刘来福

(北京师范大学数学系, 北京 100875)

摘要 对土豆和生菜, 分别建立了产量与施肥水平之间的多元二次回归模型。运用 SAS/STAT 软件依次采用全回归、逐步回归和二次响应面回归。在确认模型具完美适度性基础上, 进行线性相关、交互作用、最佳响应水平、强影响变量、回归曲面形状等分析。同时, 将两种作物进行比较, 得出一系列颇有实用价值的结论。分析结果表明: 土豆的产量对 N 具有强线性依赖性, 而生菜是对 P ; 施肥的交互作用对土豆影响较大, 对生菜则无强影响; 最佳施肥方案中 N, P, K 的用量土豆为 292, 246, 542 (公斤/公顷), 生菜为 213, 667, 427 (公斤/公顷) 对应产量为 45.18 和 23.13 吨/公顷, 且均在试验范围内达到, 可信性强; 对土豆, 强影响因子依次为 $N \rightarrow K \rightarrow P$, 对生菜为 $P \rightarrow N \rightarrow K$; 回归曲面上凸, 沿 $(N, P, K) = (1, 0, 0)$ 方向下降迅速。因此, 施肥中应特别注意 N 的使用量。

一、问题重述

对某地区作物进行施肥水平对产量影响的实验, 以土豆和生菜为例, 试验中, 分别取了 N, P, K 的十种水平, 在将其中两种的用量固定在第 7 水平时, 对第三种分别取十种水平。也就是, 对土豆和生菜分别给出了 30 组观测数据(见表 1)。其中, 可控制变量为 N, P, K 的施用量, 单位 kg/ha , 响应变量为产量, 单位 t/ha 。由此试验结果出发, 分析施肥量与产量之间关系, 并找出最优的施肥方案。

表 1 原始数据(以土豆为例)

序号	产量	N	P	K	序号	产量	N	P	K
1	18.98	259	196	0	16	40.09	259	147	372
2	27.35	259	196	47	17	41.26	259	196	372
3	34.86	259	196	93	18	42.17	259	245	372
4	38.52	259	196	140	19	40.36	259	294	372
5	38.44	259	196	186	20	42.73	259	342	372
6	37.73	259	196	279	21	15.18	0	196	372
7	38.43	259	196	372	22	21.36	34	196	372
8	43.87	259	196	465	23	25.72	67	196	372
9	42.77	259	196	558	24	32.29	101	196	372
10	46.22	259	196	651	25	34.03	135	196	372
11	33.46	259	0	372	26	39.45	202	196	372
12	32.47	259	24	372	27	43.15	259	196	372
13	36.06	259	49	372	28	43.46	336	196	372
14	37.96	259	73	372	29	40.83	404	196	372
15	41.04	259	98	372	30	30.75	471	196	372

从农作物栽培学的原理和经验,采用多元二次回归函数,一般可刻划施肥量与产量关系(见文献[1])。

二、假 设

1. 在实验中,除施肥量,其它影响因子:如环境条件、种植密度、土壤肥力等,均处于同等水平;

2. 各次实验独立,误差项均服从 $N(0, \sigma^2)$, $\sigma > 0$, 自变量 N, P, K 的观测无误差;

3. 符号说明:

Q ——产量;

N, P, K ——氮,磷,钾肥的施用量;

$\bar{Q}(\bar{N}, \bar{P}, \bar{K})$ —— $Q(N, P, K)$ 的样本均值;

$S_Q(S_N, S_P, S_K)$ —— $Q(N, P, K)$ 的样本标准差;

VIF——方差膨胀因子 (Variance Inflation Factor);

R ——复相关系数。

三、模型建立与模型分析

在固定其它生产条件下,仅考虑施肥量与产量间的关系,用三元二次多项式进行拟合。回归模型为

$$E(Q) = b_0 + b_N N + b_P P + b_K K + b_{NN} N^2 + b_{NP} NP + b_{NK} NK + b_{PP} P^2 + b_{PK} PK + b_{KK} K^2, \quad (1)$$

其观测为

$$Q_{n \times 1} = X_{n \times p} \beta_{p \times 1} + \epsilon_{n \times 1}. \quad (2)$$

其中, $Q_{n \times 1}$ 为可观测随机向量; $X_{n \times p}$ 为观测阵, 其中元素为原始数据中各变量及由其生成的二次项, 交互项; $\beta_{p \times 1}$ 为未知参数向量; $\epsilon_{n \times 1}$ 不可观测的随机误差, 服从 $N(0, \sigma^2 I_n)$ 。

对土豆和生菜分别讨论。(2)中, $n = 30, p = 10$ 。

(一) 全回归

1. 数据变换。以 VIF 作为检验共线性影响的标准(见文献[2]和[3])。记 $C = (c_{ij}) = (X'X)^{-1}$, $R(i)$ 为变量 x_i 对其余 $p - 1$ 个自变量的复相关系数, 则有

$$c_{ii} = (1 - R^2(i))^{-1} \quad (i = 1, 2, \dots, p),$$

称 c_{ii} 为变量 x_i 的方差膨胀因子。一般认为, 当 VIF 最大值接近或超过 10, 共线性显著, 最小二乘所得结果失真; 而当均趋于 1 时, 认为共线性影响很弱。

直接对原始数据拟合, VIF 最大值接近 10, 故将原始数据进行相关变换。即

$$\text{令 } Q' = (Q - \bar{Q})/S_Q, N' = (N - \bar{N})/S_N, P' = (P - \bar{P})/S_P, K' = (K - \bar{K})/S_K.$$

以下仍以 Q, N, P, K 记标准化后的变量。

2. 全回归 先以土豆为例。结果表明, 系数矩阵不满秩, 相关变换仅消除了一次项与平方项间共线性影响, 而交互项仍可表为其它项的线性组合。故仅对一次项和平方项

表 2 参数估计

Variable	Parameter Estimate	Standard Error	T for H_0 : Parameter = 0	Prob > T
INTERCEP	0.617546	0.09903863	6.235	0.0001
N	0.416751	0.06230654	6.689	0.0001
P	0.204971	0.06510116	3.148	0.0045
K	0.467305	0.06487691	7.203	0.0001
NN	-0.375899	0.03662213	-10.264	0.0001
NP	0	0.00000000	.	.
NK	0	0.00000000	.	.
PP	-0.108386	0.03873551	-2.798	0.0102
PK	0	0.00000000	.	.
KK	-0.454556	0.02853910	-16.010	0.0005

进行回归。参数估计与显著性检验结果见表 2

(1) 模型的适度性。VIF 均接近于 1, 相关变换后的模型基本上消除了多重共线性影响; 方差分析中, F 检验的 P 值为 0.0001, 回归方程极其显著, 真实值与回归值间 R 达 95.9%, 拟合度极高; 再由残差对观测的散点图, 残差均匀分布在零值两侧, 无系统偏差。因此, 模型是适度的。

(2) 参数估计。表 2 给出了参数估计值、估计方差及显著性检验结果。其中, t -检验的 P 值均小于 0.05, 参数显著不为 0, 含 N 的参数更为显著; 参数估计的方差均很小, 因此参数的置信区间相对于参数值很窄(如 b_{NN} 的 95% 置信区间上、下限为 (-0.375899 ± 0.0483)), 参数估计有较强实用价值。

(3) 相关分析。由 Q 与含 N 项间相关系数均很大, 如与 N 为 0.54, 与 NN 为 -0.62, 与 NK 为 0.72 等, 知土豆产量对含 N 项线性依赖性极强; 而对含 P 的较弱, 如和 P 为 0.12, 和 PP 为 0.07 等。 N, P, K 之间仅有极弱负相关, 如 N 和 P 为 -0.05, N 和 K 为 -0.05, P 和 K 为 -0.06。

(4) 预测与回判。模型可以给出对给定 N, P, K 水平下, Q 的预测值和预测区间。对原始数据作回判表明, 原始数据均落在置信水平为 0.95 的预测区间内。

对生菜可作类似分析, 其结果是: VIF 接近 1, 方差分析中 F 检验的 P 值为 0.0001, R 为 92.7%, 残差均匀分布, 无系统偏向, 模型适度; 参数推断中, 取显著性水平 0.05, $b_0, b_P, b_{NN}, b_{PP}, b_{KK}$ 显著不为 0, 参数估计方差均较小, P 的作用最为显著; 生菜的产量对含 P 项线性依赖性极强, N, P, K 之间仅有极弱的负相关; 原始数据均落入置信水平为 0.95 的预测区间。

从以上分析中初步得出: 土豆对 N 依赖强, 对 P 弱; 而生菜对 P 依赖强。这与土豆是块茎生长作物, 需 N 量高, 生菜是蔬菜作物, 需 P 量高的规律是吻合的。

(二) 逐步回归

逐步回归结果表明: 对于土豆, 首先进入模型的是 NK 项, 其次是 NN 项; 参数检验中, $b_{NN}, b_{NK}, b_{PK}, b_{KK}, b_{PP}$ 显著不为 0; 将保留在模型中和进入模型的显著性水平均置为 0.99 时, 一次项中只有 N 进入, 但在参数检验中不显著, 此时 R 仍为 95.9%; 同(一)

中进行适度检验,模型是适度的,且原始数据的回判结果表明,逐步回归模型与全回归模型是一致的。回归模型的不唯一性是由自变量间存在较强共线性所引起的。

因此,初步推断:施肥交互作用对土豆的产量影响较大,主要是 NK 和 PK 。逐步回归也再次证实(一)中结论:土豆对 N 依赖性较强。土豆产量的回归模型为

$$Q' = 0.73 + 0.12N' + 1.27N'K' + 0.88P'K' - 0.38(N')^2 - 0.11(P')^2 - 0.15(K')^2, \quad (3)$$

(3)式中, $Q' = (Q - 36.03)/7.73$, $N' = (N - 239.63)/94.47$,

$K' = (K - 341.03)/152.77$, $P' = (P - 179.6)/69.97$ 。

Q, N, P, K 为实际值。

对生菜,首先进入模型的是 P ,模型中包含 P, NN, K, PP, KK, N, R 达92.7%;将两种显著性水平均置为 0.99 时,交互项仍不能进入模型。说明施肥交互作用对生菜产量无显著影响;同时再次表明生菜对 P 依赖较强。生菜的回归模型为

$$Q' = 0.60 + 0.06N' + 0.70P' + 0.16K' - 0.34 \cdot (N')^2 - 1.40(P')^2 - 1.40(K')^2, \quad (4)$$

(4)式中, $Q' = (Q - 17.14)/4.18$, $N' = (N - 205.30)/79.99$,

$P' = (P - 358.53)/139.64$, $K' = (K - 341.03)/132.77$ 。

Q, N, P, K 为实际值。

逐步回归分析表明:交互作用对土豆和生菜是不同的。在建模时应考虑此因素,从而得到合理的模型。

(三) 二次响应面回归

在拟合过程中,用编码值使各变量水平在 $[-1, 1]$ 以消除共线性影响,而后拟合二次回归模型。

对土豆, R 仍为 95.9%。因子分析(见表 3)说明: N 的影响最大,其次是 K ,而 P 的影响相对很小。对回归曲面典型特征的分析表明:最佳响应水平为 $(N, P, K) = (292, 246, 542)$,最优产量为 45.18,最大值在试验范围内达到,可信性强;回归曲面上凸,沿 $(N, P, K) = (1, 0, 0)$ 方向下降迅速,沿着 $(N, P, K) = (0, 1, 0)$ 方向变化最为平缓。因此,当 K, P 取最佳水平时,产量对 N 在最佳水平附近取值敏感,而当 K, N 取最佳水平时,对 P 在最优值附近的波动不敏感。

表 3 因子分析

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio	Prob>F
N	2	1207.178634	603.589317	98.8	0.0000
K	2	621.884601	310.942300	50.895	0.0000
P	2	170.807278	85.403639	13.979	0.0001

对生菜, R 为 92.8%。因子分析中, P 影响最大, N 次之, K 最小。回归曲面上,最佳响应水平为 $(N, P, K) = (213, 667, 427)$ 对应最优产量 23.13,在试验范围内达到;沿 $(N, P, K) = (1, 0, 0)$ 方向下降迅速。故当 P, K 取最优量时,产量对 N 的取值敏感。

因此,在施肥中,应特别注意 N 的使用量不能过高。

四、模型的应用分析

前面已用统计方法讨论了土豆、生菜对 N, P, K 的不同反应,得出比较满意的结果。下面,我们对模型换个角度加以讨论。

(一) 边际产量(以土豆为例)

施肥是否增产,幅度如何?这是科学施肥管理中应明确的。从生产函数派生出的边际产量方程,说明了因营养元素投入的微小变化而引起产量的变化率或斜率。

$$\begin{cases} \frac{\partial Q}{\partial N} = S_Q \left[\frac{b_{NN}}{S_N} + \frac{b_{NK}}{S_N} \left(\frac{K - \bar{K}}{S_K} \right) - 2 \frac{b_{NN}}{S_N^2} (N - \bar{N}) \right] \\ \frac{\partial Q}{\partial P} = S_Q \left[\frac{b_{PK}}{S_P} \left(\frac{K - \bar{K}}{S_K} \right) - 2 \frac{b_{PP}}{S_P^2} (P - \bar{P}) \right] \\ \frac{\partial Q}{\partial K} = S_Q \left[\frac{b_{NK}}{S_K} \left(\frac{N - \bar{N}}{S_N} \right) + \frac{b_{PK}}{S_P} \left(\frac{P - \bar{P}}{S_P} \right) - 2 \frac{b_{KK}}{S_K^2} (K - \bar{K}) \right] \end{cases} \quad (5)$$

在土豆的回归方程中,各营养元素的交互作用,使任何一种的边际产量都包含自身和另两种元素的固定量。从方程组(5)中可看出,交互作用对土豆的影响较大。因土壤中本身所含三种营养元素有限,单独使用一种或两种对应的产量都不很高。换句话说,一种营养元素的生产率是受另两种与之配合使用的营养素高度制约的。

明确肥料用量对施肥效果的影响,有助于实现施肥的定量管理。现用生产函数的二阶导数作为指标进行分析。

$$\begin{cases} \frac{\partial^2 Q}{\partial N^2} = -2 \frac{b_{NN} S_Q}{S_N^2} = -0.000651 \\ \frac{\partial^2 Q}{\partial P^2} = -2 \frac{b_{PP} S_Q}{S_P^2} = -0.00034244 \\ \frac{\partial^2 Q}{\partial K^2} = -2 \frac{b_{KK} S_Q}{S_K^2} = -0.000135618 \end{cases} \quad (6)$$

我们把肥效作为作物对施肥量敏感程度的一个指标。它指的是在原有基础上,每增施单位肥料所增加的作物产量。从(6)式中可知,肥效随用量的递减速率是 $N > P > K$ 。因此,种植土豆时,钾肥可采用较高的施用量,而氮肥施用量不能过高。当(5)式中

$$\frac{\partial Q}{\partial N} = \frac{\partial Q}{\partial P} = \frac{\partial Q}{\partial K} = 0$$

时,得到最优产量及相应的 N, P, K 水平,与统计结果相符。

对生菜的边际产量和肥效递减性也可作同样讨论。

(二) 营养元素的最佳组合

对于土豆,从上面分析知,对 N, P, K 的交互作用很敏感。在考虑某一元素最佳投入时,必须考虑另两种对其的交互作用。当生产资源的边际替代率(即当产量一定时,一种营

养元素对另一种的导数)等于它们的价格之反比时,可取得资源或投入的最小成本组合。

$$\begin{cases} \frac{\partial P}{\partial N} = \frac{\partial Q}{\partial N} / \left(\frac{\partial Q}{\partial P} \right) = -\frac{P_N}{P_P} \\ \frac{\partial K}{\partial N} = \frac{\partial Q}{\partial N} / \left(\frac{\partial Q}{\partial K} \right) = -\frac{P_N}{P_K} \\ \frac{\partial K}{\partial P} = \frac{\partial Q}{\partial P} / \left(\frac{\partial Q}{\partial K} \right) = -\frac{P_P}{P_K} \end{cases} \quad (7)$$

从(7)中即可求得一定产量下使成本最小的营养元素量。

五、模型优缺点与改进

模型最大优点在于对原始数据拟合时,采用多种方法进行,使之愈来愈完善,具有很高的拟合精度和适度性。在此基础上,对模型作进一步讨论便可得到一系列可靠而实用的信息。并且,所得结论与客观事实很好地吻合,从而进一步说明模型是合理的。

在实际问题中,产量受作物种类、植株密度、施肥量、气候条件等各种因素的作用。我们仅考虑了施肥量影响,但稍加修改便能适合不同情况,如:

1. 考虑植株密度: 在原有数据基础上,加上一组植株密度变化数据,用同样方法建立四元二次模型,并加以讨论。

2. 土壤肥力影响: 在实际环境中,每块地肥力不等,有高产田与低产田之分。将土壤肥力也当作影响作物产量的一个因子,同样可进行分析。

在模型建立中,还可进行异常值检验,将其删除或加权,重新拟合后讨论。

参 考 文 献

- [1] 张乃生等,晋东南旱地玉米“产量——施肥”多元回归模型及其应用分析,数理统计与管理,1(1989),10—13.
- [2] 约翰·内特(美)等,应用线性回归模型,中国统计出版社,1990.
- [3] 北京大学概率统计系,SAS/STAT 软件“回归分析过程”,1991.
- [4] J·法兰士(英)等,农业中的数学模型,农业出版社,1991.
- [5] 厄尔 O·黑迪(美)等,农业生产函数,农业出版社,1991.

关于施肥效果分析问题的评注

项 可 风

(中国科学院系统科学研究所,北京 100080)

1992 年全国大学生数学模型竞赛,北京赛区共有 46 个队参赛,其中有 26 个队选做《施肥效果分析》题。我参加了本题的阅卷工作,总的情况很不错,都抓住问题的实质。应用回归方法去建立模型,而后用统计方法分析施肥效果。北京师范大学数学系队,获得北京赛区的特等奖,本期发表该队喻梅,金青松,唐福明等三位同学的文章,作为本题最优秀的一份答卷,供读者参阅。

下面就本次竞赛中被普遍忽视的几个问题提出一点看法。本文所使用符号与数据可