


# 统计模型——施肥效果分析

原创七七喝椰奶于 2023-09-22 20:08:11 发布阅读量1k收藏11点赞数

分类专栏：数学建模数学建模案例文章标签：数学建模

版权

数学建模 同时被 2 个专栏收录

4 订阅28 篇文章

订阅专栏

## 背景

某地区作物生长所需的营养主要是氮（N）、磷（P）、钾（K）。某作物研究所在该地区对土豆与生菜做了一定数量的实验，实验数据如表4-1记录了土豆的实验数据，表4-2记录了生菜的实验数据。其中hm2表示公顷，t表示吨，kg表示千克。当一个营养的施肥发生变化时，总将另外两个营养施肥量保持在第七个水平。如对土豆产量关于N的施肥做实验时，P和K的施肥量分别取196kg/hm2与372kg/hm2。

目标：试分析施肥量与产量之间的关系，并对所有结果从应用价值与如何改进方案方面做出估计。

表4-1土豆的N、P、K效果：

施肥量(N) Kg/hm2	产量 t/hm2	施肥量 (P) Kg/hm2	产量 t/hm2	施肥量(K) Kg/hm2	产量 t/hm2
0	15.81	0	33.46	0	18.98
34	21.36	24	32.47	47	27.35
67	25.72	49	36.06	93	34.86
101	32.29	73	37.96	140	39.52
135	34.03	98	41.04	186	38.44
202	39.45	147	40.09	279	37.73
259	43.15	196	41.26	372	38.43
336	43.46	245	42.17	465	43.87
404	40.83	294	40.36	558	42.77
471	30.75	342	42.73	651	46.22

表4-2生菜的N、P、K效果：

施肥量(N) Kg/hm2	产量 t/hm2	施肥量 (P) Kg/hm2	产量 t/hm2	施肥量(K) Kg/hm2	产量 t/hm2
0	11.2	0	6.39	0	15.75
28	12.70	49	9.48	47	16.67
56	14.56	98	12.46	93	16.89
84	16.27	147	14.33	140	16.24
112	17.75	196	17.10	186	17.56
169	22.59	294	21.94	279	19.20
224	21.63	391	22.64	372	17.97
280	19.34	489	21.34	465	15.84
336	16.12	587	22.07	558	20.11

392

14.11

685

24.53

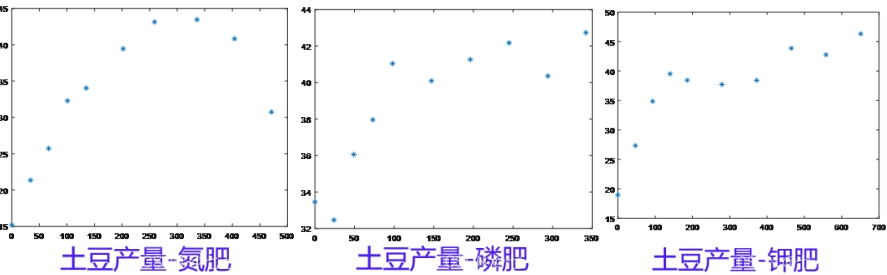
651

19.40

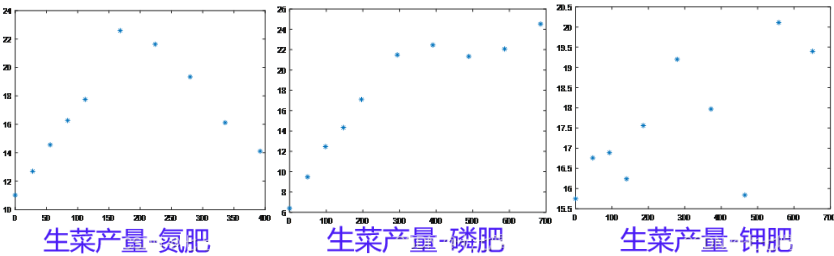
```
1 >> tn=[0 34 67 101 135 202 259 336 404 471];
2 yn=[15.18 21.36 25.72 32.29 34.03 39.45 43.15 43.46 40.83 30.75];
3 tp=[0 24 49 73 98 147 196 245 294 342];
4 >> yp=[33.46 32.47 36.06 37.96 41.04 40.09 41.26 42.17 40.36 42.73];
5 tk=[0 47 93 140 186 279 372 465 558 651];
6 yk=[18.98 27.35 34.86 39.52 38.44 37.73 38.43 43.87 42.77 46.32];
7 >> sn=[0 28 56 84 112 168 224 280 336 392];
8 xn=[11.02 12.70 14.56 16.27 17.75 22.59 21.63 19.34 16.12 14.11];
9 sp=[0 49 98 147 196 294 391 489 587 685];
10 >> xp=[6.39 9.48 12.46 14.33 17.10 21.49 22.46 21.34 22.07 24.53];
11 sk=[0 47 93 140 186 279 372 465 558 651];
12 xk=[15.75 16.76 16.89 16.24 17.56 19.20 17.97 15.84 20.11 19.40];
```

1、土豆产量与氮肥的关系

1.1 绘制散点图



以土豆产量与氮肥施肥量为例，初步观察，用二次函数作为经验方程。



1.2 回归模型

$$\begin{cases} y_n = a_0 + a_1 t_n + a_2 t_n^2 + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

【3.1】

1.3 模型参数求解与显著性检验

调用matlab的回归 **求解函数** regress,下面逐步解释。

```
1 >> clear %清除内存里的一切变量;
2 >> load d:\tudou %调回存储的数据;
3 >> [b,bt,r,rt,st]=regress(yn',[ones(length(tn),1),tn',(tn.^2)']) %调用格式
```

- b 回归系数，升幂排列
- bt 回归系数的置信区间，默认95%
- r 残差，即理论值与测量值的差
- rt 残差的95%的置信区间

- st 模型检验参数 (R2,F,P,sig2). (R2表示方程因变量与自变量之间的相关性检验, R也称为可决系数; F, p是方程显著性检验; sig2是σ2的估计值。)

regress      matlaba回归命令

yn' 因变量 (列)

[ones(length(tn),1),tn',(tn.^2)')] 矩阵, 形式是

$$\begin{bmatrix} 1 & x_1 & \cdots & x_1^p \\ 1 & x_2 & \cdots & x_2^p \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_n & \cdots & x_n^p \end{bmatrix}$$

回归系数

```
b =
14.7416 a0
0.1971 a1
-0.0003 a2
```

回归系数的95%的置信区间

```
bt =
12.6301 16.8532
0.1736 0.2207
-0.0004 -0.0003
```

回归系数的置信区间不应包括0; 若包括0, 说明该系数不显著 (100次实验, 95次以上离0非常近, 非常危险 (或者贡献不大) )

残差	残差的95%的置信区间	
r =	rt =	
0.4384	-1.9167	2.7934
0.3078	-2.5441	3.1596
-0.7065	-3.6736	2.2607
1.0998	-1.7844	3.9841
-1.1388	-3.9623	1.6847
-1.2616	-3.8888	1.3656
0.1228	-2.7587	3.0043
0.8080	-2.0382	3.6541
1.8571	-0.3055	4.0196
-1.5269	-2.5722	-0.4816

模型总体检验参数

```
st =
0.9863 251.7971 0.0000 1.6189
```

方程显著性指标F=251.7971>Fα(2,7)说明回归结果显著。

残差的置信区间以包含0为好，说明残差总在0附近，意味着残差较小。如果出现残差置信区间不包括零，这个样本观测值极可能是极端值。

残差平方和记为SSE,利用SSE/(n-p-1)可以作为σ<sup>2</sup>的无偏估计值。

在该问题里，R<sup>2</sup>=0.9863表示控制氮肥施肥量，在100次种植实验里，至少有98次土豆的产量可以用该经验公式解释

1.4 模型修正或改进

根据上面的计算，发现第10个样本的残差置信区间右端点小于0，即第10个样本可能不正常。去掉第10个样本后（可以再线性插值替换），回归得到结果为R<sup>2</sup>=0.9956;F=678.5246;p<0.000

回归方程极显著：
$$Y_n \sim N(15.3653 + 0.1825t_n - 0.0003t_n^2, 0.6036)$$

1.5 预测预报

根据上面的分析当土豆的施肥量t<sub>n</sub>给定时，土豆的产量Y<sub>n</sub>服从分布

$$Y_n \sim N(15.3653 + 0.1825t_n - 0.0003t_n^2, 0.6036)$$

$$\hat{\sigma}^2 = \frac{SS_e}{10-2-1} = 0.6036$$

其中σ<sup>2</sup>根据残差平方和作点估计

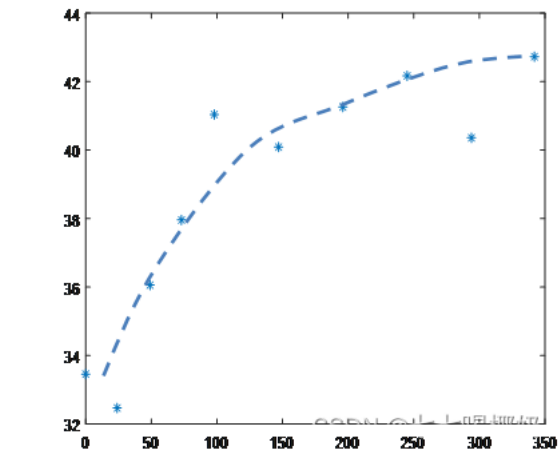
那么，在土豆的种植中，给定每公顷施肥量为t<sub>n</sub>=380kg时，土豆产量Y<sub>n</sub>的分布为

$$Y_n \sim N(41.3953, 0.6036)$$

即土豆产量的95%的置信区间为( 39.8725 , 42.9181)

2、土豆产量与磷肥的关系（1）

2.1 模型猜测



磷肥施肥量过了100以后，土豆的产量几乎没有怎么增加，即土豆与磷肥的关系里，有一个上界，下面几个经验函数可以试试：

1. 
$$y = \frac{\beta_2}{\beta_1 e^{x+\beta_3}}, y = \beta_1(1 - e^{-\beta_2 x})$$
 (指数函数)

2. 
$$y = \frac{x + \beta_1}{\beta_2 x + \beta_3}$$
 (双曲函数)

3. 
$$y = \frac{\beta_1}{1 + \beta_2 e^{-(x+\beta_3)}}$$
 (Logistic曲线)

4. 
$$y = \beta_1 + \beta_2 \ln(x + \beta_3)$$
 (对数函数)

5.  $y = pm(x)$  (多项式函数)

6.

$$y = \beta_1(x + \beta_2)^{\beta_3}$$

(幂函数)

不管用哪个，都是非线性回归，误差和可信度分析都不如线性回归那么成熟好用。

2.2 数据预处理

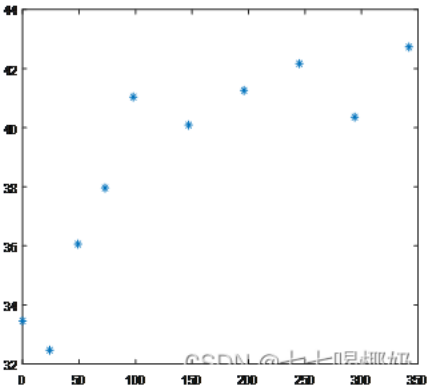
(1) 阅读数据和散点图发现，磷肥施肥量为24kg/hm2时的土豆产量反而不如不施肥的产量，认为第二个数据被污染了，用第一和第三个数据线性插值替代。

$$y_p(2) = \frac{y_p(1) + y_p(3)}{2}$$

上述几个非线性模型，不管用哪个，都会涉及到自变量取负指数函数，由于自变量太大，取负指数后在0附近，太敏感。所以先对磷肥施肥量标准化：

$$t_p' = \frac{t_p - \bar{t}_p}{\sigma(t_p)}$$

2.3 经验公式线性化



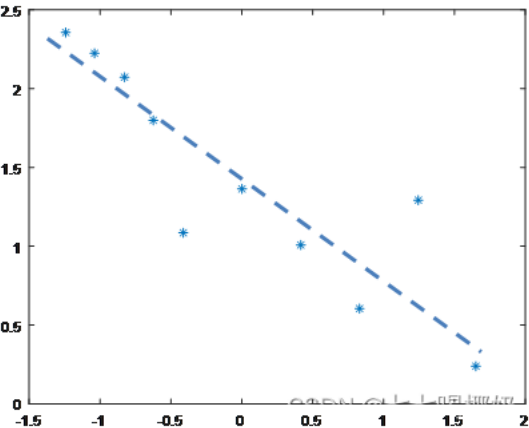
如上图所示，土豆的最大产量不超过44，而韦布函数  $y = A - Be^{-Cx}$  的最大值就是A，不妨假定A=44，则

$$y = A - Be^{-Cx} \Rightarrow 44 - y = Be^{-Cx} \Rightarrow \ln(44 - y) = \ln B - Cx$$

$$Y = a_0 + a_1x$$

取 $Y = \ln(44 - y)$ ,  $a_0 = \ln B$ ,  $a_1 = -C$ , 则韦布函数就线性化为 绘制 $(t_p', Y)$ 散点图验证猜测。

如下图所示，猜测正确。



2.4 建立模型

$$\begin{cases} Y_p = 44 - Be^{-C \cdot t_p'} + \varepsilon \\ t_p' = \frac{t_p - \bar{t}_p}{\sigma(t_p)}, \varepsilon \sim N(0, \sigma^2) \end{cases}$$

Y<sub>p</sub>表示施磷肥时土豆的产量，t<sub>p</sub>表示磷肥的施肥量，模型设为 【3.2】

$$\begin{cases} Y = a_0 + a_1 t_p' + \varepsilon' \\ \varepsilon' \sim N(0, \sigma_1^2) \end{cases}$$

取Y=ln(44-Y<sub>p</sub>),a<sub>0</sub>=lnB,a<sub>1</sub>=-C, 【3.2】就等价于下面的线性回归模型 【3.3】

2.5 参数求解

matlab计算 代码如下

```
1 >> clear
2 >> load d:\tudou
3 >> yp(2)=(yp(1)+yp(3))/2;
4 >> tp1=(tp-mean(tp))/std(tp);
5 >> Y=log(44-yp);
6 >> plot(tp1,Y,'*')
7 >> [mean(tp),std(tp)]
8 ans =
9 146.8000 118.2547
10 >> [a,at,r,rt,st]=regress(Y',[ones(length(tp),1),tp1']);
11 >> a
12 a =
13 1.4041
14 -0.6211
15 >> at
16 at =
17 1.1495 1.6587
18 -0.8895 -0.3527
19 >> st
20 st =
21 0.7807 28.4829 0.0007 0.1219
22 >> rt
23 rt =
24 -0.5561 0.9161
25 -0.5894 0.9383
26 -0.6351 0.9434
27 -0.8102 0.8236
28 -1.2249 0.0744
29 -0.8761 0.7971
30 -0.9568 0.6814
31 -1.0436 0.4756
32 0.1942 1.1279
33 -0.8104 0.5307
```

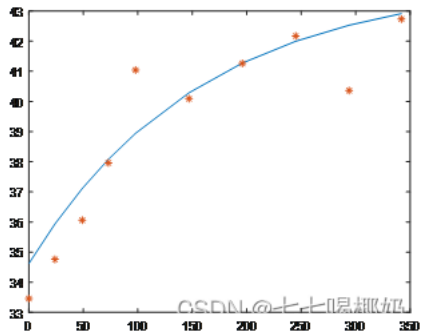
第9个样本异常，修正后计算

```
1 >> Y(9)=[];tp1(9)=[];
2 [a,at,r,rt,st]=regress(Y',[ones(length(tp1),1),tp1']);
3 >> st
4 st =
5 0.9154 75.7884 0.0001 0.0536
6 (极显著)
7 a =
8 1.3133
9 -0.7467
```

回归结果为

$$\begin{cases} Y_p = 44 - 3.7184e^{-0.0063(t_p - 146.8)} + \varepsilon \\ \varepsilon \sim N(0, 1.6346) \end{cases}$$

实验值与理论值对比如下图所示。



理论值	实验值
34.6242	33.4600
35.9398	34.7600
37.1144	36.0600
38.0806	37.9600
38.9432	41.0400
40.2863	40.0900
41.2726	41.2600
41.9970	42.1700
42.5290	40.3600
42.9129	42.7300

3、土豆产量与磷肥关系（2）

在上面的土豆产量与磷肥关系（1）里，我们假定土豆的最大产量为44后才得以线性化（这个假设有时候不太合理），如果不能线性化，就只能非线性拟合。

3.1 经验模型

$$\begin{cases} y = \beta_1 + \beta_2 \ln(x + \beta_3) + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

3.2 模型求解

调用matlab的最小二乘曲线拟合函数lsqcurvefit,来求解模型参数beta。先编写模型的经验函数的m文件tpfun.m。

```
Editor - D:\MATLAB\R2017a\bin\tpfun.m
tpfun.m x +
function y=tpfun(beta, x)
y=beta(1)+beta(2)*log(x+beta(3));
```

调用格式为：

```
1 >> clear
2 >> load d:\tudou
3 >> x=tp';y=yp';
4 >> [beta,res]=lsqcurvefit(@tpfun,[40,0.2,2],x,y)
```

- [beta,res]=lsqcurvefit(@tpfun,[40,0.2,2],x,y)
- beta为返回的参数的估计值，按照tpfun.m中顺序给出
  - res 返回残差平方和，即经验值与实验值之间的误差平方和

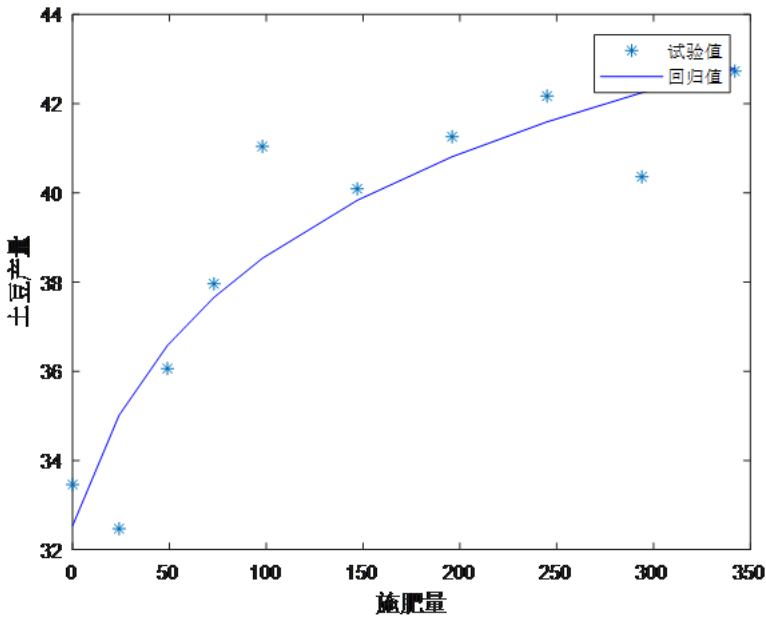
- @tpfun 调用函数名为tpfun.m的m文件
- [40,0.2,2] 是参数beta的初始值
- X 土豆试验中的磷肥施肥量（列向量）
- Y 土豆实验中的土豆产量(列向量)

计算结果为

```
1  beta =
2      19.4643    3.9464   27.4523
3  res =
4      18.1411
```

即土豆对磷肥的拟合方程为： $\hat{y} = 19.4643 + 3.9464\ln(x + 27.4523)$

3.3 误差分析

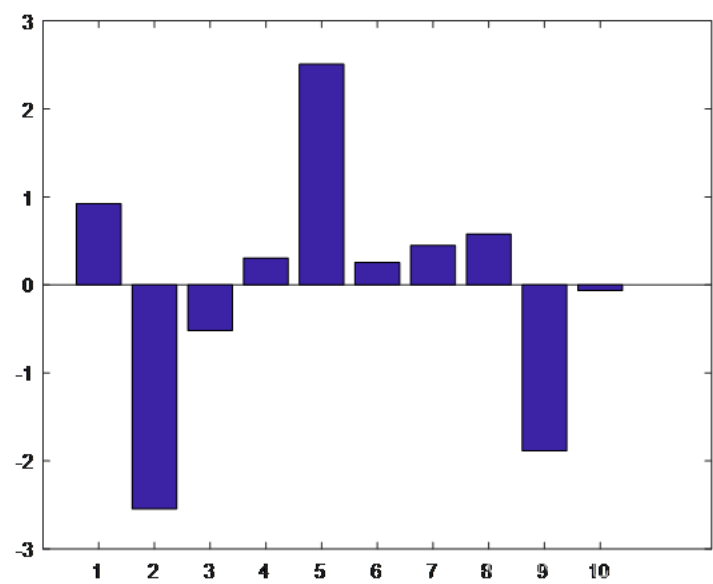


土豆对磷肥的拟合误差为  $E = y - \hat{y}$

经验值与实验值对比图像

1	实验值y	回归值y*
2	33.4600	32.5365
3	32.4700	35.0157
4	36.0600	36.5785
5	37.9600	37.6559
6	41.0400	38.5330
7	40.0900	39.8342
8	41.2600	40.8111
9	42.1700	41.5935
10	40.3600	42.2462





土豆产量的实验值与拟合值差的分布图

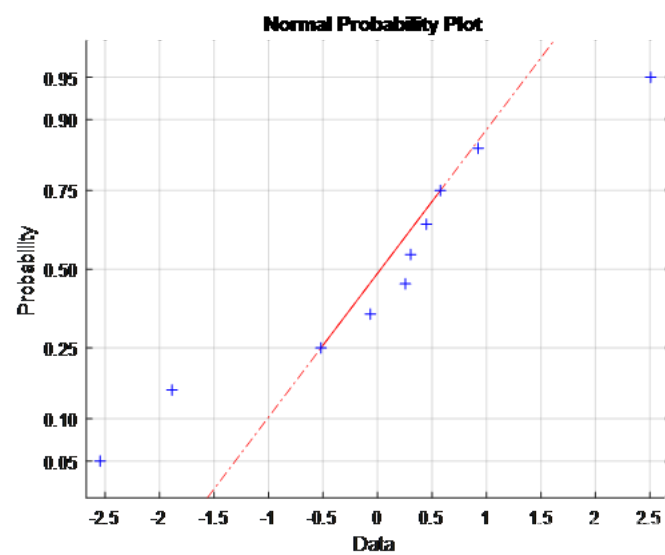
```
1 >> E=y-tpfun(beta,x)
2 E =
3     0.9235
4    -2.5457
5    -0.5185
6     0.3041
7     2.5070
8     0.2558
9     0.4489
10    0.5765
11    -1.8862
12    -0.0654
```

3.4 正态性假设检验

$H_0 : E \sim N(0, \sigma^2)$

(1) 绘制正态概率图

```
>> normplot(E)
```



除了3个点外，其余的点都均匀分布在直线两侧，即数据近似服从正态分布

(2)Jbtest正态检验

```
1 >> [h,p,jbstat,critval]=jbtest(E,0.05)
2 h =
3     0
4 p =
5     0.5000
6 jbstat =
7     0.1221
8 critval =
9     2.5239
```

Jbtest检验：h=1时拒绝原假设H0；当jbstat>critval时拒绝原假设。计算结果显示，没有理由拒绝H0，即接受H0。  
认为E服从均值为0的正态分布。

显示推荐内容



七七喝椰奶