SAS软件应用实习4

实验要求：

1. 一年多点和多年多点试验的统计分析；

2. 正交设计试验统计分析;

3. 数据转换与含缺区试验结果分析;

4. 双变量直线回归分析;

1. **一年多点试验的方差分析**

（资料来源：袁志发、贠海燕主编《试验设计与分析》page139）

【例3-6-1】



Data maize1;

Do V=1 to 5;

Do U=1 to 4;

Do R=1 to 2;

Input y@@; output;

End; end; end;

Cards;

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 6 | 7 | 7 | 8 | 10 | 12 | 5 | 6 |
| 9 | 8 | 9 | 9 | 9 | 10 | 8 | 9 |
| 7 | 8 | 3 | 4 | 4 | 2 | 5 | 4 |
| 14 | 14 | 20 | 18 | 13 | 14 | 15 | 15 |
| 8 | 6 | 9 | 5 | 8 | 8 | 7 | 6 |

;

Proc print; run;

Proc anova;

Class V U R;

Model y= R(U) U V V\*U;

Means V V\*U/duncan;

Run;

注：此处将地点和品种按固定因素对待！

1. **多年多点试验的方差分析**

（资料来源：袁志发、贠海燕主编《试验设计与分析》page144）

【例3-6-2】



Data maize2;

Do w=1 to 3;

Do V=1 to 5;

Do U=1 to 4;

Do R=1 to 2;

Input y@@; output;

End; end; end; end;

Cards;/\*3年4地5品种2区组\*/

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 6 | 7 | 7 | 8 | 10 | 12 | 5 | 6 |
| 9 | 8 | 9 | 9 | 9 | 10 | 8 | 9 |
| 7 | 8 | 3 | 4 | 4 | 2 | 5 | 4 |
| 14 | 14 | 20 | 18 | 13 | 14 | 15 | 15 |
| 8 | 6 | 9 | 5 | 8 | 8 | 7 | 6 |
| 5 | 6 | 7 | 6 | 13 | 11 | 6 | 6 |
| 15 | 13 | 14 | 11 | 13 | 15 | 10 | 11 |
| 9 | 9 | 3 | 5 | 6 | 4 | 3 | 4 |
| 16 | 12 | 16 | 18 | 15 | 11 | 14 | 13 |
| 7 | 8 | 8 | 8 | 6 | 7 | 9 | 8 |
| 6 | 4 | 8 | 6 | 9 | 10 | 7 | 5 |
| 10 | 11 | 10 | 11 | 9 | 10 | 10 | 10 |
| 7 | 7 | 3 | 3 | 4 | 5 | 4 | 4 |
| 13 | 12 | 19 | 18 | 17 | 14 | 16 | 14 |
| 6 | 8 | 7 | 6 | 8 | 4 | 8 | 6 |

;

Proc print; run;

Proc anova;

Class W V U R;

Model y= R(W\*U) W U W\*U V V\*W V\*U V\*U\*W;

Test H=U e= U\*W;

Test H=V e= V\*W;

Test H=U\*V e= V\*U\*W;

Run;

注：方差分析模型为混合模型，地点（U）和品种（V）为固定因素，年份（W）为随机因素。

关于交互效应的分析待后补！

1. **正交试验设计(**Orthogonaldesign**)的方差分析**

（资料来源：袁志发、贠海燕主编《试验设计与分析》page153）

【例3-7-1】啤酒酵母的最适自溶条件试验（无重复值的正交设计试验）

Data OD1;

Input A& B& C& O& x@@;

Cards;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 1 | 6.25 |
| 1 | 2 | 2 | 2 | 4.97 |
| 1 | 3 | 3 | 3 | 4.54 |
| 2 | 1 | 2 | 3 | 7.53 |
| 2 | 2 | 3 | 1 | 5.54 |
| 2 | 3 | 1 | 2 | 5.50 |
| 3 | 1 | 3 | 2 | 11.40 |
| 3 | 2 | 1 | 3 | 10.90 |
| 3 | 3 | 2 | 1 | 8.95 |

;

Proc anova;

Class A B C O;

Model x=A B C;

Run;

补充：经检验A列方差检验显著，可添加Means A/Duncan;语句进行多重比较。

**【例3-7-4】金属去油去锈试验（无重复值的正交设计试验）**

（资料来源：袁志发、贠海燕主编《试验设计与分析》page157）

Data OD2;

Input c1-c7 x;

Cards;

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7.7 |
| 1 | 1 | 1 | 2 | 2 | 2 | 2 | 6.1 |
| 1 | 2 | 2 | 1 | 1 | 2 | 2 | 6.0 |
| 1 | 2 | 2 | 2 | 2 | 1 | 1 | 17.7 |
| 2 | 1 | 2 | 1 | 2 | 1 | 2 | 17.3 |
| 2 | 1 | 2 | 2 | 1 | 2 | 1 | 10.5 |
| 2 | 2 | 1 | 1 | 2 | 2 | 1 | 13.3 |
| 2 | 2 | 1 | 2 | 1 | 1 | 2 | 16.2 |

;

Proc anova;

Class C1-C7;

Model x=C1-C7;

Run;

根据MS大小，修改model x=C1 C2 C5 C6;

**设置重复的正交试验**

【例3-7-5】L16（45）正交试验结果，完全随机重复3次。

（资料来源：袁志发、贠海燕主编《试验设计与分析》page158）

Data OD3;

Input c1-c7@@;

Do r=1 to 3;

Input x@@; output;

end;

Cards;

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 1 | 1 | 2.0 | 2.0 | 2.0 |
| 1 | 2 | 2 | 2 | 2 | 4.0 | 4.5 | 4.0 |
| 1 | 3 | 3 | 3 | 3 | 5.5 | 6.0 | 6.0 |
| 1 | 4 | 4 | 4 | 4 | 6.0 | 6.5 | 6.7 |
| 2 | 1 | 2 | 3 | 4 | 6.3 | 6.5 | 6.7 |
| 2 | 2 | 1 | 4 | 3 | 5.1 | 4.8 | 4.6 |
| 2 | 3 | 4 | 1 | 2 | 7.0 | 7.4 | 7.2 |
| 2 | 4 | 3 | 2 | 1 | 8.0 | 8.5 | 8.7 |
| 3 | 1 | 3 | 4 | 2 | 7.0 | 7.1 | 7.3 |
| 3 | 2 | 4 | 3 | 1 | 8.4 | 8.5 | 8.9 |
| 3 | 3 | 1 | 2 | 4 | 6.5 | 6.3 | 6.1 |
| 3 | 4 | 2 | 1 | 3 | 7.0 | 7.3 | 7.1 |
| 4 | 1 | 4 | 2 | 3 | 5.0 | 4.5 | 4.7 |
| 4 | 2 | 3 | 1 | 4 | 6.0 | 6.5 | 6.7 |
| 4 | 3 | 2 | 4 | 1 | 8.5 | 8.5 | 8.7 |
| 4 | 4 | 1 | 3 | 2 | 7.0 | 6.5 | 6.9 |

；

Proc print; run;

Proc anova;

Class C1-C5 R;

Model x=C1-C5;

Run;

补充：

如果重复是按区组排列的，则模型中增加区组因素即可（Model x=C1-C5 R;）。

合并误差模型: model x=C1-C4;

如果要进行多重比较，模型下面添加语句means C1-C4/Duncan;

**混合水平的正交试验**

【例3-7-6】安排混合水平（41X23）正交试验并分析交互作用

（资料来源：袁志发、贠海燕主编《试验设计与分析》page160）

Data OD5;

Input new C1-C15& x@@;

Cards;

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.41 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0.25 |
| 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 0.37 |
| 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 0.30 |
| 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 0.13 |
| 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 0.25 |
| 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 0.08 |
| 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 0.31 |
| 3 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 0.34 |
| 3 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 0.58 |
| 3 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 0.39 |
| 3 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 0.51 |
| 4 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 0.29 |
| 4 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 0.48 |
| 4 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 0.35 |
| 4 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 0.44 |

;

Proc print; run;

Proc anova ;

Class new C1-C15;

Model x= new C1-C15;

Run;

补充：结合混合正交设计，将相应的平方和和自由度进行合并，并做F测验。

**拟水平的正交试验**

【例3-7-7】用高效液相色谱法测定食品中β-胡萝卜素的研究。

（资料来源：袁志发、贠海燕主编《试验设计与分析》page162）

Data OD5;

Input C1-C5 x@@;

Cards;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 1 | 1 | 90.5 |
| 1 | 2 | 2 | 2 | 2 | 90.0 |
| 1 | 3 | 2 | 3 | 3 | 95.0 |
| 2 | 1 | 1 | 2 | 3 | 85.0 |
| 2 | 2 | 2 | 3 | 1 | 92.0 |
| 2 | 3 | 2 | 1 | 2 | 75.0 |
| 3 | 1 | 1 | 3 | 2 | 100.0 |
| 3 | 2 | 2 | 1 | 3 | 80.0 |
| 3 | 3 | 2 | 2 | 1 | 90.0 |

;

Proc print; run;

Proc anova ;

Class C1-C5;

Model x=C1-C5;

Run;

根据需要修改模型为：Model x=C1 C3 C4;

如果需要多重比较，增加语句：means C1 C4/Duncan;

注意：C1方差分析检验不显著，但多重比较的结果并非如此，这种情况在F检验接近显著水平时比较常见。

1. **满足方差分析基本假定的检验方法，及数据转换**

方差分析要满足三个基本假定：处理效应线性可加，误差具有独立性、正态性和同质性。

具体做法：误差正态性一般会满足。（1）处理效应线性可加，可通过均值与变异数的二维图进行观察；（2）方差同质性假定可以通过Bartlette卡方检验。

如果处理效应与误差存在关联，或者方差异质，数据转换常用平方根转换、对数转换和反正弦转换。具体的SAS函数如下：

SQRT(x) 求x的平方根；

LOG10(x) 求x的常用对数；

ARSIN(y) 计算函数y=sin(x)在区间的反函数，y取[-1,1]间值。

**均值与变异数（方差或标准差，极差）作图**

【表3-8-2】

Data new1;

Do A=1 to 5;

Do i=1 to 6;

Input x@@; output;

End; end; drop i;

Cards;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 6 | 8 | 7 | 7 | 10 | 8 |
| 4 | 6 | 6 | 3 | 5 | 6 |
| 6 | 4 | 4 | 5 | 3 | 2 |
| 7 | 4 | 6 | 6 | 3 | 5 |
| 9 | 4 | 5 | 7 | 7 | 6 |

;

PROC means;

CLASS A;

VAR x;

OUTPUT OUT=stat MEAN=Xmean range=Xrange STD=Xstd;

PROC gplot DATA=stat;

plot Xrange\*Xmean;

plot Xstd\*Xmean;

RUN;

Proc print; run;

说明：不过在Excel表中做二维图更简单些。请尝试！

数据经反正弦转换后进行方差分析

表3-8-4

Data zhuanhuan;

Do A=1 to 4;

Do i=1 to 4;

Input x@@;

Y=SQRT(x); output;

End; end; drop i;

Cards;

|  |  |  |  |
| --- | --- | --- | --- |
| 2152 | 1688 | 1508 | 1260 |
| 308 | 244 | 628 | 208 |
| 68 | 124 | 348 | 64 |
| 78 | 104 | 308 | 80 |

;

Proc print; run;

Proc anova;

Class A;

Model x y=A;/\*比较数据转换前后的差异\*/

Run;

1. **含缺区的试验分析**

录入数据时将缺区数据用“.”取代，调用Proc glm;进行方差分析。具体实例略。

1. **双变量直线回归分析**

具有线性回归功能的SAS过程有REG和GLM，其中REG专门用于回归分析。

REG过程的语法格式如下：

PROC REG [DATA=<数据集名> [选项] ] ;

　MODEL 应变量名=自变量名列/ [选项] ;

　[VAR 变量名列;

　FREQ 变量名;

　WEIGHT 变量名;

　BY 变量名列;

　OUTPUT <OUT=新数据集名 关键字=新变量名> ...;

　PLOT <纵坐标变量\*横坐标变量[=绘图符号]...> / [选项];

...

可选语句共有16项，具体可参见帮助。

【过程选项】

OUTEST=数据集名 指定统计量和参数估计输出的新数据集名。

NOPRINT 禁止统计结果在OUTPUT视窗中输出。

SIMPLE 输出REG过程中所用的每个变量的基本统计量。

CORR 输出MODEL语句或VAR语句中所列变量的相关矩阵。

ALL 等价于MODEL语句加上全部选项，即输出该语句所有选项分析结果。

MODEL语句中常用的选项有：

NOINT 在模型中不拟合常数项。

STB 输出标准化回归系数。

CLI 输出个体预测值的95%可信区间上下限。

CLM 输出因变量期望值(均值)的95%可信区间上下限。

R 输出个体预测值、残差及其标准误。

P 输出实际值Yi、预测值 和残差等。如已选择了CLI、CLM和R，则无需选择P。

PLOT语句用于输出变量间的散点图，其用法和GPLOT过程中的PLOT语句非常相似。PLOT语句定义的两变量可为MODEL语句或VAR语句中定义的任何变量。SYMBOL选项可定义散点图中点的标记，如SYMBOL=’\*’，则每个点以“\*”表示。

|  |  |
| --- | --- |
| 【例4-1-3】 |  |
| DATA new;  INPUT x y;  CARDS;   |  |  | | --- | --- | | 70 | 1616.3 | | 67 | 1610.9 | | 55 | 1440.0 | | 52 | 1440.7 | | 51 | 1423.3 | | 52 | 1471.3 | | 51 | 1421.8 | | 60 | 1547.1 | | 64 | 1533.0 |   ;  RUN;  PROC GPLOT data=new;  PLOT Y\*X;  RUN;  PROC REG DATA=new;  MODEL Y = X;  RUN; | 给出预测值并绘制图形：  PROC REG DATA=new;  MODEL Y = X/clm cli;  Plot y\*x /Conf95;  RUN;  将回归估计值和因变量预报值的0.95置信区间绘在一张图上：  symbol1 ci=black i=rlclm95 width=1;  symbol2 ci=red i=rlcli95 width=2;  proc gplot;  plot y\*x y\*x/overlay;  run; |

**有重复值的双变量直线回归**

（资料来源：《试验设计与分析》袁志发、周静芋主编Page 158）

data new;

input x@@;

do i=1 to 2;

input y@@; output;

end;

cards;

49.0 16.6 16.7

49.3 16.8 16.8

49.5 16.8 16.9

49.8 16.9 17.0

50.0 17.0 17.1

50.2 17.0 17.1

;

proc print; run;

PROC GPLOT data=new;

PLOT Y\*X;

RUN;

PROC REG DATA=new;

MODEL Y = X/noint;

RUN;

proc glm;

class x;

model y=x;

run;

根据REG过程输出的结果，可以获得回归Qe=0.03197，dfe=10；

根据glm方差分析获得的y真正误差为Qe1=0.025，dfe1=6；

据此可以进行失拟性检验：~F(4,6)

**data** new;

pr=probf(**0.396**,**4**,**6**);

putpr=;

**run**;

运行结果：pr=0.1945569943，失拟性检验不显著。

或者与下面的F0.05（4,6）结果比较：

data new;

F=FINV(0.95,4,6);

put F=;

run;

运行结果：F=4.5336769503

然后可以进行直线回归检验（略）。

**作业：**

Page172-173: 16，17，18，19，20（或21）

Page 228: 5,6