

实验8、多因素方差分析和协方差分析

年级：15级 专业：生信 学号：1513401013 姓名：郑磊

编号 一 二 三 四 总分 评阅人

得分

软硬件平台：

1. 硬件平台：（硬件配置）i5，2.9HZ处理器，16G内存，64位操作系统
2. 系统平台：（操作系统及其版本号）Windows10 企业版
3. 软件平台：（软件系统及其版本号，若是在线分析平台，还需要提供URL地址）R3.4.1 ， Rstudio

一、目的要求：

- 1、加深对多因素交互作用的理解；
- 2、从统计学角度来理解基因表达谱中设定的不同实验因素对基因转录的影响；
- 3、熟悉并掌握多因素方差分析和协方差分析中所涉及的R语言函数和脚本。

二、实验内容：

1、基因表达谱数据中的双因素方差分析

根据教师提供的基因表达谱数据（GDS6100），进行双因素方差分析；该数据表中一共涉及两种不同的因素，分别是protocol和time。

1.1、从该基因表达谱数据表中随机抽取一行数据x，注意行列转换问题；

```
dir="D:/RFile/实验八"

setwd(dir)

library(GEOquery)

gds6100 <- getGEO("GDS6100",destdir=dir)

#查看数据类型
```

```

mode(gds6100)

#查看注释信息

Meta(gds6100)

#查看列注释信息=》用来确定哪些列是肿瘤，哪些列是正常对照

Columns(gds6100)

#提取数据

data<-Table(gds6100)

#查看数据表的列名

colnames(data)

#查看数据表行列数

ncol(data)

nrow(data)

#前面两列是标题列，分别为探针id和基因名称

#3：8列是 miRNA135b transfected， 9：14列是 scrambled transfected

rownames(data)<-data[,1]

#随机抽取至少1列数据

n=1

#按行随机抽样

row.names<-rownames(data)

sam.row.name <- sample(row.names,n,replace=F)

sam.row.name

```

1.2、为该行数据设定两个因子，因素A为protocol、因素B为time，两个因素的

具体信息使用Columns函数查看；

```
ge<-data.frame(x=-t(data[sam.row.name,3:14]), A=-  
factor(substring(COLUMNS(gds6100)$protocol,1,9)), B=-  
factor(substring(COLUMNS(gds6100)$time,6)))  
  
ge
```

1.3、对数据x进行正态性检验（**shapiro.test函数**）；

```
#正态性检验
```

```
shapiro.test(x)
```

1.4、分别对数据x和因素A、B之间进行方差齐性检验（**bartlett.test函数**）；

```
#方差齐性检验
```

```
bartlett.test(x~A,data=ge)
```

```
bartlett.test(x~B,data=ge)
```

1.5、绘制直方图（plot函数）和交互作用图（**interaction.plot**），查看数据分布规律；

```
#绘图查看数据分布规律
```

```
par(mfrow=c(2,2),las=2, cex.axis=1.2, cex.lab=1.2)
```

```
plot(x~A+B,data=ge)
```

```
interaction.plot(A,B,x,legend=F)
```

```
interaction.plot(B,A,x,legend=F)
```

1.6、在不考虑交互作用的情况下，进行双因素方差分析（**aov函数**），并查看统计分析结果摘要（**summary函数**）；

```
#不考虑交互作用
```

```
ge.aov<-aov(x~A+B,data=ge)
```

```
summary(ge.aov)
```

1.7、在考虑交互作用的情况下，进行双因素方差分析（aov函数），并查看统计分析结果摘要（summary函数）；

```
#考虑交互作用
```

```
ge.aov2<-aov(x~A*B,data=ge)
```

```
summary(ge.aov2)
```

1.8、重复1.1~1.7过程，找到以下3种数据行：

(1) 齐方差、双因素方差分析中的所有 $p > 0.1$ ；

(2) 齐方差、F检验，双因素方差分析中的所有 $p < 0.1$ ，但交互作用 $p > 0.1$ (即无明显交互作用)；

(3) 齐方差、F检验，双因素方差分析中的所有 p 无要求，但交互作用 $p < 0.1$ (即有明显交互作用)。

(4) 然后将上述3种数据行的1.5步绘图结果，进行对比分析和讨论。

三、实验结果：

1.1

```
> sam. row. name
```

```
[1] "ILMN_3294213"
```

1.2

	ILMN_3294213	A....factor.substring.Columns.gds6100..protocol..1..9..	B....factor.substring.Columns.gds6100..time..6..
GSM1394594	9.20695	miRNA135b	12
GSM1394595	9.21754	miRNA135b	12
GSM1394596	8.73678	miRNA135b	24
GSM1394597	9.18711	miRNA135b	24
GSM1394598	9.95496	miRNA135b	36
GSM1394599	8.62198	miRNA135b	36
GSM1394600	8.93269	scambled	12
GSM1394601	9.05696	scambled	12
GSM1394602	8.59094	scambled	24
GSM1394603	9.04363	scambled	24
GSM1394604	9.50025	scambled	36
GSM1394605	8.84070	scambled	36

1.3

```
> shapiro.test(x)
```

Shapiro-Wilk normality test

data: x

W = 0.93035, p-value = 0.3839

1.4

```
> #方差齐性检验
```

```
> bartlett.test(x~A, data=ge)
```

Bartlett test of homogeneity of variances

data: x by A

Bartlett's K-squared = 0.88026, df = 1, p-value = 0.3481

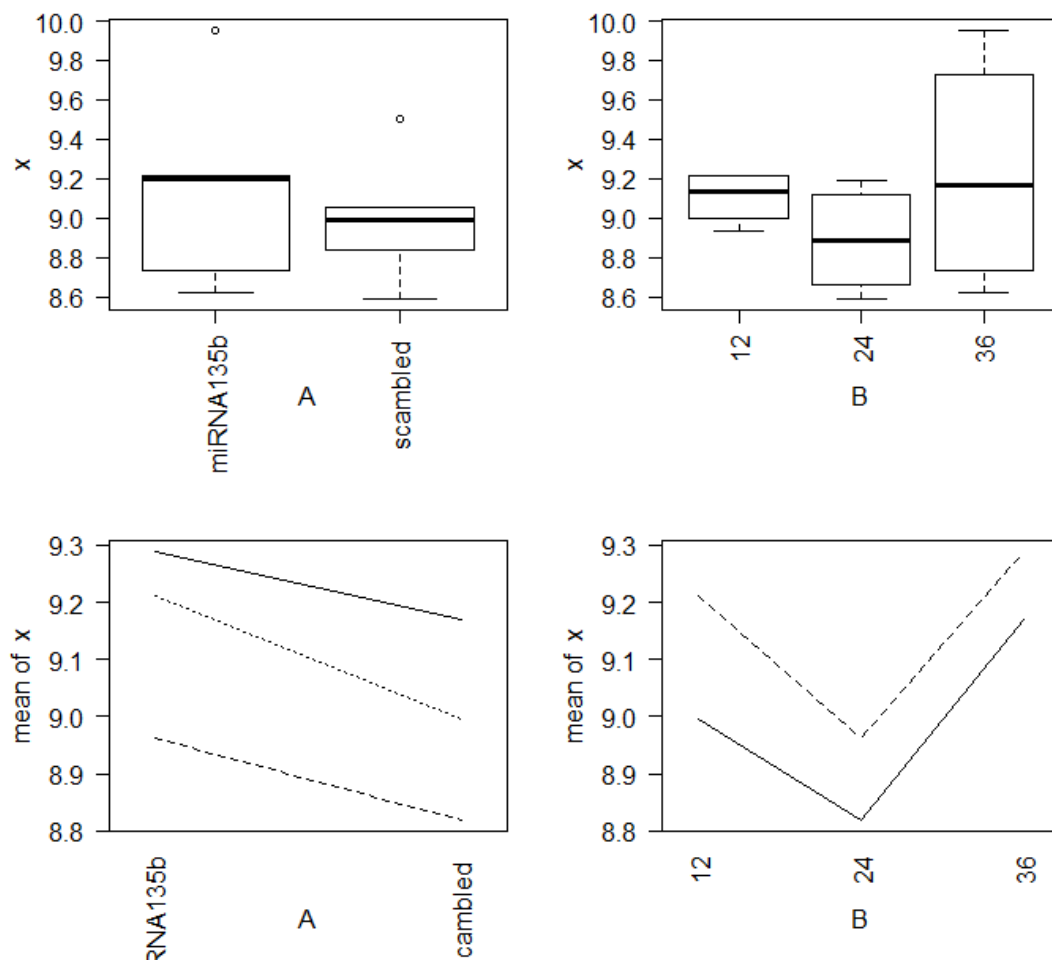
```
> bartlett.test(x~B, data=ge)
```

Bartlett test of homogeneity of variances

data: x by B

Bartlett's K-squared = 5.2045, df = 2, p-value = 0.07411

1.5



1.6

```
> ge.aov<-aov(x~A+B, data=ge)
```

```
> summary(ge.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	0.0768	0.07682	0.465	0.515
B	2	0.2362	0.11808	0.714	0.518
Residuals	8	1.3229	0.16536		

1.7

```
> #考慮交互作用
```

```
> ge.aov2<-aov(x~A*B, data=ge)
```

```
> summary(ge.aov2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	0.0768	0.07682	0.350	0.576
B	2	0.2362	0.11808	0.538	0.610

```
A:B          2 0.0053 0.00265 0.012 0.988
Residuals    6 1.3176 0.21959
```

1.8

(1)

```
> sam.row.name
[1] "ILMN_3279927"
> ge<-data.frame(x<-t(data[sam.row.name,3:14]), A<-factor(substring(Columns(gds6100)$protocol,1,9)), B<-factor(substring(Columns(gds6100)$time,6)))
> bartlett.test(x~A, data=ge)
```

Bartlett test of homogeneity of variances

data: x by A

Bartlett's K-squared = 0.14084, df = 1, p-value = 0.7074

```
> bartlett.test(x~B, data=ge)
```

Bartlett test of homogeneity of variances

data: x by B

Bartlett's K-squared = 1.3605, df = 2, p-value = 0.5065

```
> #不考虑交互作用
```

```
> ge.aov<-aov(x~A+B, data=ge)
```

```
> summary(ge.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	0.00352	0.003524	0.182	0.681
B	2	0.01443	0.007213	0.373	0.700
Residuals	8	0.15458	0.019322		

```
>
```

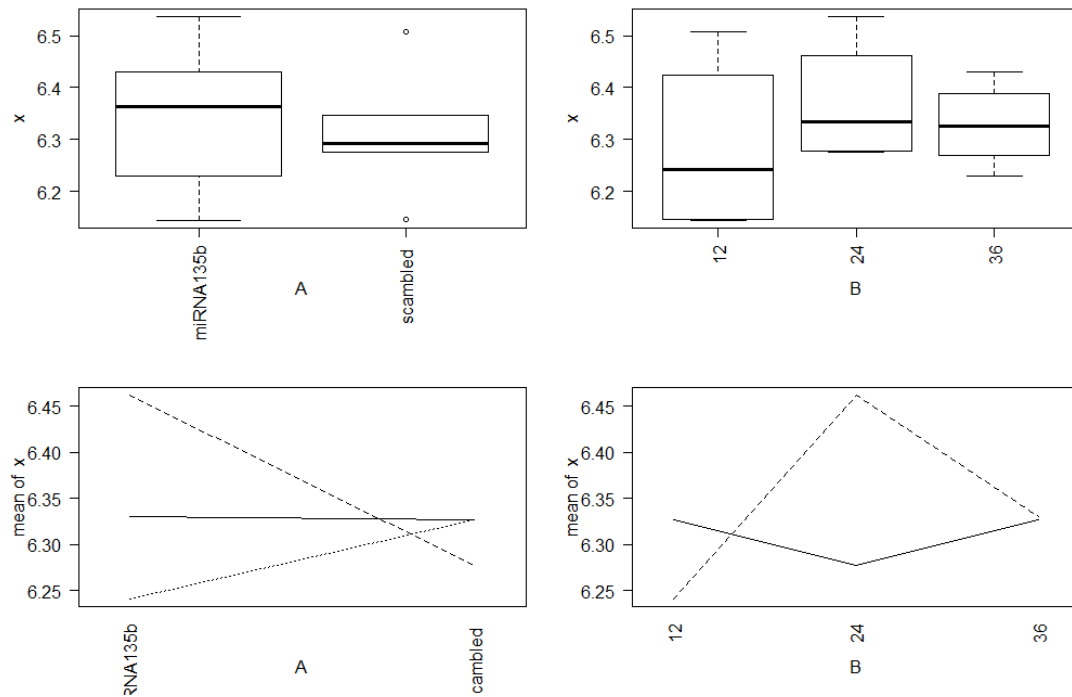
```
> #考虑交互作用
```

```
> ge.aov2<-aov(x~A*B, data=ge)
```

```
> summary(ge.aov2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
--	----	--------	---------	---------	--------

A	1	0.00352	0.003524	0.181	0.685
B	2	0.01443	0.007213	0.371	0.705
A:B	2	0.03805	0.019024	0.980	0.428
Residuals	6	0.11653	0.019422		



(2)

```
> sam.row.name
[1] "ILMN_1723674"
> ge<-data.frame(x<-t(data[sam.row.name, 3:14]), A<-factor(substring(Columns(gds6100)$protocol, 1, 9)), B<-factor(substring(Columns(gds6100)$time, 6)))
> #方差齐性检验
> bartlett.test(x~A, data=ge)
```

Bartlett test of homogeneity of variances

data: x by A

Bartlett's K-squared = 0.67791, df = 1, p-value = 0.4103

```
> bartlett.test(x~B, data=ge)
```


Bartlett test of homogeneity of variances

data: x by B

Bartlett's K-squared = 0.18249, df = 2, p-value = 0.9128

> #不考虑交互作用

> ge.aov<-aov(x~A+B, data=ge)

> summary(ge.aov)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	0.03333	0.03333	4.797	0.0599 .
B	2	0.05482	0.02741	3.946	0.0642 .
Residuals	8	0.05558	0.00695		

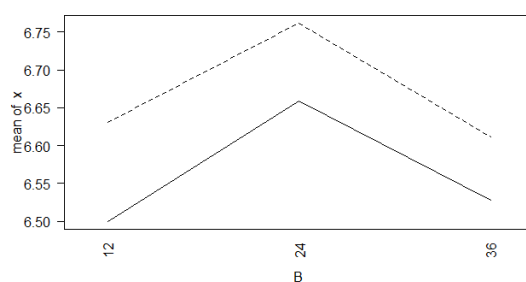
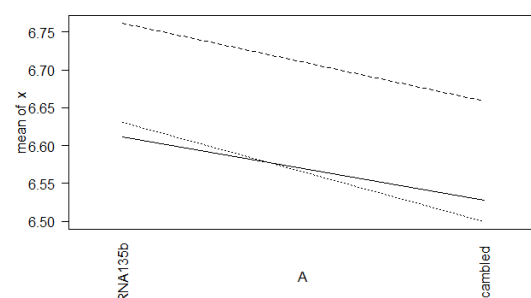
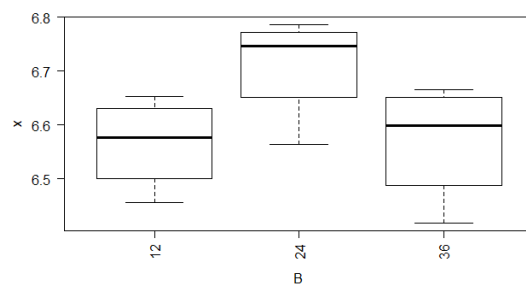
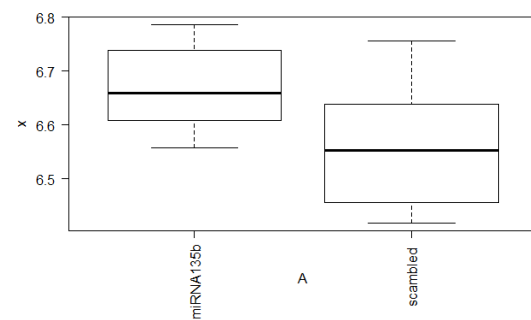
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #考虑交互作用

> ge.aov2<-aov(x~A*B, data=ge)

> summary(ge.aov2)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	0.03333	0.03333	3.673	0.104
B	2	0.05482	0.02741	3.021	0.124
A:B	2	0.00114	0.00057	0.063	0.940
Residuals	6	0.05445	0.00907		



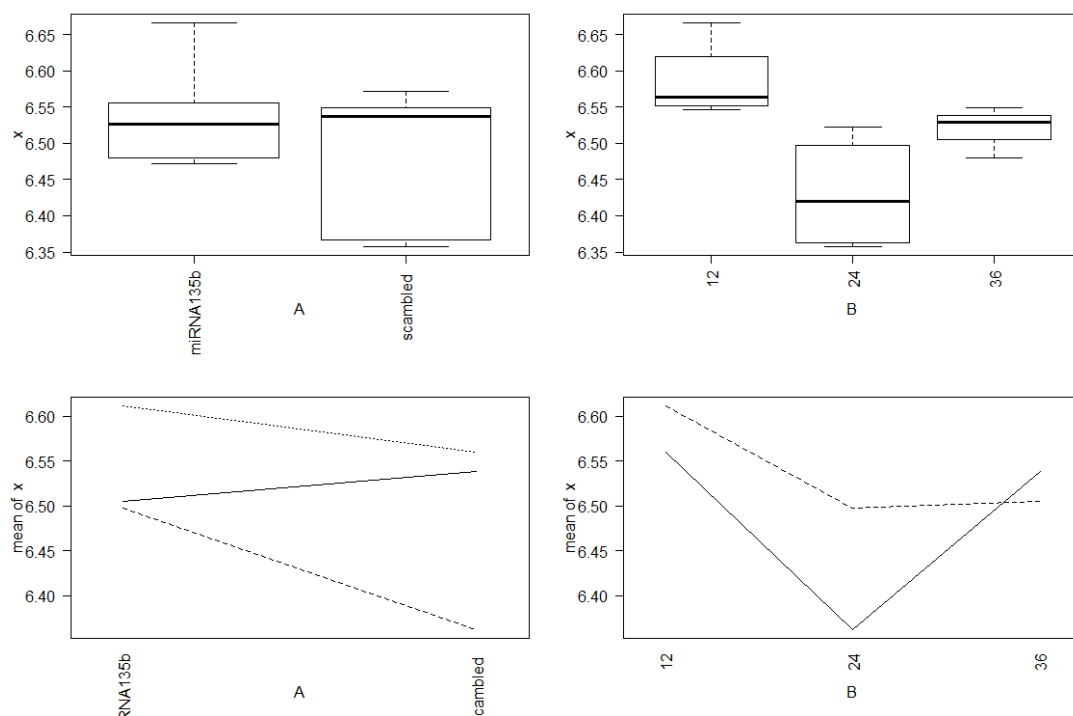
(3)

```
> sam.row.name  
[1] "ILMN_3277152"  
> ge<-data.frame(x<-t(data[sam.row.name,3:14]), A<-factor(substring(Columns(gds6100)  
(substring(Columns(gds6100)$time,6)))  
> #考虑交互作用  
> ge.aov2<-aov(x~A*B, data=ge)  
> summary(ge.aov2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	0.00785	0.007852	5.165	0.06343 .
B	2	0.04886	0.024429	16.070	0.00389 **
A:B	2	0.01429	0.007143	4.699	0.05917 .
Residuals	6	0.00912	0.001520		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>



(4) 由上图可以看出若plot箱型图有明显的包含关系，线图有平行关系，则可以看出因子之间无明显的交互作用，反之，若其有明显

的参差不齐，以及相交趋势，则其存在一定的交互作用。