**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра ВМ-2**

отчет

**по индивидуальному заданию №5**

**по дисциплине «Статический анализ»**

Вариант №13

|  |  |  |
| --- | --- | --- |
| Студент гр. 8382 |  | Мирончик П.Д. |
| Преподаватель |  | Малов С.В. |

Санкт-Петербург

2020

**Постановка задачи**

****

# Ход решения

**1. Сформулировать модели двухфакторного дисперсионного анализа зависимости значений Y от уровней фактора A и B в центральной параметризации. Является ли дизайн данного эксперимента сбалансированным? Построить МНК оценки параметров и несмещенную оценку дисперсии. В предположении нормальности ошибок построить доверительные интервалы для параметров уровня доверия**

Модель имеет вид:

Где - среднее взвешенное, - главные эффекты вектора A, - главные эффекты вектора B, - взаимодействия.

Занесем данные:

## y = c(47.89, 48.26, 47.39, 48.14, 47.71, 47.41, 49.01, 48.95, 48.25, 49.16, 48.69, 49.66, 48.51, 47.79, 48.36, 50.18, 48.62, 48.84, 50.07, 50.25, 49.46, 50.81, 50.64, 49.86, 52.05, 51.04, 51.83, 51.84, 52.08, 53.34, 53.00, 53.72, 54.33, 53.16, 52.97, 52.63, 44.21, 44.51, 45.39, 44.81, 43.08, 45.31, 45.22, 46.48, 45.78, 45.28, 46.76, 45.51)

## z2 = c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4)

## z1 = c(1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2)

здесь A и B поменяли местами: A - z2, B = z1.

Приведем к модели однофакторного анализа:

## # Linear model (code)

## k<-n.lev1\*n.lev2

## z1.l<-as.numeric(t(matrix(levels(as.factor(z1)),nrow=n.lev1,ncol=n.lev2)))

## z2.l<-as.numeric(matrix(levels(as.factor(z2)),nrow=n.lev2,ncol=n.lev1))

## t.code<-matrix(ncol=3,nrow=n.lev1\*n.lev2)

## #

## k<-0

## for (i in 1:n.lev1)

## for (j in 1:n.lev2){

## k<-k+1

## dat1$x[dat1$x1==i & dat1$x2==j]<-k

## t.code[k,1]<-i

## t.code[k,2]<-j

## t.code[k,3]<-k

## }

## t.code<-as.data.frame(t.code)

## names(t.code)<-c("z1","z2","x")

получили соответствие

## > t.code

## z1 z2 x

## 1 1 1 1

## 2 1 2 2

## 3 1 3 3

## 4 1 4 4

## 5 2 1 5

## 6 2 2 6

## 7 2 3 7

## 8 2 4 8

Найдем значения параметров:

## n<-length(y)

## lev<-levels(as.factor(dat1$x))

## n.lev<-length(lev)

## Y<-as.matrix(dat1$y)

## X<-matrix(0,nrow=n.lev,ncol=n)

## for (i in 1:n){

## X[dat1$x[i],i]<-1

## }

## S<-X%\*%t(X)

## S1<-solve(S)

## bhat<-S1%\*%X%\*%Y

## res<-Y-t(X)%\*%as.matrix(bhat)

## SS<-sum(res^2)

## s2<-SS/(n-k)

Несмещенная оценка дисперсии: 0.414725

## v1<-array(1/n.lev1,dim=n.lev1)

## v2<-array(1/n.lev2,dim=n.lev2)

## C0<-as.matrix(as.numeric(v1%o%v2))

## C1<-matrix(0,nrow=k,ncol=n.lev1)

## for (i in 1:n.lev1){ C1[t.code$x[t.code$z1==i],i]<-v2 }

## C1<-C1-matrix(C0,nrow=k,ncol=n.lev1)

## C2<-matrix(0,nrow=k,ncol=n.lev2)

## for (i in 1:n.lev2){ C2[t.code$x[t.code$z2==i],i]<-v1 }

## C2<-C2-matrix(C0,nrow=k,ncol=n.lev2)

## C12<-diag(k)

## for (j in 1:k){

## C12[,j]<-C12[,j]-C1[,t.code$z1[t.code$x==j]]

## C12[,j]<-C12[,j]-C2[,t.code$z2[t.code$x==j]]

## C12[,j]<-C12[,j]-C0

## }

## muhat<-t(C0)%\*%bhat

## V.mu<-t(C0)%\*%S1%\*%C0

## ahat1<-t(C1)%\*%bhat

## V.a1<-t(C1)%\*%S1%\*%C1

## V1<-diag(V.a1)

## ahat2<-t(C2)%\*%bhat

## V.a2<-t(C2)%\*%S1%\*%C2

## V2<-diag(V.a2)

## itr<-t(C12)%\*%bhat

## V.itr<-t(C12)%\*%S1%\*%C12

## V12<-diag(V.itr)

Среднее взвешенное: 48.92167

Главные эффекты A: [-0.545000, 0.527500, 3.744167,3.726667]

Главные эффекты B: [-0.6470833, 0.6470833]

Матрица взаимодействий:

## [,1] [,2]

## [1,] 0.07041667 -0.07041667

## [2,] -0.08541667 0.08541667

## [3,] 0.01125000 -0.01125000

## [4,] 0.00375000 -0.00375000

Доверительный интервал для взвешенного среднего:

## parameter cntr lw up

## 1 mu 48.92167 48.80054 49.04279

Доверительный интервал для главных эффектов A:

## parameter cntr lw up

## 1 A1 -0.545000 -0.7547926 -0.3352074

## 2 A2 0.527500 0.3177074 0.7372926

## 3 A3 3.744167 3.5343740 3.9539593

## 4 A4 -3.726667 -3.9364593 -3.5168740

Доверительный интервал для главных эффектов B:

## parameter cntr lw up

## 1 B1 -0.6470833 -0.7682072 -0.5259595

## 2 B2 0.6470833 0.5259595 0.7682072

Доверительный интервал для взаимодействий:

## parameter cntr lw up

## 1 int1:1 0.07041667 -0.1393760 0.2802093

## 2 int2:1 -0.08541667 -0.2952093 0.1243760

## 3 int3:1 0.01125000 -0.1985426 0.2210426

## 4 int4:1 0.00375000 -0.2060426 0.2135426

## 5 int1:2 -0.07041667 -0.2802093 0.1393760

## 6 int2:2 0.08541667 -0.1243760 0.2952093

## 7 int3:2 -0.01125000 -0.2210426 0.1985426

## 8 int4:2 -0.00375000 -0.2135426 0.2060426

**2. Проверить визуально согласование исходных данных с предположением аддитивности влияния факторов. Построить графически оценку зависимости уровней фактора A при каждом фиксированном значении фактора B. Наблюдается ли эффект пересечения факторов.**

Построим график.

## beta<-t(matrix(bhat,nrow=n.lev2,ncol=n.lev1))

## xx<-c(1:n.lev2)

## yx1<-beta[1,]

## yx2<-beta[2,]

## sym<-c(1:2)

## plot(NULL,NULL,"n",xlab="A",ylab="Y",col="blue",main="ANOVA TWO WAY",xlim=c(min(xx),max(xx)), ylim=c(min(y),max(y)))

## points(z2[z1==1],y[z1==1],col="red",pch=2)

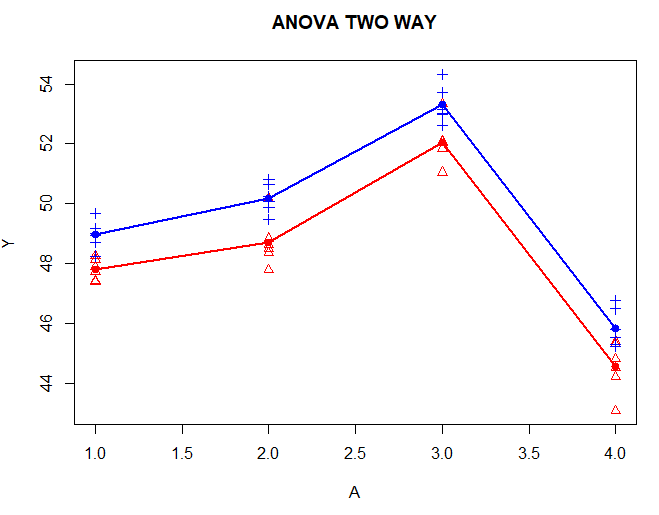
## points(z2[z1==2],y[z1==2],col="blue",pch=3)

## points(xx,yx1,col="red","l",lwd=2)

## points(xx,yx1,col="red",lwd=2,pch=19)

## points(xx,yx2,"l",col="blue",lwd=2)

## points(xx,yx2,col="blue",lwd=2,pch=19)



Заметно, что модель аддитивна, эффекта пересечения факторов не наблюдается.

**3. Сформулировать модель двухфакторного дисперсионного анализа когда пара наибольших уровней факторов A и B рассматривается как базовая. Построить МНК оценки параметров и несмещенную оценку дисперсии. В предположении нормальности ошибок построить доверительные интервалы для параметров уровня доверия**

Зададим базовым уровнем наибольший:

Зададим веса:

## v1<-c(array(0,dim=n.lev1-1),1)

## v2<-c(array(0,dim=n.lev2-1),1)

Остальной код при этом менять не нужно.

Получим значения:

Среднее взвешенное: 45.83833

Главные эффекты A: [3.115000, 4.343333,7.463333,0.000000]

Главные эффекты B: [-1.286667, 0]

Матрица взаимодействий:

## [,1]

## [1,] 0.1333333

## [2,] -0.1783333

## [3,] 0.0150000

## …далее нулевые значения

Доверительный интервал для взвешенного среднего:

## parameter cntr lw up

## 1 mu 45.83833 45.49574 46.18092

Доверительный интервал для главных эффектов A:

## parameter cntr lw up

## 1 A1 3.115000 2.630505 3.599495

## 2 A2 4.343333 3.858838 4.827829

## 3 A3 7.463333 6.978838 7.947829

Доверительный интервал для главных эффектов B:

## parameter cntr lw up

## 1 B1 -1.286667 -1.771162 -0.8021714

Доверительный интервал для взаимодействий:

## parameter cntr lw up

## 1 int1:1 0.1333333 -0.5518465 0.8185132

## 2 int2:1 -0.1783333 -0.8635132 0.5068465

## 3 int3:1 0.0150000 -0.6701798 0.7001798

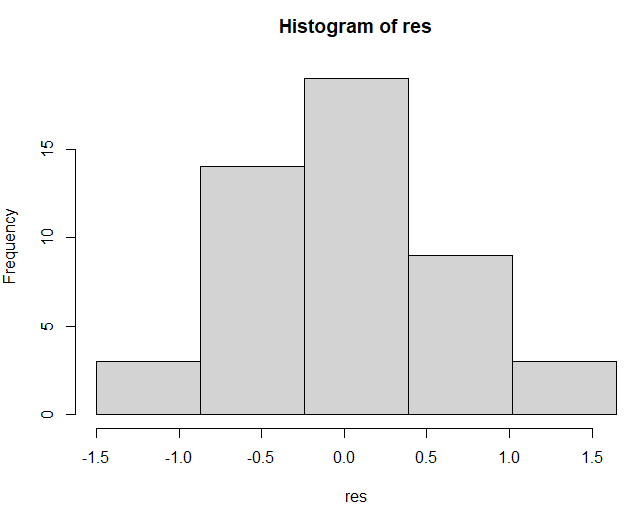
**4. Провести анализ ошибок. На базе ошибок построить гистограмму с шагом h. Оценить расстояние полученной оценки до класса центрированных нормальных распределений по Колмогорову.**

Построим гистограмму ошибок с шагом h=0.63:

## hh<-hist(res,breaks=seq(from=-1.5, to=2, by=0.63),plot=FALSE)

## nu<-hh$counts

## brk<-hh$breaks



Найдем расстояние полученной оценки до класса центрированных нормальных распределений по Колмогорову:

## kolm.stat<-function(s){

## sres<-sort(res)

## fdistr<-pnorm(sres,0,s)

## max(abs(c(0:(n-1))/n-fdistr),abs(c(1:n)/n-fdistr))

## }

## ks.dist<-nlm(kolm.stat,p=sqrt(s2))$minimum

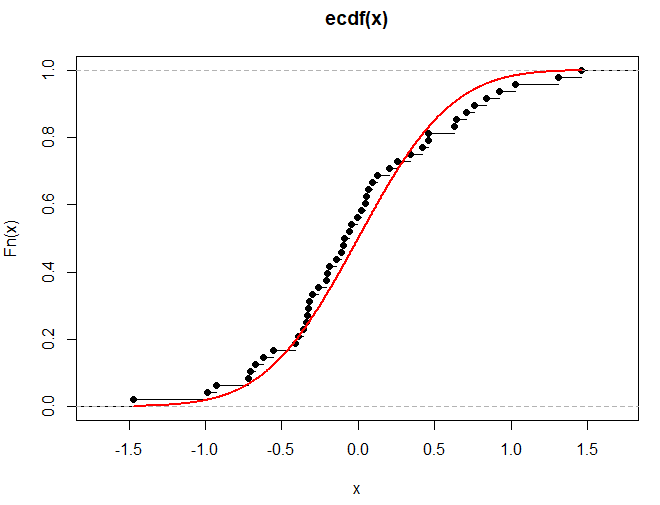
## plot.ecdf(res)

## x2<-c(0:1000)\*(max(res)-min(res))/1000+min(res)

## y2<-pnorm(x2,0,nlm(kolm.stat,p=sqrt(s2))$estimate)

## points(x2,y2,"l",col="red",lwd=2)

получим 0.09228533



**5. Составить таблицу дисперсионного анализа. Провести дисперсионный анализ, начиная с проверки значимости взаимодействий факторов на результаты эксперимента**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Источник дисперсии** | **SS** | **df** |  | **F** | **p** |
| **A:B** | 0.148 | 3 | 0.0495 | 0.119 | 0.948 |
| **A** | 171.630 | 3 | 57.210 | 137.947 | 2.144e-20 |
| **B** | 4.966 | 1 | 4.9665 | 11.975 | 7.2e-07 |
| **Ошибки** | 16.589 | 40 | 0.414 |  |  |

Видно, что взаимодействия факторов не оказывают значительного влияния, в то время как сами факторы заметно влияют на результат.

Выполним проверку:

## q<-lm(y~x1\*x2)

## qa<-lm(y~x1+x2)

## qz1<-lm(y~x2)

## qz2<-lm(y~x1)

Получим

## > anova(qa,q)

## Res.Df RSS Df Sum of Sq F Pr(>F)

## 1 43 16.738

## 2 40 16.589 3 0.14874 0.1196 0.9481

## > anova(qz1,q)

## Res.Df RSS Df Sum of Sq F Pr(>F)

## 1 44 36.836

## 2 40 16.589 4 20.247 12.205 1.412e-06

## > anova(qz2,q)

## Res.Df RSS Df Sum of Sq F Pr(>F)

## 1 46 358.52

## 2 40 16.59 6 341.93 137.41 < 2.2e-16

Заметно, что полученные значения соответствуют записанным выше.