

ANCOVA

Kovariančių analizė

TRUMPAI

ANCOVA –kelių grupių vidurkiai lyginami, atsižvelgiant į papildomą informaciją.

ANCOVA – tai ANOVA su kovariantėmis.

PAVYZDŽIAI

- Norima nustatyti, kuris iš dviejų mokymo metodų geresnis, kai, vertinant rezultatus, atsižvelgiama į mokinių motyvaciją ir IQ.
- Tiriama, ar besimokantieji vairuoti, važiuodami dideliais automobiliais, padaro daugiau klaidų nei važiuodami mažais. Papildomai atsižvelgiama į vairuotojo amžių.

KO TIKIMĖS

- Manome, kad grupių skirtumai išryškės, jei lyginsime to paties intelekto ir vienodos motyvacijos mokinių rezultatus. Nebus taip, kad mažai motyvuoti vienos grupės mokiniai bus lyginami su stipriai motyvuotais kitos grupės mokiniais.
- Tikimės, kad skirtumai išryškės, kai lyginsime maždaug to paties amžiaus vairuotojus.

PALYGINIMAS SU ANOVA

- ANOVA nulinė hipotezė:

$$H_0 : \mu_1 = \mu_2 = \mu_3.$$

- ANCOVA nulinė hipotezė:

$$H_0 : \mu_{1a} = \mu_{2a} = \mu_{3a};$$

čia μ_{1a} , μ_{2a} , μ_{3a} yra grupių vidurkiai,
„pataisyti“, atsižvelgus į kovariantes.

PASTABA

Visai įmanoma situacija, kai taikant ANOVA statistiškai reikšmingi skirtumai randami, o taikant ANCOVA – ne. Pavyzdžiui, gali būti, kad pritaikius ANOVA vidutinis vaikų skaičius katalikiškose ir protestantiškose šeimose skirsis. Kita vertus, atsižvelgus į šeimos galvos socialinį ir ekonominį statusą, šie skirtumai dings.

KINTAMIEJI

- **Priklausomas kintamasis Y** (t.y. tas kintamasis, kurio vidurkius lyginame) yra intervalinis, normaliai pasiskirstęs.
- **Kovariantė X** (t.y. Kintamasis, į kurį reikia atsižvelgti) yra intervalinė, normaliai pasiskirsčiusi.
- **Grupavimo** (nepriklausomas, faktorius) **kintamasis** – kategorinis. Jo dėka žinome, kuriai grupei koks respondentas priklauso.

PAVYZDYS

Tiriame, ar statistiškai reikšmingai skiriasi trijų grupių testo rezultatai (**REZ**, intervalinis kintamasis). Atsižvelgsime į respondentų motyvaciją (**MOT**, intervalinis kintamasis). Grupę nurodyto kintamasis **Grupe** (trireikšmis kategorinis kintamasis).

Duomenis galima rasti knygos *V. Čekanavičius, G. Murauskas, Statistika ir jos taikymai. III*, pirmos dalies šeštajame skyrelyje (1.1.9 pavyzdys).

DUOMENŲ ĮVEDIMAS

GRUPE =c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2 ,
2, 2, 2,2, 3, 3, 3, 3,
3,3, 3, 3, 3, 3)

REZ =c(4,
5,6,8,6,9,10,11,12,13,6,7,8,9,10,12,12,
13, 14, 15, 7,8,
9, 9, 13, 12, 13, 14, 15, 16)

MOT=c(1,4,3,4,5,6,8,8,9,10,2,4,3,4,5,6 ,7,
8, 9, 9, 3, 2, 3, 4, 6,
6,7, 8, 9, 10)

REIKŠMIŲ IŠSIBARSTYMO GRAFIKAS

Norint nustatyti testo rezultatų ir motyvacijos priklausomybę, verta nubraižyti abiejų kintamųjų grafiką. Parodysime, kaip tą padaryti R programa.

REIŠMIŲ IŠSIBARSTYMO GRAFIKAS

#pradinis išsibarstymo grafikas.

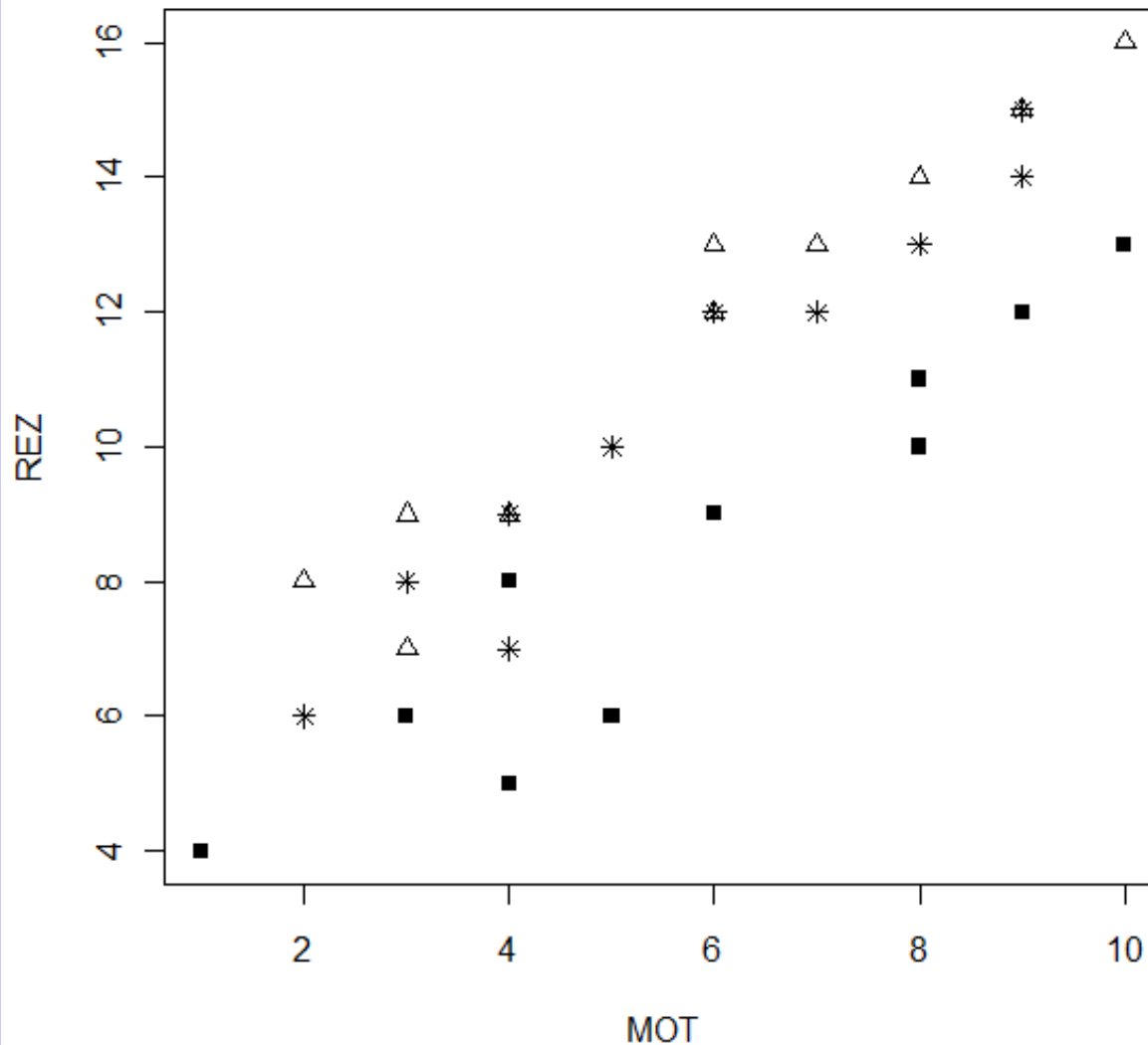
```
plot(MOT,REZ,type="n")
```

```
points(MOT[GRUPE==1],REZ[GRUPE==1],pch=15)
```

```
points(MOT[GRUPE==2],REZ[GRUPE==2],pch=8)
```

```
points(MOT[GRUPE==3],REZ[GRUPE==3],pch=2)
```

REIŠMIŲ IŠSIBARSTYMO GRAFIKAS



Į ką grafike atkreipti dėmesį

1. Visose grupėse matyti ta pati tendencija – kuo didesnė motyvacija, tuo geresni rezultatai.
2. Matome, kad skirtingų grupių rezultatų „debesėliai beveik lygiagretūs“. Vadinasi grupės ir motyvacijos sąveikos (interakcijos) nėra. Visose grupėse motyvacijos poveikis maždaug vienodas.
- 3) Kuo labiau debesėliai atsiskiria, tuo tikėtiniau, taikant ANCOVA, rasti statistiškai reikšmingų vidurkių skirtumų.

ANCOVA PROGRAMA

```
# duomenų paruošimas
dat=data.frame(REZ,MOT,GRP=factor(GRUPE))
attach(dat)
# ANCOVA be sąveikos
ancv_be=lm(REZ~MOT+GRP)
summary(ancv_be)
summary.aov(ancv_be)
```

Sakinyje *lm* užrašome tik modelio pavidalą. Grupavimo kintamasis (faktorius) GRP aprašomas išraiška $GRP = \text{factor}(\text{GRUPE})$.

ANCOVA REZULTATAI

Visus rezultatus galima skirti į dvi grupes:

- a) „pataisytų“ vidurkių lyginimą,
- b) tiesinio modelio parametru įverčių radimą ir hipotezių apie jų lygybę nuliui tikrinimą.

Toliau laikoma, kad pasirinktasis reikšmingumo lygmuo yra 0,05.

„PATAISYTŲ“ VIDURKIŲ LYGINIMAS

Df	Sum Sq	Mean Sq	F value	Pr(>F)
MOT	1 239.250	239.250	327.532	2.956e-16 ***
GRP	2 54.558	27.279	37.344	2.269e-08 ***
Residuals	26 18.992	0.730		

P reikšmės užrašas su e raide – tradicinis:

$$2,269e-08 = 2,269 \times 10^{-8} = 0,00000002269 < 0,05.$$

Kai atsižvelgiama į motyvaciją, tarp trijų grupių vidutinių testo rezultatų yra statistiškai reikšmingai besiskiriančių.

TIESINIS MODELIS

Galima ir kitokia rezultatų interpretacija, kai nagrinėjame ANCOVA taip, kaip ir bet kurį kitą tiesinį Gauso modelį. Trumpai aptarsime gautus rezultatus.

TIESINIS MODELIS

Principinis modelio užrašas atrodo taip:

$$\text{REZ} = \beta_0 + \beta_1 \text{GRP} + \beta_2 \text{MOT} + e.$$

Taigi manome, kad rezultatas priklauso nuo to, kokia respondento motyvacija ir nuo to, iš kokios jis grupės.

TIESINIS MODELIS

Tikslesniame modelio užrašė reikėtų kategorinį kintamąjį GRP pakeisti dviem pseudokintamaisiais. Nurodžius, kad kintamasis yra kategorinis, tą automatiškai atliks programa.

TIESINIO MODELIO REZULTATAI

Residual standard error: 0.8547 on 26 degrees of freedom
Multiple R-squared: 0.9393, Adjusted R-squared: 0.9323
F-statistic: 134.1 on 3 and 26 DF, p-value: 6.195e-16

Matome, kad bent vienas parametras β statistiškai reikšmingai skiriasi nuo nulio ($6,195e-16 < 0,05$).

R kvadratas rodo labai gerą bendrąjį modelio tinkamumą duomenims (R kvadrato reikšmė artima vienetui). Primename, kad tai dar negarantuoja jog modelyje reikalingi abu kintamieji (GRUPE ir MOT).

PARAMETRŲ ĮVERČIAI

Coefficients:

	Estimate	Std. Error	tvalue	Pr(> t)	
(Intercept)	2.03251	0.44311	4.587	1e-04	***
MOT	1.09784	0.06054	18.134	2.82e-16	***
GRP2	2.30978	0.38227	6.042	2.21e-06	***
GRP3	3.20000	0.38222	8.372	7.46e-09	***

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

Visi parametrai statistiškai reikšmingi. Kintamajam GRP vertinti koeficientai prie atitinkamų pseudokintamųjų.

PARAMETRŲ ĮVERČIAI

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.03251	0.44311	4.587	1e-04	***
MOT	1.09784	0.06054	18.134	2.82e-16	***
GRP2	2.30978	0.38227	6.042	2.21e-06	***
GRP3	3.20000	0.38222	8.372	7.46e-09	***

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

Naudojantis parametrų įverčiais, galima užrašyti modelio lygtis kiekvienai grupei. Gautieji modeliai gali būti naudojami prognozėms. Lentelėje nurodytos β_1 reikšmės kiekvienai grupei.

MODELIS PIRMAJAI GRUPEI

Coefficients:

	Estimate	Std. Error	tvalue	Pr(> t)
(Intercept)	2.03251	0.44311	4.587	1e-04 ***
MOT	1.09784	0.06054	18.134	2.82e-16 ***
GRP2	2.30978	0.38227	6.042	2.21e-06 ***
GRP3	3.20000	0.38222	8.372	7.46e-09 ***

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

$$\hat{R}EZ = 2,03 + 1,09MOT.$$

MODELIS ANTRAJAI GRUPEI

Coefficients:

	Estimate	Std. Error	tvalue	Pr(> t)
(Intercept)	2.03251	0.44311	4.587	1e-04 ***
MOT	1.09784	0.06054	18.134	2.82e-16 ***
GRP2	2.30978	0.38227	6.042	2.21e-06 ***
GRP3	3.20000	0.38222	8.372	7.46e-09 ***

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

$$\hat{R}EZ = 2,03 + 2,31 + 1,09MOT$$
$$= 4,34 + 1,09MOT.$$

MODELIS TREČIAJAI GRUPEI

Coefficients:

	Estimate	Std. Error	tvalue	Pr(> t)
(Intercept)	2.03251	0.44311	4.587	1e-04 ***
MOT	1.09784	0.06054	18.134	2.82e-16 ***
GRP2	2.30978	0.38227	6.042	2.21e-06 ***
GRP3	3.20000	0.38222	8.372	7.46e-09 ***

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

$$\begin{aligned}\hat{R}EZ &= 2,03 + 3,20 + 1,09MOT \\ &= 5,23 + 1,09MOT.\end{aligned}$$

TRUMPOS IŠVADOS

Tyrėme, ar statistiškai reikšmingai skiriasi trijų grupių vidutiniai testo rezultatai. Atsižvelgėme į respondentų motyvaciją.

Nagrindėdami ANCOVA, kaip tiesinį Gauso modelį kiekvienai grupei suradome lygtį, aprašančią rezultatų priklausomybę nuo motyvacijos.

PAPILDOMI TYRIMAI

PALYGINIMAS SU ANOVA

Taikant ANCOVA patartina vidurkius palyginti ir be kovariantės. Tam naudojama vienfaktorė dispersinė analizė (ANOVA).

Palyginimas su ANOVA

```
#
```

```
anov=aov(REZ~ GRP)
```

```
summary(anov)
```

PALYGINIMAS SU ANOVA

Vidurkių skirtumai statistiškai nereikšmingi ($p > 0,05$).

Primename, kad atsižvelgus į motyvaciją, statistiškai reikšmingų skirtumų atsirado.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
GRP	2	53.6	26.8	2.7917	0.07906
Residuals	27	259.2	9.6		

KINTAMŲJŲ SAŲVEIKA

Statistiškai reikšminga motyvacijos ir grupės sąveika reiškia, kad skirtingose grupėse motyvacija skirtingai veikia rezultatus.

Klasikinė ANCOVA daroma, kai nėra grupės ir kovariantės sąveikos (interakcijos).

Parodysime, kaip naudojant **R** patikrinti, ar sąveikos nėra.

ANCOVA SU KINTAMŲJŲ SAŲEIKA

```
#su sąveika  
ancv_su=lm(REZ~MOT*GRP)  
summary(ancv_su)  
summary.aov(ancv_su)
```

ANCOVA SU KINTAMŲJŲ SĄVEIKA

Df	Sum Sq	Mean Sq	F value	Pr(>F)
MOT	1 239.250	239.250	319.2543	2.275e-15 ***
GRP	2 54.558	27.279	36.4006	5.395e-08 ***
MOT:GRP	2 1.006	0.503	0.6715	0.5203
Residuals	24 17.986	0.749		

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

Statistiškai reikšmingos kintamųjų sąveikos nėra (p reikšmė viršija 0,05).

Ką daryti, jeigu sąveika yra

Tuo atveju, interpretuojant rezultatus, reikia kalbėti ne apie ANCOVA, bet apie tiesinį Gauso modelį (LNM). Reikėtų užrašyti modelį

$$\text{REZ} = \beta_0 + \beta_1 \text{MOT} + \beta_2 \text{Grupe} + \beta_3 \text{Grupe} \cdot \text{MOT} + e$$

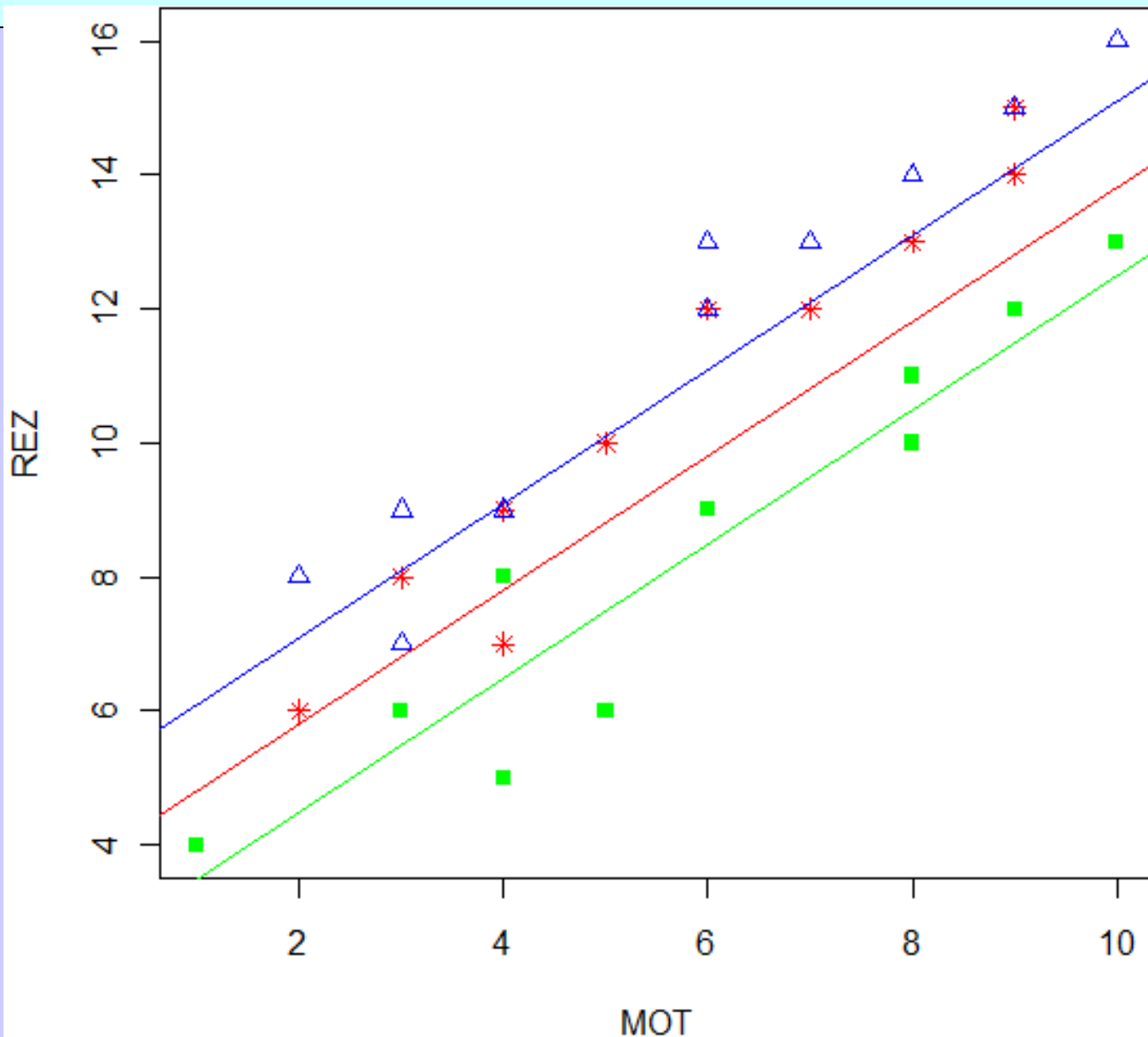
ir jį komentuoti (priklauso nuo motyvacijos, grupės ir motyvacijos sąveikos su grupe ir pan.).

GRAFIKO PERPIEŠIMAS

Galima nubraižyti pakoreguotą grafiką, su atidėtomis regresijos tiesėmis kiekvienai grupei.

```
# grafiko perpiešimas po rezultatų.  
plot(MOT,REZ,type="n")  
points(MOT[GRUPE==1],REZ[GRUPE==1],pch=15,col="green")  
points(MOT[GRUPE==2],REZ[GRUPE==2],pch=8,col="red")  
points(MOT[GRUPE==3],REZ[GRUPE==3],pch=2,col="blue")  
abline(2.5,1,col="green")  
abline(2.5+1.3,1,col="red")  
abline(2.5+2.6,1,col="blue")
```

REIŠMIŲ IŠSIBARSTYMO GRAFIKAS



PASIKLIAUTINIEJI INTERVALAI

Galima suskaičiuoti pasikliautinius intervalus „pataisytiems“ vidurkiams. Parodysime, kaip tą padaryti ANCOVA atveju, kai tiriama, ar skiriasi vidutiniai grupių rezultatai, papildomai atsižvelgiant į motyvaciją.

PASIKLIAUTINIEJI INTERVALAI

R programa:

```
#POST HOC (pataisytų vidurkių pasikliautinieji intervalai)  
  
library(effects)  
summary(effect("GRP",ancv_be,confidence.level=.975))
```

PASIKLIAUTINIEJI INTERVALAI

Rezultatai:

a) „pataisytieji“ vidurkliai

\$effect

GRP

1

2

3

8.363405

10.673189

11.563405

PASIKLIAUTINIEJI INTERVALAI

Rezultatai:

a) „pataisytieji“ vidurkliai

\$effect

GRP

1

2

3

8.363405

10.673189

11.563405

PASIKLIAUTINIEJI INTERVALAI

b) apatiniai pasikliautinųjų intervalų rėžiai

\$lower
GRP

1	2	3
7.72047	10.03020	10.92047

PASIKLIAUTINIEJI INTERVALAI

b) viršutiniai pasikliautinių intervalų rėžiai

\$upper
GRP

1	2	3
9.00634	11.31618	12.20634

PASIKLIAUTINIEJI INTERVALAI

Kompaktiškai pasikliautinuosius intervalus „pataisytiems“ vidurkiams užrašome taip:

Pirmoji grupė: [7,72; 9,00].

Antroji grupė: [10,03; 11,31].

Trečioji grupė: [10,92; 12,21].

PRIELAUDŲ TIKRINIMAS

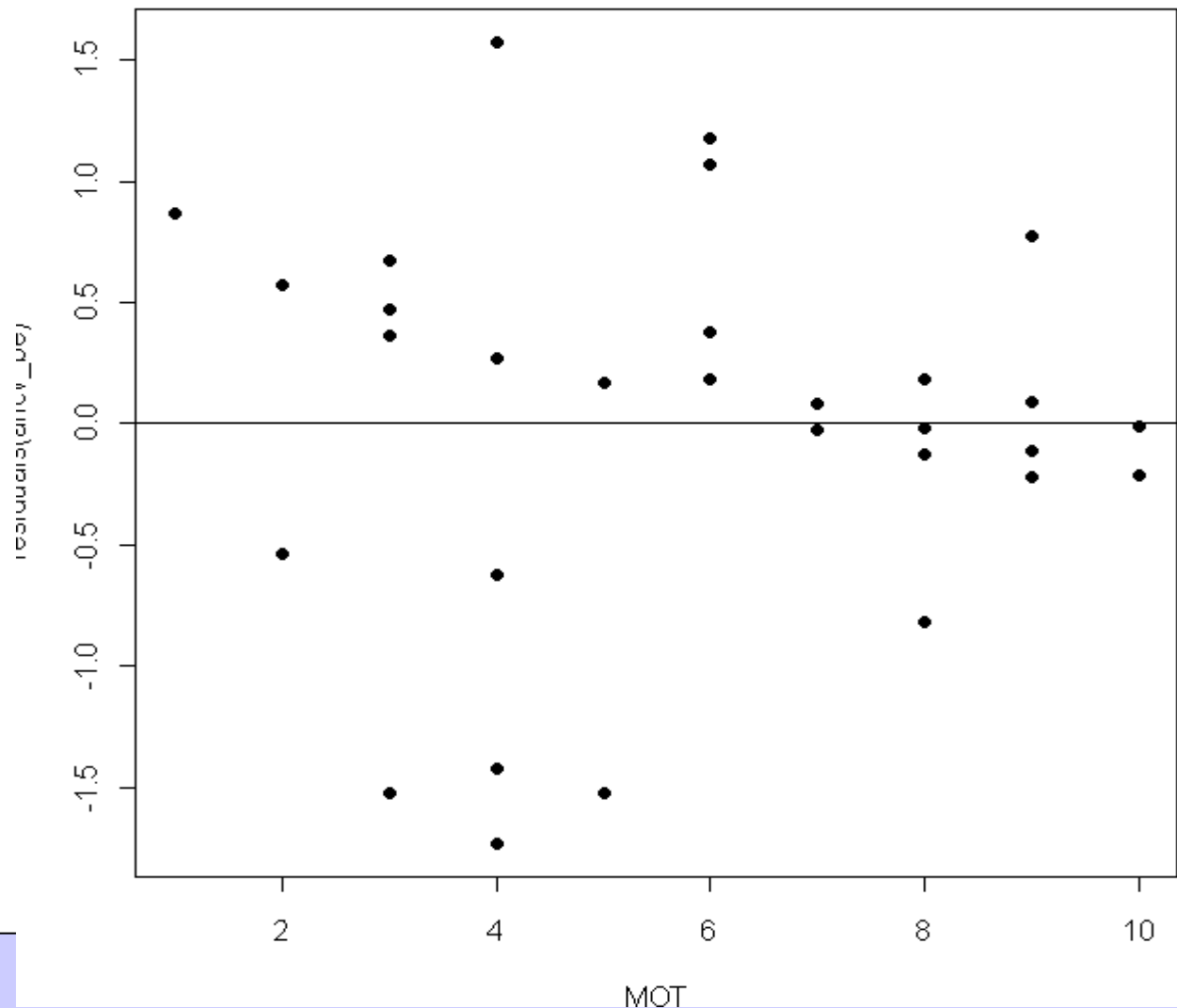
Vizualiai patikrinsime, kad liekamosios paklaidos turi tą pačią dispersiją.

```
# vizualiai:liekanų dispersijos yra pastovios  
plot(MOT, residuals(ancv_be))  
abline(h = 0)  
title(main = "ancova: liekanos vs. kovariante")
```

PRIELAUDŲ TIKRINIMAS

ancova: liekanos vs. kovariante

Gerai, kai
taškai
išsibarstę
tolygiai
(juostele)
apie x- ašį.
Yra ne visai
taip.



PRIELAUDŲ TIKRINIMAS

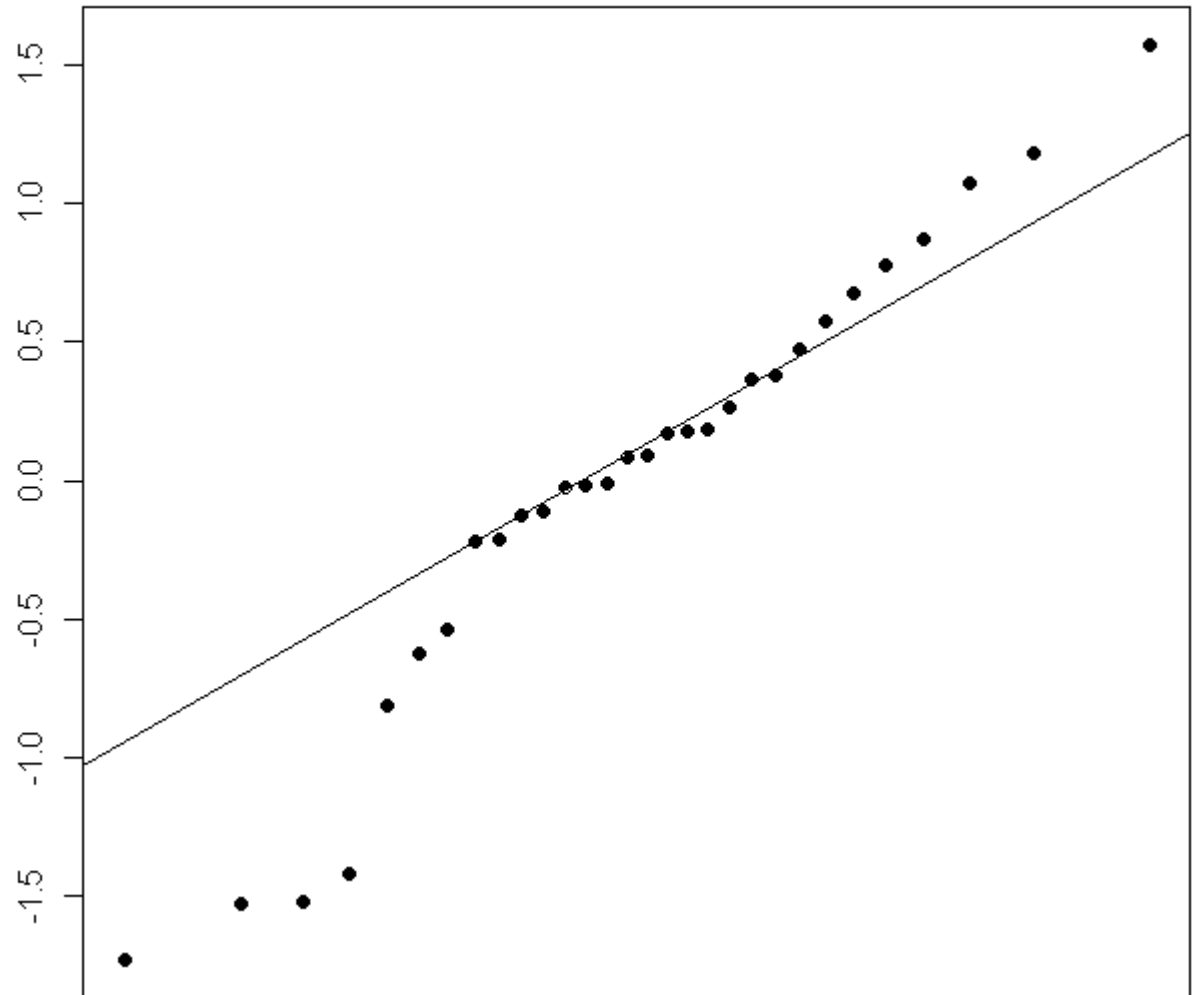
Patikrinsime ar liekamosios paklaidos normaliai pasiskirsčiusios.

```
# normalumas  
qqnorm(residuals(ancv_be), main = "ancova: qq liekanu grafikas")  
qqline(residuals(ancv_be))  
# normalumo testas  
shapiro.test(residuals(ancv_be))
```

PRIELAUDŲ TIKRINIMAS

Iš Q-Q grafiko
matyti, kad
ne
(taškai toli
nukrypę
nuo tiesės).

ancova: qq liekanu grafikas



PRIELAUDŲ TIKRINIMAS

Šapiro testas nerado statistiškai reikšmingo nukrypimo nuo normalaus skirstinio.

Shapiro-Wilk normality test

data: residuals(ancv_bc)
W = 0.9511, p-value = 0.1804

PRIELAUDŲ TIKRINIMAS

Imtyje stebėjimų mažai. Todėl labiau reikėtų tikėti grafikais, o ne Šapiro testu (kai imtis maža, nulinė hipotezė atmetama labai retai).

Kita vertus, o ANCOVA kriterijus pakankamai atsparus saikingiems normalumo prielaidų pažeidimams.

Norint, kad nagrinėtam pavyzdžiui tyrimas būtų visiškai korektiškas, reikėtų didesnės imties. Vis dėlto, preliminariom išvadoms padaryti ANCOVA visiškai tiko.

Baigiamosios pastabos

Tyrėme vienfaktorę ANCOVA, t.y. modelį su viena kovariante. Kovariančių gali būti ir daugiau. Modelio analizė nuo to nesikeičia. Vis dėlto, kuo mažiau kovariančių, tuo modelį lengviau interpretuoti.