

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. Čekanaavičius, G. Murauskas, Statistika ir jos taikymai. III*, pirmos dalies šeštajame skyrelyje (1.1.9 pavyzdys).

DUOMENŲ ĮVEDIMAS

data motyvacija;

input grupe rez mot@@@;

datalines;

1	4	1	1	5	4	1	6	3
1	8	4	1	6	5	1	9	6
1	10	8	1	11	8	1	12	9
1	13	10	2	6	2	2	7	4
2	8	3	2	9	4	2	10	5
2	12	6	2	12	7	2	13	8
2	14	9	2	15	9	3	7	3
3	8	2	3	9	3	3	9	4
3	13	6	3	12	6	3	13	7
3	14	8	3	15	9	3	16	10

run;

ANCOVA PROGRAMA

```
PROC GLM; CLASS GRUPE;  
MODEL REZ = MOT GRUPE /  
SOLUTION;  
LSMEANS GRUPE / TDIFF  
ADJ=TUKEY;  
/* contrast ir estimate sakiniai  
remiasi koreguotais vidurkais */  
CONTRAST '1 vs 3' GRUPE 1 0 -1;  
ESTIMATE '1 vs 3' GRUPE 1 0 -1;  
run;
```

PROGRAMOS KOMENTARAI

```
PROC GLM; CLASS GRUPE;  
MODEL REZ = MOT GRUPE /  
SOLUTION;
```

Naudojama procedūra **PROC GLM**. Grupių skaičius nurodytas kintamajame **GRUPE** (sakinyje **CLASS** išvardijami kategoriniai kintamieji).

Modelio pavidalas nurodomas sakiniu **MODEL**.

PROGRAMOS KOMENTARAI

```
MEANS GRUPE ;  
LSMEANS GRUPE / TDIFF  
DJ=DUKEY;
```

Tukey kriterijumi nustatomi skirtumai tarp grupių „pataisytų“ rezultatų vidurkių, t.y. tarp vidurkių, kai atsižvelgiama į motyvaciją.

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

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MOT	1	240.2079277	240.2079277	328.84	<.0001
GRUPE	2	54.5575014	27.2787507	37.34	<.0001

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

PASTABA

Modelio liekamašias paklaidas galima vertinti įvairiais būdais. „Trečiojo tipo kvadratų suma“ (*Type III SS*) yra pats populiariausias būdas.

SAS programa dar pateikia statistines išvadas gautas, nagrinėjant „Pirmojo tipo kvadratų sumą“ (*Type I SS*).

„PATAISYTI“ VIDURKIAI

*Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer*

GRUPE	REZ LSMEAN	LSMEAN Number
1	8.3634053	1
2	10.673895	2
3	11.5674053	3

Grupių rezultatų vidurkiai, kai atsižvelgiama į motyvaciją.

TUKEY KRITERIJUS „PATAISYTIEMS“ VIDURKIAMS

Gauta, kad pirmoji grupė statistiškai reikšmingai skiriasi nuo likusiųjų, o antroji ir trečioji grupės nesiskiria.

Least Squares Means for Effect GRUPE t for $H_0: LS\text{Mean}(i)=LS\text{Mean}(j) / Pr > t $			
Dependent Variable: REZ			
i/j	1	2	3
1		-6.0423 <.0001	-8.37212 <.0001
2	6.042302 <.0001		-2.32877 0.0694
3	8.372121 <.0001	2.328768 0.0694	

REZULTATŲ APRAŠYMO PAVYZDYS

Tyrėme, ar statistiškai reikšmingai skiriasi trijų grupių vidutiniai testo rezultatai. Atsižvelgėme į respondentų motyvaciją. Tyrimui taikyta kovariančių analizė (ANCOVA) ir Tukey kriterijus. Gavome, kad pirmoji grupė statistiškai reikšmingai skiriasi nuo likusiųjų, o antroji ir trečioji grupės nesiskiria.

ANCOVA rezultatų aptarimas, nagrinėjant ją, kaip tiesinį Gauso modelį

TIESINIS MODELIS

Pateikėme tipinį ANCOVA taikymo ir interpretavimo pavyzdį, kai tyrėją domina tik grupių rezultatų vidurkių skirtumai, papildomai atsižvelgiant į kovariantes.

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:

$$REZ = \beta_0 + \beta_1 GRUPE + \beta_2 MOT + e.$$

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

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	293.8079277	97.9359759	134.07	<.0001
Error	26	18.9920723	0.7304643		
Corrected Total	29	312.8000000			

Matome, kad bent vienas parametras β statistiškai reikšmingai skiriasi nuo nulio.

TIESINIO MODELIO REZULTATAI

R-Square	Coeff Var	Root MSE	REZ Mean
0.939284	8.379138	0.854672	10.20000

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

PARAMETRŲ ĮVERČIAI

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	5.232513798	B	0.44310503	11.81	<.0001
MOT	1.097842449		0.06054048	18.13	<.0001
GRUPE 1	-3.200000000	B	0.38222096	-8.37	<.0001
GRUPE 2	-0.890215755	B	0.38226890	-2.33	0.0279
GRUPE 3	0.000000000	B	.	.	.

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

PARAMETRŲ ĮVERČIAI

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	5.232513798	B	0.44310503	11.81	<.0001
MOT	1.097842449		0.06054048	18.13	<.0001
GRUPE 1	-3.200000000	B	0.38222096	-8.37	<.0001
GRUPE 2	-0.890215755	B	0.38226890	-2.33	0.0279
GRUPE 3	0.000000000	B	.	.	.

Naudojantis parametrų įverčiais galima užrašyti modelio lygtis kiekvienai grupei. Lentelėje nurodytos β_1 reikšmės kiekvienai grupei.

MODELIS PIRMAJAI GRUPEI

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	5.232513798	B	0.44310503	11.81	<.0001
MOT	1.097842449		0.06054048	18.13	<.0001
GRUPE 1	-3.200000000	B	0.38222096	-8.37	<.0001
GRUPE 2	-0.890215755	B	0.38226890	-2.33	0.0279

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

MODELIS ANTRAJAI GRUPEI

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	5.232513798	B	0.44310503	11.81	<.0001
MOT	1.097842449		0.06054048	18.13	<.0001
GRUPE 1	-3.200000000	B	0.38222096	-8.37	<.0001
GRUPE 2	-0.890115755	B	0.38226890	-2.33	0.0279

$$\hat{R}EZ = 5,23 - 0,89 + 1,09MOT$$
$$= 4,34 + 1,09MOT.$$

MODELIS TREČIAJAI GRUPEI

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	5.232513798	B	0.44310503	11.81	<.0001
MOT	1.097842449		0.06054048	18.13	<.0001
GRUPE 1	-3.200000000	B	0.38222096	-8.37	<.0001
GRUPE 2	-0.890215755	B	0.38226890	-2.33	0.0279
GRUPE 3	0.000000000	B	.	.	.

$$\hat{R}EZ = 5,23 + 1,09MOT.$$

PAPILDOMI TYRIMAI

PALYGINIMAS SU ANOVA

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

SAS programa ANOVA galima atlikti ne vienu būdu. Naudosime procedūrą glm.

PALYGINIMAS SU ANOVA

```
/* ANOVA */  
PROC GLM; CLASS GRUPE;  
MODEL REZ = GRUPE / SOLUTION;  
RUN;
```

Programos sakiniai praktiškai nesiskiria nuo rašytųjų ANCOVA modeliui. Tiesiog nebėra kovariantės.

PALYGINIMAS SU ANOVA

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GRUPE	2	53.60000000	26.80000000	2.79	0.0791

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

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

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 **SAS** galima patikrinti, ar sąveikos nėra.

ANCOVA SU KINTAMŲJŲ SĄVEIKA

```
/* ANCOVA su sąveika */  
PROC GLM; CLASS GRUPE;  
MODEL REZ = MOT GRUPE GRUPE*MOT/  
SOLUTION;  
LSMEANS GRUPE / TDIFF ADJ=TUKEY;  
/* CONTRAST IR ESTIMATE SAKINIAI  
REMIASI KOREGUOTAIS VIDURKIAIS */  
CONTRAST ' 1 VS 3' GRUPE 1 0 -1;  
ESTIMATE '1 VS 3' GRUPE 1 0 -1;  
RUN;
```

Programa praktiškai nesiskiria nuo ankstesniųjų.
Papildomai įtraukiama kintamųjų sąveika.

ANCOVA SU KINTAMŲJŲ SĄVEIKA

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MOT	1	240.2832436	240.2832436	320.63	<.0001
GRUPE	2	6.1018716	3.0509358	4.07	0.0300
MOT*GRUPE	2	1.0063777	0.5031889	0.67	0.5203

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.).

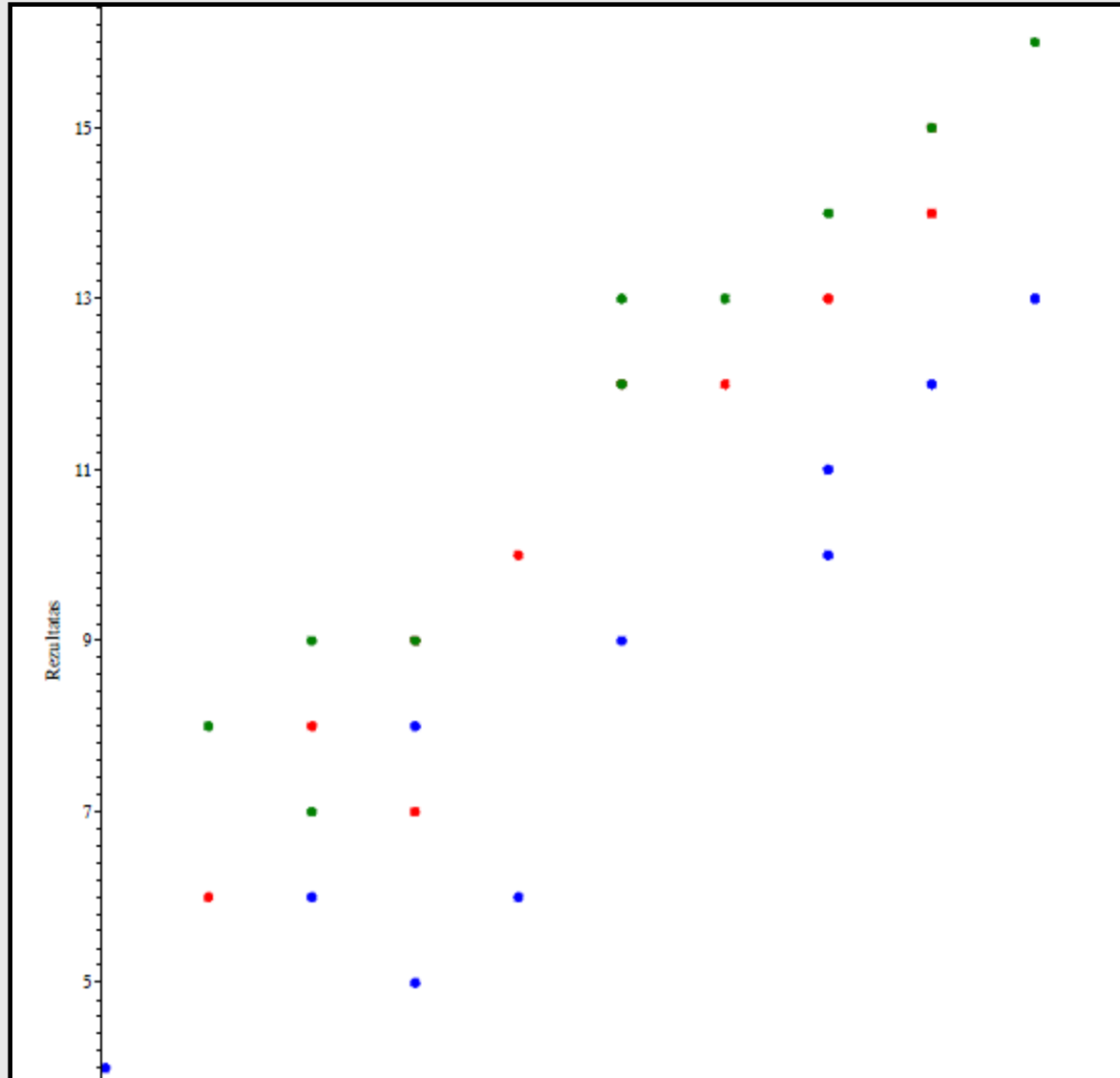
REIKŠMIŲ IŠSIBARSTYMO GRAFIKAS

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

REIŠMIŲ IŠSIBARSTYMO GRAFIKAS

```
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



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 daugmaž lygiagretūs“. Tai reiškia, kad nėra grupės ir motyvacijos sąveikos (interakcijos). Visose grupėse motyvacijos poveikis maždaug vienodas.
- 3) Kuo labiau atsiskiria debesėliai, tuo tikėtiniau, taikant ANCOVA, rasti statistiškai reikšmingų vidurkių skirtumų.

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.