

# SPSS - 1

Duomenų vedimas, aprašomoji statistika, duomenų atranka, perkodavimas

## Struktūra

SPSS sudaro trys blokai (langai):

1. Duomenų langas (**data**). Jame galima matyti:
  - a) duomenis;
  - b) kintamųjų aprašymą.
2. Rezultatų langas (**output**).
3. Sintaksės langas (**syntax**).

© V.Čekanaivičius, G.Murauskas

2

## Struktūra

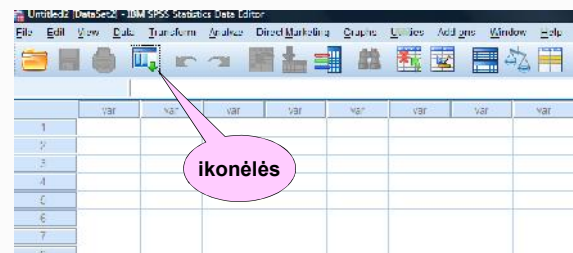
SPSS Rezultatų langas (**output**) gaunamas atlikus skaičiavimus. Rezultatus galima perkelti į Word arba Powerpoint failus.

SPSS Sintaksės langas (**syntax**) naudojamas, kai rašoma programa. Dažniausiai sintaksinės programos neprireikia.

© V.Čekanaivičius, G.Murauskas

3

## Duomenų langas

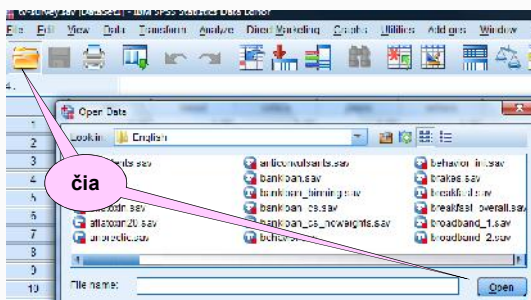


tipiškas vaizdas paleidus SPSS programą – tuščias duomenų langas.

© V.Čekanaivičius, G.Murauskas

4

## Failo atidarymas

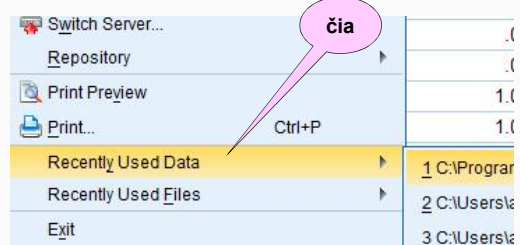


Failas atidaromas – iš bendrojo sąrašo...

© V.Čekanaivičius, G.Murauskas

5

## Failo atidarymas

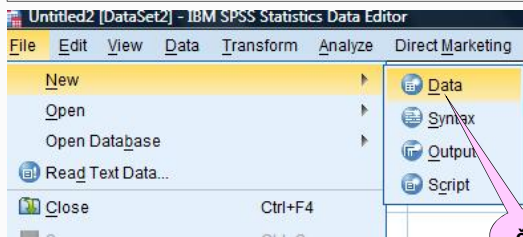


Failas atidaromas – iš neseniai naudotų failų sąrašo

© V.Čekanaivičius, G.Murauskas

6

## Tuščio duomenų lauko gavimas



Tuščio naujo failo atidarymas

© V.Čekanaavičius, G.Murauskas

7

## Matome duomenis, jei

36	37	2	3	24	4
37	38	2	3	24	4
38	39	1	3	24	4
39	40	1	3	24	4
40	41	1	3	24	4
41	42	1	3	24	4
42	43	1	3	24	4
43	44	1	3	24	4
44	45	1	3	24	4

čia

kairiajame kampe dega 'data view' ir...

© V.Čekanaavičius, G.Murauskas

8

## Matome duomenis, jei

peers	writers	director	cast
1.00	1.00	1.00	1.00
.00	1.00	.00	1.00

čia

....ikonėlė 'atspausta'

© V.Čekanaavičius, G.Murauskas

9

## Duomenų struktūra

SPSS ideologija:

1. Vienam respondentui viena eilutė (case).
2. Kintamieji (klausimai) - stulpeliai (variables).

© V.Čekanaavičius, G.Murauskas

10

## Duomenų struktūra

critics	peers	writers	director	cast
1.00	1.00	1.00	1.00	1.00
.00	.00	1.00	.00	1.00
.00	.00	.00	.00	1.00
.00	.00	.00	1.00	.00
.00	.00	.00	1.00	1.00

kintamųjų vardai

duomenys

© V.Čekanaavičius, G.Murauskas

11

## Kintamųjų charakteristikos

1. Vardas (name).
2. Tipas (type).
3. Kintamojo 'etiketė' (variable label).
4. Praleistų reikšmių kodai (missing value).
5. Kintamojo reikšmių 'teiktės' (value label).

© V.Čekanaavičius, G.Murauskas

12

## Kintamojo vardas (name)

1. Prasideda raide.
2. Neturi tarpų.
3. K1; K1.1; Lmnsvs - gerai;
4. Ali Baba, 1grupė – blogai.
5. Kintamojo vardas – stulpelio viršuje.

© V.Čekanavičius, G.Murauskas

13

## Kintamojo tipai (type)

1. Simbolinis (string) -pavardėms ir pan. Turi tikrai ilgį;
2. Skaitinis (numeric) – skaičiams, bei kodams. Standartas (pagal nutylėjimą) 8 pozicijos, iš jų 2 po kablelio.Pvz. 313,72.
3. Kiti tipai – naudojami retai.

© V.Čekanavičius, G.Murauskas

14

## Kintamojo aprašas (etiketė) (variable label)

1. Tai tekstinis komentaras, išsamiau aprašantis kintamąjį.
2. Pvz. kintamojo vardas – K1, o aprašas (label) –'gimimo vieta'.
3. Statistinei analizei – nebūtinai.
4. Naudojamas tik, kad būtų patogiau dirbti.

© V.Čekanavičius, G.Murauskas

15

## Kintamojo praleistos reikšmės (missing values)

1. Tai neatsakytų klausimų kodai.
2. Pvz. kai respondentas nežino atsakymo vedame kodą 0.
3. Jei 0 bus aprašytas, kaip 'missing value', tai SPSS programa traktuos jį taip, kaip ir tuščią langelį.

© V.Čekanavičius, G.Murauskas

16

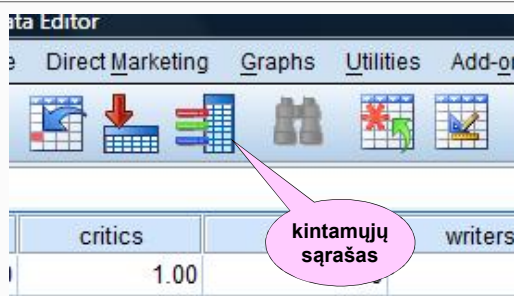
## Kintamojo reikšmių aprašai (etiketės) - value labels

1. Tai tekstiniai komentarai išsamiau aprašantys kintamojo reikšmių kodus.
2. Pvz. kintamajame (stulpelyje) naudosime tik kodus 1 arba 2, o 'value labels' paaiškins, kad 1- vyr., o 2 – mot.
3. Statistinei analizei – nebūtinai.
4. Naudojamas tik, kad būtų patogiau dirbti.

© V.Čekanavičius, G.Murauskas

17

## Visą informaciją apie kintamuosius galime rasti kintamųjų sąrašas:



© V.Čekanavičius, G.Murauskas

18

## Kintamųjų sąrašas

Variable Information:

any

Label: Any reason

Type: F8

Missing Values: none

Measurement Level: Nominal

Value Labels:

.00 Nu

1.00 Yoc

19

## Kintamųjų sąrašas

Variable Information:

any

Label: Any reason

Type: F8

Missing Values: none

Measurement Level: Nominal

Value Labels:

.00 Nu

1.00 Yoc

20

## Kintamųjų sąrašas

Variable Information:

any

Label: Any reason

Type: F8

Missing Values: none

Measurement Level: Nominal

Value Labels:

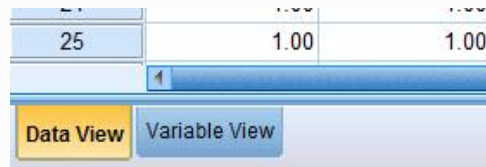
.00 Nu

1.00 Yoc

21

## Duomenų vedimas

1. Aprašome kintamuosius ('variable view' lange).
2. Vedame duomenis ('data view' lange).



© V.Čekanaavičius, G.Murauskas

22

## Kintamųjų aprašymas 'variable view' lange

	Name	Type	Width	Decimals	Label
1	lytis	Numeric	8	2	lytis
2					
3					
4					
5					

23

## Kintamųjų aprašymas 'variable view' lange

Type	Width	Decimals	Label	Values
Numeric	8	2	lytis	None

Value Labels dialog box:

Value: ?

Value Label: mot

Add: 1.00 = "vr."

24

## Duomenų vedimas 'data view' lange

	lytis	var
1	1.00	
2	1.00	
3	2.00	

naudojame klaviatūrą

© V.Čekanaavičius, G.Murauskas

25

## Duomenų vedimas

Praktiniai patarimai:

1. Duomenys vedami klaviatūra.
2. Pelė naudojama saikingai.
3. Vedant duomenis eilute - naudoti 'tab' klavišą.
4. Vedant duomenis stulpeliu – naudoti 'enter' klavišą.

© V.Čekanaavičius, G.Murauskas

26

## Aprašomoji statistika

Susipažinsime su:

1. Dažnių lentelė.
2. Vidurkiu.
3. Dispersija.
4. Standartiniu nuokrypiu.

Pastaba. Yra ir kitų charakteristikų.

© V.Čekanaavičius, G.Murauskas

27

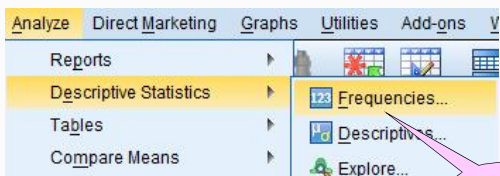
## Dažnių lentelė su SPSS

- Suskaičiuosime, kiek procentų tirtų amerikiečių turi tris vaikus.
- naudosime failą 'GSS93subset' ir kintamąjį
- 'childs'

© V.Čekanaavičius, G.Murauskas

28

## Dažnių lentelė su SPSS

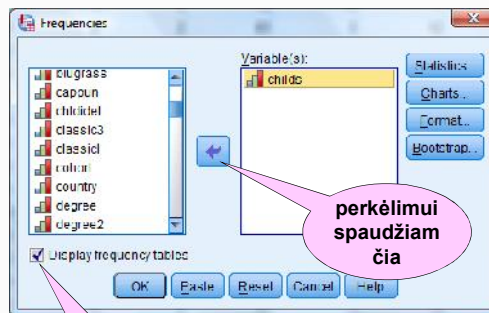


čia

© V.Čekanaavičius, G.Murauskas

29

## Dažnių lentelė su SPSS



varna tupti

perkėlimui spaudžiam čia

© V.Čekanaavičius, G.Murauskas

30

## Dažnių lentelė su SPSS

childs		Number of Children			
	Frequency	Percent	Valid Percent	Cumulative Percent	
Valid 0	414	27,6	27,7	27,7	
1	242	16,1	16,2	43,9	
2	398	26,5	26,6	70,5	
3	226	15,1	15,1	85,6	
4	115	7,7	7,7	93,3	
5	58	3,9	3,9	97,2	
6	14	,9	,9	98,1	
7	7	,5	,5	98,6	
8 Eight or More	21	1,4	1,4	100,0	
Total	1495	99,7	100,0		
Missing 9 NA	5	,3			
Total	1500	100,0			

'percent' – jei su praleistom reikšmėm, 'valid percent'- jei be

© V.Čekanaavičius, G.Murauskas

31

## Dispersija

- 1) Atspindi reikšmių išsibarstymą apie vidurkį.
- 2) kuo didesnė – tuo didesni skirtumai tarp reikšmių.
- 3) pvz. iš dviejų mokymo programų, užtikrinančių vienodus vidutinius pasiekimus, geresnė ta, kurios mažesnė dispersija.

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

© V.Čekanaavičius, G.Murauskas

32

## Standartinis nuokrypis

$$s = \sqrt{\text{dispersija}}$$

© V.Čekanaavičius, G.Murauskas

33

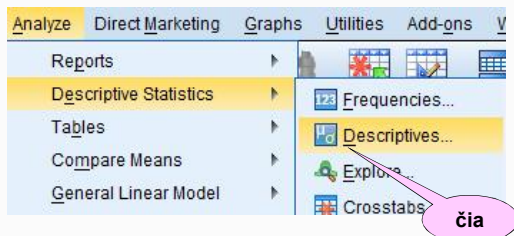
## Vidurkis ir Std su SPSS

- Failas: [GSS93 subset](#)
  - Kintamasis **agewed** (amžius pirmą kartą tuokiantis)
  - Rasime šio kintamojo vidurkį bei standartinį nuokrypį.
- Pastaba:** parodysime tik vieną iš kelių galimų būdų.

© V.Čekanaavičius, G.Murauskas

34

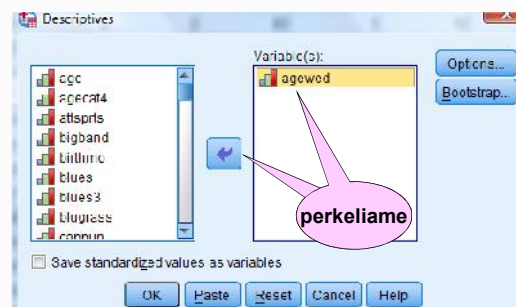
## Vidurkis ir Std su SPSS



© V.Čekanaavičius, G.Murauskas

35

## Vidurkis ir Std su SPSS



© V.Čekanaavičius, G.Murauskas

36

## Vidurkis ir Std su SPSS

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
agedwed Age When First Married	1202	13	58	22.79	5.033
Valid N (listwise)	1202				

vidutinis amžius 22,79 m

vidurkis

std

© V.Čekanavičius, G.Murauskas

37

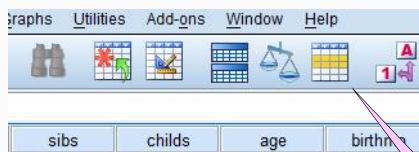
## Duomenų atranka (select cases)

- Paliekame tik dalį duomenų tolesnei analizei;
- Pvz. norime sužinoti vidutinį vyrų amžių pirmą kartą tuokiantis.
- 'Select cases' statistinės analizės neatlieka!

© V.Čekanavičius, G.Murauskas

38

## Duomenų atranka SPSS (select cases)

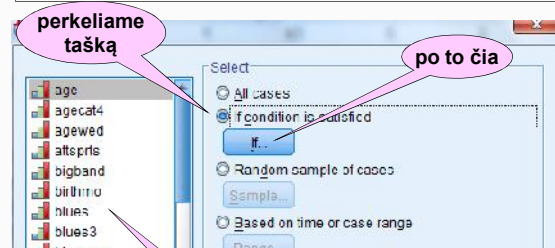


čia

© V.Čekanavičius, G.Murauskas

39

## Duomenų atranka SPSS

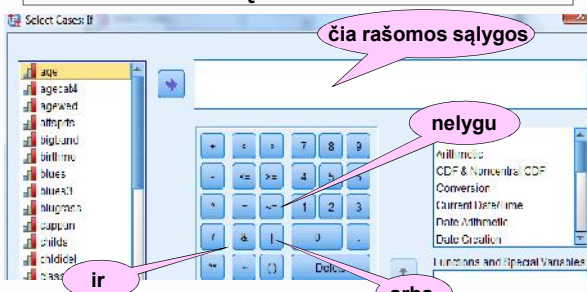


nekišame nagų

© V.Čekanavičius, G.Murauskas

40

## Duomenų atranka SPSS



select cases if = atrinkti eilutes, jei ....

© V.Čekanavičius, G.Murauskas

41

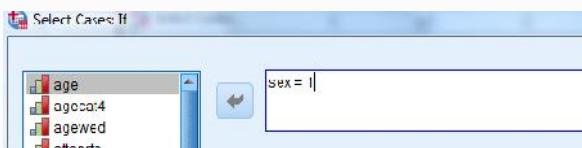
## 'select cases' sintaksė

- kintamųjų vardai sujungiami aritmetinėmis ir (arba) loginėmis operacijomis.
- Pvz., jei norime pasilikti tik vyrų duomenis, (lytis užkoduota kintamuoju sex, vyrų kodas 1), tai rašome
- sex=1

© V.Čekanavičius, G.Murauskas

42

## Duomenų atranka SPSS



Tada 'Continue' ir 'OK'

## Duomenų atranka SPSS

	id	wrkstat	marital
1	1	1	3
2	2	1	5
3	3	1	3
4	4	2	5
5	5	5	5
6	6	5	1

'išbraukta'

Eilutės nedingsta, tik pažymimos, kaip išbrauktos. Tolesnėje analizėje jos nebenaudojamos.

## Duomenų atranka SPSS

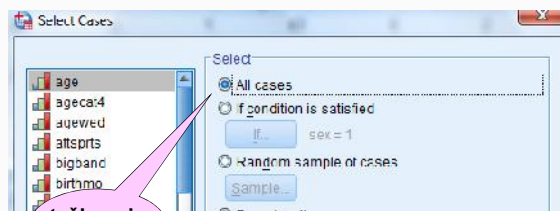
- Pvz. dabar suskaičiavę 'agewed' vidurkį, gautume

Descriptive Statistics

	N	Minimur	Maximur	Mean	Std. Deviat
agewed Age W					
First Married	492	16	50	24,16	4,867
Valid N (listwise)	492				

## Duomenų atranka SPSS

- Norint 'atžymėti' visas eilutes, reikia



tašką prie 'all cases'

## select cases - pastabos

- sąlyga taikoma kiekvienam respondentui.
- pvz. norime atrinkti baltus (race=1) vyrus.
- rašome: (sex=1) & (race=1).
- jaunesnis, nei 50 metų: age < 50.
- tarp 20 ir 50 m: (age >= 20) & (age <= 50).
- baltieji vyrai, jaunesni nei 50 m, arba juodos moterys, vyresnės, nei 30m:
- (sex=1 & race=1 & age < 50) | (sex=2 & race=2 & age > 30).

## select cases - pastabos

- **Tačiau**
- jei norime, kad liktų baltieji bei negrai: tai
- rašome: (race=1) | (race=2).
- jei parašysime (race=1) & (race=2), kompiuteris ieškos žmonių, kurie kartu ir balti ir juodi (zebrų).
- Beje, kintamųjų vardus geriau kopijuoti, o ne rinkti patiemis (mažiau klaidų).



## select cases - pastabos

- **BLOGAI**
  - race = 1|3
  - 30<age<40
  - 30 < age
  - age =15 / age =20
  - [race=1] & [age =50]
  - (race=1) (age =50)
  - (age >50) | (age <60)
- **GERAI**
  - race =1 | race =3
  - age>30 & age <40
  - age >30
  - age=15 | age=20
  - (race=1) &(age=50)
  - (race=1) &(age=50)
  - tai visi respondentai- beprasmė sąlyga.

© V.Čekanavičius, G.Murauskas

49

## select cases – simboliniai kint.

- simbolinių kintamųjų reikšmės rašomos kabutėse:
- gender='M'.
- simbolis turi būti **TIKSLIAI** nukopijuotas.
- gender='m' blogai.
- gender=' M' blogai (tarpas po kabutės).

© V.Čekanavičius, G.Murauskas

50

## COMPUTE

## Ką gali komanda Compute:

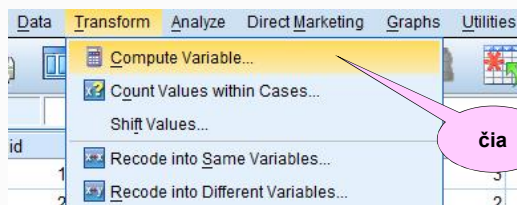
1. Sukurti kintamojo (stulpelio) kopiją.
2. Pakeisti kintamąjį (stulpelį).
3. Sukurti naują kintamąjį (stulpelį).
4. Sukurti 0 – 1 stulpelį, atsižvelgiant į sąlygas.
5. Sukurti sudėtingesnių kodų stulpelį, atsižvelgiant į sąlygas.

Visada transformuoja stulpelius.

© V.Čekanavičius, G.Murauskas

52

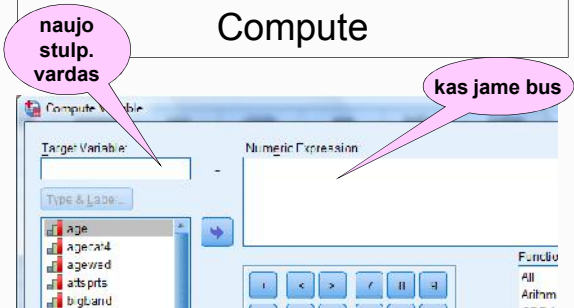
## Compute



© V.Čekanavičius, G.Murauskas

53

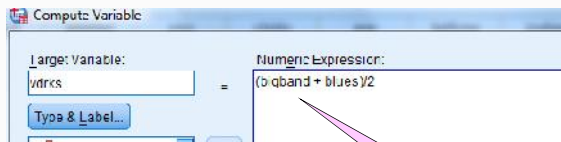
## Compute



© V.Čekanavičius, G.Murauskas

54

## Compute: naujas kintamasis

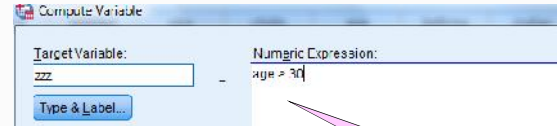


dviejų kint.  
vidurkis

© V.Čekanaavičius, G.Murauskas

55

## Compute: 0 - 1 kintamasis



loginė  
išraiška

naujas kintamasis zzz=1, jei age>30 ir zzz=0, jei age <=30.

© V.Čekanaavičius, G.Murauskas

56

## Compute: 0 - 1 kintamasis

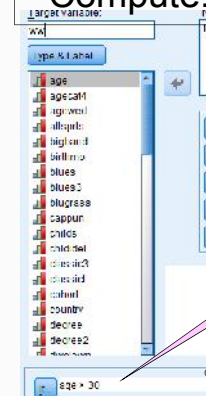
blues3	muzika	zzz	vat	vgf
1,00	3,00	1,00		
1,00	1,00	1,00		
2,00	2,00	1,00		
2,00		1,00		
1,00	1,00	1,00		
1,00	1,00	1,00		
1,00	1,00	1,00		
1,00	1,00	1,00		
1,00	1,00	1,00		
1,00	2,00	1,00		
1,00	2,00	1,00		
2,00		1,00		
1,00	2,00	1,00		

naujas kintamasis zzz=1, jei age>30 ir zzz=0, jei age <=30.

© V.Čekanaavičius, G.Murauskas

57

## Compute: sudėtingesni kodai



paspaudžiam  
ir užrašom  
sąlygą

naujas kintamasis  
ww=1, jei age>30 .

© V.Čekanaavičius, G.Murauskas

58

## AUTOMATIC RECODE

© V.Čekanaavičius, G.Murauskas

59

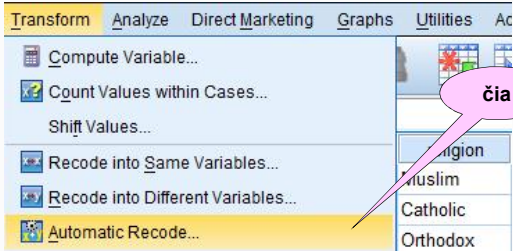
## Automatic recode

1. **Simbolinį (string)** –kintamąjį paverčia skaitiniu;
2. Senosios reikšmės virsta naujojo kintamojo 'value labels'.

© V.Čekanaavičius, G.Murauskas

60

## Automatic recode

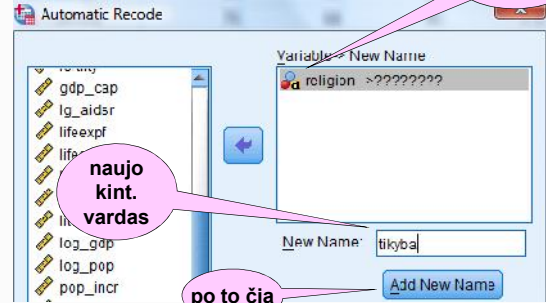


čia

© V.Čekanaavičius, G.Murauskas

61

## Automatic recode



simbolinis kint.

naujo kint. vardas

po to čia

© V.Čekanaavičius, G.Murauskas

62

## Automatic recode

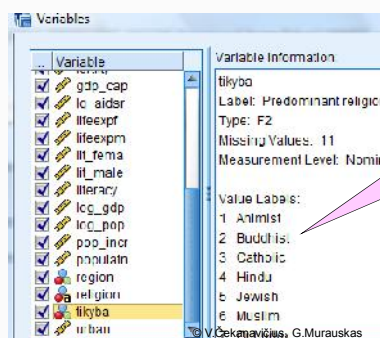
le	it_tema	climate	tikyba	%
44	14	3	6	
9E	35	8	3	
107	170	7		
10L	1.0	3	8	
.	.	8	3	
10C	1.0	3	6	
5E	55	3	6	
47	77	5	6	
4J	39	5	8	
10C	1.0	8	7	
.	.	8	3	
8E	77	4	3	
.	.	0	6	

atsirado skaitinis kintamasis

© V.Čekanaavičius, G.Murauskas

63

## Automatic recode



senos reikšmės virto 'etiketėmis'

© V.Čekanaavičius, G.Murauskas

64

## RECODE

© V.Čekanaavičius, G.Murauskas

65

## Recode

1. Iš skaitinio kintamojo gauna naują skaitinį su kitokiais kodais.
2. Gali ir iš simbolinio padaryti skaitinį, tačiau sudėtingiau, nei automatic recode.
3. Pvz. naujame kintamajame 'islam' musulmoniškas valstybes pažymėsime 1, budistines 2, o likusias 3.

© V.Čekanaavičius, G.Murauskas

66

## Recode

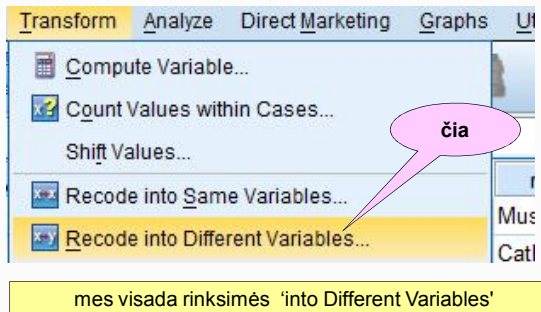
1. Kintamajame 'tikyba': musulmonai – 6, budistai – 2. Įsidėmime.

Value Labels:	
1	Animist
2	Buddhist
3	Catholic
4	Hindu
5	Jewish
6	Muslim
7	Orthodox

© V.Čekanačiūsis, G.Murauskas

67

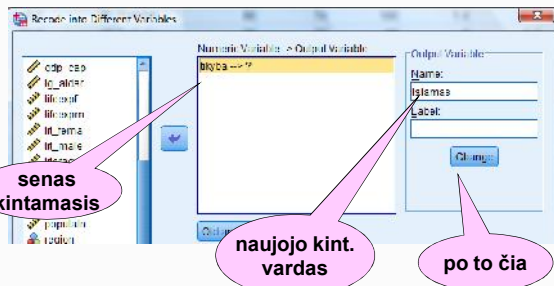
## Recode



© V.Čekanačiūsis, G.Murauskas

68

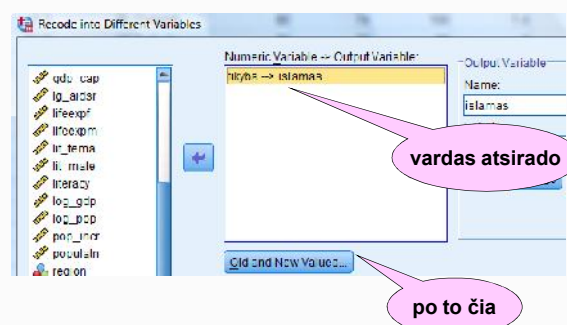
## Recode



© V.Čekanačiūsis, G.Murauskas

69

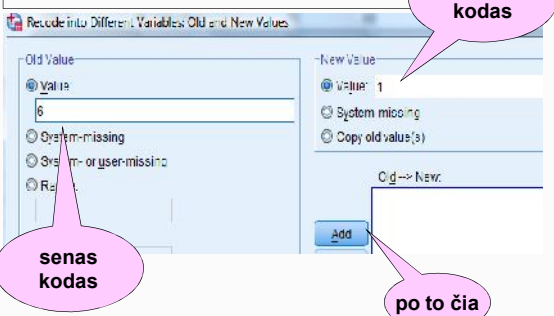
## Recode



© V.Čekanačiūsis, G.Murauskas

70

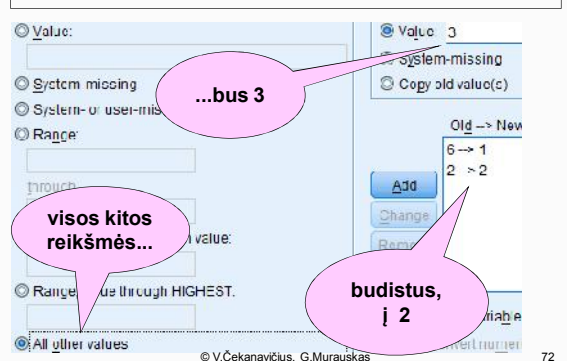
## Recode



© V.Čekanačiūsis, G.Murauskas

71

## Recode



© V.Čekanačiūsis, G.Murauskas

72

## Recode

Atsirado naujas stulpelis:

	likyba	islamais	var	var
3	6	1,00		
3	3	3,00		
.	7	3,00		
3	8	3,00		
0	3	3,00		
3	6	1,00		

© V.Čekanaavičius, G.Murauskas

73

## Recode: sudėtingesnis pvz.

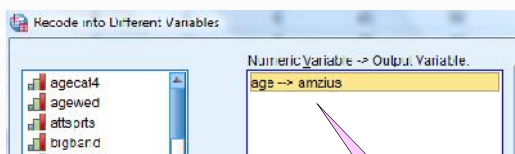
Pvz. Sukursime naują kintamąjį **amzkat**, kuriame pagal amžių respondentus suskirstysime taip:

- 1, jei amžius <30;
- 2, jei amžius 31 – 50;
- 3, jei virš 50.

© V.Čekanaavičius, G.Murauskas

74

## Recode: sudėtingesnis pvz.



nepamirštam  
naujo vardo

© V.Čekanaavičius, G.Murauskas

75

## Recode: sudėtingesnis pvz.

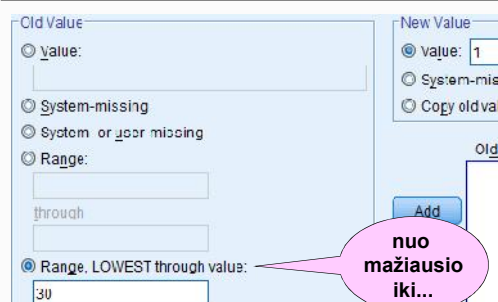


Tam, kad koduotos praleistos reikšmės ir liktų praleistomis reikšmėmis.

© V.Čekanaavičius, G.Murauskas

76

## Recode: sudėtingesnis pvz.

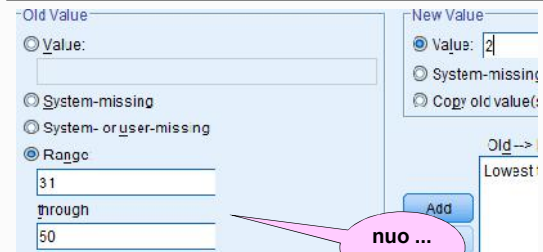


nuo  
mažiausio  
iki...

© V.Čekanaavičius, G.Murauskas

77

## Recode: sudėtingesnis pvz.



nuo ...  
...iki

© V.Čekanaavičius, G.Murauskas

78

## Recode: sudėtingesnis pvz.

through

Range, LOWEST through value:

Range, value through HIGHEST:

51

nuo ... iki didžiausio

© V.Čekanaavičius, G.Murauskas

79

## Recode: sudėtingesnis pvz.

Recode into Different Variables: Old and New Values

Old Value

Value:

System-missing

Copy old value(s)

New Value

Old -> New

MISSING -> SYSMIS

Lowest thru 50 -> 1

31 thru 50 -> 2

51 thru Highest -> 3

taip:

© V.Čekanaavičius, G.Murauskas

80

## Stjudento kriterijus vienai imčiai

Hipotezė apie vidurkio lygybę skaičiui

© V.Čekanaavičius, G.Murauskas

81

## Statistinė hipotezė:

$$\begin{cases} H_0 : \mu = a \\ H_1 : \mu \neq a \end{cases}$$

(duomenyse **vienas** intervalinis stulpelis, skaičius **a** duotas sąlygoje.)

© V.Čekanaavičius, G.Murauskas

82

## Statistinė išvada su p - reikšme

$H_0$  **atmetame** (vidurkis stat. reikšmingai skiriasi nuo  $a$ ), jei

$$p < 0.05$$

$H_0$  **neatmetame** (vidurkis stat. reikšm. nesiskiria nuo  $a$ ), jei

$$p \geq 0.05$$

Čia **0.05** - reikšmingumo lygmuo

© V.Čekanaavičius, G.Murauskas

83

## SPSS

T-testas vienai imčiai, kai

Edukologas nori sužinoti ar **20** studentų SAT skiriasi nuo **1250**.  
 $\alpha = 0.05$

© V.Čekanaavičius, G.Murauskas

84

**SPSS**

Duomenys

	SAT	var	var
1	800,00		
2	1200,00		
3	1400,00		
4	1600,00		
5	950,00		
6	1000,00		
7	1050,00		
8	1150,00		
9	1150,00		
10	1250,00		
11	1300,00		
12	1450,00		

© V.Čekanaavičius, G.Murauskas 85

**SPSS**

Štai čia!

© V.Čekanaavičius, G.Murauskas 86

**SPSS**

kintamasis

reikšmė

© V.Čekanaavičius, G.Murauskas 87

**SPSS**

imties vidurkis

ir std nuokrypis

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
SAT	20	1162,5000	279,97885	62,60517

© V.Čekanaavičius, G.Murauskas 88

**SPSS**

reikšmė

p-reikšmė

One-Sample Test						
Test Value = 1250						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
SAT	-1,398	19	,178	-87,50000	-218,5341	43,5341

kadangi  $p \geq 0.05$ ,  
tai st.reikšm. nesiskiria

© V.Čekanaavičius, G.Murauskas 89

**Išvados formulavimas**

- Pritaikę Stjudento testą gavome, kad imties vidurkis (1162,5) statistiškai reikšmingai nesiskiria nuo 1250.
- Skirtumas pakankamai didelis. Galbūt jis statistiškai nereikšmingas dėl mažo stebėjimų skaičiaus (jū buvo tik 20).

© V.Čekanaavičius, G.Murauskas 90

Stjudento t kriterijus,  
taikomas **priklausomoms**  
imtims (porinis t testas)

Dviejų vidurkių lygybė

© V.Čekanaavičius, G.Murauskas

91

## Kada taikomas

- ◆ **Dvi imtys**, duomenys gaunami poromis (tiriam vyrą-žmoną, motiną – dukterį, ekspertai vertina dvi gražuoles).
- ◆ Dažnai duomenys gaunami, dukart išmatavus tą patį respondentą.
- ◆ Duomenyse du stulpeliai. **Abiejuose intervaliniai kintamieji**. Jų vidurkiai ir lyginami.

© V.Čekanaavičius, G.Murauskas

92

## Statistinė hipotezė:

$$H_0: \mu_x = \mu_y$$

$$H_1: \mu_x \neq \mu_y$$

(hipotezė formuluojama **populiacijoms**)

© V.Čekanaavičius, G.Murauskas

93

## Statistinės išvados su **p - reikšme**

**H<sub>0</sub> atmetame** (vidurkiai stat. reikšmingai skiriasi), jei

$$p < 0.05$$

**H<sub>0</sub> neatmetame** (vidurkiai stat. reikšm. nesiskiria), jei

$$p \geq 0.05$$

Čia **0.05** - reikšmingumo lygmuo.

© V.Čekanaavičius, G.Murauskas

94

## Pavyzdys

- ◆ ar pirmadieniais ir penktadieniais vienodai ilgai kalbama telefonu?
- ◆ abi dienas (**p, pt**) slapta stebėti tie patys vienos firmos darbuotojai.
- ◆ duomenys: (20,32), (40,45), (15,11), (120,110), (34,39), (27,24), (33,35), (40,36), (20,19), (10,15) min.

© V.Čekanaavičius, G.Murauskas

95

## SPSS

Stjudento t kriterijus,  
taikomas **priklausomoms** imtims

(pvz. apie telefono pokalbius  
pirmadienį ir penktadienį)

- ◆ duomenys **greta**

- ◆  **$\alpha = 0.05$**

© V.Čekanaavičius, G.Murauskas

96



### SPSS

© V.Čekanaavičius, G.Murauskas 97

### SPSS

Pair	Variable1	Variable2
1	penktad	pirmad
2		

© V.Čekanaavičius, G.Murauskas 98

### SPSS

	Mean	N	Std. Deviat	Std. Err
Pair 1 PIRMAI	35.900	10	31.2888	9.8943
PENKTA	36.600	10	28.0444	8.8684

© V.Čekanaavičius, G.Murauskas 99

### SPSS

	Paired Differences						t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference					
				Lower	Upper				
Pair 1 PIRMAI - PENKTA	-.7000	6.32543	2.00028	-5.2249	3.8249	-.350	9	.734	

© V.Čekanaavičius, G.Murauskas 100

### SPSS

	Paired Differences						t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference					
				Lower	Upper				
Pair 1 PIRMAI - PENKTA	-.7000	6.32543	2.00028	-5.2249	3.8249	-.350	9	.734	

© V.Čekanaavičius, G.Murauskas 101

### Išvados formulavimas

- Pritaikę Stjudento testą gavome, kad nėra statistiškai reikšmingo skirtumo tarp to, kiek vidutiniškai laiko prakalbama pirmadieniais ir penktadieniais.

© V.Čekanaavičius, G.Murauskas 102

Stjudento t kriterijus,  
taikomas **nepriklausomoms**  
imtims

(dviejų vidurkių lygybė)

© V.Čekanačiūsis, G.Murauskas

103

## Pastabos

- Lyginame **dviejų nepriklausomų** grupių vidurkius,
- pvz., dviejų grupių vidutinius pažymius.
- Duomenyse visada **DU** skirtingo tipo stulpeliai:
  - viename – **intervalinis** rezultatas (pvz. pažymys),
  - **kitame - grupės kodas** (pvz., 1 arba 2)

© V.Čekanačiūsis, G.Murauskas

104

## Statistinė hipotezė:

$$H_0: \mu_x = \mu_y$$

$$H_1: \mu_x \neq \mu_y$$

(hipotezė formuluojama **populiacijoms**)

© V.Čekanačiūsis, G.Murauskas

105

## Statistinės išvados, kai reikšmingumo lygmuo 0,05

**H<sub>0</sub> atmetame** (vidurkiai stat. reikšmingai skiriasi), jei

$$p < 0,05$$

**H<sub>0</sub> neatmetame** (vidurkiai stat. reikšm. nesiskiria), jei

$$p \geq 0,05$$

© V.Čekanačiūsis, G.Murauskas

106

## SPSS

Stjudento t kriterijus,  
taikomas **nepriklausomoms**  
imtims

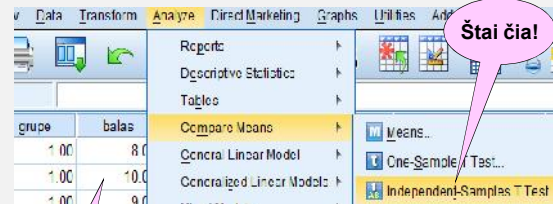
(pvz. apie egzamino pažymio  
viešinimą)

- ◆ magistras: užkoduoti 1
- ◆ fukseliai: užkoduoti 2
- ◆  **$\alpha = 0.05$**

© V.Čekanačiūsis, G.Murauskas

107

## SPSS



duomenys

© V.Čekanačiūsis, G.Murauskas

108

**SPSS**

Independent-Samples T Test

Test Variable(s):  
balas

Grouping Variable:  
grupe(??)

Define Groups...

Options...  
Bootstrap...

po to čia

grupė

© V.Čekanaavičius, G.Murauskas 109

**SPSS**

Define Groups

Use specified values

Group 1: 1

Group 2: 2

Cut point:

Continue Cancel Help

Po to 'Continue' ir 'OK'.

© V.Čekanaavičius, G.Murauskas 110

**SPSS**

Group Statistics

GRUPE	N	Mean	Std. Deviation	Std. Error Mean
BALAS 1.00	10	7.5000	1.58114	.50000
2.00	10	5.4000	1.64655	.52068

imčių vidurkiai

ir std nuokrypiai

© V.Čekanaavičius, G.Murauskas 111

**SPSS**

Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means								
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
BALAS Equal variances assumed	.058	.813	2.909	18	.009	2.1000	.72188	-.58339	3.61667	
Equal variances not assumed			2.909	17.971	.009	2.1000	.72188	-.58321	3.61673	

dispersijas laikome lygiomis

jei  $p \geq 0.05$ , tai lygios

© V.Čekanaavičius, G.Murauskas 112

**SPSS**

Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means								
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
BALAS Equal variances assumed	.058	.813	2.909	18	.009	2.1000	.72188	-.58339	3.61667	
Equal variances not assumed			2.909	17.971	.009	2.1000	.72188	-.58321	3.61673	

p-reikšmė vidurkiams

būtų čia, jei dispersijos nelygios

© V.Čekanaavičius, G.Murauskas 113

**Išvados formulavimas**

- (kadangi  $p = 0.009 < 0.05$ , tai vidurkiai skiriasi statistiškai reikšmingai)
- Pritaikę Stjudento testą gavome, kad magistrai pažymių viešinimą vertina statistiškai reikšmingai palankiau (vidutiniškai 7,5 balo) nei pirmakursiai (vidutiniškai 5,4 balo),  $p = 0,009$ .

© V.Čekanaavičius, G.Murauskas 114

## Hipotezė apie koreliacijos koeficiento lygybę nuliui

Ar du **intervaliniai** kintamieji koreliuoja

## Statistinė hipotezė

$$\begin{cases} H_0 : \text{nekoreliuoja} \\ H_1 : \text{koreliuoja} \end{cases}$$

**KINTAMIEJI DU IR ABUDU INTERVALINIAI**

## Statistinė išvada su $p$ - reikšme

$H_0$  **atmetame** (kintamieji stat. reikšmingai koreliuoja), jei

$$p < 0.05$$

$H_0$  **neatmetame** (kintamieji stat. reikšm. nekoreliuoja), jei

$$p \geq 0.05$$

Čia  $\alpha$  - reikšmingumo lygmuo

## Pastabos

- ♦ kuo koreliacijos koeficientas toliau nuo nulio (absoliutiniu didumu didesnis), tuo koreliacija stipresnė.
- ♦ 0,2 (-0,2) – silpna koreliacija; 0,8 (-0,8) stipri.
- ♦ Koreliacijos koeficiento ženklas parodo kaip koreliuoja:
- ♦ teigiamas- vienam kintamajam didėjant, kitas irgi didėja.
- ♦ neigiamas – vienam didėjant, kitas mažėja.
- ♦ **Koreliacija dar nereiškia priežastingumo!**

## SPSS

- ♦ Koreliacija
- ♦ Turime vadybininkų patirtį ir jų pradinį atlyginimą.
- ♦ Ar koreliuoja?
- ♦  $\alpha = 0.05$

## SPSS

Menu	Sub-menu	Value 1	Value 2
Reports			
Descriptive Statistics			
Tables			
Compare Means	salary	st. lėgin	
General Linear Model	57000.00	27000.00	
Generalized Linear Models	40200.00	18750.00	
Mixed Models	21450.00	12000.00	
Correlate	Bivariate...		

**SPSS**

**kintamieji**

Eivariate Correlations

Variables:

bdste  
educ  
ind  
jobcat  
jobtime  
mirccrty  
salary

prevep  
salbeg

Correlation Coefficients:

Pearson  Kendall's tau-b  Spearman

**varna**

© V.Čekanaavičius, G.Murauskas 121

**SPSS**

**raštingiems**

**koreliacija**

**Correlations**

		Beginning Salary	Previous Experience (months)
Beginning Salary	Pearson Correlation	1	.479**
	Sig. (2-tailed)	.	.000
	N	84	84
Previous Experience (months)	Pearson Correlation	.479**	1
	Sig. (2-tailed)	.000	.
	N	84	84

\*\* Correlation is significant at the 0.01 level (2-tailed).

© V.Čekanaavičius, G.Murauskas 122

**SPSS**

**Correlations**

		Beginning Salary	Previous Experience (months)
Beginning Salary	Pearson Correlation	1	.479**
	Sig. (2-tailed)	.	.000
	N	84	84
Previous Experience (months)	Pearson Correlation	.479**	1
	Sig. (2-tailed)	.000	.
	N	84	84

\*\* Correlation is significant at the 0.01 level (2-tailed).

**p-reikšmė**

**raštingiems**

© V.Čekanaavičius, G.Murauskas 123

**SPSS**

**Correlations**

		Beginning Salary	Previous Experience (months)
Beginning Salary	Pearson Correlation	1	.479**
	Sig. (2-tailed)	.	.000
	N	84	84
Previous Experience (months)	Pearson Correlation	.479**	1
	Sig. (2-tailed)	.000	.
	N	84	84

\*\* Correlation is significant at the 0.01 level (2-tailed).

**žmonių skaičius**

© V.Čekanaavičius, G.Murauskas 124

**SPSS**

**Correlations**

		Beginning Salary	Previous Experience (months)
Beginning Salary	Pearson Correlation	1	.479**
	Sig. (2-tailed)	.	.000
	N	84	84
Previous Experience (months)	Pearson Correlation	.479**	1
	Sig. (2-tailed)	.000	.
	N	84	84

\*\* Correlation is significant at the 0.01 level (2-tailed).

Stat. reikšm. koreliuoja. Kuo labiau patyrę, tuo pradinė alga didesnė.

© V.Čekanaavičius, G.Murauskas 125

**Išvados formulavimas**

- Darbuotojo patirtis (išdirbti mėnesiai) statistiškai reikšmingai koreliuoja su pradiniu atlyginimu,  $r=0,479$ ;  $p<0,01$ . Labiau patyrę gauna didesnius pradinius atlyginimus.

© V.Čekanaavičius, G.Murauskas 126

## SPSS – ranginiai kriterijai

## Turinys

- Mann Whitney ..... 1 p.
- Wilcoxon ..... 3 p.
- Spearman ..... 5 p.
- Kruskal – Wallis ..... 7 p.
- Friedman ..... 9 p.

© V. Čekanavičius, G. Murauskas

2

## Mann - Whitney kriterijus

© V. Čekanavičius, G. Murauskas

3

## Mann-Whitney kriterijus

1. Stjudento t kriterijaus nepriklausomoms imtims analogas;
2. bet nelygina vidurkių;
3. lygina skirstinius;
4. kuris kintamasis 'linkęs būti didesniu' parodo didesnis vidutinis rangas.

© V. Čekanavičius, G. Murauskas

4

## Duomenys

1. dvi nepriklausomos imtys, gautos matuojant intervalinius arba ranginius kintamuosius.
2. imčių didumai gali skirtis.
3. skirtingų ranginio kintamojo reikšmių turi būti bent 5.

© V. Čekanavičius, G. Murauskas

5

## Statistinė hipotezė:

$H_0$  : kintamųjų skirstiniai **nesiskiria**  
 $H_1$  : kintamųjų skirstiniai **skiriasi**.

© V. Čekanavičius, G. Murauskas

6

## Kriterijaus idėja:

- Visas lyginamų kintamųjų reikšmes surašome į vieną variacinę eilutę.
- Suranguojame tą eilutę.
- Lyginame kiekvienos imties elementų vidutinius rangus.
- Didesnis rangas- 'kintamasis linkęs būti didesniu'.

© V. Čekanavičius, G. Murauskas

7

## Statistinė išvada su $p$ - reikšme

$H_0$  **atmetame** (kintamųjų skirstiniai stat. reikšmingai skiriasi), jei  $p < \alpha$

$H_0$  **neatmetame** (kintamųjų skirstiniai stat. reikšm. nesiskiria), jei  $p \geq \alpha$

Čia  $\alpha$  - reikšmingumo lygmuo

© V. Čekanavičius, G. Murauskas

8

## Pavyzdys

- Ar vyresni, nei 40 m. juodaodžiai ir baltieji vienodai palankiai vertina klasikinę muziką?
- ranginių kintamųjų reikšmės: 1-labai palankiai, 2-palankiai, ..., 5-labai nepalankiai.
- Tarkime, kad reikšmingumo lygmuo yra **0,05**.

© V. Čekanavičius, G. Murauskas

9

## SPSS

Štai čia!

Classify	▶	1	0
Data Reduction	▶		
Scale	▶	2	2
<b>Nonparametric Tests</b>	▶		
Chi-Square...			
Binomial...			
Runs...			
1-Sample K-S...			
<b>2 Independent Samples...</b>			
K Independent Samples...			
2 Related Samples...			
K Related Samples...			

© V. Čekanavičius, G. Murauskas

10

## SPSS

ką matuojam

rasė

varna

© V. Čekanavičius, G. Murauskas

11

## SPSS

rasių kodai

© V. Čekanavičius, G. Murauskas

12

## SPSS

vidutiniai rangai

Ranks				
	RACE Racew	N	Mean Rank	Sum of Ranks
CLASSICL	1 white	674	371.97	250710.00
	2 black	75	402.20	30165.00
	Total	749		

baltieji rašė mažesnius balus (t.y. vertino palankiau, nes tokia kodavimo sistema)

## SPSS

p-reikšmė

Test Statistics <sup>a</sup>	
	CLASSICL Classical Music
Mann-Whitney U	23235.000
Wilcoxon W	250710.000
Z	-1.183
Asymp. Sig. (2-tailed)	.237

a. Grouping Variable: RACE Racew of Respondent

skirtumas stat. nereikšmingas

## Wilcoxon kriterijus

## Wilcoxon kriterijus

1. Stjudento t kriterijaus porinėms (priklausomoms) imtims analogas;
2. bet nelygina vidurkių;
3. lygina skirstinius;
4. kuris kintamasis 'linkęs būti didesniu' parodo didesnis vidutinis kintamųjų skirtumo rangas.

## Duomenys

1. dvi priklausomos imtys, gautos matuojant intervalinius arba ranginius kintamuosius.
2. imčių didumai vienodi
3. dažniausiai stebime tuos pat respondentus dukart;
4. skirtingų ranginio kintamojo reikšmių turi būti bent 5.

## Statistinė hipotezė:

$H_0$  : kintamųjų skirstiniai nesiskiria  
 $H_1$  : kintamųjų skirstiniai skiriasi.



### Kriterijaus idėja:

- Suskaičiuojame reikšmių skirtumus kiekvienai porai.
- Suranguojame skirtumų modulius.
- Lyginame vidutinius teigiamų ir neigiamų skirtumų rangus.

© V. Čekanavičius, G. Murauskas

19

### Kriterijaus idėja:

- (12,13), (20,30), (5,3), (4,2), (15,17), (25,30).
- Skirtumai: -1, -10, 2, 2, -2, -5.
- Moduliai: 1(-), 2(-), 2(+), 2(+), 5(-), 10(-).
- Rangai: 1(-), 3(-), 3(+), 3(+), 5(-), 6(-).
- Vidutiniai rangai: (+): 3, (-): 3.75

© V. Čekanavičius, G. Murauskas

20

### Statistinė išvada su $p$ - reikšme

$H_0$  atmetame (kintamųjų skirstiniai stat. reikšmingai skiriasi), jei  $p < \alpha$

$H_0$  neatmetame (kintamųjų skirstiniai stat. reikšm. nesiskiria), jei  $p \geq \alpha$

Čia  $\alpha$  - reikšmingumo lygmuo

© V. Čekanavičius, G. Murauskas

21

### Pavyzdys

- Ar vyresni, nei 50 m. respondentai vienodai palankiai vertina klasikinę muziką ir džiazą?
- ranginių kintamųjų reikšmės: 1-labai palankiai, 2-palankiai, ..., 5-labai nepalankiai.
- Tarkime, kad reikšmingumo lygmuo yra 0,05.

© V. Čekanavičius, G. Murauskas

22

### SPSS

Scale

- Nonparametric Tests
  - Chi-Square...
  - Binomial...
  - Runs...
  - 1-Sample K-S...
  - 2 Independent Samples...
  - K Independent Samples...
  - 2 Related Samples...**
  - K Related Samples...

Štai čia!

1	5	0
1	1	31
5	4	24
		3
		4

© V. Čekanavičius, G. Murauskas

23

### SPSS

keliam ABUDU

varna

Two-Related-Samples Tests

Test Pair(s) List: classical jazz

Current Selections

Variable 1:

Variable 2:

Test Type

Wilcoxon  Sign  McNemar

Marginal Homogeneity

Exact... Options...

© V. Čekanavičius, G. Murauskas

24

**SPSS**

**Ranks**

	N	Mean Rank	Sum of Ranks
JAZZ - CLASSIC Negative Ranks	138 <sup>a</sup>	157.43	21725.00
Positive Ranks	198 <sup>b</sup>	176.22	34891.00
Ties	161 <sup>c</sup>		
Total	497		

a. JAZZ Jazz Music < CLASSICL Classical Music  
 b. JAZZ Jazz Music > CLASSICL Classical Music  
 c. CLASSICL Classical Music = JAZZ Jazz Music

**didesni skirtumai buvo, kai palankiau vertino klasiką (jai rašė mažesnius balus).**

© V. Čekanavičius, G. Murauskas 25

**SPSS**

**Test Statistics<sup>b</sup>**

Z	-3.782 <sup>a</sup>
Asymp. Sig. (2-tailed)	.000

a. Based on negative ranks.  
 b. Wilcoxon Signed Ranks Test

**skirstiniai skiriasi stat. reikšmingai**

© V. Čekanavičius, G. Murauskas 26

**Spearman koreliacija**

© V. Čekanavičius, G. Murauskas 27

**Spearman'o kriterijus**

1. Pearson'o koreliacijos analogas.
2. Interpretuojamas visiškai analogiškai;
3. skaičiuoja Pearson'o koreliaciją tarp imčių rangų;
4. tai leidžia naudoti ir specialiu atveju, kai duomenys jau suranguoti.

© V. Čekanavičius, G. Murauskas 28

**Duomenys**

1. porinė imtis, gauta matuojant intervalinius arba ranginius kintamuosius.
2. galima naudoti ir imčių rangus.
3. skirtingų ranginio kintamojo reikšmių turi būti bent 5.

© V. Čekanavičius, G. Murauskas 29

**Statistinė hipotezė:**

$H_0$  : kintamieji **nekoreliuoja**.

$H_1$  : kintamieji **koreliuoja**.

© V. Čekanavičius, G. Murauskas 30

## Statistinė išvada su $p$ - reikšme

$H_0$  **atmetame** (kintamieji stat. reikšmingai koreliuoja), jei

$$p < \alpha$$

$H_0$  **neatmetame** (kintamieji stat. reikšm. nekoreliuoja), jei

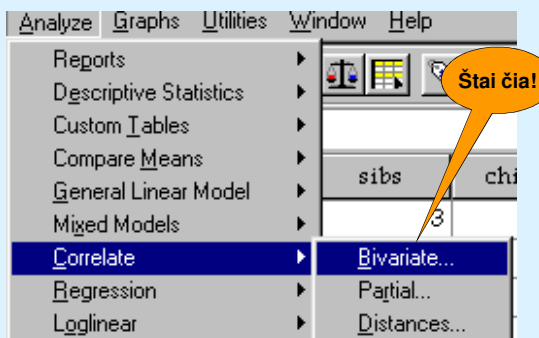
$$p \geq \alpha$$

Čia  $\alpha$  - reikšmingumo lygmuo

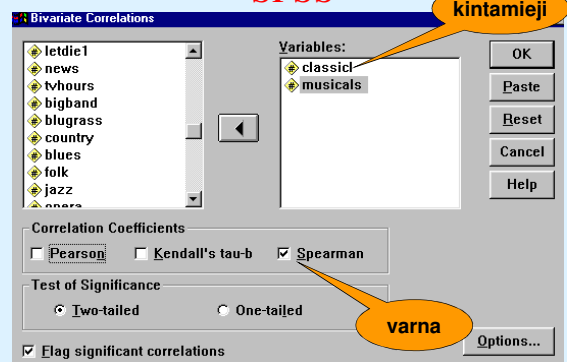
## Pavyzdys

- Ar vyresni, nei 50 m. respondentai, palankiau vertinantys miuziklus, palankiau vertina ir klasikinę muziką?
- ranginių kintamųjų reikšmės: 1-labai palankiai, 2-palankiai, ..., 5-labai nepalankiai.
- Tarkime, kad reikšmingumo lygmuo yra **0,05**.

## SPSS



## SPSS



## SPSS

		CLASSICL Classical Music	MUSICALS Broadway Musicals
Spearman's rho	CLASSICL	1.000	.399**
	Correlation Coefficient		
	Sig. (2-tailed)		.000
	N	504	489
MUSICALS	MUSICALS	.399**	1.000
	Correlation Coefficient		
	Sig. (2-tailed)	.000	
	N	489	499

\*\* . Correlation is significant at the .01 level (2-tailed).

Stat. reikšmingai koreliuoja. Palankiau vertinantys miuziklus, palankiau vertins ir klasiką.

## Spearman specialus atvejis

1. Kiekviena iš dviejų mokytojų surikiavo mokinius pagal gabumus.
2. Pirma: A, B, C, D, E, F, G, H, I, J, K, L.
3. Antra: B, C, A, D, H, E, F, G, K, I, J, L.
4. Ar mokytojų nuomonės koreliuoja?

## Statistinė hipotezė:

$H_0$  : nuomonės nekoreliuoja.

$H_1$  : nuomonės koreliuoja.

## SPSS

•Pirma: A,B,C,D,E,F,  
G,H,I,J,K,L

•Antra: B, C, A,D,  
H,E, F,G,K,I,J,L.

moki	mokyt1	mokyt
A	1.00	3.00
B	2.00	1.00
C	3.00	2.00
D	4.00	4.00
E	5.00	6.00
F	6.00	7.00
G	7.00	8.00
H	8.00	5.00
I	9.00	10.00
J	10.00	11.00
K	11.00	9.00
L	12.00	12.00

šis stulpelis  
nebūtinai

## SPSS

### Correlations

		MOKYT1	MOKYT2
Spearman's rho	MOKYT1	1.000	.916**
	Correlation Coefficient		
	Sig. (2-tailed)	.	.000
N		12	12
MOKYT2	Correlation Coefficient	.916**	1.000
	Sig. (2-tailed)	.000	.
	N	12	12

\*\* . Correlation is significant at the .01 level (2-tailed).

nuomonės labai stipriai teigiamai koreliuoja

## Kruskal - Wallis kriterijus

## Kruskal-Wallis kriterijus

1. Mann-Whitney kriterijaus apibendrinimas keletui (2 ir daugiau) imčių.
2. Interpretuojamas visiškai analogiškai;
3. vidutinis rangas rodo didesnius imčių balus.
4. nepasako, kurie kintamieji skiriasi.
5. yra ANOVA neparametrinis analogas.

## Duomenys

1. keletas nepriklausomų intervalinių arba ranginių imčių.
2. skirtingų ranginio kintamojo reikšmių turi būti bent 5.

## Statistinė hipotezė:

$H_0$  : visi skirstiniai nesiskiria

$H_1$  : kažkurie skirstiniai skiriasi.

## Statistinė išvada su **p - reikšme**

$H_0$  **atmetame** (kažkurie skirstiniai stat. reikšmingai skiriasi), jei  $p < \alpha$

$H_0$  **neatmetame** (visi skirstiniai stat. reikšm. nesiskiria), jei  $p \geq \alpha$

Čia  $\alpha$  - reikšmingumo lygmuo

## Pavyzdys

- Ar tarp prasimokiusių nemažiau 13m. respondentų visų rasių atstovai vienodai vertina rap'ą?
- ranginių kintamųjų reikšmės: 1-labai palankiai, 2-palankiai, ..., 5-labai nepalankiai.
- Tarkime, kad reikšmingumo lygmuo yra **0,05**.

## SPSS

Štai čia!

## SPSS

rap

rasė

varna

## SPSS

kodai

## SPSS

### Ranks

	RACE Racew	N	Mean Rank
RAP Rap Music	1 white	617	372.20
	2 black	65	254.05
	3 other	34	309.59
	Total	716	

palankiausiai (kodavimas!) vertina juodieji.

## SPSS

### Test Statistics<sup>a,b</sup>

	RAP Rap Music
Chi-Square	23.311
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: RACE Racew of Respondent

p-reikšmė

ne visos rasės vienodai vertina rap'ą

## Friedman kriterijus

## Friedman'o kriterijus

1. Wilcoxon kriterijaus apibendrinimas keletui (2 ir daugiau) imčių.
2. jei imtys 2, Wilcoxon kriterijus geriau.
3. interpretacija paprastesnė už Wilcoxon
4. nepasako, kurie kintamieji skiriasi.
5. yra 'Repeated measures' neparametrinis analogas

## Rangų interpretacija

1. Tarkime respondentas parašė trims aktoriams pažymius: aktoriui A - 10, aktoriui B - 6, aktoriui C - 8.
2. Pažymiai **ranguojami**. Aktorių gauti rangai: aktoriaus A - 3, aktoriaus B - 1, aktoriaus C - 2.
3. vidutiniai rangai atspindi tą patį.

## Duomenys

1. keletas imčių, gautų matuojant priklausomus intervalinius arba ranginius kintamuosius.
2. skirtingų ranginio kintamojo reikšmių turi būti bent 5.
3. galima kiekvieno respondento atsakymus suranguoti ir inaudoti tuos rangus.

## Statistinė hipotezė:

$H_0$  : visų kintamųjų skirstiniai nesiskiria.

$H_1$  : kažkurių kintamųjų skirstiniai skiriasi.

## Statistinė išvada su $p$ - reikšme

$H_0$  **atmetame** (kažkurie skirstiniai stat. reikšmingai skiriasi), jei  $p < \alpha$

$H_0$  **neatmetame** (visi skirstiniai stat. reikšm. nesiskiria), jei  $p \geq \alpha$

Čia  $\alpha$  - reikšmingumo lygmuo

## Pavyzdys

- Ar prasimokę daugiau nei 15 m. respondentai, vienodai palankiai vertinana miuziklus, klasiką ir bigbendus?
- ranginių kintamųjų reikšmės: 1-labai palankiai, 2-palankiai, ..., 5-labai nepalankiai.
- Tarkime, kad reikšmingumo lygmuo yra **0,05**.

## SPSS

Classify	4	0
Data Reduction	1	0
Scale	2	2
<b>Nonparametric Tests</b>		
Time Series		
Survival		
Multiple Response		
Missing Value Analysis...		
	5	0
	1	31
	4	0

Štai čia!

## SPSS

Tests for Several Related Samples

Test Variables:

- classicl
- musicals
- bigband

Test Type

Friedman  Kendall's W  Cochran's Q

varna

kintamieji

## SPSS

Ranks		Mean Rank
CLASSICL	Classical Music	1.87
MUSICALS	Broadway Musicals	2.05
BIGBAND	Bigband Music	2.08

palankiausiai vertinta (kodavimas!) klasika

**SPSS**

**Test Statistics<sup>a</sup>**

N	343
Chi-Square	14.286
df	2
Asymp. Sig.	.001

a. Friedman Test

**ne visi stiliai vienodai populiarūs**

© V. Čekanavičius, G. Murauskas 61

**Friedman spec. atvejis**

- Penki ekspertai vertino alų A,B ir C.
- Pirmas: B, C, A (t.y. geriausias B,...)
- Antras: B, C, A
- Trečias: A arba C, B
- Ketvirtas: A, B,C
- Penktas: B, A,C
- Ar visos alaus rūšys vienodai populiarios?

© V. Čekanavičius, G. Murauskas 62

**SPSS**

ekspert	a	b	c
Pirmas	3.00	1.00	2.00
Antras	3.00	1.00	2.00
Trečias	1.50	3.00	1.50
Ketvirtas	1.00	2.00	3.00
Penktas	2.00	1.00	3.00

šis stulpelis nebūtinai rangai!

© V. Čekanavičius, G. Murauskas 63

**SPSS**

**Ranks**

	Mean Rank
A	2.10
B	1.60
C	2.30

**populiariausias B alus**

© V. Čekanavičius, G. Murauskas 64

**SPSS**

**Test Statistics<sup>a</sup>**

N	5
Chi-Square	1.368
df	2
Asymp. Sig.	.504

a. Friedman Test

**vertinimų skirtumai stat. nereikšmingi**

© V. Čekanavičius, G. Murauskas 65



## Chi kvadrato suderinamumo kriterijus

- Spėjame proporciją.
- Jeigu nespėjame, tai čia ne tas chi kvadratas, kurio norisi (yra dar vienas chi kvadrato kriterijus)

© V.Čekanaavičius, G.Murauskas

127

## SPSS

$\chi^2$  suderinamumo kriterijus,

- ♦ Tikriname, ar klerkų nac. daugumos ir nac. mažumos atstovų proporcija yra 7:2
- ♦  $\alpha = 0.05$

© V.Čekanaavičius, G.Murauskas

128

## SPSS

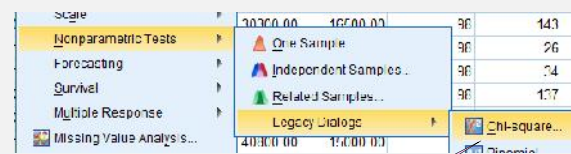
educ	jobcat	minority	total
15	3	0	
16	1	0	
12	1	0	
8	1	0	
15	1	0	
15	1	0	
12	1	0	
15	1	0	
12	1	0	
15	1	0	
16	1	0	

duomenys

© V.Čekanaavičius, G.Murauskas

129

## SPSS

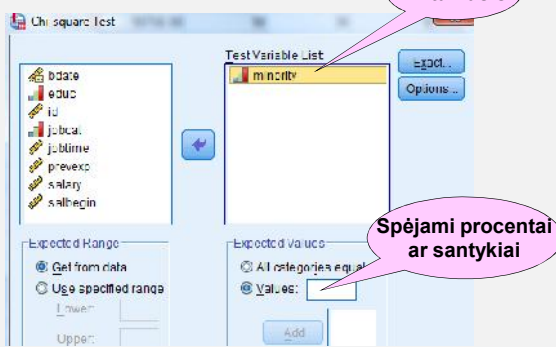


Štai čia!

© V.Čekanaavičius, G.Murauskas

130

## SPSS

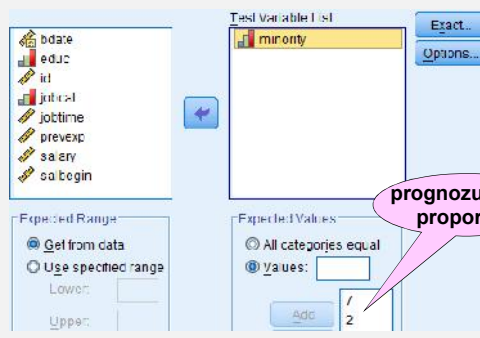


kintamasis

Spėjami procentai ar santykiai

Spėjame, pradėdami nuo mažiau minority kodo

## SPSS



prognozuojama proporcija

© V.Čekanaavičius, G.Murauskas

132

**SPSS**

**Frequencies**

		minority Minority Classification		
	Category	Observed N	Expected N	Residual
1	0 No	276	282,3	-6,3
2	1 Yes	87	80,7	6,3
Total		363		

© V.Čekanaavičius, G.Murauskas 133

skirtumas

Kiek yra

Kiek tikėjomės

**SPSS**

**Test Statistics**

		minority Minority Classification
Chi-Square <sup>a</sup>	df	,639 1
Asymp. Sig.		,424

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 80,7.

© V.Čekanaavičius, G.Murauskas 134

$\chi^2$  reikšmė

p-reikšmė

duomenys nereti

duomenys stat. reikšm. neprieštarauja spėjamam santykiui 7 nac. daug. : 2 nac. maž.

**Išvados formulavimas**

- Pritaikę chi kvadrato kriterijų, gavome, kad duomenys statistiškai reikšmingai neprieštarauja spėjamam nac. daugumos ir nac. mažumos dabuotojų santykiui.

© V.Čekanaavičius, G.Murauskas 135

**SPSS**

**Specialus atvejis**

Saldainiai parduodami 4 skirtingose dėžutėse. 25% dėžučių didelės, 25% vidutinės, 30% mažos ir 20% suvenyrinės. Pirkėjai nupirko 50 didelių, 40 vidutinių, 72 mažas ir 19 suvenyrinių dėžučių. Ar kažkurių dėžučių reikia gaminti daugiau?

♦  $\alpha = 0.05$

© V.Čekanaavičius, G.Murauskas 136

**SPSS**

rusis	kiek
1.00	50.00
2.00	40.00
3.00	72.00
4.00	19.00

© V.Čekanaavičius, G.Murauskas 137

duomenys tik skaitiniai

**SPSS**

© V.Čekanaavičius, G.Murauskas 138

tada čia

**SPSS**

© V.Čekanavičius, G.Murauskas 139

**SPSS**

© V.Čekanavičius, G.Murauskas 140

**SPSS**

© V.Čekanavičius, G.Murauskas 141

**SPSS**

**RUSIS**

	Observed N	Expected N	Residual
1.00	50	45.3	4.8
2.00	40	45.3	-5.3
3.00	72	54.3	17.7
4.00	19	36.2	-17.2
Total	181		

© V.Čekanavičius, G.Murauskas 142

**SPSS**

**Test Statistics**

	RUSIS
Chi-Square <sup>a</sup>	15.050
df	3
Asymp. Sig.	.002

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 36.2.

**Išvada: reikia keisti gamybą**

© V.Čekanavičius, G.Murauskas 143

**Išvados formulavimas**

- Pritaikę chi kvadrato kriterijų, gavome, kad duomenys statistiškai reikšmingai prieštarauja spėjamam pakuočių santykiui,  $p= 0,002$ . Žymiai daugiau, nei tikėtasi, nuperkama trečios rūšies pakuočių ir mažiau ketvirtos rūšies pakuočių.

© V.Čekanavičius, G.Murauskas 144

## Chi kvadrato homogeniškumo kriterijus

### Porinių dažnių lentelės

© V.Čekanaavičius, G.Murauskas

145

### Porinių dažnių lentelės

- Visi imties respondentai suskirstomi į grupes pagal **du** požymius.
- Pvz. pagal lytį ir požiūrį į gyvenimą.
- Gyvenimas: žavus (exciting),
- rutina (routine)
- pilkas (dull)

© V.Čekanaavičius, G.Murauskas

146

### Pavyzdžiai

- Ar nusikalstamumo lygis priklauso nuo bedarbystės lygio?
- Ar tikiama susijusi su požiūriu į abortus (pritariu-nepritariu)?
- Ar perkamo automobilio spalva priklauso nuo perkančiojo lyties?
- Ar studentės dažniau renkasi magistro studijas, nei studentai?

© V.Čekanaavičius, G.Murauskas

147

### $\chi^2$ kriterijus. Požymių homogeniškumo tikrinimas

- vienas požymis stebimas keliose populiacijose;
- tikriname, ar visose populiacijose požymis vienodai pasiskirstęs;

© V.Čekanaavičius, G.Murauskas

148

### Duomenys

- ♦ Vienas kategorinis kintamasis matuotas keliose nepriklausomose populiacijose.
- ♦ Žinome kiekvienos populiacijos sudėtį pagal kategorinį kintamąjį.
- ♦ Kategorijų gali būti keletas.

© V.Čekanaavičius, G.Murauskas

149

### Statistinė hipotezė:

$H_0$  : pagal tiriamą požymį populiacijos **nesiskiria**.  
 $H_1$  : populiacijos **skiriasi**.

© V.Čekanaavičius, G.Murauskas

150

## Statistinė išvada, kai reikšmingumo lygmuo 0,05

$H_0$  **atmetame** (populiacijos stat.  
reikšmingai skiriasi), jei  $p < 0,05$

$H_0$  **neatmetame** (populiacijos stat. reikšm.  
nesiskiria), jei  $p \geq 0,05$

© V.Čekanavičius, G.Murauskas

151

## Pavyzdys

- Ar vienodą procentą vadybininkų ir klerkų sudaro moterys?
- Tarkime, kad reikšmingumo lygmuo yra **0,05**.

© V.Čekanavičius, G.Murauskas

152

## SPSS

	jobcat	gender	sal
15	3	m	18
16	1	m	18
12	1	f	12
8	1	f	18
15	1	m	21
15	1	m	18
15	1	m	18

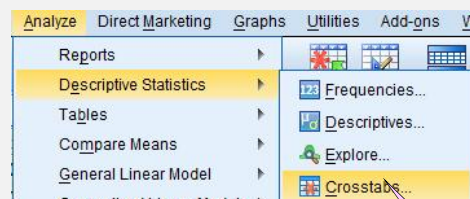
duomenys

tinka ir  
simboliniai

© V.Čekanavičius, G.Murauskas

153

## SPSS

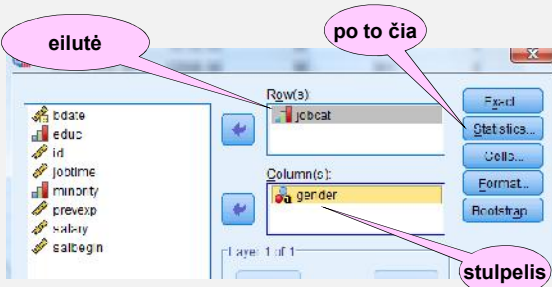


Štai čia!

© V.Čekanavičius, G.Murauskas

154

## SPSS



eilutė

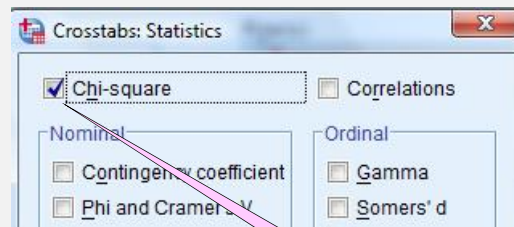
po to čia

stulpelis

© V.Čekanavičius, G.Murauskas

155

## SPSS

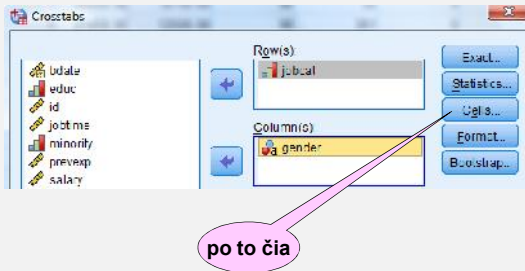


varna

© V.Čekanavičius, G.Murauskas

156

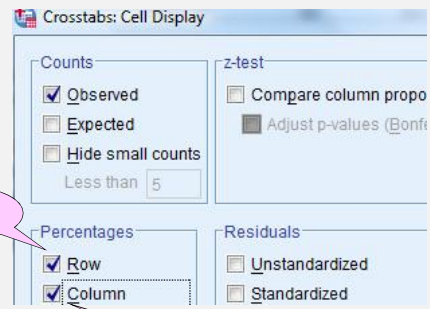
## SPSS



© V.Čekanaavičius, G.Murauskas

157

## SPSS



© V.Čekanaavičius, G.Murauskas

158

## SPSS

JOB CAT Employment Category \* GENDER Gender Crosstabulation

			GENDER Gender		Total	
			f Female	m Male		
JOB CAT Employment Category	1 Clerical	Count	206	157	363	
		% within JOB CAT Employment Category	56.7%	43.3%	100.0%	
		% within GENDER Gender	95.4%	68.0%	81.2%	
3 Manager		Count	10	74	84	
		% within JOB CAT Employment Category	11.9%	88.1%	100.0%	
		% within GENDER Gender	4.6%	32.0%	18.8%	
Total			216	231	447	
			% within JOB CAT Employment Category	48.3%	51.7%	100.0%
			% within GENDER Gender	100.0%	100.0%	100.0%

© V.Čekanaavičius, G.Murauskas

159

## SPSS

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	54.938	1	.000		
Continuity Correction	53.154	1	.000		
Likelihood Ratio	61.256	1	.000		
Fisher's Exact Test				.000	.000
N of Valid Cases	447				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count

$p < 0.05$ , todėl proporcijos sat. reikšmingai skiriasi

© V.Čekanaavičius, G.Murauskas

160

## Išvados formulavimas

- Pritaikę chi kvadrato kriterijų, gavome, kad tarp klerkų moterų yra statistiškai reikšmingai daugiau (56,7%), nei tarp vadybininkų (11,9 %),  $p < 0,01$ .

© V.Čekanaavičius, G.Murauskas

161

## SPSS

### Specialus atvejis

- grupė vaikų žiūrėjo smurtines TV laidas, grupė - nesmurtines. Po to buvo stebėtas jų elgesio agreyvumas.
- 63 žiūrėjo smurtines ir agresyvūs, 37 smurt. ir neagr.; 30 nesmurt. ir agresyvūs ir 70 nesmurt. ir neagresyvūs.
- ar TV ir elgesys susiję?

$\alpha = 0.05$

© V.Čekanaavičius, G.Murauskas

162

## SPSS

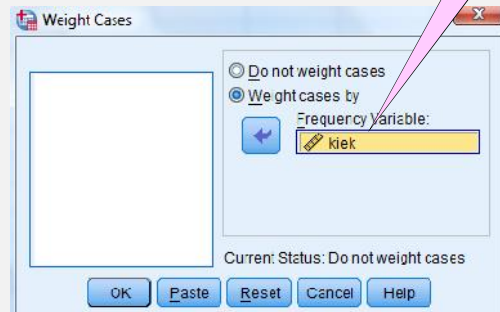
elgesys	tv	kiek
agres	smurt	63.00
neagres	smurt	37.00
agres	nesmurt	30.00
neagres	nesmurt	70.00

gali būti ir simboliniai skaitiniai

© V.Čekanavičius, G.Murauskas

163

## SPSS

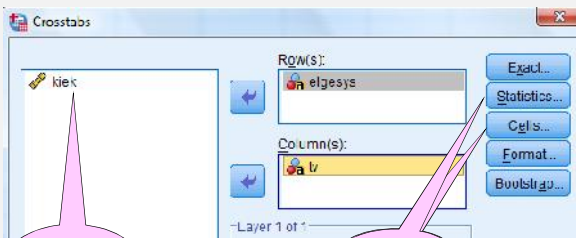


Pasveriams pagal 'kiek'

© V.Čekanavičius, G.Murauskas

164

## SPSS



naų nebekišam!

Po to čia!

Statistics ir Cells užpildome, kaip anksčiau.

© V.Čekanavičius, G.Murauskas

165

## SPSS

		TV		Total	
		nesmurt	smurt		
ELGESYS	agres	Count	30	63	93
		% within ELGESYS	32.3%	67.7%	100.0%
		% within TV	30.0%	63.0%	46.5%
neagr	Count	70	37	107	
	% within ELGESYS	65.4%	34.6%	100.0%	
	% within TV	70.0%	37.0%	53.5%	
Total	Count	100	100	200	
	% within ELGESYS	50.0%	50.0%	100.0%	
	% within TV	100.0%	100.0%	100.0%	

smurtinę TV žiūrėję agresyvesni

© V.Čekanavičius, G.Murauskas

166

## SPSS

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	21.887 <sup>a</sup>	1	.000		
Continuity Correction <sup>b</sup>	20.581	1	.000		
Likelihood Ratio	22.314	1	.000		
Fisher's Exact Test				.000	.000
N of Valid Cases	200				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 46.50.

stat. reikšmingai

© V.Čekanavičius, G.Murauskas

167

## Išvados formulavimas

- Pritaikę chi kvadrato kriterijų, gavome, kad tarp žiūrėjusių smurtinę TV buvo statistiškai reikšmingai daugiau agresyvių vaikų (63%), nei tarp žiūrėjusių nesmurtinę TV (30%),  $p < 0,01$ .

© V.Čekanavičius, G.Murauskas

168

## Mc Nemar kriterijus

© V.Čekanavičius, G.Murauskas

169

## Mc Nemar kriterijus

dažniausiai (bet ne visada):

- taikomas dvireikšmiams kintamiesiems
- gautiems apklausus **tuos pačius** respondentus
- prieš poveikį respondentams (terapiją, reklamą, įkalbinėjimą, filmą ir t.t.)
- ir po poveikio.
- domina, kaip paveikė respondentus

© V.Čekanavičius, G.Murauskas

170

## Mc Nemar kriterijus

Kitas tipinis variantas: ar vienodos proporcijos priklausomiems kategoriniams kintamiesiems.

- Duomenys gauti, apklausus **tuos pačius** respondentus.
- Pvz., ar toks pat procentas respondentų rūko, kaip ir vartoja alkoholį.

© V.Čekanavičius, G.Murauskas

171

## Duomenys

- ♦ Vienas dvireikšmis (už-prieš, taip-ne) kategorinis kintamasis matuotas dviejose priklausomose populiacijose
- ♦ arba vienoje populiacijoje dukart.

© V.Čekanavičius, G.Murauskas

172

## Duomenys

	Prieš	
	+	-
Po	+	-
	<i>a</i>	<i>b</i>
	<i>c</i>	<i>d</i>

© V.Čekanavičius, G.Murauskas

173

## Statistinė hipotezė:

$H_0$  : poveikio **nebuvo** (iš 'už' į 'prieš' nuomonę pakeitė tiek pat, kiek iš 'prieš' į 'už')

$H_1$  : poveikis **buvo**

© V.Čekanavičius, G.Murauskas

174



## Statistinė išvada su **p - reikšme**

$H_0$  **atmetame** (poveikis buvo stat.  
reikšmingas), jei

$$p < 0.05$$

$H_0$  **neatmetame** (poveikis stat.  
nereikšmingas), jei

$$p \geq 0.05$$

Čia 0.05 - reikšmingumo lygmuo

© V.Čekanaavičius, G.Murauskas

175

## SPSS

- dukart apklausėme tuos pačius rinkėjus, prieš TV laidą ir po TV laidos.
- prieš TV už, po TV už 200
- prieš TV už, po TV prieš 30
- prieš TV prieš, po TV už 60
- prieš TV prieš, po TV prieš 100
- ar TV laida paveikė rinkėjus.

© V.Čekanaavičius, G.Murauskas

176

## SPSS

kaip ir  
 $\chi^2$  atveju

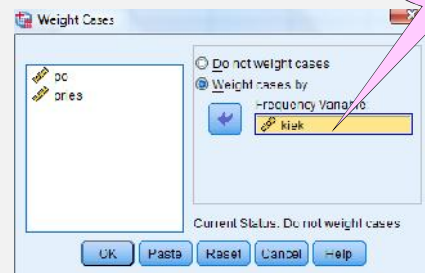
	pries	po	kiek
už	už		200.00
už	Ne		30.00
Ne	už		60.00
Ne	Ne		100.00

© V.Čekanaavičius, G.Murauskas

177

## SPSS

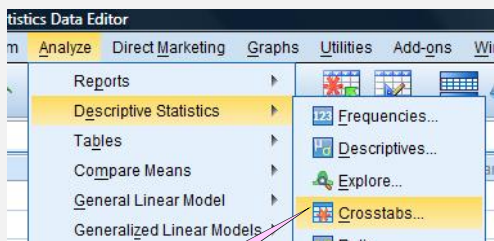
pasveriams



© V.Čekanaavičius, G.Murauskas

178

## SPSS



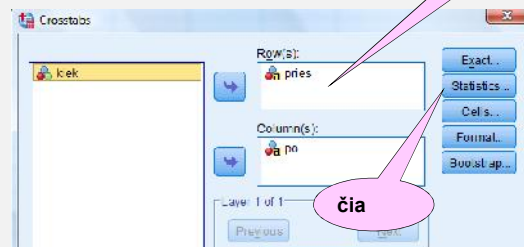
Štai čia!

© V.Čekanaavičius, G.Murauskas

179

## SPSS

kintamieji



čia

© V.Čekanaavičius, G.Murauskas

180

## SPSS

© V.Čekanavičius, G. Murauskas 181

## SPSS

pries \* po Crosstabulation

Count		po		Total
		Ne	už	
pries	Ne	100	60	160
	už	30	200	230
Total		130	260	390

© V.Čekanavičius, G. Murauskas 182

## SPSS

	Value	Exact Sig. (2-sided) <sup>a</sup>
McNemar Test N of Valid Cases	390	.002 <sup>a</sup>

stat. reikšmingai padidėjo rėmėjų

© V.Čekanavičius, G. Murauskas 183

## ANOVA

### Vienfaktorinė dispersinė analizė

#### ANalysis Of VAriance

© V.Čekanavičius, G. Murauskas 184

## ANOVA yra t kriterijaus apibendrinimas keletui nepriklausomų imčių.

Jei dispersijos lygios, **dviems** imtims ANOVA ir t kriterijus duoda tą patį atsakymą.

© V.Čekanavičius, G. Murauskas 185

## PAVYZDŽIAI

- ar lietuviai, latviai ir kareliai tiek pat laiko naršo internete,
- ar priklauso klaidų skaičius nuo išgerto vyno rūšies,
- ar visų kursų studentai **vienodai** intelektualūs,
- ar vidutiniškai vienodai garsiai rėkia L. Ryto ir K. Žalgirio sirgaliai.

© V.Čekanavičius, G. Murauskas 186

## DUOMENYS

- **VISADA DU** stulpeliai
- Vienas – **intervalinis**. Jame yra rezultatas, kurio vidurkius lyginame. (Pvz., 32, 17, 22,...)
- Kitas – **kategorinis**. Jame - grupės kodas. (pvz., 1, 2, 3).

© V. Čekanavičius, G. Murauskas

187

## Statistinė hipotezė:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

$H_1$ : **ne visi vidurkiai lygūs**

© V. Čekanavičius, G. Murauskas

188

## Statistinė išvada su **p - reikšme**

$H_0$  **atmetame** (ne visi vidurkiai lygūs), jei

$$p < 0.05$$

$H_0$  **neatmetame** (visi vidurkiai stat. reikšm. nesiskiria), jei

$$p \geq 0.05$$

Čia **0.05** - reikšmingumo lygmuo

© V. Čekanavičius, G. Murauskas

189

## ANOVA skirtumus rado, kas toliau ??????

Toliau : **POST HOC** testai. Pvz. Bonferroni testas.

Tikslas: Nustatyti **kurių** grupių vidurkiai skiriasi statistiškai reikšmingai.

© V. Čekanavičius, G. Murauskas

190

## Statistinė išvada su **p - reikšme**

**Vidurkiai stat. reikšmingai skiriasi, jei**

$$p < 0.05$$

**Vidurkiai stat. reikšmingai, nesiskiria, jei**

$$p \geq 0.05$$

Čia **0.05** - reikšmingumo lygmuo

© V. Čekanavičius, G. Murauskas

191

## Pavyzdys

Ar visų rasių respondentai vienodai palankiai vertina džiazą?

Tiriame vyresnius nei 60 m respondentus(es)

© V. Čekanavičius, G. Murauskas

192

### SPSS

© V. Čekanavičius, G. Murauskas 193

**Štai čia!**

### SPSS

© V. Čekanavičius, G. Murauskas 194

**Rasė (kategorinis)**

**Jazz (intervalinis)**

**opcijos**

### SPSS

© V. Čekanavičius 195

**aprašoma**

**grafikas**

### SPSS

© V. Čekanavičius, G. Murauskas 196

**post hoc**

### SPSS

© V. Čekanavičius, G. Murauskas 197

**post hoc**

### SPSS

© V. Čekanavičius, G. Murauskas 198

**vidurkiai**

Descriptives										
jazz Jazz Music										
				% Confidence Interval for Mean						
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum		
1 white	304	2.98	1.118	.064	2.85	3.10	1	5		
2 black	32	2.34	1.208	.214	1.91	2.78	1	4		
3 other	8	3.38	.744	.263	2.75	4.00	2	4		
Total	344	2.93	1.134	.061	2.81	3.05	1	5		

## SPSS

### ANOVA

jazz Jazz Music

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13.251	2	6.625	5.279	.006
Within Groups	427.933	341	1.255		
Total	441.183	343			

Požiūris į džiazą statistiškai reikšmingai priklauso nuo respondento rasės ( $p < 0.05$ ).

© V. Čekanaavičius, G. Murauskas

199

## SPSS

### Multiple Comparisons

Dependent Variable: jazz Jazz Music

	(I) race	(J) race	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval		
						Lower Bound	Upper Bound	
Tukey HSD	1 white	2 black	.633	.208	.007	.14	1.12	
		3 other	-.398	.401	.582	-1.34	.55	
		2 black	1 white	-.633	.208	.007	-1.12	-.14
	3 other	1 white	-.398	.401	.582	-1.34	.55	
		2 black	1 white	1.031	.443	.053	-.01	2.07
		3 other	1 white	.398	.401	.582	-.55	1.34
Bonferroni	1 white	2 black	.633	.208	.008	-.13	1.13	
		3 other	-.398	.401	.966	-1.36	.57	
		2 black	1 white	-.633	.208	.008	-1.13	-.13
	3 other	1 white	-.398	.401	.966	-.57	1.36	
		2 black	1 white	1.031	.443	.061	-.03	2.10
		3 other	1 white	.398	.401	.966	-.57	1.36

\*. The mean difference is significant at the 0.05 level.

© V. Čekanaavičius, G. Murauskas

200

## SPSS

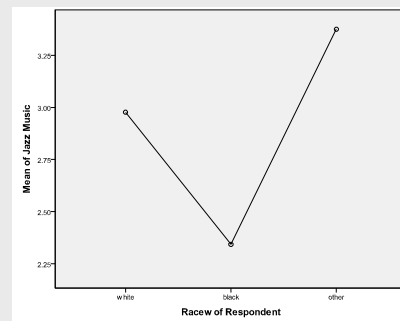
	(I) race	(J) race	Mean Difference (I-J)	Std. Error	Sig.	
Tukey HSD	1 white	2 black	.633*	.208	.007	
		3 other	-.398	.401	.582	
		2 black	1 white	-.633*	.208	.007
	3 other	1 white	-.398	.401	.582	
		2 black	1 white	1.031	.443	.053
		3 other	2 black	.398	.401	.582

Statistiškai reikšmingai skiriasi juodieji nuo baltųjų. Kiti skirtumai statistiškai nereikšmingi.

© V. Čekanaavičius, G. Murauskas

201

## SPSS



© V. Čekanaavičius, G. Murauskas

202

## Blokuotų duomenų dispersinė analizė

Repeated measures ANOVA

© V. Čekanaavičius, G. Murauskas

203

Blokuotų duom. ANOVA yra porinio  $t$  kriterijaus apibendrinimas keletui priklausomų imčių.

Neparametrinis analogas - Friedman testas.

© V. Čekanaavičius, G. Murauskas

204

## PAVYZDŽIAI

- kraujospūdis matuojamas po mėnesio, po pusmečio ir po metų. Ieškoma skirtumų.
- stebime studento IQ kitimą pereinant iš kurso į kursą.
- ta pati žiuri vertina kelias gražuoles.

© V.Čekanavičius, G.Murauskas

205

## Duomenys

Keli intervaliniai kintamieji (stulpeliai).

Būtent jų vidurkius ir lyginame.

© V.Čekanavičius, G.Murauskas

206

## Statistinė hipotezė:

$H_0$ : visi vidurkiai lygūs

$H_1$ : ne visi vidurkiai lygūs

© V.Čekanavičius, G.Murauskas

207

## Statistinė išvada su **p - reikšme**

$H_0$  **atmetame** (ne visi vidurkiai lygūs), jei

$$p < 0.05$$

$H_0$  **neatmetame** (visi vidurkiai stat. reikšm. nesiskiria), jei

$$p \geq 0.05$$

Čia **0.05** - reikšmingumo lygmuo

© V.Čekanavičius, G. Murauskas

208

## POST HOC testai

Taikomi ir interpretuojami, kaip ir vienfaktorinėje ANOVA.

© V.Čekanavičius, G.Murauskas

209

## Statistinė išvada su **p - reikšme**

Vidurkiai stat. reikšmingai skiriasi, jei

$$p < 0.05$$

Vidurkiai stat. reikšmingai, nesiskiria, jei

$$p \geq 0.05$$

Čia **0.05** - reikšmingumo lygmuo

© V.Čekanavičius, G. Murauskas

210

## Pavyzdys

Tris gražuoles įvertino 15 ekspertų.  
Arvisos gražuolės įvertintos vienodai gerai?

H0: vidutiniai gražuolių įvertinimai stat.  
reikšmingai nesiskiria

H1: skiriasi

## SPSS

Duomenys

	graz1	graz2	graz3
1	15.00	18.00	10.00
2	8.00	6.00	11.00
3	20.00	18.00	20.00
4	17.00	12.00	17.00
5	11.00	7.00	8.00
6	15.00	11.00	17.00
7	19.00	18.00	17.00
8	16.00	10.00	18.00
9	18.00	14.00	18.00
10	20.00	16.00	16.00
11	17.00	10.00	18.00
12	19.00	15.00	18.00
13	16.00	16.00	19.00

© V.Čekanašičius, G.Murauskas

212

## SPSS

Analyze Direct Marketing Graphs Utilities Add-ons Window

- Reports
- Descriptive Statistics
- Tables
- Compare Means
- General Linear Model**
  - Univariate...
  - Multivariate...
  - Repeated Measures...**
  - Variance Components...
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression

Štai čia!

© V.Čekanašičius, G.Murauskas

213

## SPSS

pseudo-vardas

pseudo-vardas:  
tai kas vienija  
kintamuosius  
(nesvarbus)

Repeated Measures: Define Factor(s)

Within-Subject Factor Name: factor1

Number of Levels: 3

Add Change Remove

Tada Define

Kintamųjų (stulpelių) skč.

Add

214

## SPSS

Repeated Measures

Within-Subjects Variables (factor1):

- graz1(1)
- graz2(2)
- ...?(3)

Model... Contrasts... Plots... Post Hoc... Save... Options...

Between-Subjects Factor(s):

opcijos

kintamieji

© V.Čekanašičius, G.Murauskas

215

## SPSS

Repeated Measures: Post Hoc

Factor(s) and Factor interactions:

- (OVERALL)
- factor1

Display Means for factor1

Compare main effects

Confidence interval adjustment

Descriptive statistics

Transformation matrix

perkėlėm

aprašoma

varna

post hoc

© V.Čekanašičius, G.Murauskas

216

## SPSS

© V.Čekanaavičius, G.Murauskas 217

## SPSS

© V.Čekanaavičius, G.Murauskas 218

## SPSS

kaip užkoduoti

Measure: MEASURE_1	
FACTOR1	Dependent Variable
1	GRAZ1
2	GRAZ2
3	GRAZ3

aprašoma

	Mean	Std. Deviation	N
GRAZ1	15.8667	3.96172	15
GRAZ2	12.2667	3.99046	15
GRAZ3	15.4000	4.08482	15

© V.Čekanaavičius, G.Murauskas 219

## SPSS

Tests of Within-Subjects Effects

Measure: MEASURE_1						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
FACTOR1	114.978	2	57.489	19.705	.000 <sup>a</sup>	
Greenhouse-Geisser	114.978	1.777	64.687	19.705	.000	
Huynh-Feldt	114.978	2.000	57.489	19.705	.000	
Lower-bound	114.978	1.000	114.978	19.705	.001	
Error(FACTOR1)	81.689	28	2.917			
Greenhouse-Geisser	81.689	24.884	3.283			
Huynh-Feldt	81.689	28.000	2.917			
Lower-bound	81.689	14.000	5.835			

p-reikšmė

© V.Čekanaavičius, G.Murauskas 220

## SPSS

post hoc

Measure: MEASURE_1						
(I) FACTOR	(J) FACTOR	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval Difference	
					Lower Bound	Upper Bound
1	2	3.600	.542	.000	2.128	5.072
	3	-.467	.593	1.000	-1.144	2.077
2	1	-3.600	.542	.000	-5.072	-2.128
	3	-3.133	.723	.002	-5.098	-1.169
3	1	-.467	.593	1.000	-2.077	1.144
	2	3.133	.723	.002	1.169	5.098

prisimenam kodus

© V.Čekanaavičius, G.Murauskas 221

## Išvada

- Pritaikius blokuotųjų duomenų ANOVA gavome, kad ne visos gražuolės vertinimas vienodai ( $p < 0,05$ ). Bonferoni kriterijus parodė, kad antroji gražuolė vertinama statistiškai reikšmingai blogiau nei pirmoji ir antroji.



**ANOVA: faktorių priklausomybės stiprumo matavimas eta kvadratu**

**Koeficientas  $\eta^2$  :**

$$\eta^2 = \frac{\text{skirtumai tarp grupių}}{\text{visi skirtumai}}$$

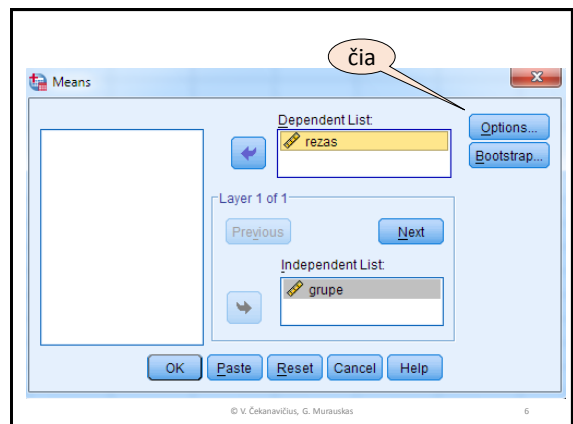
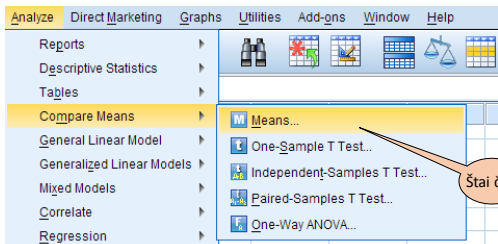
**Koeficientas  $\eta^2$  :**

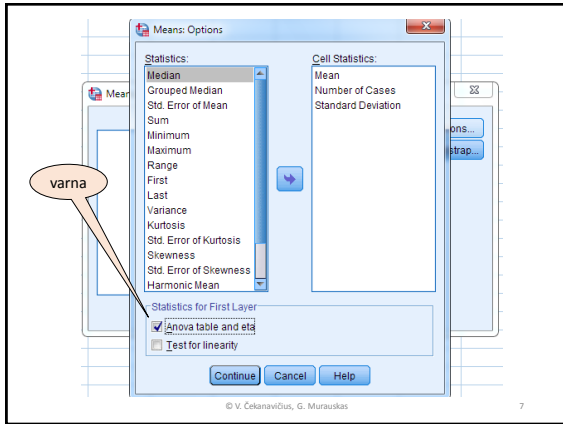
$$\eta^2 = \frac{SSB}{SST} = \frac{SSB}{SSB + SSW}$$

**Koeficiento  $\eta^2$  interpretacija:**

- parodo, kurią dalį duomenų sklaidos lemia populiacijų skirtumai.
- ~ kiek rezultatai įtakoja grupę.

**SPSS**





© V. Čekanaivičius, G. Murauskas

7

**SPSS**

čia

**Measures of Association**

	Eta	Eta Squared
rezas * grupe	,887	,786

© V. Čekanaivičius, G. Murauskas

© V. Čekanaivičius, G. Murauskas

8

## VIDURKIŲ TRENDAI

**Vidurkių trendas** - tai vidurkių kitimo tendencija (ką primena vidurkių grafikas).

© V. Čekanaivičius, G. Murauskas

© V. Čekanaivičius, G. Murauskas

9

Tiesinis

Kvadratinis

Tiesinis

Kvadratinis

© V. Čekanaivičius, G. Murauskas

© V. Čekanaivičius, G. Murauskas

10

### Kada taikoma trendų analizė?

- Kai nepriklausomas grupavimo kintamasis yra gautas iš intervalinio.
- Grupės NR atspindi vis didesnę (mažesnę) kažkokios savybės kieki.

**Fux, oberfux, senior, student (patirtis).  
1bokalas, 2bok., 3bok. (alkoholis).**

© V. Čekanaivičius, G. Murauskas

© V. Čekanaivičius, G. Murauskas

11

### Statistinė hipotezė:

**H<sub>0</sub>: vidurkiai nesudaro tiesinio trendo,**

**H<sub>1</sub>: vidurkiai sudaro tiesinį trendą.**

© V. Čekanaivičius, G. Murauskas

© V. Čekanaivičius, G. Murauskas

12

**Statistinė hipotezė:**

$H_0$ : vidurkiai nesudaro kvadratinio trendo,  
 $H_1$ : vidurkiai sudaro kvadratinį trendą.

ir t.t. Kiekvienam trendui sava hipotezė.

**Statistinė išvada su p - reikšme**

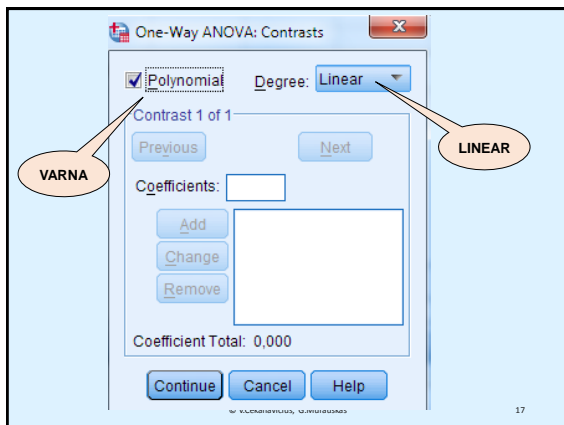
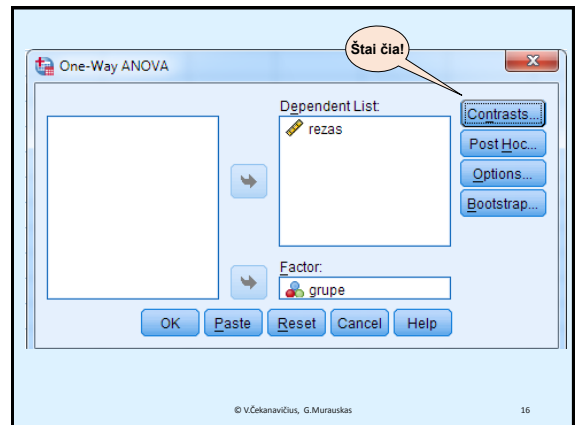
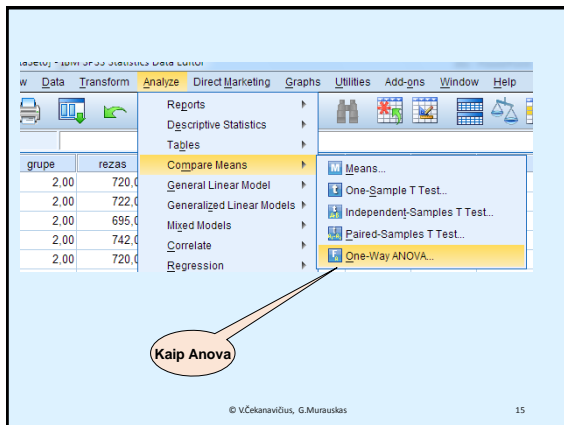
$H_0$  atmetame (tendras TINKA), jei

$$p < 0.05$$

$H_0$  neatmetame (tendras netinka), jei

$$p \geq 0.05$$

Čia 0.05 - reikšmingumo lygmuo

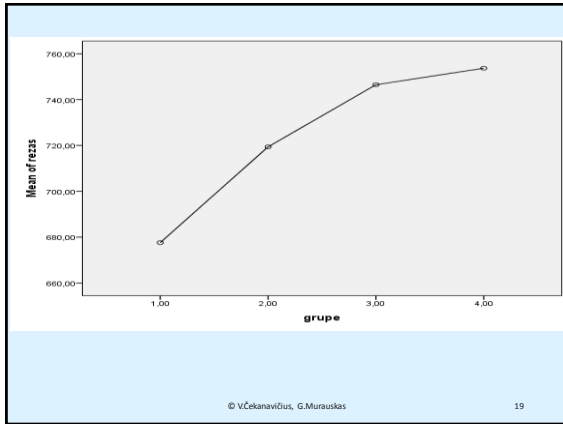


**SPSS (grupės lygūs)**

**ANOVA p-reikšmė**

		Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)	35621,000	3	11873,667	58,899	.000
	Linear Term	32614,580	1	32614,580	161,783	.000
	Contrast	3006,420	2	1503,210	7,457	.002
Within Groups		7257,400	36	201,594		
Total		42878,400	39			

**Tendras tinka**



### SPSS (grupės nelygios)

ANOVA

rezas		Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)	38474.821	3	12158.274	81.430	.000
	Linear Term	32962.061	1	32962.061	166.541	.000
	Unweighted					.000
	Weighted	33221.961	1	33221.961	167.855	.000
	Deviation	3252.859	2	1626.430	8.218	.001
Within Groups		7323.082	37	197.921		
Total		43797.902	40			

Arba visi respondentai lygiavertčiai (unweighted), arba visos grupės (weighted). Renkamės tą atsakymą, kuris mums atrodo logiškescnis.

© V.Čekanaavičius, G.Murauskas 20

One-Way ANOVA: Contrasts

Polynomial Degree: Quadratic

Contrast 1 of 1

menys Coefficients:

Buttons: Previous, Next, Add, Change, Remove, Contrasts..., Post Hoc..., Options..., Bootstrap...

Tiesinis ir kvadratinis

© V.Čekanaavičius, G.Murauskas 21

### SPSS

ANOVA

rezas		Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)	35621.000	3	11873.667	58.899	.000
	Linear Term	32614.580	1	32614.580	161.783	.000
	Deviation	3006.420	2	1503.210	7.457	.002
	Quadratic Term	2992.900	1	2992.900	14.846	.000
	Deviation	13.520	1	13.520	.067	.797
Within Groups		7257.400	36	201.594		
Total		42878.400	39			

© V.Čekanaavičius, G.Murauskas 22

## 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  $\mu_{1a}$ ,  $\mu_{2a}$ ,  $\mu_{3a}$  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 pasiskirstęsi.
- **Grupavimo** (nepriklausomas, faktorius) kintamasis – kategorinis. Jo dėka žinome, kuriai grupei koks respondentas priklauso.

## PAVYZDYS

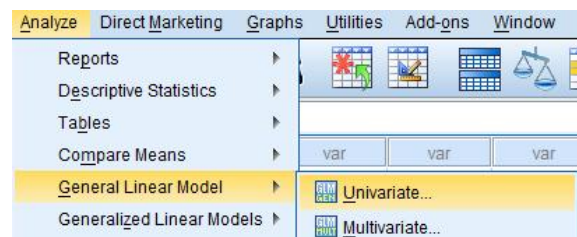
- Tiriame, ar statistiškai reikšmingai skiriasi trijų grupių testo rezultatai (REZ, intervalinis kintamasis). Atsižvelgsime į respondentų motyvaciją (MOT, intervalinis kintamasis). Grupę nurodyto kintamasis Grupe (trireikšmis kategorinis kintamasis).
- Duomenis galima rasti knygos V. Čekanavičius, G. Murauskas, Statistika ir jos taikymai. III, pirmos dalies šeštajame skyrelyje (1.1.9 pavyzdys).

## SPSS duomenys

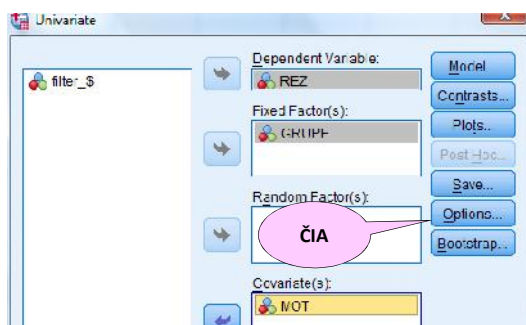
- Reikia 3 stulpelių. Juose nurodoma kiekvieno respondento grupė (Grupe), testo rezultatas (REZ) ir motyvacija (MOT).

	GRUPE	REZ	MOT
1	1.00	4.00	1.00
2	1.00	6.00	4.00
3	1.00	6.00	3.00
4	1.00	8.00	4.00
5	1.00	6.00	5.00
6	1.00	9.00	6.00
7	1.00	10.00	8.00
8	1.00	11.00	0.00

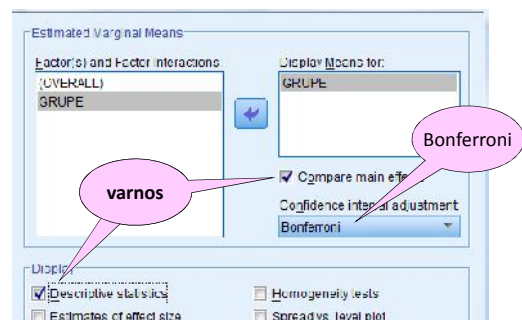
## ANCOVA SU SPSS



## ANCOVA SU SPSS



## ANCOVA SU SPSS



## ANCOVA rezultatai

### Descriptive Statistics

Dependent Variable: REZ

GRUPE	Mean	Std. Deviation	N
1.00	8.4000	3.09839	10
2.00	10.6000	3.06232	10
3.00	11.6000	3.13404	10
Total	10.2000	3.28424	30

## ANCOVA rezultatai

### Tests of Between-Subjects Effects

Dependent Variable: REZ

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	293.808 <sup>a</sup>	3	97.936	134.074	.000
Intercept	74.779	1	74.779	102.872	.000
MOT	240.208	1	240.208	328.843	.000
GRUPE	54.558	2	27.279	37.344	.000
Error	18.992	26	.730		
Total	3434.000	30			
Corrected Total	312.800	29			

a. R Squared = .939 (Adjusted R Squared = .932)

p-reikšmė

## ANCOVA rezultatai

- Tarkime, kad pasirinkome reikšmingumo lygmenį  $\alpha = 0,05$ . Matome, kad  $p < 0,05$
- Darome išvadą, kad tarp tarp trijų grupių vidutinių testo rezultatų yra statistiškai reikšmingai besiskiriančių.
- Liko atsakyti į klausimą, kurių grupių vidutiniai rezultatai skiriasi, t.y. pritaikyti post hoc testą.

## ANCOVA rezultatai

Naudojantis Bonferonio kriterijumi, lyginami „pataisyti“ rezultatų vidurkiai, kai atsižvelgiama į motyvacia

### Estimates

Dependent Variable: REZ

GRUPE	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	8.363 <sup>a</sup>	.270	7.808	8.919
2.00	10.673 <sup>a</sup>	.270	10.118	11.229
3.00	11.563 <sup>a</sup>	.270	11.008	12.119

a. Covariates appearing in the model are evaluated at the following values: MOT = 5.766

## Post hoc vidurkių lyginimai

Ką lyginame

### Pairwise Comparisons

(I) GRUPE	(J) GRUPE	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>b</sup>	
					Lower Bound	Upper Bound
					1.00	2.00
1.00	3.00	-3.200 <sup>*</sup>	.392	.000	-4.178	-2.222
2.00	1.00	2.310 <sup>*</sup>	.392	.000	1.332	3.288
2.00	3.00	-.890	.392	.084	-1.868	.088
3.00	1.00	3.200 <sup>*</sup>	.392	.000	2.222	4.178
3.00	2.00	.890	.392	.084	-.088	1.968

Based on estimated marginal means

\*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni

## ANCOVA rezultatai

- Statistiškai reikšmingai skiriasi tų grupių vidurkiai, kuriuos lyginant  $p < 0,05$ . Tokių vidurkių skirtumus SPSS automatiškai pažymi žvaigždute.
- Gauta, kad pirmoji grupė statistiškai reikšmingai skiriasi nuo likusiųjų, o antroji ir trečioji grupės nesiskiria.

## ANCOVA rezultatų aprašymo pavyzdys

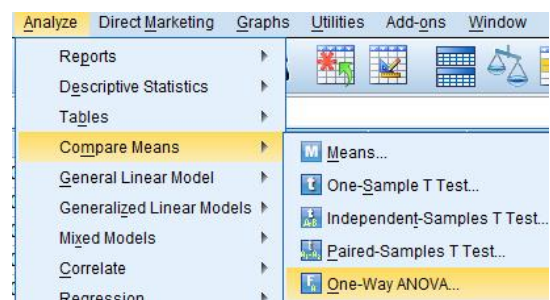
Tyrėme, ar statistiškai reikšmingai skiriasi trijų grupių vidutiniai testo rezultatai. Atsižvelgėme į respondentų motyvaciją. Taikėme kovariančių analizę (ANCOVA) ir Bonferonio kriterijų. Gavome, kad pirmoji grupė statistiškai reikšmingai skiriasi nuo likusiųjų, o antroji ir trečioji grupės nesiskiria.

## PAPILDOMI TYRIMAI

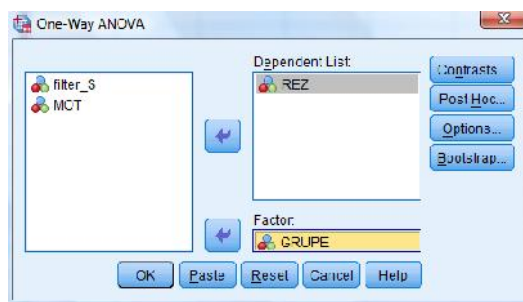
## Palyginimas su ANOVA

- Taikant ANCOVA patartina vidurkius palyginti ir be kovariantės. Tam naudojama vienfaktorė dispersinė analizė (ANOVA).
- SPSS programa ANOVA galima atlikti ne vienu būdu. Naudosime pačią paprasčiausią parinktį.

## Palyginimas su ANOVA



## Palyginimas su ANOVA



## Palyginimas su ANOVA

Vidurkių skirtumai statistiškai nereikšmingi. Primename, kad, atsižvelgus į motyvaciją, statistiškai reikšmingų skirtumų atsirado.

ANOVA

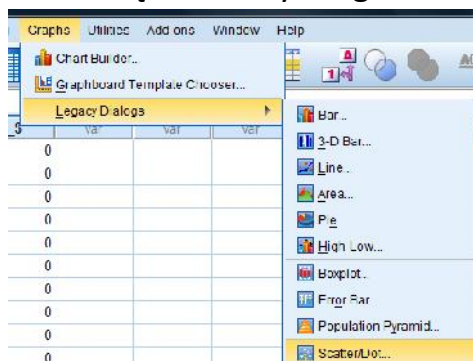
REZ	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	53.600	2	26.800	2.792	.079
Within Groups	259.200	27	9.600		
Total	312.800	29			



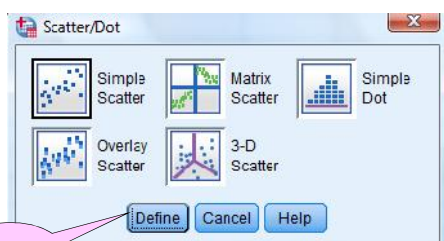
## Reikšmių išsibarstymo grafikas

- Norint nustatyti testo rezultatų ir motyvacijos priklausomybę, verta nubraižyti abiejų kintamųjų grafiką. Parodysime, kaip tai galima padaryti naudojantis SPSS 17 programa.

## Reikšmių išsibarstymo grafikas

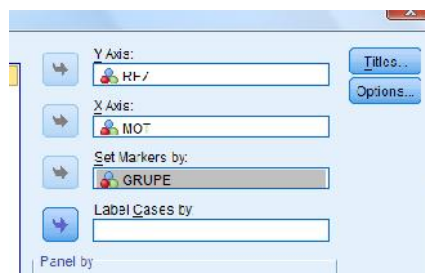


## Reikšmių išsibarstymo grafikas



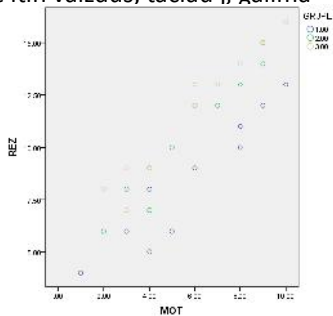
Čia

## Reikšmių išsibarstymo grafikas

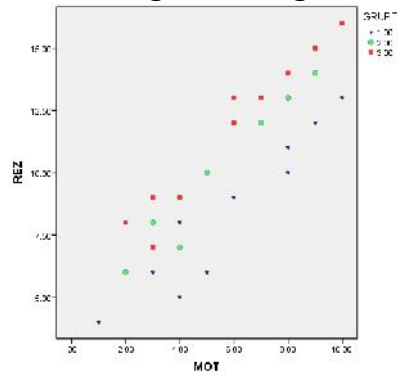


## Reikšmių išsibarstymo grafikas

- Grafikas – ne itin vaizdus, tačiau jį galima redaguoti.



## Suredaguotas grafikas



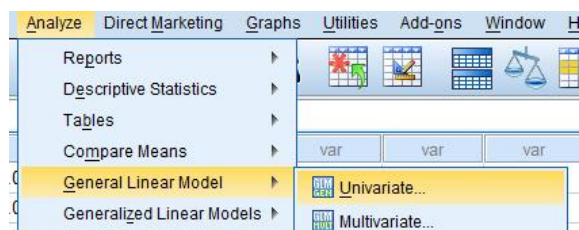
## Į ką grafike atkreipti dėmesį

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

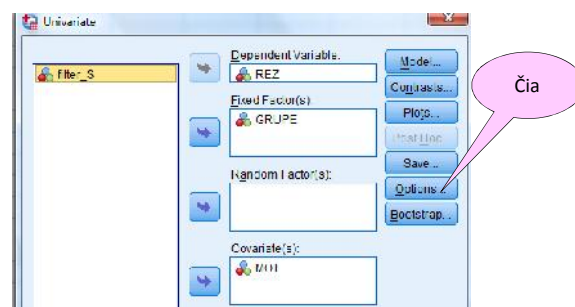
## Kintamųjų sąveika

- Statistiškai reikšminga motyvacijos ir grupės sąveika parodo, kad skirtingose grupėse motyvacija skirtingai veikia rezultatus.
- Klasikinė ANCOVA daroma, kai nėra grupės ir kovariantės sąveikos (interakcijos).
- Parodysime, kaip naudojant SPSS galima patikrinti, ar sąveikos nėra.
- Pradžia nesiskiria nuo įprastinės ANCOVA.

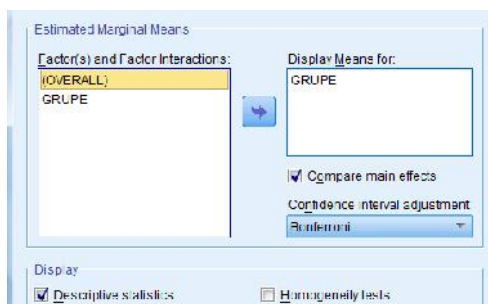
## ANCOVA su kintamųjų sąveika



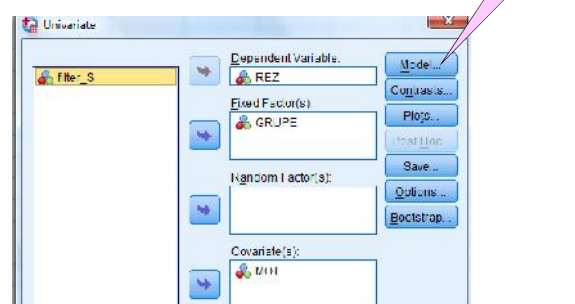
## ANCOVA su kintamųjų sąveika



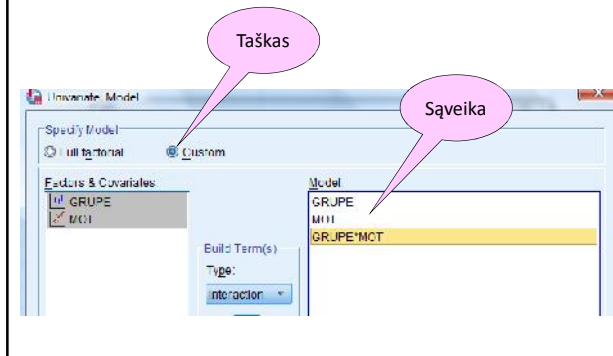
## ANCOVA su kintamųjų sąveika



## ANCOVA su kintamųjų sąveika



## ANCOVA su kintamųjų sąveika



## Sąveikos nėra

Tests of Between-Subjects Effects

Dependent Variable: REZ

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	294.814 <sup>a</sup>	5	58.963	78.680	.000
Intercept	72.200	1	72.200	96.344	.000
GRUPE	6.102	2	3.051	4.071	.030
MOT	240.283	1	240.283	320.632	.000
GRUPE * MOT	1.005	2	.503	.67	.520
Error	17.986	24	.749		
Total	3434.000	30			
Corrected Total	312.800	29			

a. R Squared = .943 (Adjusted R Squared = .931)

## Ką daryti, jeigu sąveika yra

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

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

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

## 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.

## Tiesinė regresija

© V.Čekanašius, G.Murauskas

223

**Tiesinė regresija** - toks modelis, kai vieną (**priklausomą**) kintamąjį įtakoja vienas arba keletas (**nepriklausomų**) kintamųjų.

© V.Čekanašius, G.Murauskas

224

### PAVYZDŽIAI

- ar poegzamininis dirglumas priklauso nuo gauto pažymio, nuo egzamino trukmės ir nuo ko labiau,
- kaip antikvarinės keramikos kaina priklauso nuo jos senumo ir aukciono dalyvių skaičiaus,
- ar diplomo pažymių vidurkis ir komunikabilumas gali padėti prognozuoti būsimą atlyginimą.

© V.Čekanašius, G.Murauskas

225

### PAVYZDŽIAI

- išmatavome IQ pirmame kurse. Ar galima prognozuoti **koks** bus studento diplomo pažymių vidurkis?
- **kiek** padidės pelnas, padidinus reklamos išlaidas 10 000 Lt?
- **kokį** vidutinį valstybinio egzamino balą galima prognozuoti mokiniui per baigiamąjį kontrolinį surinkusiam 87 tšk.?

© V.Čekanašius, G.Murauskas

226

### Duomenys

- $(X_1, X_2, X_3, \dots, X_k, Y)$
- tai, ką prognozuosime - **priklausomas kintamasis** (**Y**, kaina)
- pagal ką - **nepriklausomi kintamieji** (regresoriai) (**X-ai**, metai, dalyvių skaičius).
- **Keli (dažniausiai intervaliniai) kintamieji (stulpeliai).**

© V.Čekanašius, G.Murauskas

227

### MODELIS

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + e$$

Y priklauso nuo stebimų X-ų ir nuo dar kažkokių nestebimų dalykų e.  
e – atsitiktinis dydis (normalus) su 0 vidurkiu ir nežinoma dispersija.

© V.Čekanašius, G.Murauskas

228

## Visi kintamieji **intervaliniai\***.

- \* išskyrus **pseudokintamuosius**.
- **X**-ai vienas kito neveikia,
- **X**-ai veikia **Y**-ą.
- vieno **X**-o per kitus neišreikši.

© V.Čekanaavičius, G.Murauskas

229

## Standartinis tyrimas:

- $R^2$ , ANOVA, t-testai, VIF ir grafikai padės nustatyti reikalingus X-us.
- R parodys, kaip Y priklauso nuo visų X-ų iškart.
- B-koeficientai padės sudaryti regresijos funkciją.
- **Beta-koeficientai** padės nustatyti, kurie kintamieji svarbesni.

© V.Čekanaavičius, G.Murauskas

230

## Geram regresijos modeliui:

- $R^2 \geq 0,20$ .
- ANOVA  $p < 0,05$ .
- Visų t testų  $p < 0,05$ .
- Visi VIF  $\leq 4$  (nėra multikolinearumo problemos).
- Visos Kuko mato reikšmės  $\leq 1$ .
- Koeficientų ženklai atitinka koreliacijas.
- Sprendžiant pagal histogramą, liekamosios paklaidos normalios.

© V.Čekanaavičius, G.Murauskas

232

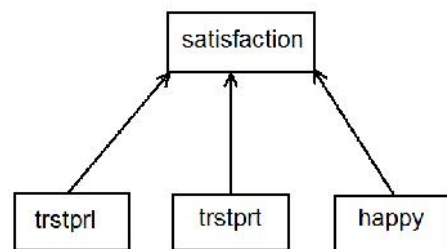
## Pavyzdys:

- Tirsime, kaip 20-40 m estų politinės padėties šalyje vertinimas (satisfaction) priklauso nuo požiūrio į parlamentą, partijas ir laimingumo (*trstprl*, *trstprt*, *happy*). Visi kintamieji matuojami 10 balų skalėje (1 – labai nepasitiki).

© V.Čekanaavičius, G.Murauskas

232

## Pavyzdys:



© V.Čekanaavičius, G.Murauskas

233

## SPSS

Štai čia!

© V.Čekanaavičius, G.Murauskas

234

### SPSS

235

### SPSS

© V.Čekanaavičius, G.Murauskas 236

### SPSS

237

### SPSS

© V.Čekanaavičius, G.Murauskas 238

### SPSS

Descriptive Statistics			
	Mean	Std. Deviation	N
satisfaction	4.0397	2.16872	517
trstprt	3.80	2.287	517
happy	7.07	1.837	517
trstprt	3.24	1.941	517

Respondentai gana laimingi ir nelabai patenkinti institucijomis.

© V.Čekanaavičius, G.Murauskas 239

### SPSS

Model Summary <sup>a</sup>				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.703 <sup>a</sup>	.495	.492	1.54608

$R^2 = 0.495$ , pakankamas (>> 0.20)

© V.Čekanaavičius, G.Murauskas 240

## SPSS

### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1200.681	3	400.227	167.434	.000 <sup>a</sup>
	Residual	1226.256	513	2.390		
	Total	2426.937	516			

Yra statistiškai reikšmingų kintamųjų

gerai, nes  $p < 0.05$

© V.Čekanavičius, G.Murauskas

241

## SPSS (koeficientai)

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients		Collinearity Statistics		
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	-.039	.278		-.139	.889		
	trstprl	.374	.042	.394	8.838	.000	.496	2.017
	happy	.228	.038	.193	5.983	.000	.942	1.062
	trstprt	.322	.050	.288	6.439	.000	.492	2.032

© V.Čekanavičius, G.Murauskas

242

## SPSS (modelis)

### Regresijos modelio lygtis

Model		Unstandardized Coefficients
	B	
1	(Constant)	-.039
	trstprl	.374
	happy	.228
	trstprt	.322

satisfaction =  
- 0,039  
+ 0,374 trstprl  
+ 0,228 happy  
+ 0,322 trstprt

© V.Čekanavičius, G.Murauskas

243

## Modelio lygtis

satisfaction = - 0,039 + 0,374 trstprl + 0,228 happy + 0,322 trstprt

- Kai trstprl padidėja vienetu, satisfaction padidėja 0,374 balo.
- Prognozuojant, į lygtį tiesiog įstatomos konkrečios kintamųjų reikšmės.

## SPSS (koeficientai)

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients
		B	Std. Error	Beta
1	(Constant)	-.039	.278	
	trstprl	.374	.042	.394
	happy	.228	.038	.193
	trstprt	.322	.050	.288

Beta koeficientai

Matome, kad modelyje mažiausiai svarbus kintamasis happy

© V.Čekanavičius, G.Murauskas

245

## SPSS (koeficientai)

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.
		B	Std. Error	Beta			
1	(Constant)	-.039	.278		-.139	-.139	.889
	trstprl	.374	.042	.394	8.838	8.838	.000
	happy	.228	.038	.193	5.983	5.983	.000
	trstprt	.322	.050	.288	6.439	6.439	.000

t testai

Visi kintamieji statistiškai reikšmingi ( visi  $p < 0.05$ ).

## SPSS (koeficientai)

Coefficients <sup>a</sup>					
	Standardized Coefficients			Collinearity Statistics	
	Beta	t	Sig.	Tolerance	VIF
78		-.139	.889		
12	.394	8.838	.000	.496	2.017
38	.193	5.983	.000	.942	1.062
50	.288	6.439	.000	.492	2.032

Multikolinearumo nėra (visi VIF < 4).

## SPSS (koeficientai)

Residuals Statistics <sup>a</sup>					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-.4182	8.5548	4.0397	1.52542	517
-----	-----	-----	-----	-----	-----
Mahal. Distance	.077	25.501	2.994	2.663	517
Cook's Distance	.000	.138	.002	.008	517
Centered Leverage Value	.000	.049	.006	.005	517

Visos Kuko mato reikšmės < 1.  
Sprendžiant pagal Kuko matą, išskirčių nėra.

© V.Čekanašičius, G.Murauskas

248

## SPSS

Galutinė modelio tyrimo išvada:  
regresijos modelis tinka duomenims.

© V.Čekanašičius, G.Murauskas

249

## Logistinė regresija

### Logistic Regression

© V.Čekanašičius, G.Murauskas

250

**Logistinė regresija** - toks modelis, kai vieną (priklusomą) dvireikšmį kintamąjį įtakoja vienas arba keletas (nepriklusomų) kintamųjų.

© V.Čekanašičius, G.Murauskas

251

## PAVYZDŽIAI

- pagal paciento svorį ir kraujo tyrimus nustatyti tikimybę susirgti diabetu,
- pagal testų rezultatus nustatyti reiks kompiuteriui garantinio remonto,
- ar pagal agresyvumą bei plepumą galima atskirti vyrą nuo moters.
- ar žinant rinkėjo pajamas ir amžių, galima numatyti balsuos už kandidatą ar nebalsuos.

© V.Čekanašičius, G.Murauskas

252



## Duomenys

- $(X_1, X_2, X_3, \dots, X_k, Y)$
- **Vienas** kintamasis (**stulpelis**) **Y dvireikšmis** – jo elgesį ir modeliuojame. Dar vadinamas **priklausomu kintamuoju**
- Kiti (kitas) kintamieji intervaliniai (dažniausiai) - **nepriklausomi kintamieji**. Jie, mūsų nuomone, daro įtaką Y'ui.

© V.Čekanavičius, G.Murauskas

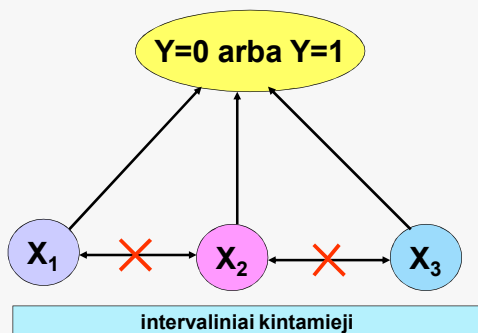
253

- priklausomas kintamasis **Y dvireikšmis** (0 arba 1).
- **X**-ai intervaliniai.
- prognozuojama tikimybė  **$P(Y=1)$** .
- jei **Y** įgyja kitokias dvi reikšmes - jis perkoduojamas.
- Nei viena iš dviejų Y reikšmių neturi kartotis daugiau nei **80 %** duomenų.

© V.Čekanavičius, G.Murauskas

254

## Modelio schema:



© V.Čekanavičius, G.Murauskas

255

Prognozuojame **tikimybę**  $P(Y=1)$ .  
Surandame:

$$z(x) = a + b_1x_1 + \dots + b_kx_k$$

$$\hat{P}(Y_m = 1) = \frac{e^{z(x)}}{1 + e^{z(x)}}$$

© V.Čekanavičius, G.Murauskas

256

Kategoriją **prognozuoti** galima pagal  **$z(x)$**

- jei  **$z(x) > 0$** , tai  **$Y=1$**  (ir pasižiūrime, ką tai reiškia).
- jei  **$z(x) < 0$** , tai  **$Y=0$**  (ir pasižiūrime, ką tai reiškia).
- Jei  **$z(x)=0$** , tai metame monetą.

© V.Čekanavičius, G.Murauskas

257

## Modelio suderinamumas su duomenimis

Ar tinka logistinė regresija nustatome analizuodami

- klasifikacinę lentelę,
- $\chi^2$  kriterijų,
- Voldo kriterijų,
- Determinacijos koeficientus.

© V.Čekanavičius, G.Murauskas

258

## $\chi^2$ kriterijus:

- Tikrina hipotezę:  
 $H_0$ : visi  $b_m = 0$   
 $H_1$ : ne visi  $b_m = 0$ .
- Kitais žodžiais:  
 $H_0$ : tikimybė nuo x-ų nepriklauso  
 $H_1$ : tikimybė nuo x-ų priklauso.
- Tik nežinome nuo kurių x-ų.

© V.Čekanaavičius, G.Murauskas

259

## Statistinės išvados su p- reikšme:

$H_0$  atmetame (logistinė regresija galbūt tinka), jei

$$p < 0.05$$

$H_0$  neatmetame (logistinė regresija netinka), jei

$$p \geq 0.05$$

Čia 0.05 - reikšmingumo lygmuo.

© V.Čekanaavičius, G.Murauskas

260

## Voldo testai koeficientams $b_m$

- Ieškome nesvarbių x-ų.
- Tikrina hipotezę:  
 $H_0$ :  $b_m = 0$   
 $H_1$ :  $b_m \neq 0$
- Jei nulinės hipotezės neatmetame-tai kintamasis 'įtartinas' (modelyje galbūt nereikalingas).

© V.Čekanaavičius, G.Murauskas

261

## Statistinės išvados apie kintamuosius (Wald kriterijus) su p- reikšme:

$H_0$  atmetame (kintamasis modeliui tinka), jei

$$p < 0.05$$

$H_0$  neatmetame (kintamasis 'įtartinas'), jei

$$p \geq 0.05$$

Čia 0.05 - reikšmingumo lygmuo.

© V.Čekanaavičius, G.Murauskas

262

## Ką daryti su 'įtartinais' kintamaisiais?

- pakartojame regresijos modelį be 'įtartino' kintamojo.
- tiriamo klasifikacinę lentelę.
- jei klasifikavimo tikslumas praktiškai nepakito - kintamąjį šaliname.

© V.Čekanaavičius, G.Murauskas

263

## Determinacijos koeficientai:

- Jų yra net keli - ir visi pseudo.
- Dažniausiai taikomi Kokso-Snelo arba Nagelkerkės koeficientai.
- Kuo  $R^2$  didesnis, tuo modelis geresnis.
- Pavyzdyje apie įskaitą Nagelkerkės koeficientas yra 0,565.

© V.Čekanaavičius, G.Murauskas

264

## Pseudokintamieji

- Pseudokintamieji naudojami, kaip ir tiesinėje regresijoje.
- Pseudokintamieji koduojami, kaip ir tiesinėje regresijoje (0 ir 1).
- Jei kategorijų daugiau nei 2, reikia net kelių pseudokintamųjų.
- SPSS perkoduoja automatiškai, jei paprašome.

© V.Čekanašvičius, G.Murauskas

265

## Labai gerai duomenims tinkančiame modelyje:

- Chi kvadrato  $p < 0,05$ .
- Visiems regresoriams Voldo kriterijaus  $p < 0,05$ .
- Koeficientų ženklai neatrodo nelogiški.
- Teisingai klasifikuojama ne mažiau 50% atvejų, kai  $Y = 1$  ir kai  $Y = 0$ .
- Pasirinktas determinacijos koeficientas  $\geq 0,20$ .

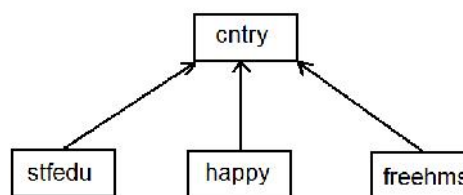
## Pavyzdys:

- Ar padeda atskirti Estijos ir Portugalijos 20 – 30 metų vyrus tokie klausimai: *stfedu* – požiūris į šalies švietimo sistemą, *happy* – laimingumas, *freehms* – gėjai ir lesbietės gali gyventi, kaip nori.

© V.Čekanašvičius, G.Murauskas

267

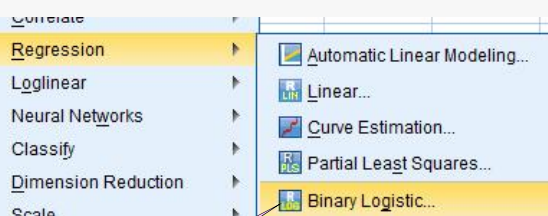
## Modelio schema:



© V.Čekanašvičius, G.Murauskas

268

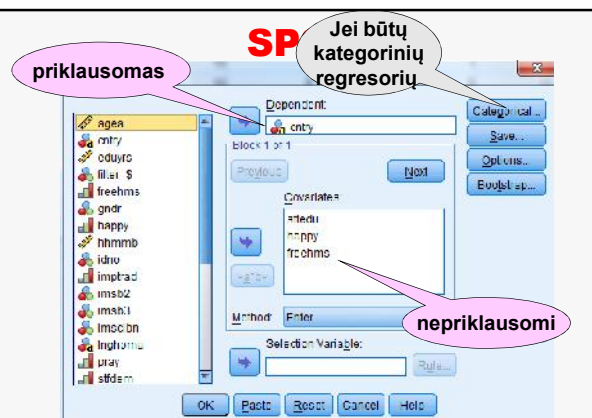
## SPSS



Štai čia!

© V.Čekanašvičius, G.Murauskas

269



© V.Čekanašvičius, G.Murauskas

270

**SPSS**

Dependent Variable Encoding

Original Value	Internal Value
EE Estonia	0
PT Portugal	1

Modelyje  $Y = 1$  atitiks Portugaliją,  
o  $Y = 0$  – Estiją.

© V.Čekanavičius, G.Murauskas 271

**Klasifikacinė lentelė**

SPSS pateikia **dvi** klasifikacines lenteles:  
Pirmoji skirta atsitiktinio spėjimo efektyvumui įvertinti.  
**Antroji** rodo modelio gerumą (būtent ji ir yra svarbiausia).  
Aptarsime antrąją klasifikacinę lentelę.

**SPSS**

Classification Table<sup>a</sup>

Observed	Predicted			Percentage Correct
	Country			
	EE Estonia	PT Portugal		
Step 1 Country	93	36		72.1
Overall Percentage	33	102		75.6
				73.9

a. The cutvalue is 500.

Teisingai klasifikuota 72,1 % estų ir 75,6% portugalų.

© V.Čekanavičius, G.Murauskas 273

**SPSS**

Omnibus Tests of Model Coefficients

Step	Chi-square	df	Sig.
Step 1	99.458	3	.000
Block	99.458	3	.000
Model	99.458	3	.000

Visos trys eilutės lentelėje yra identiškos.  
Gerai, kai  $p < 0,05$ . Kadangi  $p = 0,000\dots$ , tai modelis gerai tinka.

© V.Čekanavičius, G.Murauskas 274

**SPSS**

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke X-Square
1	286.388	.314	.419

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Nagelkerkės determinacijos pseudokoeficientas 0,419 pakankamai didelis ( $>0,20$ ).

© V.Čekanavičius, G.Murauskas 275

**SPSS**

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup> stredu	-.537	.081	43.471	1	.000	.585
happy	.325	.090	13.019	1	.000	1.384
freehms	-.752	.145	27.018	1	.000	.471
Constant	2.498	.831	9.038	1	.000	12.157

a. Variable(s) entered on step 1: stredu, happy, freehms.

**Visi kintamieji statistiškai reikšmingi.**

© V.Čekanavičius, G.Murauskas 276

## Bendra išvada

Logistinės regresijos modelis gerai tinka duomenims.

## Modelio lygtis

Variables in the Equation		B
Step 1 <sup>a</sup>	<b>stfedu</b>	-.537
	<b>happy</b>	.325
	<b>freehms</b>	-.752
	Constant	2.498

a. Variable(s) entered on step 1: stfedu, happy, freehms.

$$\begin{aligned} Z &= \ln P(Y=1)/P(Y=0) \\ &= \ln P(PT)/P(EE) = \\ &= -0,537 \text{ stfedu} \\ &+ 0,325 \text{ happy} \\ &- 0,752 \text{ freehms} \\ &+ 2,498 \end{aligned}$$

© V.Čekanavičius, G.Murauskas

278

## Modelio lygtis

$$Z = \ln P(PT)/P(EE) = -0,537 \text{ stfedu} + 0,325 \text{ happy} - 0,752 \text{ freehms} + 2,498$$

- Didėjant **stfedu**, didėja tikimybė, kad iš **EE**.
- Didėjant **happy**, didėja tikimybė, kad iš **PT**.
- Didėjant **freehms**, didėja tikimybė, kad iš **EE**.

O ką tas didėjimas reiškia, paaiškės iš kintamųjų kodavimo.

© V.Čekanavičius, G.Murauskas

279

## Prognozavimas

Tarkime, kad domina iš kokios valstybės yra respondentas, kuriam **stfedu = 2, happy = 2, freehms = 1**

Įstatome šias reikšmes į

$$\begin{aligned} Z &= \ln P(PT)/P(EE) = -0,537 \text{ stfedu} \\ &+ 0,325 \text{ happy} - 0,752 \text{ freehms} + 2,498 = \\ &= -0,537 * 2 + 0,325 * 2 - 0,752 * 1 + 2,498 = \\ &= 1,322. \end{aligned}$$

© V.Čekanavičius, G.Murauskas

280

## Prognozavimas

Įstatę visas reikšmes gauname:  
 $Z = \ln P(PT)/P(EE) = 1,322$ .

Kadangi  $z > 0$ , tai prognozuojame  $Y=1$  (cntry=PT), t.y., kad respondentas yra iš Portugalijos.

© V.Čekanavičius, G.Murauskas

281

## Prognozavimas

**Z NĖRA** tikimybė. **Z yra** tikimybių santykis.

**Tikimybė**, kad respondentas iš Portugalijos randama taip:

$$P(Y=1) = 2,71^{1,322} / (1 + 2,71^{1,322}) = 0,79$$

© V.Čekanavičius, G.Murauskas

282

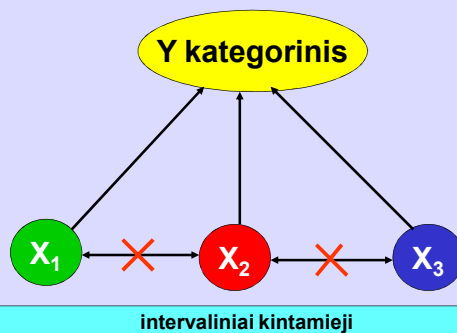
# Diskriminantinė analizė

Discriminant Analysis

Diskriminantinė analizė - toks modelis, kai vieną (priklausomą) kategorinį kintamąjį įtakoja keletas (nepriklausomų) intervalinių kintamųjų.

## Duomenys

- $(X_{11}, X_{21}, X_{31}, \dots, X_{k1}, Y_1), \dots, (X_{1n}, X_{2n}, X_{3n}, \dots, X_{kn}, Y_n)$ .
- tai, ką prognozuosime - priklausomas kintamasis ( $Y$ , mokslinis laipsnis)
- pagal ką - nepriklausomi kintamieji (regresoriai) ( $X$ -ai, arogantiškumas, įkyrumas).



## Kintamųjų diskriminacinės savybės parodo:

- Vilksio lambda - kuo mažesnė, tuo geriau diskriminuoja.
- Vilksio kriterijaus p-reiškė (jei  $p < \alpha$ , tai gerai).
- Reikia abiejų, nes p-reiškė gali būti maža ir dėl duomenų gausos.
- Atskirų kintamųjų tyrimas - pagalbinė analizė (ne pati svarbiausia).

## Kanoninės funkcijos

- Diskriminavimui naudojama ne atskiri  $X$ -ai, o jų tiesiniai dariniai:  
 $f_1(x) = a_1 + b_{11}X_1 + b_{21}X_2 + \dots + b_{k1}X_k,$   
 $f_2(x) = a_2 + b_{12}X_1 + b_{22}X_2 + \dots + b_{k2}X_k, \dots$
- Šios funkcijos sudaromos specialiu būdu ir vadinamos kanoninėmis funkcijomis.
- Jų analizė - svarbiau, nei atskirų  $X$ -ų.

### Kanoninių funkcijų vertinimas:

- Kanoninėmis funkcijomis paaiškinamos kintamųjų dispersijos procentinė sudėtis.
- Vilksio  $\lambda$ .
- Vilksio statistikos  $p$ -reikšmė.

Apsiribosime proc. sudėtimi

### Kanoninių funkcijų % sudėtis:

- Žiūrime kiek % visomis kanoninėmis funkcijomis paaiškinamos dispersijos tenka kiekvienai funkcijai.
- Tai - santykinis rodiklis. Viena funkcija gali būti svarbesnė už kitą, bet abi netikusios.
- Jei yra tik viena kanoninė funkcija, tai rodiklis beprasmiškas, nes =100%.

### Kintamųjų įtaka kanoninėms funkcijoms

- Standartizuotųjų (t.y. z-reikšmių) kanoninių funkcijų koeficientai (kuo moduliu didesni, tuo X-as svarbesnis). Tai  $\beta$  analogai.
- Atskirų X-ų ir kanoninių funkcijų koreliacijos.
- Padeda nustatyti 'įtartinus' X-us.

### Klasifikacinė lentelė

- tai pagrindinis diskriminantinės analizės tikimo indikatorius.
- joje nusprendžiame kurios Y kategorijos priskirtinos esamos imties duomenims.
- kadangi tikrąsias Y reikšmes žinome, galime nustatyti, kiek kartų suklydome.

### Standartinis tyrimas:

- Klasifikacinė lentelė.
- Kanoninių funkcijų svarba.
- Kintamųjų įtaka, kanoninėms funkcijoms.
- Grafikas.
- Vilksio kriterijus 'įtartinėms' X-ams rasti.
- (Prognozavimas).

### Pavyzdys:

- Ar galima atskirti lietuvių nuo latvių ir esto pagal testų
  - Jūra (test1),
  - Sportas (test2),
  - Kaimynai (test3)
- rezultatus?

## Duomenys

salis	test1	test2	test3	var
1	27.00	83.00	61.00	
1	19.00	86.00	57.00	
1	37.00	88.00	62.00	
1	27.00	86.00	61.00	
2	101.00	77.00	69.00	
1	29.00	84.00	65.00	
2	109.00	81.00	75.00	
1	27.00	92.00	66.00	
2	99.00	72.00	71.00	
3	77.00	77.00	69.00	

© V.Čekanaavičius, G.Murauskas

295

## Analyze -> Classify -> Discriminant

- Classify
  - TwoStep Cluster...
  - K-Means Cluster...
  - Hierarchical Cluster...
  - Tree...
  - Discriminant...**
  - Nearest Neighbor...
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response

© V.Čekanaavičius, G.Murauskas

296

## Analyze -> Classify -> Discriminant

Nurodome su Define Range

čia

kintamieji

© V.Čekanaavičius, G.Murauskas

297

## Statistics

varnos

© V.Čekanaavičius, G.Murauskas

298

## Classify -> Discriminant

Tada čia

© V.Čekanaavičius, G.Murauskas

299

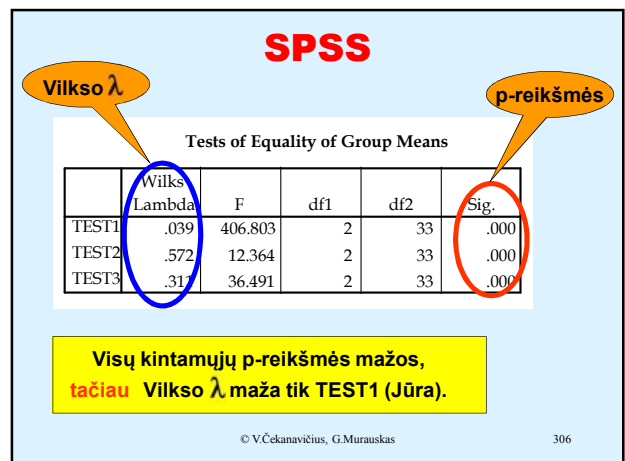
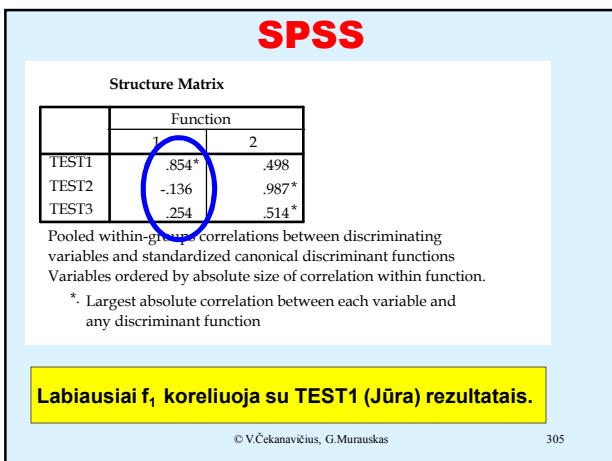
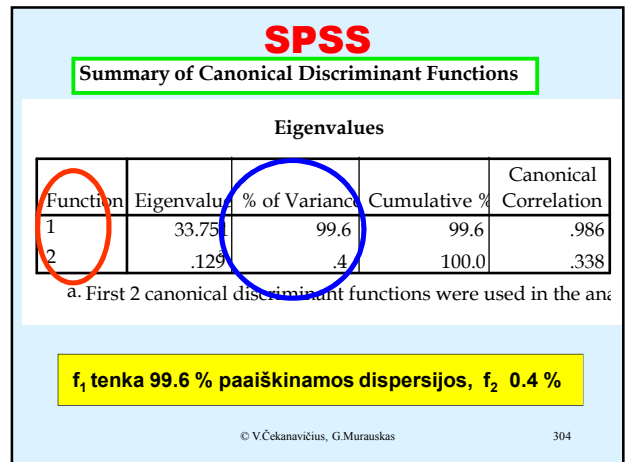
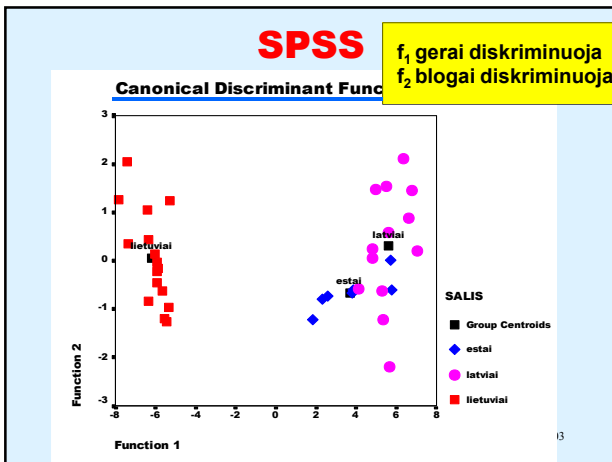
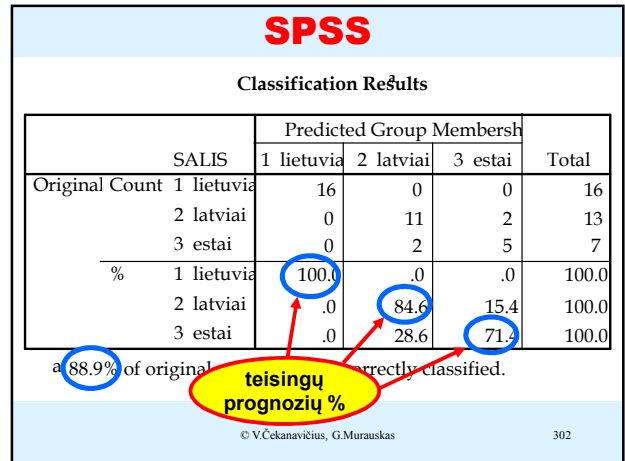
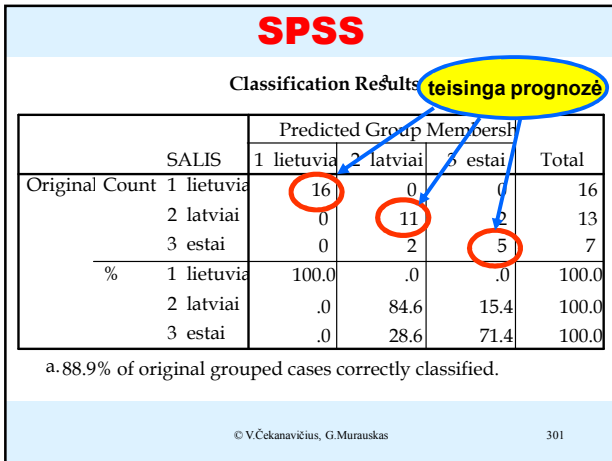
## Classify

varnos

© V.Čekanaavičius, G.Murauskas

300





## SPSS (prognozavimas)

Kintamieji

Classification Function Coefficients

	SALIS		
	1 lietuviai	2 latviai	3 estai
TEST1	-1.234	.461	.163
TEST2	7.881	6.221	6.221
TEST3	1.101	.685	.780
(Constant)	-351.724	-301.126	-278.343

Fisher's linear discriminant functions

lietuviui Fišerio funkcija =  
 $-1.23 \cdot \text{TEST1} + 7.88 \cdot \text{TEST2} + 1.10 \cdot \text{TEST3} - 351.72$

© V.Čekanaavičius, G.Murauskas

307

## SPSS (prognozavimas)

Classification Function Coefficients

	SALIS		
	1 lietuviai	2 latviai	3 estai
TEST1	-1.234	.461	.163
TEST2	7.881	6.221	6.221
TEST3	1.101	.685	.780
(Constant)	-351.724	-301.126	-278.343

Fisher's linear discriminant functions

latviui Fišerio funkcija =  
 $0.46 \cdot \text{TEST1} + 6.22 \cdot \text{TEST2} + 0.68 \cdot \text{TEST3} - 301.12$

© V.Čekanaavičius, G.Murauskas

308

## SPSS (prognozavimas)

Classification Function Coefficients

	SALIS		
	1 lietuviai	2 latviai	3 estai
TEST1	-1.234	.461	.163
TEST2	7.881	6.221	6.221
TEST3	1.101	.685	.780
(Constant)	-351.724	-301.126	-278.343

Fisher's linear discriminant functions

estui Fišerio funkcija =  
 $0.16 \cdot \text{TEST1} + 6.22 \cdot \text{TEST2} + 0.78 \cdot \text{TEST3} - 278.34$

© V.Čekanaavičius, G.Murauskas

309

## Prognozavimas:

- Tarkime TEST1=30, TEST2= 80, TEST3=70.
- Fišerio funkcijos:
- Lietuviui= 318.78.
- Latviui= 257.91.
- Estui= 269.75.
- Prognozuojame, kad respondentas lietuvis.

© V.Čekanaavičius, G.Murauskas

310

## Klasterinė analizė

Cluster Analysis

© V.Čekanaavičius, G.Murauskas

311

**Klasterinė analizė** - objektus pagal panašumą suskirstome į grupes - klasterius.

**Klasteris** - panašių objektų grupė.

© V.Čekanaavičius, G.Murauskas

312

## PAVYZDŽIAI

- pagal testų rezultatus grupuojame tiriamuosius;
- pagal sesijos rezultatus sugrupuojame II kurso psichologus;
- pagal socio-ekonominius parametrus norime suskirstyti valstybes į grupes;
- pagal nuobodumo laipsnį grupuojame pažįstamus.

© V.Čekanaavičius, G.Murauskas

313

## Krepšinio komandų klasterizavimo rezultato pavyzdys

- I klasteris: Žalgiris, L.Rytas
- II klasteris: Šiauliai, Prienai
- III klasteris: Sakalai, Neptūnas

© V.Čekanaavičius, G.Murauskas

314

## Duomenys

- $(X_{11}, X_{21}, X_{31}, \dots, X_{k1}), \dots,$   
 $(X_{1n}, X_{2n}, X_{3n}, \dots, X_{kn})$ .
- yra tik kintamųjų **X-ų** reikšmės!
- kintamieji X- intervaliniai, ranginiai, kategoriniai.

© V.Čekanaavičius, G.Murauskas

315

## Klasterinės analizės metodai

- Hierarchiniai (jungimo arba skaidymo).
- Nehierarchiniai (**K-vidurkių** ir kt.)

Mes mokysimės tik k-vidurkių klasterizaciją.

© V.Čekanaavičius, G.Murauskas

316

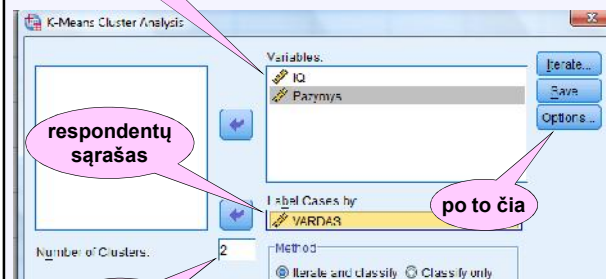
## SPSS



© V.Čekanaavičius, G.Murauskas

317

## SPSS



© V.Čekanaavičius, G.Murauskas

318

**SPSS**

pradiniai centrai

K-Means Cluster Analysis: Opti...

Statistics

- Initial cluster centers
- ANOVA table
- Cluster information for each case

Missing Values

- Exclude cases listwise
- Exclude cases pairwise

Continue Cancel Help

klasterizavimo rezultatai

© V.Čekanaavičius, G.Murauskas 319

**SPSS**

informacija apie pradinius klasterių centrus

Initial Cluster Centers

	Cluster	
	1	2
IQ	97.00	95.00
PAZYMYS	10.00	5.00

© V.Čekanaavičius, G.Murauskas 320

**SPSS**

kokiam klasteriui priklauso

Cluster Membership

Case Number	VARDAS	Cluster	Distance
1	A	1	1.000
2	B	2	1.944
3	C	1	1.000
4	D	2	1.054
5	E	1	1.333

atstumas iki klasterio centro

© V.Čekanaavičius, G.Murauskas 321

**SPSS**

informacija apie galutinius klasterių centrus

Final Cluster Centers

	Cluster	
	1	2
IQ	97.00	95.00
PAZYMYS	9.00	6.33

© V.Čekanaavičius, G.Murauskas 322

**SPSS**

Number of Cases in each Cluster

Cluster	1	2.000
	2	3.000
Valid		5.000
Missing		.000

© V.Čekanaavičius, G.Murauskas 323

**Faktorinė analizė**

**FACTOR ANALYSIS**

© V. Čekanaavičius, G. Murauskas 324

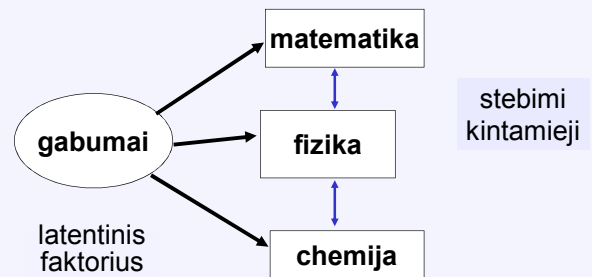
**Faktorinė analizė** - toks modelis, kai vienas tiesiogiai nestebimas kintamasis (**latentinis faktorius**) įtakoja keletą stebimų intervalinių kintamųjų.

Dažniausiai būna **keli** faktoriai.  
Nemaišyti su **dispersine** analize ANOVA

© V. Čekanavičius, G. Murauskas

325

## Faktorinė analizė



© V. Čekanavičius, G. Murauskas

326

- **Tikslas** - išskirti kuo geriau kintamųjų elgesį aprašančius faktorius.
- Naudojant faktorių reikšmes sumažėja tiriamų kintamųjų.
- Taip pat sumažėja ir turimos informacijos.

© V. Čekanavičius, G. Murauskas

327

- Latentinio faktoriaus **pavadinimą sugalvoja tyrėjas**.
- Faktorių **skaičių** parenka tyrėjas.
- Faktorinė analizė tinka **ne visada**.
- Skirtingos faktorinės analizės išskiria skirtingus faktorius.
- Ne visada lengva interpretuoti faktorius.

© V. Čekanavičius, G. Murauskas

328

## PAVYZDŽIAI

- tiriame, kas gali vienyti politikus (gal pažiūrų kairumas, gal autoritetas)
- ką atpindi pažymiai (gabumus apskritai, ar gabumus humanitariniams bei tiksliesiems mokslams;
- kas vienija respondentų požiūrį į skirtingus muzikinius stilius;
- kokios klausimų grupės išsiskiria klausimyne.

© V. Čekanavičius, G. Murauskas

329

## PASTABA

- faktorinė analizė geriau tinka vieno reiškinio skirtingų aspektų analizei; nei
- kelių visiškai skirtingų reiškinų grupavimui;
- jei turime testą skirtą gabumams ir fiziniams parametrams matuoti-
- geriau taikyti dvi faktorines analizes, kiekvienam pustesčiui.

© V. Čekanavičius, G. Murauskas

330

### Faktorinės analizės prielaidos

- visi X-ai normalūs;
- visi F-ai nekoreliuoti ir turi vienetines dispersijas;
- paklaidos  $e_j$  nekoreliuotos;
- F-ai ir paklaidos nekoreliuoja.

© V. Čekanaavičius, G. Murauskas

331

### Faktorinės analizės etapai

- Patikrinimas ar tinka duomenys.
- Faktorių išskyrimas.
- Faktorių sukimas ir interpretavimas.
- Faktorių reikšmių skaičiavimas.

© V. Čekanaavičius, G. Murauskas

332

### Duomenų tikimas

- Kintamieji X-ai turi koreliuoti. Koreliuotumą vertina:
- Bartleto kriterijus (gerai, kai  $p < 0.05$ ).
- Kaizerio-Mejerio-Olkinio matas. Blogai, kai  $KMO < 0.50$  (mažiau už vieną antrąją).

© V. Čekanaavičius, G. Murauskas

333

### Kiek faktorių imti?

- Dažniausiai žiūrima, kiek kintamųjų elgesio dispersijos paaiškina faktoriai.
- Nėra griežtos taisyklės, kiek kintamųjų dispersijos turi paaiškinti faktoriai.
- Interpretacija - daugiau dispersijos paaiškina, daugiau ir informacijos apie kintamųjų elgesį išlaikome, pereidami prie faktorių.

© V. Čekanaavičius, G. Murauskas

334

### Faktorių nustatymas

- Faktorių paaiškina tuos kintamuosius, kuriems jo svoris didelis.
- Vertinimas 'iš akies': svoris didelis, jei absoliutiniu didumu  $> 0.4$ .
- Gali tą patį kintamąjį paaiškinti keli faktoriai.
- Tada kintamąjį priskiriame tam, kuris geriau tinka pagal prasmę.

© V. Čekanaavičius, G. Murauskas

335

### Faktorių interpretavimas

- sugalvoti faktoriui pavadinimą - tyrėjo reikalas.
- tai priklauso nuo tyrėjo išsilavinimo.
- tą patį faktorių vienas įvardins, kaip percepcinę organizaciją, o kitas, kaip objekto-visumos suvokimą.
- teisybės tas, kuris sugeba įtikinti kitus.

© V. Čekanaavičius, G. Murauskas

336

### Pastaba

- pagrindinių komponentų analizė reiškia, kad pirmas faktorius svarbiausias, po to antras ir t.t.
- dominuojantis pirmasis faktorius liudija, kad matuojame vieno reiškinio skirtingus aspektus (tai gerai).

© V. Čekanavičius, G. Murauskas

337

### Faktorių sukimas

- tačiau dominuojantis pirmasis faktorius labai apsunkina faktorių interpretaciją.
- sunku interpretuoti ir kai tas pats X pakliūna į keletą faktorių.
- išeitis faktorių sukimas.
- nauji faktoriai - senųjų tiesinės daugdaros.

© V. Čekanavičius, G. Murauskas

338

### Faktorių sukimas

- po sukimo faktorių interpretacija gali keistis.
- yra visokių sukimų.
- populiariausias **VARIMAX**
- gali netikti nė vienas sukimas.

© V. Čekanavičius, G. Murauskas

339

### Faktorių reikšmės

- nėra vieningos skaičiavimo metodikos.
- naudojant SPSS siūlomas technikas rezultatai normuoti;
- dažnai faktoriaus reikšmė laikomas atitinkamų kintamųjų reikšmių vidurkis.

© V. Čekanavičius, G. Murauskas

340

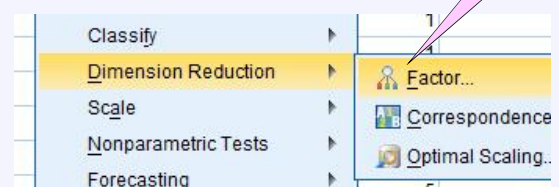
### Pavyzdys

- failas B1994, politikų vertinimai.
- kokios politikų grupės išsiskiria.
- kas keičiasi po sukimo.

© V. Čekanavičius, G. Murauskas

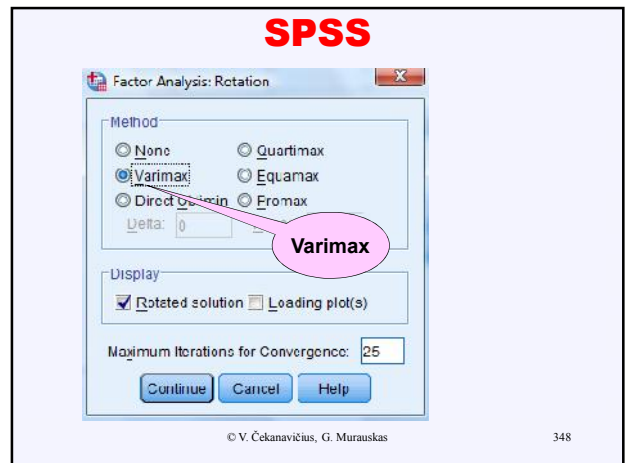
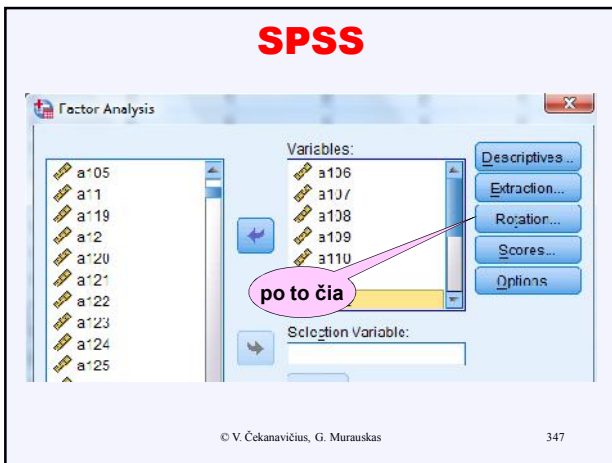
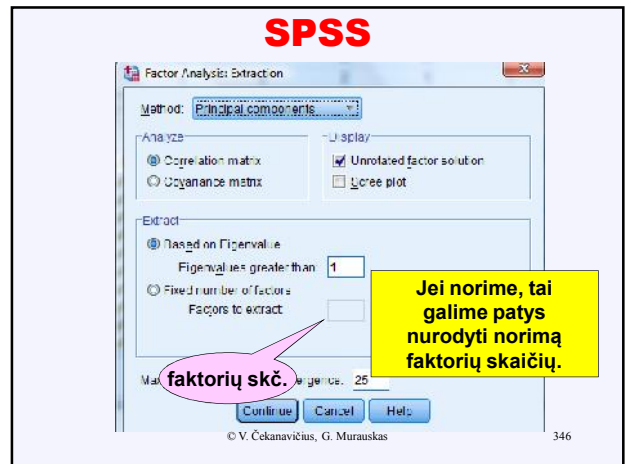
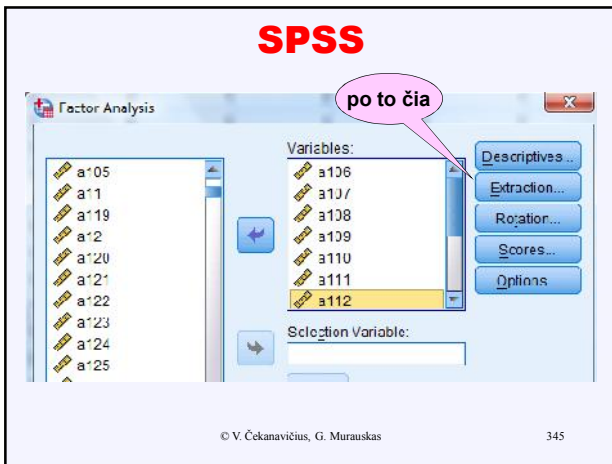
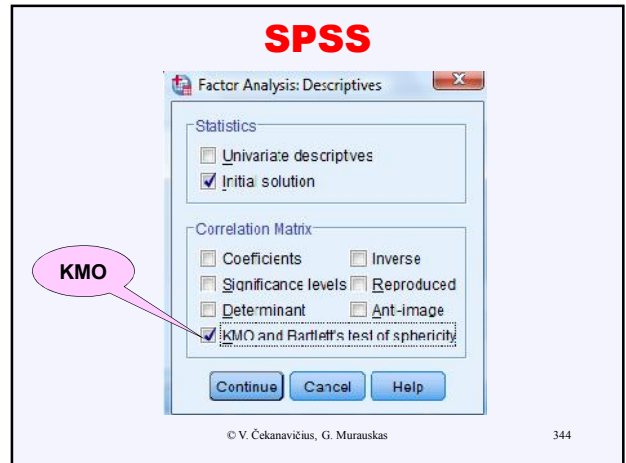
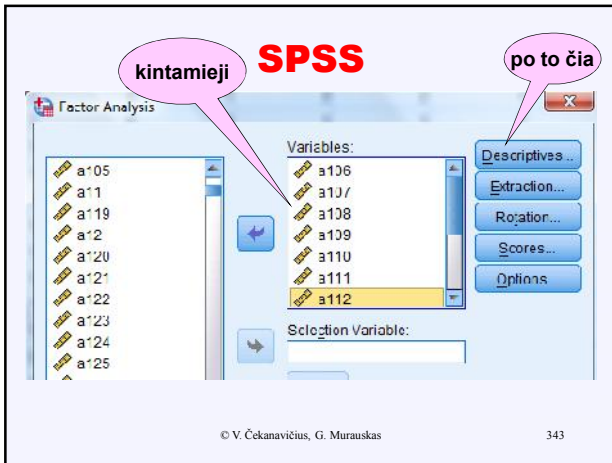
341

### SPSS



© V. Čekanavičius, G. Murauskas

342





### SPSS

© V. Čekanaavičius, G. Murauskas 349

### SPSS

© V. Čekanaavičius, G. Murauskas 350

### SPSS

#### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		
		.851
Bartlett's Test of Sphericity		
Approx. Chi-Square	4360.774	
df	78	
Sig.		.000

Duomenims FA labai gerai tinka.

© V. Čekanaavičius, G. Murauskas 351

### SPSS

#### trys faktoriai

#### Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative	Total	% of Variance	Cumulative	Total	% of Variance	Cumulative
1	4.366	33.585	33.585	4.366	33.585	33.585	2.941	22.626	22.626
2	1.664	12.800	46.384	1.664	12.800	46.384	2.327	17.897	40.523
3	1.214	9.335	55.719	1.214	9.335	55.719	1.976	15.196	55.719
4	.914	7.033	62.752						
5	.763	5.867	68.619						
6	.672	5.16	73.784						
7	.605	4.657	78.442						
8	.563	4.328	82.770						
9	.517	3.974	86.744						
10	.465	3.576	90.320						
11	.450	3.460	93.780						
12	.438	3.366	97.146						
13	.371	2.854	100.000						

Extraction Method: Principal Component Analysis.

© V. Čekanaavičius, G. Murauskas 352

### SPSS

#### faktorių svorių matrica prieš sukimą

#### Component Matrix<sup>a</sup>

	Component		
	1	2	3
a106 KANTANAVICIUS	.510		
a107 EBICKAUSKAS	.553		
a108 ABRAZAUSKAS	.427	-.001	.427
a109 P.GVILYS	.674		
a110 CILURAFNAF	.610		
a111 GILINKILAS	.610		
a112 VLANDSBERGIE		.705	
a113 RUCOLAV	.517	.420	
a114 JPALECHIE	.625		
a115 A SAKA AF	.627		
a116 A SAKARAS	.653		-.162
a117 A SLEZEVICIUS	.620		
a118 G.VAGNORIUS	.417	.648	

Extraction Method: Principal Component Analysis.  
a. 3 components extracted.

© V. Čekanaavičius, G. Murauskas 353

### SPSS

#### faktorių svorių matrica po sukimo

#### Rotated Component Matrix<sup>a</sup>

	Component		
	1	2	3
a106 KANTANAVICIUS			
a107 EBICKAUSKAS		.476	
a108 ABRAZAUSKAS		.733	
a109 P.GVILYS	.660		
a110 CILURAFNAF		.718	
a111 GILINKILAS	.726		
a112 VLANDSBERGIE			.627
a113 RUCOLAV			.610
a114 JPALECHIE	.627		
a115 A SAKA AF	.747		
a116 A SAKARAS	.781		
a117 A SLEZEVICIUS		.732	
a118 G.VAGNORIUS			.802

Extraction Method: Principal Component Analysis.  
Rotation Method: Varimax with Kaiser Normalization.  
a. Rotation converged in 5 iterations.

© V. Čekanaavičius, G. Murauskas 354

## HLM

### Tiesiniai hierarchiniai modeliai

## KĄ DARO?

- Regresijos išplėtojimas, kai nurodome koks intervalinis nuo kokių kitų intervalinių kintamųjų priklauso.
- Regresijos lygties daugikliai yra naujų kintamųjų priklausomi kintamieji.

V.Čekanašius, G. Murauskas

2

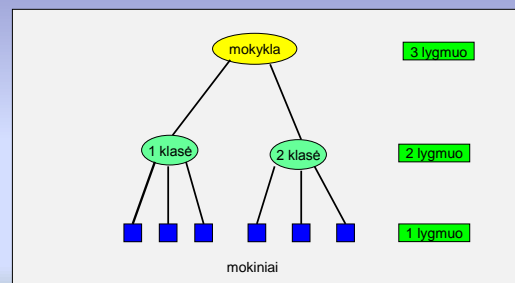
## PAVYZDYS

- Norime išsiaiškinti, kaip matematikos testo rezultatus veikia socialinė-ekonominė aplinka ir klasė.
- Duomenys turi hierarchinę struktūrą, nes klasės charakteristikos daro įtaką visiems tos klasės mokiniams.

V.Čekanašius, G. Murauskas

3

## Idėja:



V.Čekanašius, G. Murauskas

4

## PASTABOS:

- Ir pirmojo ir antrojo lygmens stebėjimų turi būti nemažai – negalima hierarchinio modelio taikyti tik porai mokyklų.
- Matuojami kintamieji – normalūs arba kategoriniai.
- Kategorinių kintamųjų neturėtų būti labai daug.

V.Čekanašius, G. Murauskas

5

## Tipinė HLM tyrimo eiga

- Pradžioje visada sudaromas besąlyginis (nulinis) modelis.
- Sudaromas sudėtingesnis modelis.
- Modeliai lyginami, aiškinamasi kuris geresnis.
- Visi modelių lyginimai prasmingi tik tada, kai jie sudaryti tiems patiems duomenims.

V.Čekanašius, G. Murauskas

6

## Besąlyginis HLM modelis

- Besąlyginis (nulinis) modelis svarbus tuo, kad kiti hierarchiniai modeliai su juo lyginami.
- Besąlyginio modelio analizė padeda atsakyti į klausimą ar apskritai galima kalbėti apie hierarchinių modelių tinkamumą.

V.Čekanašius, G.Murauskas

7

## Besąlyginis HLM modelis

- HLM iliustracijai pasirinktas duomenų poaibis – kaimo ir Vilniaus miesto mokyklų 8 klasės moksleivių tyrimo rezultatai.
- Turime 559 mokinio iš 27 mokyklų matematikos testo rezultatus ir įvairius mokyklų bei mokinį apibūdinančius duomenis.
- Besąlyginį HLM modelį sudaro lygtys abiem – mokinio ir mokyklos – lygmenims.

V.Čekanašius, G.Murauskas

8

## Besąlyginis HLM modelis

- Vadovausimės tokia logika:
- mokinio rezultatų skirtumas nuo visos mokyklos vidurkio atspindi tai, kad būdinga tik tam mokiniui – jo gabumus, namų aplinką, repetitoriaus patirtį ir pan.

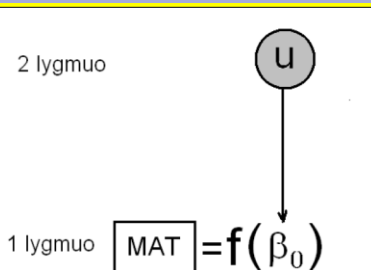
Mokinio rezultatas = mokyklos vidurkis + individualūs skirtumai nuo vidurkio.

- Analogiškai aprašome mokyklos lygmenį:
- Mokyklos rezultatas = visų mokyklų vidurkis + šios mokyklos įtaka.

V.Čekanašius, G.Murauskas

9

## MODELIO IDĖJA



V.Čekanašius, G.Murauskas

10

## Besąlyginis HLM modelis

- Mokinio lygmuo:

$$MAT = \beta_0 + e$$

- Mokyklos lygmuo:

$$\beta_0 = \gamma_{00} + u_0$$

- Jungtinis modelis:

$$MAT = \gamma_{00} + u_0 + e$$

V.Čekanašius, G.Murauskas

11

## Besąlyginis HLM modelis

$$MAT = \gamma_{00} + u_0 + e$$

- Parametras  $\gamma_{00}$  nusako visų mokyklų matematikos testo rezultatų vidurkį.
- Mokyklų skirtumus rodo kintamojo  $u_0$  dispersija  $\tau_{00}$ .
- Mokinių individualius skirtumus – kintamojo  $e$  dispersija  $\sigma^2$ .
- Kuo dispersijos didesnės, tuo didesni ir atitinkami skirtumai.

V.Čekanašius, G.Murauskas

12

## Besąlyginis HLM modelis

$\gamma_{00}$  - fiksuoto poveikio parametras,

$\tau_{00}$  ir  $\sigma^2$  - atsitiktinio poveikio parametrai.

V.Čekanauskas, G.Murauskas

13

## Besąlyginio modelio tyrimas su SPSS

- Duomenys turi tradicinę struktūrą – vienam mokiniui skiriama viena eilutė.
- Be to, yra kintamasis **IDMOK**, kuriame nurodytas mokyklos kodas.

V.Čekanauskas, G.Murauskas

14

## Besąlyginio modelio tyrimas su SPSS

duomenys

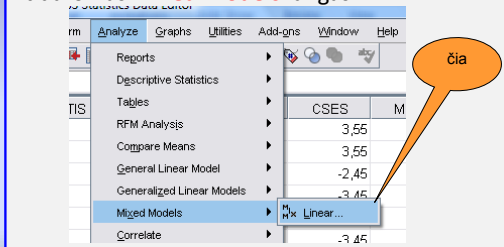
	IDKLA	LYTIS	VK	IDMOK	CSES	MSES	MAT
1	4601	2	0,00	1	3,55	20,45	553,11
2	4601	1	0,00	1	3,55	20,45	536,88
3	4601	2	0,00	1	-2,45	20,45	499,71
4	4601	1	0,00	1	-3,45	20,45	510,02
5	4601	2	0,00	1	1,55	20,45	529,04
6	4601	1	0,00	1	-3,45	20,45	418,79
7	4601	2	0,00	1	0,55	20,45	520,45
8	4601	1	0,00	1	-2,45	20,45	596,07
9	4601	2	0,00	1	-1,45	20,45	702,19
10	4601	1	0,00	1	4,05	20,45	645,75

V.Čekanauskas, G.Murauskas

15

## Besąlyginio modelio tyrimas su SPSS

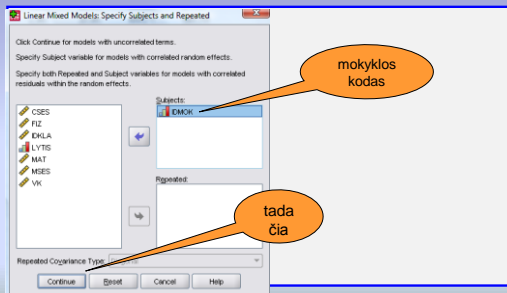
Atidaromas **Mixed Models** langas.



V.Čekanauskas, G.Murauskas

16

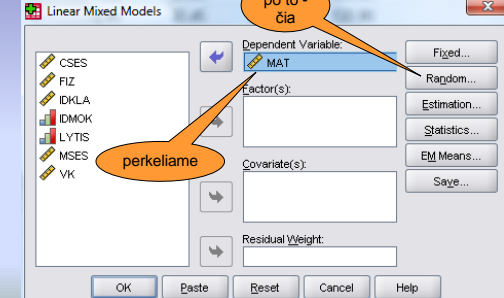
## Besąlyginio modelio tyrimas su SPSS



V.Čekanauskas, G.Murauskas

17

## Besąlyginio modelio tyrimas su SPSS



V.Čekanauskas, G.Murauskas

18

### Besąlyginio modelio tyrimas su SPSS

keičiame į Unstructured

pažymime

perkeliamo

V.Čekanauskas, G.Murauskas

### Besąlyginio modelio tyrimas su SPSS

Tada, čia

V.Čekanauskas, G.Murauskas

### Besąlyginio modelio tyrimas su SPSS

varnos

V.Čekanauskas, G.Murauskas

### Besąlyginio modelio tyrimas su SPSS

OK

V.Čekanauskas, G.Murauskas

### Besąlyginio modelio tyrimas su SPSS

- SPSS programa pateikia informacinių
- **AIC** (Akaikės),
- **AICC** (Hurvičiaus ir Tsai),
- **CAIS** (Bozdogano),
- **BIC** (Švarco – Bajeso) kriterijų reikšmes.
- Į šias reikšmes atsižvelgiama, kai lyginami du modeliai.
- Modelis, turintis mažesnes informacinių kriterijų reikšmes, geriau suderintas su duomenimis.

V.Čekanauskas, G.Murauskas

### Besąlyginio modelio tyrimas su SPSS

- Informacinių kriterijų reikšmės. Reikės vėliau – lyginant du modelius.
- mažesnė reikšmė – geriau.

Information Criteria <sup>a</sup>	
-2 Restricted Log Likelihood	6415,181
Akaike's Information Criterion (AIC)	6419,181
Hurvich and Tsai's Criterion (AICC)	6419,202
Bozdogan's Criterion (CAIC)	6429,829
Schwarz's Bayesian Criterion (BIC)	6427,829

The information criteria are displayed in smaller-is-better forms.  
a. Dependent Variable: MAT.

V.Čekanauskas, G.Murauskas

## Besąlyginio modelio tyrimas su SPSS

Pateikiamas parametro  $\gamma_{00}$  įvertis  $\hat{\gamma}_{00} = 499,47$

Be to patikrinama hipotezė

Kadangi  $p < 0,05$ , tai

stat.reikšmingai  $\gamma_{00} \neq 0$ .

$$\begin{cases} H_0 : \gamma_{00} = 0 \\ H_1 : \gamma_{00} \neq 0 \end{cases}$$

Estimates of Fixed Effects<sup>a</sup>

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	499,471114	9,876646	25,493	50,671	,000	479,150324	519,793104

a. Dependent Variable: MAT.

V.Čekanašius, G.Murauskas

25

## Besąlyginio modelio tyrimas su SPSS

• Primename, kad  $\gamma_{00}$  - tai bendras visų mokyklų matematikos testo MAT vidurkis.

• Statistinės hipotezės išvada, kad jis nelygus nuliui nėra itin vertinga (niekas ir nesitikėjo, kad visi bus beraščiai).

• Įvertis  $\hat{\gamma}_{00} = 499,47$  -- daug svarbesnis.

V.Čekanašius, G.Murauskas

26

## Besąlyginio modelio tyrimas su SPSS

Tikrinamos dvi hipotezės apie I ir II lygmenų paklaidų dispersijas:

$$\begin{cases} H_0 : \tau_{00} = 0 \\ H_1 : \tau_{00} > 0 \end{cases}$$

$$\begin{cases} H_0 : \sigma^2 = 0 \\ H_1 : \sigma^2 > 0 \end{cases}$$

Taip pat pateikiami šių **atsitiktinių** parametų įverčiai.

V.Čekanašius, G.Murauskas

27

## Besąlyginio modelio tyrimas su SPSS

• Abi nulinės hipotezės atmetamos, darome išvadą, kad  $\tau_{00} > 0$  ir  $\sigma^2 > 0$

• Gavome patvirtinimą, kad modelis nepilnai paaiškina rezultatų skirtumus.

Estimates of Covariance Parameters<sup>a</sup>

Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		5117,191389	313,883040	16,303	,000
Intercept[subject = IDMOK]	Variance	2356,106890	736,855937	3,198	,001

a. Dependent Variable: MAT.

V.Čekanašius, G.Murauskas

28

## Besąlyginio modelio tyrimas su SPSS

• Gauti dispersijų įverčiai yra dideli:

•  $\hat{\tau}_{00} = 2356,11$  ir  $\hat{\sigma}^2 = 5117,19$ .

Taigi, modelis duomenims tinka blogai.

Estimates of Covariance Parameters<sup>a</sup>

Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		5117,191389	313,883040	16,303	,000
Intercept[subject = IDMOK]	Variance	2356,106890	736,855937	3,198	,001

a. Dependent Variable: MAT.

V.Čekanašius, G.Murauskas

29

## Besąlyginio modelio tyrimas su SPSS

• Nagrinėjant besąlyginį modelį patartina suskaičiuoti tarpklausinės koreliacijos koeficientą **ICC** (Intraclass correlation coefficient), kuris parodo, kaip stipriai skiriasi grupių (mokyklų) rezultatai, lyginant su rezultatais grupėse.

$$ICC = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2}$$

• Kuo **ICC** didesnis, tuo labiau įtikėtina, kad reikia atsižvelgti į hierarchinę duomenų struktūrą.

V.Čekanašius, G.Murauskas

30

## Besąlyginio modelio tyrimas su SPSS

- Pavyzdyje

$$ICC = \frac{2356,11}{2356,11 + 5117,19} = 0,315...$$

- Tai nėra itin didelė reikšmė. Ji interpretuojama taip: apie 31,5 procento mokinio rezultatų priklauso nuo mokyklos.

V.Čekanauskas, G.Murauskas

31

## Besąlyginio modelio tyrimas su SPSS

- Ir ICC ir atsitiktinių parametų įverčiai rodo, kad reikia sudaryti sudėtingesnį modelį.
- Įtraukdami naujus I ir II lygmenų kintamuosius, siekiame sumažinti įverčių

$$\hat{\tau}_{00} = 2356,11 \quad \hat{\sigma}^2 = 5117,19$$

didumus (jie parodo tuos mokinių ir mokyklų skirtumus, kurių nepaaiškina modelis).

V.Čekanauskas, G.Murauskas

32

## HLM modelis su antrojo lygmens intervaliniu kintamuoju

Manome, kad :

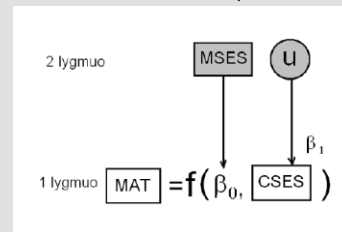
- mokinio matematikos testo rezultatas priklauso nuo jo socialinio ir ekonominio statuso (CSES);
- mokyklos įtaka visiems mokiniams vienoda;
- mokyklos įtakos stiprumą lemia jos mokinių socialinio ir ekonominio statuso vidurkis (MSES).

V.Čekanauskas, G.Murauskas

33

## HLM modelis su antrojo lygmens intervaliniu kintamuoju

- Grafiškai modelis atrodo taip:



V.Čekanauskas, G.Murauskas

34

## HLM modelis su antrojo lygmens intervaliniu kintamuoju

Reikalavimai kintamiesiems:

- Priklausomas kintamasis (MAT) būtina intervalinis.
- Kiti kintamieji intervaliniai arba kategoriniai (pseudokintamieji).
- Šiame pvz. visi kiti kintamieji CSES, MSES yra intervaliniai.
- Paklaidos (liekanos) – normaliai pasiskirsčiusios ir nepriklausomos.

V.Čekanauskas, G.Murauskas

35

## HLM modelis su antrojo lygmens intervaliniu kintamuoju

- Pasižiūrime, kurio lygmens kintamųjų turime:
- pirmojo (mokinio) lygmens kintamieji: CSES.
- antrojo (mokyklos) lygmens kintamieji: MSES.

V.Čekanauskas, G.Murauskas

36

## HLM modelis su antrojo lygmens intervaliniu kintamuoju

- Mokinio lygmens lygtis gana akivaizdi:

$$MAT = \beta_0 + \beta_1 CSES + e.$$

- Čia  $e$  žymi atsitiktinę paklaidą, kuri yra normaliai pasiskirsčiusi su nuliniu vidurkiu ir nežinoma dispersija  $\sigma^2$ . Ši dispersija atspindi skirtumus tarp mokinių, atsiradusius ne dėl mokyklų įtakos.

V.Čekanauskas, G.Murauskas

37

## HLM modelis su antrojo lygmens intervaliniu kintamuoju

- Antrojo lygmens kintamieji į modelį įtraukiami, tariant, kad atitinkami  $\beta$  koeficientai yra atsitiktiniai.
- Laikomasi tokio principo:
- Kintamieji, kurie **veikia visus mokinius vienodai**, įtraukiami į lygtį, aprašančią laisvą narį  $\beta_0$ .
- Kintamieji, kurie **turi sąveiką su CSES**, įtraukiami į lygtį, aprašančią koeficientą  $\beta_1$ .

V.Čekanauskas, G.Murauskas

38

## HLM modelis su antrojo lygmens intervaliniu kintamuoju

- Taigi, antrojo lygmens lygtys yra tokios:

$$\beta_0 = \gamma_{00} + \gamma_{01} MSES + u_0,$$

$$\beta_1 = \gamma_{10} + u_1.$$

- Čia  $\gamma_{ij}$  nežinomos konstantos.
- Paklaidos  $u_0, u_1$  turi dispersijas  $\tau_{00}, \tau_{11}$  ir gali koreliuoti (kovariacija  $\tau_{01}$ ).

V.Čekanauskas, G.Murauskas

39

## HLM modelis su antrojo lygmens intervaliniu kintamuoju

- Primerkime, kad dispersijos  $\tau_{00}, \tau_{11}$  atspindi rezultatų skirtumus, atsirandančius dėl mokyklų įtakos.
- Tačiau tik tuos skirtumus, kurių nepaaiškina pasirinktas modelis.
- Kuo šios dispersijos mažesnės, tuo modelis geriau suderintas su duomenimis.

V.Čekanauskas, G.Murauskas

40

## HLM modelis su antrojo lygmens intervaliniu kintamuoju

- Galima sudaryti jungtinę lygtį:

$$MAT = \left[ \begin{array}{l} \gamma_{00} + \gamma_{01} MSES + \gamma_{10} CSES \\ + [e + u_0 + u_1 CSES] \end{array} \right]$$

- Matome, kad modelyje išsiskiria fiksuotų poveikių komponentas ir
- atsitiktinių poveikių komponentas.

V.Čekanauskas, G.Murauskas

41

## HLM modelis su antrojo lygmens intervaliniu kintamuoju

- Fiksuoto (pastovaus) poveikio parametrai  $\gamma_{00}, \gamma_{01}, \gamma_{10}$ .
- atsitiktinio poveikio parametrai (liekanų  $e, u_0, u_1$  dispersijos ir kovariacijos):

$$\tau_{00}, \tau_{11}, \tau_{01}, \sigma^2.$$

V.Čekanauskas, G.Murauskas

42



## HLM modelis su antrojo lygmens intervaliniu kintamuoju

- Kintamieji prie  $\gamma$  vadinami fiksuoto poveikio kintamaisiais, o kintamieji prie  $u$  – atsitiktinio poveikio kintamaisiais.
- Fiksuoto poveikio kintamieji: **CSES** ir **MSES**.
- Atsitiktinio – kintamasis **CSES**.

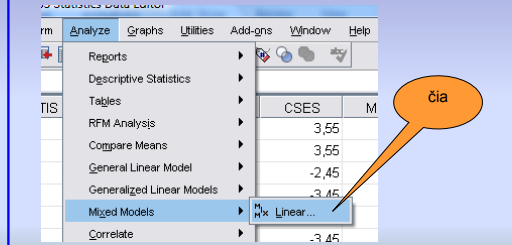
$$MAT = [\gamma_{00} + \gamma_{01}MSES + \gamma_{10}CSES] + [e + u_0 + u_1CSES]$$

V.Čekanašius, G.Murauskas

43

## Modelio tyrimas su SPSS

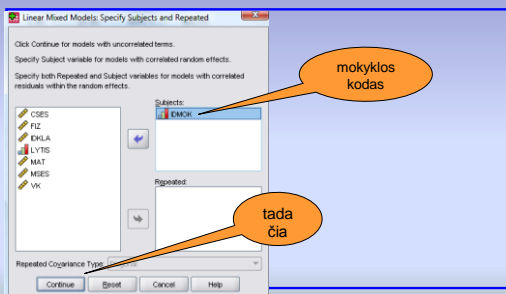
Atidaromas **Mixed Models** langas.



V.Čekanašius, G.Murauskas

44

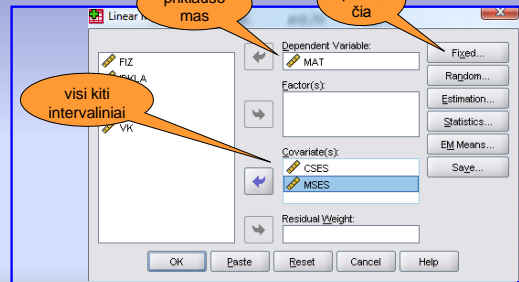
## Modelio tyrimas su SPSS



V.Čekanašius, G.Murauskas

45

## Modelio tyrimas su SPSS



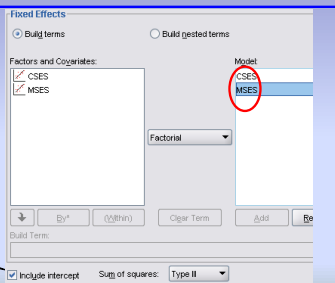
V.Čekanašius, G.Murauskas

46

## Modelio tyrimas su SPSS

Perkeliam visus fiksuoto poveikio kintamuosius

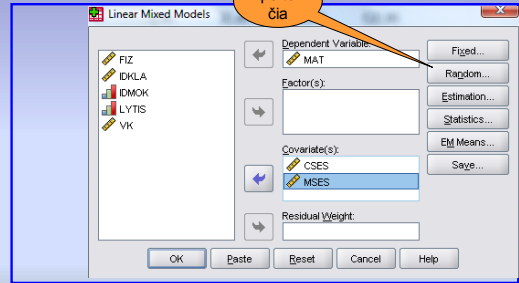
vama



V.Čekanašius, G.Murauskas

47

## Modelio tyrimas su SPSS



V.Čekanašius, G.Murauskas

48

## Modelio tyrimas su SPSS

Keičiam į Unstructured

varna

keliam atsitiktinio poveikio kintamąjį

perkeliam

V.Čekanaavičius, G.Murauskas 49

## Modelio tyrimas su SPSS

po to čia

V.Čekanaavičius, G.Murauskas 50

## Modelio tyrimas su SPSS

varnos

V.Čekanaavičius, G.Murauskas 51

## Modelio tyrimas su SPSS

OK

V.Čekanaavičius, G.Murauskas 52

## Modelio tyrimas su SPSS

Planas:

- Lyginsime AIC ir kitus informacinius indeksus su besąlyginu modeliu.
- Tikrinsime, kurie fiksuoti parametrai statistiškai nereikšmingi.
- Tikrinsime, kurie atsitiktiniai parametrai statistiškai nereikšmingi.
- Tikrinsime, kiek sumažėjo atsitiktinių parametru dispersijų įverčiai.

V.Čekanaavičius, G.Murauskas 53

## Modelio tyrimas su SPSS

• Akaičės ir kitų indeksų sumažėjimas bent per 10 vienetų iau laikomasis svarbiu.

Information Criteria <sup>a</sup>		Information Criteria <sup>b</sup>	
-2 Restricted Log Likelihood	6109,830	-2 Restricted Log Likelihood	6415,181
Akaike's Information Criterion (AIC)	6117,830	Akaike's Information Criterion (AIC)	6419,181
Hurvich and Tsai's Criterion (AICC)	6117,904	Hurvich and Tsai's Criterion (AICC)	6419,202
Bozdogan's Criterion (CAIC)	6138,996	Bozdogan's Criterion (CAIC)	6429,829
Schwarz's Bayesian Criterion (BIC)	6134,996	Schwarz's Bayesian Criterion (BIC)	6427,829

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: MAT.

b. a. D

Nujasis modelis      Besąlyginis

V.Čekanaavičius, G.Murauskas 54

## Modelio tyrimas su SPSS

Visi pastovaus poveikio kintamieji statistiškai reikšmingi.

$$\hat{\gamma}_{01} = 16,56 \quad \hat{\gamma}_{00} = 175,94 \quad \hat{\gamma}_{10} = 8,04$$

Estimates of Fixed Effects<sup>a</sup>

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	175,943870	42,631950	24,688	4,127	,000
CSES	8,044992	1,227057	21,239	6,556	,000
MSES	16,564816	2,135069	24,060	7,757	,000

a. Dependent Variable: MAT.

V.Čekanaivičius, G.Murauskas

55

## Modelio tyrimas su SPSS

- Gauti atsitiktinių parametų įverčiai:

$$\hat{\tau}_{00} = 559,48 \quad \hat{\tau}_{01} = 0,371 \quad \hat{\tau}_{11} = 9,19$$

$$\hat{\sigma}^2 = 4290,92$$

Estimates of Covariance Parameters<sup>a</sup>

Parameter	Estimate	Std. Error	Wald Z	Sig.	
Residual	4290,921188	274,026086	15,659	,000	
Intercept + CSES	UN (1,1)	559,476865	227,054045	2,464	,014
[subject = IDMOK]	UN (2,1)	,371530	37,426627	,010	,992
	UN (2,2)	9,191158	11,104040	,828	,408

V.Čekanaivičius, G.Murauskas

56

## Modelio tyrimas su SPSS

- Matome, kad  $\beta_0$  ir  $\beta_1$  statistiškai reikšmingai nekoreliuoja, nes atitinkama p reikšmė lygi 0,99; o ir pats kovariacijos įvertis mažas:  $\hat{\tau}_{01} = 0,371$ .
- Taip pat matome, kad parinktoji antrojo lygmens lygtis parametą  $\beta_1$  aprašo pakankamai išsamiai (atitinkama p reikšmė 0,408). To negalima pasakyti apie  $\beta_0$ .

V.Čekanaivičius, G.Murauskas

57

## Modelio tyrimas su SPSS

Įsitikiname, kad respondentų skirtumus atspindinti dispersija nėra itin maža:

$$\hat{\sigma}^2 = 4290,92$$

Vis dėlto, lyginant su besąlyginio modelio analogiška dispersija ( $\sigma^2 = 5117,19$ ), aišku, kad dispersija ženkliai sumažėjo.

Galima iširti, kuri dalis dispersijos dingo.

V.Čekanaivičius, G.Murauskas

58

## Modelio tyrimas su SPSS

Formulė:

$$\frac{\hat{\sigma}_{sena}^2 - \hat{\sigma}_{nauja}^2}{\hat{\sigma}_{sena}^2}$$

Gauname: 
$$\frac{5117,19 - 4290,92}{5117,19} = 0,161\dots$$

Interpretuojame taip: lyginant su besąlyginio modeliu, 16% sumažėjo modelių nepaaiškintų mokinių rezultatų skirtumų. Gana nemenkas išlošimas.

V.Čekanaivičius, G.Murauskas

59

## Modelio tyrimas su SPSS

Reziumuodami, galime pasakyti, kad nagrinėjamas modelis daug geriau tinka duomenims nei besąlyginis modelis.

Testo rezultatai matuojami šimtais baly. Todėl rezultatų paklaidos standartinis nuokrypis nėra labai didelis  $\hat{\sigma} = 65,49\dots$

Tai įgalina modelį laikyti pakankamai geru. Žinoma, tai nereiškia, kad nėra geriau duomenims tinkančių modelių.

V.Čekanaivičius, G.Murauskas

60

## Modelio tyrimas su SPSS

Jei norime daryti prognozes, tai tiesiog įstatome fiksuotų parametru įvertčius į jungtinę lygtį:

$$MAT = \hat{\gamma}_{00} + \hat{\gamma}_{01}MSES + \hat{\gamma}_{10}CSES.$$

Nagrinėtiems duomenims gauname:

$$MAT = 175,94 + 16,56MSES + 8,04CSES.$$

V.Čekanašius, G.Murauskas

61

## Modelio tyrimas su SPSS

Įstatę konkrečias kintamųjų CSES ir MSES reikšmes, gautume testo rezultato prognozę konkrečiam mokiniui.

Be to, ši regresinė lygtis leidžia įvertinti mokyklos įtaką. Iš tikro, kiekvienas papildomas MSES balas padidina testo rezultatą 16,56 taško.

$$MAT = 175,94 + 16,56MSES + 8,04CSES.$$

V.Čekanašius, G.Murauskas

62

## Kategorinis kintamasis

- Iširsime modelį su antrojo lygmens kategoriniu kintamuoju VK, nurodančiu mokyklos priklausomybę Vilniaus miestui.
- Kategorinis (nebūtinai dvireikšmis) kintamasis, patalpinamas į laukelį Factor(s).
- Tirsime ankstesnį modelį, papildytą šiuo antrojo lygmens kintamuoju.
- Suformuluosime sąlygas tiksliau.

V.Čekanašius, G.Murauskas

63

## Kategorinis kintamasis

- Mokinio matematikos testo rezultatas priklauso nuo jo socialinio ir ekonominio statuso:

$$MAT = \beta_0 + \beta_1CSES + e.$$

- Mokyklos įtaka priklauso nuo mokyklos socialinio ir ekonominio statuso vidurkio ir mokyklos lokalizacijos:

$$\beta_0 = \gamma_{00} + \gamma_{01}MSES + \gamma_{12}VK + u_0.$$

V.Čekanašius, G.Murauskas

64

## Kategorinis kintamasis

- Vilniaus mokyklose socialinio ir ekonominio statuso reikšmė kitokia nei kitose mokyklose:

$$\beta_1 = \gamma_{10} + \gamma_{11}VK + u_1.$$

- Čia  $e$  turi dispersiją  $\sigma^2$ .

- Paklaidos  $u_0, u_1$  turi dispersijas  $\tau_{00}, \tau_{11}$  ir gali koreliuoti (kovariacija  $\tau_{01}$ ).

V.Čekanašius, G.Murauskas

65

## Kategorinis kintamasis

- Jungtinė lygtis:

$$MAT = \gamma_{00} + \gamma_{01}MSES + \gamma_{10}CSES + \gamma_{02}VK + \gamma_{12}VK \times CSES + [u_1CSES + u_0 + e].$$

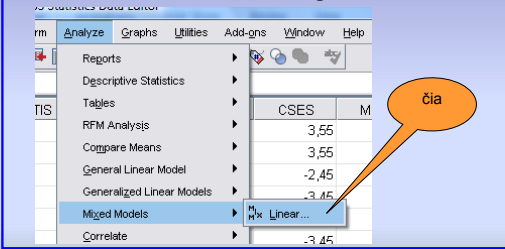
- Fiksuotieji kintamieji: CSES, MSES, VK, VK×CSES.
- Atsitiktiniai kintamieji: CSES ir postūmis.

V.Čekanašius, G.Murauskas

66

## Modelio tyrimas su SPSS

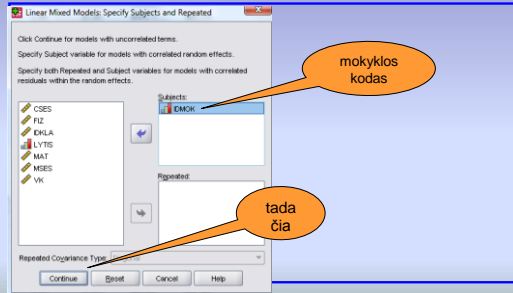
Atidaromas **Mixed Models** langas.



V.Čekanaivičius, G. Murauskas

67

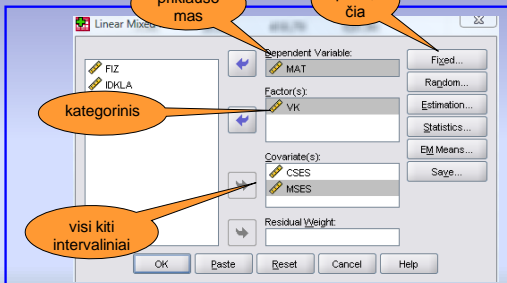
## Modelio tyrimas su SPSS



V.Čekanaivičius, G. Murauskas

68

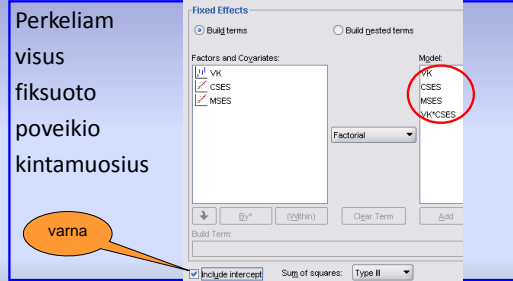
## Modelio tyrimas su SPSS



V.Čekanaivičius, G. Murauskas

69

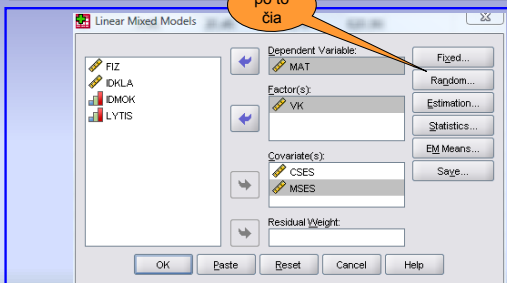
## Modelio tyrimas su SPSS



V.Čekanaivičius, G. Murauskas

70

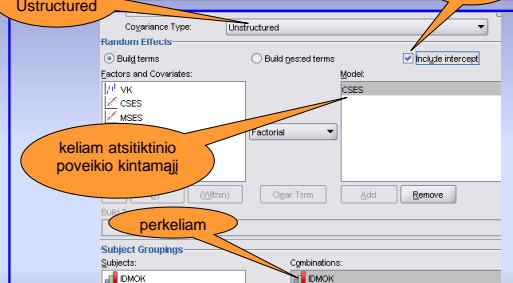
## Modelio tyrimas su SPSS



V.Čekanaivičius, G. Murauskas

71

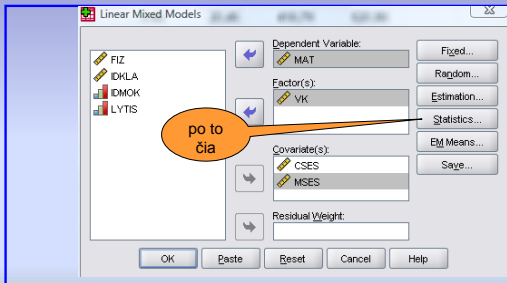
## Modelio tyrimas su SPSS



V.Čekanaivičius, G. Murauskas

72

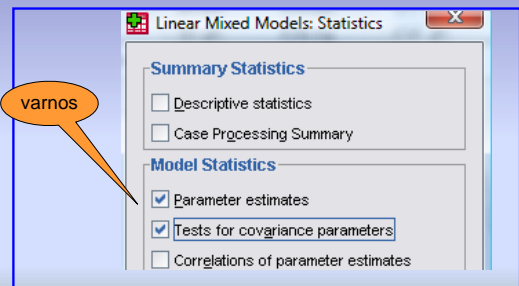
## Modelio tyrimas su SPSS



V.Čekanašius, G.Murauskas

73

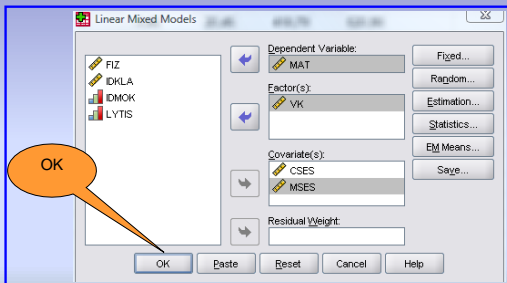
## Modelio tyrimas su SPSS



V.Čekanašius, G.Murauskas

74

## Modelio tyrimas su SPSS



V.Čekanašius, G.Murauskas

75

## Modelio tyrimas su SPSS

### Planas:

- Lyginsime AIC ir kitus informacinius indeksus su ankstesniu modeliu.
- Tikrinsime, kurie fiksuoti parametrai statistiškai nereikšmingi.
- Tikrinsime, kurie atsitiktiniai parametrai statistiškai nereikšmingi.
- Tikrinsime, kiek sumažėjo atsitiktinių parametru dispersijų įverčiai.

V.Čekanašius, G.Murauskas

76

## Modelio tyrimas su SPSS

- Akaikės ir kitų indeksų sumažėjimas bent per 10 vienetų jau laikomas svarbiu.

Information Criteria <sup>a</sup>		Information Criteria <sup>a</sup>	
-2 Restricted Log Likelihood	6091,135	-2 Restricted Log Likelihood	6109,830
Akaike's Information Criterion (AIC)	6099,133	Akaike's Information Criterion (AIC)	6117,830
Hurvich and Tsai's Criterion (AICC)	6099,210	Hurvich and Tsai's Criterion (AICC)	6117,904
Bozdogan's Criterion (CAIC)	6120,286	Bozdogan's Criterion (CAIC)	6138,996
Schwarz's Bayesian Criterion (BIC)	6116,286	Schwarz's Bayesian Criterion (BIC)	6134,996

The information criteria are displayed for two forms.  
a. Dependent variable: MAT

Nujasis modelis      Ankstesnis modelis

V.Čekanašius, G.Murauskas

77

## Modelio tyrimas su SPSS

Ne visi fiksuoto poveikio kintamieji statistiškai reikšmingi.

Type III Tests of Fixed Effects <sup>a</sup>				
Source	Numerator df	Denominator df	F	Sig.
Intercept	1	21,942	,117	,735
VK	1	20,264	7,497	,013
CSES	1	19,950	41,030	,000
MSES	1	21,875	52,062	,000
VK * CSES	1	19,947	,638	,434

V.Čekanašius, G.Murauskas

78

## Modelio tyrimas su SPSS

- Matome, kad reikėtų pabandyti iš modelio pašalinti kintamųjų sąveiką VKxCSES . Galbūt ir atsisakyti konstantos  $\gamma_{00}$ .
- Išsamesnė pavyzdžio analizė pateikta knygos skyrelyje *HLM modelių tyrimas SPSS programa*.

V.Čekanašius, G.Murauskas

79

## Svarbi problema

- Jei I lygmenyje įtrauksime labai stipriai su rezultatu koreliuojantį kintamąjį, tai paaiškės, kad joks HLM modelis netinka.
- Pvz., jei įtrauksime fizikos testo rezultatą FIZ:

$$MAT = \beta_0 + \beta_1 FIZ + e.$$

- Paaiškės, kad geriau mokantys fiziką, geriau ir matematiką išmano. Ir kas iš to? Norint pagerinti matematikos žinias, reikia sustiprinti fizikos mokymą? Kažin, kažin.

V.Čekanašius, G.Murauskas

80

## Svarbi problema

- Modelį reikia sudaryti sąmoningai mąstant apie tyrimo tikslus, o ne pagal stipriausias koreliacijas.
- Norime sužinoti, **kas daro įtaką** matematikos žinioms.
- Pavyzdžiui, koks yra mokyklos vaidmuo (II lygmuo).
- Arba koks socioekonominio statuso vaidmuo (I lygmuo).

V.Čekanašius, G.Murauskas

81

## METODINIAI PATARIMAI

- Pradedame nuo besąlyginio modelio.
- Kurie kintamieji yra pastovaus poveikio, o kurie – kintamojo padės nustatyti padės toks algoritmas:
  - 1) Nustatomi pirmojo ir antrojo lygmens kintamieji.
  - 2) Apibrėžiamos spėjamos priklausomybės.
  - 3) Lygtimis užrašoma, kas nuo ko priklauso. Reikia nepamiršti antrojo lygmens lygtyje pridėti atsitiktinę paklaidą **u**. Pirmojo lygmens lygčiai pridedama **e**.
  - 4) Antrojo lygmens lygtis įstatoma į pirmojo lygmens lygtį.
  - 5) Jungtinėje lygtyje surenkami nariai prie **u** ir **e**.
  - 6) Kintamieji be **u** - pastovaus poveikio; kintamieji su **u** - atsitiktinio poveikio. Beje, tas pats kintamasis gali įeiti į abu sąrašus.

V.Čekanašius, G.Murauskas

82