

LEKCE 06a

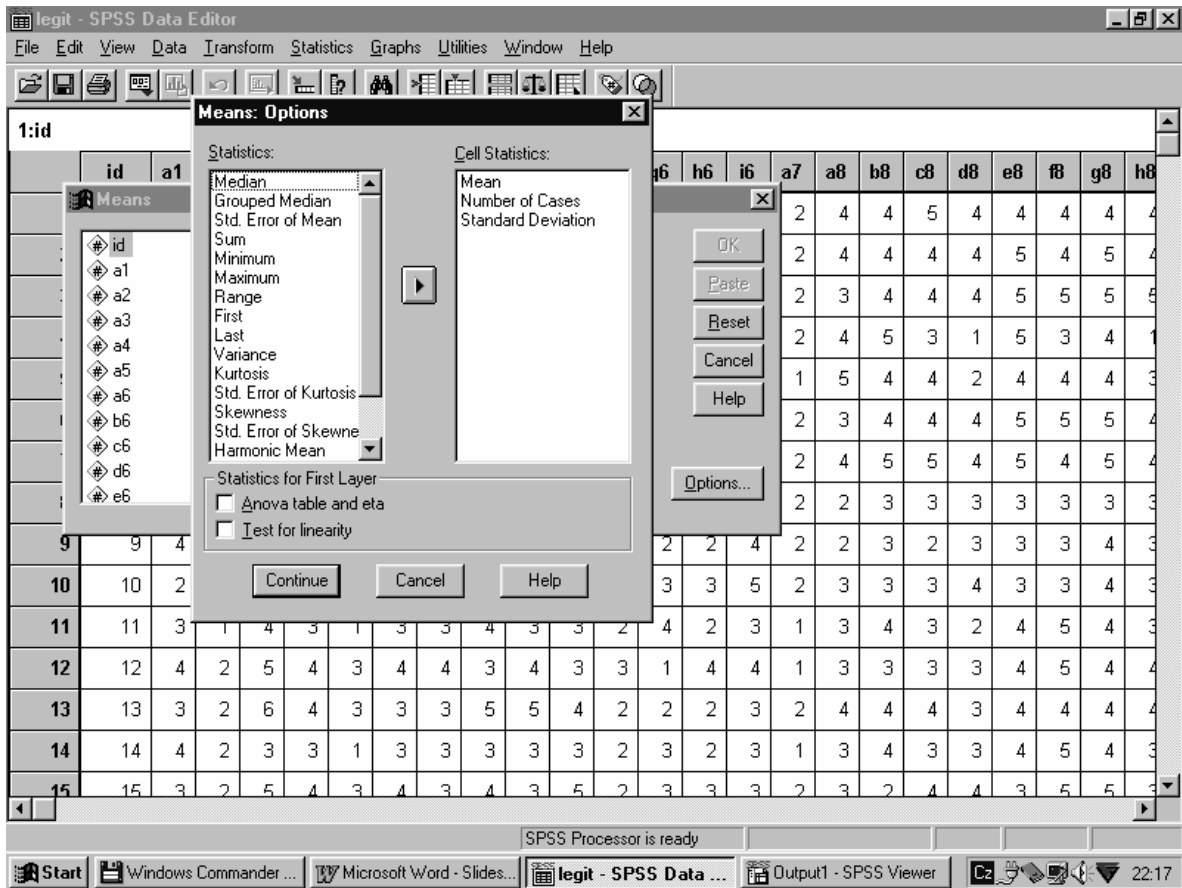
SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT JEJICH KARDINÁLNÍCH CHARAKTERISTIK

MEANS

The screenshot shows the SPSS Data Editor window with the 'Statistics' menu open. The 'Compare Means' option is selected, and a sub-menu is visible with options: 'Means...', 'One-Sample T Test...', 'Independent-Samples T Test...', 'Paired-Samples T Test...', and 'One-Way ANOVA...'. The data table below shows the following columns: 'age', 'sex', 'educ', 'impjob', 'hrs1', 'degree', and 'anomia'. The first row of data is: 1, 43, 1, 11, 2, 35, 1.

The screenshot shows the SPSS Data Editor window with the 'Means' dialog box open. The 'id' variable is selected in the 'Dependent List'. The data table below shows the following columns: 'id', 'a1', 'a2', 'a3', 'a4', 'a5', 'a6', 'b6', 'c6', 'd6', 'e6', 'f6', 'g6', 'h6', 'i6', 'a7', 'a8', 'b8', 'c8', 'd8', 'e8', 'f8', 'g8'. The first row of data is: 1, 9, 4, 2, 10, 4, 3, 5, 5, 4, 4, 5, 4, 2, 2, 4, 2, 2, 3, 2, 3, 3, 3, 4.

**LEKCE 6: SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT JEJICH KARDI-
NÁLNÍCH CHARAKTERISTIK**



POROVNÁVÁNÍ PRŮMĚRŮ - PROCEDURA MEANS

JEDNODUCHÁ VERZE

Q110A Čistý měsíční rodinný příjem

Q84 Pohlaví	Mean	N	Std. Deviation
1 muž	14977.36	514	6573.85
2 žena	13914.52	606	6935.39
Total	14402.53	1120	6789.49

nebo s více charakteristikami

Q110A Čistý měsíční rodinný příjem

Q84 Pohlaví	N	Mean	Std. Deviation	Minimum	Maximum
1 muž	514	14977.36	6573.85	400	45000
2 žena	606	13914.52	6935.39	3200	60000
Total	1120	14402.53	6789.49	400	60000

LEKCE 6: SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT JEJICH KARDI-
NÁLNÍCH CHARAKTERISTIK

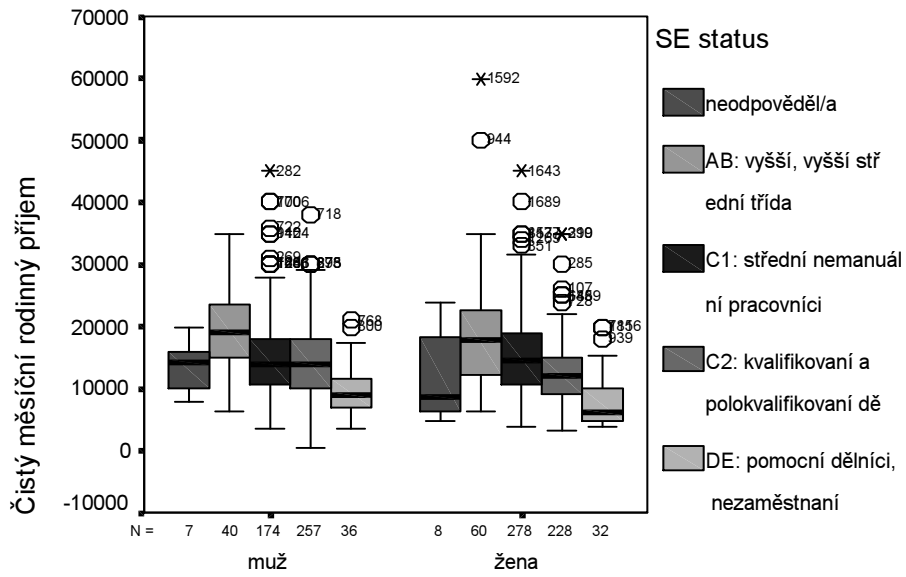
VÍCESTUPŇOVÁ VERZE - první pohlaví

Q110A Čistý měsíční rodinný příjem

Q84 Pohlaví	Q111B SE status	N	Mean	Std. Deviation	Minimum	Maximum
1 muž	1 AB: vyšší, vyšší střední třída	40	19589.76	7158.62	6400	35000
	2 C1: střední nemanuální pracovníci	174	15626.29	7372.92	3500	45000
	3 C2: kvalifikovaní a polokvalifikovaní dělníci	257	14545.74	5621.51	400	38000
	4 DE: pomocní dělníci, nezaměstnaní	36	10091.20	4626.45	3483	21000
	Total	507	14994.16	6604.38	400	45000
2 žena	1 AB: vyšší, vyšší střední třída	60	18885.53	8734.63	6400	60000
	2 C1: střední nemanuální pracovníci	278	14945.75	7030.42	4000	45000
	3 C2: kvalifikovaní a polokvalifikovaní dělníci	228	12227.90	5229.96	3200	35000
	4 DE: pomocní dělníci, nezaměstnaní	32	8085.76	4899.91	4000	20000
	Total	598	13935.16	6929.92	3200	60000
Total	1 AB: vyšší, vyšší střední třída	100	19166.48	8111.85	6400	60000
	2 C1: střední nemanuální pracovníci	452	15207.94	7163.86	3500	45000
	3 C2: kvalifikovaní a polokvalifikovaní dělníci	485	13456.72	5557.46	400	38000
	4 DE: pomocní dělníci, nezaměstnaní	69	9146.47	4828.19	3483	21000
	Total	1105	14421.22	6799.92	400	60000

GRAFICKÁ PODOBA - (GRAPHS - BOXPLOT - CLUSTERED)

LEKCE 6: SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT JEJICH KARDI-
NÁLNÍCH CHARAKTERISTIK



Pohlaví

Analysis weighted by W

LEKCE 06b

TESTOVÁNÍ VÝZNAMNOSTI ROZDÍLŮ STŘEDNÍCH HODNOT PROMĚNNÝCH U DVOU RŮZNÝCH SKUPIN

Existují ovšem rozdíly, které jsme našli ve výběrovém souboru i v souboru základním? Nebo také, porovnáváme-li výsledky dvou výběrových souborů, byly vybrány ze stejného základního souboru?

T-TEST

T-test je testem významnosti rozdílů dvou středních hodnot (aritmetických průměrů). Je založen na STUDENTOVĚ ROZDĚLENÍ (t rozdělení).

Pomocí t-testu řešíme vlastně:

- Buď otázku zda jsou dva populační průměry stejné neboli zda mezi nimi není rozdílu = nulová hypotéza.

Např. průměrné mzdy mužů a žen v základním souboru.

- Nebo otázku, zda jsou či nejsou porovnávány výběry z jednoho základního souboru.

FORMY T - testu

- T-TEST pro jediný výběr neboli ONE-SAMPLE T TEST (jak se liší statistika a parametr nebo jak se liší statistika od nějakého standardu)
- T-TEST pro dva nezávislé výběry (jak se liší 2 parametry – populační průměry) Je modifikací předchozího, zahrnuje informaci o variabilitě dvou nezávislých průměrů (průměrů z nezávislých výběrů). Standardní chyba průměru rozdílů je odhadována z variancí a velikosti výběrů dvou nezávislých skupin.
- T-TEST pro párovaná data

INDEPENDENT SAMPLES T-TEST

SPSS Processor is ready

Start Windows Commander ... Microsoft Word - doku... gss - SPSS Data E... Output1 - SPSS Viewer 18:38

SPSS Processor is ready

Start Windows Commander ... Microsoft Word - doku... gss - SPSS Data E... Output1 - SPSS Viewer 18:39

LEKCE 6: SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT JEJICH KARDI-NÁLNÍCH CHARAKTERISTIK

POSTUP:

Otestujeme shodu variancí obou rozložení pomocí F-testu.

- Nelze-li zamítnout nulovou hypotézu o shodě variancí (signifikance u F větší jak 0,01), pak použijeme T-testu pro případ EQUAL VARIANCES ASSUMED.
- Lze-li zamítnout nulovou hypotézu o shodu variancí (signifikance u F menší jak 0,01), pak použijeme T-testu pro případ EQUAL VARIANCES NOT ASSUMED.

Otestujeme shodu průměrů pomocí T-testu (použití jedné či druhé verze určí F-test - viz výše). Testujeme nulovou hypotézu neboli předpoklad shody. Signifikance T-testu nižší než 0,01 nám ji umožní odmítnout (je-li vyšší jak 0,01, odmítnout ji nemůžeme).

Příklad

- Existuje statisticky významný rozdíl mezi průměrným věkem spokojených a nespokojených se svou prací?
 - Neboli existuje rozdíl, který vznikl při porovnání těchto průměrných věků, spočítaných z pozorovaných hodnot v našem výběrovém souboru, také v základní populaci, z níž byl výběr proveden?
 - Nebo je riziko, že tento rozdíl vznikl jen díky výběrové chybě nepřijatelně vysoké?
- Výpočet deskriptivních statistik:

Group Statistics

		SATJOB2 Job Satisfaction	N	Mean	Std. Deviation	Std. Error Mean
AGE	Age of Respondent	1 Very satisfied	325	41,50	11,54	,64
		2 Not very satisfied	419	39,57	10,79	,53

Výpočet t-statistiky a její otestování:

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
AGE	Age of Respondent	Equal variances assumed	,377	,540	2,347	742	,019	1,93	,82	,32	3,54
		Equal variances not assumed			2,327	672,439	,020	1,93	,83	,30	3,56

Předpoklad shody variancí (NULOVOU HYPOTÉZU) nelze odmítnout (54% riziko chyby je příliš vysoké).

T- statistika se získá: $1,93/0,82 = 2,35$ neboli použitím EQUAL VARIANCES ASSUMED.

df = N-2

Průměrný věk obou výběrů se liší o 1,93 let.

Existuje jen 1,9% pravděpodobnost pozorovaného rozdílu mezi průměry, přinejmenším této velikosti, pokud jsou si oba populační průměry rovny (H_0) Nulovou hypotézu lze tedy zamítnout.

PAIRED T-TEST

SPSS Data Editor window showing a data grid with columns labeled 'id', 'a1', 'a2', 'a3', 'a4', 'a5', 'a6', 'b6', 'c6', 'd6', 'e6', 'f6', 'g6', 'h6', 'i6', 'a7', 'a8', 'b8', 'c8', 'd8', 'e8', 'f8', 'g8', 'h8'. The data grid contains numerical values for each cell. A 'Paired-Samples T Test' dialog box is overlaid on the grid. The dialog box has a list of variables on the left: id, a1, a2, a3, a4, a5, a6, h6. The 'Paired Variables' field is empty. Below the list, 'Current Selections' shows 'Variable 1:' and 'Variable 2:'. Buttons for 'OK', 'Paste', 'Reset', 'Cancel', 'Help', and 'Options...' are visible on the right side of the dialog box.

SPSS Data Editor window showing the same data grid as above. The 'Paired-Samples T Test: Options' dialog box is now open. It shows 'Confidence Interval: 95 %'. Under 'Missing Values', the radio button for 'Exclude cases analysis by analysis' is selected. The 'Current Selections' section shows 'Variable 1:' and 'Variable 2:'. Buttons for 'Continue', 'Cancel', and 'Help' are visible on the right side of the dialog box.

ONE-SAMPLE T-TEST

SPSS Processor is ready

Start Windows Commander ... Microsoft Word - Slides... legit - SPSS Data ... Output1 - SPSS Viewer 22:05

SPSS Processor is ready

Start Windows Commander ... Microsoft Word - Slides... legit - SPSS Data ... Output1 - SPSS Viewer 22:05

LEKCE 6: SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT JEJICH KARDI- NÁLNÍCH CHARAKTERISTIK

PŘÍKLAD

Pro testování hypotézy, že výběr pochází z populace se známým průměrem respektive standardem (zde je to 40 pracovních hodin týdně), ale neznámou standardní odchylkou, musíme VYPOČÍTAT T-STATISTIKU.

Výpočet je stejný jako u standardizovaného skóre, jen místo standardní odchylky pro populaci (parametru) je použita její hodnota ve výběru (statistika).

$$t = (47,73-40)/0,795 = 9,722$$

One-Sample Test

	Test Value = 40					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
HRS1 Number of Hours Worked Last Week	9,722	247	,000	7,73	6,16	9,30

pravděpodobnost hodnoty T
větší jak +9,72 nebo menší
jak -9,72 je menší jak 0,0001
 t - hodnota
 stupeň volnosti
 interval spolehlivosti pro rozdíl
mezi průměry
 rozdíl mezi
průměrem
ve výběru a
hypotetickým
průměrem
v populaci

Je nepravděpodobné, že graduované osoby (v základním souboru) pracují v průměru 40 hodin týdně. Riziko chyby při odmítnutí nulové hypotézy „že se jejich pracovní týdenní doba neliší od standardu“ je velmi nízké (méně než 0,1%).

Pravděpodobnost je 2-tailed, protože sčítá obě oblasti t- rozložení (menší jak -9,72 a větší jak +9,72): průměrný počet hodin graduovaných může teoreticky být menší i větší než 40 hodin týdně.

LEKCE06c

JAK TESTOVAT NULOVOU HYPOTÉZU O SHODĚ NĚKOLIKA POPULAČNÍCH PRŮMĚRŮ

The screenshot shows the SPSS Data Editor window with the 'Statistics' menu open. The 'One-Way ANOVA...' option is highlighted. The data table below shows the following values:

	filter_\$	satjobre	agegroup
1	1	2	3
2	1	2	3
3	1	1	3
4	0	2	3
5	0	.	4
6	2	0	4
7	1	1	4
8	0	0	4
9	1	1	2
10	0	1	4
11	1	2	1
12	1	2	1
13	1	2	4
14	0	.	4
15	0	2	2

The screenshot shows the 'One-Way ANOVA: Options' dialog box. The 'Statistics' section has 'Descriptive' and 'Homogeneity-of-variance' checked. The 'Missing Values' section has 'Exclude cases analysis by analysis' selected. The data table below shows the following values:

	husbhr	wifehr	husbft	wifft	educdiff	income4	filter_\$	satjobre	agegroup
1	3	1	2	3
2	3	1	2	3
3	3	1	1	3
4	0	2	3
5	4	0	.	4
6	2	0	.	4
7	40	40	1	1	.	4	1	1	4
8	1	0	.	4
9	-2,00	4	1	1	2
10	4,00	1	0	1	4
11	6,00	3	1	2	1
12	3	1	2	1
13	40	40	1	1	.	2	1	2	4
14	1	0	.	4
15	1	0	2	2

LEKCE 6: SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT JEJICH KARDI-

5:agegroup

	husbhr	wifehr	husbft	wifeft	educdiff	income4	filter_\$	satjobre	agegroup
1						3	1	2	3
2						3	1	2	3
3						3	1	1	3
4						.	0	2	3
5						4	0	.	4
6						2	0	.	4
7						4	1	1	4
8						1	0	.	4
9					-2,00	4	1	1	2
10					4,00	1	0	1	4
11					6,00	3	1	2	1
12					.	3	1	2	1
13	40	40	1	1	.	2	1	2	4
14					.	1	0	.	4
15					.	1	0	2	2

5:agegroup

	husbhr	wifehr	husbft	wifeft	educdiff	income4	filter_\$	satjobre	agegroup
1						3	1	2	3
2						3	1	2	3
3						3	1	1	3
4						.	0	2	3
5						4	0	.	4
6						2	0	.	4
7						4	1	1	4
8						1	0	.	4
9						4	1	1	2
10					4,00	1	0	1	4
11					6,00	3	1	2	1
12					.	3	1	2	1
13	40	40	1	1	.	2	1	2	4
14					.	1	0	.	4
15					.	1	0	2	2

**LEKCE 6: SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT JEJICH KARDI-
NÁLNÍCH CHARAKTERISTIK****ONE-WAY ANALÝZA - (analýza rozptylu)**

T-TEST DOVOLUJE POROVNAT POUZE DVA PRŮMĚRY.

Pokud bychom chtěli porovnat průměrů více musely bychom provést celou řadu t-testů pro všechny možné dvojice jež připadají v úvahu.

Příklad:

Otázka: *Liší se v základním souboru průměrné platy u různých vzdělanostních kategorií?*

1. základní, nevyučen;
2. základní vyučen;
3. střední bez maturity;
4. střední s maturitou;
5. vysokoškolské bakalářské;
6. vysokoškolské magisterské a vyšší)?

V TOMTO PŘÍPADĚ BYCHOM MUSELI PROVĚST 15 T-TESTŮ neboť $(6 \times 6 - 6) / 2 = 15$

ONE-WAY ANALÝZA (analýza rozptylu) dovoluje provést mnohonásobné porovnání.

Její aplikace zahrnuje:

- Výpočet deskriptivních statistik (popis dat) pro porovnávané skupiny
- Určení intervalů spolehlivosti průměrů jednotlivých porovnávaných skupin
- Testování předpokladů použití
- Testování nulové hypotézy neboli analýza variancí (ANOVA, zde ONE-WAY ANOVA).
- Porovnávání dvou odhadů variability.
- Vícenásobné porovnávání.

TESTOVANÁ NULOVÁ HYPOTÉZA

NULOVÁ HYPOTÉZA: Populační průměry všech srovnávaných skupin jsou shodné.

ALTERNATIVNÍ HYPOTÉZA: Rozdíly mezi jednotlivými průměry - přinejmenším mezi dvěma z nich - existují (nevíme však mezi kterými).

LEKCE 6: SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT JEJICH KARDI- NÁLNÍCH CHARAKTERISTIK

Příklad:

Porovnání průměrných hodnot míry ANOMIE (Sroleho škála anomie) mezi voliči jednotlivých politických stran.

MNOHONÁSOBNÉ SROVNÁNÍ - Bonferroni test

Multiple Comparisons

Dependent Variable: ANOMIE SROLLUV INDEX ANOMIE

Bonferroni

		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval		
					Lower Bound	Upper Bound	
(I) A66RED Volená strana - red		(J) A66RED Volená strana - red					
liší se od	1 Republ	1 Republ					
		2 ODS	2,11*	,37	,000	1,07	3,16
		3 KDU-CSL	1,29*	,43	,031	6,62E-02	2,51
		4 CSSD	,85	,37	,228	-,20	1,90
		5 KSCM	,54	,40	1,000	-,59	1,66
liší se od všech ostatních	2 ODS	1 Republ	-2,11*	,37	,000	-3,16	-1,07
		2 ODS					
		3 KDU-CSL	-,83*	,26	,018	-1,57	-8,28E-02
		4 CSSD	-1,26*	,15	,000	-1,68	-,85
		5 KSCM	-1,58*	,20	,000	-2,15	-1,00
liší se od	3 KDU-CSL	1 Republ	-1,29*	,43	,031	-2,51	-6,62E-02
		2 ODS	,83*	,26	,018	8,28E-02	1,57
		3 KDU-CSL					
		4 CSSD	-,43	,27	1,000	-1,18	,31
		5 KSCM	-,75	,30	,129	-1,59	9,75E-02
liší se od	4 CSSD	1 Republ	-,85	,37	,228	-1,90	,20
		2 ODS	1,26*	,15	,000	,85	1,68
		3 KDU-CSL	,43	,27	1,000	-,31	1,18
		4 CSSD					
		5 KSCM	-,31	,20	1,000	-,89	,26
liší se od	5 KSCM	1 Republ	-,54	,40	1,000	-1,66	,59
		2 ODS	1,58*	,20	,000	1,00	2,15
		3 KDU-CSL	,75	,30	,129	-9,75E-02	1,59
		4 CSSD	,31	,20	1,000	-,26	,89
		5 KSCM					

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

INTERPRETACE MNOHONÁSOBNÉHO SROVNÁNÍ

- Průměrná míra indexu anomie voličů ODS se statisticky významně liší ($HV=0,05$) od průměrných měr tohoto indexu ve všech skupinách voličů ostatních politických stran.
- Průměrná míra indexu anomie voličů KDU-ČSL se statisticky významně liší ($HV=0,05$) od průměrných měr tohoto indexu voličů ODS na jedné straně a voličů Republikánů na druhé straně. Neliší se od průměrných měr tohoto indexu mezi voliči ČSSD ani mezi voliči KSČM.
- Průměrná míra indexu anomie voličů ČSSD a voličů KSČM se statisticky významně liší ($HV=0,05$) jen od průměru tohoto indexu u voličů ODS (neliší se od průměrných měr tohoto indexu mezi voliči KDU-ČSL ani mezi voliči Republikánů).
- Průměrná míra indexu anomie voličů Republikánů se statisticky významně liší ($HV=0,05$) od průměrných měr tohoto indexu voličů ODS a KDU-ČSL (neliší se od průměrných měr tohoto indexu mezi voliči ČSSD ani mezi voliči KSČM).

