

# **Непараметарски тестови за $k$ независних узорака**

Никола Томовић

Марија Здолшек

18.5.2015.

## *Непараметарски тестови за $k$ независних узорака*

- непараметарске методе су развијене да се користе у случајевима када истраживач не зна ништа о параметрима обележја у популацији као ни о расподели обележја (отуда име непараметарски).
- непараметарске методе се не ослањају на процене параметара (као што су средња вредност или стандардна девијација) које описују расподелу обележја.
- карактеришу се тиме да расподела тест статистике која се у њима користи не зависи од расподеле и параметра посматраног обележја
- хипотезе које се тестирају непараметарским тестовима односе се на врсту расподеле обележја популације (тестира се сагласност узорка са претпостављеном расподелом), могу да се односе на случајност узорка, зависност два обележја итд.

## *Непараметарски тестови за $k$ независних узорака*

- *Тест медијане*
- *Kruskal-Wallis* тест
- *Jonckheere-Terpstra* тест

## *Тест медијане*

*Тест медијане* је непараметарски статистички тест који је посебан случај Пирсоновог  $\chi^2$  теста.

За датих  $k$  узорака обима  $n_1, n_2, \dots, n_k$  опсервација, рачуна се главна медијана свих  $n_1 + n_2 + \dots + n_k$  опсервација. Затим се формира табела  $2 \times k$ , где први ред представља број опсервација већих од главне медијане, док други ред представља број опсервација мањих или једнаких главној медијани.

Тада се на ту табелу може применити  $\chi^2$  тест независности.

$H_0$ : свих  $k$  узорака имају исту медијану

$H_1$ : најмање два узорка имају различите медијане

Тест статистика:

$$T = \frac{N^2}{ab} \sum_{i=1}^k \frac{(A_i - an_i/N^2)^2}{n_i}$$

Где је  $a$  број опсервација већих од главне медијане за све узорке,  $b$  број опсервација мањих или једнаких од главне медијане за све узорке,  $N$  укупан број опсервација и  $A_i$  број опсервација већих од главне медијане за узорак  $i$ . У случају  $H_0$  хипотезе, та статистика има  $\chi_{k-1}^2$  расподелу.

Ако је  $T > \chi_{k-1}^2(1 - \alpha)$ , где је  $\alpha$  ниво значајности, одбацује се хипотеза  $H_0$

## *Kruskal-Wallis test*

Je непараметарски метод за тестирање да ли је к независних узорака извучено из исте популације или идентички истих популација са једнаким медијанама на основу рангова. Прво се изврши ранкирање свих узорака заједно, нека је  $R_i$  сума ранкова за узорак  $i$ . Тада је Kruskal-Wallis test статистика дата са:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1)$$

У случају  $H_0$ : свих  $k$  узорака има исту медијане, тест статистика  $H$  има  $\chi_{k-1}^2$  расподелу.

Хипотеза  $H_0$  се одбацује ако је  $H > \chi_{k-1}^2(1 - \alpha)$  где је  $\alpha$  ниво значајности.

## *Jonckheere- Terpstra test*

Je непараметарски тест за  $k$  независних узорака.

$$H_0: \theta_1 = \theta_2 = \dots = \theta_k$$

$H_1: \theta_1 \leq \theta_2 \leq \dots \leq \theta_k$ , где је бар једна од неједнакости строга неједнакост.

$\theta_i$  представља медијану  $i$ -тог узорка.

Тест статистика је:

$$T = \frac{\sum U_{xy} - \frac{N^2 - \sum_{j=1}^k n_j^2}{4}}{\sqrt{\frac{N^2(2N + 3) - \sum_{j=1}^k n_j^2(2n_j + 3)}{72}}}$$

Где је  $U_{xy}$  број опсервација узорка  $y$  које су веће од сваке опсервације узорка  $x$ . Овако дата тест статистика има стандардну Нормалну расподелу.

# Медијана тест

## Пример

Analyze -> Nonparametric Tests-> Legacy Dialogs -> K Independent Samples...

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the path 'Nonparametric Tests' -> 'Legacy Dialogs' -> 'K Independent Samples...' is highlighted. The data table in the background contains 11 variables: 'id', 'pplata', 'plata', 'obrazovanje', 'kat\_posla', 'rasa', 'pol\_rasa', 'starost', 'r\_staz', and five unlabeled 'var' variables. The 'K Independent Samples...' option is selected in the 'Legacy Dialogs' submenu.

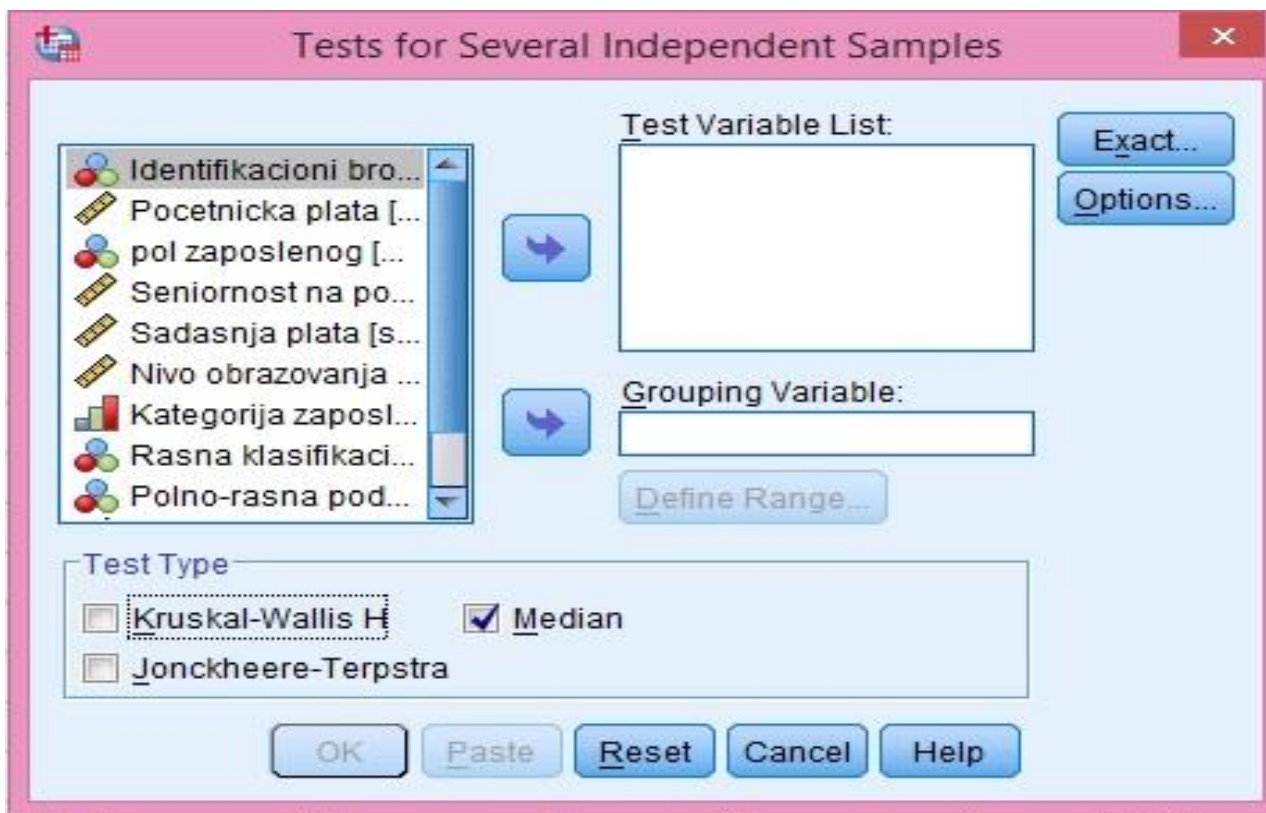
	id	pplata	plata	obrazovanje	kat_posla	rasa	pol_rasa	starost	r_staz	var	var	var	var	var
1	748	400	6300	8	1	0	3	63.83	22.92					
2	832	400	6360	8	1	0	3	55.25	3.58					
3	754	390	6480	8	1	0	3	55.50	.00					
4	869	400	6480	12	1	0	3	60.00	.00					
5	969	400	6480	12	1	0	3	62.33	15.00					
6	825	400	6540	12	1	1	4	60.50	13.58					
7	1083	400	6600	12	1	0	3	62.42	24.00					
8	1107	390	6660	8	1	0	3	62.50	34.33					
9	886	400	6720	8	1	0	3	59.08	6.33					
10	1127	400	6780	8	1	0	3	56.92	26.58					
11	921	360	6780	12	1	1	4	60.67	10.33					
12	935	400	6960	8	1	1	4	51.50	22.58					
13	831	400	6960	12	1	0	3	55.08	6.00					
14	940	400	7080	8	1	1	4	51.50	.00					
15	1128	400	7260	12	1	1	4	53.92	29.83					
16	749	400	7380	12	1	1	4	57.83	32.50					
17	647	400	7500	15	1	1	4	46.50	9.67					
18	826	400	7680	8	1	1	4	61.50	15.33					
19	995	390	7860	12	1	1	4	62.00	6.00					
20	1121	400	7860	15	1	1	4	51.00	19.00					
21	1034	420	7860	15	1	1	4	58.00	4.50					
22	1096	360	7860	15	1	1	4	60.50	1.92					
23	784	520	7860	8	1	1	4	55.92	8.50					



# Медијана тест

## Пример

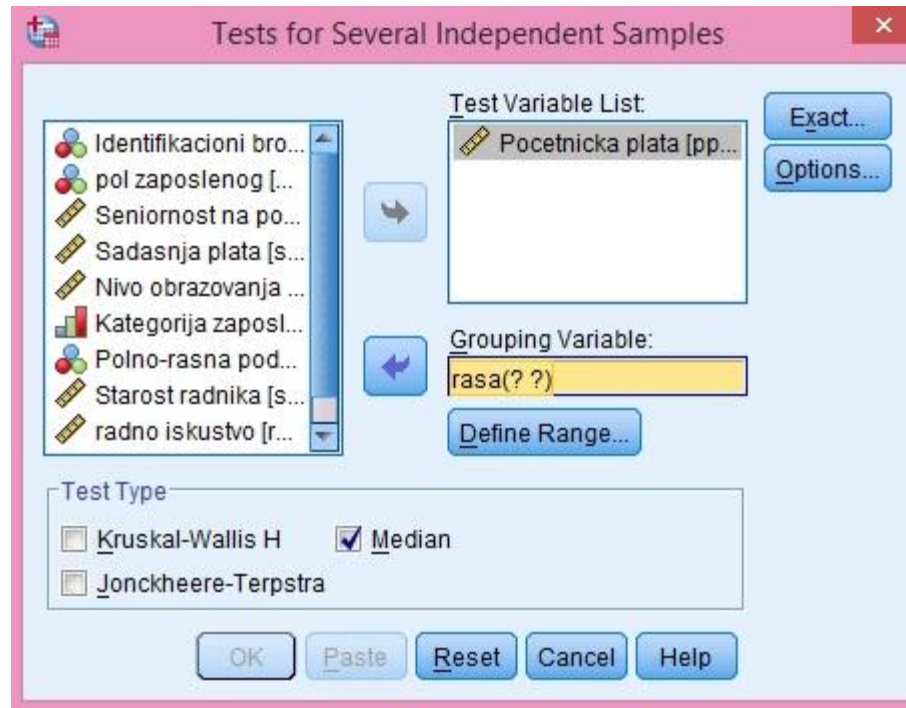
Отвориће се прозор као што је приказан на слици



# Медијана тест

## Пример

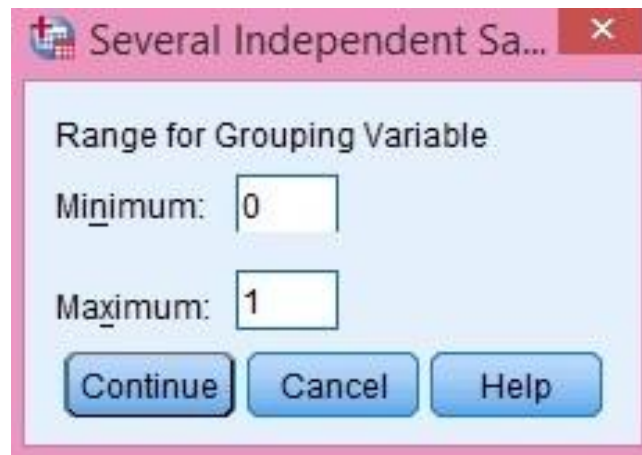
У прозорчић *Test Variable List* можемо убацити било који тип променљиве, док у прозорчић *Grouping Variable* убацујемо искључиво категоријску променљиву. У овом примеру ми ћемо радити на променљивима *početnicka plata* и *rasa*.



# Медијана тест

- *Пример*

- Кликом на дугме *Define Range* отвара се прозор дат на слици. У поље *minimum* и *maximum* уносе се најмања и највећа вредност категоричке променљиве. Као резултат посматраће се све групе које се налазе у опсегу [*minimum*, *maximum*] за које категоричка променљива има вредност из тог интервала. Специјално, може се унети подинтервал, (нпр. ако категоричка променљива узима вредности 1, 2, 3, 4 и 5 ако унесемо 2 у *minimum* и 4 у *maximum* посматраће се три групе.)



# Медијана тест

## Пример

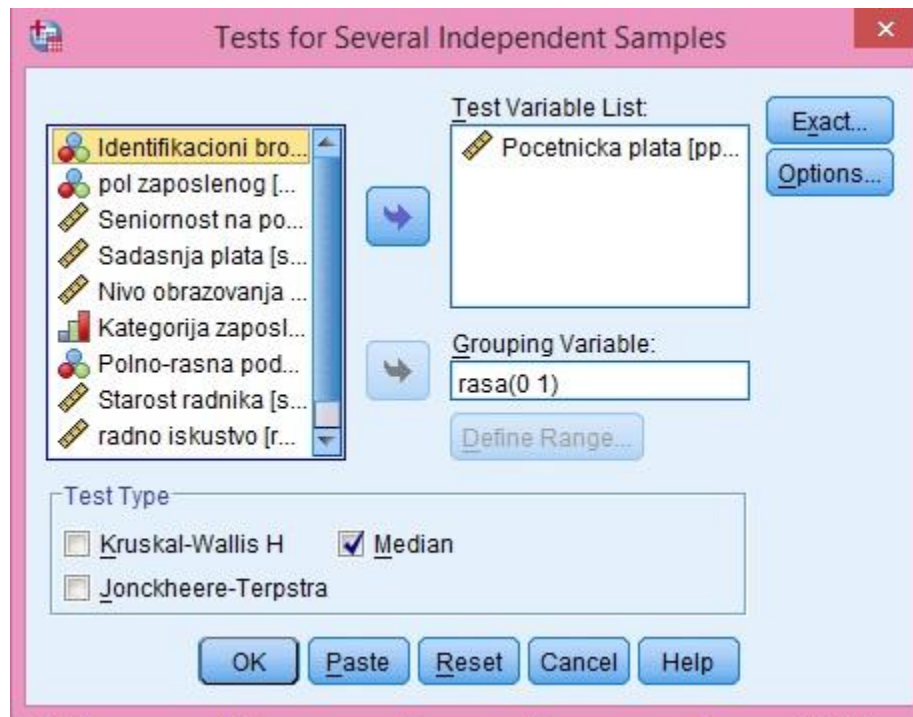
Кликом на дугме *Options* отвара се прозор приказан на слици. Можемо одабрати *Descriptive* и/или *Quartiles* које враћају дескриптивне статистике и /или квантите.



# Медијана тест

## Пример

Кликом на дугме *Continue* се враћамо у стари прозор



Кликом на *OK* се покреће *Медијана* тест.

# Медијана тест

## Пример

SPSS ће генерисати 3 табеле у зависности од тога шта је селековано у прозору *Options*. Табела *Test Statistics* даје информације о резултатима *Медијана* теста. *P-vrednost* теста је 0,000, што значи да се одбацује нулта хипотеза и постоји значајна разлика у медијанама ове две групе.

### Median Test

Frequencies		
	Rasna klasifikacija	
	bela rasa	druge rase
Pocetnicka plata > Median	181	29
<= Median	189	75

Test Statistics <sup>a</sup>		Pocetnicka plata
N		474
Median		6000,00
Chi-Square		14,556
df		1
Asymp. Sig.		,000
Yates' Continuity Correction	Chi-Square	13,716
	df	1
	Asymp. Sig.	,000

Double-click to activate

a. Grouping Variable: Rasna klasifikacija

# Kruskal-Wallis test

## Пример

- Проверићемо да ли почетничка плата има приближно исту медијану у групама раса-пол. Користићемо *Kruskal-Wallis* тест.

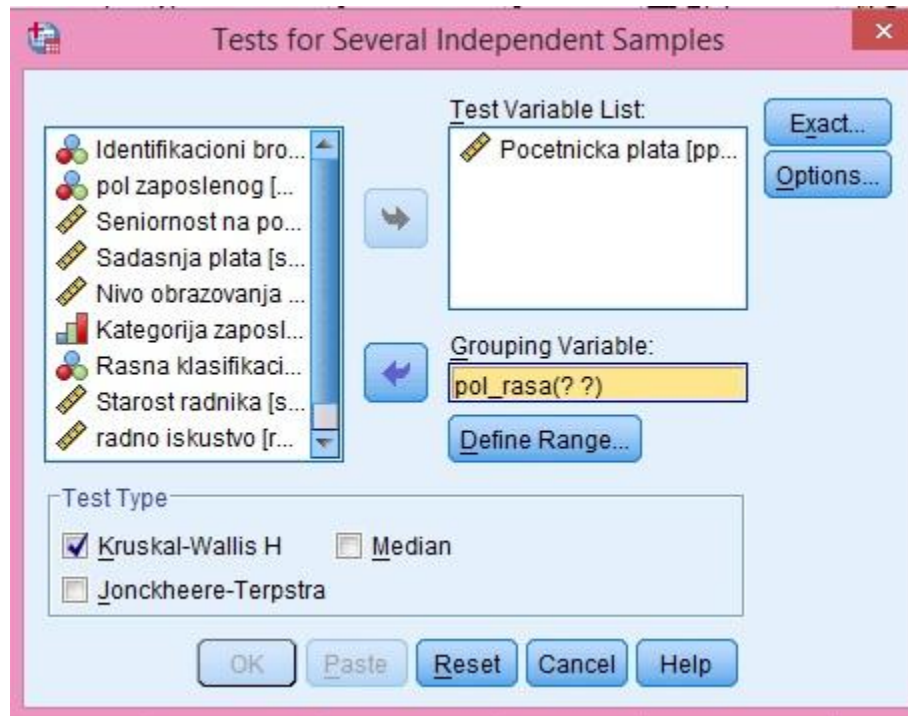
Analyze -> Nonparametric Tests-> Legacy Dialogs -> K Independent Samples...

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the path 'Nonparametric Tests' -> 'Legacy Dialogs' -> 'K Independent Samples...' is highlighted. The data table in the background has the following structure:

	id	pplata	plata	obrazovanje	kat_posla	rasa	pol_rasa	starost	r_staz	var	var	var	var	var
1	748	406	6300	8	1	0	3	63.83	22.92					
2	832	406	6360	8	1	0	3	55.25	3.58					
3	754	390	6480	8	1	0	3	55.50	.00					
4	869	406	6480	12	1	0	3	60.00	.00					
5	969	406	6480	12	1	0	3	62.33	15.00					
6	825	406	6540	12	1	1	4	60.50	13.58					
7	1083	406	6600	12	1	0	3	62.42	24.00					
8	1107	390	6660	8	1	0	3	62.50	34.33					
9	886	406	6720	8	1	0	3	59.08	6.33					
10	1127	406	6780	8	1	0	3	56.92	26.58					
11	921	360	6780	12	1	1	4	60.67	10.33					
12	935	406	6780	12	1	1	4	51.50	22.58					
13	831	406	6780	12	1	0	3	55.08	6.00					
14	940	406	6780	12	1	1	4	51.50	.00					
15	1128	406	6780	12	1	1	4	53.92	29.83					
16	749	406	6960	8	1	1	4	57.83	32.50					
17	647	406	6960	12	1	1	4	46.50	9.67					
18	826	406	7080	8	1	1	4	61.50	15.33					
19	995	390	7260	12	1	1	4	62.00	6.00					
20	1121	406	7380	12	1	1	4	51.00	19.00					
21	1034	420	7500	15	1	1	4	58.00	4.50					
22	1096	360	7680	15	1	1	4	60.50	1.92					
23	784	520	7860	8	1	1	4	55.92	8.50					

# *Kruskal-Wallis test*

Изаберемо *Kruskal-Wallis test* и *pol\_rasa* за Grouping Variable.





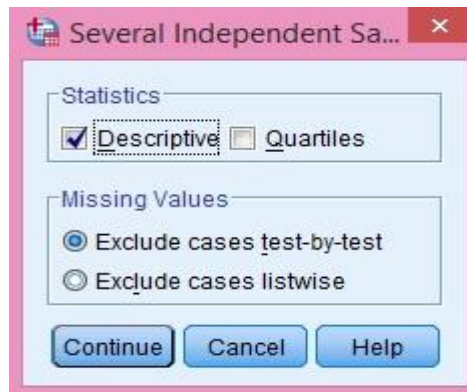
# Kruskal-Wallis test

## Пример

Minimum je 1 a Maximum 4.



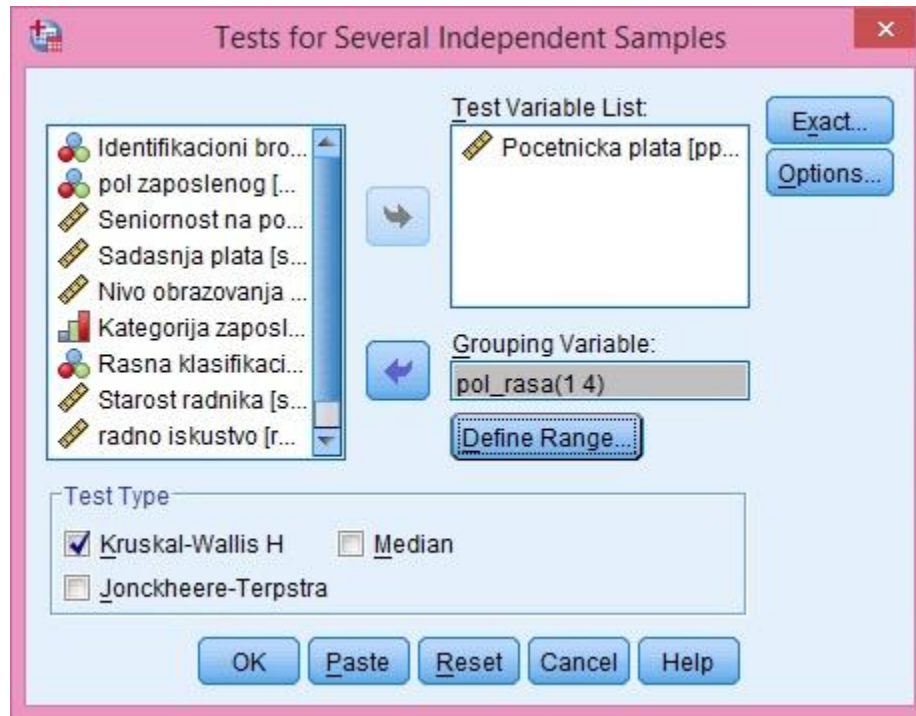
Као и код медијана теста може кликом на *Options* можемо одабрати *Descriptive* и/или *Quartiles*.



# *Kruskal-Wallis test*

## *Пример*

Кликом на дугме *Continue* се враћамо у стари прозор.



Кликом на *OK* се покреће *Kruskal-Wallis* тест.

# *Kruskal-Wallis test*

## *Пример*

Табела *Test Statistics* даје информације о резултатима *Kruskal-Wallis* теста. *P-vrednost* теста је 0,000, што значи да се одбацује нулта хипотеза и постоји значајна разлика у медијанама ове четири групе.

### **Kruskal-Wallis Test**

		<b>Ranks</b>	
	<b>Polno-rasna podela</b>	<b>N</b>	<b>Mean Rank</b>
Pocetnicka plata	muskarci bele rase	194	334,73
	muskarci drugih rasa	64	255,42
	zene bele rase	176	154,79
	zene drugih rasa	40	101,19
	Total	474	

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

	<b>Pocetnicka plata</b>
Chi-Square	203,326
df	3
Asymp. Sig.	,000

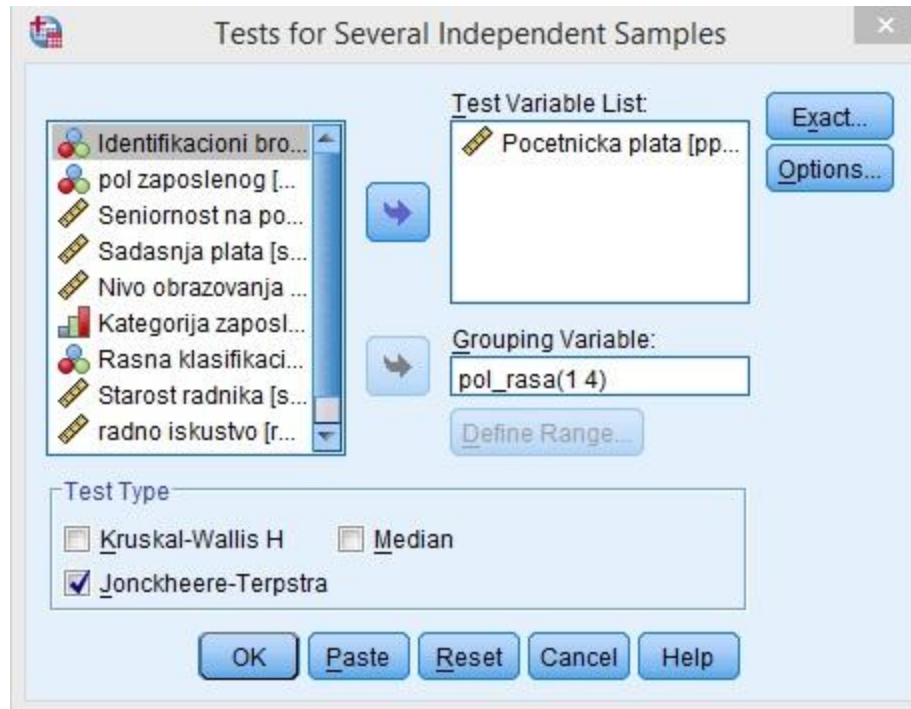
a. Kruskal Wallis Test

b. Grouping Variable:  
Polno-rasna podela

# *Jonckheere-Terpstra test*

## *Пример*

Сада желимо да проверимо да ли су медијане почетничких плата у групама у односу на пол-расу једнаке. Поступак је исти као код претходна два теста само ћемо изабрати *Jonckheere-Terpstra* тест.



# *Jonckheere-Terpstra test*

## *Пример*

Табела *Test Statistics* даје информације о резултатима *Kruskal-Wallis* теста. *P-vrednost* теста је 0,000, што значи да се одбацује нулта хипотеза и постоји значајна разлика у медијанама ове четири групе.

**Jonckheere-Terpstra Test<sup>a</sup>**

	Pocetnicka plata
Number of Levels in Polno-rasna podela	4
N	474
Observed J-T Statistic	14004,000
Mean J-T Statistic	37592,000
Std. Deviation of J-T Statistic	1610,057
Std. J-T Statistic	-14,650
Asymp. Sig. (2-tailed)	,000

a. Grouping Variable: Polno-rasna podela

**ХВАЛА НА ПАЖЊИ!**