

# Статистички софтвер 4

## Трећи час

Марија Радичевић

Математички факултет, Београд

2015.

# Садржај

- 1 *Анализа података*
  - Припремање података за анализу
  - Прелиминарна анализа података
- 2 *Примери*

Анализа података обухвата низ техника и метода, које служе као моћни алат за добијање значајних информација из података који су добијени било из примарних или секундарних извора података.

Комплетан истраживачки пројекат укључује:

- припремање прелиминарног плана анализе података
- преглед упитника и валидацију података
- едитовање (кориговање)
- кодирање
- пречишћавање и табелирање података

# Припремање података за анализу

Квалитет улазних података одређује квалитет информација које се добијају у поступку анализе података, а на основу којих се доносе одговарајуће пословне одлуке.

Поступак припремања података за анализу:

- 1 Прелиминарни план анализе података
- 2 Преглед упитника и валидација података
- 3 Уређивање података
- 4 Кодирање података
- 5 Унос података
- 6 Пречишћавање и статистичко модификовање података
- 7 Табелирање података
- 8 Селекција стратегије анализа података

# Преглед упитника и валидација података

## Провера прихватљивости упитника

- непотпуни делови упитника
- системски или случајно прескочена питања
- неадекватност упитника у смислу задатког типа узорка
- мала варијанса одговора

Валидација података подразумева проверавање опсервација, упитника или целокупног испитивања.

Обухвата:

- 1 скрининг
- 2 процедура
- 3 потпуност
- 4 превара

## Уређивање (едитовање) података

Проверавање података у смислу постојања грешака, било да се ради о грешкама испитивача или испитаника. Проблеми који се могу идентификовати током уређивања података:

- грешке испитивача
- испитаник није одговорио на поједина питања
- нејасан одговор испитаника
- логички неконзистентни одговори
- невољност испитаника да одговори на питања
- неадекватан испитаник

Третирање незадовољавајућих вредности

- 1 поновити контакт на терену
- 2 додељивање третмана недостајућих вредности
- 3 искључивање испитаника

# Шифровање (кодирање) података

Додељивање одређене шифре сваком могућем одговору на постављено питање. Услови који морају бити испуњени:

*услов искључености* категорије одговора морају бити међусобно искључиве

*услов потпуности* сваки одговор на постављено питање може да се додели одређеној категорији

## Унос података

- електронски облик
- остварује се преносом кодираних података из упитника у рачунарски систем
- посебно писани документи - шифарници
- формат радних табела (*Excel*, разни софтверски пакети)



# Пречишћавање и статистичка модификација података

Пречишћавање података се односи на проверу конзистентности података и третман недостајућих вредности.

Провера конзистентности:

- логичка неконзистентност
- вредности ван дозвољеног интервала
- екстремне вредности

Третман недостајућих вредности:

- замена недостајућих вредности неком неутралном вредношћу
- брисање опсервација
- брисање података само за појединачне променљиве у оквиру посматране опсервације

# Пречишћавање и статистичка модификација података

Статистичко модификовање података може се вршити на различите начине, а најчешти су:

- пондерисање података
- прекодирање података
- трансформација података мерених на различитим скалама мерења

# Табелирање података

Процес рачунања броја опсервација које припадају одређеној категорији посматраног обележја (променљиве).

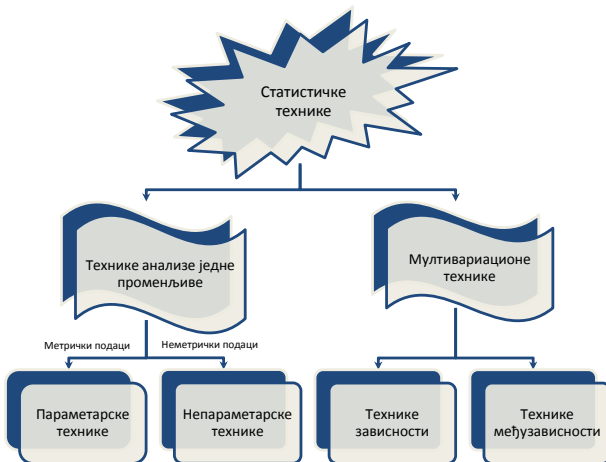
- једноструко табелирање
- унакрсно табелирање

# Избор стратегије анализе података

Фактори који утичу на избор стратегије анализе података:

- дефинисање проблема истраживања, приступ истраживању и дизајн истраживања
  - међусобна (не)зависност опсервација
  - број група које се посматрају
  - број мерења (променљивих) по једном објекту посматрања у истраживању
  - могућност контроле посматраних променљивих
- карактеристике података
- особине статистичких техника
- приступ истраживача

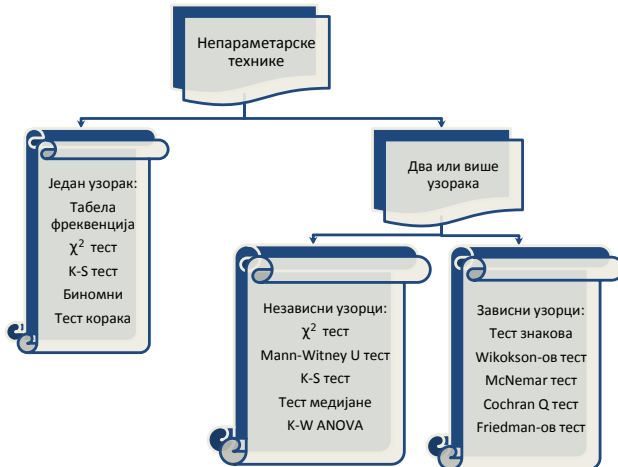
# Статистичке технике



# Параметарске технике



# Непараметарске технике



# Мултивариационе технике





# Прелиминарна анализа података

- Процедуре за сумарно приказивање података
- Истраживачку анализа података

# Frequencies

- табела фреквенција за изабрану променљиву
- низ статистичких дескриптивних величина
- дијаграм

Табела фреквенција садржи:

- апсолутне фреквенције
- проценат
- валидни проценат
- кумулативни проценат

## Пример

## Frequency Table

Често купујем на распродаји

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid apsolutno se ne slazem se	16	9,9	9,9	9,9
ne slazem se	35	21,6	21,6	31,5
neutralan	56	34,6	34,6	66,0
slazem se	34	21,0	21,0	87,0
apsolutno se slazem	21	13,0	13,0	100,0
Total	162	100,0	100,0	

Где живите?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid svoja kuca	96	59,3	59,3	59,3
rentira	66	40,7	40,7	100,0
Total	162	100,0	100,0	

Каква је Вас радни статус?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid radi puno radno vreme	77	47,5	47,5	47,5
radi pola radnog vremena	40	24,7	24,7	72,2
penzioner/ne radi	45	27,8	27,8	100,0
Total	162	100,0	100,0	

# Пример са недостајућим вредностима

Da li prolazite pored marketa1 i marketa2 kada idete/vracate se sa posla?

N	Valid	133
	Missing	29

**Da li prolazite pored marketa1 i marketa2 kada idete/vracate se sa posla?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ne	25	15,4	18,8	18,8
	da	108	66,7	81,2	100,0
	Total	133	82,1	100,0	
Missing	System	29	17,9		
Total		162	100,0		

# Descriptives

- *Variable* - променљива(е) непрекидног типа
- *Options* - избор статистике централне тенденције, дисперзије, облика расподеле и редоследа приказивања резултата
- *Save standardized values as variables* - добија се нова променљива са стандардизованим вредностима

## Explore

- информације о расподели података
- робусне оцене централне тенденције појава
- тестове о испуњености претпоставки основних статистичких техника анализе података

### Статистичке мере:

- *Descriptives*: аритметичка средина, интервали поверења за аритметичку средину, медијана, варијанса,...
- *M – estimators*: робусне статистике максималне веродостојности
- *Percentiles*: 5%, 10%, 25%, 50%, 75%, 90%, 95%
- *Extreme Values*: 5 највећих и 5 најмањих вредности променљиве

# Садржај

- 1 *Анализа података*
  - Припремање података за анализу
  - Прелиминарна анализа података
  
- 2 *Примери*

# Пример 1

Применом одговарајућих процедура одговорити на следећа питања:

- a) У којој категорији посла има највише (односно најмање) запослених.
- b) Приказати просечно радно искуство запослених за сваку категорију посла посебно.
- c) Колико процената запослених ради у прве две категорије посла (приказати посебно за мушкарце и жене)?

Користити базу са претходног часа (*baza.sav*).



a) Analyze  $\Rightarrow$  Descriptive Statistics  $\Rightarrow$  Frequencies за  
променљиву kat\_posla

- a) Analyze⇒Descriptive Statistics⇒Frequencies за променљиву kat\_posla
- b) Поделити датотеку према категоријама посла:  
Data⇒Split File, а затим применити процедуру Analyze⇒Desriptive Statistics⇒Descriptive за променљиву r\_staz

- a) Analyze⇒Descriptive Statistics⇒Frequencies за променљиву kat\_posla
- b) Поделити датотеку према категоријама посла: Data⇒Split File, а затим применити процедуру Analyze⇒Desriptive Statistics⇒Descriptive за променљиву r\_staz
- c) Поделити датотеку према полу: Data⇒Split File и приказати табелу фреквенција за променљиву kat\_posla

## Пример 2

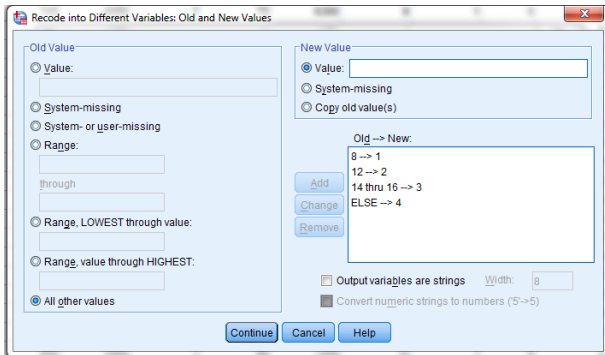
Приказати *Box plot* за променљиву *splata* по категоријама запослених: 1 - завршена основна школа; 2 - завршена средња школа; 3 - завршена виша, висока школа или факултет; 4 - мастер, специјалистичке и докторске студије.

Из табеле фреквенција за променљиву образовање виде се вредности које се појављују у оквиру ове променљиве (*Analyze*  $\Rightarrow$  *Descriptive Statistics*  $\Rightarrow$  *Frequencies*).

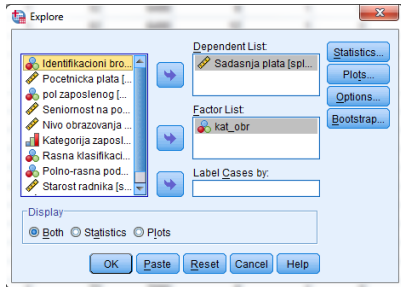
Nivo obrazovanja

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	8	53	11,2	11,2	11,2
	12	190	40,1	40,1	51,3
	14	6	1,3	1,3	52,5
	15	116	24,5	24,5	77,0
	16	59	12,4	12,4	89,5
	17	11	2,3	2,3	91,8
	18	9	1,9	1,9	93,7
	19	27	5,7	5,7	99,4
	20	2	,4	,4	99,8
	21	1	,2	,2	100,0
Total		474	100,0	100,0	

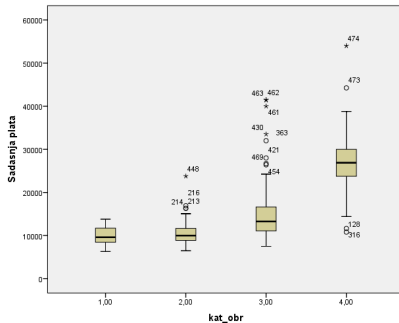
*Transform* ⇒ *Recode into Different Variables* -  
формирање нове променљиве која указује на 4  
категије образовања.



# Analyze ⇒ Descriptive Statistics ⇒ Explore



Sadasnja plata



# Хвала на пажњи!