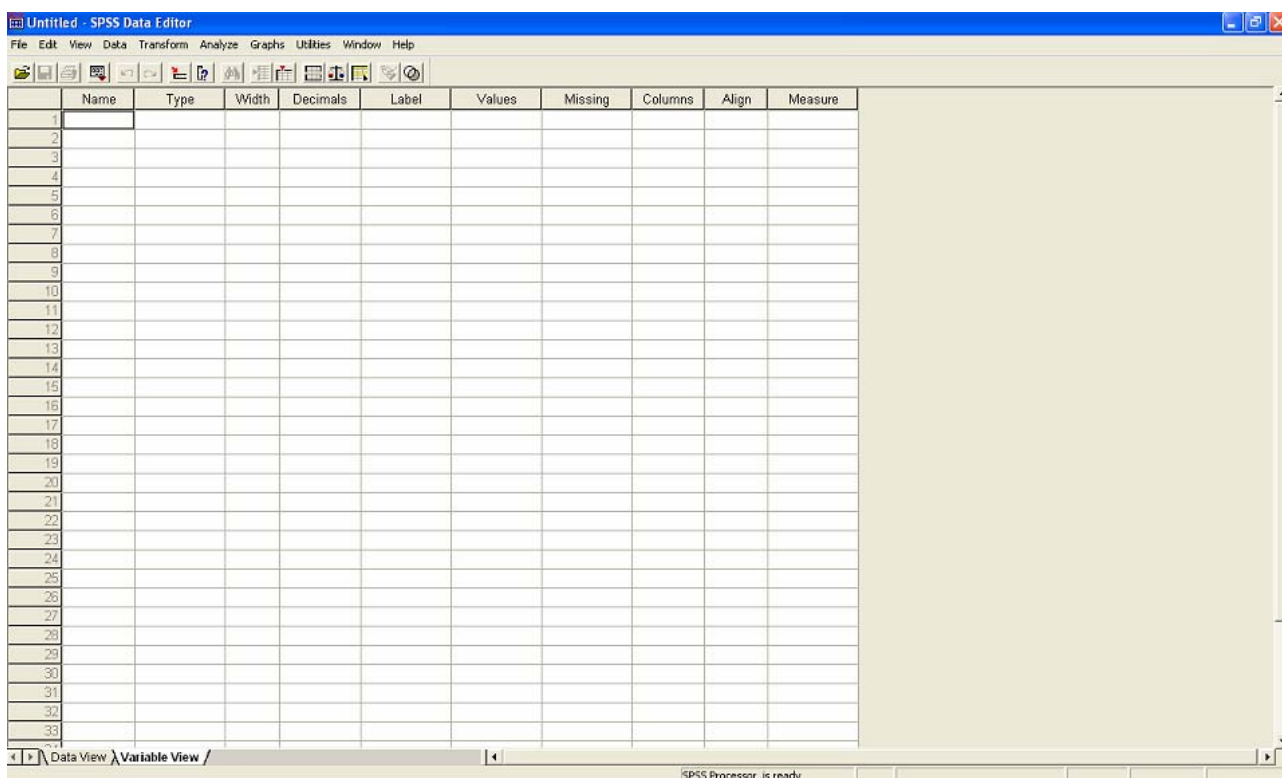


Първи стъпки в SPSS

Statistical Package for Social Science (SPSS – статистически пакет за социални науки) е компютърна програма, работеща в средата на операционна система Windows, която е специализирана за систематизиране, обработка и анализ на статистическа информация. Както подсказва наименованието, тя е насочена предимно към изследвания в областта на социологическите и маркетинговите проучвания.

Стартираме програмата SPSS от операционната система Window по следния начин: Start → Programs → SPSS for Windows → SPSS 13.0 for Windows. След отварянето виждаме прозореца SPSS Data Edition, който изглежда така:

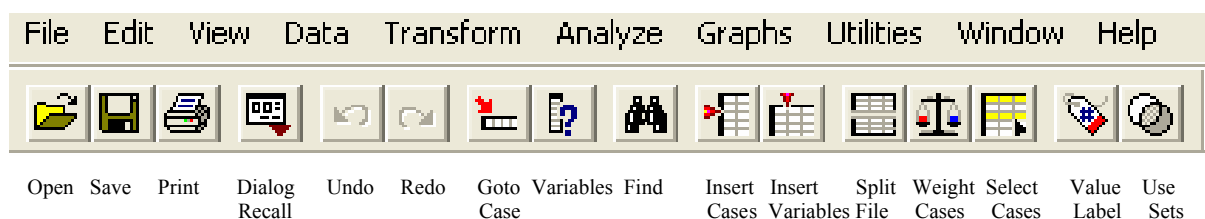


фигура 1

SPSS работи с два прозореца **Data Editor** - за въвеждане и редактиране на данни и **Output Viewer** - за отпечатване на резултатите от статистическата обработката - таблици и графики. Като за да превключваме между двата

прозореца използваме табулаторите в долната лента на екрана. В прозореца за данни (SPSS Data Editor) се дефинират, въвеждат и редактират статистическите данни. Output Viewer не е видим още, защото той се активира автоматично след осъществяване на статистическа обработка. В него се извеждат изчислените статистически показатели или графични изображения.

Менюто и на двата прозореца е еднакво, но иконите са различни. Менюто на SPSS съдържа някои от основните за Windows програмите подменюта, както и някои специализирани функции. Така изглежда то в прозореца Data Editor:



фигура 2

За по-добро и нагледно описание на клъстеризиращите процедури, където е нужно в настоящата глава, ще си служим със следния пример:

Пример: В Таблица 1 са представени основните показатели на UniCredit Bulbank. Всяка година ще бъде разглеждана като отделна единица, а показателите като отделни величини. Матрицата на данните, която е представена в таблицата е с размерност 7×5 .

Таблица 1

Основни показатели (в милиони лева) за периода 2000-2006 година на UniCredit Bulbank

№	Година	Нетна печалба	Собствен капитал	Активи	Клиентски депозити	Кредити
		X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}
1	2000	160,065	602,776	2559,476	1692,270	316,380
2	2001	68,912	490,479	2731,686	2021,634	362,353
3	2002	79,130	513,249	2721,980	2049,957	563,935
4	2003	89,752	550,026	2825,439	2177,781	916,634
5	2004	86,279	574,112	3614,696	2917,169	1393,968
6	2005	96,116	609,609	3474,829	2618,771	1706,858
7	2006	120,654	630,781	4346,594	3336,875	2131,577

Ако отделните променливи величини са изразени с различни мерки и мащаби, те трябва да бъдат стандартизирани. Например, когато в клъстерния анализ взимат участие променливи от различен тип – интервална и номинална, то за да ги използваме следва да направим стандартизация, привеждаща ги в един тип – интервален. Стандартизацията в SPSS може да се извърши автоматично, като се използват различни методи. Най-разпространения метод за стандартизация на данни е *Z scores*, където всички променливи преминават в един диапазон от -3 до +3 и после се преобразяват в интервални. При *Z*-метода стандартизираните стойности имат нулева средна и стандартно отклонение, равно на единица. Други начини са ранжиране в интервала от -1 до 1 или в интервала от 0 до 1, както и стандартизиране по отношение на максималната стойност, средната величина или стандартното отклонение. Ако променливите величини са изразени с еднакви мерки и мащаби не се препоръчва стандартизацията им.

В статистиката е прието променливите (признаците), чиито стойности се задават с думи, да се наричат **качествени** (неметрирани). Такъв тип са повечето променливи при анкетите, защото те са въпроси, на които отговорите се дават вербално. Другият тип променливи са **количествените** (метрираните). При тях стойностите се задават с числа. За предпочитане е да се работи с количествени променливи, защото това улеснява статистическата обработка и анализа. Затова стремежът е качествените признаци изкуствено да се превръщат в количествени. Например променливата „пол“, която е качествена, изкуствено се превръща в количествена, като мъжете се кодират с 1, а жените с 2. По този начин се подготвят анкети за статистическа обработка и анализ.

Въвеждането на данните в SPSS става като се превключи на изглед за променливите **Variable View** (фигура 3), където основно трябва да се запишат името, типът и етикетът (пояснението) на променливата. Попълнените имена

на променливите се появяват в първия ред на таблицата, при изглед за данните **Data View** (фигура 4), където преди това във всяка клетка е записано **var**.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	нетна_печалба	Numeric	8	3	Нетна печалба	None	None	15	Right	Scale
2	собствен_капитал	Numeric	8	3	Собствен капитал	None	None	15	Right	Scale
3	активи	Numeric	8	3	Активи	None	None	15	Right	Scale
4	клиентски_депозити	Numeric	8	3	Клиентски депозити	None	None	15	Right	Scale
5	кредити	Numeric	8	3	Кредити	None	None	15	Right	Scale
6										
7										

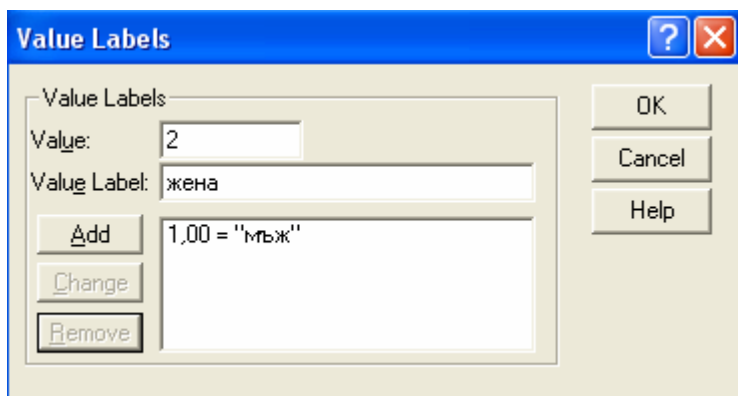
фигура 3

	нетна печалба	собствен капитал	активи	клиентски депозити	кредити	var
1	160,065	602,776	2559,476	1692,270	316,380	
2	68,912	490,479	2731,686	2021,634	362,353	
3	79,130	513,249	2721,980	2049,957	563,935	
4	89,752	550,026	2825,439	2177,781	916,634	
5	86,279	574,112	3614,696	2917,169	1393,968	
6	96,116	609,609	3474,829	2618,771	1706,858	
7	120,654	630,781	4346,594	3336,875	2131,577	
8						
9						
10						


фигура 4

В полето **Name**, при изглед за променливите **Variable View**, се изписва името на променливата накратко, като то трябва да бъде уникално, т.е. не може две променливи да имат едно и също име. След това се отива с мишката върху **Type** и се маркира нейния тип. В нашия случай е **Numeric**. Модифицирането на основния тип на променливата, когато има въведени данни в нея може да доведе до загубата им. Например променлива, която е само от числа, дефинирана като **String** (тя е идентификатор и за нея не се прави анализ) и превърната в **Numeric** изгубва наличните до момента на промяната стойности. Също така може да се промени ширината на колоната чрез **Width** и десетичните знаци след запетаята посредством **Decimal Places**. Полето **Label** служи за въвеждане на етикета към променливата, т.е. нейното разширено наименование, а **Values** е за етикет на стойностите на променливата.

Когато се превръща количествена променлива в качествена, кодовете и етикетите на категориите променливи се задават в колона **Values**. След като се кликне върху сивия бутон в дясната част на колоната се появява диалоговия прозорец **Value Labels**. В полето Value се въвежда кодът, с който е означена категорията на признака, например 1, а във Value Label съответния му етикет се изписва текстово - „мъж”. С натискане на Add се присвоява на 1=„мъж” и тя се записва в полето. След това се повтаря и за жените - във Value се въвежда 2, а във Value Label „жена” и отново Add. Процедурата се повтаря толкова пъти, колкото категории има признакът, след което се потвърждава с ОК.



фигура 5

Когато се въвеждат стойностите при количествените променливи, във всяка клетка се вписва съответната числова стойност, а при качествените – съответния код. Например за качествения признак пол, ако първият анкетиран е жена се вписва код 2. Това може да стане и по друг по-лесен начин – активизира се от менюто прозорчето с нарисован етикет . Отива се в съответната клетка от качествения признак и от падащия списък с възможните определения на признака – в случая „мъж” и „жена” се прави избор. По същия начин се постъпва за всички качествени променливи. Особеното е, че при количествените променливи не се дефинират етикети, защото няма кодове.

Посредством **Missing Values** се определят някои стойности като липсващи, включването им в други и отделно анализ само на тях. Използва се при кодиране на въпроси с отговор като "не знам", които не би трябвало да се загубят, а при анализа не са необходими. Ширината на колоните се определя в

Columns, подравняването от **Align**, а с **Measure** се дефинира скалата на променливите. Числовите са интервални/скалови (scale), а символните - номинални (nominal). Възможно е да се определят някои числа, които са кодове за номинални, а други за които наредбата на стойностите има значение за наредени/ординални (ordinal). След като се завърши дефинирането на променливата тя се появява в първата клона на таблицата. По същия начин се дефинират и останалите променливи.

Състави: Десислава Войникова