Appendix F

How to Use a Data Spreadsheet

Excel

ne does not necessarily have special statistical software to perform statistical analyses. Microsoft Office Excel can be used to run statistical procedures. Although in some respects Excel is not as preferable for data analyses as IBM SPSS, it is very user-friendly with simpler statistical procedures. This appendix describes how to use Excel to execute basic statistical calculations. Data from the 2004 version of the General Social Survey (GSS) is used for examples. This Appendix is based on Excel 2007 version, which differs in certain aspects from Excel's previous versions. The most notable change that affects the exercises presented in this appendix concerns the pivot table feature.

2Basic Procedures

Starting Excel

To start Excel using Windows, click on the **Start** button at the bottom left corner of the screen. Under **Programs** locate and click the **Microsoft Excel** icon.

The layout of the Excel program has changed substantially for the Microsoft Office 2007 edition compared to its predecessors. Commands are now grouped in ribbons that are accessed by clicking on a

specific tab. Thus, the **Home** tab grants access to a ribbon of several command groups: **Clipboard**, **Font**, **Alignment**, **Number**, **Styles**, **Cells**, and **Editing**.

Once the program is started you will see a Worksheet Area that consists of cells forming columns and rows. Rows are identified by numbers, and columns are identified by letters. Consequently, each cell has its own unique address – a combination of letters and numbers. For example, cell C6 is in column C, row6. The dark rim around a cell means that the cell is highlighted or active. You can highlight a range of cells by clicking and dragging the cursor across several adjacent cells.

At the bottom left of the screen you will find worksheet tabs labeled Sheet 1, Sheet 2, etc. You can rename these worksheets, add additional ones or delete ones you do not need.

Opening data

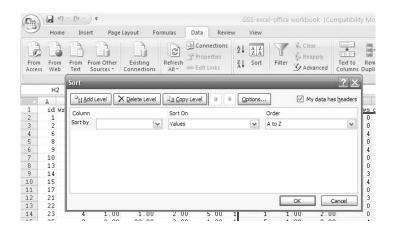
The data you need to use might be saved in a format other than Excel (file.xls). The data file we will be using in this tutorial is saved in SPSS format (file.sav). This is not a problem. Open the data file in SPSS. Click **File** – **Save As** – choose *Excel 97 and later (.xls)* in the **Save as Type** drop down menu. The file is ready to be open in Excel. To do so – either double-click on the icon of the saved Excel file or, if Excel program is already open, click **Office** button at the top left of the screen - **Open** – locate the data file you need to work with.

Looking at the Data

Once the data is open you will discover that the view is somewhat different than what you saw in SPSS program. Variable names are displayed in the first row (row A). Unfortunately, there is no variable view worksheet in Excel, and transferring the file from SPSS to Excel results in a loss of variable labels and value labels. Keep the codebook for the data on hand!

Sorting Data

As you are exploring the data you might want to take advantage of the Sorting tool, which allows you to sort data by two or more variables in ascending or descending order. To sort your data, click the **Data** tab, find the **Sort and Filter** command group, and click on the **Sort** icon. A dialog window will open where you will specify the variables and order you want your data to be sorted by. If the names of variables (variable labels) are entered in the first row of your data make sure you check the "*My data has headers*" box at the top right of the dialog window. You can sort numeric data as well as text (In A to Z order). If you want to sort by more than one variable click **Add Level** button at the top of the dialog window. Click **OK**.



Missing Data

If you have converted an SPSS file to be used in Excel, be mindful of how missing data was coded in the original file. If missing data was entered as blanks, you have nothing to worry about. However, if missing data in the original data file was entered as a certain numerical expression (e.g. zeroes or 99) this might create problems for running certain statistical calculations in Excel (e.g. computing mean).

Loading Data Analysis Toolpak

Before we start with running some basic analysis make sure that Excel's Data Analysis Toolpak is loaded on your computer when you start the program. The Data Analysis Toolpak comes as an Add-In to your Excel program and is found on the installation disk. To see if the Toolpak is loaded in Excel 2007 – check if there is an Analysis icon under the Data tab.

Data Rei	iew View												0
Connection Properties Se Edit Links	⁵ A ↓ A Z A Z ↓ Sort	Filter	Clear	Text to Columns	Remove Duplicates	Data Validation	Consolidate	What-If Analysis *	Group	Ungroup	Subtotal	🕒 Data A	nalysis
onnections		Sort & Fi	lter			Data Too	5			Outline	Γ _M	Analy	sis
F	G	Н	1	J	К	L	M	N	0	Р	Q	R	S

If you do not see one click **Office** button – **Excel Options** – **Add-Ins**. In the dialog window that opens check if Analysis Toolpak and Analysis Toolpak-VBA are listed under *Active Application Add-ins*.

cel Options				?
Popular	View and manage Microsoft Office	add-ins.		
Formulas	in the second se			
Proofing	Add-ins			
Save	Name	Location	Туре	^
Advanced	Active Application Add-ins			
	Acrobat PDFMaker Office COM Addin	C:\0\PDFMaker\Office\PDFMOfficeAddin.dll	COM Add-in	
Customize	Analysis ToolPak Analysis ToolPak - VBA	C:\fice\Office12\Library\Analysis\ANALYS32.XLL	Excel Add-in Excel Add-in	
Add-Ins	Analysis ToolPak - VBA	C:\\Office12\Library\Analysis\ATPVBAEN.XLAM	Excel Add-In	
Aud-ms	Inactive Application Add-ins			
Trust Center	Conditional Sum Wizard	C:\rosoft Office\Office12\Library\SUMIF.XLAM	Excel Add-in	
Resources	Custom XML Data	C:\iles\Microsoft Office\Office12\OFFRHD.DLL	Document Inspector	
Resources	Date (Smart tag lists)	C:\iles\Microsoft Shared\Smart Tag\MOFL.DLL	Smart Tag	
	Euro Currency Tools	C:\ft Office\Office12\Library\EUROTOOL.XLAM	Excel Add-in Smart Tag	
	Financial Symbol (Smart tag lists) Headers and Footers	C:\iles\Microsoft Shared\Smart Tag\MOFL.DLL C:\iles\Microsoft Office\Office12\OFFRHD.DLL	Document Inspector	
	Hidden Rows and Columns	C:\iles\Microsoft Office\Office12\OFFRHD.DLL	Document Inspector	
	Hidden Worksheets	C:\iles\Microsoft Office\Office12\OFFRHD.DLL	Document Inspector	
	Internet Assistant VBA	C:\icrosoft Office\Office12\Library\HTML.XLAM	Excel Add-in	
	Invisible Content	C:\iles\Microsoft Office\Office12\OFFRHD.DLL	Document Inspector	
	Lookup Wizard	C:\soft Office\Office12\Library\LOOKUP.XLAM	Excel Add-in	
	Person Name (Outlook e-mail recipients)	C:\es\Microsoft Shared\Smart Tag\FNAME.DLL	Smart Tag	
	Solver Add-in	C:\fice\Office12\Library\SOLVER\SOLVER.XLAM	Excel Add-in	
	Document Related Add-ins			۵
	No Document Related Add-Ins			
				~
	Add-in: Acrobat PDFMaker Office CO	MAddin		
	Publisher: Adobe Systems, Incorporated			
	Location: C:\Program Files\Adobe\Adob	e Acrobat 7.0\PDFMaker\Office\PDFMOfficeAddin.dll		
	Description: Acrobat PDFMaker Office CO	MAddin		
	Manage: Excel Add-ins	0		_

	Add-Ins ?	Clear
f Existi Connec	Add-Ins available:	Filter
	Analysis ToolPak Provides data analysis tools for statistical and engineering analysis	

If they are listed as inactive select both of them and click **GO** at the bottom of the window. Next, check the boxes for both Add-Ins in the dialog box that opens. Click **OK**.

■2Univariate Statistics

Frequencies

PivotTable I		
a invocidute i	Field	List 👻
		dd to report:
Choose field	s to a	dd to report:
satjob		
satfin		
maeduc race		
pres00		
relig		
attend		
absingle		Move Up
satself		Move Down
moregoo		Move to Beginning
rincom98		
		Move to <u>E</u> nd
Drag fields b	Y	Move to <u>End</u> Move to Report Filter
Drag fields b	Y	Move to <u>End</u> Move to Report Filter Move to Row Labels
Drag fields b	V	Move to End Move to Report Filter Move to Row Labels Move to Column Label
Drag fields b	Σ 2	Move to End Move to Report Filter Move to Row Labels Move to Column Label Move to Values
Drag fields b	γ Ξ Ξ Σ Χ	Move to End Move to Report Filter Move to Row Labels Move to Column Label Move to Values Remove Field
Drag fields b	γ Ξ Ξ Σ Χ	Move to End Move to Report Filter Move to Row Labels Move to Column Label Move to Values

The easiest way to create frequency distributions in Excel is by using the Pivot Table. To create a frequency distribution of the variable ABRAPE (pregnant as a result of rape) click on any cell in your data. Next click **Insert Tab** - **Pivot Table**. In the opened dialog window you will see a *Table/Range specified* that includes the call range for all of your data. It is also possible to specify the range manually by typing A1:AW1500. Next, lace your Pivot Table: Click **New Worksheet** - **OK**.

A new worksheet will open with newly created Pivot Table. In a **Pivot Table Field List** on the right side choose and click the variable we are interested in – ABRAPE. Do not be alarmed that your newly created Pivot table displays the sum oall the ABRAPE values in the data. Since you are interested in the count or each of the values of the variable and not the sum of the values, click a dropdown menu located under the **Pivot Table Field List** and make sure that the **Fields Section and Areas Section Stacked** is selected. Next click on the ABRAPvariable and drag it down to the **Row Labels** box. Next click the dropdown menu to the rigof Sum of Ab **ettings** – change **Sum** to **Count** - **OK**. Now you have tions for the variable ABRAPE. Unfortunately, Excel does not dispvalue labels like SPSS does, so your table can look a little bit confusing with numbers in place of the actual values of the es"). To rectify this problem, you can either correct the codebook for the data or recode the variable prior to creatinga Pivot table (substituting numerical values with text).

Excel does not offer an option to simultaneously display both the count and percentage within the same Pivot table. However, you can view the percentage distriby clicking **Value Field Settings – Show values as –** choose **% of total**.

Descriptive Statistics

The easiest way to obtain descriptive measures on a variable in Excel is by using the Data Analysis Tool. Under the **Data** tab click **Data Analysis** - **Descriptive Statistics** - **OK**. In the descriptive statistics dialog window specify the range of the variable you are interested in. Variable EDUC occupies the range O1: O1500. You can either specify the range manually by typing it into the *Input Range* Dialog box or highlight the column that the variable occupies in the dataset.

Excel offers the option of grouping each variable in its own column or its own row, with former being the default option. Click *Labels in First Row* option or you are risking receiving an error message "Input range contains nonnumeric data" since the program will get confused with the name of the variable being the fist cell in the column.

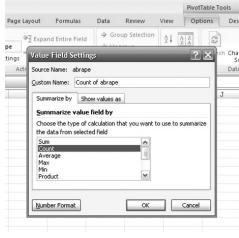
Next you can choose to either place the output into a new worksheet (default option), new workbook, or in the same worksheet as the data (Output Range). In the later case you have to specify the upper left cell "address" where you want your output to be pasted. Click the *Summary Statistics* box - **OK**. You will be presented with information on a range of statistical measures: mean, standard error, median, mode, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, sum, count.

If you are interested in obtaining only a certain descriptive statistic you can take advantage of Excel's functions instead of using the Descriptive Statistics tool. For example, let's calculate the Mode of the EDUC variable. Either below or to the side of you data find some empty cells. In one of the cells type EDUC Mode. Then activate the cell immediately to the right, click **Formulas - Insert Function**. In the dialog window specify *Statistical* and then choose **MODE - OK**. In the

Function Arguments window specify the range of the variable of interest (EDUC occupies O2:O1500). Click **OK.**

1.12	A	В	
1	Frequency distri	bution	
2	Count of abrape		
3	abrape 🔭	Total	
4	YES	365	
5	NO	102	
6	N/A	29	
7	NAP	1003	
8	Grand Total	1499	
9			
10			
11	Count of abrape		
12	abrape 📑	Total	
13	YES	24.35%	
14	NO	6.80%	
15	N/A	1.93%	
16	NAP	66.91%	
17	Grand Total	100.00%	
18			

	Tables	Illustrations						
	A6	~ (3	f _x	1			
2	A		В		С			
1	Drop Page	Fields	Here					
2								
3	Count of al	brape	-					
4	abrape	-	Total					
5		0	1003					
6		1	365					
7		2	102					
8		9	29					
9	Grand Tota	1	1499					
10								



Re	view	View				
ties nks		$\begin{array}{c} A \downarrow \\ Z \downarrow \\ Z \downarrow \\ A \downarrow \\ A \downarrow \\ S ort \end{array}$	Filter Advanced	cordinit's bupileares in		at-If ysis -
	0	R	Descriptive Statistics		? ×	Z
ICT		vote00	Descriptive statistics	ł		onlab
00	52	2	Input		OK	onitop
00	43	1	Input Range:	\$0\$1:\$0\$1500		
00	33		e		Cancel	
00	22	2	Grouped By:	⊙ <u>C</u> olumns		
00	19	3		○ <u>R</u> ows	Help	
00	26	1 2 3 1 2	Labels in First Row		6	
00	32	2			6	
00	40	1	Output options			
00	42		O Output Range:		D D	
00	39	1 2 1			6	
00	29	1	New Worksheet Ply:	EDUC descriptives		
00	67		O New Workbook		D	
00	57	1 2 3 1	Summary statistics		b	
00	19	3			в	
00	81	1	Confidence Level for	Mean: 95 %	2	
00	35	2	Kth Largest:	1	B	
00	47	4	Kth Smallest:	1	b	
00	58	1			þ	
00	29	1			þ	
00	48	1	2 2	3 3	2 3 3	
00	23	2	2 2	3 2	2 3 2	

	A1	• (0)	
	A	В	С
1	ed ed	luc	
1 2 3			
3	Mean	13.71047	
4	Standard	0.091856	
5	Median	13	
6	Mode	12	
- 7	Standard	3.556369	
8	Sample Va	12.64776	
9	Kurtosis	220.2888	
10	Skewness	9.109408	
11	Range	99	
12	Minimum	0	
13	Maximum	99	
14	Sum	20552	
15	Count	1499	
16		· · · · ·	
17			

Graphing

There are two ways to create graphical representation of your data in Excel: using the Pivot Table or the Histogram Analysis Tool. First, let's try using the Pivot Table option. Follow the steps to create a frequency distribution Pivot Table for the variable ABRAPE. It is very easy to create a bar chart from here. Click on any cell in the Pivot Table, then click **Insert tab** - **column** - **2D column**. You can format your chart using a **Chart Styles** section under the **Design Tab**.

Now let's practice creating a histogram using the Histogram Analysis Tool. In order to create a histogram we have to specify "bin values" that represent the entire range of values of your variable. By specifying bins we are creating slots to gauge how many times a specific value appears in our data. Our histogram is supposed to convey the age distribution of our respondents. First, create a frequency distribution of the respondent's age variable. You will see that our dataset the youngest respondent is 18 and the oldest is 89. Let's specify eight bins: 28, 38, 48, 58, 68, 78, and 89 with each bin being the upper limit for a particular age group. Insert a new column in you data, title it AGEBINS and enter the specified numbers.

	C150	5 🗸 🌘	×	✓ f _x =									
1	A	В		С	D		E	F	G	H	I	J	K
1493	2801		1	1.00	2	. 00	2.00	4.00	1	1	1.00	2.00	
1494	2802		5	1.00	1	.00	3.00	6.00	2	4	1.00	1.00	
1495	2803		2	9.00	99	.00	3.00	3.00	2	1	1.00	1.00	
496	2804		2	1.00	1	.00	3.00	3.00	2	3	1.00	2.00	
1497	2805		1	1.00	1	.00	2 00	2 00	2	3	1 00	2 00	_
1498	2807		2	2.00	4	Inser	t Function					?	X
1499	2811		7	9.00	99	moet	e i uniscioni		-		_	لنف	200
1500	2812		1	2.00	4	Searc	h for a function	n:					
1501						Tve	pe a brief desci	intion of w	nat vou	want to do a	nd then click	Go	
1502						Go		ipcion or m	lac you	Marie Co do a	nd then thek	Go	
1503								S			-	1	
1504						Ors	elect a <u>c</u> atego	ry: Statisti	cal		~		
1505		EDIIC Mode	-			Select	a function:						
1506				Ī		19221							
1507						MO	GBINOMDIST						
1508							RMDIST						
1509							RMINV						_
1510							RMSDIST						
1511							RMSINV ARSON					ſ	~
1512						- Andrews							
1513						1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	DE(number1						
1514						data		requently o	occurring	, or repetitiv	e, vaiue in an	array or range	or
1515						uau							
1516													
1517													
1518													_
1519						Help	on this function				OK	Cancel	
1520												10 T	_

Click Data - Data Analysis - Histogram - OK.

ta Review	v View						
Connections Properties dit Links ctions	A*	K Clear Reapply Iter Advanced & Filter	Text to Columns D	Remove Ouplicates	Data Validation + Data Tools	Consolidat	e What-If Analysis *
0	R	S	Т	U	V	ប	X
age	AGEBINS			-	confinan		
52	28		2 0	2	0	0	0
43	38		1 0	2	0	0	0
33	48		1 0	2	0	0	0
22	58	6					0
19	68	Data Analysis					<u> 2 × 2</u>
26	78	Analysis Tools					0
32	89	Histogram			0		0
40		Moving Average				and a second sec	ncel 2
42		Random Number Ge					0
39		Rank and Percentile Regression	9			н	elp 0
29		Sampling			ſ		2
67		t-Test: Paired Two					0
57		t-Test: Two-Sample			ices		0
19		t-Test: Two-Sample z-Test: Two Sample		Jnequal Var	iances	~	2
81		2-rest; two sample	or means		U		1 2
35		C.	2 2	2	3	3	
47			4 0	2	0	0	0

In the dialog window specify the *Input Range* (the range of the variable you are interested in) and the *Bin Range* (cell range where the bins values are specified). Make sure that the *Labels* box is checked. Choose the location of your output histogram (*New Worksheet Ply* is default). Check *Chart Output*. Click **OK**.

C Be	Home	Insert P	age Layout	Formulas	Data Revi	ew View	Design	Layout	Format		
Fro	ess Web Te	om From Otherst Sources	- Connection		Connections Properties See Edit Links Connections	$\begin{array}{c} A \downarrow \\ Z \downarrow \\ A \downarrow \\ A \downarrow \\ A \downarrow \\ Sort \end{array}$	Filter	< Clear Reapply Advanced	Text to Columns	Remove Duplicates	Data Validatio Data To
	Chart 1	+ ()	f_{x}								
1	A	В	С	D	E	F	G	Н	I	J	
1	AGEBINS I	Frequency	mulative %			P-		-9			1
2	28	258	17.21%			- F	listogra	am			
3	38	286	36.29%			6		0			
4	48	312	57.10%		× 400 T	_					
5	58	273	75.32%	:	200						:
6	68	201	88.73%	:	e 0 +						
7	78	95	95.06%		28	3 38 48	58 68	78 89 1	More	Frequency	/
8	89	72	99.87%		L 20	5 36 46	58 08	/8 85 1	NOIE		
9	More	2	100.00%				AGEBINS				
10											1
11											
12											

Your histogram will open in a new worksheet. It does not look perfect, but you can modify it to make it look better. First, make the histogram taller by clicking on it and dragging its lower border down. Second, change the axis titles and the title of the histogram. Third, you can remove the space between the bars: right-click on one of the bars and select **Format Data Series**, then drag the *Gap Width* cursor all the way to the left. The possibilities of formatting a chart in Excel are endless.

■2Recoding Variables

Recoding EDUC (highest level of education completed) variable. To recode EDUC into a new categorical variable you have to create a reference table that lists your new categories. Find a range of empty cells below your data and create a reference table where you specify:

Those who have 0-11 years of education are put in category 1

Those who have exactly 12 years of education put in category 2

Those who have 13 to 15 years of education are put in category 3

Those who have 16 or more years of education are put into category 4.

	Tables	Illustrations			
	B1532 -	(fx			
1	A	В	С	D	E
1499	2811	7	9.00	99.00	1.0
1500	2812	1	2.00	4.00	3.0
1501					
1502	Yrs of Educ	Category			
1503	0	1			
1504	12	2			
1505	13	3			
1506	16	4			
1507					
1508					
1509					
1510					

Next insert a row next to the EDUC variable – this is where you will compute your new variable - ED4CAT. Highlight the cell right under the variable label. Click: **Formulas** - **Insert Functions**. Select *Lookup & Reference category* and choose **VLOOKUP** function in the menu below. Click **OK**.

More Functions *	Define Name * Manager B Create from Selection Defined Names	न्द्री Trace Dependents २ २२ Remove Arrows २ (hecking * e Formula	Watch Window		ition 🛅 c	alculate No alculate Shi		
	Defined Names	Form		ng			Calcula	uon	_	_
Insert	Function		?×	N		0	P	0	R	
	for a function:			lincome		-	ED4CAT	educr		vo
5.				1	19	9 =		2.00	52	
	e a brief description of what you want to d	o and then click Go		1	5	14		4.00	43	
GO				1	12	16		5.00	33	
2. Or se	Or select a <u>c</u> ategory: Lookup & Reference 🛛 🗸					14		4.00	22	
Select a function:					23	12		3.00	19	
2. Select a	a runctio <u>n</u> :				20	14		4.00	26	
MAT			<u>~</u>	1	13	10		2.00	32	
. OFF:	OFFSET				21	18		6.00	40	
. ROW					15	14		4.00	42	
B. RTD					99	9		2.00	39	
	NSPOSE				99	13		4.00	29	
	OKUP				20	14		4.00	67	
)KUP(lookup_value,table_array,col_				10	12		3.00	57	
	for a value in the leftmost column of a tal				99	13		4.00	19	
· accor	row from a column you specify. By defaul nding order.	t, the table must be sorted i	- P		13	16		5.00	81	
B. Cascer			0		16	13		4.00	35	
1.1			- P		99	17		6.00	47	
5.(!		23	16		5.00	58	
	this function	OK Can			23	19		6.00	29	
			· · · · · · · · · · · · · · · · · · ·		10	16		5.00	48	
.00 2	1 2.00 2.00	3 1	2.00	J	11	13		4.00	23	

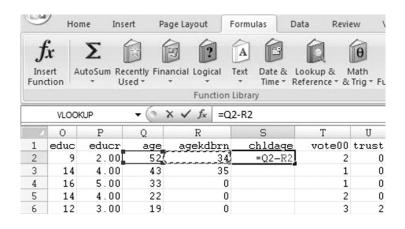
In the dialog window first specify the *Lookup_value*. The lookup value is the value of your original variable you want to lookup and replace with some value of a new variable. *Table_array* is the location of the reference table we created to lookup values for the new variable. You can either highlight the location of the reference table you're your mouse or type the cell range manually. *Col_index_num* refers to the position of the reference table column we want to lookup the new variable values from. In our case the column that contains all the potential values of the new variable is column number 2 (column B). Click **OK**.

(La)	Home Inset	Page Layout	Formulas Data	Review View			
<i>fx</i> Insert Function	AutoSum Recentl			up & Math More nce * & Trig * Functions *	Name Manager	Define Name * Ouse in Formula * Create from Selection tined Names	ई अञ्च Trace Precedents ⇔ई Trace Dependent ्रद्ध Remove Arrows Fr
	VLOOKUP -	● X ✓ Jx	=VLOOKUP(P2,A1503	B:B1506)			
X	A	В	(1 1		
1490	2795	1	Function Arguments				? × 1
1491	2798	1	VLOOKUP				
1492	2800	2		·			
1493	2801	1	Lookup_value	P2	- E	0	
1494	2802	5	Table array	A1503:B1506	ES =	{0,1;12,2;13,3;16,4}	
1495	2803						
1496	2804	2	Col_index_num		ES =	number	
1497	2805	1	Range_lookup		E65 =	logical	
1498	2807	2	-				
1499	2811	7			-		
1500	2812	1		tmost cclumn of a table, and ble must be sorted in an asce		value in the same row from	m a column you
1501			speciry, by derault, the tai	Die musc de sorteu in an asce	haing order.		
	Yrs of Educ	Category		Table_array is a table of I			
1503	0			Table_array	can be a refere	nce to a range or a range	name.
1504	12						
1505	13	3					
1506			Formula result =				
1507							
1508			Help on this function			OK	Cancel
1509							
1510							

This gives you your first value of the new variable ED4CAT. To avoid repeating the above procedure for each case in the dataset, simply copy the formula from the cell just completed to other cells downward by clicking on the right bottom corner and dragging the mouse cursor down.

Computing a New Variable

It's quite easy to compute a new variable in Excel. All it takes is recollection of a few basic algebra rules. For example, to determine the age of each respondent's eldest child we need to compute the difference between the respondents' age and the age of the respondent when his or her first child was born from the respondent's. Let's start by inserting a new column anywhere in the worksheet and typing the name of our new variable (CHLDAGE) in the first row. In the empty cell right below type a formula for calculating the age of the respondent's age (AGE) (Q2 in our example); and (2) respondent's age when his or her first child was born (AGEKDBRN) (R2). Our formula will be: = Q2 - R2. Press Enter, and there you have it – the first respondent's eldest child is 18 years old. Copy the formula for the respondents.



2Bivariate and Multivariate Statistics

Crosstabulation

Exploring the relationship between two or more variables in Excel is done with the help of pivot tables. You will not find a Crosstab option like in SPSS. Let's say we are interested in the relationship between individual's level of education and whether or not he or she participates in elections. We will create a pivot table for two variables: EDUCR3 is a trichotomized measure of education and variable VOTE00, which assesses whether the respondent participated in 2000 Presidential elections.

Click **Insert** X **Pivot Table**. Select data range and the location for the Crosstab table (new worksheet). Click **OK**. On the right side of the new worksheet, in the *Pivot Table Field List* area first select **Field Section and Area Section Side-By-Side** option for more convenient display, then select (check the boxes) the two variables of interest VOTE00 and EDUCR3. Place VOTE00, the dependent variable, in the *Row Labels* box, and EDUCR3 in the *Column Labels*. Drag VOTE00 into the *Values* box, and change the **Value Field Settings** to *Count* instead of *Sum*. Your crosstab will look a little confusing due to presence of numbers instead of actual variable values. If you have the codebook for you data on hand, you can quickly change this manually. For example, we know that a value of 1 is assigned to VOTE00 variable if the respondent voted in the 2000 Presidential elections. We can change the cell accordingly.

Excel does not offer an option to simultaneously display both the count and percentage within the same crosstab. You can observe percentages (of rows or columns) by clicking **Value Field Settings** – % of column - OK.

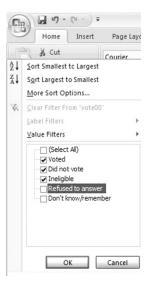
1	A	В			С		D]	Ξ
1		Dr	op Pa	ige Fiel	ds Here				
2									
3	Count of vote00	educr3	4						
4	vote00	Grade Sch	ool	High	School	Some			Total
5	Voted	34.	27%		63.68%	:	69.99%		63.22%
6	Did not vote	51.	64%		29.35%	:	22.42%		28.44%
7	Ineligible	12.	21%		5.72%	:	7.13%		7.48%
8	Refused to answer	0.	47%		0.00%	:	0.11%		0.13%
9	Don't know/remember	1.	41%		1.24%	:	0.34%		0.73%
10	Grand Total	100	00%		100.00%	:	100.00%	1	00.00%
11									
12									
13									
14									
15			_						
16	Count of vote00	educr3	4						
17	vote00	Grade Sch	ool	High	School	Some	College	Grand	Total
18	Voted		73		256		618		947
19	Did not vote		110		118		198		426
20	Ineligible		26		23		63		112
21	Refused to answer		1				1		2
22	Don't know/remember		3		5		3		11
23	Grand Total		213		402		883		1498

Since there were so few people that either refused to answer the question about voting or cannot recollect whether they voted you have an option of dropping these two rows from the table (this will, of course, change the grand total counts, so you might not want to do this). Click the little arrow next to the VOTE00 cell in the table and uncheck these two rows. Click **OK**.

Three-Variable Crosstabs

It is very easy to convert the crosstab you have created into a three-variable crosstab.

The two-variable cross-tab we have created demonstrated that individuals with higher levels of education (those who have completed high school and/or some college) were more actively involved in voting in the 2000 Presidential election than individuals who have completed grade



Σ AutoSum Fill ≁ Clear * Ec	Sort & Find & Filter * Select *
PivotTable Field List	•
Choose fields to add to	report:
Tid 🔊	Report Filter
wrkstat	a report the
incomr2	
incom98r	
educr3 7	
ager	
sex	
marital	Column Labels
□raced 🗉	educr3 •
vote00d	incom98r 🔻
polviews	
cappun	
trustd	
income98	Row Labels
educr	vote00
Vote00 V	
Trust	
region	
confinan	Σ Values
conbus	Count of vote00
conclerg	Count of Voteou +
coneduc	
confed	
conlabor	
conpress	
conmedic	D Update

school. But perhaps education is not ra moving force behind individuals' propensity to be politically active, and the relationship we observed is due to extraneous factor. To explore the relationship between voting and another variable whilcontrolling for education all we have to do is add a third variablthe pivot table (to the column labels).

Comparing Means

Comparing means in Excel once again requires the use of Pivot tables. Let's examine whether individuals differ in terms of years of education by race. Click:

Insert – **Pivot Table** – [make sure the cell range for your data is specified in the *Table/Range* box] – **OK**.

In the *Pivot Table Field List* area check and drag the variables you are interested in (EDUC and RACE) into *Columns* and *Row Labels* boxes respectively. With your dependent variable (EDUC) also being in the *Values* box. Click **Value Field Settings – Average – OK**. This will produce a pivot table that shows mean years of education for each group within the race variable. Let's type the value labels of the race variable instead of numbers.

To accurately assess the difference between the mean years of education of Whites, Blacks and those individuals who comprise the "Other" racial cat-

egory, we need to calculate standard deviations by racial groups. This is also done in the pivot table. Let's copy the pivot table that we just created somewhere on the worksheet you have open. Now, all you have to do to obtain standard deviations by group is click **Value Field Settings** and choose **StdDev** under the *Summarize by* tab. In the figure below we have copied and pasted the Standard Deviations column adjacent to the Means column.

-	Home Insert	Page	Layout	Formula	s Dat	ta	Review	1
	Cut	Courier		8 -	A A	= =	-	\$? ^-
Pas		B <i>I</i> <u>U</u> →			• <u>A</u> • ==		F 30	律者
	Clipboard 🕞		For	nt	5			Alig
	Y34 -	0	f_{x}					
1	A			U	U			
11								
12								
13	Average of ed	uc						
14	Race of Respond	lent -	Me	ean	St.D	ev		
15	White			13.77		2.79		
16	Black			13.02		2.57		
17	Other			13.63		3.08		
18								

■2Printing

To print portions of Excel worksheets click the **Office** button – **Print**. Before printing the **Page Setup** button under the **Page Layout** tab allows you to manipulate the orientation of the page (portrait or land-scape), the margins of the page, rows and columns you want to print, as well as printing the output with or without gridlines.