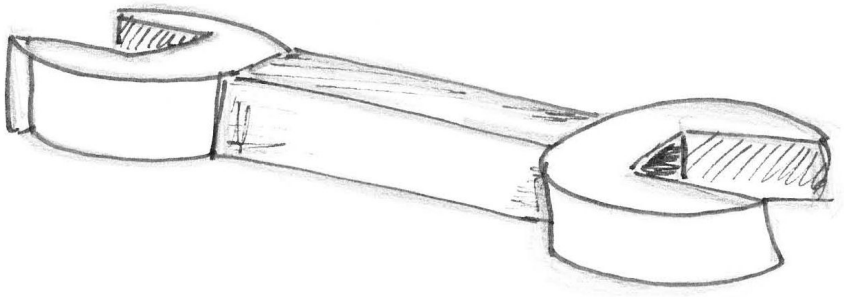
Chapter 9
Tools



Most of us who become experimental physicists do so for two reasons; we love the tools of physics because to us they have intrinsic beauty, and we dream of finding new secrets of nature as important and as exciting as those uncovered by our scientific heroes. But we walk a narrow path with pitfalls on either side. If we spend all our time developing equipment, we risk the appellation of 'plumber', and if we merely use the tools developed by others, we risk the censure of our peers for being parasitic.

Luis W. Alvarez

'Recent Developments in Particle Physics', Nobel Lecture, December 1th, 1968. In *Nobel Lectures: Physics 1963-1970* (1972), 24.

There are various tools which one can employ to help in statistical research. These range from simple tools that produce basic calculations (and are readily available in proprietary office packages) to specialized statistics tools that are developed solely for the purpose of analysing data. Another more interesting group of tools is available freeware and has the same functionality as the commercial packages. This Chapter will give an overview of the tools available and will give examples from some of the tools for ease of reference.

Prior to venturing out and buying or downloading/using online software one has to decide on what **Depth of Use** one is going to employ the tools.

A. Does one need basic tools that identify keywords and help to find links between the different inputted data? This is particular to qualitative methodology.
B. Does one need simple tools to aid in quantitative work such as summing, adding, subtracting, sorting by minimum or maximum and averaging as well as exporting in the forms of tables and graphs?
C. Does one need tools that employ the above but also allow for basic statistics, such a measures of central tendency, etc. (refer to Chapter 11)?
D. Does one need a full statistical package that employs the above as well as a whole range of statistical measures (refer to Chapter 11)?
E. Does one need to employ spatial statistical tools that deliver maps as outputs and use specialised statistical measures, inclusive of Moran's I, Standard Deviational Ellipse, Nearest Neighbour Hierarchical Analysis, etc. (refer to Chapter 11)?

An early decision will ensure that one is not sidetracked into wading through forest-equivalent paper trails or wrestling with a software tool that is as complex to manoeuvre through as it is to pilot the space shuttle. Choose wisely and the earlier the decision is taken the quicker the choice is made and the less tears to be shed!

Let us take 4 examples. Read them and try to establish which of the above would fit each example:

i) Analysing correlation of offender residential areas by poverty levels;
ii) Identifying the most common links highlighted during interviews with the elderly in a retirement age study;
iii) Attempting to establish if there is a relationship between youth leisure activities and drug use;
iv) A demographic project that aims to establish the central tendency for a number of attributes such as sex, age, citizenship and residential location; and
v) Identifying the oldest person in the electoral register.

The answers would appear as follows:

Example	Depth of Use
i	E
ii	A
iii	D
iv	C
v	B

Chose well as each choice will imply the acquisition and learning of new tools where not already acquired. This may take some time to establish so it is best to run the above check first.

Which Tools are Available?

There are many tools which one can use in statistical analysis. They are varied: some are linked to traditional approaches and do not require highly-detailed technology, others are based on multi-mode methods employing both manual and digital means, whilst others are based on high-end technology that

basically takes over an analysis function. There is no longer the need to think about how to go about running a statistical test as this is done through a simple drop-down menu command.

This progress in technological innovation is tantamount to a literal JANUS! It is both a blessing and a curse as one now can literally abuse of the test to be used. It is really easy to run a correlation test for a nominal-ordinal structure as it is easy to sum two numbers. The issue is no longer one of access to the statistical tool and it is to abuse of the tests that are relative to the variable type. Great care must be taken to ensure that the relative tests are chosen.

Once one has decided to go for a specific method of employment, one has to choose the best which fits the purpose. There are both manual and fully automatic tools.

Manual

A simple calculator or grey matter should do the trick! In terms of basic mathematical calculations there is no need for hi-end technology, A few soft taps on the simple stand-alone keyboard that is even available in mobile phones will help one to reach an answer. Or else do it the old fashioned way and count on one's fingers, sketch on a piece of paper or use an abacus. Simple – yet very effective!

One can also use a spreadsheet and input all the numbers in successive cells. The statistical measure is inputted manually.

Town	Population
Valletta	1000
Floriana	200
Qormi	300
Qala	500
Zebbug	800
Total	**2800**

Semi-Automated

Taking the spreadsheet concept a bit further, one can create a tool that carries out a number of functions in order to automate the process. These tools are called macros, where the researcher inputs the data in the individual cells and the macro runs simple statistical measures automatically.

Automated

Automated tools serve as the cherry on the statistical cake. The researcher structures the dataset and the relative variable, input or imports the data from his/her survey and runs the relative statistical test. Today's tools are a breeze compared to those employed a few years ago when one had to programme the tool and input the data within the tool, severely limiting the number of variables one could use as well as the number of interviews one could carry out.

Imagine a questionnaire with 10 questions and the researcher had 100 respondents. That equates to a cool 1,000 inputs! Imagine the probability of error generation during the input stage as a series, of 1, 4, 3, 2, 5, 6, 3, 9, 12, 1, 3, 2, 4, 5….a veritable nightmare! At a rate of 1 input every two seconds (allowing for reading from the questionnaire and inputting) and stopping for 5 minutes every 20 minutes for an eye-rest that would easily result in a 45-minute run. Run this for every query and one runs out of research time…

Researchers nowadays get a cool deal: they create the variable structure, then input the data either manually or import it from another tool such as a spreadsheet and *viola* their work is done in as little as a few seconds. The authors have run statistical tests on data sets comprised of 221 attributes and 153,080 records. That would result in inputting time of 392 solid 24-hour days!

Distributed

There are different types of tools, including the automated ones that do not reside in one's computer! Users do not even need to know where they are. The internet phenomenon has helped users to run statistical tests from their desktops using technologies that are automated but are located anywhere on the Earth (or even elsewhere, should a particular technology be located in orbit!). The user inputs the data locally and can run statistical tests across the attributes created by the researcher and even against other attributes created by other parties which again are stored somewhere else. These distributed datasets could again be anywhere and as long as they conform to the same research rules (covered in Chapter 2), they can be used. One can for example analyse Malta's epidemiological data against the WHO's (World Health Organisation) data using such tools, where such options are available.

The following sections describe the different types of tools that exist giving examples of simple outputs.

Spreadsheets

What are spreadsheets? The simple electronic tools remind us of graph paper where one could add or subtract numbers, draw or play simple games. Today's electronic version can allow for the same modes of operation but also allow for some basic statistical analysis. The tools are composed of multiple cells in what are described as rows (records) and columns (attributes).

	Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5
Record 1					
Record 2					
Record 3					
Record 4					
Record 5					

Spreadsheet cells allow for the inclusion of numbers, formulas and alphanumeric text. Formulas range from simple to very advanced and complex examples such as the following:

Example 1: Summing a set of values

Town	Population
Valletta	1000
Floriana	200
Qormi	300
Qala	500
Zebbug	800
Total	**=SUM(F2:F6)**

Town	Population
Valletta	1000
Floriana	200
Qormi	300
Qala	500
Zebbug	800
Total	**2800**

Example 2: Calculating how many persons live in a household aged over 40 years or less than 30 years

Age (Years)	Persons
46	Over 40 years
49	Over 40 years
29	Less than 30 years
38	Not Applicable
25	Less than 30 years
67	Over 40 years
12	Less than 30 years
33	Not Applicable
34	Not Applicable
67	Over 40 years
56	Over 40 years
4	Less than 30 years
78	Over 40 years
100	Over 40 years
10	Less than 30 years

=IF (E4>40,"Over 40 years", IF(E4<30,"Less than 30 years", "Not Applicable"))

Spreadsheets allow various basic statistical tools to be run and some modules also exist to expand on the tools and turn a spreadsheet into an advanced statistical tool. There are also online and standalone spreadsheets that in themselves offer very powerful features. The choice is nearly endless and the best advice one can give is, to use the one with the simplest interface and which has a modest set of commands.

The available tools

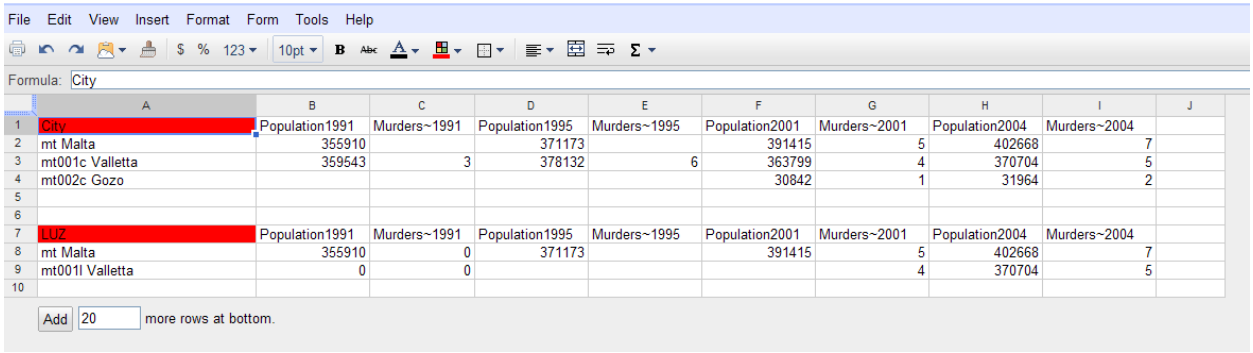
The most popular commercial Spreadsheets are IBM Lotus 123[1] and Microsoft Excel[2]. The former was originally created for DOS but has since been overtaken by Excel. Free/Opensource Desktop Spreadsheets are available that include OpenOffice Calc[3], Gnumeric[4], GNU Oleo[5], Bean Sheet[6], KSpread[7], SIAG[8], and Resolver One[9] (the latter free for personal use).

Free Online Spreadsheets include Simple Spreadsheet[10], wikiCalc[11], Google Spreadsheets[12] (Figure 9.1) and ThinkFree Online Calc[13].

---

[1] http://www-01.ibm.com/software/lotus/products/123/
[2] http://office.microsoft.com/en-us/excel/
[3] http://www.openoffice.org/
[4] http://projects.gnome.org/gnumeric/
[5] http://www.gnu.org/software/oleo/oleo.html
[6] http://bsheet.sourceforge.net/
[7] http://www.koffice.org/kspread/
[8] http://siag.nu/siag/
[9] http://www.resolversystems.com/
[10] http://www.simple-groupware.de/cms/Spreadsheet/Home
[11] http://www.softwaregarden.com/products/wikicalc/
[12] http://spreadsheets.google.com/
[13] http://member.thinkfree.com/member/goLandingPage.action#

Figure 9.1: Google Spreadsheets



	A	B	C	D	E	F	G	H	I	J
1	City	Population1991	Murders~1991	Population1995	Murders~1995	Population2001	Murders~2001	Population2004	Murders~2004	
2	mt Malta	355910		371173		391415	5	402668	7	
3	mt001c Valletta	359543	3	378132	6	363799	4	370704	5	
4	mt002c Gozo					30842	1	31964	2	
5										
6										
7	LUZ	Population1991	Murders~1991	Population1995	Murders~1995	Population2001	Murders~2001	Population2004	Murders~2004	
8	mt Malta	355910	0	371173		391415	5	402668	7	
9	mt001l Valletta	0	0				4	370704	5	
10										

Source: http://spreadsheets.google.com/

A very good site to access free opensource software is the Junauza.com[14] collection. There are many more such tools and an exhaustive list exists in the Wikipedia spreadsheet page[15].

Macros

As described earlier, some tools exist that serve as add-ons for spreadsheets called macros. These tools cater for specific requirements and are normally based on a sequence of commands that are enclosed within a 'shell' that can be run over and over as new data is inputted. Macros allow researchers to input their data in specified cells and run the resultant measure accordingly, thus drastically reducing the need for repeated work.

An excellent example of free macros can be found in the page: http://www.ozgrid.com/VBA/. Other related spreadsheets and macros webpage can be found in Matt H. Evans site[16]. Dedicated statistical macros, ranging from freeware to commercial licenses) can be found through the software.informer website[17].

Dedicated Statistical Software

Dedicated Statistical Software is available for researchers in various forms and covers both methodologies: those catering for quantitative and qualitative. In this section we cover both methodologies as well as the specialised quantitative-based spatial statistics tools. The tools are described in summary and a walkthrough has been created based on one tool for the readers serving as an introductory step towards the use of such tools.

Quantitative

A number of highly polished tools exist and are used by researchers in their process of analytical endeavour. The main tools used are PASW known as SPSS, SAS, Stata and MiniTab, with opensource tools gradually coming to the fore such as R-Commander, PSPP and Gretl. These examples are not exhaustive and many more tools are available.

1. SPSS (PASW)

SPSS[18] (Statistical Package for the Social Sciences), which was also called PASW (Predictive Analytics SoftWare) between 2009 and 2010 is a commercial statistical analytical processing tool. It has a base set of statistical features and can also be enhanced by various specialised modules.

---

[14] http://www.junauza.com/2008/04/freeopen-source-spreadsheet-programs.html
[15] http://en.wikipedia.org/wiki/List_of_spreadsheet_software
[16] http://www.exinfm.com/free_spreadsheets.html
[17] http://software.informer.com/getfree-statistical-macro/
[18] http://www.spss.com/

The user-friendly base tool caters for descriptive statistics such as frequencies and cross tabulation, bivariate statistics such as means, anova and correlations as well as nonparametric tests. It also carries out tests for linear regression, factor analysis and cluster analysis. The add-on modules include such tools as those of: tables, trends, advanced models and maps, amongst others.

2. SAS

SAS[19] (or Statistical Analysis System) is a commercial suite of statistical tools that was formed, based on the integration of a number of software tools. Since SAS works on the integration of a number of tools, the most adequate tool for statistical analysis is that called SAS/STAT[20], which together with BASE SAS and SAS GRAPH form the main components required for analysis within a tool called SAS Analysis Pro.

SAS/STAT can be employed for analysis through categorical data analysis, regression, bayesian analysis, multivariate analysis and other specialised functions as survival analysis, psychometric analysis, cluster analysis and nonparametric analysis, amongst others.

3. Stata

Stata[21] is a commercial general-purpose statistical tool, originally using a command-line interface but recent versions have been enhanced with GUI (Graphic User Interface) which makes it easier to use. Stata allows for such tests as: summary statistics, regressions, ANOVA, cluster analysis, survival models and cluster analysis, amongst others.

It is somewhat limited by its inability to load more than one file simultaneously and that it cannot load very large files. However, it has the capability to operate in the same way as opensource through the integration of online material.

4. MiniTab

MiniTab[22] is a commercial tool that together with another tool from the same company called Quality Trainer provides a range of statistical functions that are both wide-ranging and user-friendly. The statistical functions include basic statistics, descriptive statistics, regression, t-tests, variance, correlation, least squares and ANOVA, amongst others.

5. R-Commander

R-Commander[23] is an opensource tool that is deemed to be the most comprehensive, free statistical software available. Whilst the interface is a bit daunting for new users, there is a comprehensive manual and FAQs which guide the user in its usage. A number of plug-ins also enhances the product.

The tool and its plug-ins provide a wide-range of statistical tests, time-series analysis, classification, as well as linear and non-linear modeling. The outputs include very interesting 3D plots that emulate digital elevation models (DEMs) which can be used also for change analysis.

6. PSPP

PSPP[24] is a free opensource tool that was originally called Fiasco. It replicates SPSS functionality and serves as a useful tool for statistical analysis. The description states that “PSPP is a program for statistical analysis of sampled data. It is a free replacement for the proprietary program SPSS. PSPP development is ongoing. It already supports a large subset of SPSS's syntax. Its statistical procedure support is currently limited, but growing” (PSPP Installation File).

On first reviewing the tool, one would be forgiven to mistake it for SPSS, so mirrored is the whole concept. Whilst the number of statistical measures are limited, this tool serves as a very good tool for

[19] http://www.sas.com/technologies/analytics/statistics/index.html
[20] http://www.sas.com/technologies/analytics/statistics/stat/index.html
[21] http://www.stata.com/
[22] http://www.minitab.com/
[23] http://www.r-project.org/
[24] http://www.gnu.org/software/pspp/

most organisations since it covers the most used measures, inclusive of descriptive statistics, means, correlations, factor analysis, linear regression and non-parametric statistics, amongst others. The walkthrough uses this tool to aid researchers get an overview of how to use such tools.

7. Gretl

Gretl[25] is a free opensource asset that provides various statistical tools for econometrical analysis. Gretl stands for Gnu Regression, Econometrics and Time-series Library. Whilst not fully comparable to the other commercial tools as in the case of PSPP, it offers various time-series, maximum likelihood methods, least-squares based statistical estimators and econometric tests (Baiocchi et al, 2003).

A simple Walkthrough using PSPP

Part 1: Using PSPP for the first time - Values and Variables

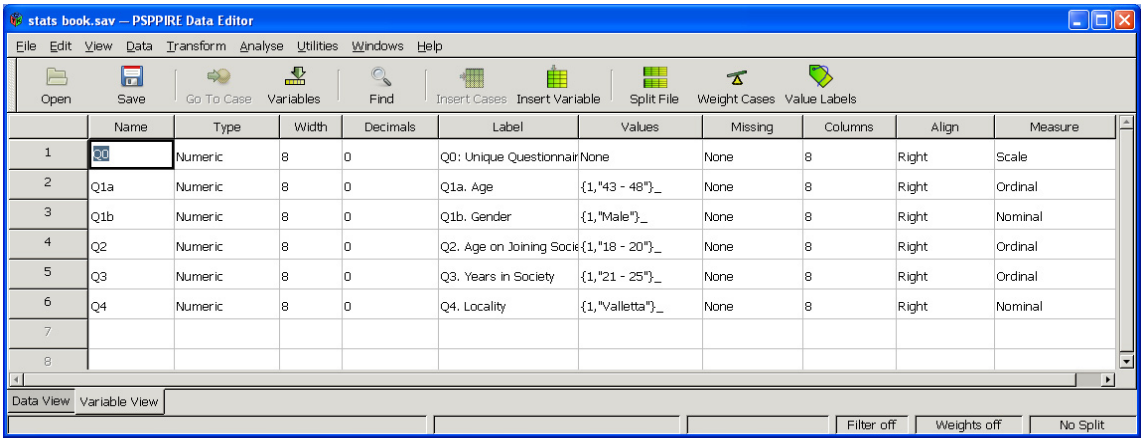
1) Open PSPP

2) Using PSPP for the first time – Inputting Data

There are two Views in PSPP, the Variable View and the Data View.

Variable View – input the variables from your survey here and list the different parameters being analysed, e.g.: for a question on sex: input 1 for male and 2 for female. This will do away with retyping each description for every questionnaire as against typing only one number (Figure 9.2).

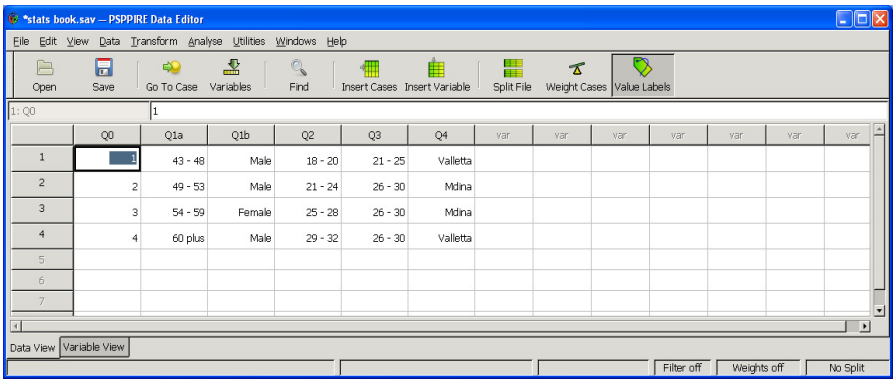
Figure 9.2: Data View



[25] http://gretl.sourceforge.net/win32/

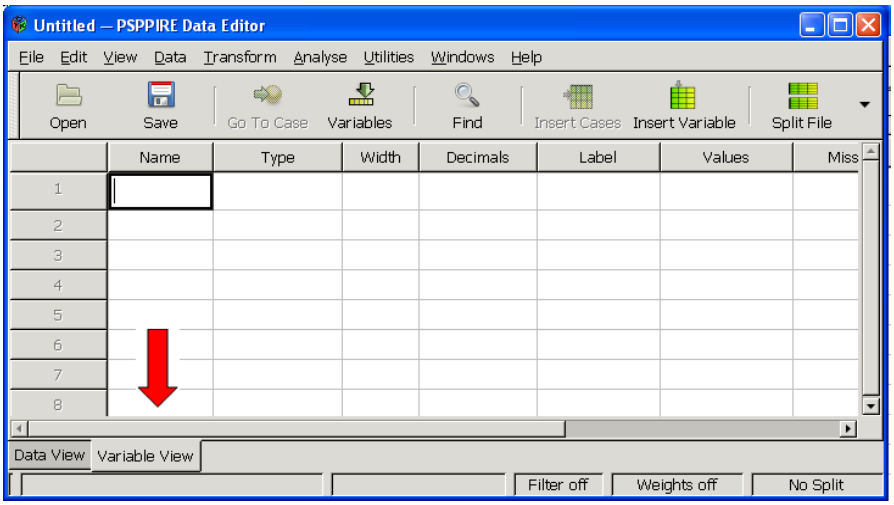
Data View – input the data from your survey in this tab sheet as per codes entered in the variable field above (Figure 9.3).

Figure 9.3: Variable View

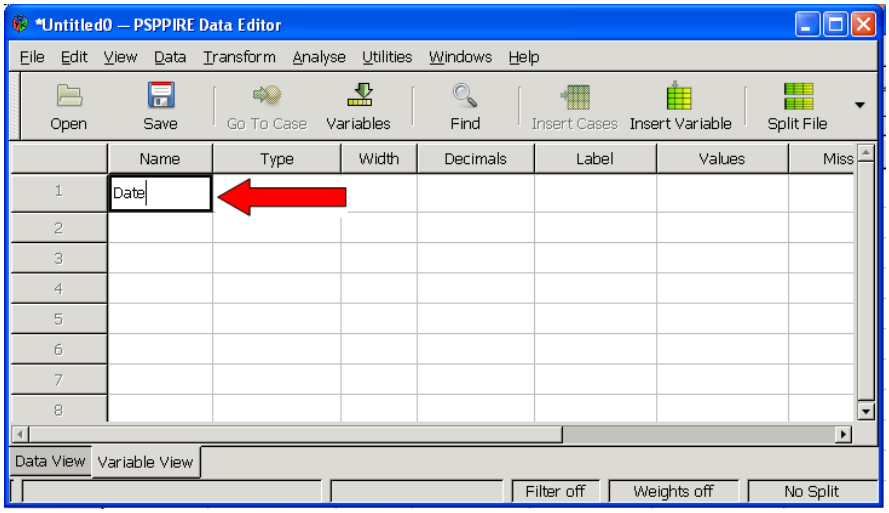


3) Using PSPP for the first time – the first variable.

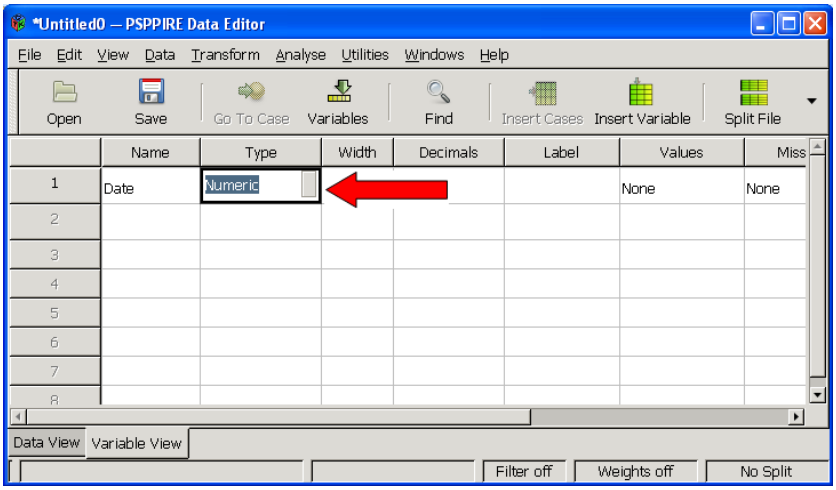
Click on Variable View option

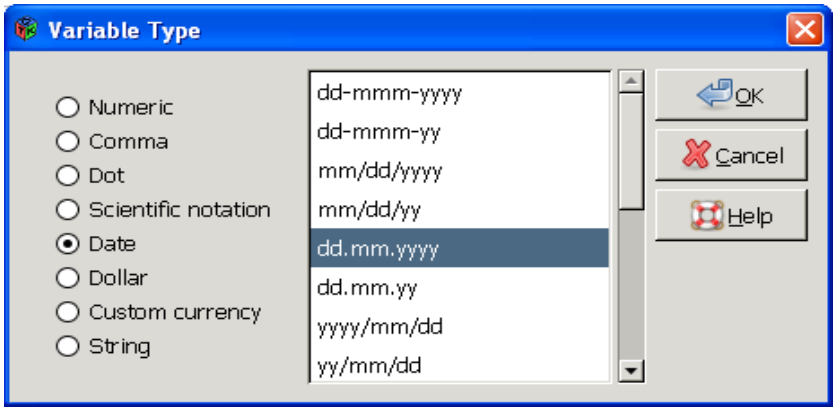


4) Double click on the top left hand cell (under Name) and type in ‘Date’.



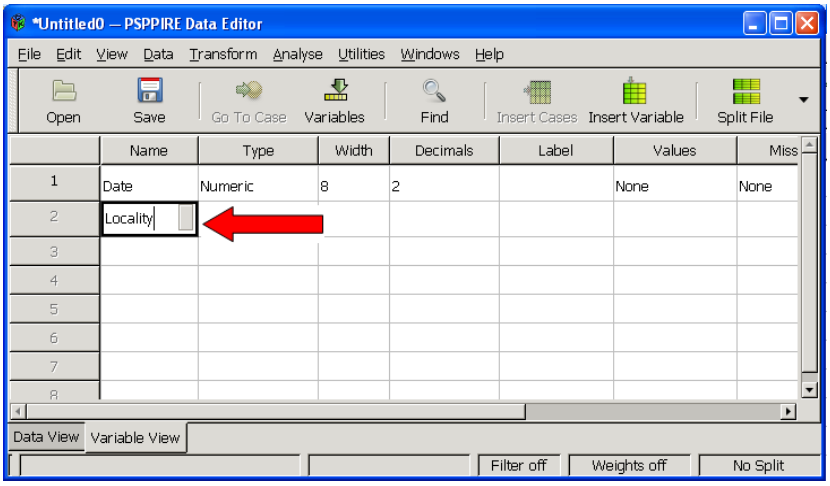
5) Double click on the top cell (under Type) cell. ‘Numeric’ and a choice box appears. Click on the grey box and then choose Date (dd.mm.yyyy) in the popup window.



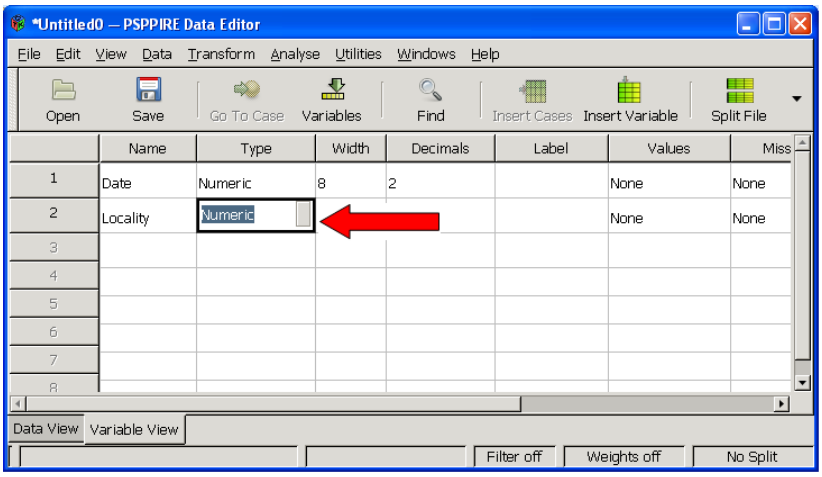


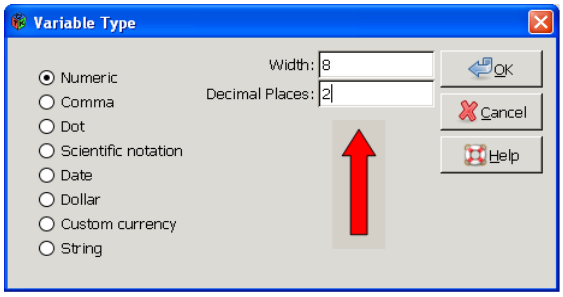
6) Using PSPP for the first time – the Variable Type.

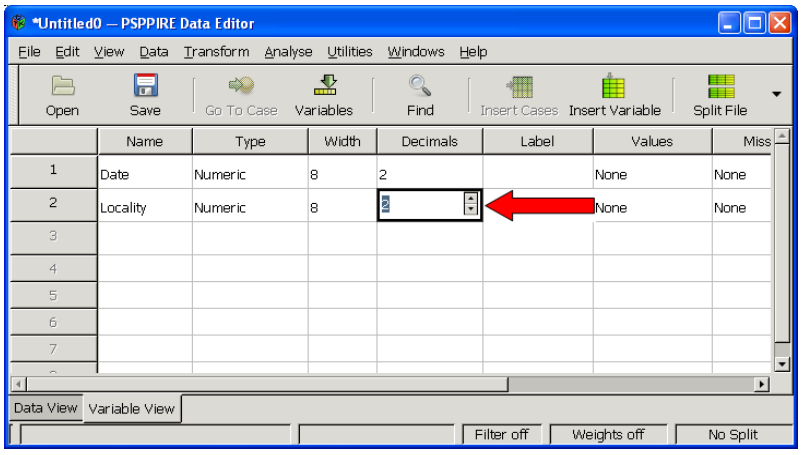
Double click on the second cell (under Name) and type in ‘Locality’



7) Press tab to go to the next parameter (Type). Ignore the number that appears under Width, Decimals, etc. In the popup window (Variable Type) type in 2 for Width and 0 for Decimal Places. Click Ok and you will notice that the numbers in under Width and Decimals change accordingly.

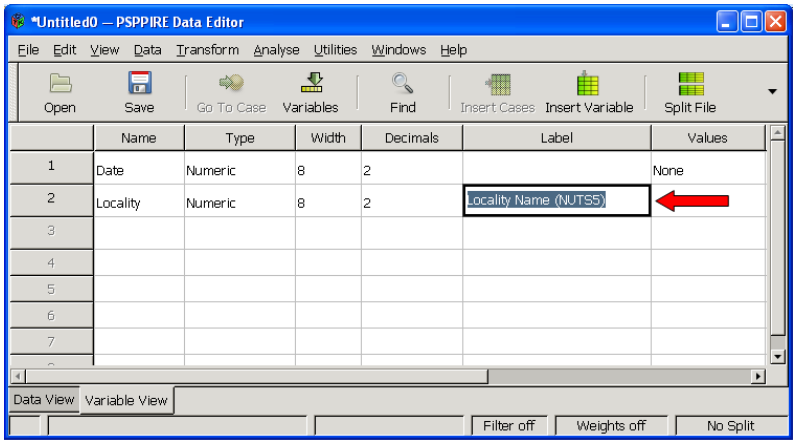




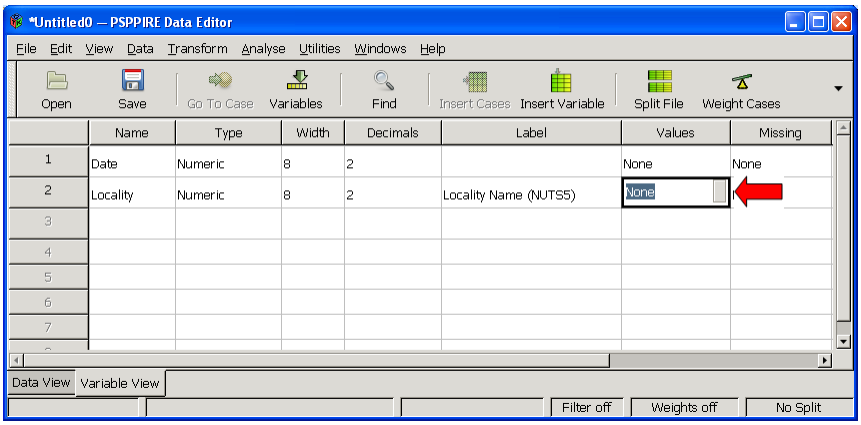


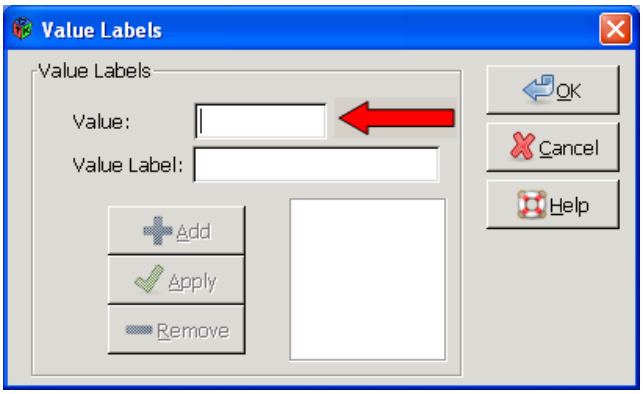
8) Using PSPP for the first time – the Value Label.

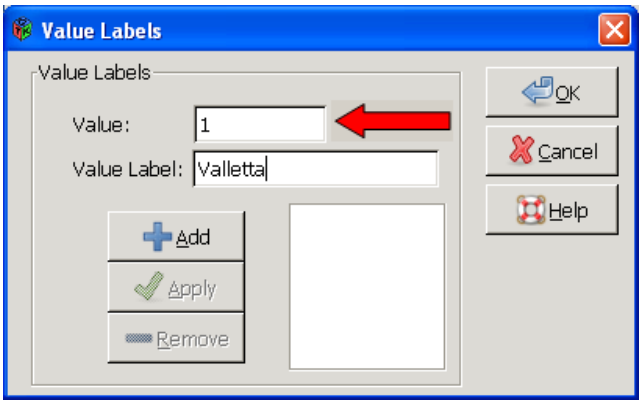
Click on Label and type in ‘Locality Name (NUTS5)’.

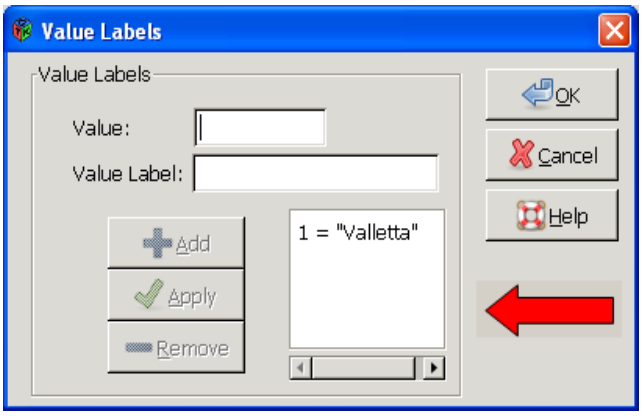


9) Click on Values and press the grey button. A Value Label window will pop up. Click on value and type in 1, then press tab and type in ‘Valletta’ in the Value Label box. Click on the add button. Note that the text ‘1 = “Valletta”’ will appear. Repeat this process for ‘Floriana’ with a Value of 2. Add some more localities.



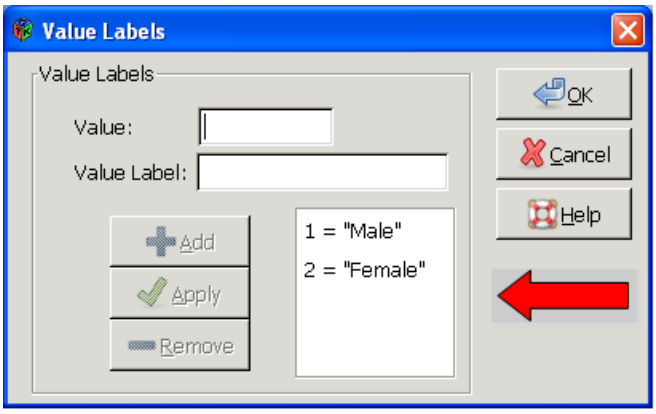






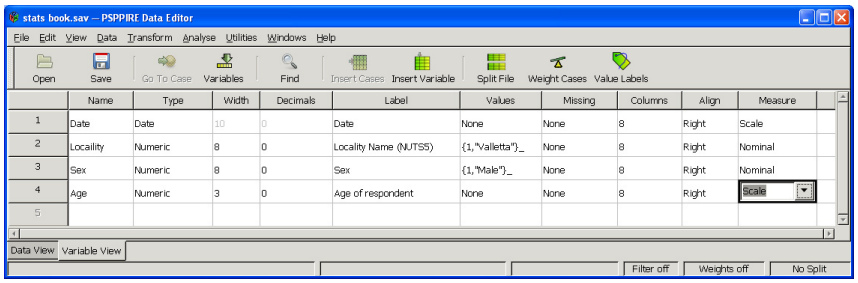
Using PSPP for the first time – some more inputs

10) Add another variable called Sex (under Name), Type (Numeric 2,0), Label (Sex of Respondent) and insert Value Labels (1 for Male and 2 for Female).



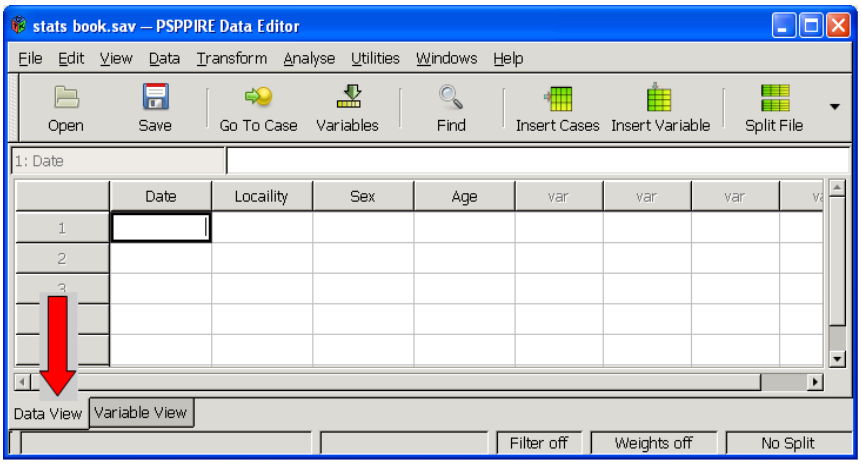
11) Insert another variable named Age, Type (numeric), Width (3) Decimals (0), Label (Age of respondent) and leave Values as none, since this is dependent on each individual's age rather than a specific category.

Note that the Measure has been inserted for the variables accordingly (Locality and Sex are Nominal whilst Age is termed as Scale).

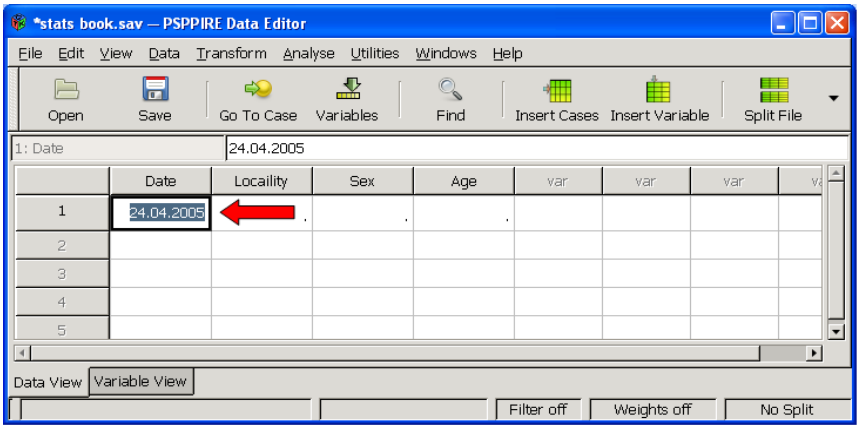


12) Using PSPP – inputting some real data now.

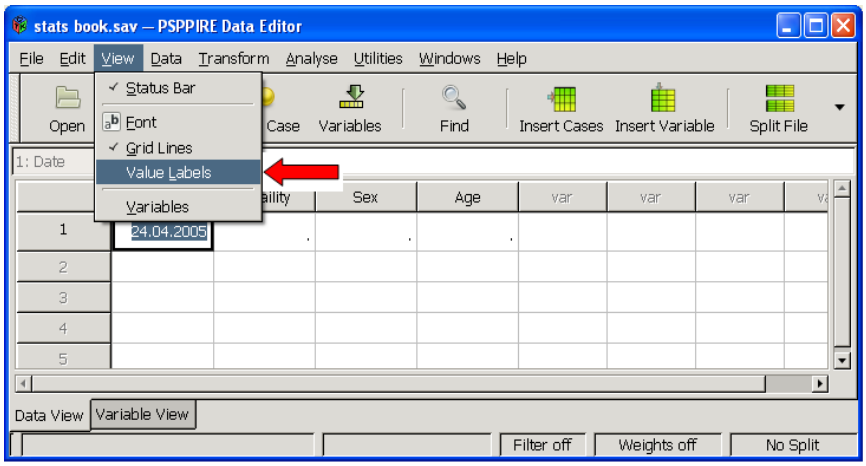
Click on Value View option



13) Type in 24.04.05 (under Date).

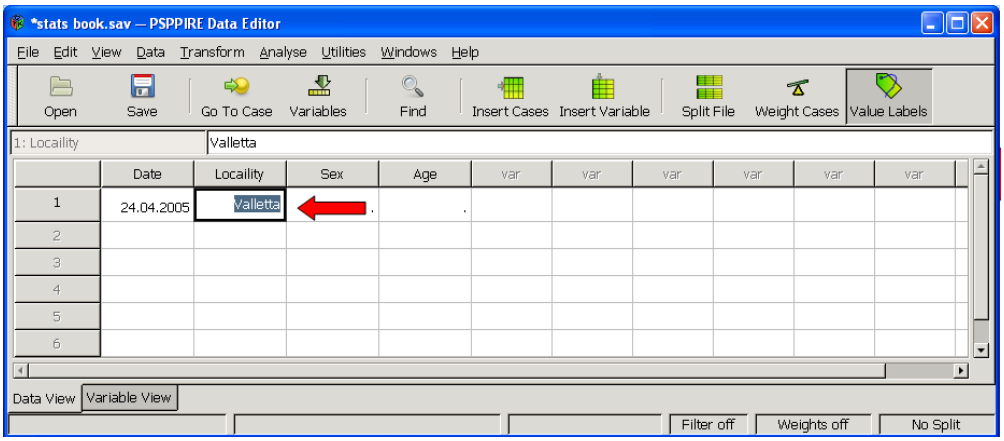


14) Click on the menu button VIEW and click Value Labels. This activates the choice button in the variable fields.

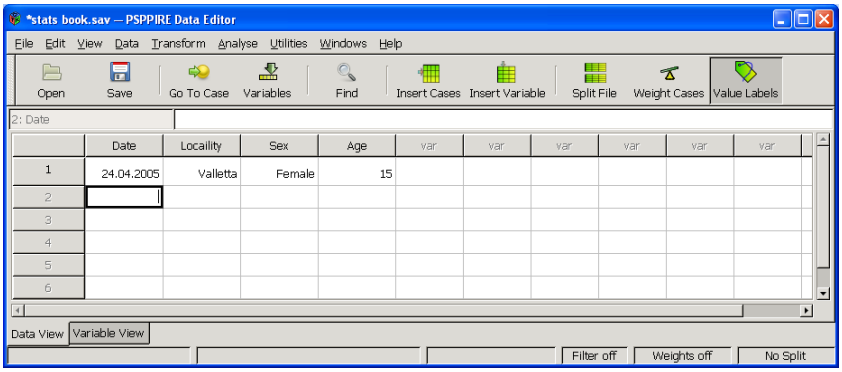


15) Using PSPP: choosing the values.

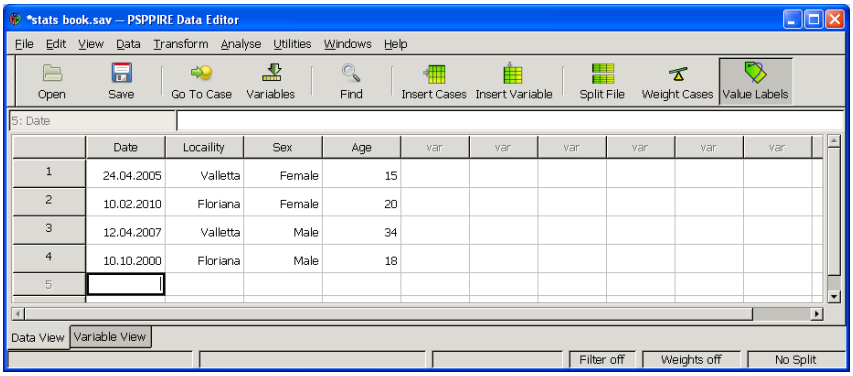
Go back to the input cells and click on Locality and insert your number that relates to the variable – e.g. typing in 1 under locality gives Valletta. Note the activations of the value Label at the Right Hand Side.



16) Repeat for Sex, whilst for Age type in 15.



17) Populate the dataset with a number of records.



Part 2: Using PSPP – The statistical measures: Frequencies

18) Running Frequencies – work instruction.

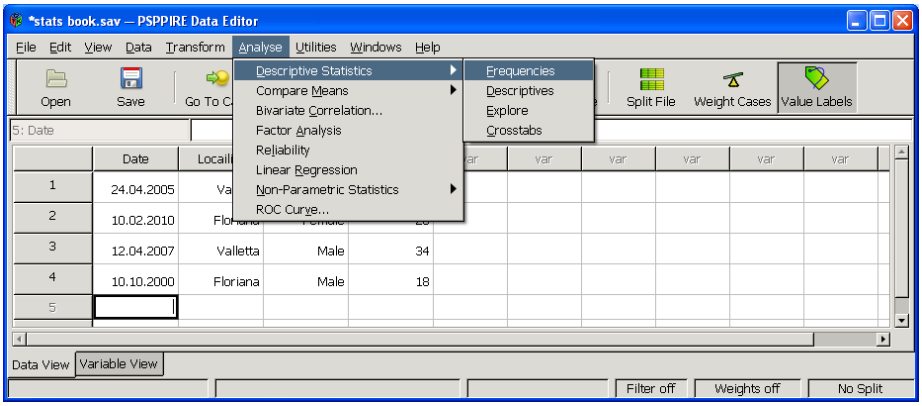
From within PSPP, Run Frequency and Descriptive Statistics

Follow these instructions.

1) Go to Analyse on the menu bar.

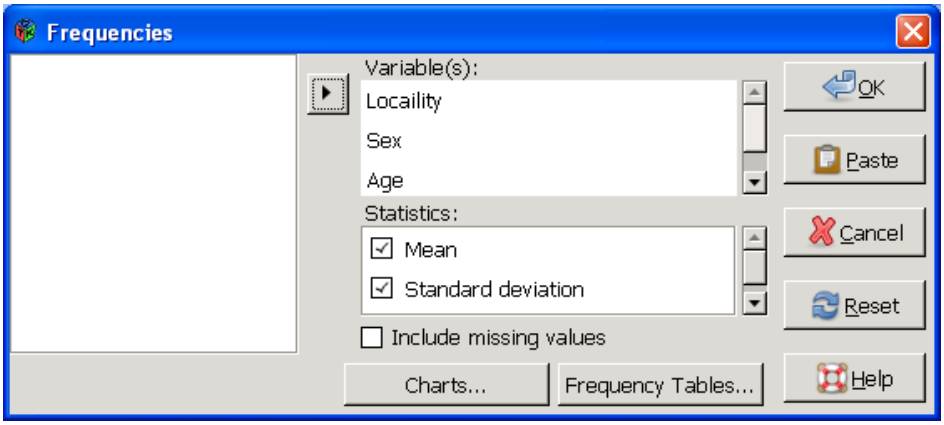
2) Click on Descriptive Statistics and then click on Frequency.

Figure 9.4: Descriptive Statistics

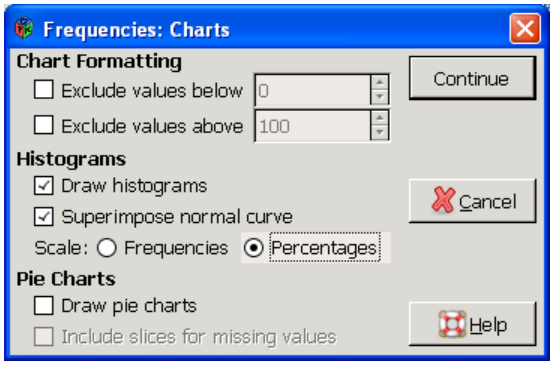


3) Select all variables shown in the left box, send to the right box (using arrow).

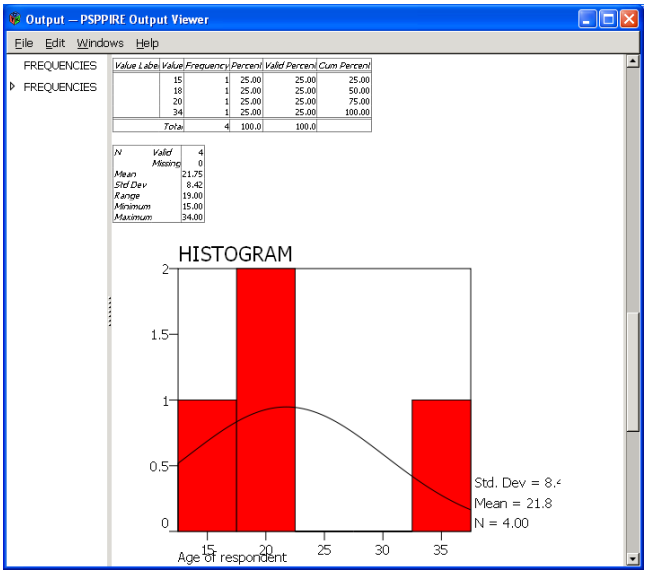
4) Select statistics (e.g., mean, minimum, maximum, standard Deviation, range).



5) Click on Charts, if you want to see charts for your variables.



6) Then click on OK. You will see the results in output file.



The above example gives a brief overview of what can be achieved with statistical tools. There are many statistical functions that can be carried out using PSPP. It is ideal to follow the PSPP tutorial that cover the different functions.

A list of free statistical tools can be downloaded from the Freestatistics website: http://www.freestatistics.info/stat.php

Qualitative tools

The software packages that are used in qualitative research analysis are called Computer-Assisted Qualitative Data Analysis Software (CAQDAS) or Qualitative Data Analysis Software (QDAS or QDA software). This kind of software help to: organize, categorize, annotate textual and visual data. This software aims at building theory while visualizing the relationships between data and/or theoretical constructs. Some of the software are the following: AnSWR[26], ATLAS.ti[27], CDC EZ-Text[28], The Ethnograph[29], Kwalitan[30], MAXqda[31], N6 or NUD*IST (which when superseded was replaced with NVivo[32] and XSight[33]), QDAMiner[34], Qualrus[35], TAMS Analyzer[36], Transana[37] and Weft QDA[38]. As an example of how these software packages function this section will provide a brief walkthrough of ATLAS.ti6 identifying some of functions that are used in qualitative research methods.

CAQDAS packages interpret the data collected through recognition and codification of the various research issues, facilitating the explanation and creation of theories. For instance, some of the approaches include grounded theory and conversation analysis. This does not exclude the possibility that qualitative researchers conjoin quantitative data with their methods. One of the most frequently asked questions in choosing one of the CAQDAS is "which is the 'best' package…?" It is practically impossible to answer this question as all the packages have tools that provide support in various stages of the analytic process and every program has its own advantages and disadvantages.

Some software packages suit certain types of approach more than other. This creates debates as to whether a particular package can be manoeuvred to suit a particular analysis. The researcher should remain in control of the interpretive process and decide if utilizing any software facilitates the chosen analysis approach. Whichever package one chooses, only a selected number of tools will be utilised in data running and analysis. The most sophisticated packages may not suit the envisaged task. Deciding which is the 'best' CAQDAS is a subjective judgment based on a numbers of factors but a more resolute choice of tools determines if a software package will serve over time.

Some similarities between CAQDAS packages

CAQDAS packages have key principles that facilitate the qualitative research process in similar ways. For instance in content analysis, transcribed interviews there is a process called KWIC (Key Words in Context). This tool offers ways to search in the text for singular words, phrases, or a collection of words on a particular theme. This function provides access to those keywords that appear in the analyzed documents.

Some of these software packages integrate code and retrieve functionalities. The user or research here can define key-words and/or theoretical categories (codes) that are embedded in the text. The researcher structures the coding and strategies to be employed. Code is simple and flexible and the researcher can modify and refine the coding as considered necessary. In most of these software packages the coding action rests entirely on the user.

All these packages offer means to control the research project and classify the data according to facts, features and data types. CAQDAS packages significantly facilitate qualitative research and analysis processes enables the researcher to focus on combinations and comparison of singular data. Qualitative data analysis is rarely a linear process. CAQDAS packages include various writing tools that enable the researcher to post comments and annotations of data that could not be reported in any other way; such as the non-verbal communications that one may observer during interviews. Some of the software

---

[26] http://www.cdc.gov/hiv/topics/surveillance/resources/software/answr/index.htm
[27] http://www.atlasti.com/
[28] http://www.cdc.gov/hiv/topics/surveillance/resources/software/ez-text/
[29] http://www.qualisresearch.com/
[30] http://www.kwalitan.nl/engels/
[31] http://www.maxqda.de/
[32] http://www.qsrinternational.com/products_nvivo.aspx
[33] http://www.qsrinternational.com/products.aspx
[34] http://www.provalisresearch.com/QDAMiner/QDAMinerDesc.html
[35] http://www.ideaworks.com/qualrus/
[36] http://tamsys.sourceforge.net/
[37] http://www.transana.org/
[38] http://www.pressure.to/qda/

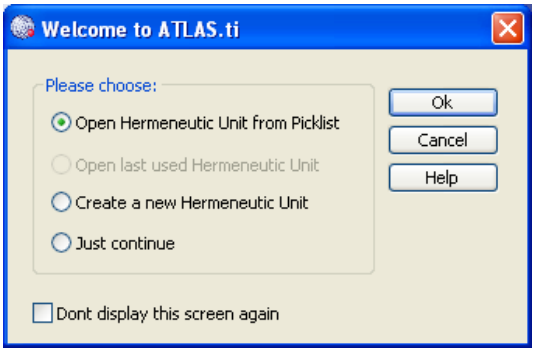
packages allow the researcher to analysis the material in hard copy or to integrate it in other applications such as Microsoft Word[39] and Excel[40]. These similarities indicate that each CAQDAS software package can be an asset that its methodical use of packages assists in continuity, increase precision and thoroughness in qualitative analysis.

A simple Walkthrough in CAQDAS using ATLAS.ti 6

http://www.atlasti.com/

Thomas Muhr from Free University, Berlin created ATLAS.ti and Scientific Software Development GmbH, Berlin is continuing to support the development of this software in the dynamic sphere of research. The features of ATLAS.ti can be applied in various fields of research such as art, social sciences, education and criminology. The last version of this software package is ATLAS.ti6 (Figure 9.5).

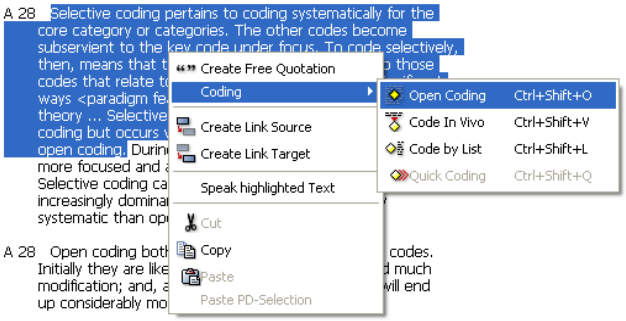
Figure 9.5: Atlas Ti6



The latest version of ATLAS.ti acts as an external database and it is possible to transcribe multimedia data using this software. This software has an electronic "room" called Hermeneutic Unit (HU), which helps text interpretation and prepares the data for each project analysed by ATLAS.ti. Thus the HU is connected between the primary data and any annotations taking in examining the collected data and keeps track of all of your data in ATLAS.ti project file. The data can include text, images audio and visual recordings, pdf files, and data extracted from Google Earth[41].

Functions operate from main menus and drop-down menus that can be access through the Manager windows. Selected quotes, text (Figure 9.6), images (Figure 9.7) or other documents enables flexibility in the use of coding.

Figure 9.6: Coded Text



[39] http://office.microsoft.com/en-gb/word-help/word-help-and-how-to-FX010064925.aspx?CTT=97
[40] http://office.microsoft.com/en-gb/excel-help/excel-help-and-how-to-FX010064695.aspx?CTT=97
[41] http://earth.google.com/

Figure 9.7: Coded Images

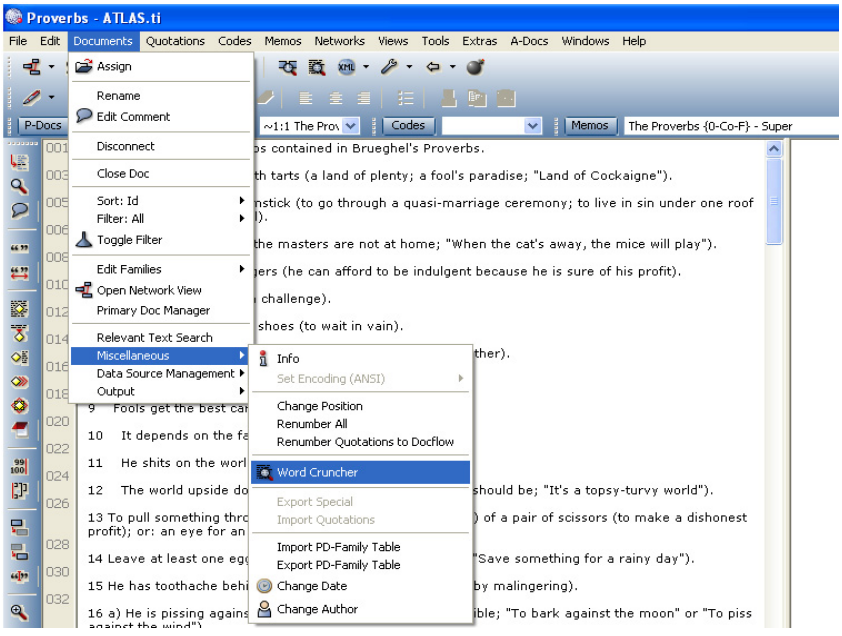
(This image is one of Hitchcock's cameos extracted from the Hitchcock's movie *North by Northwest,* adopted from http://www.filmposters.com/, accessed 9th August, 2010)

However the output results this kind of data are independent from any coding function. Quotations are not dependent on ATLAS.ti software, selected data can be pointed and separately marked without being coded. However this independence from the software does not hinder that quotations are located in the framework. This feature allows one to produce quotations in different formats such as audio and can be play-backed autonomously from the written transcript.

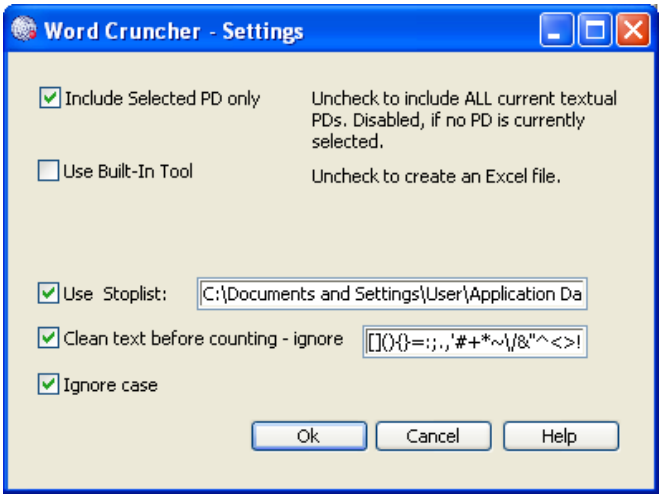
Word Cruncher

The Word Cruncher counts the number of times a word appeared in the whole collected data or a particular document. The results can be escorted in a Microsoft Excel spread sheet or in a form of a memo. After uploading the text in ATLAS.ti, these are the steps taken to use the Word Cruncher:

Step 1 – Click on **Documents**, then **Miscellaneous**, then **Word Cruncher.**

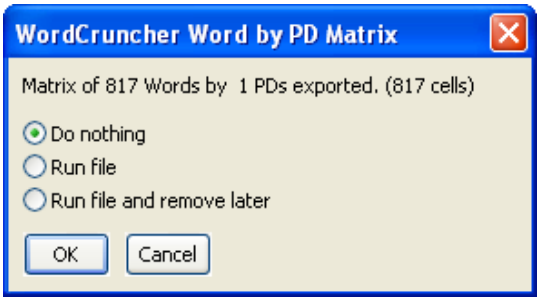


Step 2 – Click on Properties before starting the frequency count.



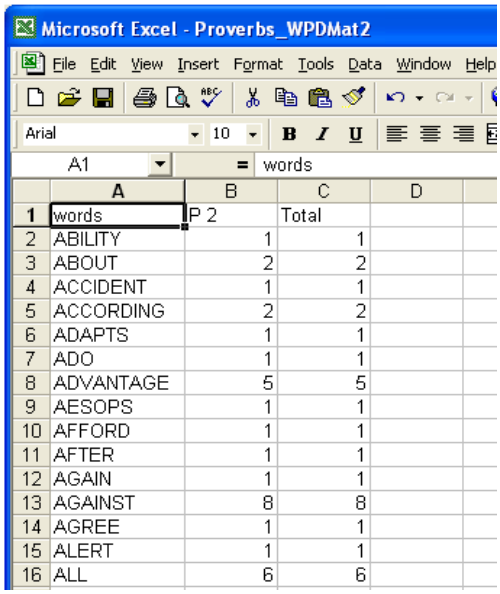
The results can be either saved in Excel or not. For this example the user opted to use Excel as the user wanted to use the data for further analysis.

Step 3 – This is the final window before the user can proceed to the final output. If you click on "**Do nothing**" the frequency count is saved in Excel. If you select "**Run file**" the file is opened in Excel, if you click on "**Run file and remove later**" you may want the file to be removed after viewing the results.



Naturally, you need to have Excel installed before being in a position to see the results.

Step 4 – Open Excel where the information is stored.



	A	B	C	D
1	words	P 2	Total	
2	ABILITY	1	1	
3	ABOUT	2	2	
4	ACCIDENT	1	1	
5	ACCORDING	2	2	
6	ADAPTS	1	1	
7	ADO	1	1	
8	ADVANTAGE	5	5	
9	AESOPS	1	1	
10	AFFORD	1	1	
11	AFTER	1	1	
12	AGAIN	1	1	
13	AGAINST	8	8	
14	AGREE	1	1	
15	ALERT	1	1	
16	ALL	6	6	

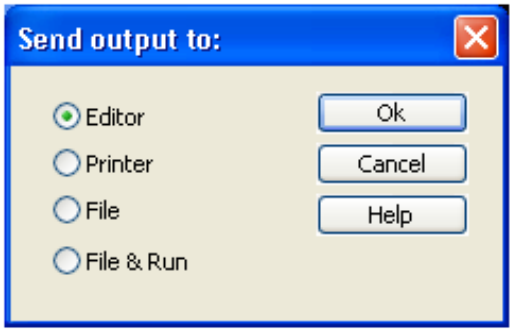
ATLAS.ti6: Remarks

Holistically ATLAS.ti is a flexible system to work with and can be engage in various qualitative projects. This software does not need codes to function but yet one can easily filter the data collected. Thus the researcher has all the liberty to manage the different linkages in the data. The latest version of this software (i.e. ATLAS.ti6) has a number of upgrades when compared to the earlier versions. ATLAS.ti6 supports pdf files and creates a very precise representation of the pdf including the layout, images and other contents. Another important feature that can be accessed through ATLAS.ti6 is Google Earth (GE). This package offers the possibility to include geographical references and navigation through GE facilitating the creation of snap-shots of the interested spots around the world.

However there are some issues that the user of ATLAS.ti6 has to be aware of. This software has no KWIC and the word frequency tool is quite basic. Besides ATLAS.ti does not have any function that integrates quantitative data with filtered qualitative data. Another function that may raise some eyebrows is that the code list does not have a function to create connections in a hierarchical structure to systematically clean up the collected data. However for some other users this is an advantage, as it does not restrict the researcher in anyway. Nevertheless ATLAS.ti6, like any other software, is continuously being updated and new tools are constantly being added.

Output generated by ATLAS.ti

Among the output of the information generated, ATLAS.ti usually creates textual information though exists the possibility of producing a graphical result. Textual reports usually include lists of codes, annotations and citations. In order to create textual output the following window is displayed providing four options (as found in the diagram) of how the user wants to extract the data.



Geo-statistical tools

A third set of tools which are available for researchers are those related to spatial statistics. Though forming a specialised methodology, they form part of the quantitative approach, but bring into context the spatial element. This element as discussed in the Visualization Chapter looks at the analysis of statistics through the spatial dimension. It employs GIS and other specialised tools.

There are various tools covering geostatistics; some commercial such as ArcGIS Geostatistical Analyst and MapInfo Vertical Mapper, whilst others were produced through grants such as CrimeStat, SAGE and STAC. Others have been developed as opensource which include Gstat and SGeMS.

1. ArcGIS Geostatistical Analyst

ArcGIS Geostatistical Analyst[42] is a commercial tool that serves as an extension to ArcGIS Desktop. It consists of a suite of geostatistical tools focusing on spatial data exploration and surface generation. The tools allows for data sampling, interpolation surfacing and prediction modelling. The tool requires the main ArcGIS software to operate.

[42] http://www.esri.com/software/arcgis/extensions/geostatistical/index.html

2. MapInfo Vertical Mapper

A commercial tool which is also an add-on to another GIS software, MapInfo Vertical Mapper[43] sits on the GIS tool MapInfo and provide a set of tools that produce: trend analysis, gridding algorithms, prediction modeling, gravity modeling, risk modelling and large dataset correlations.

Though not strictly fully statistical tools, both MapInfo Vertical Mapper and ArcGIS Geostatistical Analyst serve to provide specific geostatistical function not available in mainstream applications. The outputs are in visualised formats such as 2D and 3D maps as well as in VRML and other virtual environments.

3. CrimeStat[44] (Levine, 2002)

'*CrimeStat®* is a free spatial statistics program for the analysis of crime incident locations, developed by Ned Levine & Associates under grants from the National Institute of Justice (grants 1997-IJ-CX-0040 and 1999-IJ-CX-0044). CrimeStat allows the analysis of: standard deviation maps, attribute analysis, journey to crime, hotspot analysis and a series of spatial statistical measures (Formosa, 2007). Though the software was created to analyse crime statistics, it is a robust tool and is a veritable multi-thematic / multi-discipline tools as it can be used for both social and natural scientific analysis.

As an example of types of spatial statistics used in crime, listed below are the CrimeStat categories clustered in four-groups[45]: Spatial distribution, Distance statistics, 'Hot spot' analysis routines and, Interpolation statistics (refer to Chapter 11).

The software works as a standalone and exports its data in GIS format for further mapping through the dedicated GIS software.

4. SAGE

SAGE[46] (Spatial Analysis in a Geographical Environment) was produced under ESRC (Economic and Social Research Council) research grant R000234471 'Developing spatial statistical software for the analysis of area-based health data linked to a GIS' (Craglia M., Haining R., and Wiles P., 2000).

The idea behind the project is described by Wise, Haining; and Ma (2001). It was structured through the creation of a software for statistical spatial data analysis (SSDA) which was concatenated with ARC/INFO. This product eventually produced SAGE.

5. STAC

STAC[47] (Space and Temporal Analysis of Crime) software was developed by the Illinois Criminal Justice Information Authority (ICJIA). A Users Manual is available: Users Manual and Technical Manual, (1996), Chicago, IL: ICJIA. It is a free tool that helps spatial analysts in their statistical analysis. It achieves this through cluster mapping, employing standard deviational ellipse creation. STAC was eventually integrated into the CrimeStat II tool.

6. Gstat

Gstat[48] is a free opensource tool that war developed to enable multivariable geostatistical modeling. It also predicts and simulates modeling scenarios. The tool can now be used in conjunction with R-Commander earlier described in the quantitative section. It is capable of calculating variograms, kriging, and allows unlimited variables to be cross-correlated.

---

[43] http://www.pbinsight.com/products/location-intelligence/applications/mapping-analytical/vertical-mapper/
[44] http://www.icpsr.umich.edu/NACJD/crimestat.html
[45] http://comm-org.utoledo.edu/pipermail/announce/1999-December/000025.html
[46] http://www.informaworld.com/smpp/content~db=all~content=a713811641
[47] http://www.icjia.state.il.us/public/index.cfm?metasection=Data&metapage=StacFacts
[48] http://www.gstat.org/

7. Stanford Geostatistical Modeling Software (SGeMS)

A free opensource tool, SGeMS[49] hosts a plethora of algorithms catering for extensive multiple-point statistics simulation and 3D visualization. Apart from standard data analysis tools such as histogram, QQ-plots and variograms, the tool also provides for kriging, multi-variate kriging (co-kriging), sequential gaussian simulation, sequential indicator simulation, multi-variate sequential gaussian and indicator simulation

In addition to the above, a list of geostatistical tools is available from the http://www.brynmawr.edu/geology/GIS/geostats.html website. The GWR[50] (Geographically Wieghted Regression) website also highlights the importance of specialised tools for geostatistical analysis.

Online Tools

There are two types of online tools: those that cater for the analysis of data and those that help the researcher to create an online survey for world-wide respondent input.

Some basic online analysis tools

There is quite a number of statistical tools that run quick tests for users who do not wish to acquire or use large complex software.

This section lists some of these tools in brief.

i) GraphPad Software

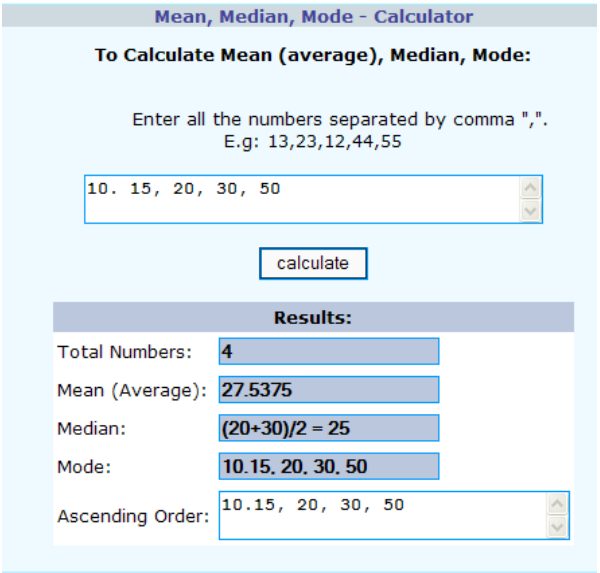
Website: http://www.graphpad.com/quickcalcs/index.cfm

This site provides tools for the following categories: categorical data, continuous data, statistical distributions and interpreting P values, random numbers and chemical and radiochemical data.

ii) Online Measures of Central Tendency Calculator

Website: http://easycalculation.com/statistics/mean-median-mode.php

This simple but very effective tool calculates the Mean, Median and Mode of a set of variables (refer to Chapter 11).



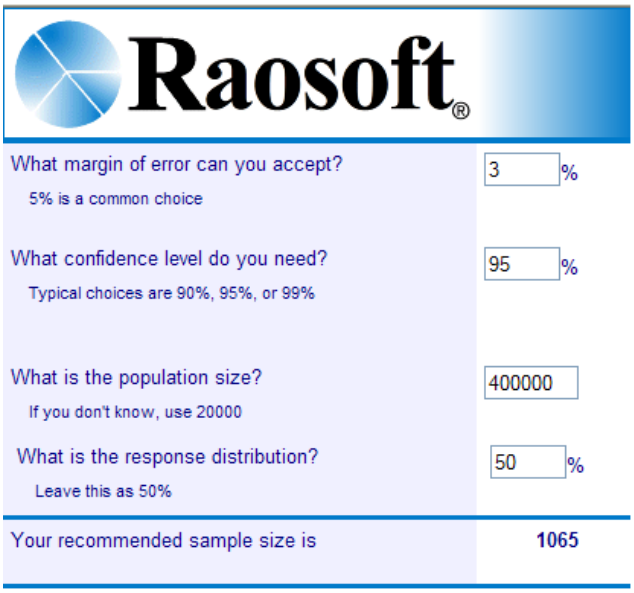
[49] http://sgems.sourceforge.net/
[50] http://ncg.nuim.ie/ncg/GWR/whatis.htm

iii) Sample Size Calculator

Website: http://www.raosoft.com/samplesize.html

This effective sampling size tool by RaoSoft is highly sought and has gone through an incarnation which tool calculates the sample size required for a population. The tool gives an example of what the Maltese islands sample population at 3% margin of error, a confidence level of 95% and a response distribution of 50%.

Interestingly the sample size stands at 1,065. Half the margin of error and the sample size quadruples to 4,224, half it again and the sample size quadruples again to a figure of 16,375.



Other online statistical tools can be found in the StaPages.org website: http://statpages.org/. One particularly interesting tool that can be investigated is StatCrunch[51] which allows online data analysis and even hosts and online mapping tool.

ii. Surveying tools – online

This final section on tools covers those ready-made online packages that help researchers, either free or at low cost to carry out their own online surveys with the tool even delivering either base statistical analysis or raw data for importation into statistical packages.

Other tools do not use ready-made online services but employ other desktop-based tools such as spreadsheets to export to an interactive online format, with replies being returned through email. A sample list of these services is covered below.

1. Survey Monkey

Survey Monkey[52] offers a reliable service, with a free basic service and a low-priced advanced service. The service offers pre-prepared templates hosting a considerable number of question types ranging from multiple choice to text boxes to demographics. Interestingly, Survey Monkey allows randomization and sorting of answers to ensure that each time the survey is run the question appears differently structures.

[51] http://www.statcrunch.com/
[52] http://www.surveymonkey.com/

2. Question Pro

Question Pro[53] provides a free service with an add-on priced account. The free version has quite a large number of features inclusive of multiple choice, Likert scale, open ended and essay open text, rank order, and template library amongst others.

3. FreeonlineSurveys

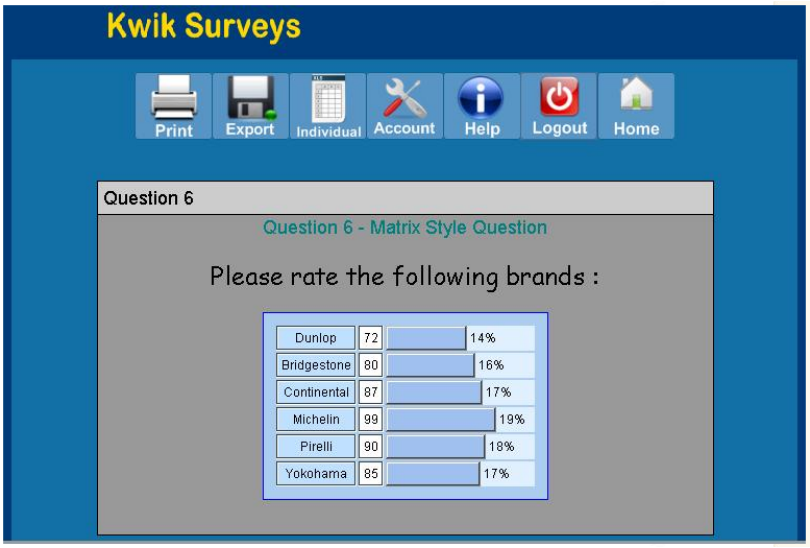
Freeonline Surveys[54] is another online survey which has the limited functionality as that of the other free services in addition to its priced services, but has an interesting output: that of the data appearing at the same time that it is inputted.

4. eSurveysPro

eSurveysPro[55] has a free online version as well as commercial options. It allows one to create free online unlimited surveys, questionnaires and responses as well as providing a survey editor and professional reporting.

5. Kwik Surveys

Advertised as a truly free online survey tool, Kwik Surveys[56] was created to serve as an opensource version of online surveys. It provides a veritable variety of options inclusive of different question types, (such as multiple choice, matrix, star rating and text input), randomization, logo upload, page skipping and other structuring features.



6. Creating one's own online survey

Tools exist that allow researchers to work through their data using proprietary software and then export to an online service for eventual data reporting and analysis. An example of such a tool is the Excel-Based Spreadsheet Converter[57] that takes an Excel survey or questionnaire and converts the relevant input cells into code for online input. The results of this tool are sent through email for eventual inputting. An example of such a survey is shown below. The highlighted colour indicates those cells that the respondent is asked to choose from.

[53] http://www.questionpro.com/
[54] http://freeonlinesurveys.com/
[55] http://www.esurveyspro.com/
[56] http://www.kwiksurveys.com/
[57] http://www.exceleverywhere.com/

Microsoft Excel Desktop Version

	A	B	C
1		**Survey: Investigating the use of the Internet**	
2			
3	1	Age:	
4			
5	2	What kinds of research studies would be worth funding in order to formulate internet policy?	
6			
7	3	Sex:	
8			
9	4	Occupation:	
10			
11	5	What is your general interest in Social Networks?	
12			
13	6	Do you have full open access to a PC with internet connection at home?	
14			
15	7	Did you make use of any of the government's Information technology internet incentives?	
16			
17	8	How many hours on average do you spend on the Internet daily?	
18			

Spreadsheet Converter Online Version

	Survey: Investigating the use of the Internet	
1	Age:	
2	What kinds of research studies would be worth funding in order to formulate internet policy?	18 yrs – 20yrs 21 yrs - 23yrs 24 yrs - 26 yrs 27 yrs – 29yrs 30 yrs +
3	Sex:	
4	Occupation:	
5	What is your general interest in Social Networks?	
6	Do you have full open access to a PC with internet connection at home?	
7	Did you make use of any of the government's Information technology internet incentives?	
8	How many hours on average do you spend on the Internet daily?	

In summary, this Chapter has given an overview of the tools available to researchers that are specifically related to statistics. Databases have not been considered as they will be covered in a separate chapter.

From spreadsheets to quantitative to qualitative and geostatistical tools, the researchers has a veritable sea of choices. The main issue is to keep to one or two tools in order to become deeply conversant with their functions. Also consider whether it is best to use desktop or online versions as these may have a relative impact on process, access and most importantly functionality.

Questions (refer to Appendix for the answers)

1. List the three main different categories of statistical tools.
2. What are “spreadsheets”?
3. What do “Macros” do?
4. What do the letters “SPSS” stand for and what is this?
5. What do the letters “SAS” stand for and what is this?
6. What is “Stata”? What does it do?
7. What is “MiniTab”? What does it do?
8. How would you briefly describe “R-Commander”?
9. What is “PSPP”?

10. What is “Gretl”?
11. What do the letters “CAQDAS”, “QDAS” and “QDA” stand for? What does this software do?
12. What do the letters “KWIC” stand for? What does this tool do?
13. What does the “Word Cruncher” do?
14. What is the “ArcGIS Geostatistical Analyst” and what does it do?
15. What does the “MapInfo Vertical Mapper” do?
16. What is “CrimeStat”? What does it do?
17. List the four main CrimeStat categories.
18. What do the letters “STAC” stand for? What is this and what does it do?
19. There are two types of online tools. Which are they?