SR501 for Windows (draft only)   © Copyright 2005 John F Hall

This document describes current work on converting and updating teaching materials from the final version of the postgraduate part-time evening course **SR501 Survey Analysis Workshop** taught at the Polytechnic of North London from 1976 to 1992.

Originally taught over 20 sessions, later reduced to 15, then 14, split between one hour of statistics (taught from 1976 to 1980 by the late John Utting[1], from 1981 to 1987 by Jim Ring[2]) and two hours of SPSS (taught throughout by John Hall[3], but with assistance from Maureen Ashman[4] 1976-78) the course was targeted primarily at trainee researchers, postgraduate students and (unemployed) social science graduates in the form of evening short courses.

A sister course, **SR502 Survey Research Practice**, comprising 10 sessions, later increased to 15, then reduced to 14, was taught entirely by practising professionals from commercial and not-for-profit research organizations. Both courses represented a continuation of the Summer Schools in Survey Methods offered by the then SSRC Survey Unit from 1970 to 1976 (initially at Oxford and latterly at Reading).

Although the courses could be taken separately, most students took both. SR501 was always taught before SR502 as it was felt that knowledge and appreciation of the technical side of data capture, processing and management, together with the emphasis on the logic of analysis, would not only help dampen enthusiasm for the generation of lengthy questionnaires, but also improve the subsequent quality of questionnaire design.

In a much condensed form, survey analysis was also taught to final year undergraduate students on the new 4-year BA Applied Social Studies (Social Research Option) from 1979. On semesterisation, the sessions in Survey Analysis were reduced from 20 to 15, and in Survey Research Practice increased from 10 to 15. Both courses were later reduced to 14 sessions to allow for assessment.

In 1976 there were no on-site computing facilities in Ladbroke House (the base for the Faculty). Indeed, since it was the very first course to be held in the evenings, there were no facilities for anything else either. Each week, coffee and biscuits were brought in by teaching staff, a couple of kettles commandeered, and on the first week a couple of cases of Bulgarian Cabernet Sauvignon served to stimulate student exchanges in the introductory session.

Student exercises in SPSS had to be written on coding sheets and sent to Computer Services to be punched on 80-column Hollerith cards, then run without being checked. The error rate was high and resulted in frustrating losses of time and motivation. Moreover, student experience of computers was virtually non-existent: many of them could not even type!

With the advent of personal computers, there was a remarkable improvement in the previous skills of students and with the installation of a computer laboratory with 16 terminals, 4 servers and a fast link to the mainframe, the whole course was revolutionised. Much less time needed to be

---

[1] John Utting was Deputy Director of the SSRC Survey Unit from 1971 to 1976, then Deputy Director of the National Childrens Bureau from 1976 until his retirement in 1981.

[2] Jim Ring was Research Officer at the SSRC Survey Unit from 1972 to 1976 and Senior Research Officer at the Polytechnic of North London Survey Research Unit from 1976 until he moved across to PNL Computer Services in 1985, where he still works, in the new London Metroplitan University.

[3] John Hall was Senior Research Fellow at the SSRC Survey Unit from 1970 to 1976, then Principal Lecturer in Sociology and Director of the Survey Research Unit at the Polytechnic of North London from 1976 until his early retirement in 1992

[4] Maureen Ashman was a programmer at PNL Computer Services with special responsibility for SPSS

spent on basic technology and keyboard skills, and much more could be devoted to hands-on data management and analysis, with a consequent benefit to work-rates and motivation. When the courier service was supplemented by two fast line-printers, no student left empty handed even from the first session.

When Jim Ring was unable to continue teaching the statistics component, he provided an early draft of his notes for distribution to students whilst John Hall revamped the sessions into one hour of presentation followed by two hours of hands-on computing covering the statistical elements as and when appropriate, but not to the same depth, and including discussion of results. The undergraduate version of survey analysis was subsequently moved into the second semester of the second year of the re-validated course and continued into the modular degree scheme for BSc Social Science (SR206: Data Management and Analysis). It was compulsory for Social Research and strongly recommended for Sociology students. It continued as a part-time evening postgraduate course (SR501) for which, from 1990, students could gain 15 points towards a CNAA Masters' degree, provided they took the assessment.

For BSc Social Research students this not only complemented their statistics course, but also prepared them for effective professional placements in their 3rd year, since many of them later got jobs with their placement agency: others even found employment as researchers ahead of candidates with Master's or PhD's from other universities. Previous students on both courses are now frequently to be found in senior positons in UK social research.

The original notes were mostly written direct on to the Vax mainframe or in WordStar4 on a PC and refer to successive releases of SPSS on a variety of mainframe computers, from the CDC2000 at ULCC (via LSE) through DEC-10 and Dec-KL20 and finally the Vax cluster at PNL. The current undertaking is to convert these files to MS-Word format and to recreate or generate parallel examples of syntax, output and system files suitable for use with SPSS for Windows. Because conversion of tabular output from WS4 to MSWord is tedious and complicated, and also because some original data has been irretrievably lost (and therefore not available to regenerate tables in Windows format), some files retain tables and figures in the original lineprinter format of SPSS output.

There are three main kinds of documents:

> General notes on the nature of survey data, use of computers and logic of analysis

> Syntax intro and summaries for SPSS commands, including tricks of trade

> Worked examples, class exercises and homework exercises

In addition there is a set of statistical notes designed to accompany the course, and to complement available texts. See **Statistics booklets** in **Survey Analysis Workshop** below.

There are two main folders: **SR501 for Windows**, which holds current editions of (revised and updated) file conversions, and **Survey Analysis Workshop**, which holds most of the files written for the original PNL course, including some still in Courier typeface and a few still in WordStar4.

The following screen dumps show what sub-folders and files are contained within each main folder: within each sub-folder, files are listed with an extended label and description of their contents.

Main folder: SR501 for Windows
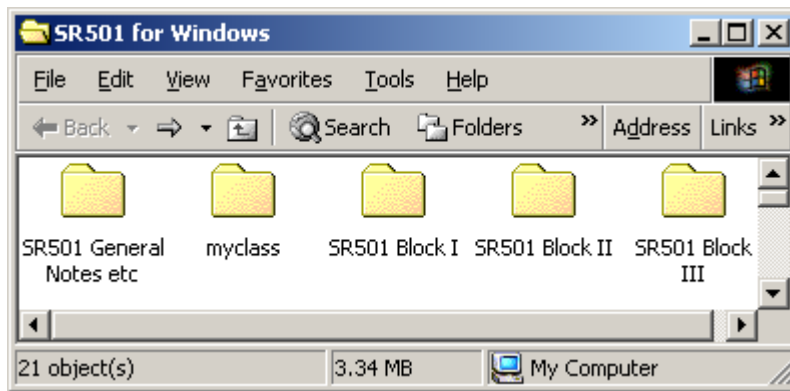
Sub-folder: SR501 General Notes etc
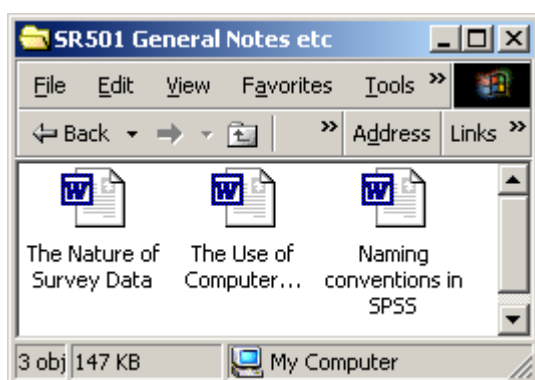
The Nature of Survey Data: originally written by John Hall in 1976 and revised 1990, in response to a constant need to explain to previously non-numerate students, as a general introduction to the idea of responses to questionnaires being tabulated in a two-dimensional **data matrix** defined by **variables** (the columns) and **cases** (the rows). Also serves indirectly as an intro to the language of SPSS (30kb, 3pp including author's note). Incorporated, with extended explanation in ????

The Use of Computers in Survey Analysis: originally written by John Hall in 1971 as an aid to academics and others engaged in survey research for the first time and explaining the process of data cleaning, editing, input and analysis. Somewhat out-dated in parts in the light of current IT and CAPI developments, but useful for beginners and an object lesson for those researchers prone to accept data as given, who then proceed unsuspecting direct to complicated multivariate analysis. (72kb, 12pp including author's note)

Naming Conventions in SPSS: paper written by John Hall and Jim Ring in 1989, based on their intensive experience with SPSS, from 1971 at the SSRC Survey Unit and from 1976 at the PNL Survey Research Unit. It also draws on their extensive research, consultancy, supervisory and advisory work with a wide range of clients and students on the design, processing and analysis of dozens of surveys on various computers, operating sytems and software, to explain the evolution of their recommended convention for naming variables in SPSS, related to the actual questionnaires used, particularly where these may be the only available documentation. (45kb, 7pp)

Sub-folder: **myclass**

class.dat	contains cumulative raw data from the fun questionnaire
myclass1.sav	first edition of SPSS saved file with variable names only
myclass2.sav	second edition of SPSS saved file with variable and value labels added
myclass2.sps	SPSS setup file for variable and value labels

These files relate to the working up of, and screen dumps for, Stage 1 and Stage 2 of the first two hands-on class exercises in generating and reading raw data, adding dictionary information and then saving the work in a system file (See next window, **SR501 Block I**)

Sub-folder: **SR501 Block I**
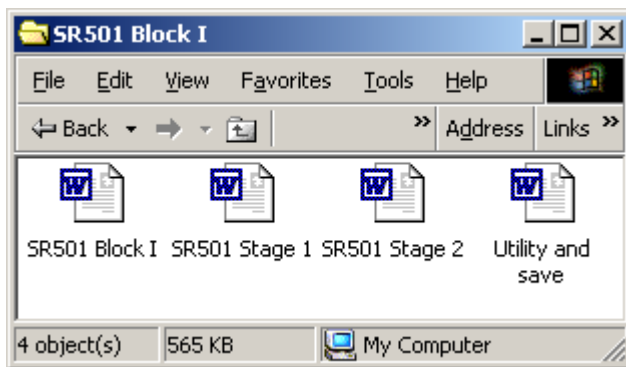
SR501 Block I	Extract from sessions 1-3 of 1991-92 syllabus using VMS and EDT on the Vax to be used as guide to sequencing of materials for Windows version (19kb, 1 page, see below).

**Block I From questionnaire to SPSS-X system file**   (Norusis 1990 Ch 1-6)

1   Data matrix. CASES, VARIABLES, VALUES. Coding of questionnaire data.   Levels of measurement.   The use of computers in survey research. Intro to Vax computer.   Use of computer terminals and printer. Simple VMS commands.   Special keys. Files on the Vax. Demonstration of SPSS-X.   Creating and editing files with the screen editor EDT.   Entering questionnaire responses into a data file.

2   Intro to SPSS-X.   Basic structure of SPSS-X language; commands, sub-commands and specifications. Using SPSS-X to read an external data file.   Records, fields, formats. Naming variables. Dictionary, active file.   Displaying contents of dictionary and active files.

3   Extending a data dictionary. Labelling variables and values.   Missing values. Saving an external system file.

4

SR501 Stage 1: Fun questionnaire, data transfer sheet and explanation of basic SPSS syntax using **commands**, **specifications** and **sub-commands**. Step by step example of data definition for the fun questionnaire in the form of a table to help ready the data for input to SPSS. Step by step displays of SPSS Data Editor windows in **Variable View** and **Data View** mode before starting and after each step. Use of **file… new… syntax…** to open a syntax file and then **data list** to read in the raw data from an external data file **'a:class.dat'**; SPSS utilities **list cases** and **display**; saving active file with **save** or point-and-click **file... save as…**. (345kb, 18pp)

SR501 Stage 2: Retrieving previously saved file and completing data dictionary with **variable labels** and **value labels**; also **missing values,** but none defined for this data set (161kb , 11pp.)

Utility and save: Remaining bits of handouts on utility commands in VMS for the Vax and SPSS commands **set**, **title**, **document** and **display**.
Also **save /drop… /keep…** plus a class exercise, specimen answer and homework exercise. (37kb, 5pp)

Sub-folder: **SR501 Block II**
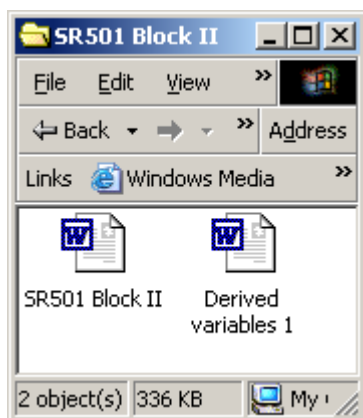
SR501 Block II   Extract from sessions 4-6 of 1991-92 syllabus (Norusis 1990 Ch 7-8,10)

Derived variables 1   Part of data transformation (session 6) changing values with recode, generating new variables with count and compute. (317kb, 19pp)

Block II   One Variable   (Norusis 1990 Ch 7-8,10)

4 Describing data. Univariate distributions. Graphical representations. Retrieving an external system file and selecting variables for analysis (1989 British Social Attitudes survey). Frequencies for nominal and ordinal variables. General and integer mode; treatment of **missing values**; absolute, relative, adjusted frequencies.   Barcharts. Utilities for printing .

SPSS commands   Sub-commands

get …   /keep …   /drop….
missing values
frequencies
frequencies   /barchart

5

5   Frequencies for interval variables. Cumulative percentages. Univariate statistics. Measures of central tendency and dispersion. Histograms, percentiles. Condensed format for variables with many values.

   SPSS commands   Sub-commands

             frequencies
                                    /statistics
                                    /percentiles
                                    /histogram   /normal
                                    /format condense

6   Data transformations. Changing the coding scheme. Derived variables. Selecting cases for analysis.   Conditional frequency distributions.
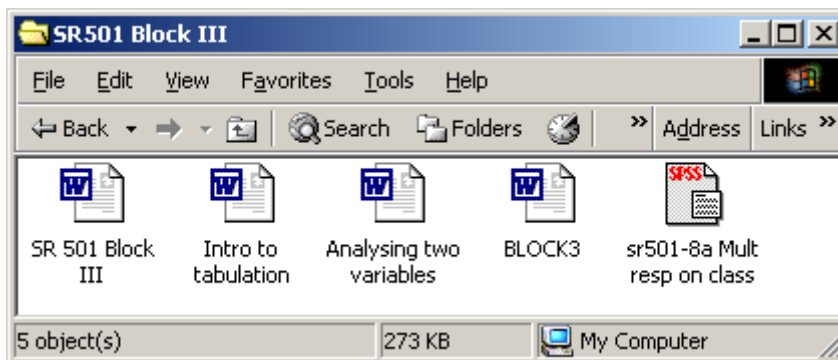
   SPSS commands   Sub-commands

             recode   into
             count
             compute
             temporary
             select if

   Sub-folder: **SR501 Block III**

SR501 Block III: Extract from sessions 7-10 of 1991-92 syllabus using VMS and EDT on the Vax to be used as guide to sequencing and sectioning of materials

Block III Two variables (and sometimes three)   (Norusis 1990 Ch 9,11,13)

Session 7   Joint frequency distributions for two variables. Contingency tables. Dependent and independent variables. Rules for percentaging. Specifying cell contents. Percentage differences (epsilon).

             SPSS commands   Sub-commands

             crosstabs   /variables =
                                        /tables =
                                        /cells =

Session 8   Introducing a third variable. Conditional contingency tables. Controlling for test variables. Elaboration.

Session 9   Handling multiple response questions. Frequencies and contingency tables using multiple response.

          SPSS commands   Sub-commands

          mult response   groups =
                                      /variables =
                                      /frequencies =
                                      /tables =

Session 10 More transformations. Creating simple scales.   Comparing averages across different groups of cases.

          SPSS commands   Sub-commands

          means

Intro to tabulation: General introduction to the idea of tabulation with worked examples starting from simple counts, building up to two-way contingency tables and ending with three-way tabulation (elaboration) including 6 exercises and specimen answers, all on data from actual surveys. **NB.** The main part of this could just as easily go at the beginning as it is not specific to SPSS. (121kb, 30pp)

Analysing two variables: Updated version of part of above using data from the fifth form survey, moving from a simple frequency count of "sexism" via conditional frequency counts to 2-way contingency tables, and then to 3-way tables to illustrate elaboration. **NB.** Again not specific to SPSS, but used to illustrate logic of analysis (34kb, 3pp)

Block3:   Condensed file comprising several sections, some of which appear elsewhere, covering derived variables and means and an introduction to inferential statistics. Worked SPSS examples using **count compute** and **recode…into** and **means** on data from the 1981 fifth form survey and the 1986 British Social Attitudes survey. Also use of sub-command **crossbreak** (not implemented yet for Windows) to tabulate means plus a crafty trick to produce elaboration output using **crossbreak**. Conditional transformations **if** and **do if …. end if**;

Hypothesis testing[5], intro to inferential statistics, **samples** and **populations**, sample **statistics** and population **parameters**; use of **seed** and **sample** to generate individual sub-samples for each student in the class, analysis of which demonstrates variability of estimates of percentages and means.

Final section covers **chi-square**, **t-test**, **oneway analysis of variance**, **linear regression** and **correlation** with worked SPSS examples (99kb, 25pp).

SR501-8a: SPSS setup file to analyse **multiple response** questions in the fun questionnaire.
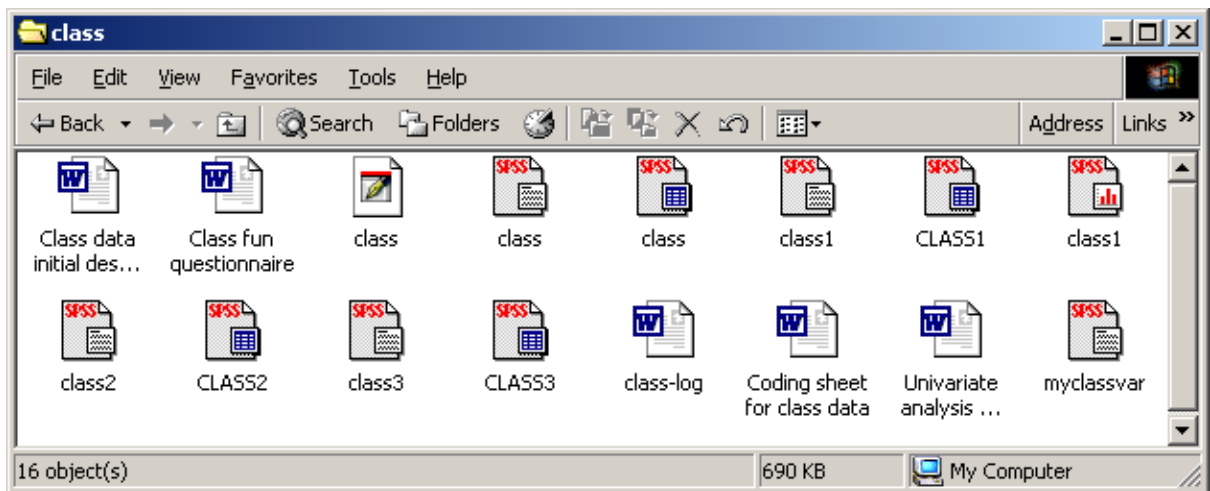
---

[5] The bit on hypothesis testing properly belongs in Block IV, and is contained in file Block4 (13pp) in that window: covers **t-test** and **anova**

Main folder: **Survey Analysis Workshop**
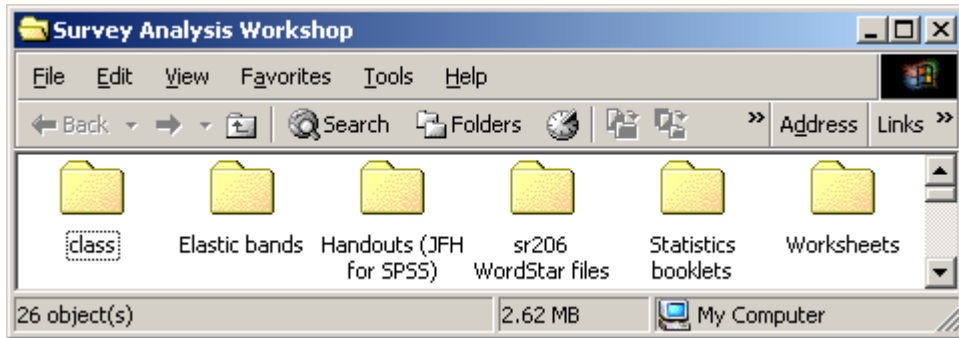
Sub-folder: class

JFH's working folder for generation of SPSS examples from the fun questionnaire

Class data initial descriptions:

Introduces notion of **data matrix** formed by **variables** and **cases**, coding for data entry, naming of variables, step-by-step worked example of use of **data list** to read raw data into SPSS. Includes the fun questionnaire at the beginning, followed by the data transfer sheet. Several screen dumps to demonstrate what everything looks like at each step, output to show what data look like inside the computer and output from simple SPSS commands to list and display file contents. Contents duplicated in **SR501 Stage 1** in folder **SR501 for Windows**, sub-folder **SR501 Block I.** Other files are stages in the build-up of the system file, and are used to display windows from the working files.

Sub-folder: Handouts (JFH for SPSS)

BLOCK4 Hypothesis testing (53.5kb, 13pp)

Intro to inferential statistics; notions of samples from populations, statistics and parameters. Content overlaps with other files and needs breaking up.

SR206-10 Multiple response: (55kb, 17pp)

Explanation of multiple responses in questionnaires and two different ways in which they can be coded in the data set. Production of frequencies and contingency tables in general or dichotomous mode (pp1-12). Worked example from fifth form survey. Notes on multiple response in the 1986 BSA survey

SPSS commands   Sub-commands

mult response   /groups =
                          /variables =
                          /frequencies =
                          /tables =

Second part includes part of the next section on derived variables (pp13-17)

SR206-11 Derived variables: (58kb, 12pp)

See above, but this version has a Vax exercise and conditional transformations as well. Comparing means. Exercises.

SPSS commands   Sub-commands

count
select if
compute
recode[6]
if
do if…   end if   else if ….
means   variables =
                          /tables =
                          /cells =
                          /crossbreak[7]

---

[6] Needs a warning about permanent effects of **recode** if the file is saved.
[7] Not implemented in Windows yet

SR206-12: (25kb, 2pp)   Testing differences between means

     SPSS commands   Sub-commands

t-test
oneway   /statistics
                          /format
                          /range

SR206-13 Chi-square: (21.5kb, 1 page)
Exercise using 1989 BSA data; two-way and three-way tables on attitude to marches and demonstrations; demonstration of effect of sample size on chi-square.

     SPSS commands   Sub-commands

crosstabs   /stat chi
seed

SR206-14 Correlation: (28kb, 3pp)
Correlation and simple linear regression. Correlation matrix example for "sexism" scale in fifth form survey; exercise on "anti-protest" scale in BSA 1989 survey; scatterplot for injury causing road accidents in December 1986 and December 1987 in 41 police authorities.

     SPSS commands   Sub-commands

correlation   /statistics
                          /print
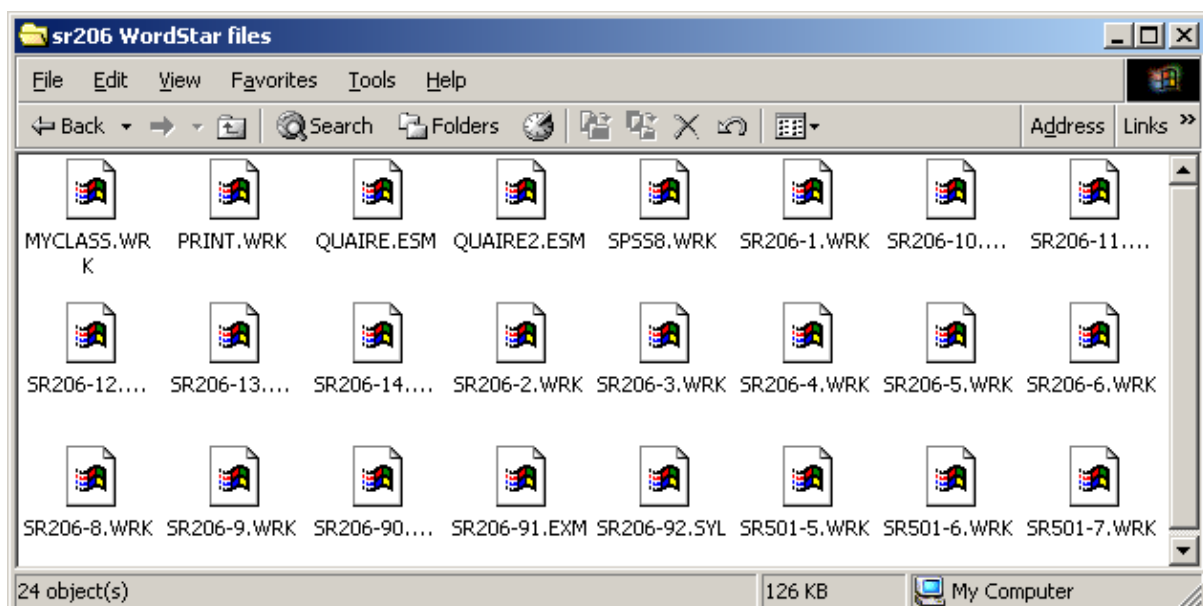**plot**   /format
                          /plot

TESTING t-test oneway: (52kb, 10pp)
Hypothesis testing (duplicates earlier sections) but with exercises, and extra section on **association, structure** and **prediction**. This is not specific to SPSS and needs splitting up.

Sub-folder: sr206 Wordstar files

These files are the original worksheets in WordStar4 format. Most are replicated in MS-Word files elsewhere, except for examinations (Check this)

Sub-folder: Statistics booklets
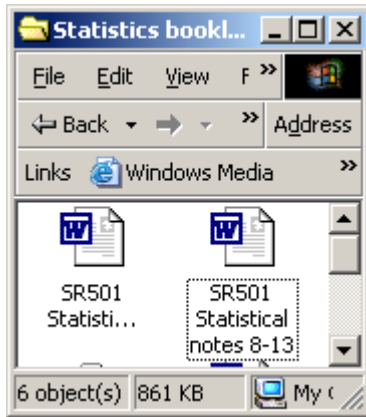
SR501 Statistical notes 1-7 Chapters 1-7 (287kb, 37pp)
SR501 Statistical notes 8-13 Chapters 8-13 (104kb, 26pp)

Originally written by Jim Ring and John Hall in 1988, these notes derive from teaching the statistical component of professional practice oriented courses in survey data collection, data management, computer processing and statistical analysis to social science students at the Polytechnic of North London from 1976 to 1992.

These notes represent an attempt to fill a gap in the textbook provision for students who find computers and statistics daunting, and were mostly written before the appearance of the SPSS Guide to Data Analysis (Norusis, M). They are not intended as a replacement, and should be used in conjuction with the recommended texts.

Sub-folder: Worksheets
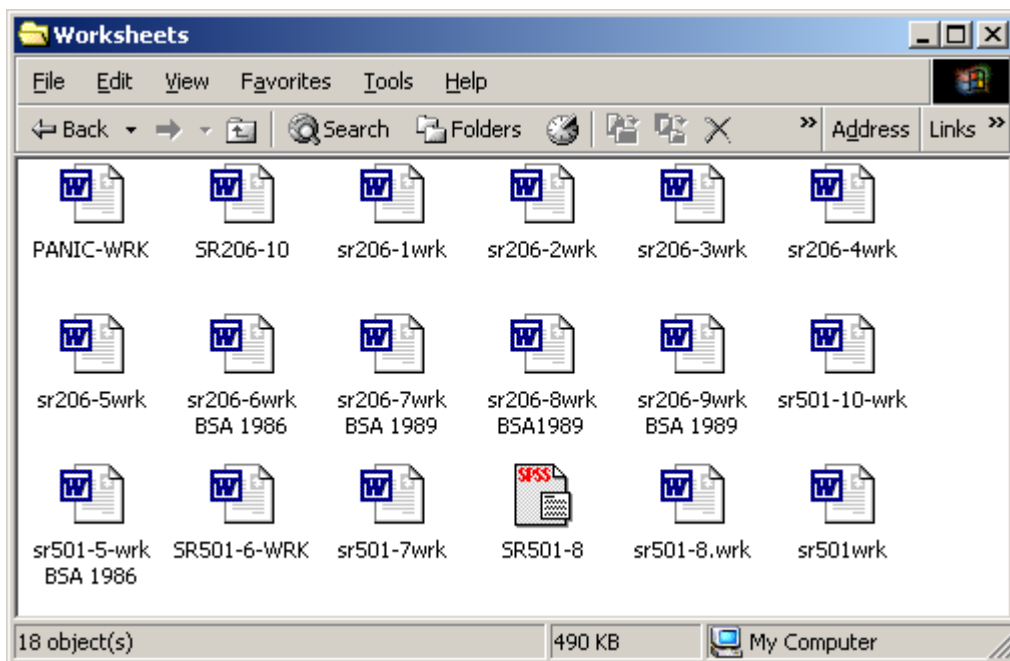
These files are conversions from WS4 or other earlier versions. The numbers refer to course sessions, but are not always the same for each course.. This is partly due to disruptions caused by student occupations, staff strikes and management lock-outs, partly to technical problems with servers and/or terminals, and partly to variations in progress related to student aptitudes and interests (hence the file **PANIC-WRK** when at one point we were about three sessions behind!).

I really need some advice on how to get some files, especially formal syntax handouts, worksheets and cribs, in the form of pop-ups by clicking on boxes or words in the text. Any offers?

hallj@wanadoo.fr

Last updated: 17 Dec 2005