Putting it all Together
Creating Master Data Sets &
Conducting *Many* Statistical Analyses

       Our procedure thus far in the manual has been to provide examples of studies involving only one dependent variable, and only one or two independent variables. Also the researcher in each of the examples thus far was interested in answering only one research question. Thus, each of the examples thus far has called for only one statistical test to be conducted. And the set up of the procedures for conducting that singular statistical test has been provided.

       Most researchers, however, collect information on a multitude of variables for each subject and then put the data into a **master data set**. Then the researcher after conducting the statistical test to answer the **main** research question of interest will sit back and generate a number of supplemental questions that might be answered by analyzing the data in the master data set. For example, a researcher may conduct a study that has two independent variables and has the participants randomly assigned to one of the four conditions generated by that design. The researcher may collect information on one main dependent variable of interest that was measured on a ratio or interval scale. Hence the researcher would conduct a 2 x 2 two way analysis of variance for that main analysis. However, in addition the researcher also may have collected a lot of supplement information on how the participant behaved on other dependent variables. Rather than putting the data into a number of **separate** data files, the researcher puts all of the information for each participant in **one main master data file**. The first step for analyzing the data would be the labeling and identification of each of the variables in the master data set. There may be a multitude of variable labels, one for each column of numbers in the master data set. For example:

       A social psychologist was interested in determining whether people were more likely to be aggressive when frustrated than when not frustrated. The psychologist was also interested in finding out if conditions of anonymity were more or less likely to facilitate people being aggressive if and when they are frustrated. In order to answer this question the social psychologist took a sample of 40 participants and randomly assigned them to one of four experimental treatments. Half of the participants (20) were randomly assigned to a condition where they were given an impossible task to solve (and hence frustrated). The other half of the participants (20) were randomly assigned to a condition where they were given a solvable problem (and hence not frustrated). After encountering the task, the experimenter asked each participant to indicate the extent to which they were angry on a 7 point anger scale that looked like that below. Specifically each participant was asked to place a check mark somewhere along the following scale that best represented the degree of anger they were feeling as a result of working on the problem. The response of participant #1 whose responses are listed in the table on the next page are presented below:

not at all angry:__ __ __ __ __ __ _x_ :very angry

The experimenter scored a check mark adjacent to the not at all angry end of the scale as a 1 and a check mark adjacent to the very angry end of the scale as a 7.

       All participants in the study were then put into a situation where they had an opportunity to evaluate the quality of performance of an experimental confederate. In particular the experimenter informed the participants that they were to provide feedback to another subject (actually a confederate) who had been asked to create a humorous essay. The real participant's task was to provide feedback to the confederate as to the quality of the humorous essay. Feedback was to be given in the form of administering shock if they thought the essay was of poor quality and no shock if they thought the essay was of good quality. Half of the participants when giving the feedback did so under conditions of anonymity and the other half under conditions of no anonymity. The experimenter simply recorded the participant's response on a dichotomous yes/no shock scale and arbitrarily assigned the number 1 if the participant administered shock and the number 2 if the participant did not.

Shock Yes _x_   Shock No __

Again, the responses of participant # 1 are listed in the shock column of the table below. In addition to observing whether the participant shocked or didn't shock the confederate, the experimenter also recorded the **duration** of the shock administered in terms of milliseconds, as well as the **intensity** of the shock administered in terms of volts.

Duration Milliseconds: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15: Duration Milliseconds

Volts: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30: Volts

The data for each participant is recorded in the table below. Thus participant # 1 did administer shock (scored as a 1), for a duration of 5 milliseconds, at an intensity of 20 volts. This participant also indicated that they were very angry (scored as a 7).

Column 1
Anonymous

Column 2
Not Anonymous

Frustrated Row 1

Subject #	Shock	Duration	Intensity	Anger
1	1	5	20	7
2	2	0	0	5
3	1	4	19	6
4	1	5	23	6
5	1	6	10	7
6	1	5	24	5
7	2	0	0	4
8	1	3	15	3
9	1	4	19	6
10	1	7	35	9

Subject #	Shock	Duration	Intensity	Anger
21	1	4	15	6
22	1	1	2	3
23	2	0	0	3
24	2	0	0	2
25	2	0	0	4
26	2	0	0	7
27	1	4	14	4
28	2	0	0	7
29	1	2	6	8
30	2	0	0	5

Not Frustrated Row 2

Anonymous

Not Anonymous

Subject #	Shock	Duration	Intensity	Anger
11	2	0	0	2
12	2	0	0	1
13	2	0	0	2
14	2	0	0	1
15	1	2	10	3
16	1	3	12	4
17	1	2	10	3
18	1	1	7	2
19	2	0	0	1
20	1	0	0	1

Subject #	Shock	Duration	Intensity	Anger
31	2	0	0	2
32	2	0	0	1
33	1	1	7	2
34	2	0	0	1
35	2	0	0	3
36	2	0	0	2
37	2	0	0	1
38	2	0	0	1
39	2	0	0	2
40	1	2	10	4

We
are going to use **all** of the data from this experiment to create one master data file. Then we are going to ask a number of different questions about the data which call for different statistical analyses.

1. To create your master **data** file for the experiment first create a **paper** version template of the data on a sheet of paper that looks like that below.
2. These column of numbers refer respectively to: a) subject #; b) shock response; c) duration of shock; d) intensity of shock; e) anger of subject; f) the row level number of the experimental frustration condition the subject was in where 1 = frustrated condition and 2 = not frustrated condition; and finally g) the column number of the experimental anonymity condition the subject was in where 1 = anonymous condition and 2 = the not anonymous condition.

Subject   Shock   Duration   Intensity   Anger   Frustration   Anonymity

1	1	5	20	7	1	1
2	2	0	0	5	1	1
3	1	4	19	6	1	1
4	1	5	23	6	1	1
5	1	6	30	7	1	1
6	1	5	24	5	1	1
7	2	0	0	4	1	1
8	1	3	15	3	1	1
9	1	4	19	6	1	1
10	1	7	35	9	1	1
11	2	0	0	2	2	1
12	2	0	0	1	2	1
13	2	0	0	2	2	1
14	2	0	0	1	2	1
15	1	2	10	3	2	1
16	1	3	12	4	2	1
17	1	2	10	3	2	1
18	1	1	7	2	2	1
19	2	0	0	1	2	1
20	1	0	0	1	2	1
21	1	4	15	6	1	2
22	1	1	2	3	1	2
23	2	0	0	3	1	2
24	2	0	0	2	1	2
25	2	0	0	4	1	2
26	2	0	0	7	1	2
27	1	4	14	4	1	2
28	2	0	0	7	1	2
29	1	2	6	8	1	2
30	2	0	0	5	1	2
31	2	0	0	2	2	2
32	2	0	0	1	2	2
33	1	1	7	2	2	2
34	2	0	0	1	2	2
35	2	0	0	3	2	2
36	2	0	0	2	2	2
37	2	0	0	1	2	2
38	2	0	0	1	2	2
39	2	0	0	2	2	2
40	1	2	10	4	2	2

The **last two column of numbers** in the data set identify which row (frustration) level and which column (anonymity) level of the experiment the participant is in. The first column of numbers is the participant number, the second column whether shock was administered, the third column the duration of the shock, the fourth column the intensity of shock, and the fifth column the anger level of the participant.

Following a procedure similar to that described on pages 1-3 to create the computerized version of the above **paper** template:

1. Logon to the system
2. **Click Start > Programs > SPSS for Windows > SPSS 10.1 for Windows**. At this point a window will appear asking you what you would like to do. Click on the circle next to Type in Data (2nd option in list) and then click OK at the bottom of the window.
3. A Data Editor will appear. Look in the lower left corner of the screen. You should see a **Data View** tab and to the right of it a **Variable View** tab. The **Variable View** tab will be used first for the Data **Definition** Phase of creating a data file. This tab will be used to define the characteristics of the seven columns of numbers listed above in the **paper** template. The **Data** View tab will be used to actually enter the raw numbers listed above.

DATA DEFINITION PHASE

4. Click on the **Variable View** tab in the lower left corner. A new screen will appear with the following words at the top of each column.

   Name Type Width Decimals **Label Values** Missing Columns Align Measure

Unless otherwise specified in the manual or by your instructor, the only columns we will be using in this manual are the **Name** column, the **Label** column, and the **Values** column.

The **Name** column allows us to put a name at the top of the **computerized** version of our data set. It is, however, limited to no more than **eight** characters. Thus, although the **paper** version of the data set above has the names **Intensity, Frustration, and Anonymity** which are more than eight characters, the **computerized** version of the names will be shortened to **Intens, Frust,** and **Anon**.

The **Label** column allows us to expand the Name of any variable beyond eight characters but only for the computer print out of the results of our analysis. Thus, we can expand the word Intens to Intensity in the Label column and the word Frust to Frustration, and the word Anon to Anonymity. Thus, when we get a computer print out we will have the more informative words of Intensity, Frustration and Anonymity rather than the less informative words of Intens, Frust, and Anon.

The **Value** column allows us to attach meaning to the 1's and 2's for the shock, frustration and anonymity columns so that the computer "knows" that a 1 = "frustrated" and a 2 = "not frustrated" for the frustration column, or that a 1 = "anonymous" and a 2 = "not anonymous" for the anonymity column, or that a 1 = "yes shock was administered" or 2 = "no shock" for the shock column.

5. Click in the white cell in **Row 1** under the word **Name** and type in the word **Subject** (for the subject number) which is the name of the first column of numbers in the **paper** template.
6. Click in the white cell in **Row 1** under the word **Label** and type in Subject Number. (Doing this will provide you with a more expansive label in the computer print out of the results of your analysis.)
7. Click in the white cell in **Row 2** under the word **Name** and type in the word **Shock** (for Shock) which is the name of the second column of numbers in the paper template.
8. Click in the white cell in **Row 2** under the word **Label** and type in **Shock** (Doing this will provide you with a more expansive label in the computer print out of the results of your analysis).
9. Click in the white cell in **Row 2** under the word **Value**. The word none will appear along with a grey box to the right.
   a. Click on the small grey box and a value labels window will appear
   b. In the white box next to the word **Value**, type in the number **1.**
   c. Click on the white box next to the word **Value Label** and type in **Yes Shock**
   d. Click on the **Add** button. 1 = "Yes Shock" should now appear in the bottom white box.
   e. In the white box next to the word **Value** type in the number **2.**
   f. Click on the white box next to the word **Value Label** and type in **No Shock**
   g. Click on the **Add** button. 1 = "Yes shock" and
      2 = "No shock" should now appear in the bottom white box.
   h. Click **OK** button (top right corner of screen)
10. Click in the white cell in **Row 3** under the word **Name** and type in the word **Duration** which is the name of the third column of numbers in the **paper** version of the data template.
11. Click in the white cell in **Row 3** under the word **Label** and type in **Shock Duration**. (Doing this will provide you with a more expansive label in the computer print out of the results of your analysis.)
12. Click in the white cell in **Row 4** under the word **Name** and type in the word **Intens** (for Intensity) which is the name of the fourth column of numbers in the paper template.
13. Click in the white cell in **Row 4** under the word **Label** and type in **Intensity.** (Doing this will provide you with a more expansive label in the computer print out of the results of your analysis).
14. Click in the white cell in **Row 5** under the word **Name** and type in the word **Anger** which is the name of the fifth column of numbers in the **paper** version of the data template.
15. Click in the white cell in **Row 5** under the word **Label** and type in **Anger Level**. (Doing this will provide you with a more expansive label in the computer print out of the results of your analysis.)
16. Click in the white cell in **Row 6** under the word **Name** and type in the word **Frust** (for Frustration) which is the name of the sixth column of numbers in the paper template.
17. Click in the white cell in **Row 6** under the word **Label** and type in **Frustration** (Doing this will provide you with a more expansive label in the computer print out of the results of your analysis).

18.   Click in the white cell in **Row 6** under the word **Value**. The word none will appear along with a grey box to the right.

        a.   Click on the small grey box and a value labels window will appear

        b.   In the white box next to the word **Value**, type in the number **1.**

        c.   Click on the white box next to the word **Value Label** and type in **Frustrated**

        d.   Click on the **Add** button. 1 = "Frustrated" should now appear in the bottom white box.

        e.   In the white box next to the word **Value** type in the number **2.**

        f.   Click on the white box next to the word **Value Label** and type in **Not Frustrated**

        g.   Click on the **Add** button. 1 = "Frustrated" and
                            2 = "Not Frustrated" should now appear in the bottom white box.

        h.   Click **OK** button (top right corner of screen)

19.   Click in the white cell in **Row 7** under the word **Name** and type in the word **Anon** (for Anonymity) which is the name of the seventh column of numbers in the paper template.

20.   Click in the white cell in **Row 7** under the word **Label** and type in **Anonymity** (Doing this will provide you with a more expansive label in the computer print out of the results of your analysis).

21.   Click in the white cell in **Row 7** under the word **Value**. The word none will appear along with a grey box to the right.

        a.   Click on the small grey box and a value labels window will appear

        b.   In the white box next to the word **Value**, type in the number **1.**

        c.   Click on the white box next to the word **Value Label** and type in **anonymous**

        d.   Click on the **Add** button. 1 = "anonymous" should now appear in the bottom white box.

        e.   In the white box next to the word **Value** type in the number **2.**

        f.   Click on the white box next to the word **Value Label** and type in **Not anonymous**

        g.   Click on the **Add** button. 1 = "Anonymous" and
                            2 = "Not anonymous" should now appear in the bottom white box.

        h.   Click **OK** button (top right corner of screen)

DATA ENTRY PHASE

22.   Click on the **Data View** Tab in the lower left corner. The data **view** screen will now appear with **Column 1** named **Subject** (for the Subject Number), **Column 2** named **Shock**, **Column 3** named **Duration**, **Column 4** named **Intens** (for Intensity), **Column 5** named **Anger**, **Column 6** named **Frust** (for the Frustration independent variable), and **Column 7** named **Anon** (for the Anonymity independent variable).

23.   Enter the data for all **40 subjects** listed in the paper template as follows. Click the white cell at Row 1, Column 1 under **Subject** and enter:

      1 tab 1 tab 5 tab 20 tab 7 tab 1 tab 1   then mouse to the second row and enter the responses of the second subject.

      2 tab 2 tab 0 tab 0 tab 5 tab 1 tab 1   then mouse to the third row and enter the responses of the third subject etc.

Follow the same procedure to enter the responses of the remaining subjects. The data could also be entered down one column at a time and then moving on to the second column etc.

Data Analysis

Whenever a researcher collects data and records the data in a data set, other researchers may come along and raise research questions that may be answered by exploring the data set. Listed below are just a few of a series of questions one might pose about the above data set. Along with each question are the data analysis procedures one would use to answer that **specific** question and a reference to the location in the manual relevant to the statistical procedure used.

Question 1:   Are people more likely to administer shock when they are frustrated than when not frustrated? This question calls for a chi-Square test. See Page ii and Pages 58-60.

1. Click on **Data** at top of screen then
   a. Click on **Weight Cases** then
   b. Click on **Circle** by Do **Not** Weight Cases (a dot will appear) then
   c. Click on **OK**
2. Click **Analyze** at top of screen then
   a. Click on **Descriptive Statistics** then
   b. Click on **Crosstabs**
3. Highlight **Shock** by clicking on it then
   a. Click on **arrow** > to transfer this name to the **Row(s)** box
4. Highlight **Frust** by clicking on it then
   a. Click on **arrow** > to transfer this name to the **Column(s)** box
5. Click on **Statistics** button at bottom of screen, then
   a. Click on white square next to **Chi-square**. This will put a check in the box.
6. Click **Continue** button
7. Click **Cells** button at bottom of screen, then
   a. Click on white squares next to Observed and Expected to put a check mark in the box if one is not already there then
8. Click **Continue** button
9. Click **OK.** Doing this will result in analysis being conducted

Question 2   Is there a relationship between the intensity of the shock administered and the duration of the shock? This question calls for a Pearson's Correlation. See Page i and Pages 10-12.

1. Click on **Analyze** at top of screen then
   a. Click on **Correlate** then
   b. Click on **bi-variate**
2. Highlight the word **intens** by clicking on it and then
   a. Click on **arrow >** to transfer this name to the **Variables** box
3. Highlight the word **duration** by clicking on it and then
   a. Click on **arrow >** to transfer this name to the **Variables** box
4. Make sure there is a check mark in the small white box next to the word **Pearson**. If not click on the small white box and a check mark should appear
5. Click on **OK.** Doing this will result in analysis being conducted.

Question 3   Do the anonymous people administer shocks of a longer duration than the people who are not anonymous? This question calls for an independent groups t-test. See Page i and Pages 4-6.

1. Click on **Analyze** at top of screen then
   a. Click on **compare means** then
   b. Click on **Independent Samples t-test**
2. Highlight the word **Anon** by clicking on it and then
   a. Click on the **lower** arrow to transfer it to the **grouping** variable box.
      This is your independent variable in the question.
   b. When the **anon ??** shows up click on the **Define Groups** box
   c. Type in **1** in Group 1 box
   d. Type in **2** in Group 2 box
   e. Click the **Continue** button
3. Highlight the word **duration** by clicking on it and then
   a. Click on the **upper** arrow to transfer it to the **Test Variable** Box
      This is your dependent variable in the question.
4. Click **OK** Doing this will result in analysis being conducted.

Question 4   Does the participant's level of anonymity interact with their level of frustration to influence their level of expressed anger? This question calls for a 2 x 2 ANOVA since there are two independent variables (anonymity and frustration) in the question and the dependent variable in the question is measured on an interval scale. See page ii and Pages 37-41.

1. Click on **Analyze** at top of screen then
   a. Click on **General Linear Model** then
   b. Click on **Univariate**
2. Highlight the word **anger** by clicking on it then
   a. Click on **arrow >** to transfer this name to the **Dependent** Box
3. Highlight the word **Anon** by clicking on it then
   a. Click on **arrow >** to transfer this name to the **Fixed Factor(s)** Box.
4. Highlight the word **Frust** by clicking on it then
   a. Click on **arrow >** to transfer this name to the **Fixed Factors(s)** Box.
5. Click on **Options** Button
   a. Highlight the word **Anon** by clicking on it then
   b. Click on **Arrow >** to transfer this name to the **Display Means Box**.
   c. Highlight the word **Frust** by clicking on it then
   d. Click on **Arrow >** to transfer this name to the **Display Means Box**.
   e. Highlight the words **anon*frust** by clicking on it then
   f. Click on **Arrow >** to transfer this name combination to the **Display Means Box**.
6. Click on white square next to the words **descriptive statistics** to put a check mark in the box
7. Click **Continue**
8. Click **OK** Doing this will result in analysis being conducted.

Question 5   **Within** the condition where participants are **both** frustrated and anonymous, is there a relationship between the level of anger expressed and the intensity of shock administered. This question calls for a Pearson's Correlation. See Pages i and Pages 10-12. However, the analysis will be performed **only** on the responses of the participants who are **both** frustrated and anonymous.

---

1. Click on **Data** at top of screen then
   a. Click on **Select Cases**
2. Click on 2nd Circle which is adjacent to **"If Condition is Satisfied"**
3. Click on **If** button
4. In white box type **Anon = 1 and Frust = 1**
5. Click **Continue** button
6. Click **OK**
7. Click on **Analyze** at top of screen then
   a. Click on **Correlate** then
   b. Click on **bi-variate**
8. Highlight the word **intens** by clicking on it and then
   a. Click on **arrow >** to transfer this name to the **Variables** box
9. Highlight the word **anger** by clicking on it and then
   a. Click on **arrow >** to transfer this name to the **Variables** box
10. Make sure there is a check mark in the small white box next to the word **Pearson**. If not click on the small white box and a check mark should appear
11. Click on **OK.** Doing this will result in analysis being conducted.

Notice in the box above, Steps 1 through 6 are the procedures used in order to select specific cases of interest for conducting an analysis. Not all cases need to enter into an analysis. Thus in the steps 1 though 6 above, the researcher is limiting his/her analysis to the responses of subjects 1 through 10 on page 83. These are the subjects who are **both** frustrated and anonymous.