

Guide to the Group Project Data Entry Plan

**Due 3/27 (10 points)**

Each group must submit a Code Book and SPSS data file (pre-data entry), which must be approved prior to data entry.

Codebook

- A codebook is an outline of all the variables that you are collecting. It tells you the **type of data** (discrete vs. continuous) you are working with, the **Variable Names** you will use, the **Variable Labels** you will use, the **Value Labels** you will use, the **range of scores** possible, and lists any special treatments required (e.g. which items need to be **reverse scored** later, **questions that had problems** during data entry, or **missing data** problems).

- Note: It is best to keep **Variable Names** short (**8 or fewer characters** - See the Introduction to SPSS 14 Guide for rules on variable names.

- **Variable Labels** can be longer and more descriptive. Just remember to put the **unique info at the beginning** of the label as it gets trimmed off in the spss windows and output: all the labels will look alike and it is a real pain. For questionnaires you do not need to give a label for every item. Label the first question but do not label them all. It is redundant and time consuming.

- You can create these in Word as Tables or in Excel as spreadsheets.

- The order you list your variables should reflect the order in which they will be entered into SPSS. I would suggest the following order: subject number, demographics, IVs, then DVs.

Survey Study Codebook Example

Questionnaire	Item #	Variable Name	Variable Label	Value Label / Range	Comment / Reversal
Participant #		parnum	Participant Number	n/a	P 15 and 22 did not complete CPL
Demographics	1	age	Age of Participant	18+	
	2	sex	Sex of Participant	1=Female 2=Male	
	2	ethnic	Ethnicity of Participants	1=Euro-Am 2=Af. Am. 3=Asian/PI-Am 4=Latino/Hispanic 5=Am. Indian 6=Other	
	2	class	Class Rank / Year in School	1= Freshman 2= Sophomore 3= Jr. 4= Sr.	
Aggression Quest	1	aq01	Item 1 - Aggression Questionnaire	1-7	
	2	aq02		1-7	
	3	aq03		1-7	reverse
	4	aq04		1-7	
Cheezy Poof Liking Quest	1	cp101	Item 1 - Aggression Questionnaire	1-5	reverse
	2	cp102		1-5	Typo in question - maybe bad data
	3	cp103		1-5	

### Experiment Codebook Example

Variable	Item #	Variable Name	Variable Label	Value Label	Comment / Reversal
Participant #		parnum	Participant Number	n/a	P 24 quit after first 10 min.
Demographics	1	age	Age of Participants	18+	
	2	sex	Sex of Participants	1= Female 2= Male	
IV1 - Video Treatment		vid_treat	Video Treatment Condition: 10 hrs of either South Park, Barney, or No Video Control	0= Control 1= South Park 2= Barney	Control group had high attrition rates may need to be dropped
IV2 - Snack Treatment		snac_trt	Snack Treatment Condition: Snack Prior to watching vs. No Snack Prior to watching	0= No Snack 1= Snack	
DV - Eric Cartman Resemblance Ratings	1	ecr-rat1	Rater 1 : Eric Cartman Resemblance Ratings	1-9	
	2	ecr-rat2	Rater 2 : Eric Carman Resemblance Ratings	1-9	

Note that I used underscores in the Variable names but not dashes or spaces (you cant use either)

### Content Analysis Codebook Example

Variable	Item #	Variable Name	Variable Label	Value Label	Comment/Reversal
Show Number		snum	Show Number	n/a	Show 16 is missing last 15 minutes - data needs to be dropped.
Show Name		sname	Show Name	n/a	This is a string variable
Training Show		train	Show Uses for Coder Training: All 3 coders rated these shows	0= no 1= yes	
Main Coder		coder-m	Main Coder	1= Timmy 2= Jimmy 3= Mr. Hat	
Secondary Coder		coder-2	Secondary Coder	1= Timmy 2= Jimmy 3= Mr. Hat	
show type (IV)		stype	Show Type (Oxygen Chanel, MTV, Spike)	1= Oxygen 2= MTV 3= Spike	
DV - Anger Expression - Coder1	1	ae_1_1	Coder 1 - 1 <sup>st</sup> half of show :Frequency of Anger Expression	0+	
	2	ae_1_2	Coder 1 - 2 <sup>nd</sup> half of show :Frequency of Anger Expression	0+	
DV - Anger Expression - Coder 2	1	ae_2_1	Coder 2 - 1 <sup>st</sup> half of show: Frequency of Anger Expression	0+	
	2	ae_2_2	Coder 2 - 2 <sup>nd</sup> half of show: Frequency of Anger Expression	0+	
DV - Anger Expression - Coder 3	1	ae_3_1	Coder 3 - 1 <sup>st</sup> half of show: Frequency of Anger Expression	0+	

	2	ae_3_2	Coder 3 - 2 <sup>nd</sup> half of show: Frequency of Anger Expression	0+	
--	---	--------	---	----	--

## SPSS - Data File

- Create an SPSS data file to enter your data in set up all the variables (Names, Labels, Value Labels, etc). You may want to get your codebook approved before you spend a lot of time creating this.
- See the Introduction to SPSS 14 for a refresher on how to use the Data Editor in SPSS. However, here is a very brief overview.
  - Access SPSS through : Start - programs - Course Software - Math Stat - SPSS 14.
  - Remember there is a Data View and a Variable View : the little tabs in the bottom left corner switch back and forth between the two. The tabs sometimes disappear, if this happens just minimize spss and then maximize it: they will come back.
    - You create and edit variables in the Variable View - Rows are different variables and columns represent aspects of the variable (name, type, label, etc.)
    - You enter data in the Data View - Rows are different cases (participants, shows, events, etc.) and the column are different variables (which you created in the variable view).
  - Be sure to save your data as you work, SPSS (like all programs) can crash and loose data.

## Data Cleaning Plan -

- You need to decide how you will check your data once it is entered.
- There are two general methods
  - 1) Enter all the data into a single data file and then two people work together: 1 reads off the scores from the data sheets (questionnaires) and the other checks the score in SPSS. A new master data file is created from the raw data file. Corrections are made directly to the master data file as you check.
  - 2) Two people enter all the data into separate (but identical) files and then we transfer those raw data files to Excel - Subtract one raw data file from the other and anything that is not a 0 was mis-entered in one file (I will walk you through this step). One raw data file is used to create a master file and corrections are made to the master.
- Both methods have their strengths and drawbacks
  - 1) is good if the team has time to meet, but it can be very tedious and fatigue can result in missing mistakes.
  - 2) is good if the team can not meet, as it can be done separately. It does require more data entry, but the result is far less error prone and checking step is quite quick.
- Decide what you want to do and let me know when you turn in your data files.

## Submitting your Data Entry Plan -

- Give me a hard copy of your code book along with a copy of your latest protocol which should include all of the measures/data sheets.
- Email your final data file to me and indicate your planned method of data checking and how you have divided up the work (who will do what).