Statistical Consulting Service
Codebook Guidelines

Codebooks, which provide technical descriptions of the datasets, should be provided with the datasets. If submitting the data in Microsoft Excel or as a text file, a Codebook can be provided in a separate file (for example, in Microsoft Excel or Microsoft Word). When using certain statistical software, such as SPSS, the codebook is integrated into the dataset file – provided the client has appropriately prepared the dataset. In this case, the Variable View must be populated with the information indicated below.

Codebook Guidelines

Codebooks can contain various details about the dataset. Typically, for a SCS project, the Codebook should be a neat table containing the following information:

- **Variable name**

- Variable **description** or label

- **Data type**, for example, text, numeric (count), numeric, date

- **Units/Range/Format**

- **Codes** for categorical variables, providing the codes used and their meaning; for example, for a Likert scale variable one could specify 1-Strongly Disagree, 2-Disagree, 3-Neutral, 4-Agree, 5-Strongly Agree

- **Calculations** used to derive 'constructed variables' contained in the dataset, or notes on the relationship between variables

Example of Dataset (Screenshot of Microsoft Excel Dataset)

This hypothetical (unusually small) dataset is for the study of Outcome two weeks after treatment initiation, for youth, in a hospital ward, with a particular disease. The role of certain variables (such as BMI and Gender) is also to be considered, and the symptoms of the patients one week after treatment are to be described.

Patient_ID	Doctor_ID	Start	Gender	Age	Weight	Height	BMI	Symptoms	Fever	Weightloss	ShortBreath	Other	OtherType	Outcome
AS002	T	2015/05/26	M	18	45	150	20.00	1	1	0	1	0		1
AT015	S	2014/07/23	F	12	37	98	38.53	1	1	0	0	1	2	1
BR002	S	2014/12/10	F	14	42	104	38.83	1	0	0	1	0		2
BR005	T	2015/03/14	M	10	28	105	25.40	1	0	0	1	0		2
SX014	V	2015/04/01	M	14				0	0	0	0	0		1
RT058	S	2015/08/12		8	12	75	21.33	0	0	0	0	0		1
AM005	Th	2014/11/08	F	9	16	80	25.00	0	0	0	0	0		1
OL058	Th	2014/09/05	F	14	25	109	21.04	0	0	0	0	0		3
OL035	V	2014/11/23	M	16	34	152	14.72	1						3
OL057	S	2015/01/01	M	21	55	142	27.28	1	0	0	0	1	1	1
RT012	V	2015/08/07	M	18	70	120	48.61	0	0	0	0	0		1
AT014	V	2015/12/12	M	17	54	160	21.09	1	0	0	1	0		1
AR099	Th	2014/04/16	M		81	139	41.92	1	1	1	1	0		2

Example of Corresponding Codebook (Screenshot of Microsoft Excel Codebook)

Name	Label	Type of data	Units/range/format	Codes	Notes
Patient_ID	Patient identifier (folder number)	Text			Each row has a unique patient
Doctor_ID	Primary doctor identifier	Text		(the IDs are used to protect doctor identies)	(Will not be used for analysis)
Start	Date of treatment initiation	Date	YYYY/MM/DD		(Will not be used for analysis)
Gender	Gender	Text category		M = Male F = Female	
Age	Age in years	Numeric	Whole number, 8-21 inclusive (youth ward only allows these ages)		
Weight	Weight in kilograms (kg)	Numeric	Positive, expected range 10-100 inclusive		
Height	Height in centimeters (cm)	Numeric	Positive, expected range 50-200 inclusive		
BMI	Body mass index (kg / m^2)	Numeric			Calculated field: Weight / (Height/100)^2
Symptoms	Symptoms one week after treatment initiation	Numeric category (binary)	0/1	0 = No 1 = Yes	If this is 0, Fever-Other should all be 0. If this is 1, then Fever-Other are used to record the particular types of symptoms.
Fever	Fever one week after treatment intitation	Numeric category (binary)	0/1	0 = No 1 = Yes	
Weightloss	Weightloss one week after treatment initiation	Numeric category (binary)	0/1	0 = No 1 = Yes	
ShortBreath	Shortness of breath one week after treatment initiation	Numeric category (binary)	0/1	0 = No 1 = Yes	
Other	Any other symptoms one week after treatment intitation	Numeric category (binary)	0/1	0 = No 1 = Yes	
OtherType	Type of any other symptoms one week after treatment intitation	Numeric category	0/1/2	0 = Likely related to disease 1 = Likely related to medication side-effects 2 = Unrelated symptoms	If Other is 1, this will be populated if known.
Outcome	Treatment outcome two weeks after treatment initiation	Numeric category (ordinal)	0/1/2/3	0 = Death 1 = Deterioration / no change 2 = Some recovery 3 = Full recovery	