

Microsoft Excel for Data Management

Dr. K. A. Korb
University of Jos

Outline

- Data Processing
- Data Checking
- Calculating Total Scores for Variables

Data Processing: Introduction

- *It is more from carelessness about truth than from intentionally lying that there is so much falsehood in the world.* Samuel Johnson
- *Garbage in, Garbage out.* Proverb
- One thoughtless mistake in data processing and analysis can result in an entirely incorrect conclusion for the study
- Check, double-check, triple-check, and quadruple-check all steps in the data analysis process

Data Processing

- Step 1: Number the completed questionnaires
 - Rationale: Link data into the computer to the actual questionnaire
 - Generally use a letter related to the data, then 3 numbers, e.g., D001, D002, D003...D010, D011...
- Step 2: Code the data in Excel
 - Create a matrix with each questionnaire item across the top (column), and each completed questionnaire going down (row).

1	5/No	Gen	Age	SchTeach	EduComp	YrsTeach	N/P	Lev	CurN/P	DayEng	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17		
2	P001	2	48	1	4	26	1		6	0	0	2	4	5	5	2	5	2	5	2	5	5	5	5	5	4	5	5	
3	P002	2	z	1	3	10	1	456	1	3	5	6	5	5	4	2	6	5	5	6	5	6	5	6	5	z	2	2	
4	P003	z	46	1	3	9	1		6	1	0	3	2	5	4	1	5	2	2	6	4	5	2	4	3	2	4	5	
5	P004	1	30	2	3	7	0	z		0	z	6	5	4	5	4	4	6	6	5	5	6	6	5	6	6	5	5	
6	P005	2	40	1	3	14	1	z		1	3	6	5	6	1	2	5	1	6	2	6	6	4	5	4	2	5	1	
7	P006	1	30	1	2	3	1		5	0	3	5	4	5	6	4	4	6	5	6	6	5	4	5	4	z	5	1	
8	P007	1	30	1	2	5	0	z		0	0	5	5	6	6	5	5	4	5	2	6	5	5	4	5	5	6	5	
9	P008	z	50	1	2	24	1	N3,123456	1	5	2	1	4	5	5	4	5	5	5	4	6	2	5	3	5	4	3		
10	P009	z	29	1	3	3	z		5	0	z	6	5	6	5	z	6	1	6	6	6	2	5	6	5	2	5	6	
11	P010	z	38	1	4	19	0	z		0	5	6	5	6	6	6	5	5	6	6	6	6	4	6	5	6	5	5	
12	P011	z	38	1	4	15	1	N1,N2,126	0	2	6	5	5	5	4	1	4	6	5	4	6	5	6	5	6	5	6	5	3
13	P012	1	49	1	2	14	1	9	0	4	6	5	5	6	5	6	5	2	2	5	2	2	5	2	4	4	4	4	
14	P013	1	42	1	4	10	1	123456	1	5	5	4	4	5	2	2	1	6	4	3	6	5	2	5	4	1	6		
15	P014	2	44	1	4	17	0	z		z	5	5	3	2	5	6	3	4	6	6	3	6	6	5	4	5	6		
16	P015	1	44	1	2	12	1	123456	1	0	1	5	6	5	1	3	2	3	4	5	5	6	6	6	5	4	3		
17	P016	z	42	1	4	17	1		6	1	5	6	4	4	5	1	2	4	5	1	4	4	4	4	4	4	3		
18	P017	z	47	1	5	1	ALL		1	0	1	2	4	5	1	5	1	6	2	6	6	1	6	5	2	5			
19	P018	z	54	1	4	26	1	1246	0	0	5	5	6	6	2	5	3	4	6	5	6	4	5	3	z	5	4		
20	P019	z	40	1	6	14	1	N1N212456	0	5	6	6	6	5	2	2	4	6	5	5	6	5	5	3	6	6	5		
21	P020	1	35	1	4	16	1		6	1	5	5	2	5	6	1	2	1	5	1	5	5	2	3	6	5	4	2	
22	P021	1	43	4	4	8	1	12345	1	5	6	6	6	6	3	6	6	6	3	6	6	6	6	5	6	5	5		
23	P022	1	45	1	3	12	1		1	4	6	5	4	6	5	1	5	6	4	1	6	4	5	4	6	4	5		
24	P023	1	34	13	2	6	1	N3,123456	0	3	6	6	1	3	1	2	6	6	6	6	6	6	3	6	6	3			
25	P024	z	27	1	2	3	1	3z		z	5	4	6	5	1	4	4	6	6	5	5	5	6	6	5	6	5		
26	P025	z	29	1	2	8	0	z		0	0	4	5	2	6	3	5	2	5	6	4	6	6	2	4	5	1	2	
27	P026	1	26	z	2	6	0	z		0	z	5	5	6	5	4	1	2	6	5	5	2	5	5	5	5	2		

Tips for Coding Data

- *Coding the Data:* Each response is given a number or letter code and entered into the computer
- Make a reminder sheet of the codes given to each response for future reference
 - Some responses are already given a code on the questionnaire (**a** for Male and **b** for Female)
 - If a code is not already given, use the first letter of the word (**M** for Males and **F** for Females)
 - Be careful that each response gets a unique code
 - Yes is coded as 1; No is coded as 0
 - For Likert-Scale items, enter the number of the response

Tips for Coding Data

- Be **very** careful when doing data entry to prevent computer typos
 - Incorrectly entering data is careless research and is therefore unethical
 - Double-check every questionnaire to ensure the correct responses are indicated in the correct columns

Missing Data

- Missing an item on the questionnaire
 - Discard the item for that particular participant
 - Enter **Z** for missing items (Microsoft Excel)
 - Do not leave empty cells
- Items where the response is unclear (e.g., circled two responses)
 - It is typically best to enter that item as missing, Z.
 - Possible Exception: If they circled 3 and 4, then you can enter 3.5

Missing Data

- Participants do not complete an entire portion of the questionnaire
 - Before the study, set criteria about when to discard data
 - *Sample criteria:* If all items for one variable are skipped, then discard that participant
 - Keep track of how many are discarded and why
 - Report this in the Methods

Missing Data

- Participants only complete one part of the study (in experiments)
 - Consider how to get their data with alternate procedures that do not confound the research design
 - If the data cannot be recovered or if they missed a substantial portion of the treatment, discard the participant from the entire study
 - Report in the Methods how missing a phase of the study was handled and how many were excluded because they did not complete the experiment

Reverse-Coded Items

- **Reverse-Coded Items:** Items that say the *opposite* of other items
 - Assign the opposite code to the response.
 - Agreeing to "I dislike school" is actually Disagreeing to the overall Enjoyment of School
 - Strongly Disagreeing to "I dislike school" is actually Strongly Agreeing to Enjoyment of School
 - Reverse coding is typically used to control for acquiescence bias
- It is easiest and more accurate to enter items exactly as they are completed on the questionnaire and then recode later
 - Recoding can be done by the *If* function in Microsoft Excel
 - SPSS has a recoding function

Reverse-Coded Items

	Typical Code I like Maltina	Reverse Code I do NOT like Maltina
Strongly Agree	4	1
Agree	3	2
Disagree	2	3
Strongly Disagree	1	4

Typical Item:	Strongly Disagree	Disagree	Agree	Strongly Agree
Reverse Coded Item:	Strongly Agree	Agree	Disagree	Strongly Disagree

Step 3: Data Checking

- After all data has been entered, run a few tests to ensure that there are no typographical errors
 - Check the maximum and minimum values for each numerical item
 - If the maximum/minimum is outside of the possible range of values, there is an error
 - Calculate the frequency of categorical values
 - Ensure the sum of frequencies adds up to the total number of participants

Step 4: Calculating Scores for Each Variable

- Total scores are needed for each variable
- Items that measure each variable are likely distributed across the questionnaire.
 - Identify which items measure each variable
 - Reverse-code items as necessary
- To avoid problems associated with missing items, calculating the mean (average) of the items that measure each variable is generally best

Calculating Total Scores

- =average()
 - If the items for one variable are together, highlight the range of items within the brackets
 - If the items are spread apart, click on each cell and separate by commas