THE SEVENTH CRIDAL WORKSHOP

Topic: Quantitative Data Analysis (Intermediate)

Date : 21 September (Monday) or
22 September (Tuesday)
Time : 2:30 p.m. - 5:30 p.m.
Facilitators: Dr. Zhang Wei-yuan, Research Fellow (CRIDAL)
Mr. Clarence Fong, Data Analyst (CRIDAL)

This is a series of workshops on quantitative data analysis using *Statistical Package for the Social Science (SPSS) for Windows*. This workshop is divided into the three sessions as follows:

Session one: 2:00 – 2:30 p.m.

Individual help for participants who need to learn how to use SPSS or who have questions about the previous workshop .

Session two: 2:30 – 3:30 p.m.

Some basic statistical concepts

- Mean, standard deviation and distribution
- Normality assumptions
- Type I and II errors in hypothesis testing
- Interpretation of critical values and p-values

Session three: 3:30 – 5:00 p.m.

Using SPSS

- Reliability Analysis
- Non-Parametric Techniques
- Multiple Response and Multiple Dichotomy analysis

Recommended reading:

- Coakes, S J & Steed, L G (1999**) *SPSS analysis without anguish: versions 7.0, 7.5, 8.0 for windows,*** Milton, Australia: John Wiley & Sons Australia, Ltd.
- Gravetter, F J & Wallnau, L B (1996) *Statistics for the behavioral sciences (fourth edition),* MN, USA: West Publishing Company.

Lesson 6: Reliability Analysis

- Reliability means consistency. It is the degree to which an instrument will give similar results for the same individuals at different times.
- Reliability can take on values of 0 to 1.0, inclusive.

Methods for checking Reliability

Test-retest reliability

The calculation of test-retest reliability is straightforward. The same test is administrated on two occasions to the same individuals under the same conditions. This yields two scores for each person and the correlation between these two sets of scores is the test-retest reliability coefficient. If the test is reliable, there will be a high positive association between the scores.

- Exercise: The scores of 20 students in language proficiency test and retest

Student	Test	Retest
1	94	96
2	92	87
3	88	91
4	87	86
5	87	89
6	86	86
7	85	89
8	95	91
9	85	84
10	83	86
11	82	84
12	81	77
13	78	81
14	76	71
15	72	76
16	68	72
17	66	66
18	65	72
19	63	59
20	58	55

- **Inputting data as above**

To conduct a reliability analysis

- **Statistics – Correlate – Bivariate - Pearson - flag**
- **Result: r = 0.945**

Split half

Only need one administration. The test items are divided into two halves, with the items of the two halves matched on content and difficulty.

Exercise: Interest Inventory (RIASEC)
Five-point scale:

- very much like me 5
- somewhat like me 4
- neither like nor unlike me 3
- somewhat unlike me 2
- very much unlike me 1

Social type

1. Easy to talk with all kinds of people
2. Good at explaining things to others
3. Enjoying working as a neighbourhood organiser
4. Teach children easily
5. Teach adults easily
6. Help people who are upset or troubled
7. Good understanding of social relationships
8. Good at teaching others
9. Making people fell at ease
10. Better at working with people than things or ideas

Inputting data

	It1	It2	It3	It4	It5	It6	It7	It8	It9	It10
1	2	5	3	1	4	5	2	1	2	4
2	2	2	2	2	2	2	2	2	2	2
3	2	2	2	1	2	1	2	1	2	2
4	1	2	1	1	1	4	2	1	1	2
5	2	5	2	2	2	3	2	1	2	2
6	1	1	1	2	2	2	3	1	1	1
7	2	3	2	1	3	2	3	1	2	3
8	1	4	1	5	1	1	2	1	1	1
9	3	2	2	4	2	2	3	1	2	2
10	1	1	1	1	1	2	1	1	1	3

To conduct a reliability analysis

- Statistics – scale – reliability analysis – split half
- Report: Spearman-Brown correction split-half reliability coefficient.

Cronbach alpha

Prof. Lee J. Cronbach, Stanford University.

Attitude scales

Five point Likert scale format

Strongly agree	Agree	Undecided	Disagree	Strongly disagree
(5)	(4)	(3)	(2)	(1)

e.g. Please circle the choice after each statement that indicates your opinion.

- Students can learn to become a scientist without losing their cultural values.
- Science alienates people from their traditional culture.

Exercise 1:

Scale for Measuring Attitudes Towards Mathematics or Science.

	Strongly agree (5)	Agree (4)	Undecided (3)	Disagree (2)	Strongly disagree (1)
1. I want to develop my mathematical (science) skills and study this subject more.					
2. Mathematics (science) is not a very interesting subject.					
3. Mathematics (science) is a very worthwhile and necessary subject.					
4. Mathematics (Science) makes me feel nervous and uncomfortable.					
5. I have usually enjoyed studying mathematics (science) in school.					
6. I don't want to take any more mathematics (science) than I absolutely have to.					
7. Other subjects are more important to people than mathematics (science).					
8. I am very calm and unafraid when studying mathematics (Science).					
9. I have seldom liked studying mathematics (Science).					
10. I am interested in acquiring further knowledge of mathematics (science).					

Positively worded items: 1, 3, 5, 8, 10
Negatively worded items: 2, 4, 6, 7, 9

- **Recoding value**

Recoding negatively worded items: 2, 4, 6, 7, 9

Data file: Reliability - Cronbach

Sample: 105 subjects

- Transform – Recode – into Same Variables – move numeric variables, i.e. 2, 4, 6, 7, 9 - Old and new variable (1=5; 2=4; 3=3; 4=2; 5=1) (add)

Or Transform – Recode – into different variables – input variable –output variable Old and new variable (1=5; 2=4; 3=3; 4=2; 5=1) (add)

- **To conduct a reliability analysis**

Statistics – **Scale** – **Reliability Analysis** – **Select the variables** (i.e. it1 –to it10) – model: **Alpha** – Statistics**: Scale and Scale if item deleted** – Inter-item: **Correlations** - **Continue** and **ok**

- Results: Alpha = 0.7684
- Delete item 7, item 10. The Alpha will be higher.

Exercise 2:

Working Example: p. 147

Data file: Reliability - Work15.sav

Exercise 3:

Practice Example: p. 150

Data file: Reliability **- Prac15.sav**

Lesson 7: Non-Parametric Techniques

The non-parametric analyses do not require interval scale measurement; ordinal and nominal scale data can be analysed. Also, for the most non-parametric analyses, assumptions about the shape of the population are not required. For that reason, they are often used when small sample sizes are involved.

Chi-square test for goodness of fit

(One-sample Chi-square test: only one variable)

Exercise 1
A researcher is interested in the factors that are involved in course selection. A sample of 50 students is asked, "Which of the following factors is most important to you when selecting a course? Students must choose one and only one of the following alternatives.

1. Interest in course topic
2. Ease of passing the course
3. Instructor for the course
4. Time of day course is offered

The frequency distribution of responses for this sample is as follows:

Interest in topic	26
Ease of passing	12
Course instructor	7
Time of day	5

- Inputting data

	factors	frequency
1	1 (Interest in topic)	26
2	2 (Ease of passing)	12
3	3 (Course instructor)	7
4	4 (Time of day)	5

- Labels
- Data - Weight Cases (freq)
- Statistics – Nonparametric Tests – Chi-square – selecting variable: factor
- Results: P=0.000 < 0.05
- Conclusion: some of the factors are more important than others in course selection.

Exercise 2: *Working example p. 198*

Data file: Non-parametric – Work19a.sav

Chi-square test for relatedness or independence (2 x 2 table)

Exercise 1:

Purpose: To determine whether or not the gender is an important factor in students' course selection

Sample: 74 male students and 72 female students

Preference for course selecting

	Male	Female
Science	46	26
Social science	28	45

- Inputting data

	Gender	Preference of courses (pref)	Number of frequency (freq)
1	1	1	46
2	1	2	28
3	2	1	26
4	2	2	45

- Labels
  - Gender: 1=male 2=female
  - Preference of courses: 1=science 2=social science

- Data - Weight Cases (freq)
- Statistics – Summarize – Corssrtabs – Select a raw variable (i.e. pref) – Select a column variable (i.e. gender) – statistics: Chi-square – cells: percentages
- Results:
  - Chi-Square test: Pearson Chi-Square Asymp Sig. (2-sided) .002, i.e. P=0.002
  - 0 cells (. 0%) have expected count less than 5. The minimum expected count is 32.26.
    - Check these values above to be sure your test is valid. (.0%) < 20% and 32.26 > 1

- Analyses
  - P=. 002 < 0.01
  - The test is valid. (.0%) < 20% and 32.26 > 1
- Conclusion: Gender is factor which influences on students' course selection.

Exercise 2: Working example: p. 200
Data file: Parametric - Work19b.sav.

	Urban readers	Rural readers
Financial review	32	24
Newsweek	28	4

Lesson 8: Multiple Response and Multiple Dichotomy Analysis

Multiple response and multiple dichotomy analysis are commonly used in the analysis of questionnaire or survey data.

- Open-ended questions
- More than one choices

Multiple Response

Exercise 1:

Open-ended questions: What important factors do you consider when you choose jobs?

Or: What the following factors do you consider when you choose jobs?

1. Use of ability
2. Working conditions
3. Secure and stable employment
4. Chance to advance
5. Status/prestage
6. Job opportunity
7. Interest
8. Benefit to society
9. Personal qualifications
10. Salary
11. Challenging
12. Independent
13. Location
14. Working time

- **Survey results:**
  - The maximum number of responses obtained from an individual was six.
  - The fourteens factors were identified.

- **Data:**
  - **Six variables**: *crit1; crit2; crit3; crit4; crit5; crit6.*
  - **99***: missing*

Participant 1: 01 06 99 99 99 99
Participant 2: 02 04 06 11 14 99
Participant 3: 01 03 08 09 10 13
Participant 4: 01 04 06 12 13 99
Participant 5: 02 05 07 09 10 12

- **Inputting data**

	Crit1	Crit2	Crit3	Crit4	Crit5	Crit6	
1	1	6	99	99	99	99	
2	2	4	6	11	14	99	
3	1	3	8	9	10	13	
4	1	4	6	12	13	99	
5	2	5	7	9	10	12	

- **Data file**: Multiple response – jobchoice

  Sample: 408 students

- **Statistics**
  - Multiple Response Define Sets – Define Multiple Response Sets
  - Select and move six variables (i.e. crit1 to crit 6) into **Variables in Set:** box.
  - In the **Variables Are Coded as**, click on the **Categories**
  - **Range**: from lowest code (i.e.1) to highest code (i.e. 14)
  - **Name**: type a suitable variable name (Crits); label: a description of this variable (selection criteria)
  - **Add**
  - **Close**
  - Statistics – Multiple Response – frequencies – select variable (i.e. $crits) – ok

  Results:
  - Percentage of responses refers to the proportion of a given response in relation to the count: ***count/total responses.***
  - Percentage of cases refers to the proportion of a given response in relation to the number of valid cases: ***count/total valid cases.***

Exercise 2:

Working example: p. 216

Data file: Multiple response – Work20a.save

Multiple Dichotomy Analysis

Multiple dichotomy analysis is very similar to the multiple response analysis.

Exercise 1:

Question: please tick the important reasons why you study at the open university.

___ Job change

___ Professional development

___ Earning university degree

___ Personal interest

__ Career advancement

- Data
  - Each item would be given a variable (labels)
  - If the item is ticked, give 1
  - If the item is not ticked, give 0
  - No items were ticked from one case, tick 9 (missing)

- The data from the first participants may look as follows:
  - Participant 1 0 1 1 0 0 (This participant ticked 2 & 3)
  - Participant 2 1 1 1 1 1 (This participant ticked all five)
  - Participant 3 9 9 9 9 9 (This participant didn't tick any items)
  - Participant 4 1 1 0 0 0 (This participant ticked the first two)

- Inputting data

  Sample: 362 students

	Job	Profe	Degree	Interest	career		
1	0	1	1	0	0		
2	1	1	1	1	1		
3	9	9	9	9	9		
4	1	1	0	0	0		
5							

- **Data file**: multiple responses – Oustudy
- **Labels:** 0=no 1=yes
- ***To conduct a multiple dichotomy analysis:***
- Statistics – Multiple Response and Define Sets
- Select five variables (i.e. job, profe, degree, interest, career) and move into the Variables in Set: box
- In the **Variables Are Coded As** box, select Dichotomies
- In the **Counted value** box, type the value that you assigned to those items which were ticked by respondents (i.e. 1)
- In the **name**, type a suitable variable name (i.e. Reasons). In the label: type a description of this variable (i.e. The reasons of study)
- Add – close
- Statistics – Multiple Response - Frequencies – Select the variable ($reasons) – ok
- **Results:**
- Percentage of responses refers to the proportion of a given response in relation to the count: ***count/total responses.***
- Percentage of cases refers to the proportion of a given response in relation to the number of valid cases: ***count/total valid cases.***

Exercise 2:

Practice example: p. 221

Data file: multiple response -- Prac20b.sav

Output: see p.223