Highlighting Missing Data in Lertap5

Larry R Nelson¹, Curtin University, Australia 20 July 2022

Enhancements have been made to Lertap5 so that assessing the presence and extent of missing data is easier to do, with "NA" used to denote missing data whenever users ask Lertap5 to output a csv (comma-separated values) file with item scores².

This document is largely designed for Lertap5 users who might desire to export data in the csv file format for use with R scripts, RMD scripts, or data analysis packages such as JASP and SPSS, paying special attention to matters related to how missing data are processed.

I'll use the 10-item "Comfort" scale from the <u>LertapQuiz dataset</u> built into Lertap5.

~~	LertapC	-	٦)	Larry	Nelso	n 🧑			(j)	Ī	3	– c	ו	×
File	Lertap H	lome	nsert	Draw	Page	Forn	nu Da	ta Re	viev V	iew D	evelc	Add-i	ı Help		ß
R1C1		:	×	~	<i>fx</i>	'Com	nfort S	Scale i	respor	nses f	rom C)tago	worksho	р	*
	1	2	3	4	5	6	7	8	9	10	11	12	1	3	
1	Comfort	Scale	respo	onse	s fro	m Ot	ago	work	shop	o par	ticipa	ants.			
2	Record	ID	026	027	028	029	030	031	032	033	034	035			
3	1	9	2	3	1	2	2	4		4	4	3			
4	2	31	3	3	4	3	3	5	4	2	4				
5	3	26	4	2	2	1	3	4	4	1	4	2			
6	4	27	2	4	3	1	2	5	3	3	4				
7	5	21	2	4	3	1	2	4	4	2	4				
8	6	59	4	3	2	2	2	3	5	1	5	4			
9	7	47	5	1	2	1	2	5	3	1	4	4			
10	8	42	5	3	2	1	1	5	4	1	5	4			
11	9	55	5	2	2	1	2	3	5	1	4	4			
12	10	51	5	3	2	1	1	4	4	1	5	4			
13	11	20	3	3	4	2	2	3	4	3	1	3		,	
14	12	41	4	2	1	2	2	2	3	1	5	4		\checkmark	
	Data	CCs	(+)	-	-	-		: 💽			1			<u> </u>
Ready	ĒŌ		Ę	Displ	ay Setti	ngs		Ξ] [<u> </u>		-		- 100	%



Figure 1 displays the first 12 records of Comfort Scale item responses. There were a total of 60 respondents.

Each of the 10 items on the scale allowed for five responses: 1, 2, 3, 4, 5. The <u>CCs sheet</u> used to process results is shown in Figure 2.

Ordinarily only the first three rows in Figure 2's CCs sheet would be used. However, later on I will be talking about "MDO" and thus have defined a second Lertap5 subtest, giving it the name of "MDO Comfort", and including the MDO keyword on the second sub Aff row.

Missing data in Lertap5 arise in just two cases: a cell in the Data sheet will be empty and/or will contain a response not seen in the Res= declaration on a *sub line.

¹ I.nelson@curtin.edu.au

² Figure 8 has an example of a csv file.

Figure 1 indicates that the first record was missing a response on item Q32. The last item, Q35, turned out to have 9 missing responses – three of them are showing in Figure 1, corresponding to records 4, 6, and 7. (Readers unfamiliar with Lertap5 may benefit from <u>this pdf document</u>.)

IJ	>> LertapCo	- ,0	Larry Nelson		<i>lä</i> [*]	T –	D	×
File	Lertap Home	Insert Draw Pa	age Layc Formul	as Data Rev	iew View	Develope H	elp 🖓	È
*col *sub *pol	(c3-c12) Aff, Res=(1,2 + +++	,3,4,5), Na	me=(User com	nfort), Ti	tle=(Con	nfort), wt	:=0	
*col *sub /	(c3-c12) Aff, Res=(1,2 + +++	,3,4,5), MD	0, Name=(MDC) comfort)	, Title	=(MDOComf)	, wt =0)
Ready	Data CCs	Freqs Score	s Stats1f Stats1f	ats1b IStat	(+) :	•	+ 10	▼

Figure 2

Lertap5's "<u>Interpret</u>" option will create a "<u>Freqs</u>" worksheet with a summary of item responses. Figure 3 shows results for Q34 and Q35. As mentioned above, Q35 was missing a response from 9 people.



Figure 3

Once the Interpret option has been run, and the Freqs report studied, the <u>Elmillon</u> option is then called on. It will add summary missing data information to the bottom of the Freqs report, as seen in Figure 4, and then proceed to create three new worksheet reports, including "<u>Stats1b</u>", found below in Figure 5.

E		ッ 、 »	Lertar	oC 🔻	م	Larry Ne	lson 🧑	•	⅌	[j]	F			×
File	Ler	tap Home	Insert	Draw Pa	ge Layc F	ormulas	Data Re	view V	/iew D)evelop	e Add	-ins Hel	p 🖓	ß
Q35 (c <u>O</u>	12) ption 1[2] 3] 4 5] ?[n 2 13 2 12 2 17 2 7 1 9 1	/60 3.3% 21.7% 20.0% 28.3% 11.7% 15.0%											
<u>Subtes</u>	<u>t 1 su</u>	10 item co Thus, in to 15 of thes As a perce (Note: *ex	olumns are otal, there e cells are ntage, th cc and *m	e defined e are 600 e missing iis is 2.5 iws CCs li	by the *c cells in th a valid re % of the c nes, if any	ol line in th le Data she sponse. (T ells. y, have not	e CCs sh et alloca he sum o been co	ieet, and ted for f if all the nsiderec	d there this su e 'othe d at thi	e are 60 ıbtest's r' cases is point) records item re in the .)	s in the E sponses. Stats-f re	Data she eport.)	et.
		Data C	Cs Fre	qs Sco	ores Sta	its1f Sta	ts1b I	Stats	(• : [•	-		
Ready	EŌ				L	🛱 Display Set	tings	====		삔			+	80%

Figure 4

>>	Lerta	pComfoi	tScal `	•	Q		ß	Ē			×
File	Lerta	Home	nsert D	raw Pag	ge Forr	nı Data	a Rev	viev Viev	v Deve	l Help	P ¢
Lertap5	brief it	em stat	s for "U	ser con	nfort", (created	1: 15	/06/202	22.		▲
Res->	1	2	3	4	5	other	pol.	mean	s.d.	cor.	<u>h</u>
Q31		3%	18%	43%	35%		+	4.10	0.81	- 0.05	
Q32			13%	53%	32%	2%	+	4.17	0.66	0.22	
Q33	40%	23%	23%	13%			-	3.90	1.08	0.65	
Q34	2%		17%	60%	22%		-	2.00	0.73	- 0.56	
Q35	3%	22%	20%	28%	12%	15%	+	3.23	1.02	0.57	
↓ →	; 5	Scores	Stats1	State	s1b	(+)	: 4				▼
Ready	Ē0	ि स्त्र Dis	play Setti	ngs	=	E	巴		-	+	90%

Figure 5

The "Other" column in a Stats1b report indicates the percentage of data records having missing data on each item. For Q35 the figure is 15%, corresponding to 9 divided by 60.

Two of Q35's results, the mean of 3.23 and "cor." of 0.57, are further discussed below. More about Stats1b figures are provided here.

Before getting into more meaty matters, now that the Interpret and Elmillon options have been taken, look anew at the Data sheet (Figure 6):

»	LertapCo	mfortS	c 🔻		Q	\Rightarrow		lä i	₫			>	<
File Le	ertaj Hor	nelns	ert Dr	aw Pa	ge F	orm	Data	Revie	View	Deve	Help	P	Ê
Comfort'	scale	resp	onse	s fro	m Ot	ago	work	kshop	o par	ticip	ants.		Ĥ
Record	ID	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35		
1	9	2	3	1	2	2	4		4	4	3		
2	31	3	3	4	З	3	5	4	2	4			
3	26	4	2	2	1	3	4	4	1	4	2		
4	27	2	4	3	1	2	5	3	3	4			
5	21	2	4	3	1	2	4	4	2	4			
6	59	4	3	2	2	2	3	5	1	5	4		
7	47	5	1	2	1	2	5	3	1	4	4		
8	42	5	3	2	1	1	5	4	1	5	4		
9	55	5	2	2	1	2	3	5	1	4	4		
10	51	5	3	2	1	1	4	4	1	5	4		
11	20	3	3	4	2	2	3	4	3	1	3	×	
· · · · •	Data	CCs	Freq	s Sc	ores	(+) :				_		
Ready 💽	La C	Display	Setting	5		E	巴			-		100%	>

Figure 6

Colouring has now been used to highlight cells with missing data. Each time Lertap5 starts up, among numerous other behind-the-scenes events, it looks at Rows 78 and 79 of its <u>System work-sheet</u> to make sure that users want to have highlighting applied to cells with missing data.

The default settings are to have missing data cell highlighting on, using a light-blue background colour. Other colours are possible.

Figure 7 displays "IStats", the report created whenever the "<u>Item scores and correlation</u>" option is taken. It indicates the scoring points each participant got for their answer on each item. The sum of all these points, the "item scores", becomes the participant's total test score.

**	Lertap		Q	Larry N	elson 🧃		\Rightarrow	[ž	T			×
File	Lertar Ho	me Inse	rt Draw	Page I	Formu	Data Re	eviev Vie	ew Dev	el Add-	i Help		Ċ
Lertap:	5 IStats m	atrix, c	reated:	15/06	/2022.			<u>h</u>				
ID	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35		
9	2.00	3.00	5.00	4.00	4.00	4.00	3.00	2.00	2.00	3.00		
31	3.00	3.00	2.00	3.00	3.00	5.00	4.00	4.00	2.00	3.00		
26	4.00	4.00	4.00	5.00	3.00	4.00	4.00	5.00	2.00	2.00		
27	2.00	2.00	3.00	5.00	4.00	5.00	3.00	3.00	2.00	3.00		
21	2.00	2.00	3.00	5.00	4.00	4.00	4.00	4.00	2.00	3.00		
59	4.00	3.00	4.00	4.00	4.00	3.00	5.00	5.00	1.00	4.00		
47	5.00	5.00	4.00	5.00	4.00	5.00	3.00	5.00	2.00	4.00		
42	5.00	3.00	4.00	5.00	5.00	5.00	4.00	5.00	1.00	4.00		
55	5.00	4.00	4.00	5.00	4.00	3.00	5.00	5.00	2.00	4.00		
51	5.00	3.00	4.00	5.00	5.00	4.00	4.00	5.00	1.00	4.00		
20	3.00	3.00	2.00	4.00	4.00	3.00	4.00	3.00	5.00	3.00		
41	4.00	4.00	5.00	4.00	4.00	2.00	3.00	5.00	1.00	4.00		
23	4.00	4.00	4.00	3.00	4.00	4.00	3.00	5.00	2.00	3.00		
24	2.00	2.00	3.00	3.00	3.00	5.00	5.00	3.00	2.00	4.00		
44	3.00	2.00	3.00	4.00	4.00	4.00	4.00	5.00	2.00	4.00		-
•	Scores	s Stats	1f Stat	s1b	Stats	+	- -					Þ
Ready	E0		୍ଲ Displa	y Setting	s 🎚						- + 100)%

Figure 7

In Figure 7, Lertap5 has made a <u>missing-data substitution</u>, putting values (scores) of 3.00³ in those cells that had missing data, leaving the cell highlighting in place. This is a default action. It may be turned off by using "MDO" on a *sub line in the CCs sheet. This is seen above in Figure 2. (MDO = missing data substitution off.)

Had MDO been used, the cells with missing data would still be highlighted in the IStats report, but cell values (scores) would be 0.00.

Stats1b reports (Figure 5) indicate the correlation between each item and the total test score, with the item omitted from it (the total test score) in order to control for part-whole inflation, referred to as "spuriousness" in some other apps.

In this case, with missing data substitution in effect (Lertap5's default action), Q35's correlation with the total test score was 0.57, based on all 60 records, with the item's mean being 3.23. With MDO active, the 9 records with missing data are excluded and the correlation drops to 0.56, with mean 3.27, based on 51 records. Not a big difference *in this example*.

Finally, I move on to using the "<u>Omega1 macro</u>", knowing that it will create a csv file with item scores, a file which I may then use with R/RMD scripts, and JASP (or SPSS).

Figure 8 has the "IScores" report created by the macro, displayed as another Excel worksheet, while Figure 9 captures the first few rows of the csv file using Windows Notepad (TextEdit is the equivalent on a Mac).

~~	Lertap	Comfort	Scale.xlsx	-	م	♦	Ċ	র্গ দ	3 –	- 0	>	<
File	Lertar Ho	me Inser	t Draw	Page I I	Form(D	ata Rev	vie Vie	w Deve	l Add-i	Help	P E	ŝ
Lertap5	IScores	matrix,	create	d: 16/	06/202	2		<u>h</u>				
ID	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35		
9	2.00	3.00	5.00	4.00	4.00	4.00	NA	2.00	2.00	3.00		
31	3.00	3.00	2.00	3.00	3.00	5.00	4.00	4.00	2.00	NA		
26	4.00	4.00	4.00	5.00	3.00	4.00	4.00	5.00	2.00	2.00		
27	2.00	2.00	3.00	5.00	4.00	5.00	3.00	3.00	2.00	NA		
21	2.00	2.00	3.00	5.00	4.00	4.00	4.00	4.00	2.00	NA		
59	4.00	3.00	4.00	4.00	4.00	3.00	5.00	5.00	1.00	4.00		
47	5.00	5.00	4.00	5.00	4.00	5.00	3.00	5.00	2.00	4.00		
42	5.00	3.00	4.00	5.00	5.00	5.00	4.00	5.00	1.00	4.00		
55	5.00	4.00	4.00	5.00	4.00	3.00	5.00	5.00	2.00	4.00		
51	5.00	3.00	4.00	5.00	5.00	4.00	4.00	5.00	1.00	4.00		
20	3.00	3.00	2.00	4.00	4.00	3.00	4.00	3.00	5.00	3.00		
41	4.00	4.00	5.00	4.00	4.00	2.00	3.00	5.00	1.00	4.00		
23	4.00	4.00	4.00	3.00	4.00	4.00	3.00	5.00	2.00	NA		
24	2.00	2.00	3.00	3.00	3.00	5.00	5.00	3.00	2.00	4.00		
44	3.00	2.00	3.00	4.00	4.00	4.00	4.00	5.00	2.00	4.00		•
	State	s1b IS	tats IS	cores	+		-				►	
Ready			🛃 Displ	ay Setting	gs 🗄]			1009	6

Figure 8

³ An item score of 3 (in this case) is the centre score in Res=(1,2,3,4,5) seen in the CCs lines.



Figure 9

Missing data have been replaced with "NA" in Figures 8 and 9. JASP and many R/RMD scripts will readily recognise NA⁴ as a missing data code and will adjust their statistics accordingly.

Lertap5 version 5.11.2 was used in this paper. It may be downloaded <u>here</u>. **JASP** is a capable, free data analysis system available <u>here</u>. Appendix B in <u>this paper</u> exemplifies the use of JASP.

Suggested reading:

Dai, S (2021). Handling missing responses in psychometrics: methods and software. Psych 2021, 3, 673-693 (https://doi.org/10.3390/psych3040043).

⁴ NA is interpreted as "value not available". I found <u>this reference</u> useful.

Appendix

The small note at the bottom of Figure 4 advises that the results in Subtest 1's summary have been derived without consideration of any <u>CCs lines</u> which may have been used to delete an item from the subtest, such as a *exc or *mws line.

Look at this little example with two subtests. It has to do with the <u>Test13 sample dataset</u>:



Figure 10

Test13 was a 13-item multiple-choice test of elementary mathematics ability given to primary school students, with items 'numbered' from Q1 to Q13. Student responses were entered in columns 2 to 14 of the Data worksheet (see line 2 in Figure 10), with 9s used if a student did not respond.

The second item, Q2, with responses found in Data column 3, was found to be of questionable relevance, making it desirable to exclude it from Lertap5's analyses. This can be done in two ways.

Line 5 uses *exc (c3) to exclude Q2 from the first subtest (c3 means column 3).

Line 7 effectively excludes Q2 from the second subtest by <u>not</u> referring to its corresponding Data column (c3). Note that the *keys seen in line 9 now have 12 entries – the key for Q2 is missing.

Figure 11 has the missing data summary seen at the bottom of the Freqs report.

The summary for the first subtest covers 13 items while that for the second subtest covers 12 items.

Both summarise the extent of missing responses in the Data sheet. However, the first summary may be of little or no interest as it references all items, including Q2. Lertap5 has ignored the *exc statement in line 5 of the CCs sheet (an oversight which may be corrected in a future revision).

Finally, compare the number of cells found to be missing a valid response: 3715 for subtest 1 and 3316 for subtest 2, a difference of 399. Then look at Figure 12. In this dataset, a code of 9 was used if a student failed to answer a question. Figure 12 shows the 399 missing on Q2, the excluded item.

	<u></u> <u></u>		Test1	• Save	ed • 🔎	Larry	Nelson			ĽŽ	শ	-		×
File	Lertap	Home	Insert	Draw	Page Layout	Formulas	Data	Review	View	Develo	per Ado	d-ins H	elp	
	1	2	3	4	5	6		7		8	9	1	0	11 📥
106 <u>Su</u>	ıbtest 1 sur	<u>mmary:</u>												
107		13 item	colum	ns are de	efined by the	*col line in	the CC	s sheet, a	and the	ere are 2	976 reco	rds in th	e Data	sheet.
108		Thus, ir	i total, t	there are	e 38688 cells i	n the Data :	sheet a	llocated	for thi	s subtes	t's item r	esponse	s.	
109		3715 of	these of	ells are	missing a vali	d response.	(The s	um of all	the 'o	ther' cas	es in the	Stats-f r	eport.)	
110		As a pe	rcentag	e, this is	9.6 % of the o	cells.								
111		(Note: *	[•] exc and	d *mws (CCs lines, if ar	ny, have not	: been	consider	ed at t	his poin [.]	t.)			
112														
113 <u>Su</u>	ıbtest 2 sur	<u>nmary:</u>												
114		12 item	colum	ns are de	efined by the	*col line in	the CC	s sheet, a	and the	ere are 2	976 reco	rds in th	e Data :	sheet.
115		Thus, ir	i total, f	there are	e 35712 cells i	n the Data :	sheet a	llocated	for thi	s subtes	t's item r	esponse	s.	
116		3316 of	these of	ells are	missing a vali	d response.	(The s	um of all	the 'o	ther' cas	es in the	Stats-f r	eport.)	
117		As a pe	rcentag	e, this is	9.29 % of the	cells.								
118		(Note: *	[•] exc and	d *mws (CCs <mark>lines, if</mark> ar	ny, have not	: been	consider	ed at t	his poin [:]	t.)			
119														-
•	▶ Da	ata CO	is Fre	eqs Sc	cores Stats	1f Stats	1k (+ : -	•					•
Ready	ĒŌ					ାର୍ଚ୍ଛ Display S	ettings	III	Ξ	E			+	100%

Figure 11

	>>		Q		lä ^s	-	0	×
Fi	ile	Lerta	Home I	nsert Dr	aw Page	Form	Data	Re
		1	2	3	4		5	
4		2	291	9.8%				
5		3	2,235	75.1%				
6		4	140	4.7%				
7]	9	124	4.2%				
8								
9	Q2 (c3)						
10		Option	n	/2976				
11		1	1,048	35.2%				
12		2	637	21.4%				
13		3	526	17.7%				
14		4	366	12.3%				
15		9	399	13.4%				
16								
17	Q3 (c4)						-
•	►	F	reqs	Sco (+ : 🖣			•
			L _A	Display Set	tings		E	巴

Figure 12