Using Practical Exhibits to Present Selected Measurement Topics

This Exhibit: "BDI-21"

Larry R Nelson[1]

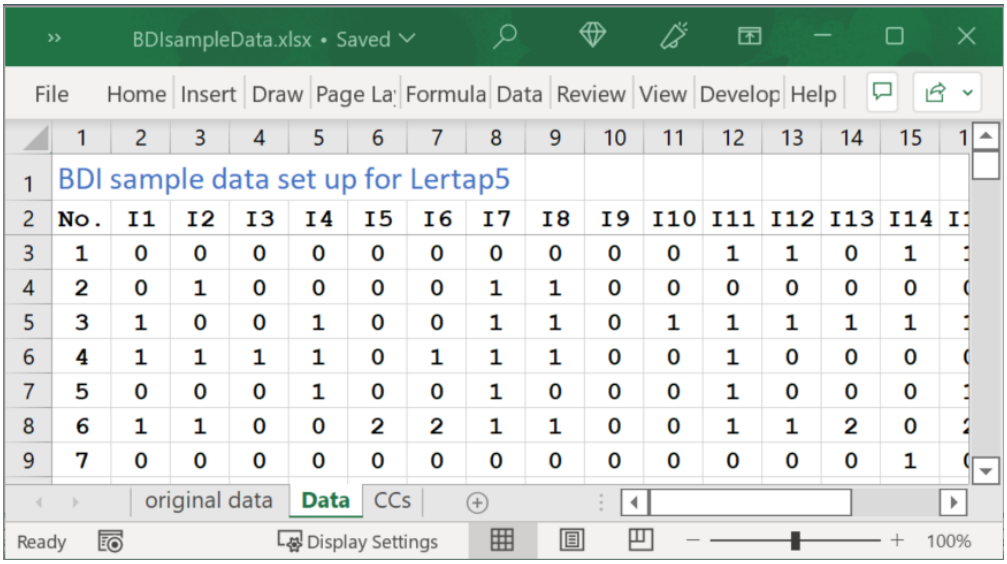
Dated: 15 May 2023 (check for update)

● About the "BDI-21"

The Beck Depression Inventory, BDI, is a self-report depression instrument with 21 questions. The BDI example featured in this document used a version of the inventory presented in a multiple-choice format with each item (question) having four options. *See the actual instrument here.*

This Wikipedia page had a good description of the BDI as of the date above. I have previously used the BDI in lectures with postgraduate classes – class notes are here.

● Student responses recorded in an Excel workbook



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	BDI sample data set up for Lertap5														
2	No.	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14
3	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1
4	2	0	1	0	0	0	0	1	1	0	0	0	0	0	0
5	3	1	0	0	1	0	0	1	1	0	1	1	1	1	1
6	4	1	1	1	1	0	1	1	1	0	0	1	0	0	0
7	5	0	0	0	1	0	0	1	0	0	0	1	0	0	0
8	6	1	1	0	0	2	2	1	1	0	0	1	1	2	0
9	7	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure 1 *(numbers denote columns)*

Figure 1 displays the first nine records of an open Excel workbook, "BDIsampleData.xlsx", displaying question responses from seven of 242 Australian university students who responded to the inventory some years ago. The workbook name is seen in the green band at the top of the figure.

The workbook columns in Figure1 are numbered. Columns 1 to 15 and rows 1 to 9 are displayed. Excel's "R1C1" reference style has been activated in this case, meaning that the first cell in the open worksheet (Data) is denoted as being in Row 1, Column 1. Often Excel users will find their copy of Excel to be using letters instead of numbers to label columns, with A being the first column, a reference style referred to as "A1" – Column A, Row 1. See Figure 2 for an example of the A1 style.

The Excel-based data analysis app used in this document, Lertap5, includes a shortcut for switching between Excel's two reference styles – see this help topic. This page from Microsoft describes another way to change the reference style. *The R1C1 reference style is preferred when working with Lertap5.*

[1] l.nelson@curtin.edu.au

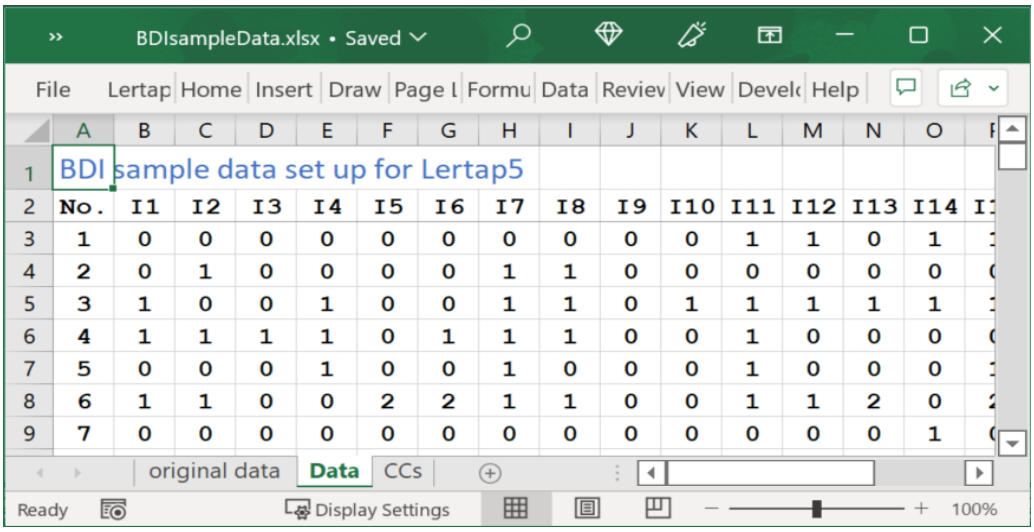


Figure 2 *(here letters denote columns)*

● Lertap5

Lertap5 is an Excel-based app used to analyse tests and surveys. An introduction to Lertap5 is found at this website. The use of Lertap5 is exemplified in a number of places. *A highly recommended start is this PowerPoint presentation (also available as a PDF).* Another suggested guide is the "Cook's Tour".

To follow the steps presented in this document, readers will want to (1) have Excel running on their computer[2]; (2) install Lertap5; (3) open the Lertap5.xlsm workbook; (3) get a copy of the BDIsample-Data.xlsx Excel workbook; and (4) open BDIsampleData.xlsx. This web page is the one to visit in order to get Lertap5; it has a link to download the Lertap511.zip file, and also a link to the "ReadMe-First.pdf" document which explains how to get Lertap5 started. A copy of the BDIsampleData.xlsx Excel workbook is available here: BDIsampleData.xlsx

● Get "Freqs"

We'll begin by getting Lertap5 to summarise item response frequencies as found in BDIsample-Data.xlsx.

To do so we need to have Lertap5.xlsm open and running in Excel, and we also need to have opened the BDIsampleData.xlsx workbook. Once this is done, a click on Lertap5's "Interpret" option will start things perking; refer to Figure 3 where the Interpret option has been highlighted.

[2] It must be a version of Excel that will run "macros". Tablet and phone versions might not support macros.

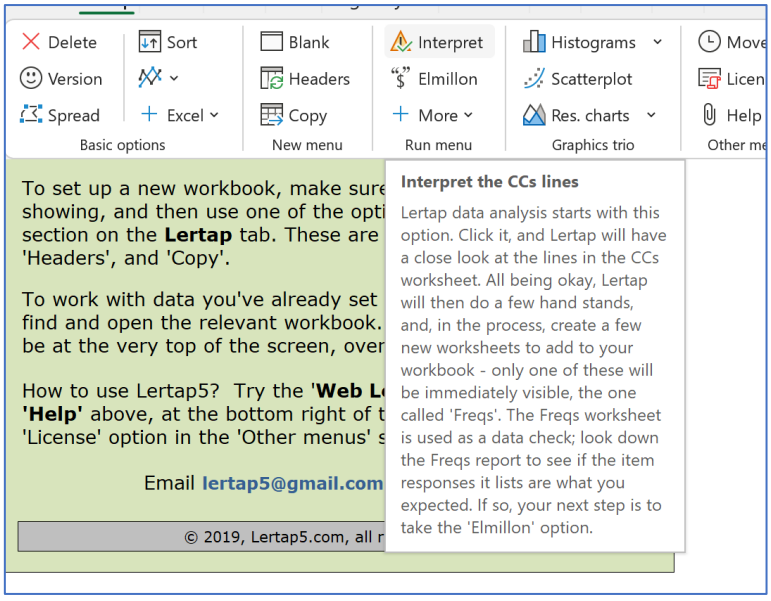


Figure 3

The Interpret option adds a worksheet called "Freqs" to the BDIsampleData.xlsx workbook. Figure 4 displays response frequencies for the first three of the 21 BDI items.

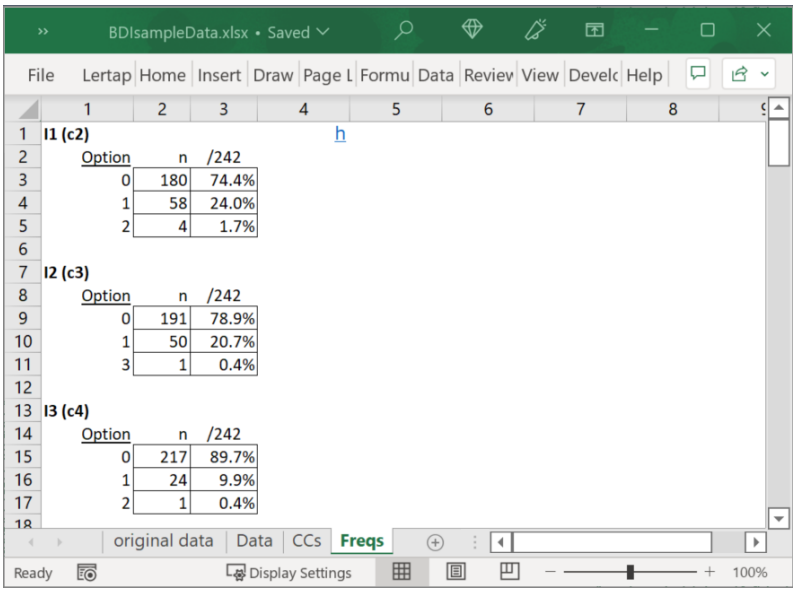


Figure 4

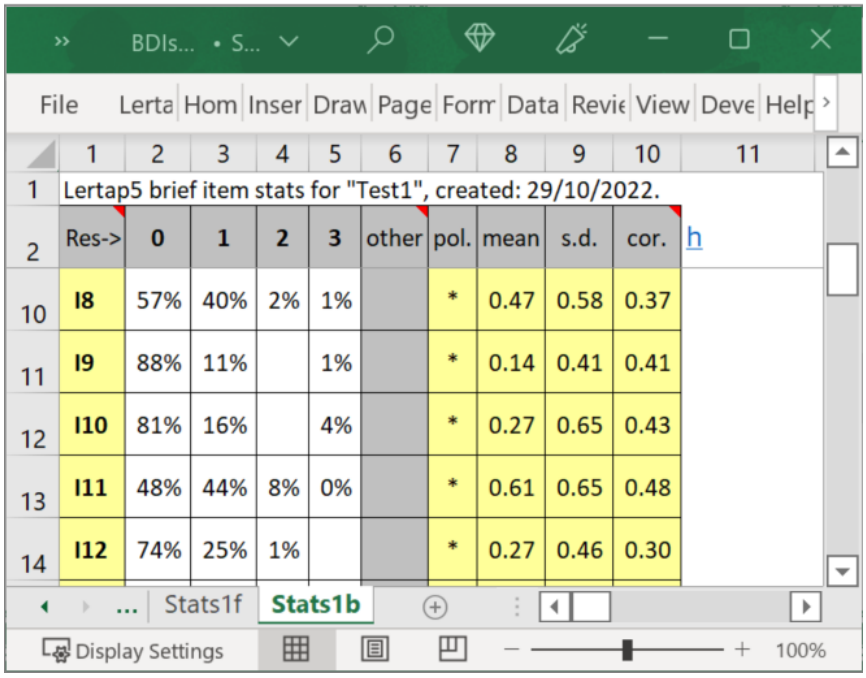
The Freqs report is clearly quite simple. Nonetheless it readily provides the means to contemplate student responses at the item level, and that will always be highly informative. Note that the number of the column each item response resides in follows the item number and is in parentheses. For example, **I1 (c2)** indicates that student responses to the first item (**I1**) are found in the second column of the workbook (**c2**).

On the first item, just under 75% of the students reported that they did not feel sad (click here to see the actual questions). As rosy as that might seem, four students indicated that they felt sad all of the time and were unable to "snap out of it". On the second item, one student indicated feeling that "the future is hopeless, and things cannot improve".

● Get item statistics and student scores with "Elmillon"

Elmillon is the name of the main item analysis routine in Lertap5. The meaning of the name is given here. In Figure 3 above the Elmillon option is seen immediately below the Interpret option.

I'll use Elmillon now. When it finishes it will have added three new worksheets (or "reports") to the workbook: Stats1f, Stats1b, and Scores.



	1	2	3	4	5	6	7	8	9	10	11
1	Lertap5 brief item stats for "Test1", created: 29/10/2022.										
2	Res->	0	1	2	3	other	pol.	mean	s.d.	cor.	h
10	I8	57%	40%	2%	1%		*	0.47	0.58	0.37	
11	I9	88%	11%		1%		*	0.14	0.41	0.41	
12	I10	81%	16%		4%		*	0.27	0.65	0.43	
13	I11	48%	44%	8%	0%		*	0.61	0.65	0.48	
14	I12	74%	25%	1%			*	0.27	0.46	0.30	

Figure 5

A portion of the Stats1b worksheet is displayed in Figure 5. This webpage describes Stats1b worksheets in some detail – note that many Lertap5 worksheets have a small blue h towards the end of row 2 – these link to help pages – one of these little blue hs is seen in Figures 4 and 5.

Look for a moment at the results for item **I9.** They're in row 11 of the Stats1b worksheet.

The average score on **I9** was 0.14 – in this study, a student's score on each question corresponded to the item option which they selected: 0, 1, 2, or 3.

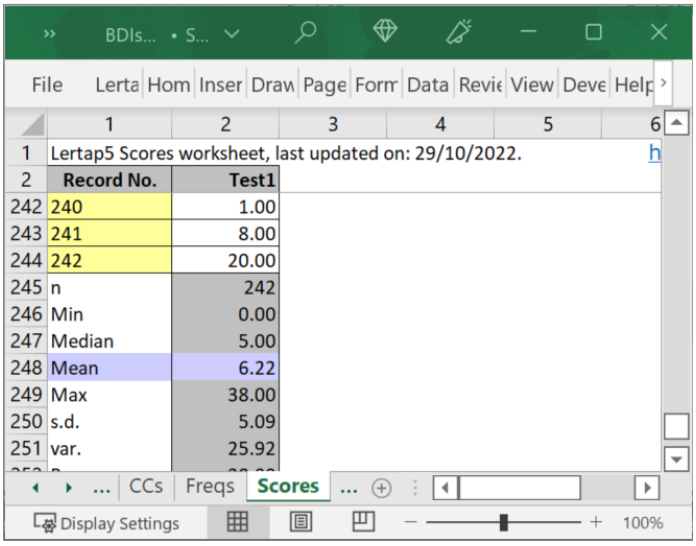
The Stats1b worksheet shows that 88% of the students (213 of them) selected a response of 0 (zero) on question **I9** ("I don't have thoughts of killing myself"); 11% (27 students) indicated that they had had thoughts of killing themselves, while, disturbingly[3], 1% (2 students) took option 3 ("I would kill myself if I had the chance").

Of the 21 questions in the instrument, **I11**, the eleventh item, had the highest mean score, 0.61 – a total of 52% of the students indicated that they were "slightly" (44%) or "quite" (8%) irritated "by things". Note the "cor." values in Column 10 – they're item correlations, plotted below in Figure 13.

The bottom of the Scores worksheet created by the Elmillon option is shown in Figure 6, with a histogram of the scores seen in Figure 7.

[3] Disturbing if we can assume that each student gave honest answers to each question.

A student's total score was computed by summing their scores on each of the 21 questions. The minimum possible total score for a student was zero. The maximum possible was 63 (there were 21 items with a maximum possible score of 3 on each).



	1	2
1	Lertap5 Scores worksheet, last updated on: 29/10/2022.	
2	**Record No.**	**Test1**
242	240	1.00
243	241	8.00
244	242	20.00
245	n	242
246	Min	0.00
247	Median	5.00
248	Mean	6.22
249	Max	38.00
250	s.d.	5.09
251	var.	25.92

Figure 6

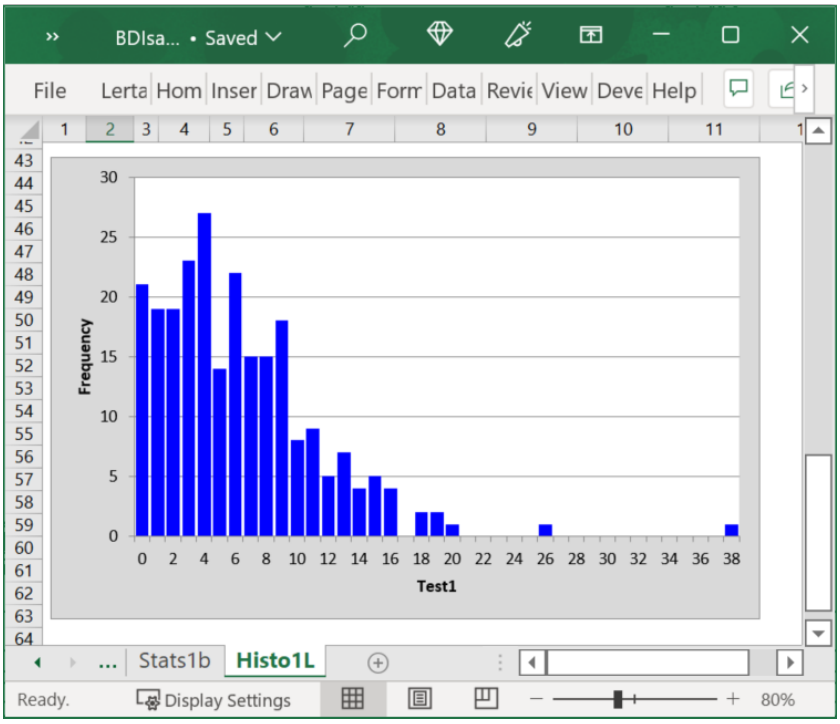


Figure 7

Total Score	Levels of Depression
1-10	These ups and downs are considered normal
11-16	Mild mood disturbance
17-20	Borderline clinical depression
21-30	Moderate depression
31-40	Severe depression
over 40	Extreme depression

Figure 8

Figure 8 displays the score "interpretation guide" given at the end of the self-scoring instructions.

In our group of 242 students, the median score was 5, as seen in Figure 6. Half of the students got a score of 5 or less -- more exactly, 123 of the 242 students had a score of 5 or less[4]. Over 80% (201 of the 242) had a score of 10 or less. Thus, according to the "scoring guide" in Figure 8, the great majority of the students had a score which would be "considered normal" (i.e., 10 or less).

No student had a score above 40; however, there was one student with a score of 38 – this student was one of the two who answered question **I9** in the disturbing manner mentioned above.

● Identifying the record with the top score

Figure 6 indicates that the "Max" score, the highest score, was 38. This is confirmed in the score histogram seen in Figure 7.

Who had this score? Well, we'll never know in this case as the respondents were not identified in any way – no Student ID, no name.

But I can find the record having a score of 38. To do so I open the Scores worksheet.

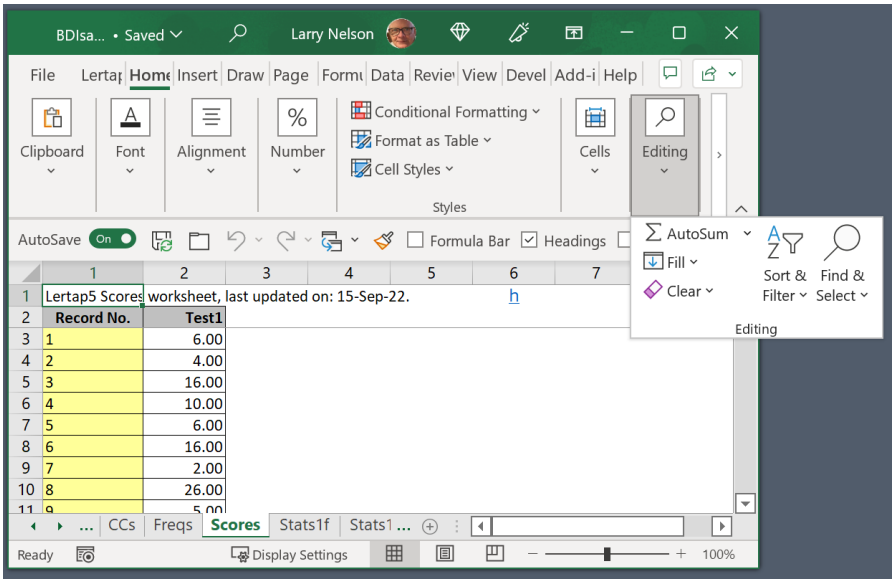


Figure 9

[4] If readers have been following the steps I've gone through here on their computers, the Histo1L worksheet has the data used to quickly find the number of students with a score of 5 or less, or 10 or less.

A “Find & Select” option is available on the Home tab, over on the right, under the “Editing” panel, as seen in Figure 9. I click on it. A new box pops up and in it, next to “Find what:”, I enter the number 38. (Figure 10.)

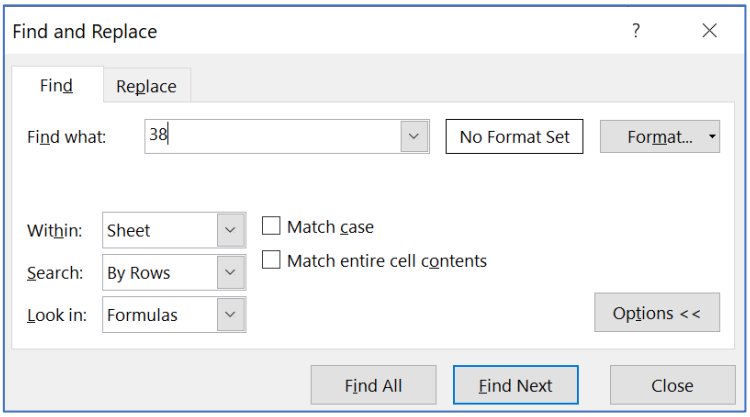
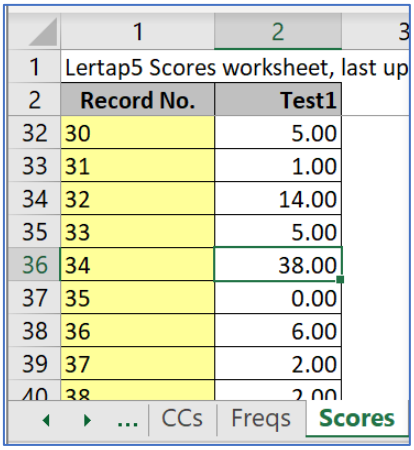


Figure 10

Then I click on “Find Next” – Figure 11 shows the outcome – the 36th row in the Scores worksheet, corresponding to Record No. 34, has a score of 38.



	1	2	3
1	Lertap5 Scores worksheet, last up		
2	**Record No.**	**Test1**	
32	30	5.00	
33	31	1.00	
34	32	14.00	
35	33	5.00	
36	34	38.00	
37	35	0.00	
38	36	6.00	
39	37	2.00	
40	38	2.00	

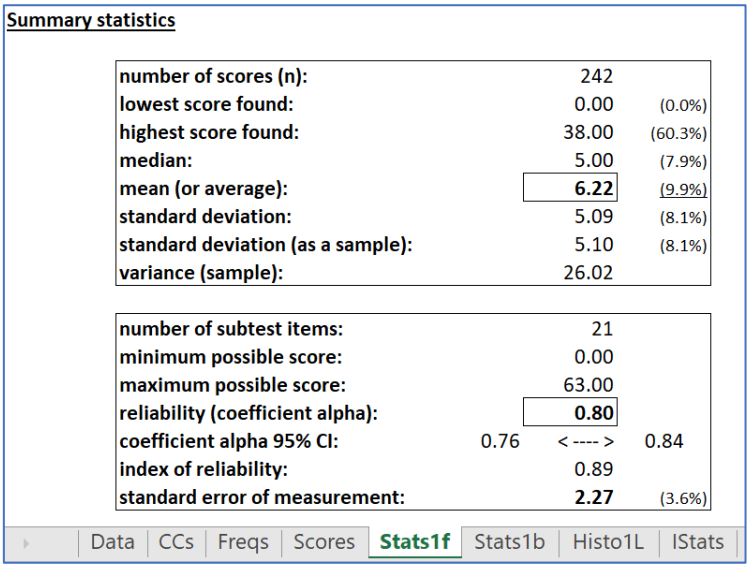
... CCs Freqs **Scores**

Figure 11

If I then went to 36th row in the Data worksheet I would be able to see this student’s responses to each of the 21 items on the inventory.

● Reliability

Lertap5 computes the value of coefficient alpha as a measure of instrument reliability – in the literature, this is also known as "Cronbach's alpha".



Summary statistics

number of scores (n):		242	
lowest score found:		0.00	(0.0%)
highest score found:		38.00	(60.3%)
median:		5.00	(7.9%)
mean (or average):		**6.22**	(9.9%)
standard deviation:		5.09	(8.1%)
standard deviation (as a sample):		5.10	(8.1%)
variance (sample):		26.02	
number of subtest items:		21	
minimum possible score:		0.00	
maximum possible score:		63.00	
reliability (coefficient alpha):		**0.80**	
coefficient alpha 95% CI:	0.76	< ---- >	0.84
index of reliability:		0.89	
standard error of measurement:		**2.27**	(3.6%)

Data | CCs | Freqs | Scores | **Stats1f** | Stats1b | Histo1L | IStats

Figure 12

Figure 12 displays that part of a Stats1f worksheet having summary test information. Coefficient alpha was found to be 0.80 for this administration of the BDI-21 with 242 university students.

Figure 13 is from Lertap5's Stats1b worksheet – its plot of item correlations shows a steady pattern of correlation values of about 0.40 for the first 17 of the instrument's 21 items.

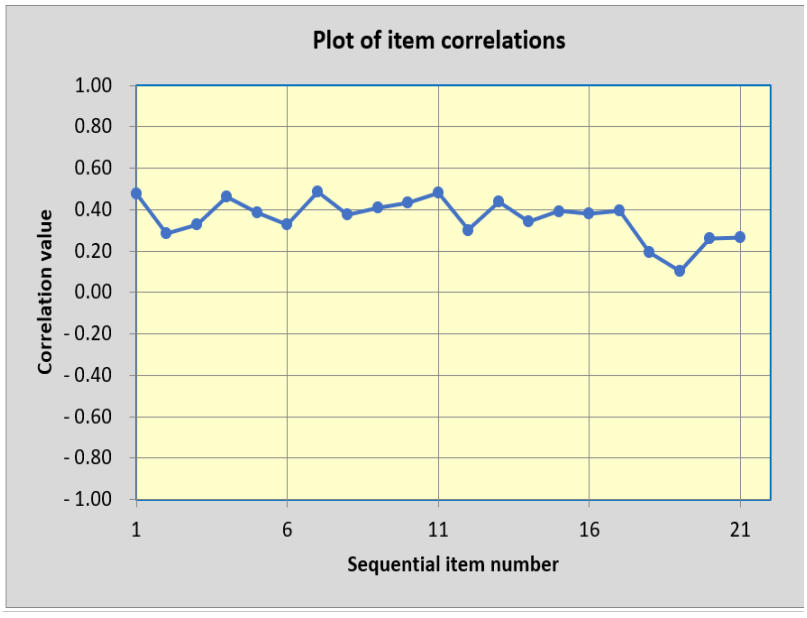


Figure 13

The lowest correlation value was 0.10 for item I19, an item which asks how much weight a respondent may have lost "lately". Keeping in mind that this scale, the Beck Depression Inventory, has been well regarded in the literature, I'd then assume that research has indicated that questions having to

do with weight loss (item I19) and appetite (item I18) are items known to often be associated with feelings of depression.

There's a small table towards the bottom of Stats1f worksheets which indicates how the value of coefficient alpha would change were an item omitted from the scale.

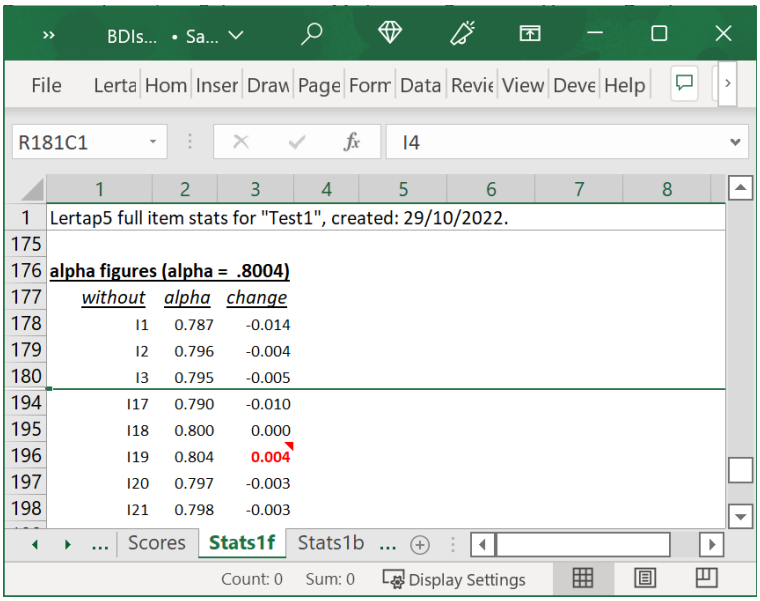


Figure 14

Figure 14 has captured that part of the Stats1f report I refer to. The original value of alpha was .8004 as shown in row 176. If item I1 were omitted from the scale, alpha would drop to 0.787. In fact, for seven of the eight items featured in Figure 11, we might[5] expect alpha to *decrease* (very slightly) if these items were taken out of the scale. This of course isn't wanted – these items are contributing to alpha, increasing its value; take one of these items out, and alpha drops.

But then there's item I19 – were that item omitted, alpha would go up, would *increase*, to 0.804. The reliability of the BDI-21 depression inventory would have been, in this case, with this sample of 242 students, very very slightly better if I19's responses were excluded from the scoring. We might then say that I19 doesn't appear to "hang" with the other items – it may be tapping a construct (*weight* in this case) not strongly related to the other 20 items.

In order for a test to have acceptable reliability, as indexed by coefficient alpha, the plot of item correlations should not show any substantial dips, all correlations will hopefully be 0.30 or more, and certainly none will fall below zero.

A value of 0.80 for coefficient alpha would be regarded as indicating adequate instrument reliability. A relevant and suggested read would be Chapter 8 of the Lertap5 manual.

It is well known that adding items to an instrument will improve its reliability as reflected in coefficient alpha – longer instruments will generally have higher reliability assuming, of course, that any added items are targeting the same topic (depression in this case).

[5] I say "might" because the items are interrelated. Take one out and all the other correlations would likely change from those seen in the figure.

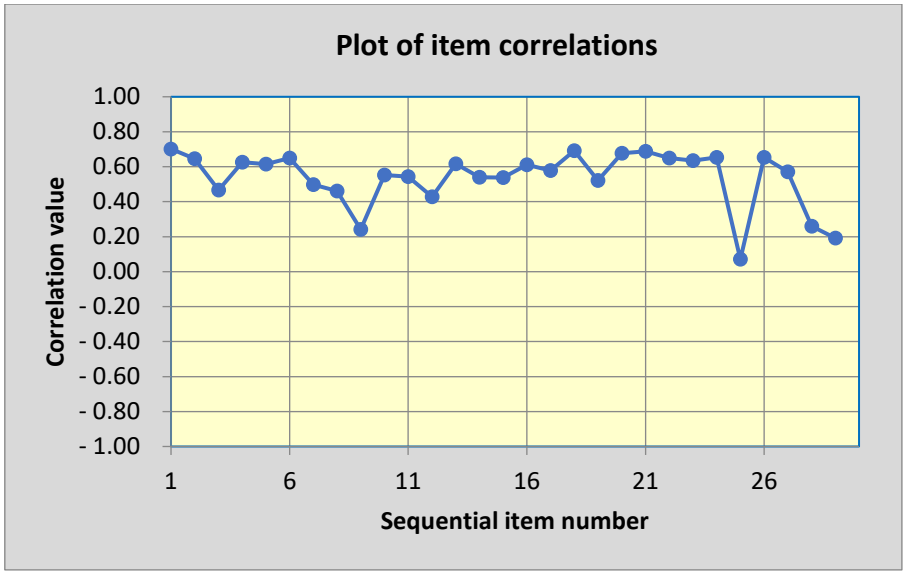


Figure 15

Figure 15 displays item correlations from an "Internalized Stigma of Mental Illness" study by Hammer & Toland. It involved an instrument with 29 items; the correlations plotted in Figure 15 resulted from an administration of the instrument to 758 adults who had been "identified as depressed" beforehand.

Coefficient alpha was found to be 0.93, considerably better than the alpha of 0.80 obtained when the BDI-21 inventory was administered to university students.

The stronger alpha from Hammer & Toland's study will have resulted from at least two factors: their instrument was longer, with 29 items as opposed to 21, and item correlations were stronger – many of them were at about the 0.60 level as compared to the BDI-21 correlations which trailed along the 0.40 level (compare Figures 13 and 15).

Now, having pointed to the advantage of having longer instruments (reliability is expected to be stronger if more items are added), we might look at another example.

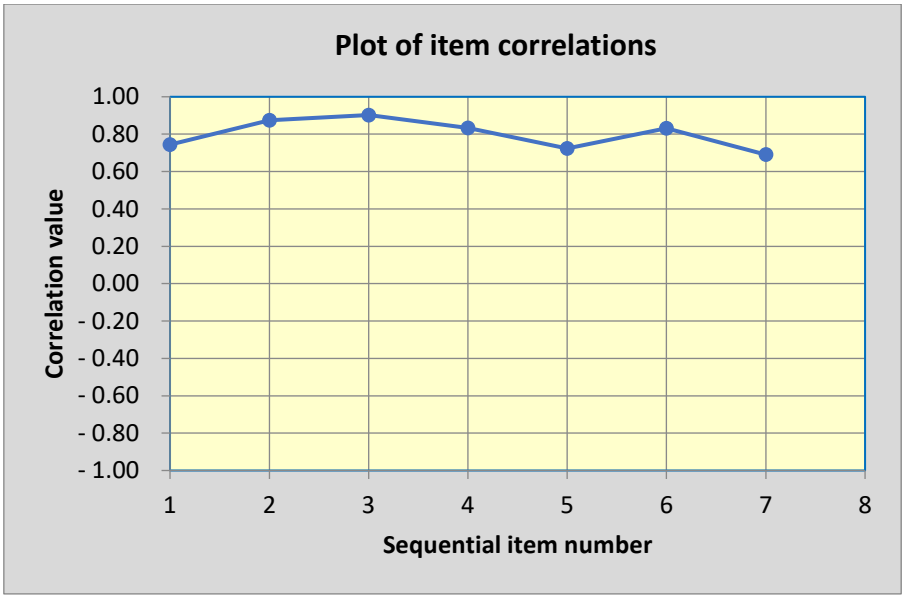


Figure 16

Figure 16 is from the "DunnSES" scale mentioned here. This scale had just seven items and yet coefficient alpha reliability was found to be 0.94 – the clue to understanding how alpha can be so high even with a small number of items is provided in the figure: the item correlations tend to hover along the 0.80 line and there aren't any big dips in the plot.

Lertap5's Stats1b reports include a small table which displays summary descriptive statistics related to item means, standard deviations, and correlations ("cor.") with the other items in the scale as shown in Figure 17.

We see that for these seven items the average item correlation was 0.80.

The BDI-21 inventory's items had an average correlation of 0.36, while the 29 items in Hammer & Toland's scale had an average correlation of 0.54. *The more the items intercorrelate, the higher alpha will tend to be. Alpha was 0.80 for BDI-21, 0.93 for Hammer & Toland, 0.94 for DunnSES.*

Lertap5 brief item stats for "Dunn & Castro scale", created: 4/11/2022.

Res->	1	2	3	4	5	other	pol.	mean	s.d.	cor.	h
tech1	4%	4%	8%	46%	39%		+	4.12	0.97	0.74	
tech2	4%	9%	17%	44%	27%		+	3.82	1.04	0.87	
tech3	5%	7%	17%	42%	29%		+	3.83	1.07	0.90	
tech4	11%	17%	23%	32%	16%		+	3.24	1.24	0.83	
tech5	14%	25%	23%	21%	17%		+	3.01	1.30	0.72	
tech6	4%	11%	21%	42%	22%		+	3.67	1.06	0.83	
tech7	7%	13%	18%	40%	22%		+	3.56	1.16	0.69	
Average:								3.61	1.12	0.80	
Std. Dev.:								0.35	0.11	0.07	

... Stats1f **Stats1b** IStat ...

Figure 17

Lertap5 has an option that will produce an extensive report of more detailed item statistics, something that may interest some readers. The report is known as "**IStats**". It's requested by taking the "Item scores and correlations" option from the Run menu on the Lertap tab. See Figure 18. Click here to read about it.

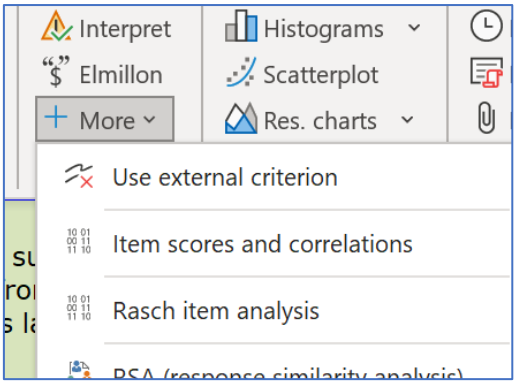


Figure 18

● From csv to xlsx: how I obtained and prepared these BDI-21 results

This section may interest mostly those readers who may like to know more about using "csv" files, files which can be read by a text editor (such as Notepad or TextEdit), a word processor (such as Microsoft Word), or a spreadsheet app (such as Excel). csv files provide a common means for researchers to share their data. "csv" means "comma-separated values".

There will be several distinct "fields" of information in csv files, with the fields delimited by commas. For example, a typical record referencing cities and towns in the U.S. state of Wisconsin, with four fields, might be: {United States, Wisconsin, Waukesha, Pewaukee}[6]. The length of each field may vary.

In the case of the BDI-21 responses I've used above, I began by downloading a csv file of the results by working through a CRAN package called "KernSmoothIRT". I have placed a copy of the BDI-ItemScores.csv made from that package here.

Figure 19 displays a few of the first rows of the csv file open in Notepad. Figure 20 shows the same file opened in Excel.

[6] {Country, State, County, Town}

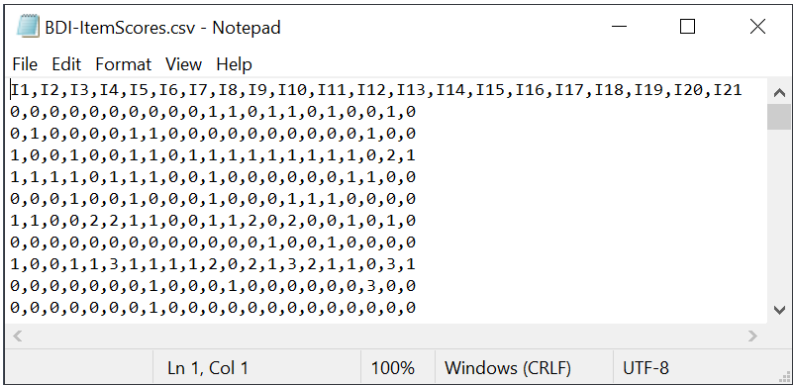


Figure 19

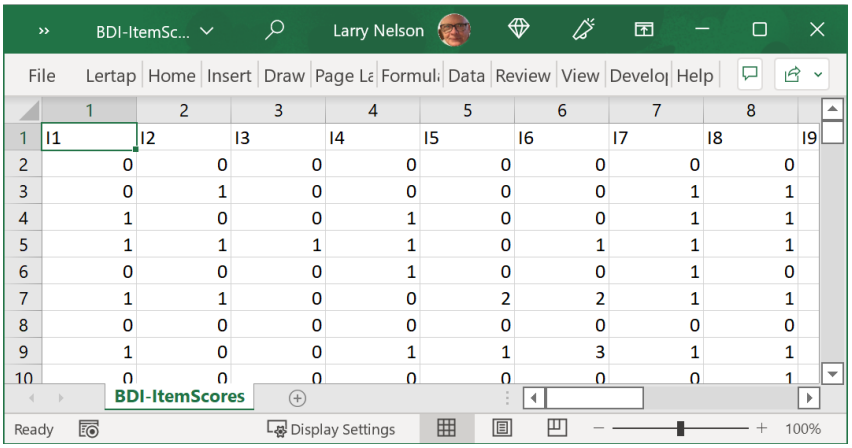


Figure 20

Lertap5 expects workbooks to have been saved as "xlsx" files. It's easy to have Excel save a csv file as an xlsx workbook: just open the csv file in Excel, and then Save As an xlsx file. I do that.

Next, Lertap5 wants xlsx workbooks to have at least two worksheets, and they can't have just any name: one must be called "Data" and the other "CCs".

To make the Data worksheet I made a copy of the BDI-ItemScores worksheet and saved it with the name "Data". Easy to do? Yes, you bet. Left-click on the BDI-ItemScores worksheet tab, take the "Move or copy" option, being sure to tick the "Create a copy" option, and select the (move to end) option too. Then rename the copy to "Data".

Next thing to do: insert a new row at the top of the Data worksheet. Lertap5 expects to have some sort of description there, in that row, at the very top of the worksheet.

See what I did by looking at Figure 2 way up above, and, while you're there, looking at Figure 2, note that I had also inserted a new column on the left which I called "**No.**" (for Number). A column such as this is very highly recommended – some csv files will come with a record number or code of some sort in the first column, and then again some won't. Lertap5 is happiest when there's some sort of record ID in the first column and I always make sure there's one there. Read more about what Lertap5 likes to see in Data worksheets by viewing this topic.

Okay? Ready to continue?

The next step is to add a new worksheet called "CCs". The letters stand for "control cards". These "cards" are really Excel rows. Each row is to contain information which will control how Excel and

Lertap5 team up to get the item analysis I want. (Incidentally, these rows used to be "punch cards" way back in the days of mainframe computers when dinosaurs still roamed. A bit of history is here.)

To add a new worksheet to the workbook, I click on the very small circle with a plus sign in its centre, just to the right of the worksheet tab for Data[7]. Excel will create a new worksheet and call it "Sheet 1" or "Sheet 2", depending on how ever many worksheets were already existing.

I change the name of this new worksheet to "CCs" by left-clicking on the sheet's tab and selecting "Rename" from the menu that appears. Then I added the control information I know Lertap5 will need.

The first row in a CCs sheet must indicate the columns in the Data worksheet where item responses are found.

In this case, the responses begin in column 2 and end in column 22.

My first row in the CCs sheet will thus be *col (c2-c22). The asterisk at the start of the row tells Excel that this row contains job control information – rows in CCs without an asterisk at the start are for comments (if any).

The second row will tell Excel what item responses it may expect to encounter, and what type of instrument is involved. In this case, it will be *sub res=(0,1,2,3), aff

The reason this "card" begins with the characters "sub" is that Lertap5 assumes that the items specified in the *col line are to be processed as a subset of all the items which might be found in the Data worksheet. Sure, true enough, the items in c2 to c22 are all of the items in this example, but Lertap5 leaves the door open to the possibility of there being more item responses beyond c22 of the Data sheet.

In this case, the subset, or, the "subtest" as Lertap5 will want to call it, has items which involve response codes of 0,1,2,3. All of these items are affective items. Affective items measure feelings, attitudes, beliefs, and so on. Lertap5 will also process cognitive items from quizzes, tests, and examinations – see an example here.

[7] See the bottom of Figure 20 -- the little circled + sign is seen to the right of BDI-ItemScores sheet tab.

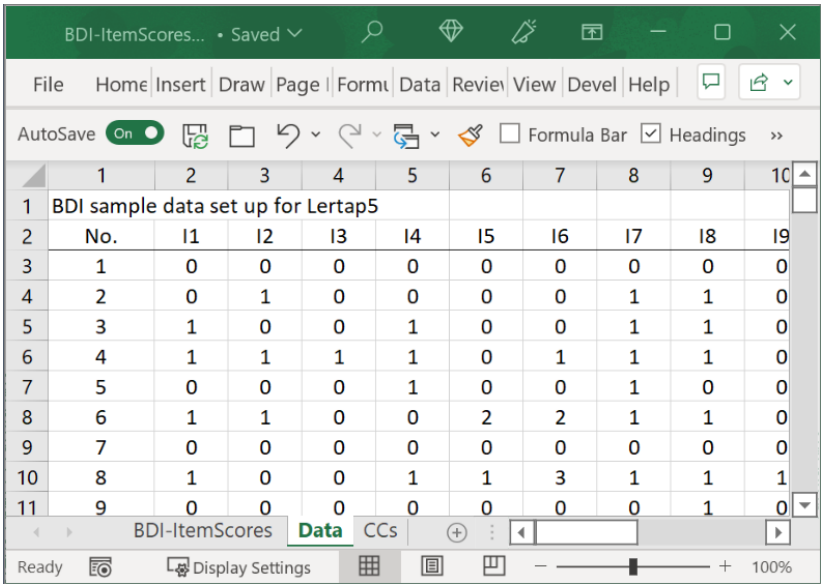


Figure 21

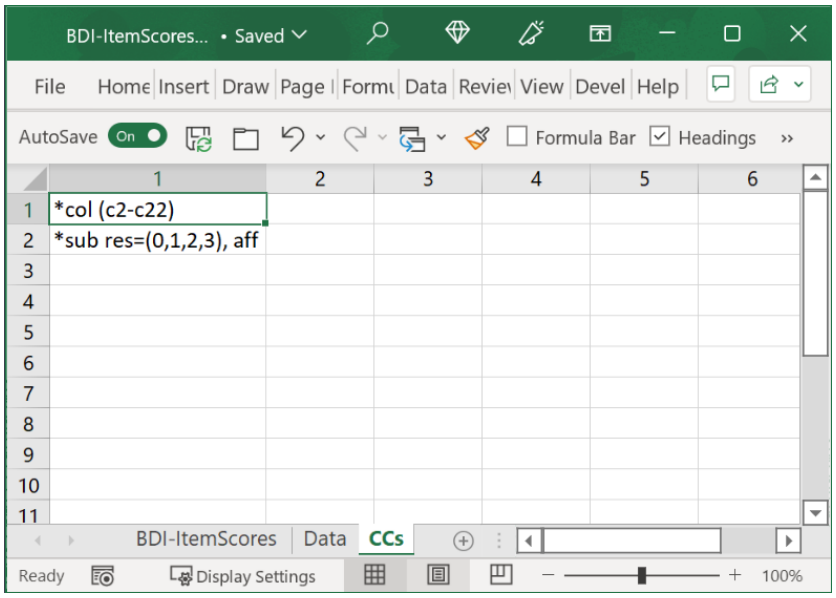


Figure 22

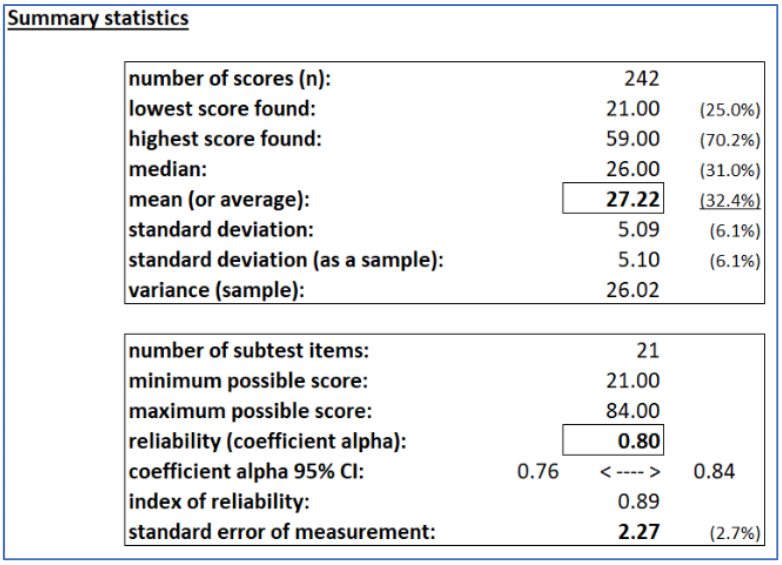
Figures 21 and 22 show how my Data and CCs sheets look at this point.

To get results, I start with the activities mentioned above on page 2 “Get Freqs”. Then I study the Freqs report produced.

I follow that by “get item statistics and student scores with “Elmillon”; refer to page 4 above. Some of Elmillon’s results are seen in Figure 23.

Now, something to notice: these results do not agree with those I obtained earlier.

Compare Figure 23 results with those in Figure 12 on page 8 above.



Summary statistics

number of scores (n):		242	
lowest score found:		21.00	(25.0%)
highest score found:		59.00	(70.2%)
median:		26.00	(31.0%)
mean (or average):		**27.22**	(32.4%)
standard deviation:		5.09	(6.1%)
standard deviation (as a sample):		5.10	(6.1%)
variance (sample):		26.02	
number of subtest items:		21	
minimum possible score:		21.00	
maximum possible score:		84.00	
reliability (coefficient alpha):		**0.80**	
coefficient alpha 95% CI:	0.76	< ---- >	0.84
index of reliability:		0.89	
standard error of measurement:		**2.27**	(2.7%)

Figure 23

In Figure 12, the "minimum possible score" is 0.00 – in Figure 23 it's 21.00.

The cause of the disagreement is that I haven't told Lertap5 how to convert the res=(0,1,2,3) specification seen in Figure 22 into item scores of zero, one, two, and three.

When Lertap5 sees res=(0,1,2,3) it doesn't really see 0,1,2,3 as numbers. It sees them as characters. Unless informed otherwise, the first character will convert to a score of one, the second to a score of two, the third to a three, the fourth to a four.

The reason it does this is because Lertap5's default assumptions are (1) the instrument is using a Likert scale, and (2), the entries following an res= speciation on a *sub line are characters – they're **never** to be seen as numbers (even when they may sure look like numbers to we humans).

If I add the following line to the CCs sheet all will be well:

*mws call, 0, 1, 2, 3

"*mws" cards in Lertap5 are ones having a "multiple-weights specification".

For example, assume res=(0,1,2,3) as above. A line with *mws c1, 0, 1, 2, 3 would be Lertap5's way of telling Excel that, for column 1, those four response possibilities (0,1,2,3) are to be interpreted as numbers equalling 0.00, 1.00, 2.00, and 3.00.

If I had had res=(A,B,C,D), the same *mws line would convert a response of A to 0.00 scoring points, B to 1.00, C to 2.00, and 3 to 3.00 points. Again I point out that that the values within the parentheses after res= are seen as characters. Even when they appear to be numbers, they're not. *mws lines are used to convert the characters to numbers so that Lertap5 can compute its item statistics and test scores.

The example I've just given starts with *mws c1, -- the "c1" means "column 1".

A line with *mws call, means "columns all", or, "all columns" as specified in the *col line of a CCs worksheet.

This help topic might be a useful reference.

The CCs lines in Figure 24 are the ones which were really used when I got Lertap5 to produce the results seen in Figure 12.

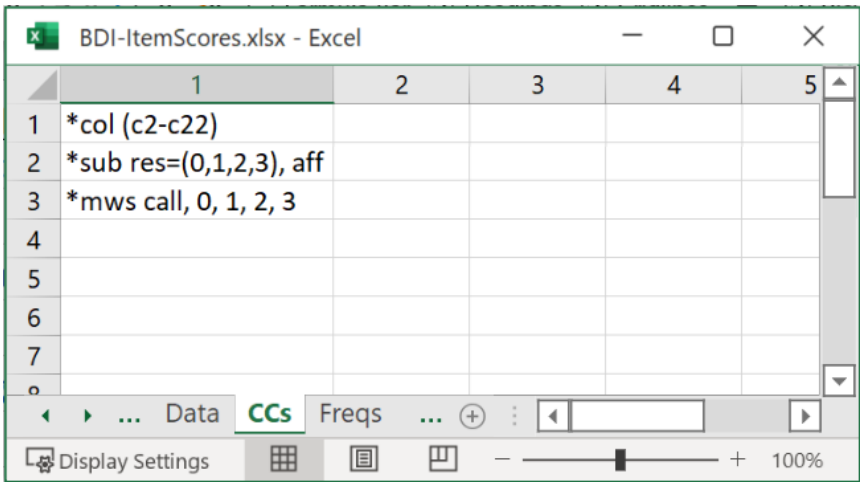


Figure 24

● **Postscript**

By far the most common scale encountered in affective instruments (in my experience) is the Likert scale, with response possibilities of “strongly disagree” to “strongly agree”.

*To me, *mws lines in Lertap5 are wonderful, I love them!*
Strongly disagree ___ ___ ___ ___ ___ *Strong agree*

Usually, should a respondent tick the first option their response would be coded as a “1”. Those taking the last option would have their response coded as a “5”. This is so common it is Lertap5’s default assumption. Instead of *sub res=(0,1,2,3), aff I would only need to have *sub aff and there would be no need at all for a *mws line. These two lines would to the job for me:

```
*col (c2-c22)

*sub aff
```

These lines would also work:

```
*columns (c2-c22)

*subtest affective
```

*mws “cards” are used only in special cases, particularly when a Likert scale has not been used. A reference is here.