Group work - field test answers

As part of your research investigating the effect of a multidisciplinary intervention for patients with chronic whiplash associated disorder (WAD) in patients seen in primary and secondary sectors of the Danish health care system, you are continuing your work on the Neck Disability Index (NDI).

Recall from the introductory course that the NDI has 10 items, each item has 6 response options, and the scale range is 0-100 (high score equals high disability). It is based on a reflective model.

(You can view the NDI in full by clicking here: Neck Disability Index)

Assignment

You have decided to perform a field test procedure on the NDI to see how the data structure looks like and how the items are performing.

For this you need to download the dataset: field_test_data.zip (unpack the zip-file).

You want to look at the properties listed below. **Try to solve the questions using Stata without looking at the hints.** However, if you run into trouble, please use them.

1. Item characteristics

1.1 Item scores

Look at the median, mean, and missings for each item.

Hint

Use the `fsum` command at baseline with the relevant stats for each item.

```
ssc install fsum //Installing the fsum ado file in Stata's command window
```

Question

1.1 Describe your findings and discuss what they mean.

Answer

Table 1.1 Item characteristics

Item	Content	Mean (N=326)	Median (N=326)	Missing items (% of 326)
1	Pain intensity	2.39	2	0.61
2	Personal care (washing, dressing etc)	0.95	1	0.31
3	Lifting	1.65	1	2.15
4	Reading	1.76	2	0.31
5	Headache	2.40	3	0.31
6	Concentration	1.19	1	0.61
7	Work	1.84	2	2.15
8	Driving	1.83	2	5.21
9	Sleeping	1.78	2	0.31
10	Recreation	2.35	2	2.15

Item scale range is [0-5]. Items 1, 5 and 10 have acceptable means and medians. The other items seem to have somewhat lower scores which needs to be analysed looking at the item distribution. The % of missing items is acceptable, however, item 8 has twice as many missing items compared to the second highest item. It is worthwhile looking into this when performing cognitive interviews.

1.2 Item distribution

Look at how the scores of each item is distributed across the answer categories.

Hint

Use the `tabulate` command with the options `missing details` at baseline for each item.

Question

1.2 Describe your findings and discuss what they mean.

Answer

Table 1.2 Item distribution

		Distribution (%) over response options							
Item	Content	0	1	2	3	4	5	Skewness	Kurtosis
1	Pain intensity	2.5	13.2	42.9	27.9	9.8	3.1	0.3	3.2
2	Personal care (washing, dressing etc)	38.3	33.4	23.3	3.9	0.6	0.3	0.8	3.4
3	Lifting	19.3	39.9	12.9	14.1	11.7	2.2	0.7	2,4
4	Reading	19.0	25.5	22.4	26.7	5.5	0.6	0.1	2.0
5	Headache	16.9	15.6	13.8	30.4	10.1	12.9	-0.0	2.0
6	Concentration	31.9	33.4	19.6	11.7	2.8	0.0	0.7	2.6
7	Work	23.9	21.8	18.4	21.2	5.2	7.4	0.5	2.3
8	Driving	16.9	25.8	20.3	23.6	5.2	3.1	0.3	2.4
9	Sleeping	15.3	31.6	24.2	17.8	9.2	1.5	0.4	2.4
10	Recreation	7.4	27.9	21.2	16.9	13.8	10.7	0.3	2.0

Most items have some degree of right skewedness (items 2,3,6,7,8,9,10) but are relatively normally distributed. Item 2 is somewhat peaked and right skewed and maybe a candidate for omission if it makes sense clinically.

2. Reproducibility

2.1 Internal consistency

Calculate the internal consistency (both overall and at item level).

Hint

Use the `alpha` command with the options `std item` at baseline for all item.

Questions

2.1.1 Describe your findings and discuss what they mean.

Answer

```
. set linesize 100

. alpha n* if bafu==1, std item label

Test scale = mean(standardized items)
```

Item	Obs	Sign	item-test corr.	item-rest corr.	interitem corr.	alpha	Label
n1	324	+	0.6857	0.5975	0.4484	0.8798	Pain intensity
n2	325	+	0.7408	0.6646	0.4376	0.8751	Personal care
n3	326	+	0.7429	0.6672	0.4370	0.8748	Lifting
n4	325	+	0.7271	0.6481	0.4401	0.8761	Reading
n5	325	+	0.4693	0.3459	0.4912	0.8968	Headache
n6	324	+	0.7212	0.6408	0.4412	0.8767	Concentration
n7	319	+	0.7904	0.7258	0.4272	0.8704	Work
n8	309	+	0.7529	0.6786	0.4355	0.8741	Driving car
n9	325	+	0.6775	0.5859	0.4504	0.8806	Sleep
n10	319	+	0.7701	0.7013	0.4305	0.8719	Recreation
Test scale					0.4439	0.8887	mean(standardized items)

> The headache item (n5) shows significantly lower item-rest correlation compared to the other items. Second, the inter-item correlation and alpha will increase substantially if this item is omitted from the scale. This is probably a (slightly) misfitting item, and we need to look carefully how it behaves in our factor analyses.

2.1.2 Should we remove an item? If so, which one(s)?

Answer

> Yes, I would remove the headache item as (a) it is probably measuring something different conceptually, and (b) it is misfitting. However, I would only do this if my factor anaysis gives a similar answer.

If you believe one or more items should be removed, please run the new analysis.

Hint

Use the same command as before without the poorly fitting item.

Question

2.1.3 Describe your findings of the new internal consistency analysis and discuss what they mean.

Answer

Lets run alpha again but without item 5:

```
. alpha n1-n4 n6-n10 if bafu==1, std item label

Test scale = mean(standardized items)

                          item-test  item-rest  interitem
Item         |  Obs  Sign   corr.      corr.       corr.     alpha   Label
-------------+-----------------------------------------------------------------
n1           |  324    +    0.6937     0.6009     0.5028    0.8900   Pain intensity
n2           |  325    +    0.7606     0.6838     0.4869    0.8836   Personal care
n3           |  326    +    0.7608     0.6843     0.4867    0.8835   Lifting
n4           |  325    +    0.7280     0.6434     0.4945    0.8867   Reading
n6           |  324    +    0.6912     0.5981     0.5031    0.8901   Concentration
n7           |  319    +    0.7940     0.7254     0.4779    0.8799   Work
n8           |  309    +    0.7654     0.6889     0.4857    0.8831   Driving car
n9           |  325    +    0.6846     0.5878     0.5052    0.8909   Sleep
n10          |  319    +    0.7915     0.7232     0.4781    0.8799   Recreation
-------------+-----------------------------------------------------------------
Test scale   |                                    0.4912    0.8968   mean(standardized items)
-------------------------------------------------------------------------------
```

This has increased the interitem corr. a fair amount (from 0.44 to 0.49) and alpha slightly (from 0.89 to 0.90). In fact our factor analyses revealed item 5 as a misfitting item, and it was one of two items that were removed in the final Danish 8-item version which was published.

NB: Remember to use standard scores by adding the option `std`. This gives the interitem correlation rather than the interitem covariance.

2.2 Reliability

Please determine the reliability (ICC-values) at item level.

Hint

Use the `icc` command (in older versions of Stata, use the `icc23` command) for each item. The `forvalues` command can be used to make the sytax shorter.

```
icc
icc23 //Stata<12: Install it by typing - scc install icc23 - in Stata's command window
```

You can use the `forvalues` command to make the syntax shorter.

```
forvalues i = 1(1)10 {
icc n`i' idnr bafu if stable==2, absolute
quietly display as text "The ICC for Question `i' = " _col(40) as result %5.4f
r(icc_i)
quietly display as text "No of stable pts for Question `i' = " _col(40) r(N_target)
}
```

The code above calculates an ICC for each item and displays the ICC and the number of stable patients for each item. If you want to see the ICC output, you can remove ‘quietly’.

Question

2.2 Describe your findings and discuss what they mean.

Answer

Table 2.1 Item reliability

Item	Content	ICC 2,1 (agreement)
1	Pain intensity	0.60
2	Personal care (washing, dressing etc)	0.72
3	Lifting	0.80
4	Reading	0.81
5	Headache	0.87
6	Concentration	0.81
7	Work	0.88
8	Driving	0.74
9	Sleeping	0.81
10	Recreation	0.76

All items have an acceptable ICC (>0.5), but the ICC for the pain item is relatively low (0.60) compared to the other items. We need to keep an eye on this item in the other analyses.

3. Floor and ceiling effect

3.1 Conventional method

Calculate the floor and ceiling effect using the 'conventional' method.

Hint

Use the `tabulate` command to find the floor and ceiling effects.

Question

3.1 Describe your findings and discuss what they mean.

Answer

See the answer below (under scale width method)

3.2 Scale width method

Calculate the floor and ceiling effect using the 'scale width method'.

Hint

Use the `concord` command. Install the command by typing `scc install concord`. The graph in the code is for illustration of the LOA and exporting it can be omitted.

Then use the following code:

```
sort idnr
keep idnr bafu stable NDI10

reshape wide NDI10 stable, i(idnr) j(bafu)
drop stable1
order idnr stable2 NDI10*

keep if stable==2

concord NDI*, loa(lopts(lp(dash..)))
graph export "B_A_plot, all pts.wmf", replace
```

NB: do not **save** the dataset after running this code as it will change the dataset. If you need a clean dataset, please download it from the webpage.

You can use the `concord` output to find the measurement error, and then find the % of patients who fall within measurement error at each end of the scale.

Question

3.2 Describe your findings and discuss what they mean.

Answer

Table 3.2 Floor and ceiling effects

	Conventional method (%)	Scale width method (%)
Floor effect	0.3	10.1
Ceiling effect	0.0	0.9

The conventional method shows no floor or ceiling effects (<15%), however, the scale width method shows a floor effect of 10.1% which is borderline. If we want to address this, we need to look at how the answer categories are formulated in the questions which have most right skewness.