Advanced couse in questionaire research
Exercises in validity

STATA
All the Stata commands that are necessary to complete the following exercises can be found on the last page of this document.

BACKGROUND INFORMATION
SF-36:
The SF-36 is a multi-purpose, short-form health survey with only 36 questions. It yields an 8-scale profile of scores (physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional and mental health) as well as physical and mental health summary measures (PCS and MCS). It is a generic measure, as opposed to one that targets a specific age, disease, or treatment group. In these exercises we will only be using the PCS and MCS. The two summary measures (PCS and MCS) have been standardized to have a mean of 50 and a standard deviation of 10 in 1998 general U.S. population, with higher scores representing better health.

Whiteley-7 index:
The Whiteley 7-Index (WI-7) measures health anxiety (aka. hypochondriasis). It consists of 7 items and is scored on a rating scale from 1-5 with a total sumscore ranging from 7-35. The scores were transformed to a 0-100 score point scale with higher scores indicating better health.

SCL-8:
The Symptom Checklist (SCL-8) consists of 8 items measuring self-reported symptoms of emotional distress. The items are scored on a rating scale from 1-5 with a total sumscore ranging from 8-40. The scores were transformed to a 0-100 score point scale with higher scores indicating better health.

EXERCISE 1.1 (criterion validity, continuous instrument and gold standard)
The data we will consider in this exercise, "*criterion validity sf12.dta*", originates from a randomized controlled study (RCT). The main purpose of the RCT was to investigate the effect of Acceptance and Commitment Therapy (ACT) on 126 patients with severe health anxiety. During the study patients were asked to complete several questionnaires (among others SF-36 and SF-12) at different timepoints.

In this exercise though, we will consider the concurrent criterion validity of the Physical Component Summary (PCS), as computed by the SF-12, by comparing it to PCS calculated by SF-36 (the gold standard). As a help you can find SF-36 norm data from a general U.S population, as well as from a Danish sample of patients with health anxiety, below. Below you find all the Stata commands that are necessary to do the analysis.

1. Go through step two and three in the process of criterion validation (see slide 13 from todays lecture). That is; discuss whether the choice of gold standard and sample is appropriate for the purpose.

It is known, that for *this group* of patients, a change in PCS larger than ±5 (equal to half a standard deviation) is considered clinically significant. Also, changes less than ±5 is considered as “no change”. Hence we choose the required level of agreement to be between -5 and 5.

2. Find the limits of agreement (aka. 95% prediction interval).

3. Test whether there is a systematic difference between the two scores.

4. Make the Bland-Altman plot.

5. Check the assumptions behind the analysis (see slide 25 from todays lecture).

6. Make a summary of your findings and discuss the results.

EXERCISE 1.2 (criterion validity, dichotomous instrument and gold standard)
In a study similar to the one used in the example of concurrent validity in section 6.4.1 in “Measurement in Medicine”,16.060 patients were screened for cancer using a gold standard and a new test instrument. The data can be found in the file “*criterion validity cancer.dta*”.

The researchers hypothesized that the new instrument is as good as the gold standard, with a sensitivity of 85%, specificity of 95%, positive predicted value (PPV) of 80% and a negative predicted value (NPV) of 90%. Below you find all the Stata commands that are necessary to do the analysis.

1. Make a crosstable comparing the two methods. Identify the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

2. Calculate the sensitivity and specificity by "hand".

3. Calculate the negative and positive predicted value by "hand".

Download the **`-diagt-`** command in Stata by typing **`-ssc install diagt-`** in Stata’s command window.

4. Use the **`-diagt-`** command to find the sensitivity, specificity, the negative predicted value, the positive predicted value, and their 95% CI's.

5. Why are the 95% CI's for the specificity and negative predicted value so narrow compared to the CI's for the sensitivity and positive predicted value?

6. Summarize and interpret your findings. Is the data in correspondence with the researchers hypotheses about the sensitivity, specificity, PPV and NPV?

EXERCISE 1.3 (Construct validity)
In this exercise we will consider only a part of the data from the RCT described in Exercise 1.1, but now our focus is on assessing the convergent, discriminant and discriminative validity of the WI-7 (see page for a description of WI-7). The data can be found in the file "*construct validity.dta*".

Hypotheses regarding convergent and discriminant validity:

- A moderate to strong positive correlation of 0.5 between WI-7 and MCS. A strong correlation of 0.8 between WI-7 and SCL-8.

- A low correlation between of 0.2 between WI-7 and PCS

Hypotheses regarding discriminative validity:

- Men had a 10 point higher (indicating better health) mean score than women on the WI-7 at baseline.
- Patients with at least one psychiatric comorbidity had a 15 point lower score (indicating worse health) than patients without psychiatric comorbidity on the WI-7 at baseline.

The **`-corrci-`** command can be used to compute 95% CI's for the Pearson correlation between two or more numerical variables (and with a little extra manual labour it can also be used to compute 95% CI's for the Spearman correlation). You can download the command bytyping **`-search corrci-`** in Stata's command window and follow the instructions.

1. Test the hypotheses regarding convergent and discriminant validity, and check the assumptions underlying the tests.

2. Test the hypotheses regarding discriminative validity, and check the assumptions underlying the tests.

3. Summarize and interpret your findings.

EXERCISE 1.4
Think or your own data and choose a specific variable/scale. Specify hypotheses regarding direction and strength of correlations between your chosen scale and other variables/scales.

Stata commands

General
Change directory: -**cd "N:\Afdeling\AUHFKLID\ …"**- (insert your own path)

Read data: -**use "nameofdata.dta", replace**-

Exercise 1.1
Generate differences: **gen diff = pcs_sf12 - pcs_sf36**

Generate averages: **gen ave = (pcs_sf12 + pcs_sf36)/2**

Limits of agreement: **centile diff, centile(2.5 97.5) meansd**

BA plot: **scatter diff ave, yline(-2.1) yline(-78.1 73.9, lcol(blue))**

QQ-plot: **qnorm diff**

Paired t-test: **ttest wpl=wml**- or -**ttest diff==0**

Exercise 1.2
2 x 2 table: **tab criterion test, mis**

Using Stata as a calculator: **display 2*2**,
display (4*2)-(8/2)

Spensitivity, specificity, PPV and NPV: **diagt criterion test**

Exercise 1.3
Scatterplots: **graph matrix whiscore7_1 scl8score_1 pcs_1 mcs_1, half**

Correlations: -**corr**- and/or -**pwcorr**-

Correlations WITH CI's: -**corrci**-

Unpaired t-test: **ttest whiscore_1, by(gender)**

Test of equal standard deviations: **sdtest whiscore_1, by(gender)**