Examining agreement

Contents

- Rather than looking for differences between groups, we may want to check whether measurements taken on the same subject show **agreement**
- We can assess agreement between continuous data (i.e. taking measurement on people/things)
  - displaying data
  - measuring agreement
- Measuring agreement for categorical data (i.e. counting people/things) or ordered categories (i.e. ordinal data)

Agreement between continuous variables

- We may want to check whether continuous measurements taken on the same subject show **agreement**
- This is different from paired data where we are testing to see if there has been any change within each subject
- For agreement studies, measurements are taken on the **same subject using different methods or observers**
- **Examples**
  - **Method comparison**: Using MRI and Echo to measure ejection fraction on the same patient
  - **Inter-observer agreement**: Two different doctors using Echo to measure the ejection fraction on the same patient
  - **Intra-observer agreement**: The same doctor assessing the ejection fraction of the same MRI scan at two different timepoints

Agreement between continuous variables

- Examining the correlation (and corresponding p-value) is not the correct approach for this type of study because we expect two measurements of same thing to be highly correlated
- We need to investigate whether the two measurements give the **same value** from the different methods or observers

- One popular method for this are **Bland-Altman plots**
- This is a **visual** method of assessing agreement, there are no p-values calculated here
- These plots look at the difference between the two measurements
- Limits of agreement are calculated, between which 95% of differences would be expected to lie

Agreement between continuous variables

- The difference is calculated in each pair of measurements
- The average of the two measurements is taken as a best estimate of the ‘true’ value
- A scatter plot is then drawn with the average on the x-axis and the difference on the y-axis
- Limits of agreement estimated using

  mean difference ± 1.96*SD of difference

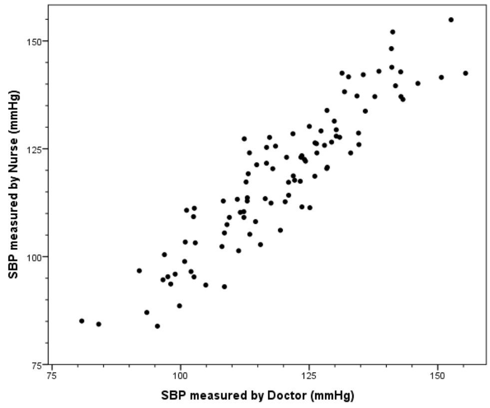
Agreement between continuous variables

- If there is good agreement between the measurements then the plot should have the following features
  - Mean difference close to zero
  - Differences randomly scattered around zero, with no obvious pattern or trend
  - Limits of agreement should be within clinically unimportant boundaries (this can be quite subjective!)

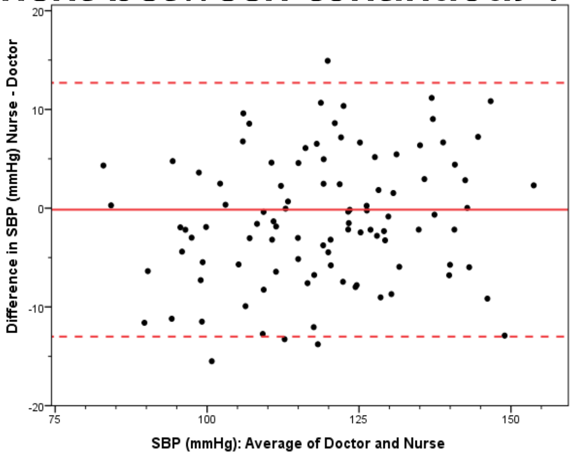
Agreement between continuous variables

- **Example**: A Dr and a Nurse in a GP surgery both measure the systolic blood pressure of 100 consecutive patients
- Table shows the results for the first 5 patients
- Scatter plot shows clear correlation (as expected), but how well do they agree?
- Mean (SD) of the difference for all 100 patients is -0.15 (6.53) mmHg

Doctor	Nurse	Difference	Average
131	133	2	132
114	120	6	117
133	128	-5	130.5
157	151	-6	154
118	122	4	120



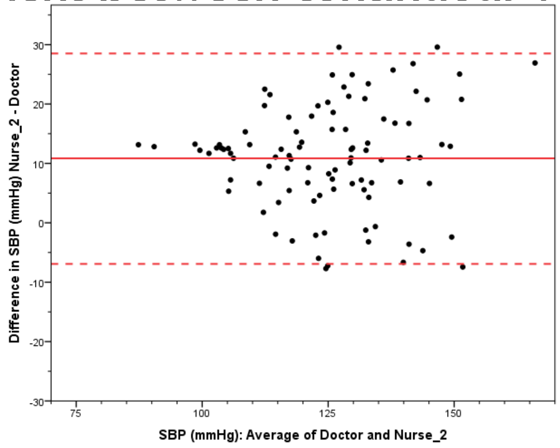
Agreement between continuous variables



This plot has a mean difference close to zero, and no clear pattern to the observed differences

Limits of agreement suggest that the Dr and nurse would disagree by around ± 13mmHg for 95% of patients – is this clinically unimportant?

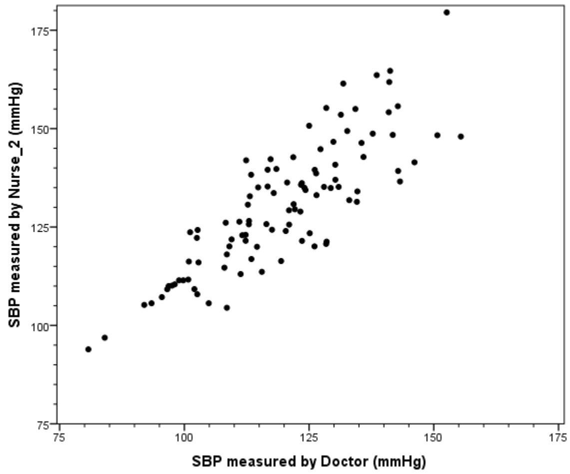
Agreement between continuous variables



This plot has a mean difference above zero (~10mmHg), and a pattern to the observed differences

Limits of agreement suggest that the Dr and nurse would disagree by around -7 to +28mmHg for 95% of patients

Agreement between continuous variables



This information (i.e. mean difference above zero (~10mmHg) and pattern to the observed differences) would not be as clear in a scatterplot.

Agreement between continuous variables

- The amount of agreement between continuous variables can also be measured quantitatively (i.e. an index of reliability)
- We use the intraclass correlation coefficient (ICC), taking a value between 0 (no agreement) and 1 (perfect agreement)
- This represents the amount of variation due to differences between pairs of measurements as a proportion of the total variation (i.e. between all observations)
- This can be calculated in software. In the previous example the nurse and doctor had very high agreement, ICC=0.91

Agreement between categorical variables

- Similarly we can examine agreement between categorical data (e.g. between two assessments using an ordinal scale)
- Typically this data is displayed in a frequency table one set of observations counted horizontally and the other vertically
- The diagonal values are then the number of times the observers agree
- Off-diagonal values are instances where the two observers have disagreed

Agreement between categorical variables

- For example, we have assessments for two observers rating the density (low, medium, high) of cells in the same 100 samples using a new highlighting technique

	Low	**Medium**	**High**	***TOTAL***
Low	24	10	2	*36*
Medium	6	32	8	*46*
High	1	4	13	*18*
TOTAL	*31*	*46*	*23*	*100*

- The observers agree in 69% of the cases, i.e. (24+32+13)/100 and disagree in 31% of the cases, i.e. (10+2+6+8+1+4)/100

Agreement between categorical variables

- Again examining the (non-parametric) correlation is not the correct approach for this type of study
- We use Cohen’s kappa statistic (in the e.g. κ=0.59), interpreted as:
  - < 0.2 poor agreement (0=no better than by chance)
  - 0.2 to 0.4 fair agreement
  - 0.4 to 0.6 moderate agreement
  - 0.6 to 0.8 good agreement
  - 0.8 to 1 very good agreement (1=perfect agreement)
- We can also take into account the extent of disagreement by using a weighted kappa (e.g. further penalises the “2” and “1”)