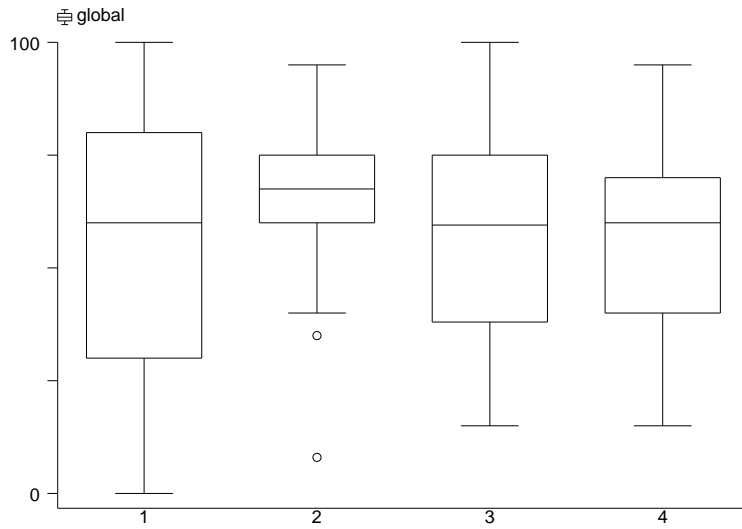


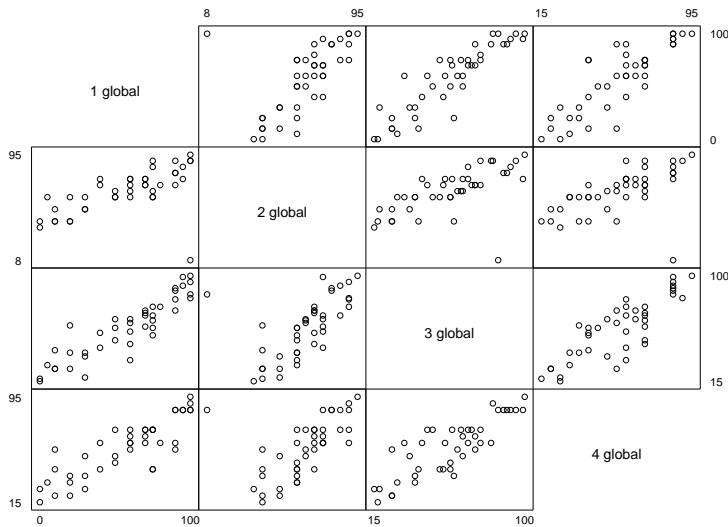
# Measuring Agreement

## The Intraclass Correlation Coefficient

In the larger scope of the previously described study of hand and wrist X-rays, 4 clinicians made independent evaluations of each of 42 X-rays. The global scores are plotted below, by "rater".



The agreement can also be assessed by looking at scatter-plots.



To obtain an overall measure of inter-rater reliability (agreement) we can calculate an intra-class correlation coefficient (ICC).

One-way Analysis of Variance for global:

Number of obs = 168  
 R-squared = 0.7751

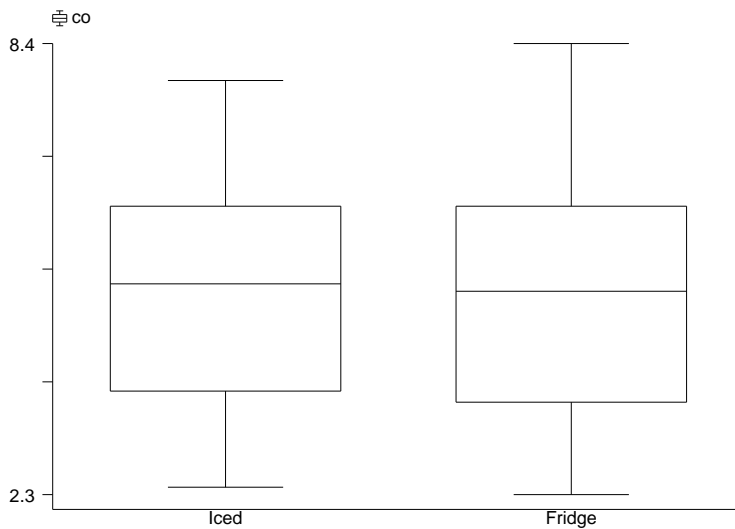
Source	SS	df	MS	F	Prob > F
Between xrayid	76067.476	41	1855.3043	10.59	0.0000
Within xrayid	22072.5	126	175.17857		
Total	98139.976	167	587.66453		

Intraclass correlation	Asy. S.E.	[95% Conf. Interval]	
0.70569	0.05832	0.59139	0.81998

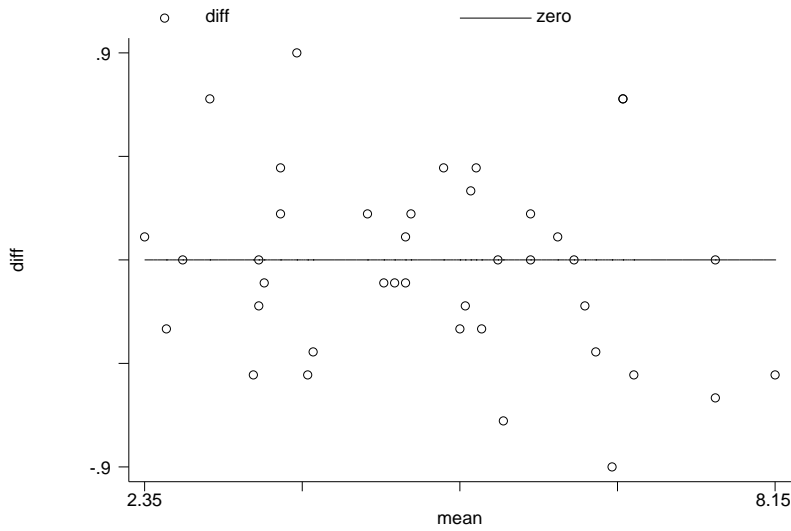
Estimated SD of xrayid effect 20.49467  
 Estimated SD within xrayid 13.2355  
 Est. reliability of a xrayid mean .9055796  
 (evaluated at n=4.00)

### Inter-method Comparison

Thermo-dilution Cardiac Output is measured by infusing cold water into the circulation and monitoring the down-stream drop in temperature. The standard protocol for doing so involved using iced water. It was suggested that cold water from the refrigerator would do just as well and be more convenient. 40 patients were assessed by both methods.



To assess the pattern of agreement, the differences between paired measures can be plotted versus subject means (Bland & Altman, Lancet 1986).



### The Kappa Coefficient

The corresponding problem can arise in the case of nominal assessments (e.g. dichotomies such as Present/Absent, Normal/Abnormal). For example, consider the table below, which represents the results of assessments of 100 liver-spleen scans as read by two physicians:

phys1	phys2		Total
	Abnormal	Normal	
Abnormal	21	14	35
Normal	8	57	65
Total	29	71	100