

Test-Retest Reliability (ICC) and Day to Day Variation

Consistency of Measurements

- Reliability refers to the consistency of a test or measurement.
- A test cannot be considered valid if it is not reliable.
- You should know the day to day variation in your dependent variable.
- How much does 1RM bench press change from day to day?
- How much does VO₂ max change from day to day?

Subject	Trial 1	Trial 2	Difference
1	146	140	-6
2	148	152	+4
3	170	152	-18
4	90	99	+9
5	157	145	-12
6	156	153	-3
7	176	167	-9
8	205	218	+13
Mean	156	153	2.75
SD	32.8	32.9	10.7

Test-Retest Reliability

- The 1RM squat strength was measured on 8 subjects on two separate days.
- In this experiment we want to measure the reliability of our measurements.

Here is the day to day variation in 1RM bench press strength.

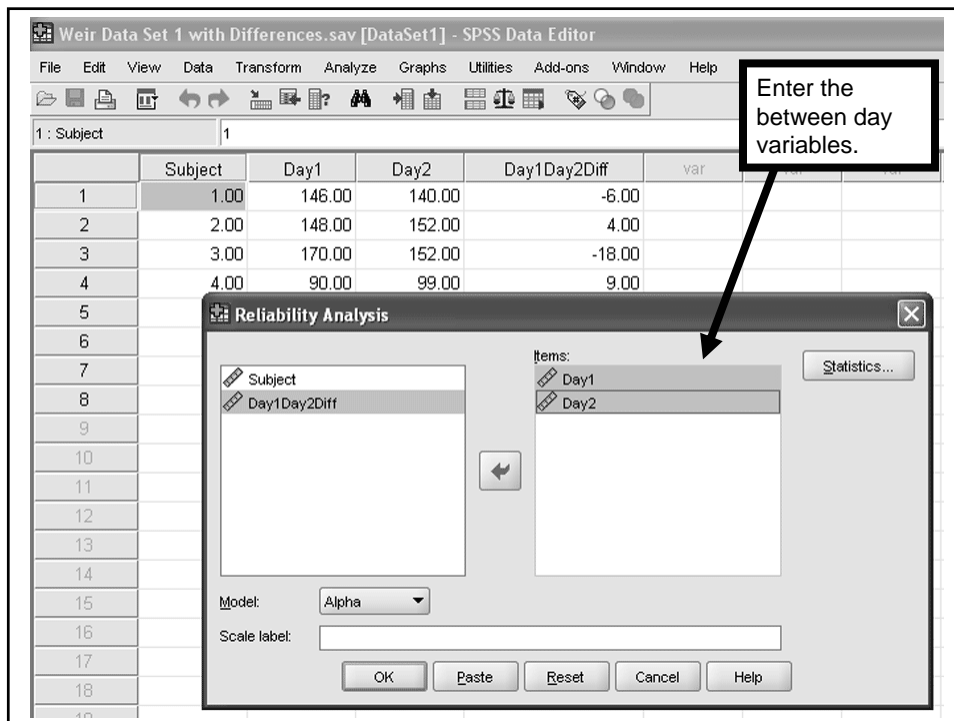
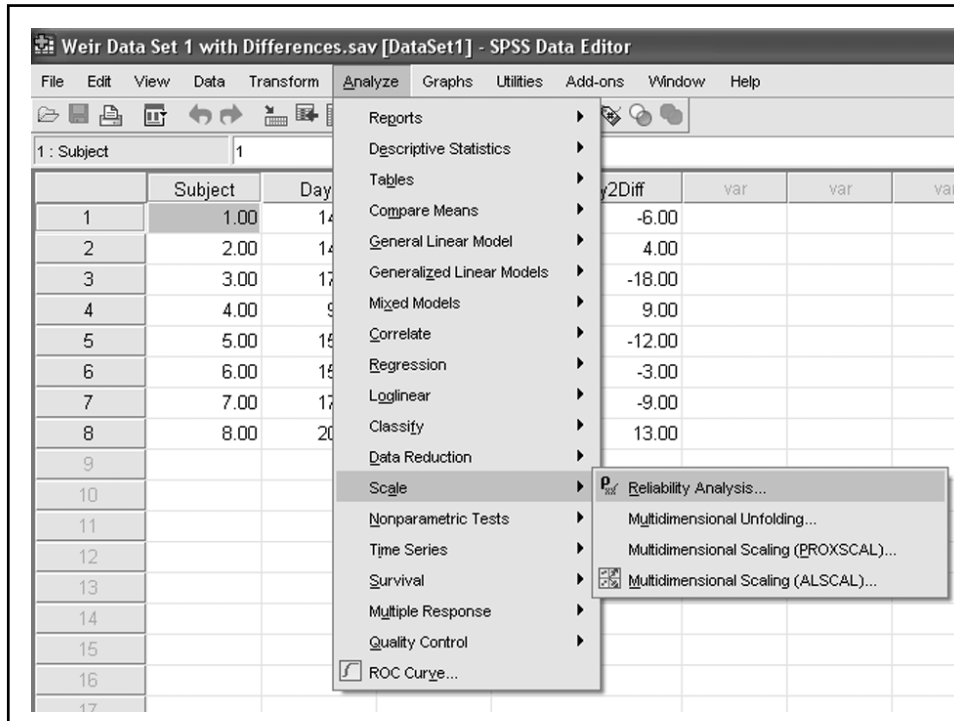
In this experiment, we plan to average trials, so we will use Averaged Measures ICC.

Weir Data Set 1 with Differences.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window

1 : Subject 1

	Subject	Day1	Day2	Day1Day2Diff	va
1	1.00	146.00	140.00	-6.00	
2	2.00	148.00	152.00	4.00	
3	3.00	170.00	152.00	-18.00	
4	4.00	90.00	99.00	9.00	
5	5.00	157.00	145.00	-12.00	
6	6.00	156.00	153.00	-3.00	
7	7.00	176.00	167.00	-9.00	
8	8.00	205.00	218.00	13.00	



Weir Data Set 1 with Differences.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1: Subject 1

Subject	Day1	Day2	Day1Day2Diff	var	var	var	
1	1.00	146.00	140.00	-6.00			
2	2.00						
3	3.00						
4	4.00						
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							

Click Statistics

Reliability Analysis: Statistics

Descriptives for:

- Item
- Scale
- Scale if item deleted

Inter-Item:

- Correlations
- Covariances

Summaries:

- Means
- Variances
- Covariances
- Correlations

ANOVA Table:

- None
- F test
- Friedman chi-square
- Cochran chi-square

Hotelling's T-square

Tukey's test of additivity

Intraclass correlation coefficient

Model: Two-Way Mixed Type: Consistency

Confidence interval: 95 % Test value: 0

Continue Cancel Help

Statistics...

Weir Data Set 1 with Differences.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1: Subject 1

Subject	Day1	Day2	Day1Day2Diff	var	var	var	
1	1.00	146.00	140.00	-6.00			
2	2.00	148.00	152.00	4.00			
3	3.00	170.00	152.00	-18.00			
4	4.00	90.00	99.00	9.00			
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							

Reliability Analysis

Model: Alpha

Scale label:

Items:

- Day1
- Day2

Subject

Day1Day2Diff

OK Paste Reset Cancel Help

Click OK to run

Statistics...

SPSS Output

Inter-Item Correlation Matrix

	Day1	Day2
Day1	1.000	.947
Day2	.947	1.000

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between People	14689.750	7	2098.536		
Within People				.530	.490
Between Items	30.250	1	30.250		
Residual	399.750	7	57.107		
Total	430.000	8	53.750		
Total	15119.750	15	1007.983		

Grand Mean = 154.6250

Intraclass Correlation Coefficient

	Intraclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig.
Single Measures	.947 ^b	.761	.989	36.747	7	7	.000
Average Measures	.973 ^c	.864	.995	36.747	7	7	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between People	14689.750	7	2098.536		
Within People				.530	.490
Between Items	30.250	1	30.250		
Residual ^a	399.750	7	57.107		
Total	430.000	8	53.750		
Total	15119.750	15	1007.983		

Grand Mean = 154.6250

a. Tukey's test for nonadditivity is undefined for dichotomous data.

There is no difference between trials $F(1,7) = .530, p = .49$
 If there is a difference between trials you may have a learning effect or a fatigue effect and you must modify your methods to control for learning and/or fatigue.

Intraclass Correlation Coefficient

	Intraclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.947 ^b	.761	.989	36.747	7.0	7	.000
Average Measures	.973 ^c	.864	.995	36.747	7.0	7	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

- a. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.
- b. The estimator is the same, whether the interaction effect is present or not.
- c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

ICC = .973 [High Degree of Reliability]
Using Mixed Model, Type Consistency, Averaged Measures.

Subject	Trial 1	Trial 2	Difference
1	146	140	-6
2	148	152	+4
3	170	152	-18
4	90	99	+9
5	157	145	-12
6	156	153	-3
7	176	167	-9
8	205	218	+13
Mean	156	153	2.75
SD	32.8	32.9	10.7

Day to Day Variation

- If you train subjects for 8 weeks and they improve their strength by 2.8 Kg have you actually done anything?
- Is the program effective?

• Here is the day to day variation in 1RM bench press strength.

Manuscript Methods & Results Sections

METHODS

Measurement of Reliability

Prior to initiating the study, 8 subjects (not in the experiment), participated in a test-retest assessment of measurement reliability. One RM bench press strength was measured on two separate days for subjects in the reliability study. Reproducibility of bench press strength was analyzed using SPSS (16.0 for Windows) to compute the intraclass correlation coefficient (ICC) using a two factor mixed effects model and type consistency (McGraw and Wong, 1996; Shrout and Fleiss, 1979).

RESULTS

A high degree of reliability was found between bench press strength measurements the single measure ICC was .997 with a 95% confidence interval from .915 - .997. The mean between day variation for 1RM strength was 2.75 ± 10.7 Kg.

REFERENCES

- McGraw KO and Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1: 30-46, 1996.
- Shrout PE and Fleiss JL. Intraclass correlations: Uses in assessing reliability. *Psychol Bull* 86: 420-428, 1979.