How to do Confidence Intervals?

First read Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 203-220.

Here is an example, from an experiment with a 2 x 2 within-subjects design. This was a sentence priming study in which the prime could either have one structure (the so-called PO) or another one (the DO). Additionally, the verb in prime and target sentence could be the the same or different. The dependent variable is the proportion of target sentences with the structure PO.

Here are the data:

subject	DO - same verb (dos)	PO -same verb (pos)	DO-different verb (dod)	PO-different verb (pod)
1	,17	1,00	,17	,83
2	,67	,83	,50	1,00
3	,00	,83	,50	1,00
4	,00	1,00	,00	,83
5	,67	,83	,67	1,00
6	,33	,83	,33	,50
7	,17	1,00	,50	,50
8	1,00	1,00	,83	1,00
9	,17	1,00	,67	1,00
10	,50	,67	,67	,67
11	,33	1,00	,83	,83
12	,50	,83	,17	,83
13	,50	,50	,33	,83
14	,00	1,00	,00	,67
15	,00	,33	,00	,00
16	,67	,33	,50	,00
17	,83	,83	,83	,83
18	,00	,83	,33	,67
19	,17	,67	,00	,50
20	,00	,83	,00	,50
21	,33	,67	,33	,33
22	1,00	,83	,83	1,00
23	,33	1,00	,83	1,00
24	,33	,83	,17	,50

These data were analysed with a repeated-measures ANOVA.
SPSS code:

```
GLM
  dod dos pod pos
  /WSFACTOR = prime 2 Polynomial verbrep 2 Polynomial
  /METHOD = SSTYPE(3)
  /EMMEANS = TABLES(prime)
  /PRINT = DESCRIPTIVE
```

/CRITERIA = ALPHA(.05)
*/WSDESIGN = prime verbrep prime*verbrep .*

This is what SPSS gives you.

Tests of Within-Subjects Effects
MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
PRIME	Sphericity Assumed	3,251	1	3,251	34,979	,000
	Greenhouse-Geisser	3,251	1,000	3,251	34,979	,000
	Huynh-Feldt	3,251	1,000	3,251	34,979	,000
	Lower-bound	3,251	1,000	3,251	34,979	,000
Error(PRIME)	Sphericity Assumed	2,138	23	9,294E-02		
	Greenhouse-Geisser	2,138	23,000	9,294E-02		
	Huynh-Feldt	2,138	23,000	9,294E-02		
	Lower-bound	2,138	23,000	9,294E-02		
VERBREP	Sphericity Assumed	1,852E-02	1	1,852E-02	,687	,416
	Greenhouse-Geisser	1,852E-02	1,000	1,852E-02	,687	,416
	Huynh-Feldt	1,852E-02	1,000	1,852E-02	,687	,416
	Lower-bound	1,852E-02	1,000	1,852E-02	,687	,416
Error(VERBREP)	Sphericity Assumed	,620	23	2,697E-02		
	Greenhouse-Geisser	,620	23,000	2,697E-02		
	Huynh-Feldt	,620	23,000	2,697E-02		
	Lower-bound	,620	23,000	2,697E-02		
PRIME * VERBREP	Sphericity Assumed	,167	1	,167	5,750	,025
	Greenhouse-Geisser	,167	1,000	,167	5,750	,025
	Huynh-Feldt	,167	1,000	,167	5,750	,025
	Lower-bound	,167	1,000	,167	5,750	,025
Error(PRIME*VERBREP)	Sphericity Assumed	,667	23	2,899E-02		
	Greenhouse-Geisser	,667	23,000	2,899E-02		
	Huynh-Feldt	,667	23,000	2,899E-02		
	Lower-bound	,667	23,000	2,899E-02		

PRIME

MEASURE_1					
	Mean	*Std. Error*	*95% Confidence Interval*		
PRIME			*Lower Bound*	*Upper Bound*	
1	*,389*	*,058*	*,268*	*,509*	
2	*,757*	*,047*	*,660*	*,854*	

Now we have sufficient information to calculate the Confidence Interval, for example for the effect of Prime.

- Find the MSE for the effect of interest (highlighted in yellow above)
- Find the DF of this Error Term (also in yellow)
- Find N, the number of within-subject observations on which mean is based. As you can see in the data table, the mean of PO is based on 24 (number of subjects) x 2 (number of PO conditions) = 48 observations.

If you use my CI-calculator, just type in the appropriate values, and your CI will be computed.

If you want to calculate it by hand do the following:

- Find the critical value of t (DF), e.g. from the web or a stats book[1]

Now compute your CI as

$$M \pm \sqrt{2} \cdot \sqrt{\frac{MSE}{N}} . tcrit \qquad (1)$$

To check if this is correct, do the following test:

- Get the F-value (highlighted in yellow in the example)
- Get the means of the two conditions and determine the difference (i.e., your numerical effect).

Now substitute *the square root of your actual F* in place of tcrit. The CI then obtained should be equal to your actual effect size.

This is what makes a CI (somewhat) meaningful. All things being equal, an effect equal to the CI would be significant at p = 0.05.

Interaction

Assuming a 2 x 2 within. Compute difference scores (A1 - A2) for each level of B. Now do a one-way on B. The F-value of its main effect should be equal to the F-value of the interaction.
Get the MSE of this main effect, and proceed using (1). Take care that while for the main effect you had Nm =2, now you have Nm = 1!

If you now do the test with the actual F, your C.I. should equal (A1B1 - A2B1) - (A2B1 - A2B2), thus the difference of the differences.

[1] Stats books have tables with critical values! This is highly useful, for example when you read something like F(1, 24) = 2.95, p = .05 and have some doubts whether this is right.