How to calculate effect sizes in JMP

(or SPSS where needed)

[by Bernhard Riecke, iSpaceLab.com]

Disclaimer: this is work in progress - in case you find any errors or have suggestions for improvement, please email me at b_r@sfu.ca

Table of Contents

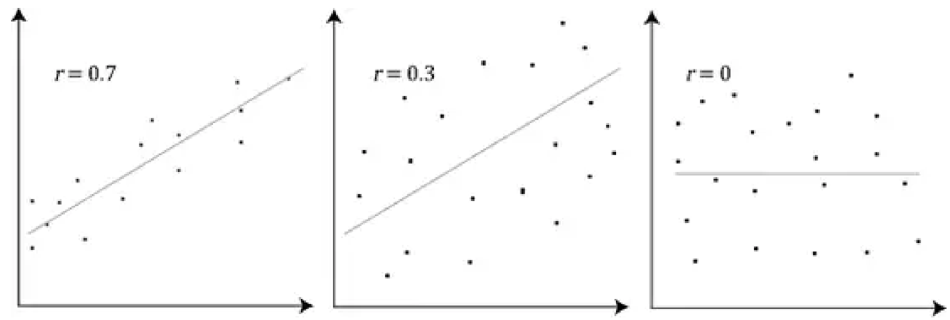
Effect Size in a Nutshell

The p-value of the inferential stats only tells you if the result is “statistically significant” in the sense that, e.g., for a p-value of p = .04 you’d expect that if you ran the same experiment 100 times all but 4 of these (100-4 = 96) would show the same effect. It does not tell you if the effect is meaningful, or of a size that matters. For this, the effect size is useful, and should always be included in your writeup. That is, the effect size is a quantitative measure of the **magnitude of the effec**t. The larger the effect size the larger the effect (e..g, of the independent on the dependent variable).

As a rule of thumb, and according to Cohen (1988, 1992), small/medium/large effects sizes refer to

- |r| >= 0.10 (**small** effect): in this case, the effect explains r^2 = .01 = 1% of the total variance or variablity in the data.
- |r| >= 0.30 (**medium** effect): the effect accounts for r^2 = .09 = 9% of the total variance.
- |r| >= 0.50 (**large** effect): the effect accounts for r^2 = .25 = 25% of the total variance.

**r is not measured on a linear scale, so r = 0.4 is NOT twice as big as r = 0.2*



Note that different effect size measures can mean different things. E.g., **Cohen's d** is an effect size used for comparing 2 means (e.g., after a t-test), and indicates the standardized difference between two means - basically the size of the difference in standard deviations. E.g., a d = 2.2 indicates that the two groups differ by 2.2 standard deviations (or z-scores),

Ok… but what does effect size even mean?

Table on **magnitude of effect sizes** for other effect size measures like Cohen's d or η2
Effect Size Calculator
Chapter on Effect Sizes

The effect size describes how big the difference is between groups. A value of 0 means there's no effect, and a value of 1 would be a perfect effect (not a linear scale). Effect sizes are important because they go beyond the simplistic 'Does it work or not?' to the far more sophisticated, 'How well does it work in a range of contexts?'. So, for example, you could have a significance (p-value) of 0.01, which is highly significant. This means this effect probably didn't happen by chance. However, if the effect size is 0.15, then that means this effect is small and not that important.

Why report effect sizes (and not just p=values)?

Well, a significant p-value (p < .05 typically) tells us that an intervention works (or in general that the IV had an effect on the DV), whereas an effect size tells us how much it works (ie., how large the effect of the IV on the DV is). Note that the effect size does NOT increase with N (the sample size), unlike significance tests which become more significant for larger N.

Which effect size should I use?

This is discussed in various resources in more detail - e.g., the Analysis Factor explains that R squared, Eta Squared, Partial Eta Squared, and Omega Squared all have the intuitive interpretation

of the proportion of the variance accounted for (compared to, e.g., Cohen's d, which indicates the size of the difference in standard deviations or z-scores, and can be larger than 1). Some effect sizes such as eta or eta squared can be biased, and often omega squared it suggested as an alternative - see https://www.theanalysisfactor.com/effect-size/ or https://daniellakens.blogspot.com/2015/06/why-you-should-use-omega-squared.html for details.
So if your stats software provides omega squared, this might be the best option as a rule of thumb.

Independent measures (between-subject) t-test

Calculate r Manually

based on r := sqrt (t^2) / (sqrt (t^2 + df))

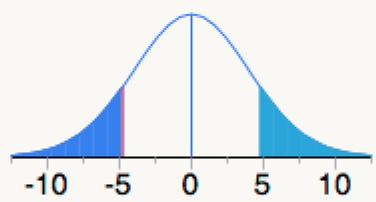
$$r^2 = \frac{t^2}{t^2 + df}$$

t Test

Male-Female

Assuming equal variances

Difference	-4.792	t Ratio	-1.19006
Std Err Dif	4.026	DF	46
Upper CL Dif	3.313	Prob > |t|	0.2401
Lower CL Dif	-12.896	Prob > t	0.8799
Confidence	0.95	Prob < t	0.1201



Here: r^2 = (-1.19^2) / (1.19^2 + 46) = .0298, hence

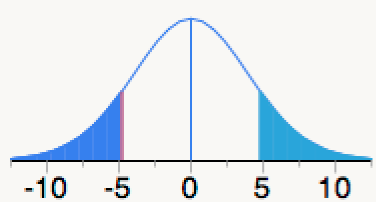
t(46) = -1.19, *p* = .24, ***r* = .17**

t Test

Male-Female

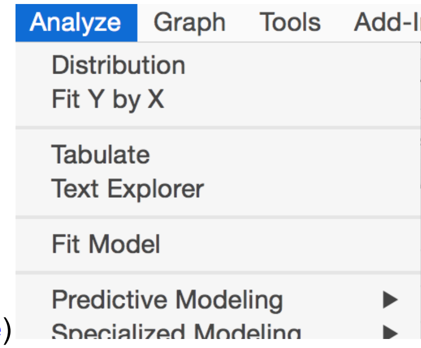
Assuming unequal variances

Difference	-4.792	t Ratio	-1.19006
Std Err Dif	4.026	DF	28.48371
Upper CL Dif	3.450	Prob > |t|	0.2438
Lower CL Dif	-13.033	Prob > t	0.8781
Confidence	0.95	Prob < t	0.1219

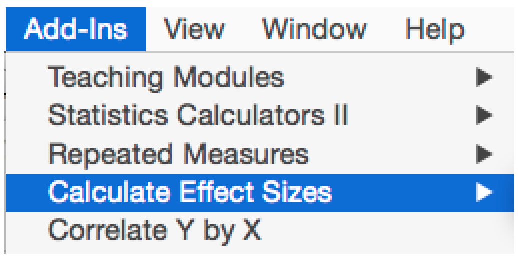


t(28.48) = -1.19, *p* = .24, ***r* = .22**

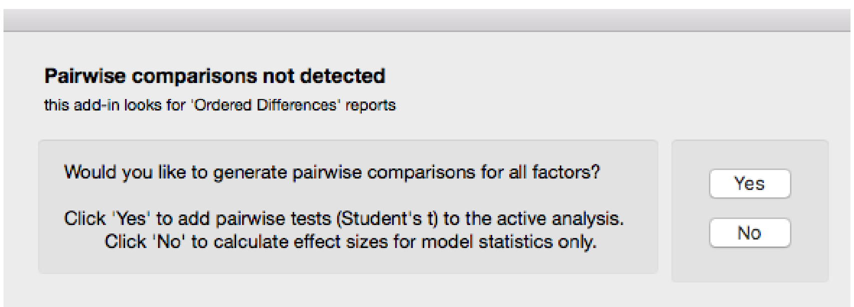
Or Calculate r with JMP



First: use Analyze > Fit Model to run ANOVA first (example)
Second: Add-ins > Calculate Effect Sizes > From Least Squares Report (Fit Model)



Third: you might see this window; click Yes



This should give you all the effect sizes needed in a separate table in JMP, for both t-test and ANOVAs

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F	Eta Squared	Partial Eta Squared	Omega Squared
Gender	1	1	3136.0000	8.8161	0.0102	0.38640	0.38640	0.32818

Formulas used:

Eta Squared

$$\eta^2 = \frac{SS_{effect}}{SS_{total}}$$

Partial Eta Squared

$$\eta^2_{partial} = \frac{SS_{effect}}{SS_{effect} + SS_{error}}$$

Omega Squared

$$\omega^2 = \frac{SS_{\text{treatment}} - df_{\text{treatment}} * MS_{\text{error}}}{SS_{\text{total}} + MS_{\text{error}}}.$$

Pairwise Effect Size: ***Cohen's d***

$$d = \frac{M_1 - M_2}{SD_{pooled}}$$

Repeated measures (within-subject) t-test

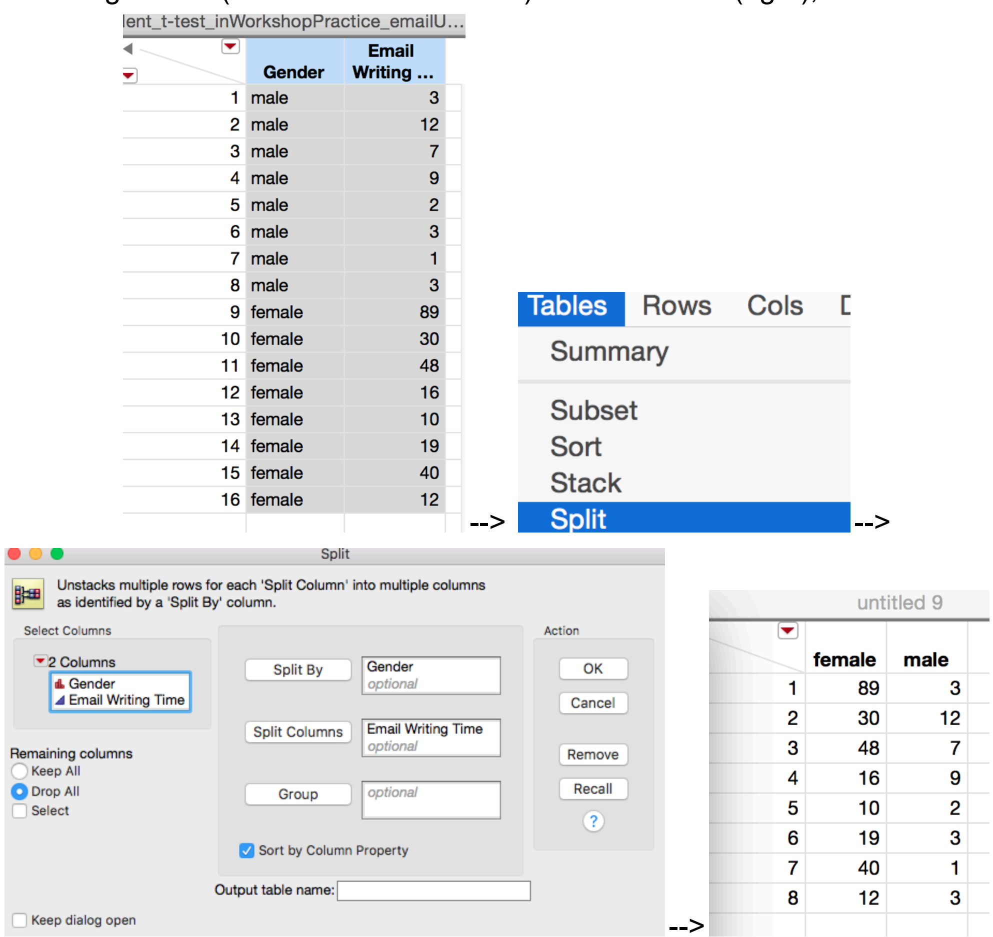
Calculate r using JMP

based on r := sqrt (t^2) / (sqrt (t^2 + df))

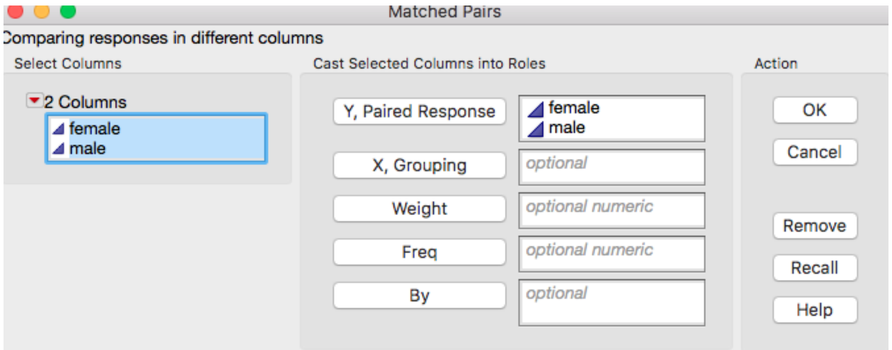
$$r^2 = \frac{t^2}{t^2 + df}$$

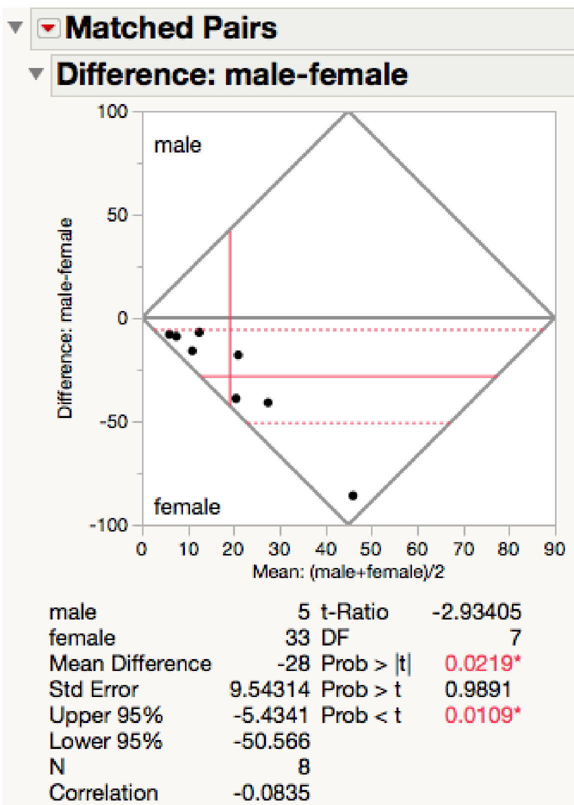
First: convert your data to **wide format**

- to convert long format (left table in screenshot) to wide format (right), use Tables > Split



Second: run Analyze > Specialized Modeling > Matched Pairs:





The provides the ***t*** and ***df*** needed to calculate ***r^2***:

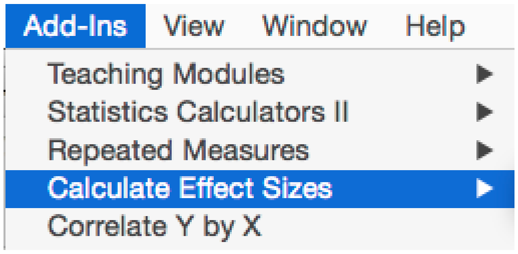
$$r^2 = \frac{t^2}{t^2 + df}$$

t(7) = -2.93, *p* = .02, ***r* = .74**

Independent measures (between-subject) >= 1-Way ANOVA

Calculate effect sizes with Add-in

1. Use Analyze > Fit Model to run ANOVA (an ordinary least squares model)
2. Use Add-Ins > Calculate Effect Size > From Least Squares Report (Fit Model)



Model Effect Sizes

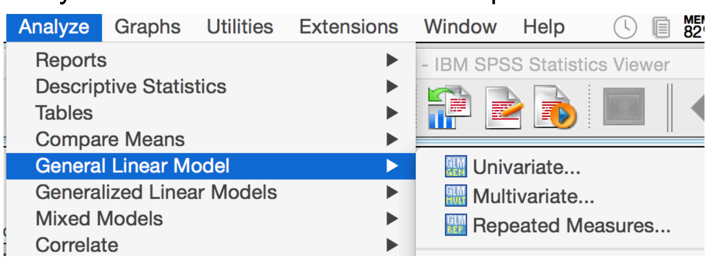
	Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F	Eta Squared	Partial Eta Squared	Omega Squared
1	Sex	1	1	275.5208	3.2116	0.0803	0.02987	0.07104	0.02038
2	Alcohol level [pints]	2	2	3463.5417	20.1865	<.0001	0.37547	0.49012	0.35358
3	Sex*Alcohol level [pints]	2	2	1882.2917	10.9705	0.0001	0.20405	0.34314	0.18374

Repeated measures (within-subject) 1-Way ANOVA

We don't even have the SS in the JMP output to calculate the effect size…
→ go to **SPSS** or another software :-(
Laerd Statistics: one-way repeated measures ANOVA with SPSS

In SPSS

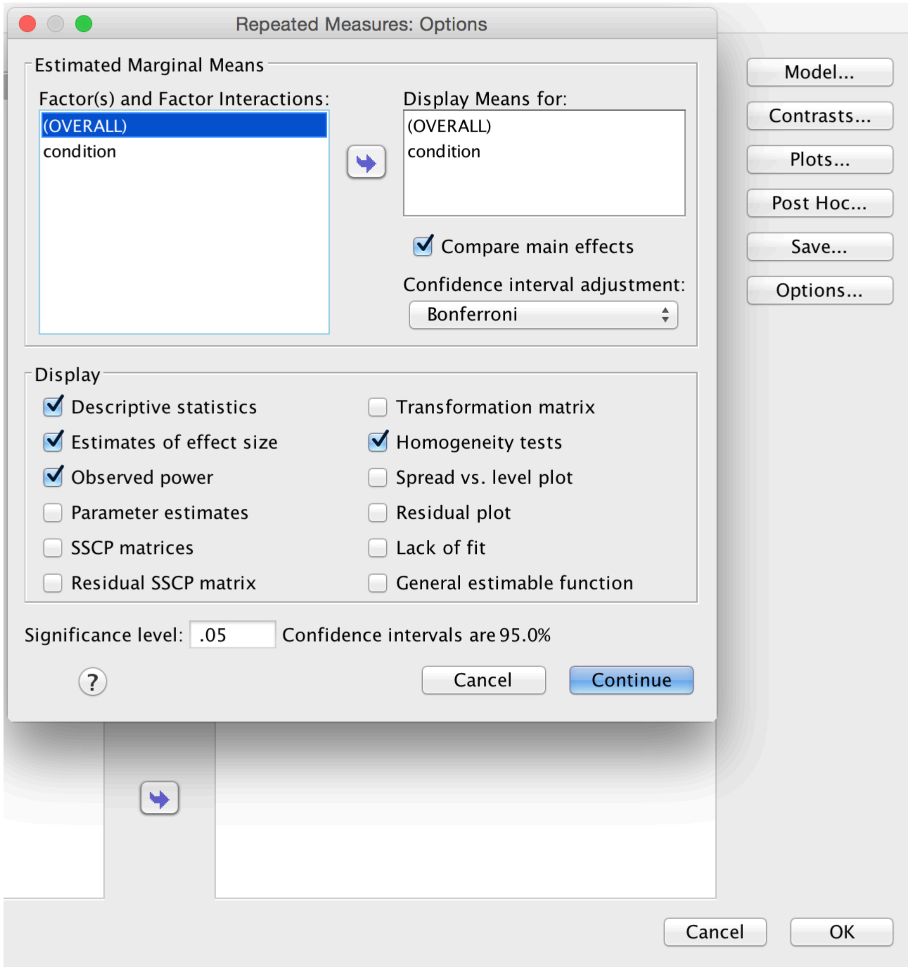
Analyze > General Linear Model > Repeated Measures



Define your levels and factors

Assign them (make sure they match)

Under "options" check the power, effect size etc. options if you need them:



Check sphericity:

Mauchly's Test of Sphericity[a]

Within Subjects Effect	Measure	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Epsilon[b] Huynh-Feldt	Epsilon[b] Lower-bound
condition	vectionOnsetLatency	.928	.747	2	.688	.933	1.000	.500

Effect size partial eta squared and power are included in table below
Use G-F or H-F if sphericity is violated

Univariate Tests

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
condition	vectionOnsetLatency	Sphericity Assumed	1902.565	2	951.283	31.021	.000	.738	62.042	1.000
		Greenhouse-Geisser	1902.565	1.866	1019.713	31.021	.000	.738	57.878	1.000
		Huynh-Feldt	1902.565	2.000	951.283	31.021	.000	.738	62.042	1.000
		Lower-bound	1902.565	1.000	1902.565	31.021	.000	.738	31.021	.999

Sample write up

Vection intensity ratings were higher for the leaning interface (*M* = 64.6, *SD* = 21.9) compared to the joystick (*M* = 49.7, *SD* = 26.4, *F*(1,15) = 10.406, *p* = .006, η^2 = .410), see also Figure 2. The effect size of η^2 = .410 indicates that 41% of the variability can be attributed to the factor interface, which is considered a large effect size (Cohen, 1988).

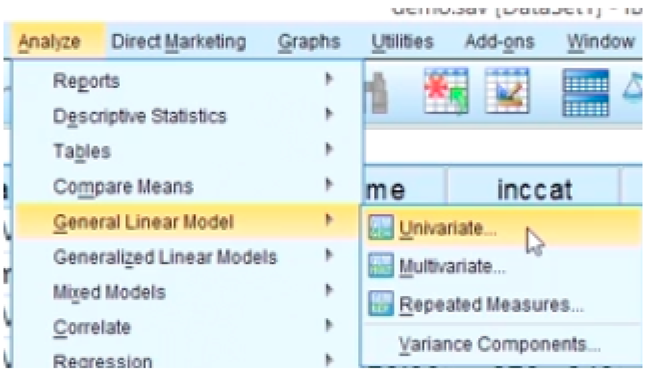
>= 2-Way Independent measures (between-subject)

In JMP: see above

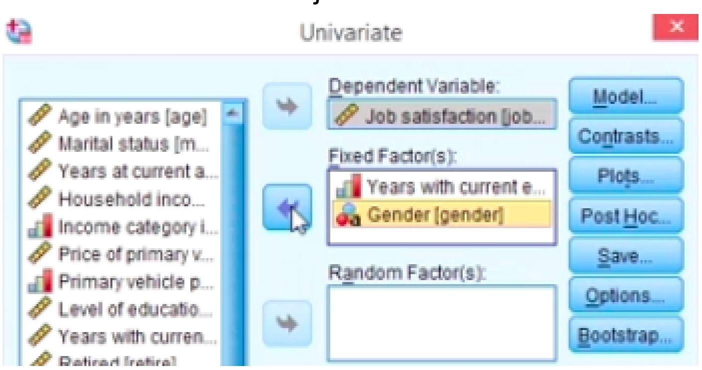
In SPSS

Laerd Statistics: [two-way repeated measures ANOVA with SPSS]
Lynda Tutorials: [SPSS two categorical variables ANOVA]

- Analyze > General Linear Model > Univariate (as we have only one DV)



- Put between-subject factors under "Fixed factors"



>= 2-Way Repeated measures (within-subject) or mixed ANOVA

In SPSS

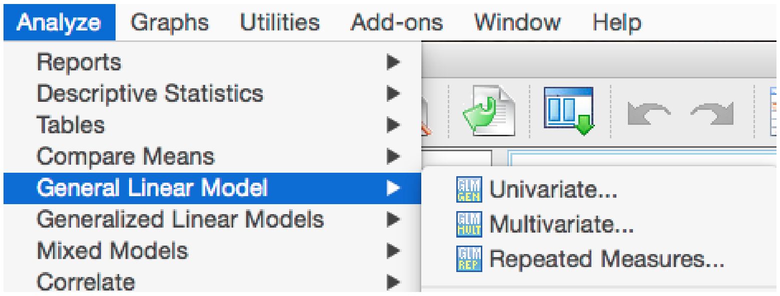
Laerd Statistics: [mixed ANOVA with SPSS]
Or just google for instructions, there's plenty of tutorials out there.

- use JMP and Tables > Split to create "wide format" table
- save as .csv or .xls file that you can import into SPSS (should be installed on the SFU lab computers)
- follow instructions in below videos. E.g., on 2-way repeated-measures ANOVA in SPSS:

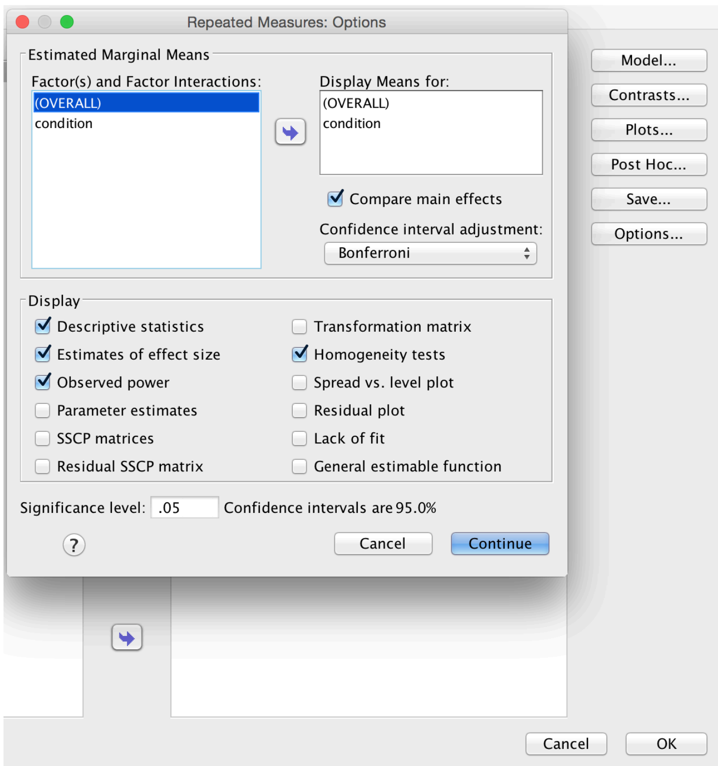
Part 1: https://www.youtube.com/watch?v=qobtGqVBFig

Part 2: https://www.youtube.com/watch?v=zBIqSRC3-Z8
Part 3: https://www.youtube.com/watch?v=VzjAM0jEDtg

- SPSS > Analyze > General Linear Model > **Repeated Measures**



Under "options" check the power, effect size etc. options if you need them:



Post-hoc Tests

Calculate effect sizes in JMP

First of all, you don't have to report effect sizes for post-hoc tests or planned contrasts.
It can still be interesting and useful though, for example if you're interested in how much variability is explained when you compare two factor levels of one IV:

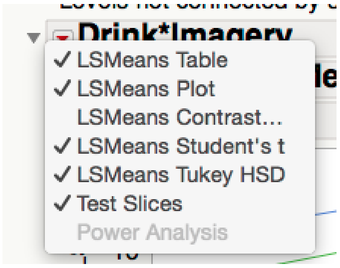
First: JMP > Analyze > Fit model

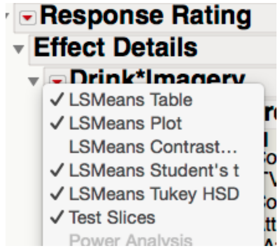
Second: Then Effect Details > Test Slices

Alternatively, if you'd like to use the t-test results:

Second: Effect Details > LS Means Student's t

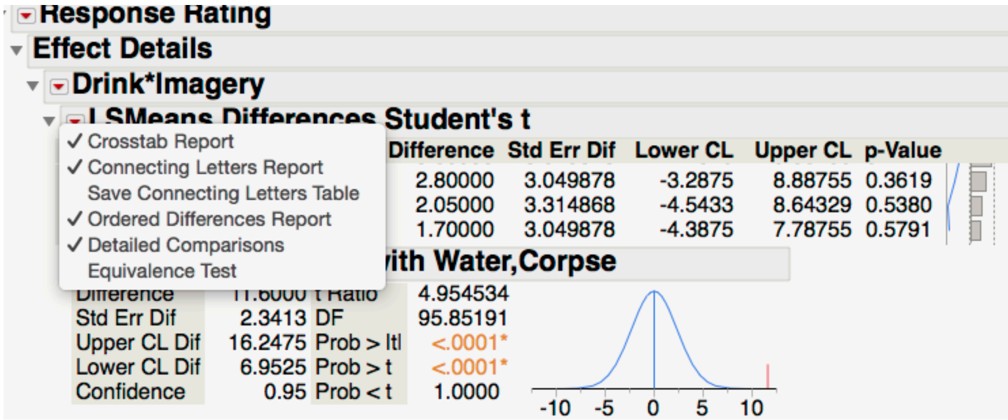
**just be sure to correct for Alpha inflation using, e.g, Bonferroni correction*





Third: to see the actual ***t*** values to compute ***r*** or ***r^2***, use the **Detailed Comparison** option:

$$r^2 = \frac{t^2}{t^2 + df}$$



Want more infos?

E.g., https://www.simplypsychology.org/statistics.html has useful and easy to understand summaries

Overview of effect sizes,

From https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize

Effect Size	Use	Small	Medium	Large
Correlation inc Phi		0.1	0.3	0.5
Cramer's V	r x c frequency tables	0.1 (Min(r-1,c-1)=1), 0.07 (Min(r-1,c-1)=2	0.3 (Min(r-1,c-1)=1), 0.21 (Min(r-1,c-1)=2)	0.5 (Min(r-1,c-1)=1), 0.35(Min(r-1,c-1)

		), 0.06 (Min(r-1,c-1)=3)	, 0.17 (Min(r-1,c-1)=3)	=2), 0.29 (Min(r-1,c-1)=3)
Difference in arcsines	Comparing two proportions	0.2	0.5	0.8
η^2	Anova	0.01	0.06	0.14
omega-squared	Anova; See Field (2013)	0.01	0.06	0.14
Multivariate eta-squared	one-way MANOVA	0.01	0.06	0.14
Cohen's f	one-way an(c)ova (regression)	0.10	0.25	0.40
η^2	Multiple regression	0.02	0.13	0.26
κ^2	Mediation analysis	0.01	0.09	0.25
Cohen's f	Multiple Regression	0.14	0.39	0.59
Cohen's d	t-tests	0.2	0.5	0.8
Cohen's ω	chi-square	0.1	0.3	0.5
Odds Ratios	2 by 2 tables	1.5	3.5	9.0
Odds Ratios	p vs 0.5	0.55	0.65	0.75
Average Spearman rho	Friedman test	0.1	0.3	0.5

Also:Haddock et al (1998) state that 3π multiplied by the log of the odds ratio is a standardised difference equivalent t