Statistics

Null hypothesis: there are no differences between groups. By convention, we use .05 as the cut-off for rejecting the null hypothesis. This means there is less than a 1 in 20 possibility that this finding occurred by chance.

Sample size:

Type 1 errors (false “positive”): reject the null hypothesis when should not, i.e., state there is a significant difference between groups when there is none. May occur when doing multiple tests (by chance) or when sample sizes are so large almost any difference becomes significant (not technically type 1 error).

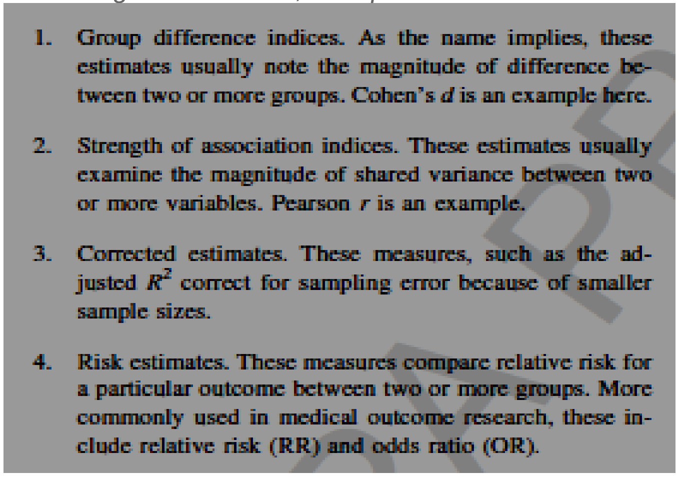
To correct for multiple comparisons we use adjustment such as the Bonferroni Correction which is: .05/number of comparisons , which then gives the adjusted p value. Thus, for 10 comparisons it would be .05/10= .005

Type 2 errors (false “negative”): accept the null hypothesis when should not, i.e., state there is no difference between groups when there is a difference. Usually occurs when sample size is too small to detect differences. To avoid these errors, by convention we use .80(called beta). That is, there is an 80 out of 100 chance that the sample size is sufficient to detect an effect. This is called “power” or the ability to detect differences if they exist.

Expected difference (P1-P2)	Total sample size required *
5%	1450-3200
10%	440-820
20%	140-210
30%	80-100
40%	50-60

**5% significance level, 80% power.*

Effect sizes



1. Group difference indices. As the name implies, these estimates usually note the magnitude of difference between two or more groups. Cohen’s *d* is an example here.
2. Strength of association indices. These estimates usually examine the magnitude of shared variance between two or more variables. Pearson *r* is an example.
3. Corrected estimates. These measures, such as the adjusted R^2 correct for sampling error because of smaller sample sizes.
4. Risk estimates. These measures compare relative risk for a particular outcome between two or more groups. More commonly used in medical outcome research, these include relative risk (RR) and odds ratio (OR).

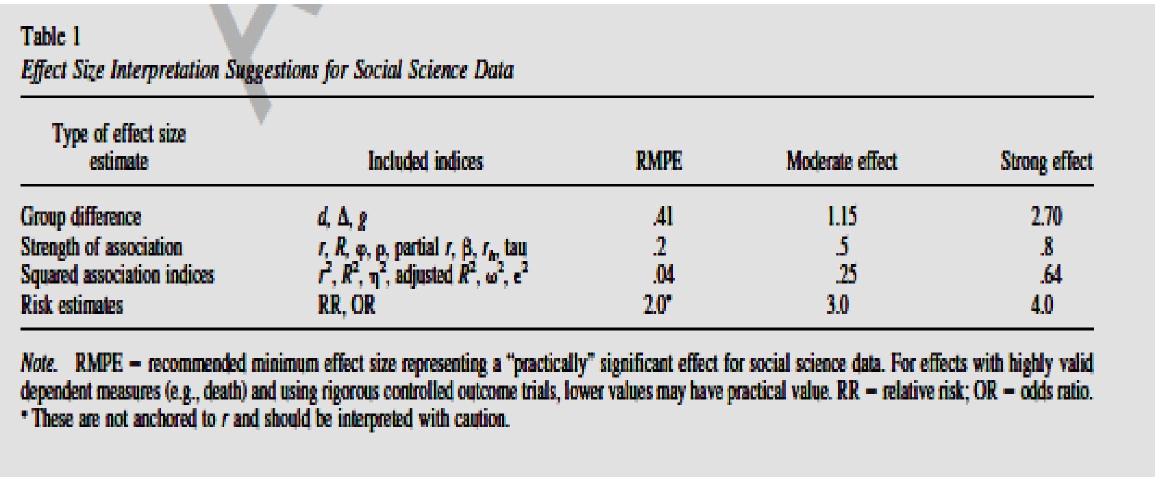


Table 1
Effect Size Interpretation Suggestions for Social Science Data

Type of effect size estimate	Included indices	RMPE	Moderate effect	Strong effect
Group difference	*d*, Δ, *g*	.41	1.15	2.70
Strength of association	*r*, *R*, φ, ρ, partial *r*, β, r_h, tau	.2	.5	.8
Squared association indices	r^2, R^2, η^2, adjusted R^2, ω^2, ϵ^2	.04	.25	.64
Risk estimates	RR, OR	2.0[a]	3.0	4.0

Note. RMPE = recommended minimum effect size representing a "practically" significant effect for social science data. For effects with highly valid dependent measures (e.g., death) and using rigorous controlled outcome trials, lower values may have practical value. RR = relative risk; OR = odds ratio.
[a] These are not anchored to *r* and should be interpreted with caution.

p-values not the same as effect sizes: although there may be significant differences between groups the impact of the intervention or variable may be minimal.
Formula= differences between two means divided by the pooled (combined) standard deviation

Effect size: 0.2=small; 0.5 =medium ; 0.8=large

Note: Effect size is for group differences.

For individual differences: "Reliable Change Index" which examines an individual patient's change (difference between 2 scores) is meaningful, e.g , 50% reduction on the HAM-D

Sensitivity and Specificity

Definitions:
Sensitivity: + / all + ex. 10/22
Specificity: - / all – ex. 11/26
Positive predictive value: percent correct + /+ and – identified ex 10/25
False +: called + when really -
False -: called – when really positive
Increase sensitivity, specificity will decline.
In rare events, specificity is always very high

	Dementia	**Non-dementia**	
MSSE ≤ 23(+)	10	15 (false positive)	25
MSSE ≥24(-)	12(false negative)	11	23
	22	26	48

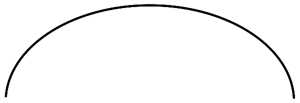
Number Needed to Treat (NNT)=
100 / difference in absolute percentages (round off to higher number, e.g. 3.2 to 4.)

(50% response to rx; 25% placebo response= 100/25 = 4)
Lower is better, and less than 10 is considered desirable.
Number Needed to Harm is the same formula except higher is better.

Frequency of an occurrence:
Note how the implications of preventative interventions are different if the condition is rare versus frequent, e.g. 50% prevention rate versus placebo in a condition that occurs in 1 in 100 persons is different for a 50% prevention rate in a condition that occurs in 1 in 10 persons.

Example: 10 persons (4 no RX, 2 with RX) = 5NNT
100 persons (2 no Rx, 1 with Rx)=100 NNT
Need to weigh benefits versus harm

Normal curve



Occurs by chance, e.g. coin toss $(H + T)^{x}$ creates a normal(bell-shaped) curve.

Deviations around the mean: each score(x) -mean X

Because this will add up to zero, we calculate sum by ignoring the signs. This is done by squaring the term, dividing by the number of scores, and then calculating its square root : sum of $(x-X)^2$. This term is called the standard deviation.

The term before the square root is taken is called the variance. However, it is difficult for most persons to conceptualize in terms of variance although it is used in statistical procedures.

The standard error is the std dev of a population, based on the notion that samples drawn from a large population tend to be normally distributed.

95% confidence interval: 95% chance the true value lies within this interval. Thus, in selecting a sample, since it is impossible to find a population’s true mean from a sample drawn from it, we want to know if we are 95% certain that the true mean lies within this interval.

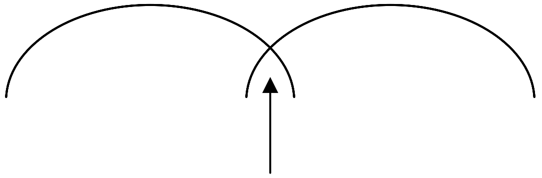
Statistical Tests

Types of variables to be analyzed:
Continuous such as scale values
Ordinal such as ordered categories (e.g, small =1, medium=2, large =3)
Categorical such as gender(male , female) or racial categories

Statistical tests for continuous variables are called “parametric”
Statistical tests for non-continuous variables are called “non-parametric”

Parametric Tests

t-tests: for continuous variables and two samples such as comparing the height(continuous) of Russians versus Spanish(2 groups)

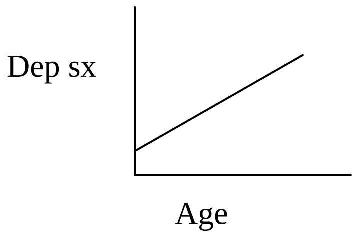


2.5% overlap(2-tailed test) or 5% overlap (1-tailed)

Analysis of Variance (ANOVA): compares more than 2 groups on continuous variables such as the heights of three nationalities or 2 x 2 tests such as the impact of two treatments on depression (i.e., each one, two together, and none)

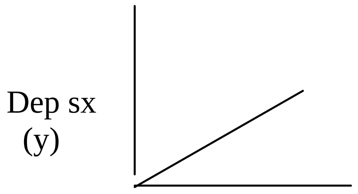
	Psychotherapy	**No psychotherapy**	***Tests drug/no drug***
Drug A	Depression score	Depression score	
No drug A	Depression score	Depression score	
Tests therapy vs non-therapy			***Also can test interactions***

Correlations: relationship between two continuous variables. Scores range from –1 to +1. The more positive the score the more two variables are positively associated (e.g., increasing age and vital capacity) and the more negative the score the more variables are inversely related (e.g., age and score on cognitive testing). Scores closer to 0 indicate no associations in either direction. Partial correlations are associations between two variables after one or more other variables are controlled for (taken into account). Ex. A partial correlation between race and IQ, controlling for socioeconomic class.



Regression Analysis
How well several variables predict an outcome variable.

Note: outcome variables are often called “dependent” variables and predictor variables are called “independent” variables. Variables that are controlled for are called “covariates.”



Age(x_1) + physical health(x_2) + smoking (x_3)

Non-Parametric Tests

Chi-square: measures differences between observed and expected frequencies of two or more variables (that are categorical). For example, we might expect that the number of males and females born in a given population should be roughly 50:50. However, we observe that there are 60 % male and 40% female. Is this a meaningful difference?

Other important tests
Data reduction: Factor analysis, which determines how well certain items cluster together

Survival Analysis

Proportional Hazards: ratio of risks of an event (e.g., death) at any particular time—compares two groups

e.g., Cox Regression

Kaplan-Meier Survival curves

Longitudinal Analyses

Path Analysis

Structural Equation Models e.g., Lisrel

Corresponding Parametric and Non-Parametric Tests

Table 1

	Parametric Test	**Non –Parametric Tests**	
	Continuous Data	**Nominal Data**	**Ordinal Data**
Two-Group Case	t-test	Chi-square	Mann-Whitney U
k-group(>2) Case	One-way ANOVA	Chi-square	Kruskal-Wallis H
Dependent Groups-2 points in time	Paired t-tests	Mc Nemar test for significance of change	Wilcoxan matched pair sign rank test(analogous to paired t-test) measures)
Dependent Groups: Repeated Measures	Repeated measures ANOVA; Linear Mixed Models		Friedman matched samples(analogous to repeated
Correlations	Pearson's R	Phi (with chi square)	Spearman's Rho or Kendall's tau
Multivariable	Linear regression	Binary logistic regression(2 groups); Nominal regression(3+ groups)	Ordinal regression