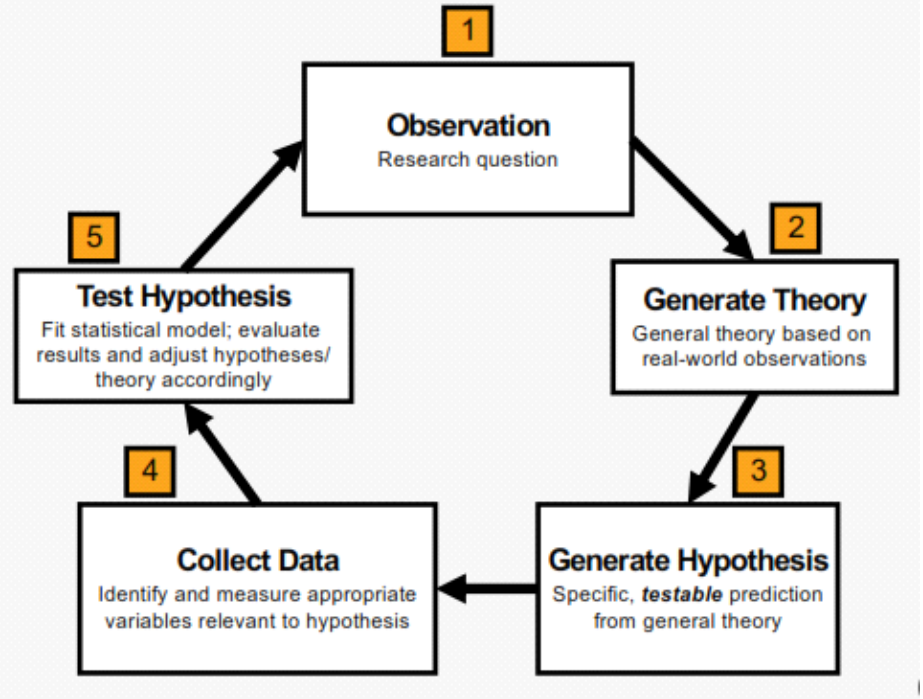
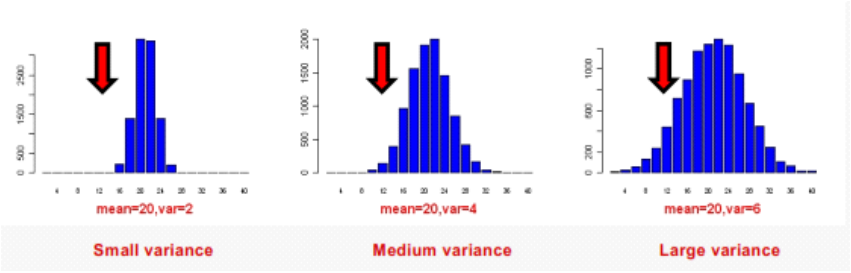
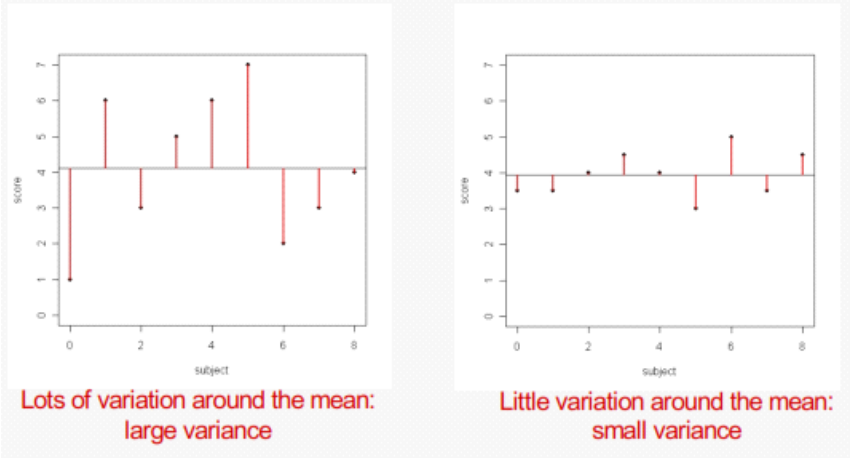
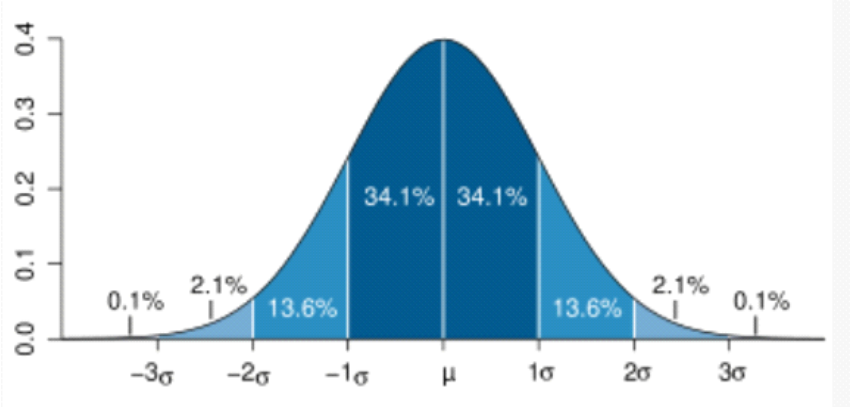
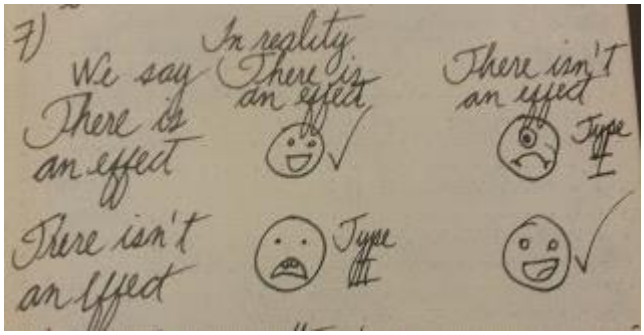
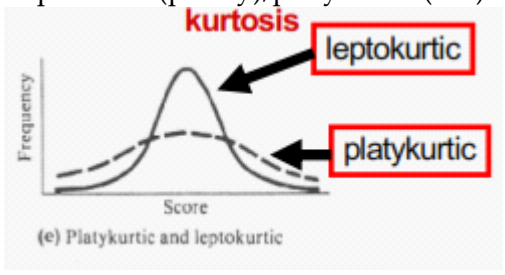
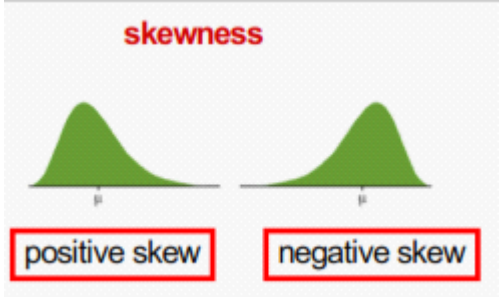
1 - Basic statistical concepts and data exploration

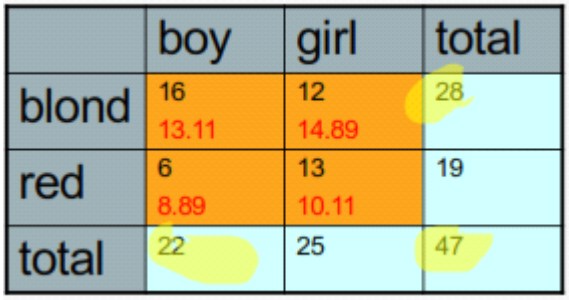
Describe the five steps of the research cycle.	**Empirical research cycle** 1) Observation - Start with a question An observation is not enough to conclude anything - Biases intrinsic to humans, no control group The question defines the scope of the investigation 2) Generate theory - Makes predictions about future observations Serves as a framework for your statiscal model 3) Generate hypothesis Specific and falsifiable 4) Collect data Increase or decrease confidence in the theory No single data set can confirm or disprove a theory 4.A - Explore your data before analysing it (data entry is prone to error) Graphs and frequency distributions (histograms) 5) Test hypothesis - Is there a significant association between two or more variables  Example:

	Research question: Do boys and girls differ with respect to hair color? Theory: If hair color genes are on the X chromosome, sex differences would be expected Hypothesis: (2-sided) H_0: Boys and girls *do not differ* in the proportions with each hair color H_A: Boys and girls *differ* in the proportions with each hair color
What are the two types of study design? What is their main difference?	**Types of study design** **Non-experimental**: Observational, non-invasive, cannot infer causation **Experimental**: Manipulation of variables, can infer causation
What is a dataset?	**Variable**: Things that vary in a scientific investigation **Dataset**: Multiple variables
What are the two variable types?	**Types of variables:** **Independent** - Predictor (cause) **Dependent** - Outcome (effect)
What are the types of qualitative data? **What are the types of quantitative data?**	**Measure data** **Qualitative**: Categorical (nominal or ordinal) **Quantitative**: Continous (ratio or interval) **Interval** - You can add and subtract, but not multiply. There is no absolute zero **Ratio** - There is an absolute zero All data could be measured in a discrete or continuous way Numbers in categorical data are arbitrary Measurement errors should always be considered in the data interpretation
What does sum of squares, variance and standard deviation describe? **What is range?**	Descriptive statistics **Central tendency** Mode (nominal) - Highest frequency Median (ordinal) - Middle value after ranking Mean/avarege **Dispersion** **Variance** - Mean deviation from the mean Needs to be square in order not to be zero **Standard deviation** - Square root of variance N -1 due to degree of freedom (you already know the mean) 

	 **Range** - Distance between lowest and highest values **Percentiles** - Each score relative to a standardized distribution (0-100)
What is the simplest statistical model?	**Statistical models** - Describe your data well (minimize the error) Outcome = model + error Simplest model = mean + standard deviation/error 
What is systematic and unsystematic variation?	**Variance** - Describes the spread of distributions **Systematic** - Due to known/manipulated factors **Unsystematic** - Due to unknown/random factors
How do you know a sample is representative?	**Sampling theory** - If we could collect all data, we would not need inferential statistics Sample estimates represent population parameters - There is an inherent variability inherent in sampling How do you know a sample is representative? Standard error = Variation in the sample means When it is large enough (over 30), sampling distribution is normal SD - Within 1 samples SE - Within multiple samples in the population
What is a test statistic?	**Test statistic** - Ratio of systematic to unsystematic variance Proportions between groups - Chi-squared Linear association

	Comparation of means
What are the parameters of a standard normal distribution	Distribution of the data Parameters of a standard normal distribution 68% within 1 SD 95% within 2 SD 99.7% within 3 SD Mean = 0 SD = 1  Z value = 1.96 for 95% confidence interval
What are degrees of freedom?	Degrees of freedom - The amount of information you have to calculate a test statistic while accounting for the amount of information you are trying to estimate
What is the p-value? **What is alpha?**	Null hypothesis: H0 Alternative hypothesis (usually two-sided): H1 What should the data look like if H0 was true? P-value = If H0 was true, what is the chance that this observation occurred? Alpha = Usually .05 P < alpha = Our result is unlikely under H0 (reject null hypothesis) P > alpha = Our result is likely under H0 (not reject null hypothesis) P = alpha?
What is beta? **Differentiate type I and type II erros**	Alpha = Chance to accept a not real effect as present (detect non-existing effect) Beta = Chance to dismiss real defect as absent (not detect an existing effect) Type I error = alpha Type II error = beta

	
	Effect size - Pearson's r or Cohen's d
	Common assumptions Parametric test - Observations are independent - Data is measured at a interval/ratio level - Normally distributed - Equal variance between groups Non-parametric test - Ordinal data
What is a leptokurtic curve? And a platykurtic?	**Skewness and kurtosis** Leptokurtic(pointy)/platykurtic (flat)  Positive or negative skew 
What are outliers? What is their relation to the standard deviation?	Outliers - Improbably large or small observations 3 SD from the mean
	How to report study results Design and sample characteristics N Mean SD

	Test statistic/p-value Effect size
What is the chi-squared formula? **How do you calculate the degrees of freedom in a chi-squared test?**	**Chi-squared test** - Difference in proportion in nominal variables H0 - There is no difference between different groups Example: Colour hair in boys and girls Expected value = marginal values (total) $\text{expected} = \frac{\text{marg(a)} * \text{marg(b)}}{\text{total}}$ Example - Expected chance of a blonde boy (22 * 28 / 47)  Chi-squared formula $\chi^2 = \sum \frac{(obs - \exp)^2}{\exp}$ Degrees of freedom = (number of rows - 1)*(number of columns - 1) Conditions for the test to be valid - No values = 0 - Few values lower than 5

Sophie Van der Sluis (s.vander.sluis@vu.nl)

Goals: Form opinions independent of others; formulate own research questions

Statiscal techiniques are stable over time

You have to think about statistics before designing the experiment

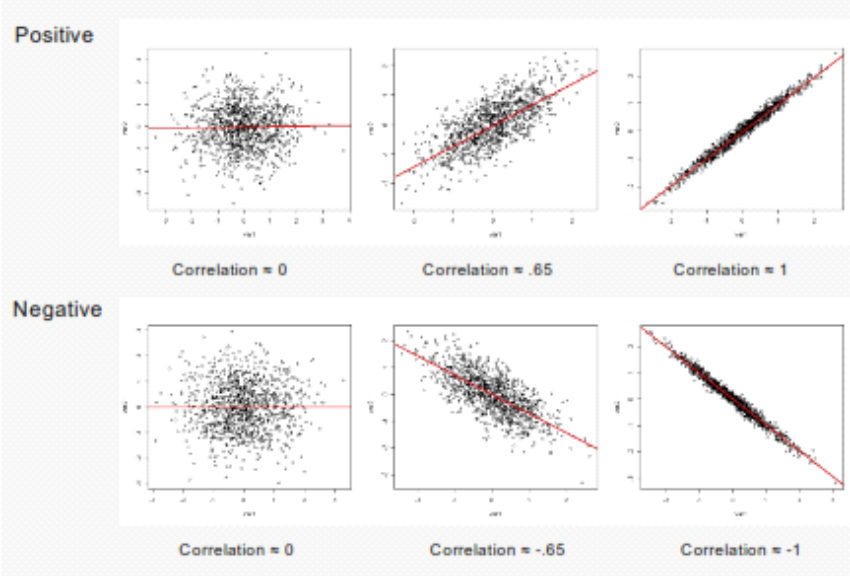
6 Sessions - Lectures, practicals, assignments, self-study
You have to pass all the assignments to be able to do the exam

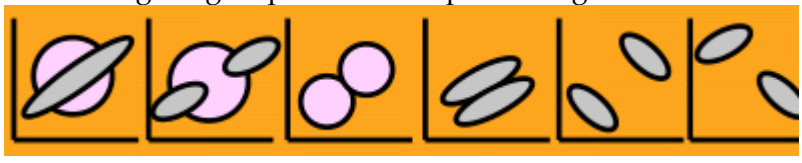
Exam - Multiple choice and open-ended
Presentation - 10% of the grade

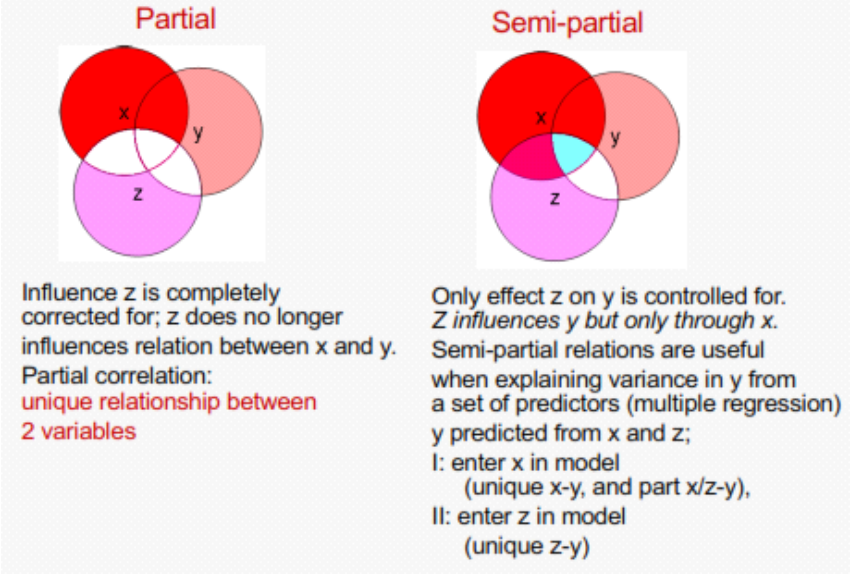
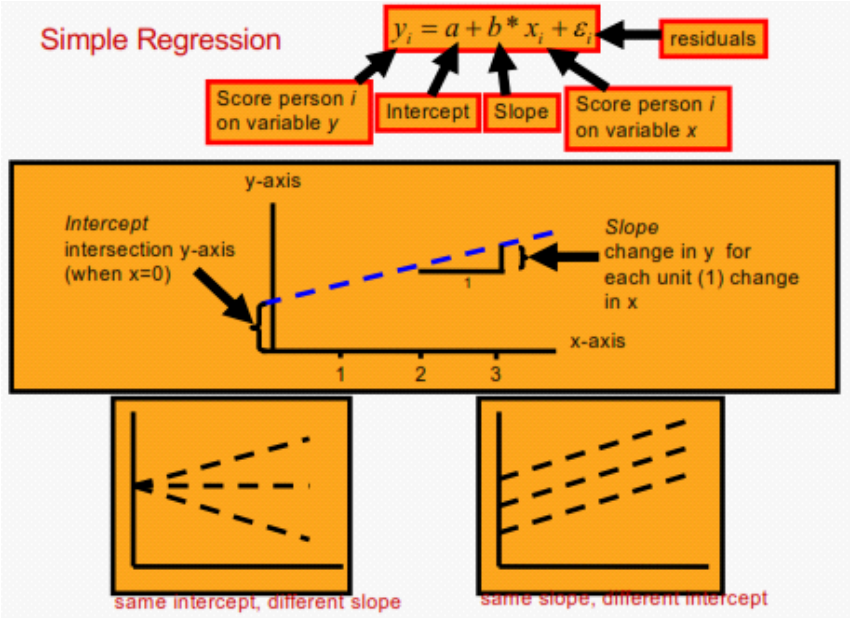
Review week 1
Sampling theory - Inferential statistics

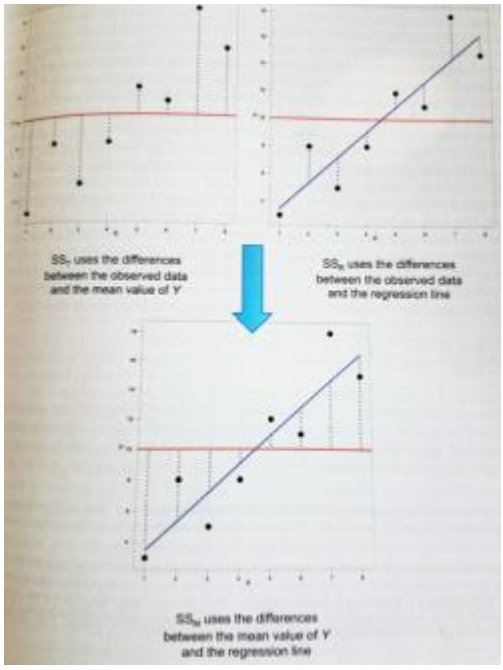
- We cannot collect data about the entire population
- If we get enough samples - The sampling distribution will be a normal curve
    - Sampling mean = Standard error (standard distribution divided by the squared root of the sample)
- Sampling theory applies to other parameters (e.g. difference between populations, linear associations)
- Central limit theorem - If you have a large sample, it is more likely to be representative of the population (it does not mean that the population itself has a normal distribution)
    - Small samples are more likely not to be representative of the population (due to random chance)
- Test statistics - Effect size (ratio of systematic to unsystematic variability)
    - Sampling variability = effect size, express uncertainty (Z, t, F)
- Confidence intervals - Known distributional properties used to express uncertainty around an estimate
    - Z-score = x - mean divided by standard deviation
- Bootstrapping confidence intervals - Construct your own confidence intervals based on your own data
    - 1000 subsamples is the standard
    - Cut off top and bottom 2.5%
    - Remaining highest and lowest values are bootstrapped confidence
- Kolmogorov-S and Shapiro-W - Difference between your distribution and a theorical distribution
    - KS - Any theorical distribution
    - SW - Only normal distributions (more reliable for small and medium samples)
- Chi-squared - Margins take initial differences into account

2 - Correlation and Regression

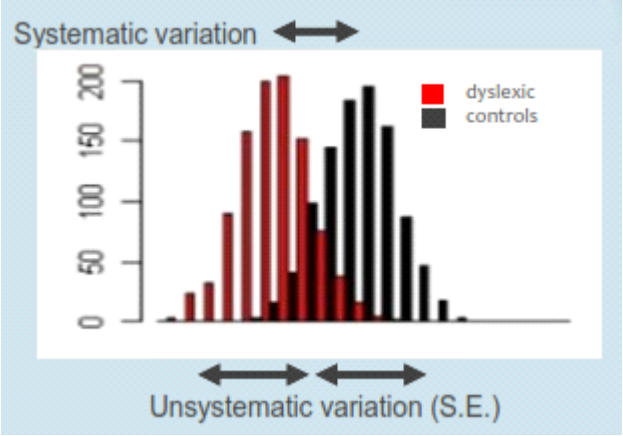
What is the main difference between correlation and regression?	Linear association between two or more variables Correlation - Non-directional Regression - Direction They have more information than a pure means model
Why correlation does not imply causation?	Correlation - Strength and direction of variation between two variables Values between +1 and -1  Does not equal causation - Third cause
How to calculate Pearson`s correlation? When to use, instead of Spearman? **How can you make the correlation result not scale dependent?**	Pearson's correlation rho Standardized covariance Calculation: Same as variance, but multiplying x by y instead of squaring it $Covariance(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1}$ Scale dependent - Cannot compare studies unless they are using the same units of me Solution - Divide this by standard deviation (correlation coefficient, not scale dependent) $z = \frac{x - \bar{x}}{s}$ Values are between +1 and -1
What is the difference between R and R squared?	Correlation significance R is a standardized effect - Not a normal distribution (converted Z or t test) R squared is an indication of how much variability in one

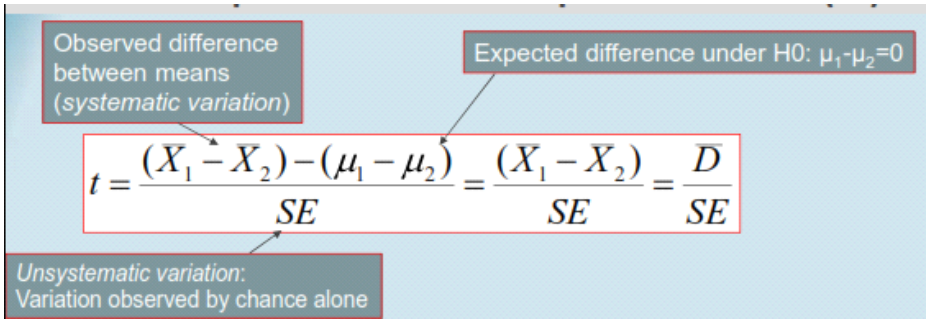
	variable is explained by another
What are the assumptions for Pearson correlation?	Assumptions for Pearson Relations are linear Variables are bivariate normally Homoscedasticity of variances (similar variances between samples)
How to calculate Spearmans' correlation? When to use it over Pearson's?	Spearmans' correlation rho Two ordinal variables or one ordinal and one interval/ratio Does not use scores, but ranks = Scale independent Calculation: Differences between ranks Squared differences between ranks Rank scores *within variables* educ | IQ | Educ_r | IQ_r | diff | diff² 5 | 121 | 5 | 6 | 1 | 1 3 | 104 | 3 | 3 | 0 | 0 2 | 78 | 2 | 1 | 1 | 1 4 | 107 | 4 | 4 | 0 | 0 1 | 85 | 1 | 2 | 1 | 1 6 | 145 | 6 | 7 | 1 | 1 7 | 120 | 7 | 5 | 2 | 4 $\rho = 1 - \frac{6\sum \text{diff}^2}{n(n^2-1)} = 1 - \frac{6*(1+1+0+0+1+1+4)}{7(49-1)} = 1 - .143 = .86$
What is a common correlation hazard?	Correlational hazards Combining subgroups in one sample is dangerous  Significant correlations do not imply causality and direction of causation (tertio quid)
What is the difference between partial and	Partial correlations Correlation corrected for relationship with a third variable

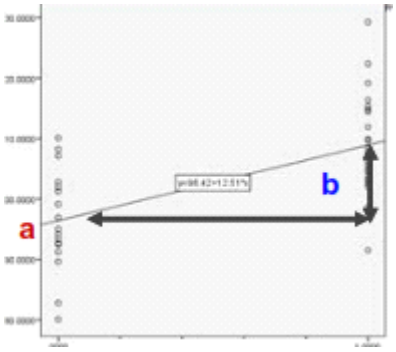
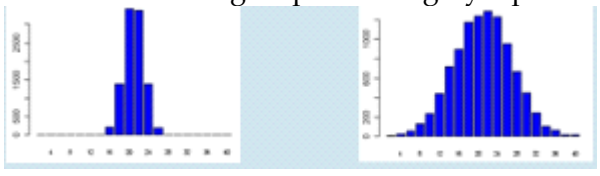
semi-partial correlations?	
	Other correlations Kendall's tau - Small samples and many ties between differences Bi-serail - Dichotomus variable Polychronic - Two dichotomus variables
What can a regression do that a correlation cannot?	Regression Similar to correlation, but directional (one variable predicts another) It is a predictive model
What each variable means in the linear model?	Two bits of information are sufficient to describe a linear model Y = ax + b  Best model minimizes total squared error (least squared estimate)

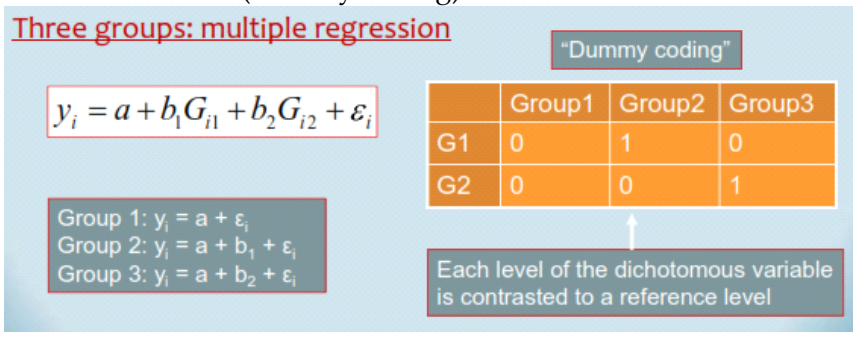
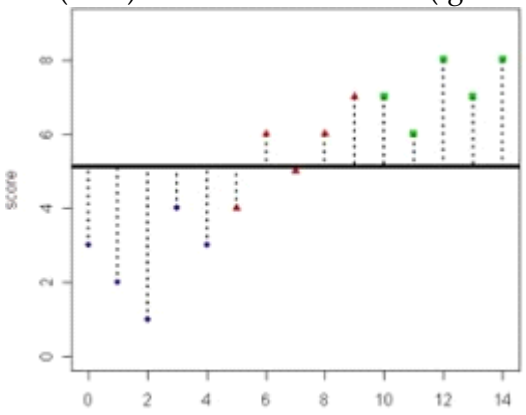
	
Why should we use linear models?	Why use a linear model? Simple model - More parcimonious You can predictive new values of the data
What is the difference between simple and multiple regression?	Simple regression - Two variables (one predictor and one outcome) Multiple regression - More than two variables
Why transforming variables does not affect the interpretation of a regression?	Transforming variables - Does not change the model or the interpretation of the data Insensitive to linear transformations Sensitive to non-linear transformations (e.g log)
Under which circumstances are correlations and regressions the same?	Standardizing both variables = Correlation and regression is the same! B = correlation r A = 0

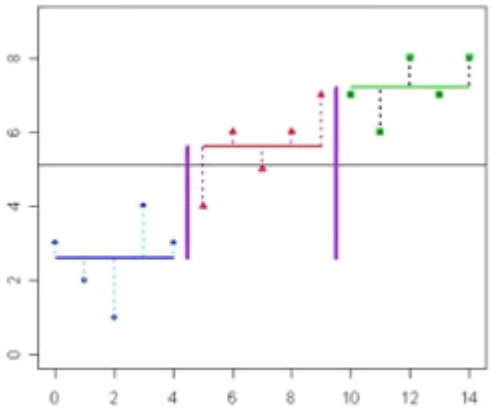
3 - Between subject designs

What is the difference between within subject design and between subject design?	**Within subject design** - Dependent, overlapping samples (same subjects are tested more than once) **Between subject design** - Independent, non-overlapping samples
When should you use a t-test?	**Independent sample t-test** Two groups - Control and experimental groups Assumption: Difference in means can be attributed to the manipulation (causal interpretation) Random assignment
What is the difference between experimental and quasi-experimental design?	**Experimental vs quasi-experimental** **Experiment** - Experimental factor is manipulated by reasearcher, random assignment, control group, equivalent control and experimental group (match) **Quasi-experiment** - Experimental factor not manipulated by researcher, no random assignment, no real control, no match between experimental and control group
What is the question that a t-test tries to answer? **How can systematic and unsystematic variation be observed in a distribution?**	Question of t-test: **Do this data come from the same population or from difference population?** Answer depends on the variance **Systemic vs unsystematic variation**  **Systematic** - Due to the fact that the groups are different **Unsystematic** - The variance between the same group (due to other factors outside from the experimental manipulation)
What the formula for a t-	Formula for t-test

test?	 (difference between observed means) - (difference between expected means)/standard error Systematic variation/unsystematic variation - Like all test statistics Under H0 - the difference between **expected means is 0** Calculation of standard error If samples are equal $SE = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$ If samples are not equal $SE = \sqrt{\frac{s_p^2}{N_1} + \frac{s_p^2}{N_2}}, where$ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$
Why should you report an effect size along with a t-test?	**Effect size** - Quantitative measure of the magnitude of a phenomenon, allows comparisons between different studies Df = N - 2 $r = \sqrt{\frac{t^2}{t^2 + df}}$ R= .1 (small effect) R = .3 (medium effect) R = .5 (large effect) It is always good to calculate the effect size and report it along with the t-test
How can you maximize the probability of observing a significant experimental effect?	How can you maximize the probability of observing a significant experimental effect? - Increase sample size - Decrease variance - Increase effect size/mean difference (the difference between the means between two groups)

	It is important to think about this before collecting data (during the experimental design)
What is the interpretation of a and b in a regression line?	Analogy t-test - regression A = Mean value of Y when group variable is 0 B = Mean value of Y with 1 increase in group variable B is the difference between the two means  Examples Overall mean Means between groups Interpretation of a - The difference between the first group and 0 is significant Interpretation of b - The difference between groups 1 and 2 is significant
What are the assumptions of a t-test	**Assumptions of t-test** Data are normally distributed Data are measured at interval/ratio level Variances in two groups are roughly equal  Probably there was some problem in the experimental design if homoscedasticity is not met Observations are independent - Come from different subjects, not correlated between groups
What a significant Levene test indicates?	Levene test less than .05 - Variances are not equivalent between groups Then you can use equal variances not assumed
	How to report a t-test

	The arithmetic capacity of children who experience reading disability (RD, N=24, M=-.03, SD=.79) was compared to that of chronological age controls (CA, N=24, M=.07, SD=.37). Because the variances of the two groups were significantly different (F(1,46)=10.05, p < .01), a t-test for unequal variances was conducted. This test was not significant (t(32.71)=.54, ns, r = .09, Cohen's d=.26). In other words, no difference in arithmetic capacity between RDs and CAs could be detected.
When should you use an one-way ANOVA?	**One-way ANOVA** Compares **more than two means** When you manipulate one or multiple variables - e.g. Control, KO, double KO One dependent variable = one independent variable with 3 or more levels
Why dummy coding is used in an one-way ANOVA?	Explained in terms of regression Y = a + bG + b2G2 (dummy coding) 
Define SSt, SSm and SSr. **What is the formula of a f-statistic?**	Means for each group **SSt** (total) = Total sum of means (ignores group)  **SSm** (model) = Total sum of mean (taking groups into account) If this is different than SSt, it means that group membership is sufficient to explain variation in the data



SSr (residual) = Residual variation unexplained by SSt and SSm

Fstatistic = SSm/SSr (what is explained by the model divided by what is not explained by the model

Define the three different types of planned contrast in an one-way ANOVA.

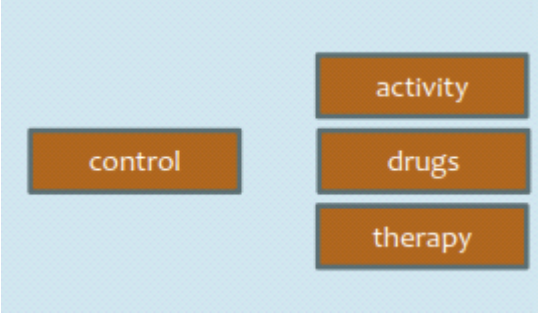
F-ratio more than 1 - More systematic than unsystematic variation

Under H0 - Systematic variation should be zero

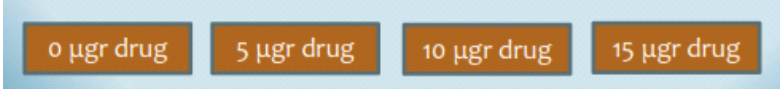
If F-ratio is significant, we know there is a different, but we do not know how (do all groups differ from each other? Only one has a significant difference?)

Planned contrast - Often one-sided

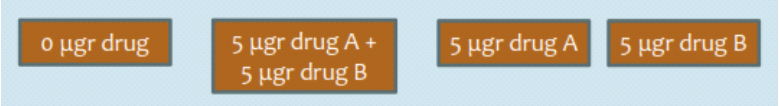
Simple - You compare experimental groups against the control group, not against each other



Repeated - Sequentially compare groups in a logical order (severity manipulation)



Helmert - A+B vs A or B (combination of manipulation)



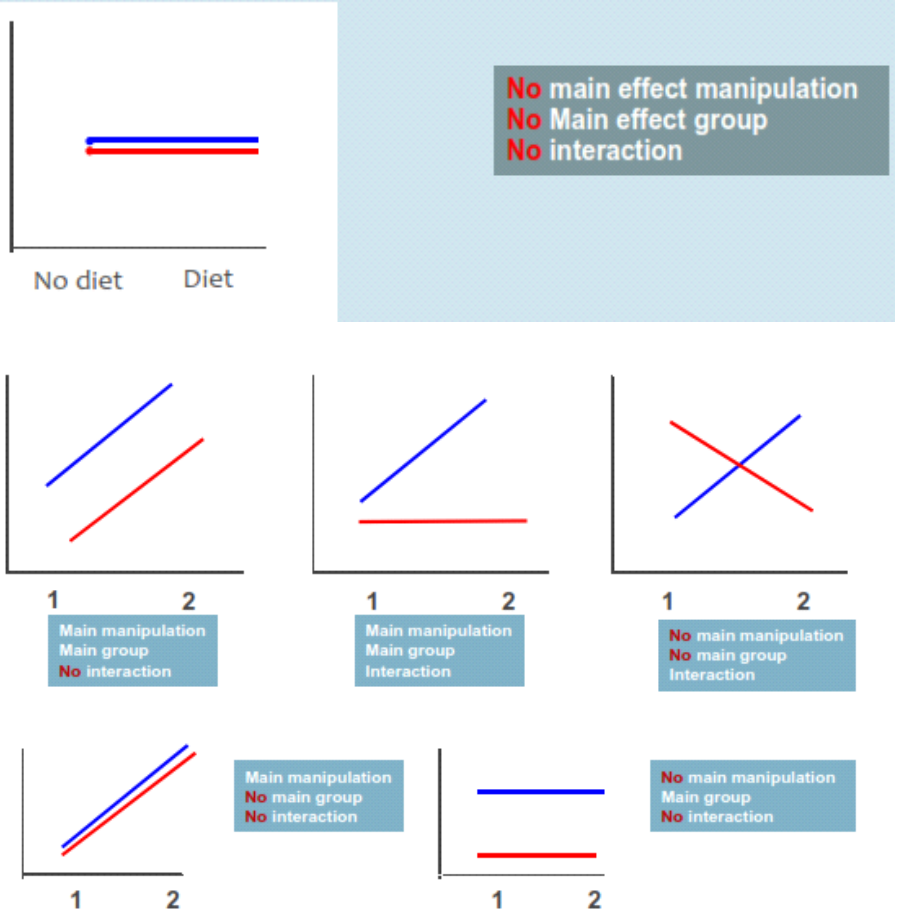
In SPSS - Analyse->GLM

Post-hoc - Often two sided

Why is ANOVA always a two-sided test?

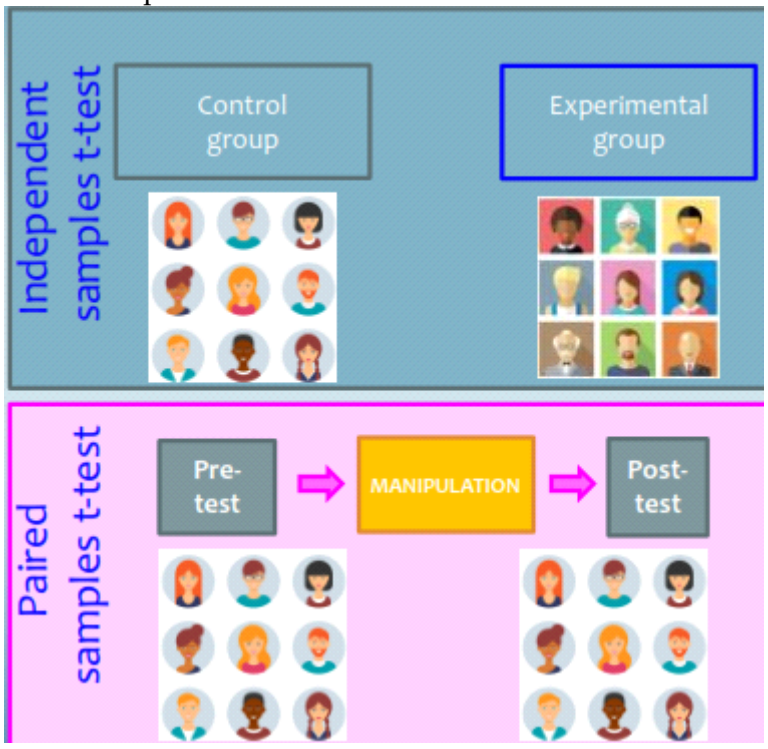
ANOVA: 1 or 2 sided?

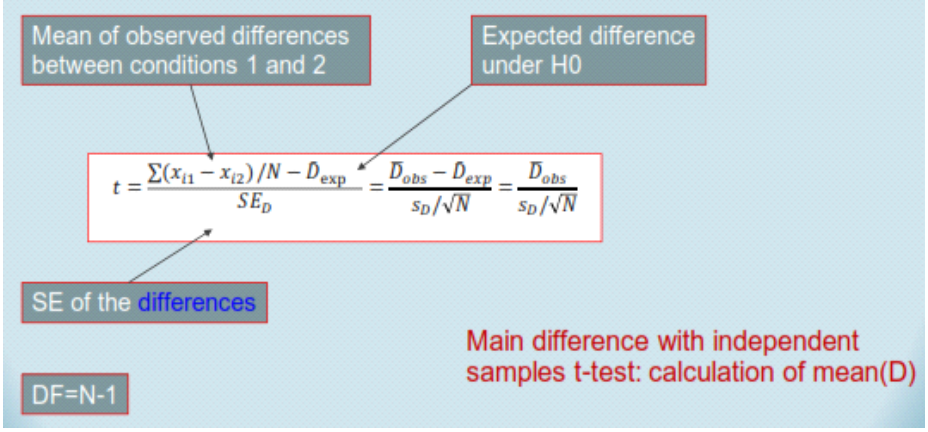
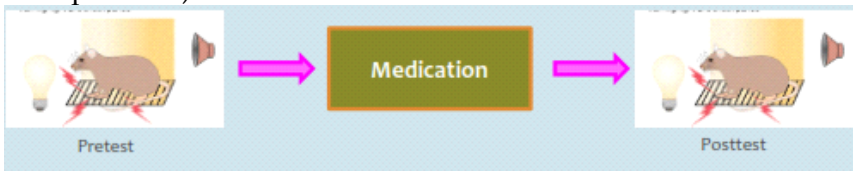
Either all means are equal or they are not equal - ANOVA is always 2-sided

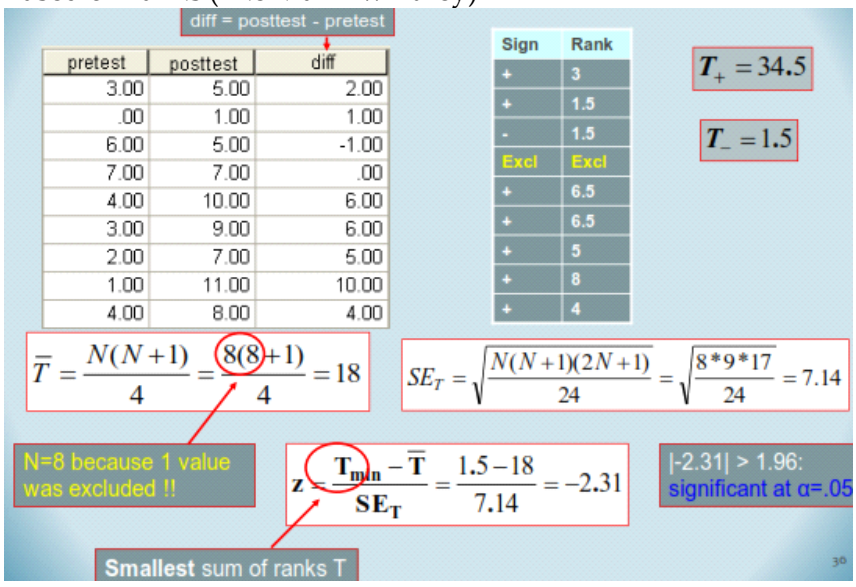
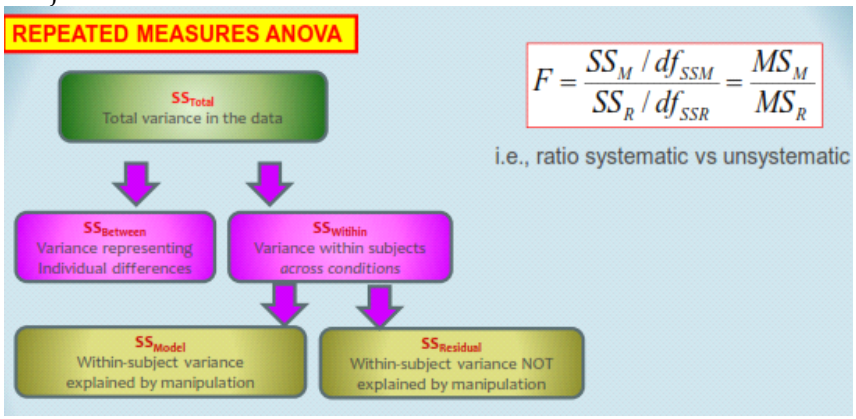
	Post-hoc tests - Can be one-sided Manipulation groups in depression - A manipulation cannot increase depression levels
What are the assumption for an ANOVA	Assumptions for ANOVA Dependent variable is normally distributed Equal variance in each group Independence observations
When should you use a factorial ANOVA?	**Factorial ANOVA (2-way, 3-way)** Multiple means and multiple predictors E.g. Control vs drug; Home vs clinique; male vs female Two-way ANOVA - Diet1/Diet2/no diet; male/female (Image) Interpretation Main effect diet - Across two gender groups Main effect sex- Across all dietary conditions Interaction diet vs sex - The effect of manipulation is the same in all groups?
How can you visually determine if there is an effect of the manipulation, an effect of the group and an effect of the interaction?	Interpretation interaction 
When should you use a non-parametric equivalent of ANOVA?	Parametric vs non-parametric Parametric - Assumptions about distribution of the data, for interval and ratio data Non-parametric - For data with outliers that you do not want to

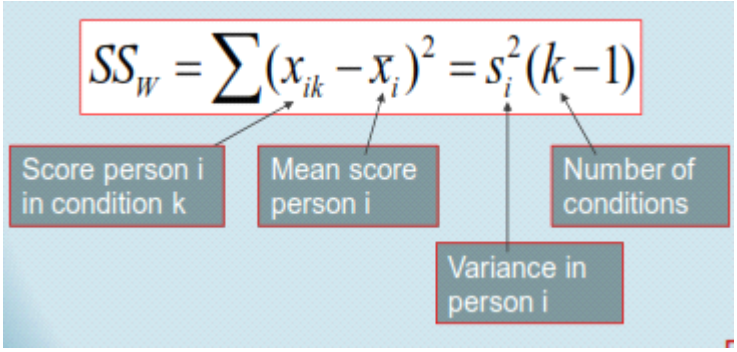
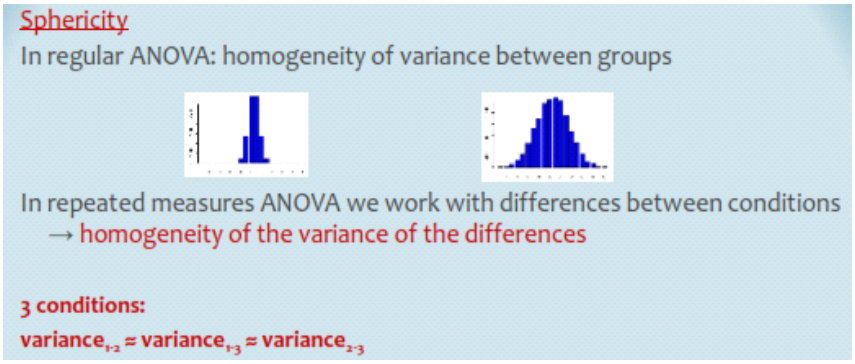
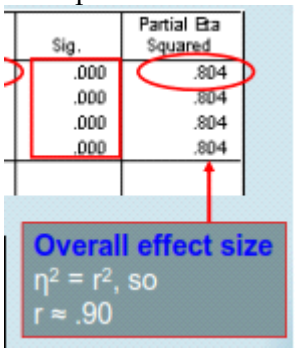
	remove; N is small - If your distribution is normal, some argue that you should use a t-test

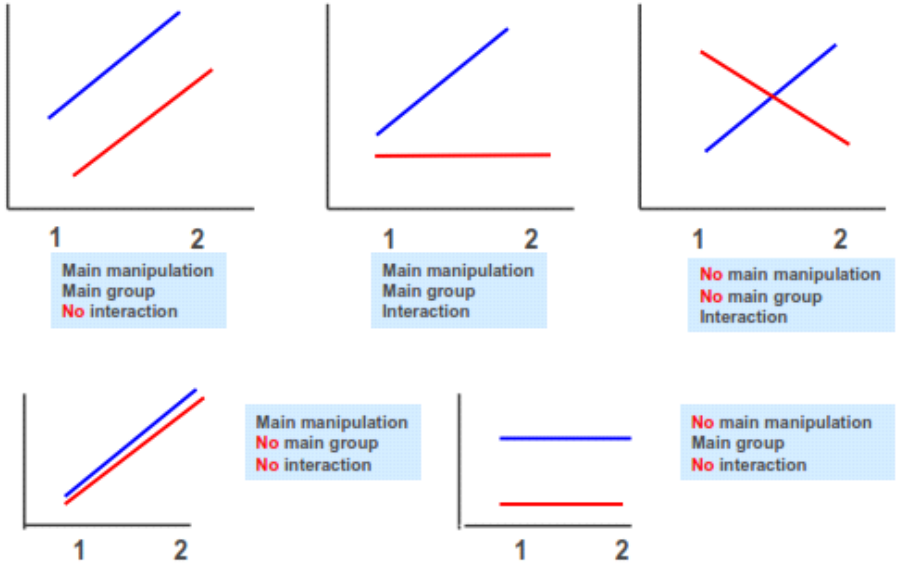
4 - Within subject design

When should you use paired sample t-test over normal t-tests?	Within subject - Multiple measures on the same subject Independent sample t-test Systematic variation - Caused by the manipulation Unsystematic variation - Not caused by the manipulation (observations at different times, genetic variability of mice) Cohen's d = Effect size for mean difference between two groups T-test cannot be used when the samples are not independent
Under which circumstances should paired samples t-test be used?	Paired samples t-test Same people in every condition -> Change is due to a manipulation  *Difference between independent and paired samples t-tests* **Exemples of experimental designs** **Pairs of cases** - Placebo and drug group with match background (age, gender, height, blood pressure, insuline levels) Matched pair parallel design Pairs of cells from the same brain region One has a manipulation, one has not

	Pre-test -> X -> Post-test Training, Intervention, Medication, Passing of time Paired t-test can be used for **different evaluations by different people** Teacher and mother evaluation of student musical performance
	Quasi-experimental designs WT and KO pups from the same litter (womb environment is the same) Pairs of males and females matched on IQ, educational background, ethnici
	Calculation  Mean of observed differences between conditions 1 and 2 divided by the standard error (df = n - 1) Cohen's d for paired samples $d = \frac{\bar{X}_1 - \bar{X}_2}{SD} = \frac{t}{\sqrt{N}}$
What are the main advantages and disadvantages of paired sample t-tests?	**Advantages of paired t-test** Perfect matching - More power to detect effect manipulation Has a smaller unsystematic variation - Variation within the same person changes less It is often one-sided, since you have a clear idea of the type of effect that the manipulation will have (divide alpha levels by 2) **Disadvantages** Depending on design - Carry-over effect (people learn from the manipulation)  - The control group has to be exposed to the pre-test and post-test but without the manipulation!
What are the assumptions of paired	**Assumptions of paired t-test** Data is interval/ratio

sample t-tests?	Differences between scores are normally distributed Observations have no carry-over effect
When should you use Wilcoxon's signed rank test instead of paired t-tests?	Non-parametric variant - **Wilcoxon's signed rank test** Data is ordinal Group is small Differences are not normally distributed
How is the 'difference' defined in Wilcoxon's signed rank test?	Calculation of Wilcoxon's Based on ranks (like Mann-whitney) 
How to report Wilcoxon's results?	How to report You report the median, not the mean! You are comparing ranks, not values Always report a z-score
When should you use repeated measures ANOVA?	Repeated measures ANOVA - **Analysis of within subjects across conditions** More than 2 measures in each subject - Plant growth over multiple days - Global temperature over time
	Calculation Variation within subjects across conditions - Variance within subjects across conditions 

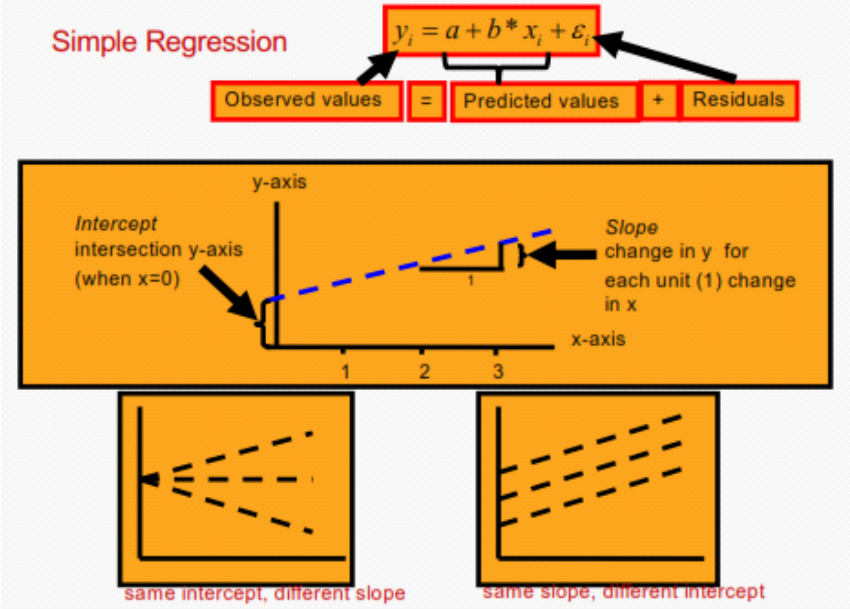
	
Define three types of Post hoc analysis to repeated measures ANOVA	Post hoc: contrasts Simple - 1 vs 2, 1 vs 3, 1 vs 4 Repeated - 1 vs 2, 2 vs 3, 3 vs 4 Deviation - Mean level to overall grand mean Helmert - Each level to later levels
Define sphericity. **Why the sphericity of an experiment with two groups cannot be calculated?**	Sphericity - Differences between conditions Determine by Mauchly's test (chi-square) Homogeneity of the variance of the differences (variance1 = variance 2 = variance3  *Sphericity cannot be done when there are only two groups!*
	SPSS gives you r squared (overall effect size) = square root of partial Eta squared 
What is the main difference between factorial repeated measures and repeated measures ANOVA?	Factorial repeated measures Within subject factor - Variable is manipulated within a subject - Time, dose, sessions Between subject factor - Variable distinguishes different groups
	Example - Diet Main diet

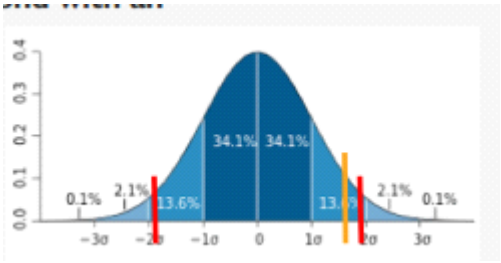
	Main gender Interaction  Each point has a research question (H0 and H1) You always need a control group
How can each of this effect be visually observed in a graph: **a) Manipulation** **b) Group** **c) Interaction**	Understand graph interactions for exam  Manipulation - The mean of pretest and posttest is different Group - There is a space between the lines Interaction - The lines are going in different ways
When should Friedman's ANOVA be used instead of factorial repeated measures?	Friedman's ANOVA - Non-parametric version Ordinal data Small sample size Sphericity assumption violated Differences are not normally distributed Does not allow the study of interactions Only allows non-parametric post-hocs
When should the Mc Nemar test be used?	Mc Nemar test - Repeated measures for dichotomous data E.g. Drinkers and non-drinkers

Assignment review

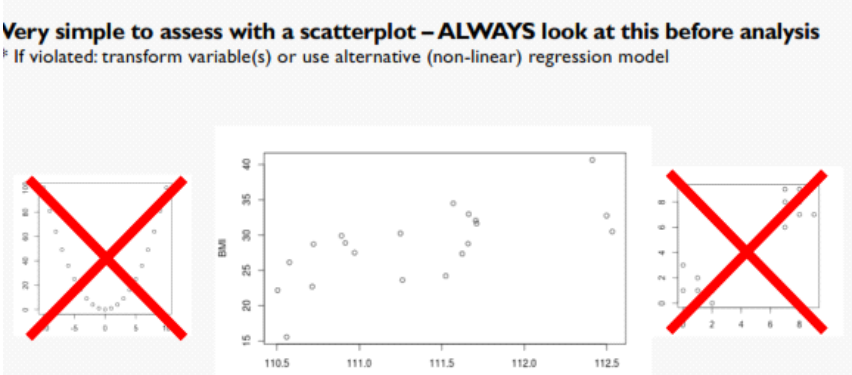
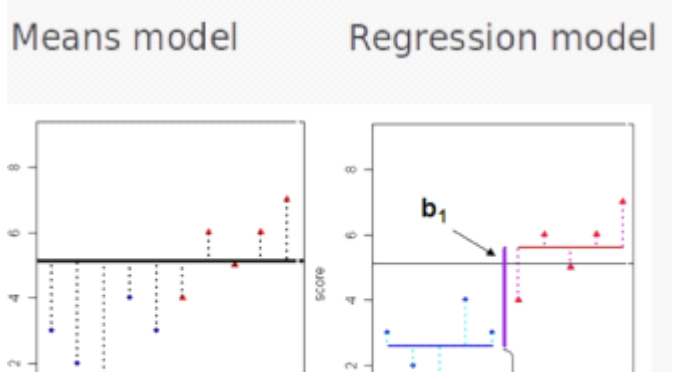
- Multiple groups - N, mean and standard deviation for each
- Always check assumption before running the analysis
  - Report df, test statistics and p-value
- F-test has two p-values - Report both (degrees of freedom of model and the distribution)
  - SSt = n - 1
  - SSr = N - k
  - SSm = k - 1
- Always explicitily state research question, H0 and H1

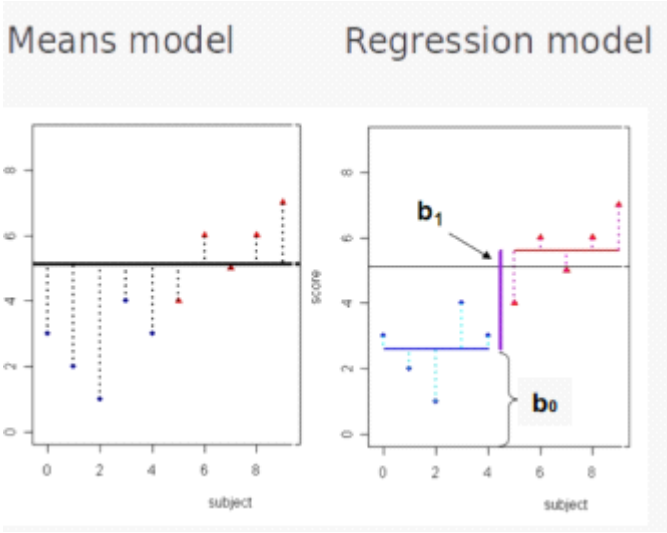
5 - Multiple regression

What is a and b in a simple regression model?	Simple regression - Describes a line Y = a + bx Explained variance in regression formulas How much of the observed values are predicted by the model?  SSr - Variance of the regression model to the means model
What is the criteria to assess if you should use a regression or not?	Criteria for regression model evaluation Hypothesis testing Regression coeffients Full model Parsimony - Simple models with few predictors are prefered to complex models
What does B represent in an one-side regression model? And in a two-sided regression?	Two-sided test Null hypothesis - Slope (B) = 0 Alternative hypothesis - B different from 0 One-sided test Null hypothesis - B <= 0 Alternative hypothesis - B > 0



What does an F-test represent?	F-test Ratio improvement of regression model compared to means model $F = \frac{MS_M}{MS_R} = \frac{SS_M / df_M}{SS_R / df_R} = \frac{(SS_T - SS_R) / df_M}{SS_R / df_R}$ Explained vs non-explained
Why is an adjusted R squared important to be reported?	Example on SPSS Arithmetic ability based on Raven task? - Adjusted R squared - As you add more predictors to the model, r squared tend to increase on its own B - constant (a), intercept Beta coefficient - for every increase of x, how much is the increase of y (standardized in stardard deviations)
	How to report a regression Arithmetic ability and IQ were measured in 170 children. Arithmetic ability was measured on a scale… ranging from… to … (M_a, SD_a); Raven was measured on… (M_{iq}, SD_{iq}). Aim of the study: to investigate whether IQ significantly predicts arithmetic ability. To this end, a regression analysis was run with arithmetic ability as dependent variable, and IQ as predictor. The F-statistic was significant ($F(1,168)=15.87$, $p < .001$), implying that… Both the intercept and the regression weight for IQ were significant ($t(168)=6.03$, $p < .001$, $t(168)=3.98$, $p < .001$, respectively). The resulting equation thus looks as follows: $Arithi = 74.17 + 1.145*Raveni + \varepsilon i$ Correlation Raven-Arith was .29, and R^2 equalled .09 (adjusted R^2 was…), meaning that… Inspection of the standardized residuals shows that less than 2% > |2.58|, and < 5% > |1.96|, implying that…
What are the implicit and explicit assumptions to perform a regression?	Implicit assumptions for regression Predictor is quantitative or categorical with 2 levels Dependent variable is continous No perfect multicollinearity between predictors Predictors are not correlated with external variable Explicit assumptions for regression Linearity - Always assess this with a scatter plot Independent observations - No repeated observations from the

	same subject, no hierarchical organization in subjects/animals Durbin-Watson test Graph residuals can identify gross deviation Homoscedasticity - Variation outcome should be constant across predictor range Boxplot Normal residuals - Mean equals 0 If violated: Non-parametric testing Outliers can influence regression model  - Checking for outliers is very important in regression analysis - Always make a scatterplot of the data
If regression and correlation are the same thing, why do regressions at all?	Why bother with regression? Mathematically, regression and correlation are the same Conceptually, you frame them differently (regression provides a scalable framework)
What is a multiple regression?	Multiple regression More complex questions require more complex models Y = b0 + b1X1 + b2X2 + b3X3
What is R-squared?	R squared - percentage of variability explained by all predictors
What are the different types of predictors entry for multiple regressions?	Enter mode - All predictors entered in the equation at once (no order) Hierarchical regression - Order of predictors is chosen by experimenter Stepwise - Decided by the computater (which generates largest R squared) Forward - Improve R square by adding predictor Reverse - Improve R square by removing predictors
How to deal with categorical variables in a multiple regression?	How to deal with categorical variables? 

categorical variables in a multiple regression?	 Dichotomous - Indicator variable (0 and 1) T-test can be coverted into a regression model More than two variables- Dummy coding Number of dummy categories equals number of levels minus 1
What is dummy coding?	What if effect of predictor depends on group - Interaction is determined by dummy coding **Condition** / **X1** / **X2** / **X3** **No intervention** 0 0 0 **Placebo** 1 0 0 **Old drug** 0 1 0 **New drug** 0 0 1
	Multiple continous variables Displacement measured as continous variables **Coefficients[a]** Unstandardized Coefficients: B, Std. Error; Standardized Coefficients: Beta; t; Sig. Model 1: (Constant) 33.403, 2.837, 11.776, .000 number of cylinders -2.340, .395, -.690, -5.919, .000 transmission 3.331, 1.463, .265, 2.276, .031 a. Dependent Variable: miles per gallon • Intercept (H_0: $b_0 = 0$) • Main effects (H_0: $b_1=0$ / $b_2=0$) • Beta & Std. beta control for effects of other predictors, so coefficients will change when they are added/removed from the model • Remove non-significant predictors and re-evaluate final model
What are the assumptions of multiple	Assumptions for multiple regressions Linearity

regressions?	Independent observations Homocedasticity Normal residuals No perfect multicollinearity
What is Variance Inflation Factor? From which value should you consider removing a variable?	Very large correlation between covariates is problematic **Variance Inflation Factor (VIF)** is measure to estimate how redundant each covariate in your regression is - It performs a regression with one covariate as outcome and all other covariates as predictors [VIF = 1 / (1-R^2)] - VIF = 1: uncorrelated with all other covariates in the original regression (R^2=0) - VIF > 5: very correlated; consider removing this variable (R^2>0.8) - VIF > 10: extremely correlated; you probably should remove this variable (R^2>0.9) Each predictor has its own VIF; if necessary remove variable with highest VIF and recompute.
What is a familywise error?	The more tests you do, the more likely you are to get a false positive Setting alpha (probability to reject the null hypothesis even though it was true = type I error) Familywise error = 1 - 0.95(n) 10 tests = 40% In neuroscience, many designs are characterized by many tests Proteomics Genomics Metabolomics
What are the classes of correction for familywise error?	Classes of correction Family-wise error rate (FWER) False discovery rate (FDR) False discovery proportion (FDP)
What is the difference between Bonferroni correction and Holm correction?	Familywise error rate - Bonferroni correction - Divide alpha by the number of tests Often too conservative when p-values are correlated Holm - Sort the p-values sequentially, than define the critical value for each p-value (ai = m-i+1) Less conservative in Bonferroni Tukey - Similar to Holm (put p-values in order)
What is false discovery rate?	False discovery rate FDR - Hits by chance alone and hits observed -> Ratio Benjamini & Hochberg Read Verhoeven (2005) 'm' - Control for the number of hypothesis

Assignment review

Errata - Effect size (Z-score/square root of N)
N is the total number of observations in the study

Normality tests - Always report df, test statistic and p-value
Always report the specific p-value

4.1.

The two hypothesis (null and alternative) need to cover all the possible outcomes

The **differences** should be normally distributed

Rsquared =

Always report effect size

4.2

Weight of three strains of mice

- Main weight - The weight changes across time
- Strain - Do the strain differs in weight
- Interaction - Do the strains gain weight in different ammounts

Always report all the effects
You don't do a post-hoc test if your effect is not significant

4.3

DBP in increasingly stressful conditions
Always report the groups, time points, what is being tested (N, median, sd)

Test the normality of differences
Friedman ANOVA is a chi-square and needs to be reported as such
Always report effect size

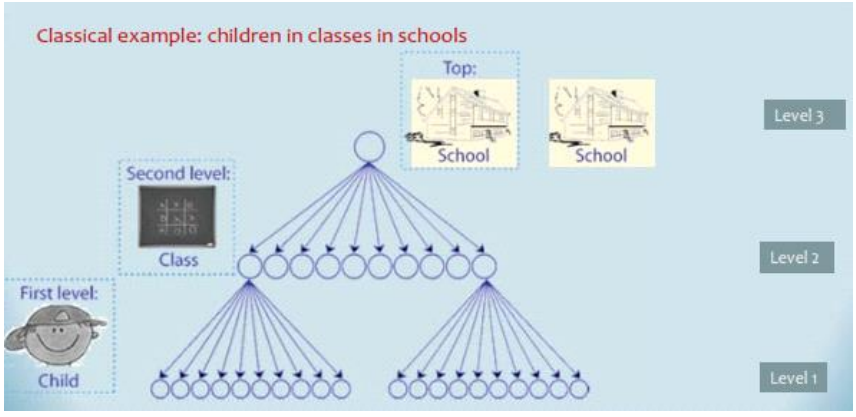
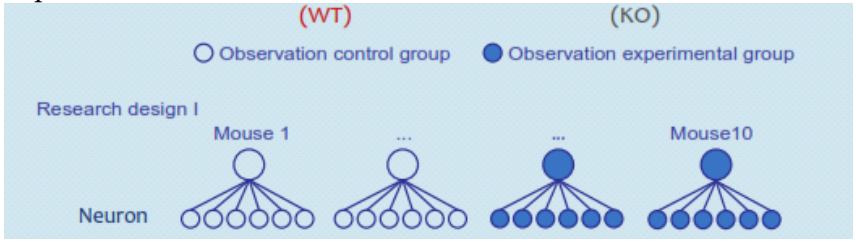
Bonferroni - Related samples
Tukey - Unrelated samples

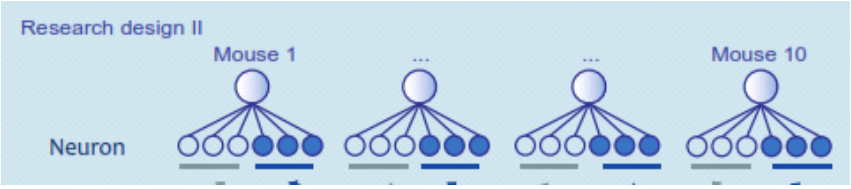
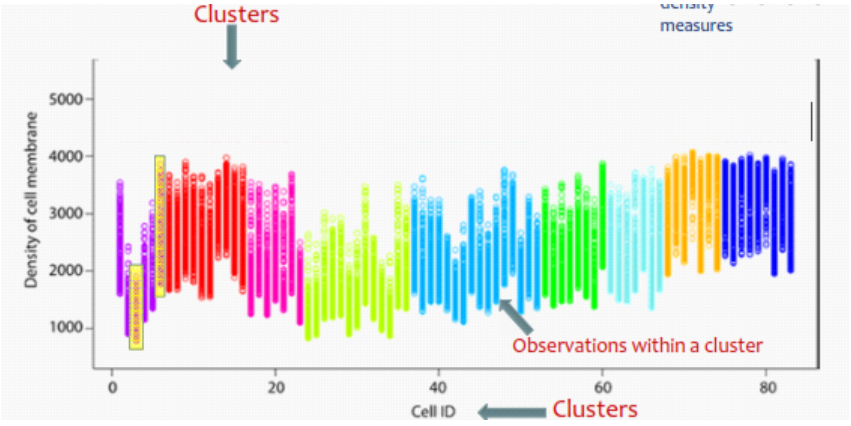
1 or 2-sided question
Dictated by your hypothesis/study design

T-test/ANOVA - Two-sided

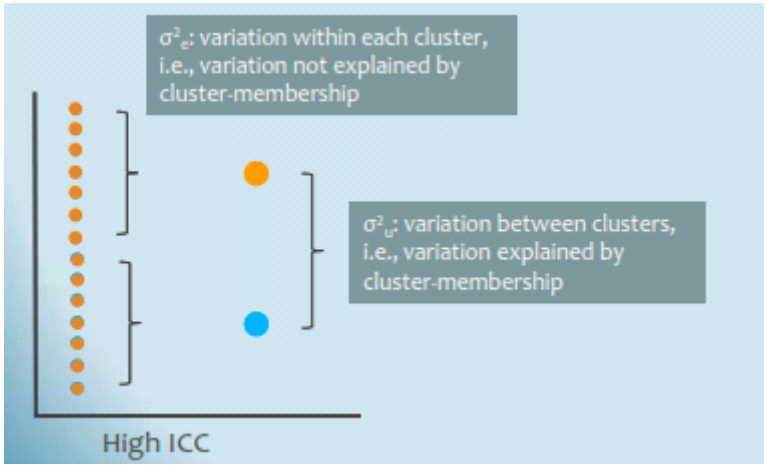
Paired t-test - One-sided

6 - Multilevel analysis, handling clustered data

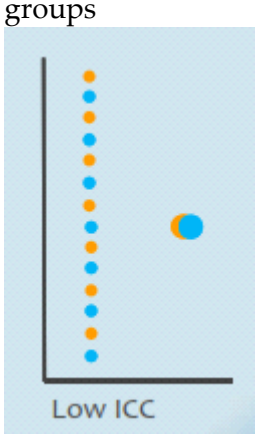
What is the difference between longitudinal data and clustered data?	**Multilevel analysis** General assumption of statistics: Data are a random sample from the population 1. Longitudinal data - Dependency over time (paired t-tests) 2. Clustered data - Data are not randomly sampled
Why the violation of assumption of independence creates more false positives?	**Violating the assumption of independence** Within the class - Scores are more alike (they are clustered), dependent by design This creates a high false positive rate
What is hierarquical data?	**Conceptual introduction** Nested/clustered/hierarquical data - Multiple measurements of the same object - Children in the same class (child is level 1, class is level 2, school is level 3)  - Neurons in the same pup, multiple axons/dendrites from the same neuron - Patients from the same doctor 53% of neuroscience data in papers included nested data
What are the two types of nested designs?	**Two types of nested designs** 1) All observations from the same cluster are in the same experimental condition  2) Observation from the same cluster can be part of different experimental conditions 

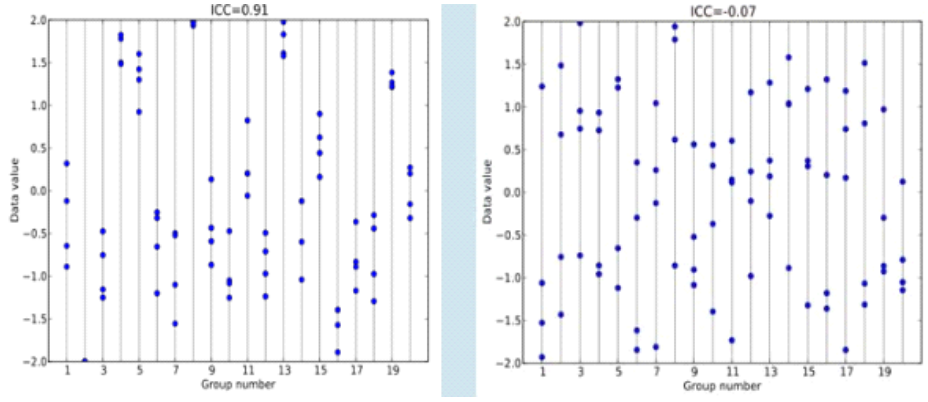
	
What is the difference between observed and effective sample size?	Observed vs effective sample size 3 WT pups: 10 neurons per group 3 KO pups: 10 neurons per group Observed sample size: 60 measures Effective sample size: Ranges between 6-60 depends on ICC ICC: Intra-cluster correlation 0 - All samples are independent 1 - All samples convey the same information
How can clusters be visually identified in a histogram?	Density of cell membrane - Observations of the same cell are more similar to each other  You cannot do ANOVA/t-tests on hierarchical data
What does ICC stand for? **What is the formula for ICC?**	**ICC** - Ratio of variation in the data explained by clustering/Total variation in the data $$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$
Visually, what is the difference between high	**High ICC** - Small variation between groups, large variation among groups

and low ICC?



Low ICC - High variation between groups, small variation among groups

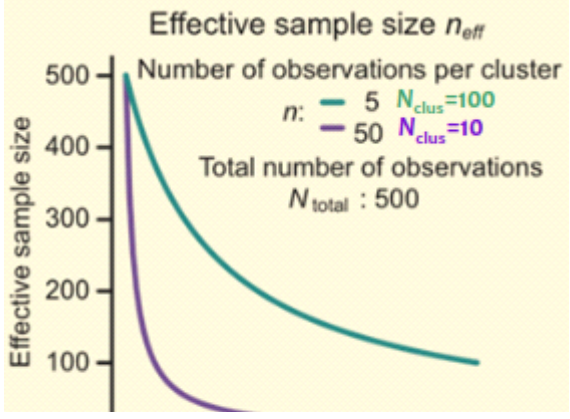


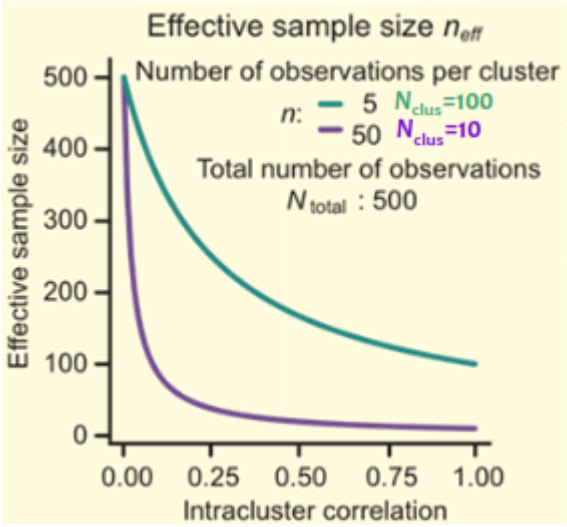
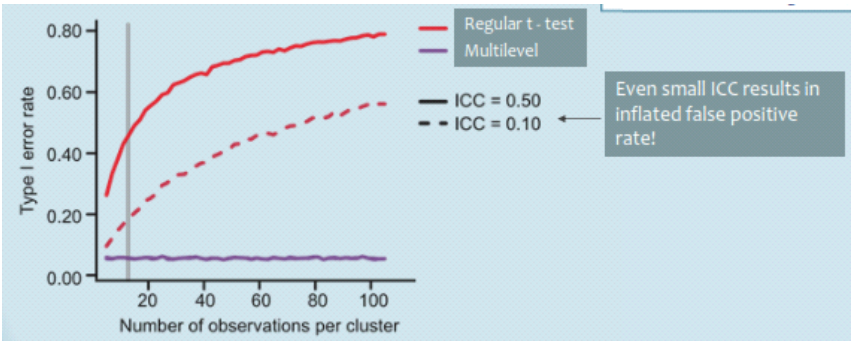
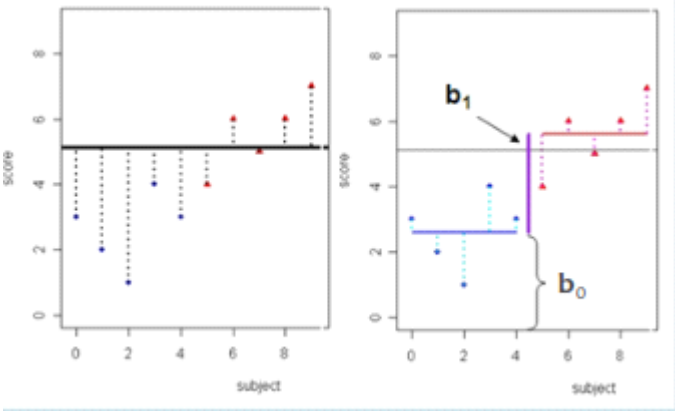


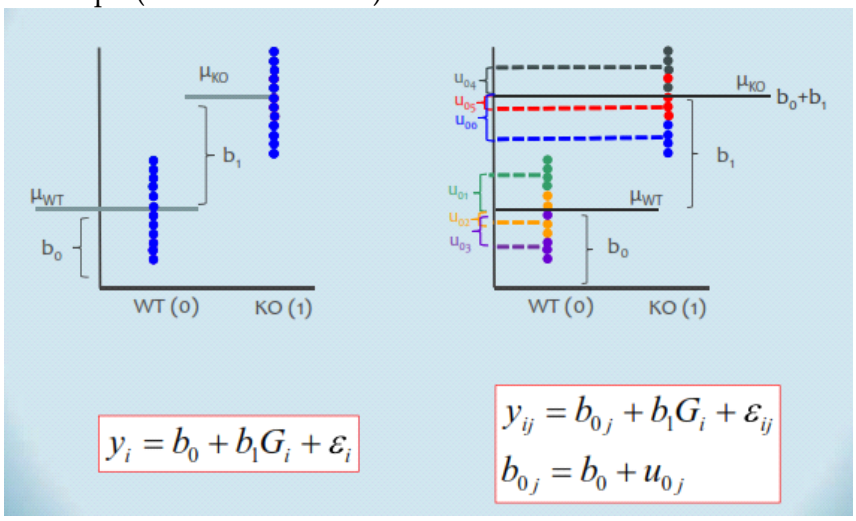
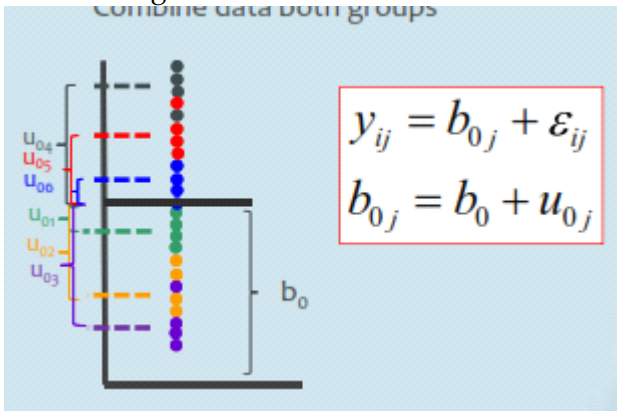
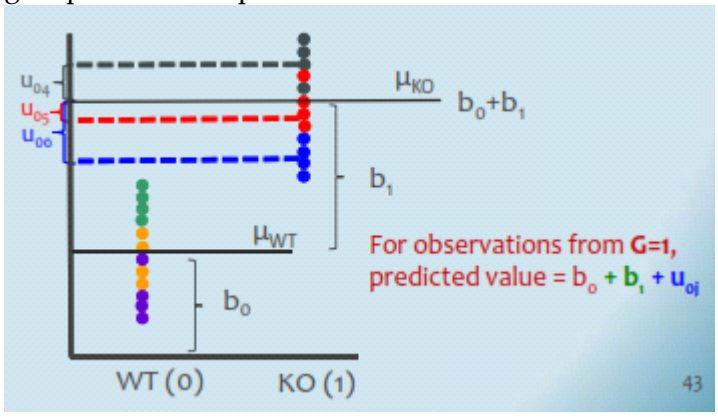
If the ICC is small, does that mean that you can perform a normal t-test/ANOVA with your data?

Effective sample size

Neff = Ntotal / 1 + (n-1)ICC



perform a normal t-test/ANOVA with your data?	 Even if ICC is small, it can have a large influence if your sample size is very large
What happens when you do ANOVA/t-test with clustered data with: **a) Standard error** **b) Degrees of freedom** **c) p-values**	If you do not do multilevel analysis with clustered data: Underestimate SE and p-value Overestimate t and df  Test is too significant - Type I error
What are the three ways that you can handle dependency in your data?	**Handling dependency** Average within each cluster - Lose information Take a random observation from each cluster - Lose even more information Multilevel analysis - Does not lose information, power of test increases
What is the only difference between normal regression and multilevel analysis?	**Normal regression** Intercept - b0 Slope - b1 

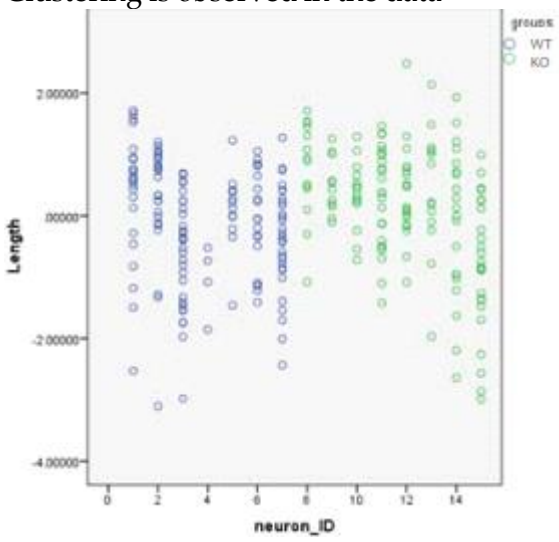
	Only difference for Multilevel analysis: Submodel for the intercept! (Mean + deviation)  Cluster specific deviation - From the group mean Sigmau - Intercept variance Sigmae - Within the clusters
What is the difference between intercept only model and multilevel models?	Multilevel: Random intercept **Overall ICC (intercept only model)**- Total variability in the data due to clustering  **Explained ICC (multilevel model)**- Variability present when taking group membership into account  **Unexplained ICC** - Variability that remains after inclusion of group membership

What are the two different types of research design in multilevel analysis?

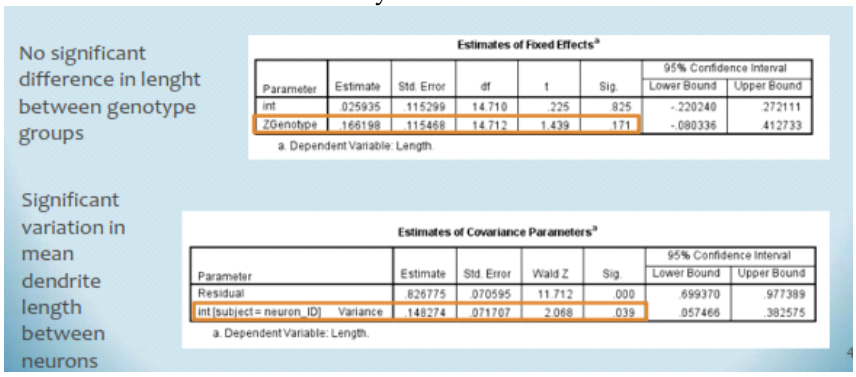
Examples of research design

Dendrite length in WT and KO

Clustering is observed in the data



Covariance is significant (difference between the clusters) - Multilevel model is necessary

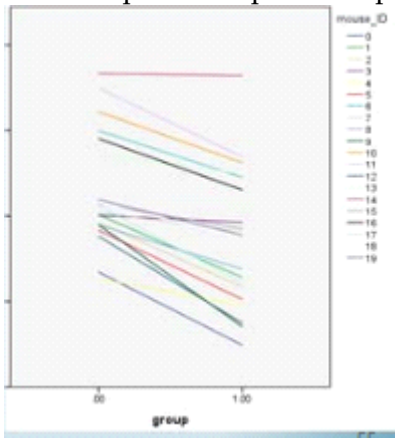


Research design II

Multiple observation from the same mouse, part is treated with virus, other part does not have virus

Cluster-specific intercept - Mean

Cluster-specific slope - Manipulation effect



Fixed effect - Same for all clusters

Random effect - Different for different clusters

	All neurons are stained with GFP - part is transfected with virus, part is not ○ Significant intercept variance ○ Significant slope variance **Always do multilevel analysis with clustered data, even if covariance parameters are not significant**
	Conclusions Clustering needs to be taken into account in statistical test ML is just an extension regression analysis Read more about multilevel modeling
	Questions: • What is meant by "dependent data" • Observed versus effective sample size • What is the ICC • Why is multi-level modelling required if data are not independent • What is menat by "clustering" or "hierarchical data"?

Assignment discussion

5.1. Is sex still a predictor of ICV once height is taken into account?

Null hypothesis - Sex does not improve the prediction once height is taken into account
Alternative hypothesis - Sex does improve the prediction once height is taken into account

Check assumptions

Mean + standard deviation of groups need to be reported separately

The fact that height is not significant does not mean that you can exclude it from the model - you need to run the regression again without the variable

5.2. Sepal length can be predicted by petal length/width

1. Dummy coding
   Two groups = a + bG
   - bG = 0; intercept straight line
   - A = 0; sloped line b

   More than two groups = a + bG1 + bG2
   - Dummy coding - Compare each variable to the reference group (does not matter which one)
   - Allows an ANOVA to be performed in the context of a regression
   - B1 - Group 1 from reference group
   - B2 - Group 2 from reference group
2. Mislabelling
3. Check for outliers

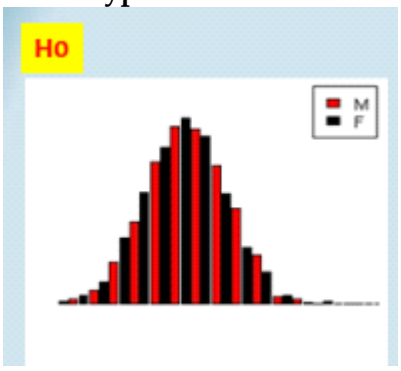
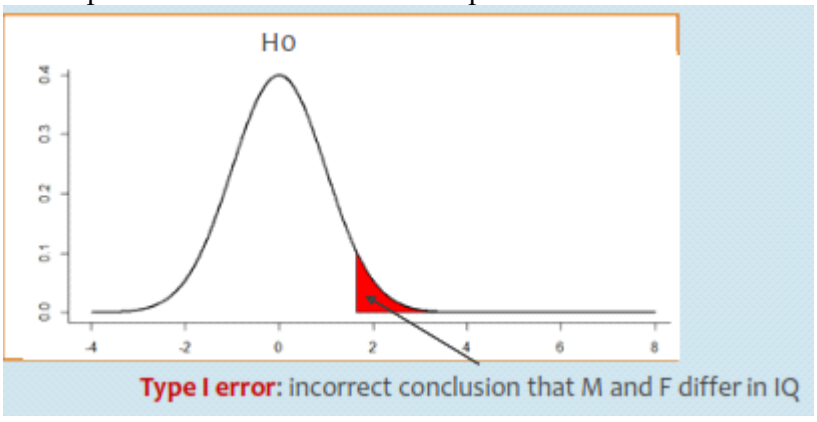
Petal length and petal width are highly correlated - Cannot both be in the model (VIF > 18)

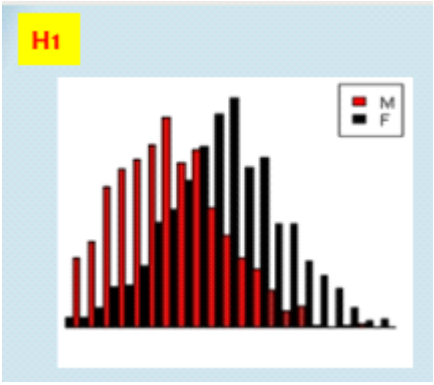
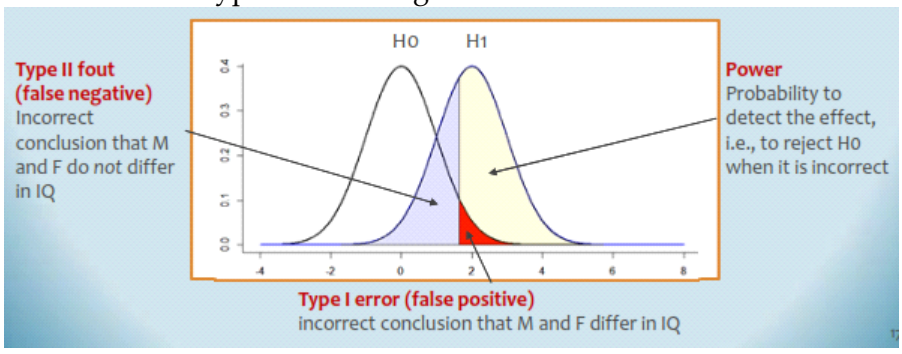
You must include all dummies or no dummies - you cannot include only some dummies

Hierarchical entry - You know the predictors (from literature)
Stepwise entry - You do not know, you let the computer decide the best model

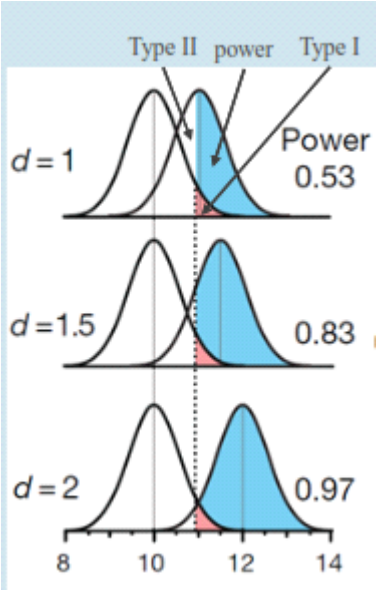
Two predictors - Petal length + species (do not count dummy codings individually)

6b. Power

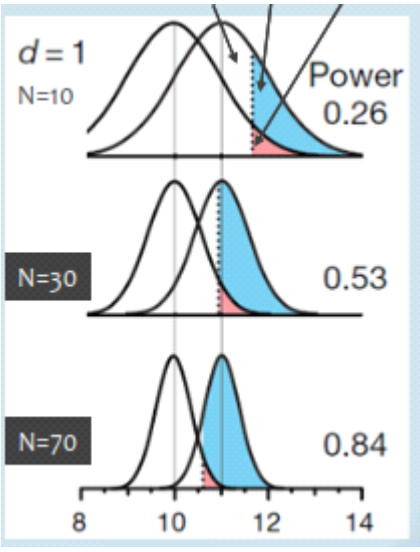
What is the definition of Power?	**Power: Probability to detect an effect that is actually there** Power = 1 - beta Probability of rejecting the null hypothesis if it is false **Possible research outcomes** / **H0 is** True / False Researcher — Accept H0: True negative 1-α / False negative Type II error (β) Researcher — Reject H0: False positive Type I error (α) / True positive 1-β: power!
Why is there no type II error under the null hypothesis? **How is type I, type II error and power represented in a t-statistic graph?**	Example - Men and women differ in IQ scores Sampling theory - Most results will be 0, but some test statistics will yield a false positive **Under null hypothesis**  Tail represents the chance of a false positive  **Under alternative hypothesis**

	 Type I error - Right tail of left curve Type II error - Left side of right curve Power - Above type I error in rightmost curve  This is the distribution of t-values (test statistics - not the distribution of the data itself!)
What is the relation between type I and type II errors?	There is an inverse relation between Type I and Type II erros If alpha gets smaller, beta gets larger
What is sensitivity and specificity?	**Sensitivity** - Of all positives, how many are true positives Can you identify the sick **Specificity** - Of all negatives, how many are true negatives Can you identify the healthy
What are the three ways you can increase power? **How is the increase of effect size represented visually?**	**How to increase power?** • Increasing sample size • Increasing difference between two means (effect size) • Decreasing variance **Increase effect size** - Make test more difficult

How is the increase of sample size represented visually?



Increase sample size - Less sampling variation, more accurate estimate of the effect

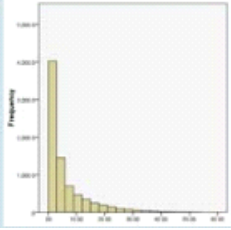


The curve becomes narrower (less variation)

Decreasing noise - Control through careful matching and selection; use of a good instrument (standardized, reliable, objective, high resolution)

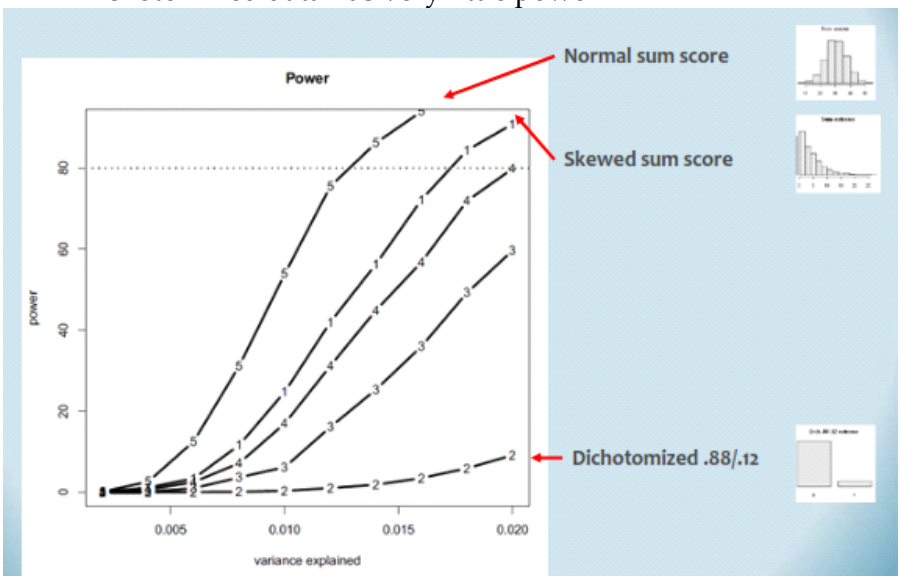
How is power related to the way you collect your data?

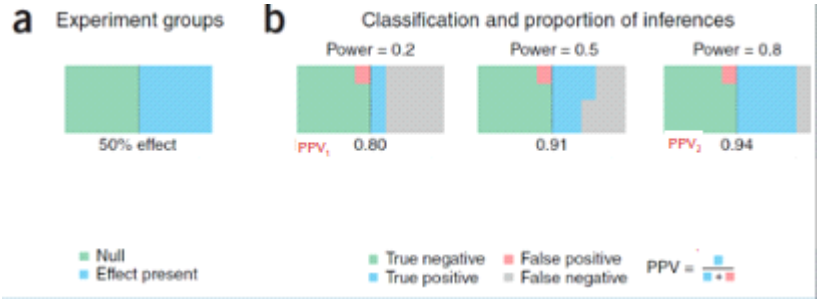
Depression questionnaire



	Rarely /never	Sometimes	Occasionally	Often/always
6) I felt depressed	0	1	2	3
9) I thought my life had been a failure	0	1	2	3
10) I felt fearful	0	1	2	3
14) I felt lonely	0	1	2	3
17) I had crying spells	0	1	2	3
18) I felt sad	0	1	2	3
19) I felt that people disliked me	0	1	2	3

Extreme questions - Most people will score 0, even though they do not feel equally well

	Controls = 0 Loss of a lot of information Cases = 1 Dichotomized data has very little power  The instrument we use and statistical treatments change the power of your data
What is censoring? What are `bottom` and `ceiling` effects?	**Censoring** - Loss of information in your data **Bottom effect** - Some could not perform the task at all **Ceiling effec**t - Some answered everything correctly
What is the mean power of neuroscience studies? Why is that a problem?	Current state of affairs Neuroscience studies - **Mean power of 21%** Power should be at least 80% You cannot trust what is discovered You cannot trust what is not discovered Underpower studies are useless
What is the problem with increasing false negatives due to underpowered studies?	Problems with underpowered studies **1. More false negatives** a. Most research focus on minimizing false positives b. False negatives - Waste of money, effort and time; negative results cannot be interpreted as the absence of an effect
What is PPV? **What is the relation between PPV and power?**	**2. Observed effects are less likely to be true** a. Positive predictive value - The effect that we observed is likely to be a true effect (depends on the knowledge of the field) Shoe sizes varies between men and women - Safe hypothesis, needs very little evidence Mind-reading - Risky study, it takes a lot of evidence to convince readers

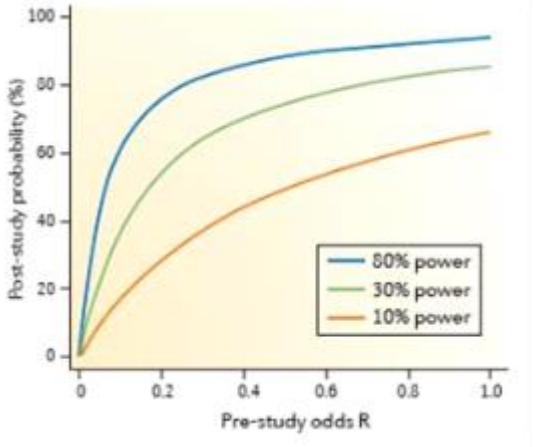


Calculation of PPV = (1-beta) * odds/(1-beta) * odds + alpha

Odds = sucesses/failutes

True positives/ true + false positives

The probability that something is true determines the likelihood of what you will find is true

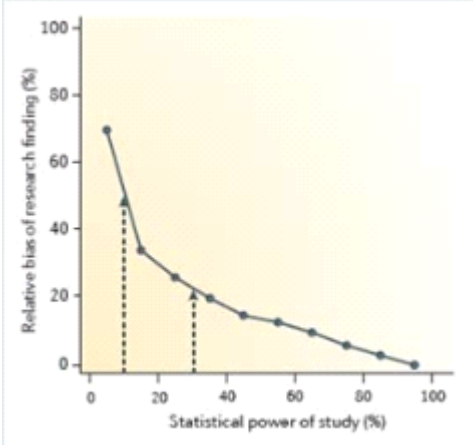


What is Winner's curse? What is its relationship with power?

3. Overestimation effects

Winner's curse - Detected effects are overestimated/inflated

Large effect in false positives are more likely to occur with a small sample



Study replication often finds contradictory results

Why are underpowered studies unethical?

4. Unethical

Low powered studies are inefficient and wasteful

Sacrifice of animals, public funding wasted

Underpowered studies do not yield any conclusions

	Conclusion Given that I want to see an effect of at least a certain size - how can I design my study A priori testing - G-power software Small power - DO not do the study or increase sample size
	Questions • What is power • False positives/false negatives (the table) • Factors that affect power (e.g., N, effect size, variance: learn to interpret statistical equations (like t-test) in terms of power) • Consequences of poor power (e.g., false negatives, Winner's curse, PPV, unethical)