Experimental Design

An Overview

First things first...

- What would you like to know?
  - Hypothesis, predictions?
    - H_0 and H_1
- Identify the "conditions", i.e. what do you compare??
- What do you manipulate, what is measured?
- What will the result let you conclude?
- Why all this... (Floyd Bullard, Duke University)

Select the right type of study

- Control
  - What can you manipulate?
- Cost & effort
  - Is it worth the time and money?
- Ecological validity
  - Does your design represent and address a realistic scenario/ problem?
  - Can you find the appropriate type of participants?

Control (Variables)

- Independent Variables
  - what you manipulate
- Dependent Variables
  - what you measure
- Control Variables
  - what you hold constant
- Random Variables
  - what you allow to vary randomly
- Confounding variable
  - correlated with independent variable

Example: Smoking

Participant	DV (gender)
1	m
2	f
3	f
4	m
5	f
...	...
196	m
197	m
198	f
199	m
200	m

Example: Smoking

- Do more men smoke compared to women?
  - Factor/IV: Smoking (yes)
  - Measure/DV: Gender (male, female)
- Sample: 200 smokers
  - 112 men, 88 women (significant, or chance??)
- Does a a certain medication help to quit smoking?
  - Factor/IV: Medication (with,without)
  - Measure/DV: Smoking (quitted, not quitted)

Baseline

- Independent variable
  - With/without medication
- Neutral and fair baseline condition
  - With medication: Give drug
  - Without medication: Give nothing / Give placebo
- Make sure only the intended manipulation distinguishes levels of independent variable
  - (Not) giving confounds with (no) drug -> Give placebo!

Participants

- Assign to conditions: Within- vs Between-subjects
  - Between: More participants
  - Within: More time per participant
- How many levels of IV? How many subjects?
- Confounding variables
  - Between: Group differences possible
    - Use randomization, many subjects, matching
  - Within: Order effects possible
    - Use counterbalancing

Confounds:
Group differences

- Sample = 200 smoking
  - 112 men, 88 women (significant, or chance??)
- Socio-economic status
  - 105 low SES (90 men), 95 high SES (73 women)
  - ?More men smoke
  - ?People with low SES smoke more often than those with high SES

Example: Smoking

Participant	DV (gender)
1	m
2	f
3	f
4	m
5	f
...	...
196	m
197	m
198	f
199	m
200	m

Example: Smoking

Participant	DV (gender)	DV (SES)
1	m	low
2	f	low
3	f	high
4	m	low
5	f	high
...	...	...
196	m	high
197	m	low
198	f	high
199	m	low
200	m	low

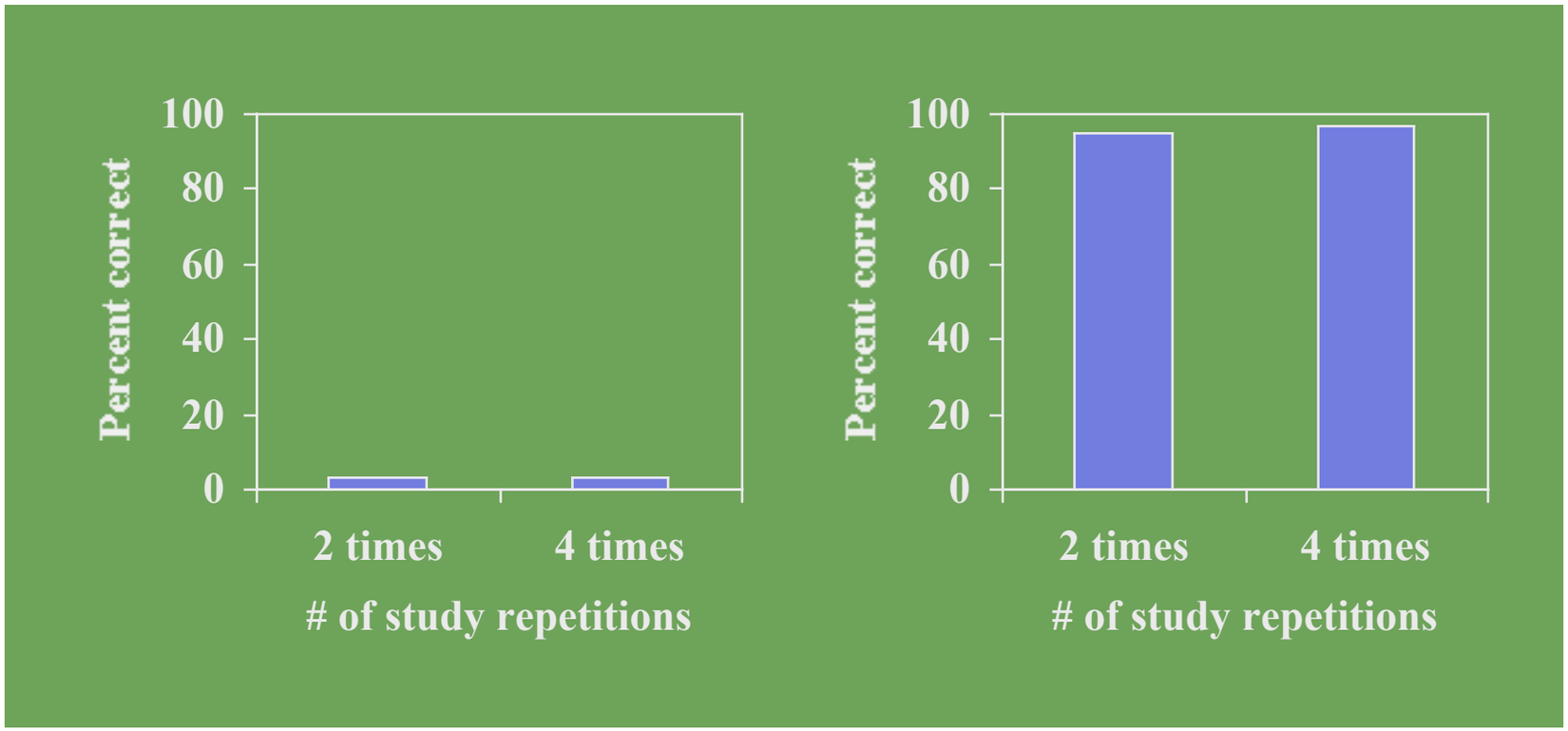
Confounds: Group differences

- Sample = 200 smoking
  - 112 men, 88 women (significant, or chance??)
- Socio-economic status
  - 105 low SES (90 men), 95 high SES (73 women)
  - ?More men smoke
  - ?People with low SES smoke more often than those with high SES
  - Correlation != Causality

Confounds: Order effects

- Adjust condition order to unconfound order effects with condition effects
  - A,B,C
  - A,C,B
  - B,A,C...
- Counter-balancing either within- or between- subjects
  - Between: Joe: A,B,C - Mary: C, A,B - John: B,C,A
  - Within: Joe: ABCCBA - Mary: ABCCBA

Floor & Ceiling effects



Effects & Interactions

- Single Variable:
  - Only one independent variable
  - Cannot look at interactions
- Multiple Variables:
  - Two or more independent variables
  - If use factorial design, can look at interactions
  - Can require a lot of participants (between) or time (within)

Effects & Interactions

- Main effect: A factor has an effect on the DV, independent of other IV
- Interaction: Two independent variables interact when the effect of one depends on the level of the other
- When is an effect/interaction significant?

Example: Smoking

Participant	DV (gender)	DV (SES)
1	m	low
2	f	low
3	f	high
4	m	low
5	f	high
...	...	...
196	m	high
197	m	low
198	f	high
199	m	low
200	m	low

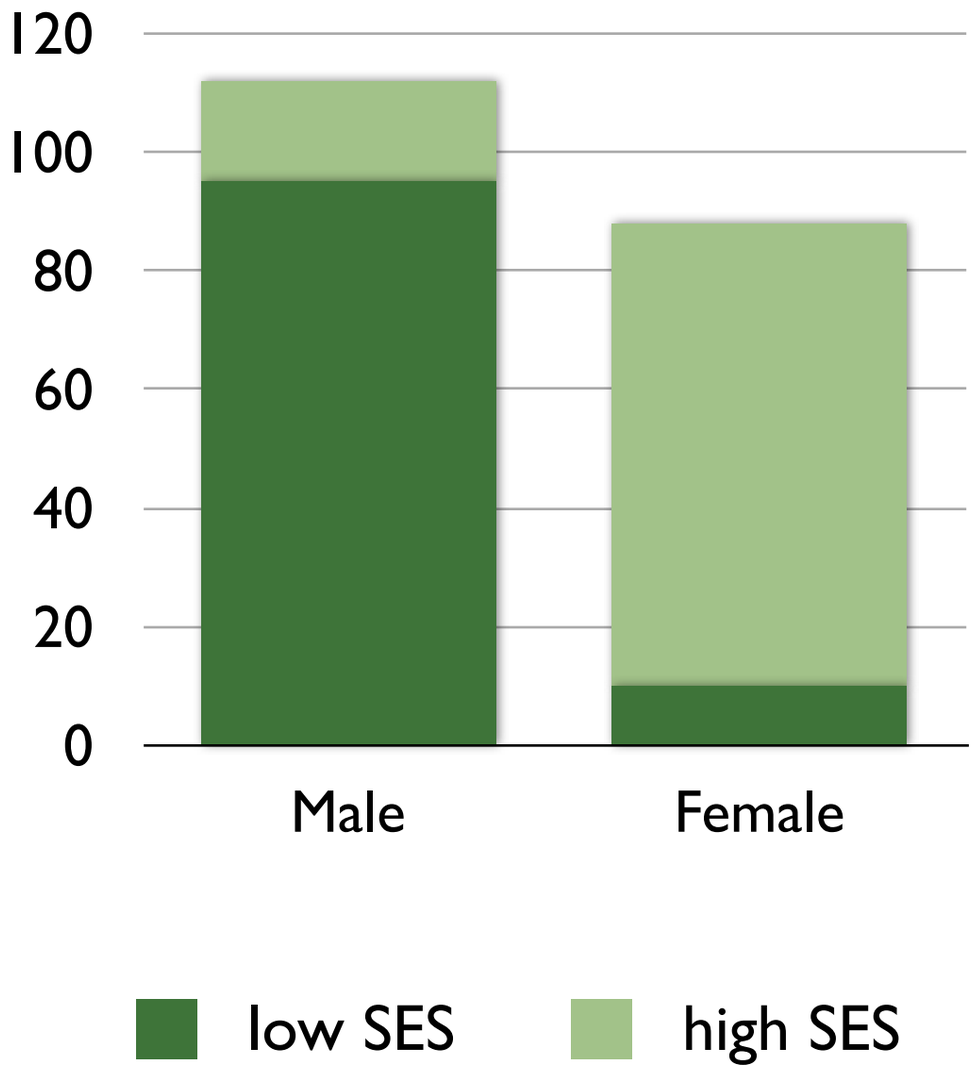
Example: Smoking

Participant	DV (gender)	DV (SES)	DV (age)
1	m	low	young
2	f	low	young
3	f	high	old
4	m	low	young
5	f	high	old
...	...	...	...
196	m	high	old
197	m	low	young
198	f	high	old
199	m	low	old
200	m	low	old

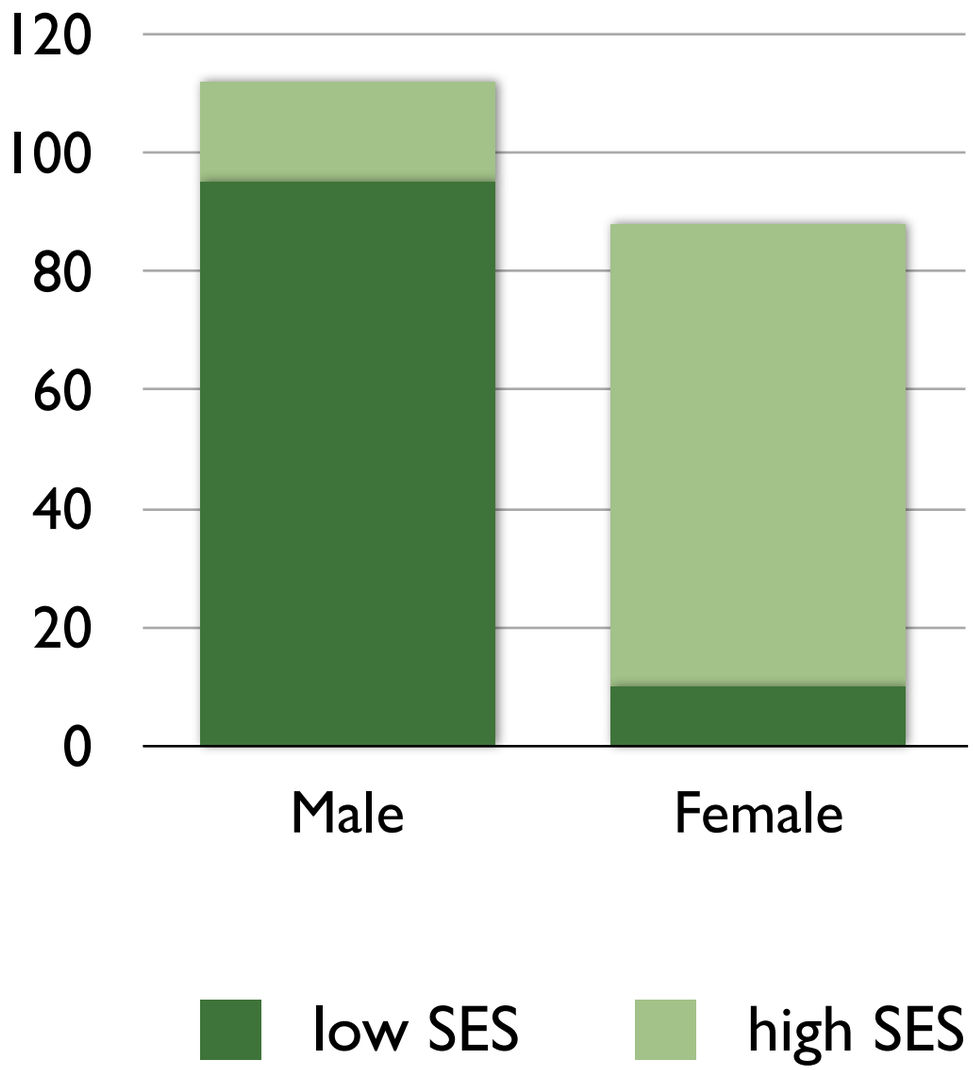
Main effect

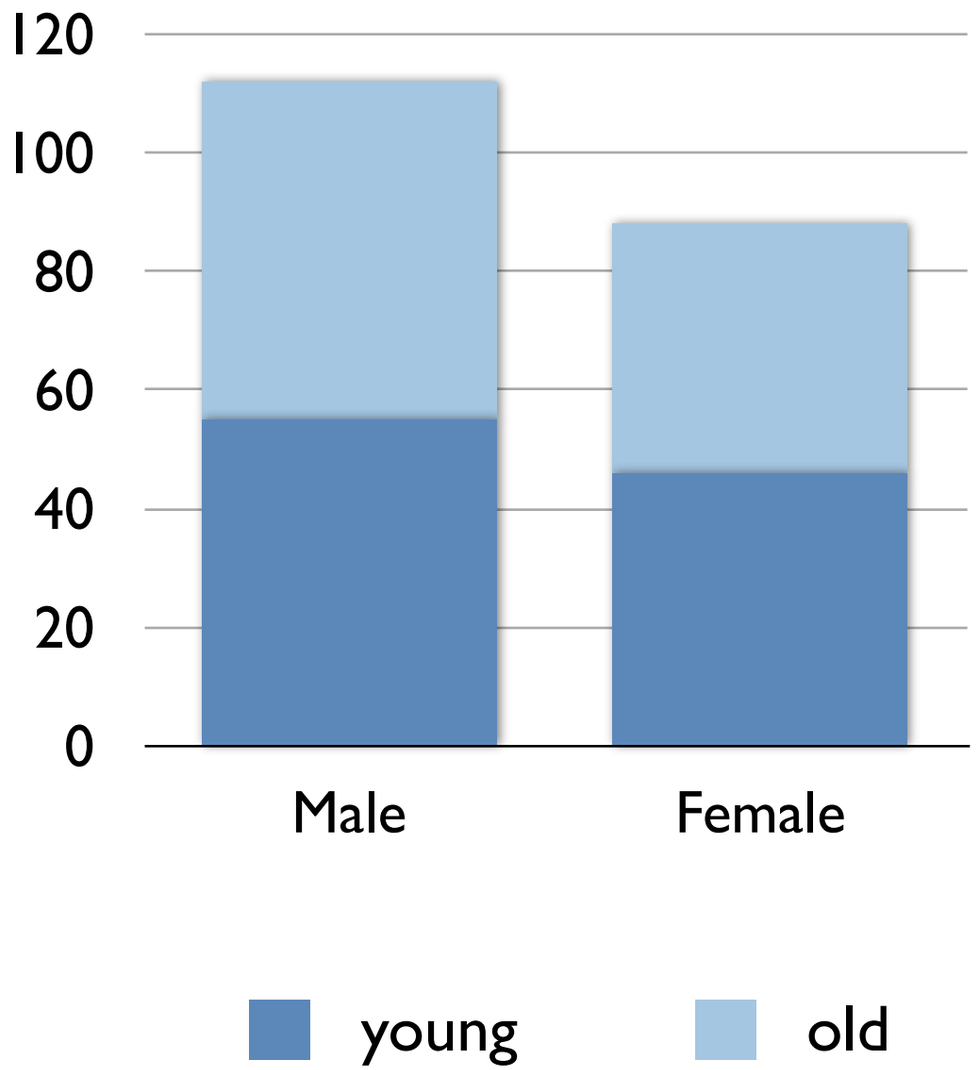
- Sample = 200 smoking
  - 112 men, 88 women (significant, or chance??)
- Socio-economic status
  - 105 low SES (95 male), 95 high SES (78 female)
- Age
  - 101 young (55 male), 99 old (57 male)
  - Main effect? Interaction?

Main effect

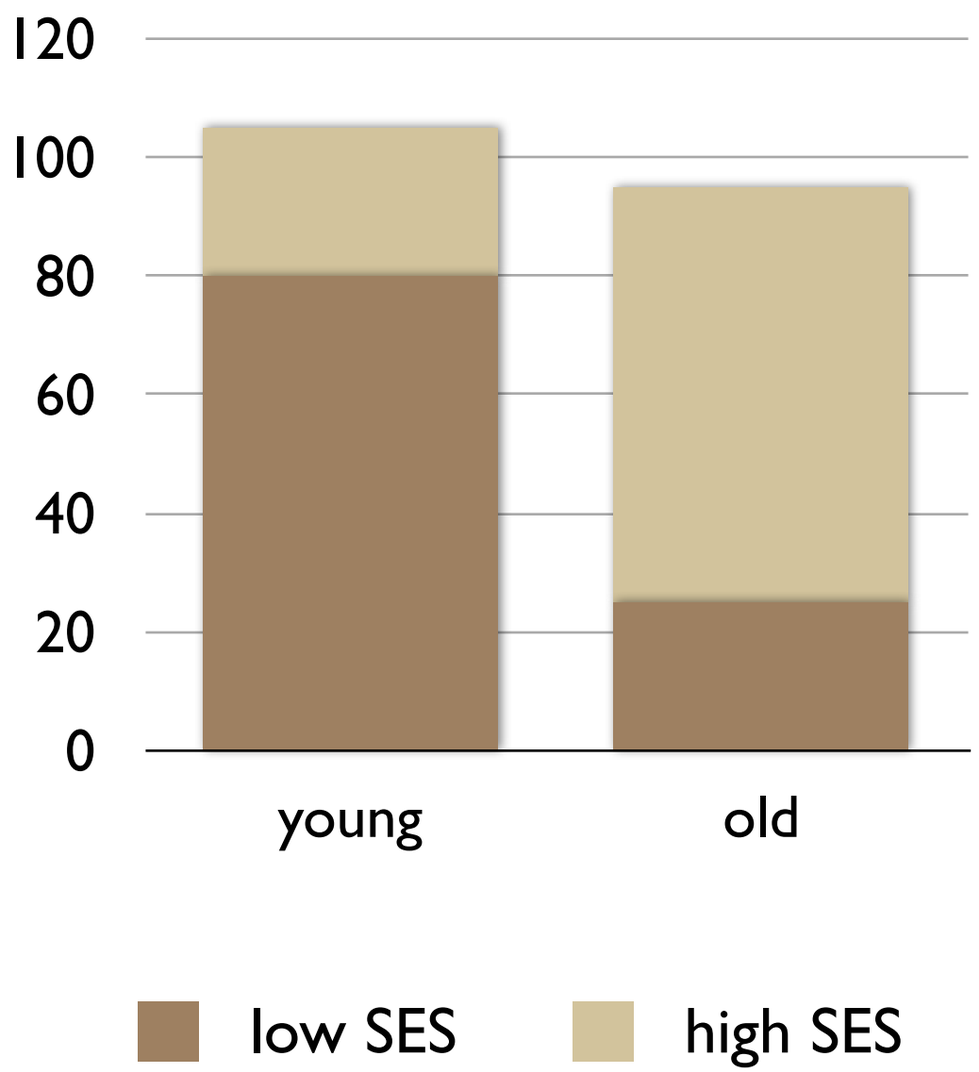


Main effect


120
100
80
60
40
20
0
Male
Female
low SES
high SES


120
100
80
60
40
20
0
Male
Female
young
old

Interaction



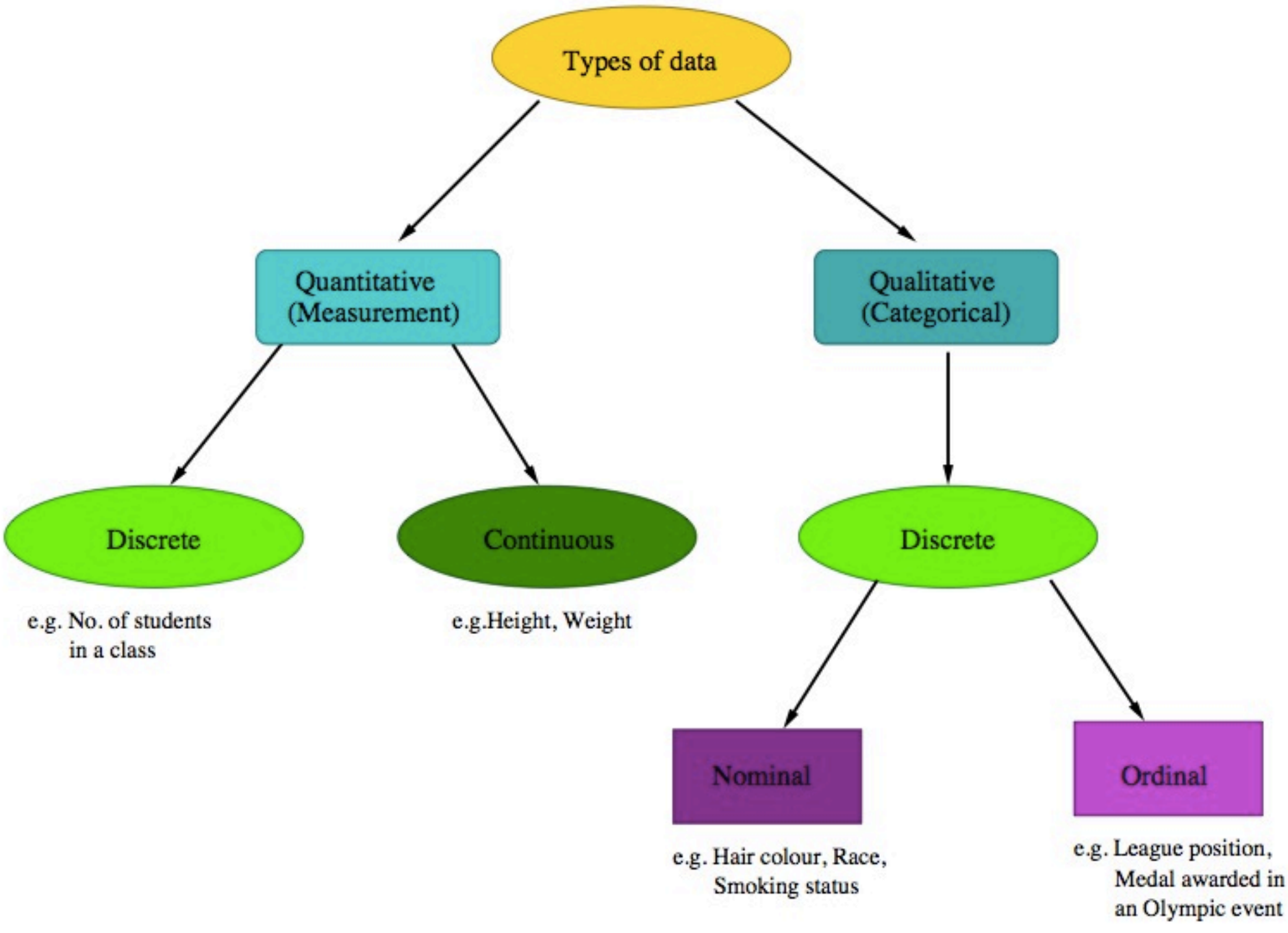
Effects & Interactions

- Main effect: A factor has its own effect on the DV, independent of other IV
- Interaction: Two independent variables interact when the effect of one depends on the level of the other
- When is an effect/interaction significant?

Statistics

- Describe data and draw inferences
- Variable type
  - Measurement vs Categorical (Ordering/Nominal)
  - Discrete vs Continuous

Statistics

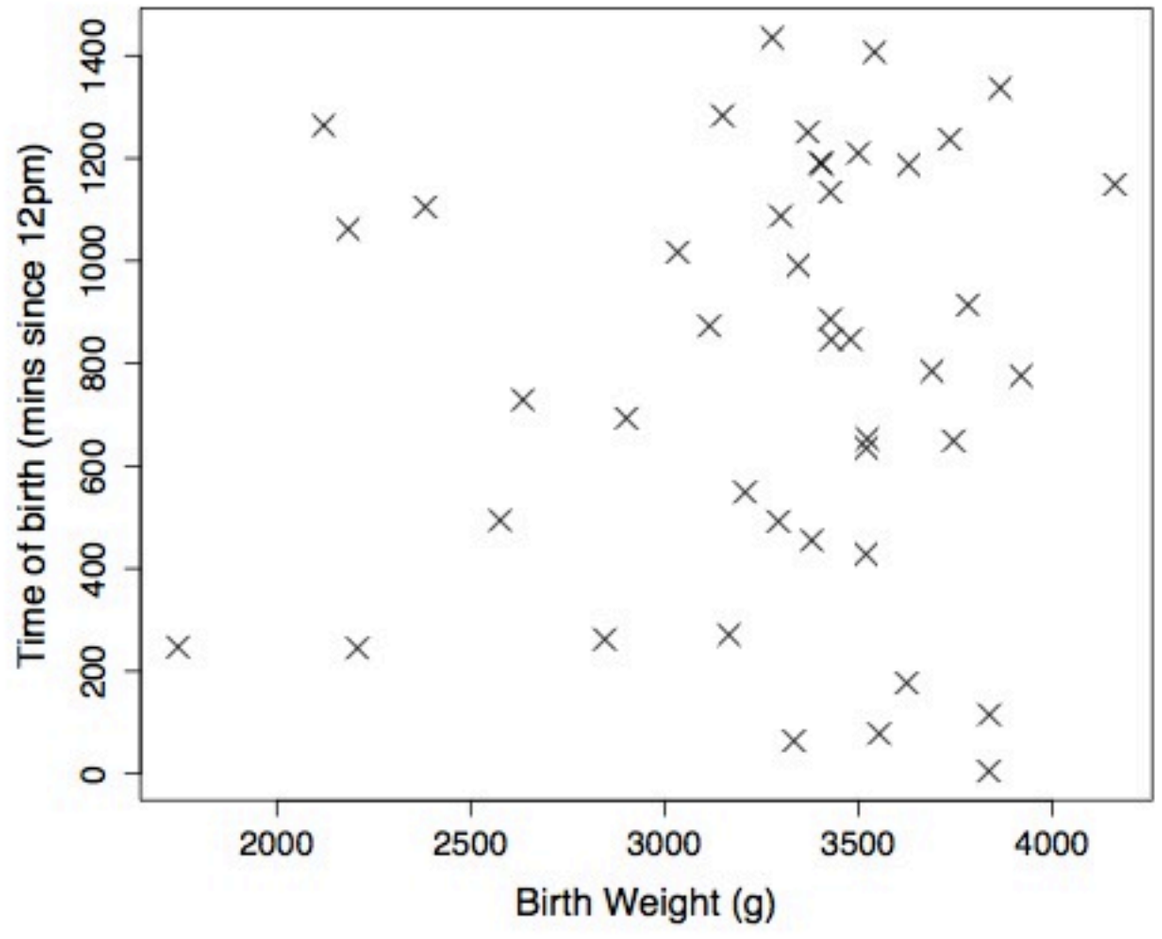

Types of data
Quantitative (Measurement)
Qualitative (Categorical)
Discrete
Continuous
Discrete
Nominal
Ordinal
e.g. No. of students in a class
e.g.Height, Weight
e.g. Hair colour, Race, Smoking status
e.g. League position, Medal awarded in an Olympic event

Statistics

- Describe data and draw inferences
- Variable type
    - Measurement vs Categorical (Ordering/Nominal)
    - Discrete vs Continuous
- Description of data
    1. Plots
    2. Mean, median, mode & variance

Scatter Plots

Scatter plots are useful when we wish to visualise the relationship between two measurement variables.

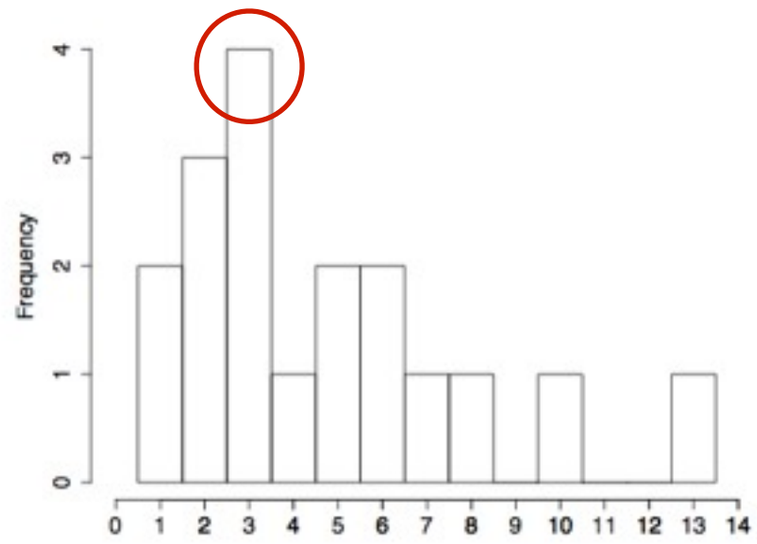


Mean and more...

- Consider the measured values: 15, 3, 9, 22, 3, 1, 8, 4
- Mode:
  - Most common value: 3
- Median:

1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 5, 5, 6, 6, 7, 8, 10, 13 is 3.

  - , 3, 3, 4, 8, 9, 15, 22
  - iers! (1,3,3,4,8,9,15,22,999)
-



  - = 8,125
  - 9 -> 118,2)

Distribution

- Mean and standard deviation / variance
- Normal distribution
- Skewedness

Data x	Deviations $x - \bar{x}$
10	10 - 48 = -38
15	15 - 48 = -33
18	18 - 48 = - 30
33	33 - 48 = -15
34	34 - 48 = -14
36	36 - 48 = -12
51	51 - 48 = 3
73	73 - 48 = 25
80	80 - 48 = 32
86	86 - 48 = 38
92	92 - 48 = 44
Sum = 528 $\sum x = 528$	Sum = 0 $\sum(x - \bar{x}) = 0$

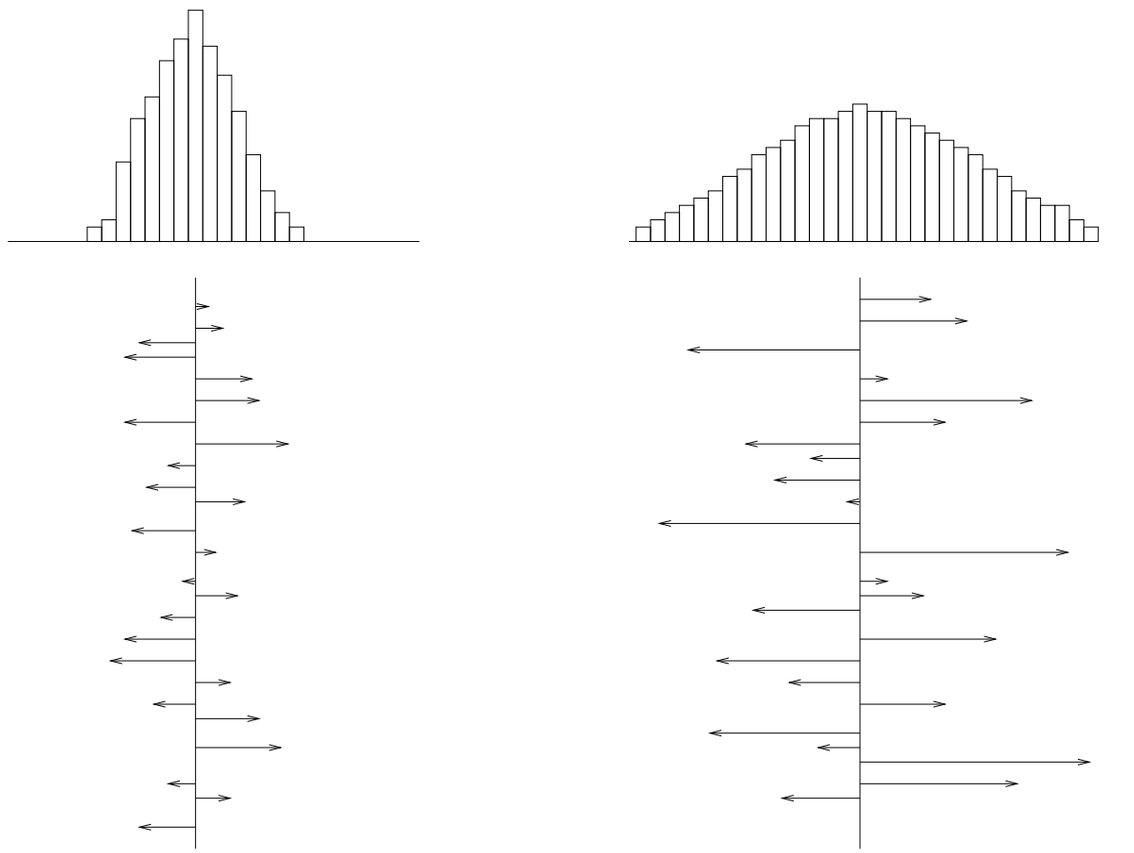
The mean is $\bar{x} = \frac{528}{11} = 48$

The Mean Deviation of a set of numbers is simply mean of deviations.

In practice, the mean deviation is *always* zero.

The Mean Deviation

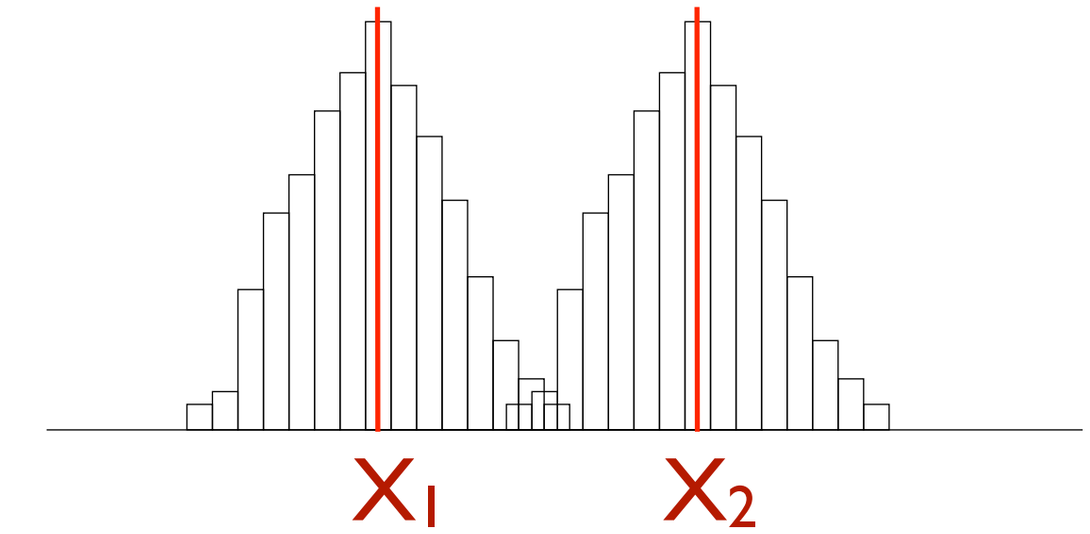
To measure the spread of a dataset it seems sensible to use the 'deviation' of each data point from the mean of the distribution. The deviation of each data point from the mean is simply the data point minus the mean.

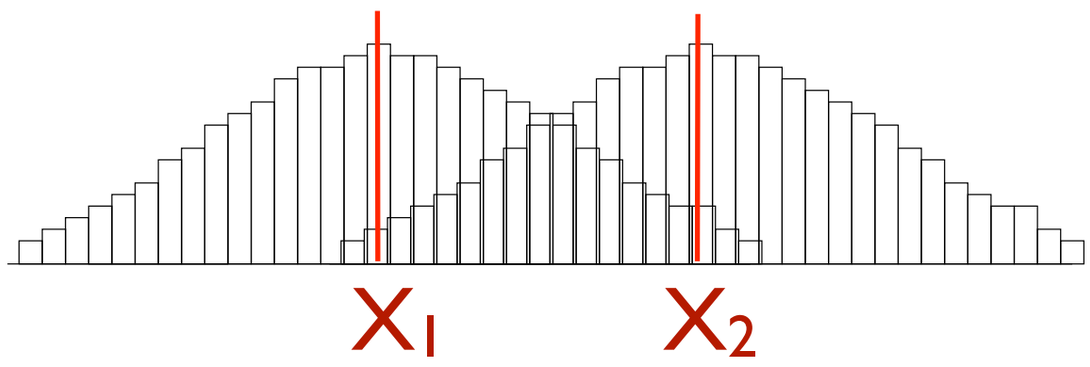


small spread = small deviations

large spread = large deviations

Means and distribution



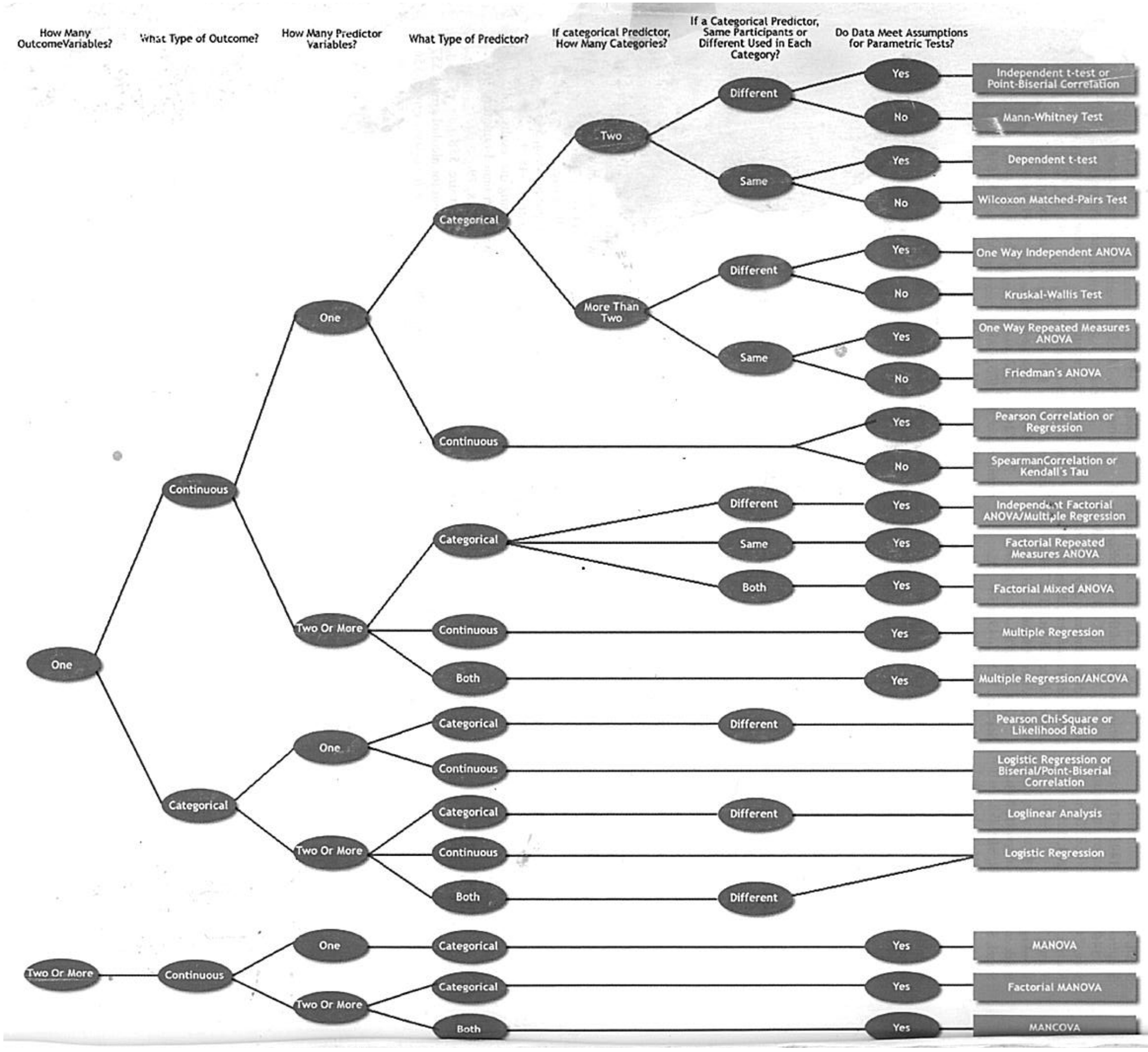


Significance or chance?

- Compare a value against chance level
  - Are male smokers significant more frequent than chance?
  - χ^2 , One-sample t-Test
- Compare two means
  - Treatment on smokers: medication group vs placebo group
  - Two-sample t-Test (also paired for repeated measures)
- Comparing several means, account for types of variance
  - ANOVA

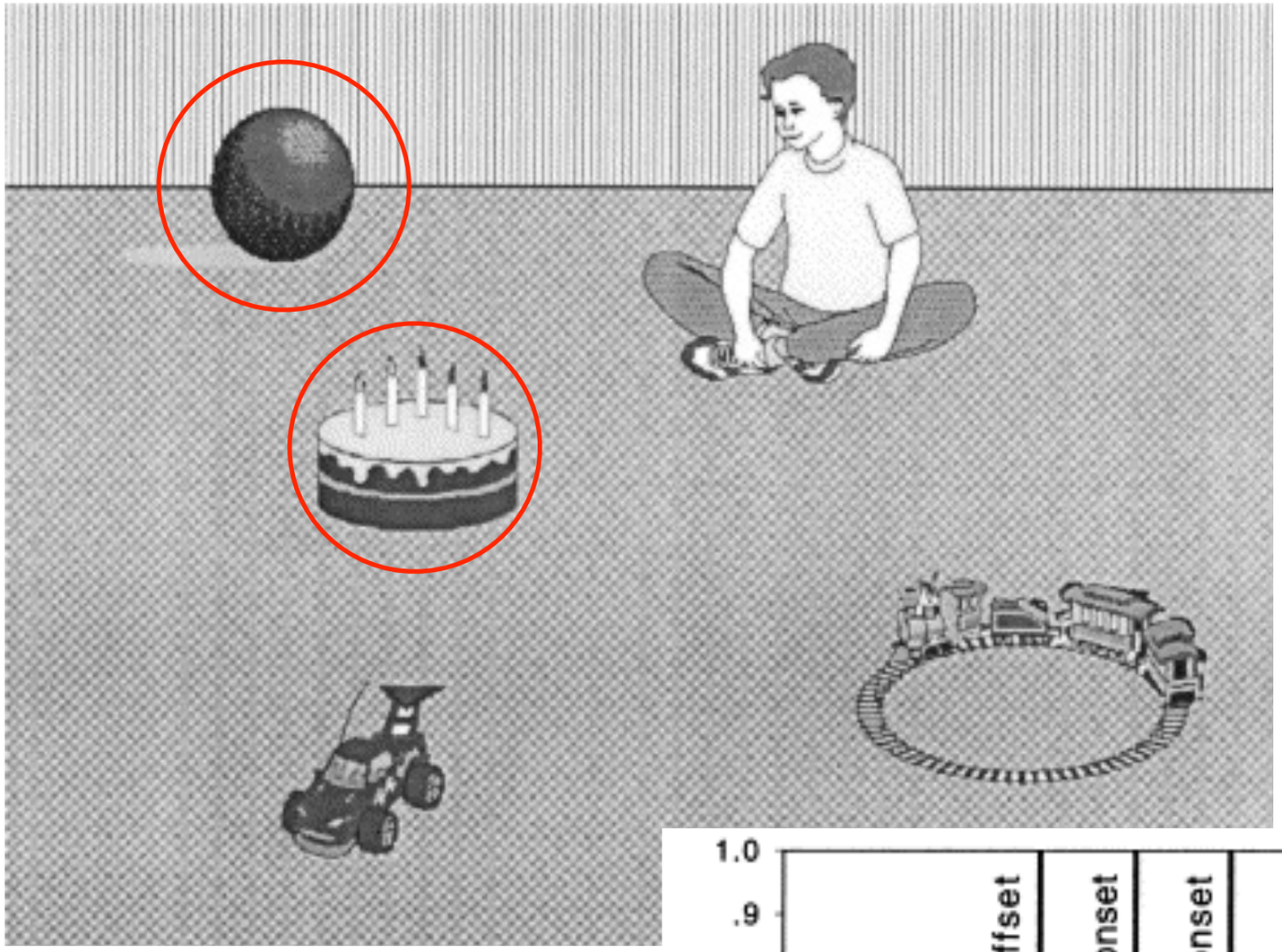
Tests

- Non-parametric tests
  - χ^2
  - Spearman / Wilcoxon rank
- Parametric Tests (assumption about distribution!)
  - ANOVA
  - Linear/Logistic Regression

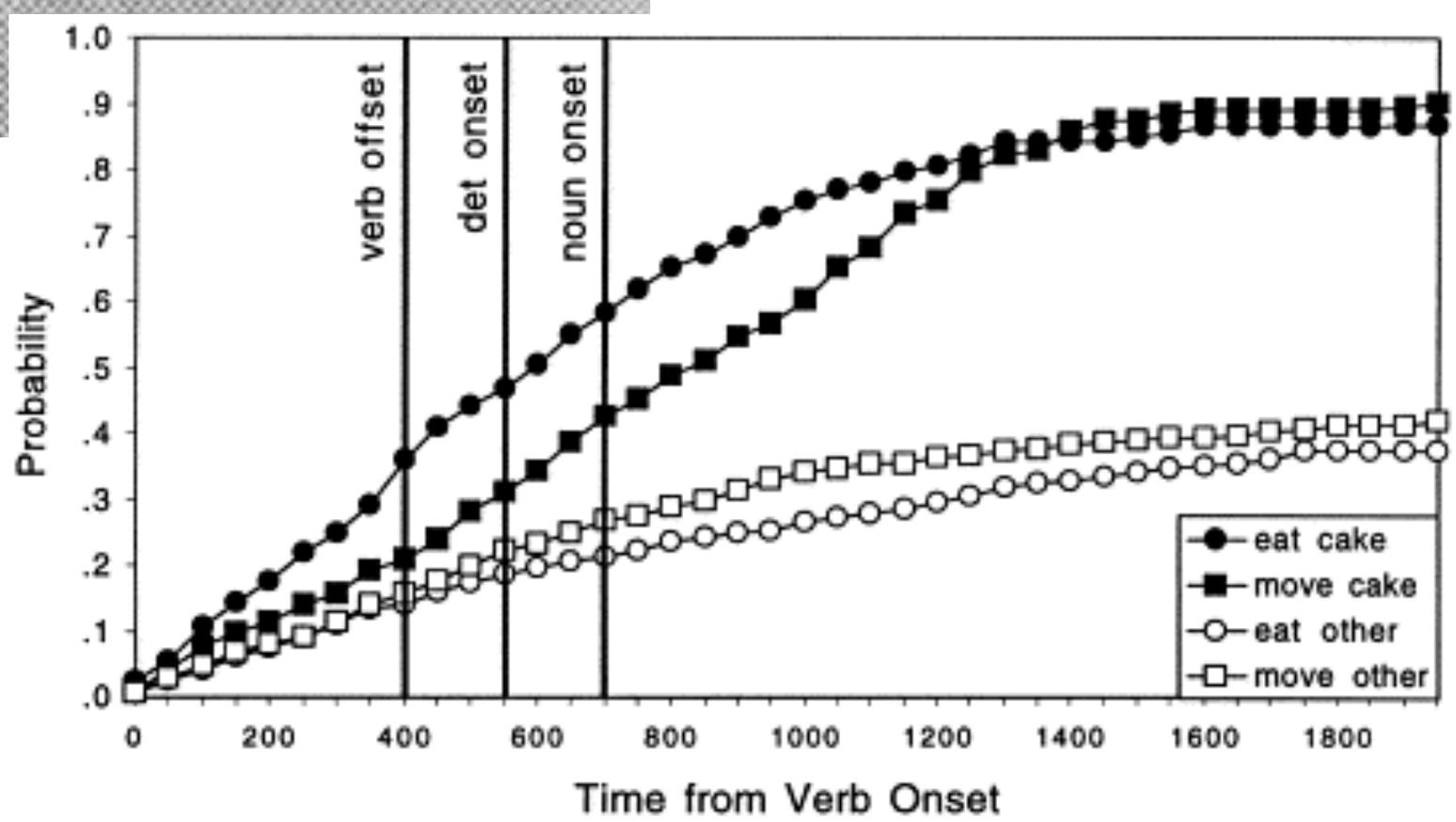

How Many OutcomeVariables?
What Type of Outcome?
How Many Predictor Variables?
What Type of Predictor?
If categorical Predictor, How Many Categories?
If a Categorical Predictor, Same Participants or Different Used in Each Category?
Do Data Meet Assumptions for Parametric Tests?
One
Continuous
Categorical
Two
More Than Two
Different
Same
Yes
No
Independent t-test or Point-Biserial CorreTation
Mann-Whitney Test
Dependent t-test
Wilcoxon Matched-Pairs Test
One Way Independent ANOVA
Kruskal-Wallis Test
One Way Repeated Measures ANOVA
Friedman's ANOVA
Continuous
Pearson Correlation or Regression
SpearmanCorrelation or Kendall's Tau
Two Or More
Categorical
Continuous
Both
Independent Factorial ANOVA/Multiple Regression
Factorial Repeated Measures ANOVA
Factorial Mixed ANOVA
Multiple Regression
Multiple Regression/ANCOVA
Categorical
One
Two Or More
Continuous
Both
Pearson Chi-Square or Likelihood Ratio
Logistic Regression or Biserial/Point-Biserial Correlation
Loglinear Analysis
Logistic Regression
Two Or More
Continuous
MANOVA
Factorial MANOVA
MANCOVA

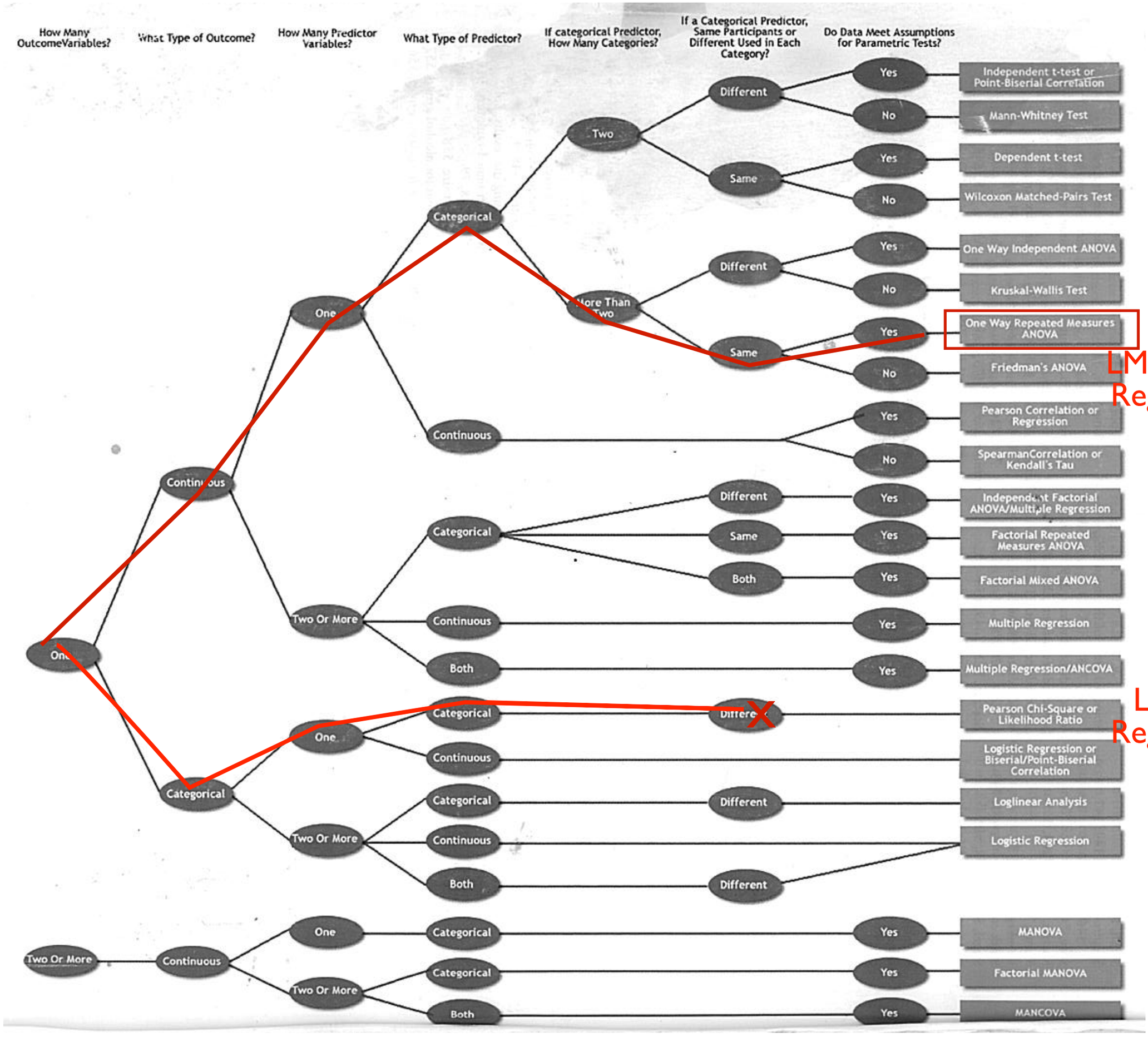
Gaze data

- Continuous or discrete?
  - Binning, Proportions
- Fixations, Saccades, blinks
  - Change vs duration
- Independent measures?
  - Within a trial or across trials, Repeated-Measures
- Random variables / subject- and item-variation
  - Average over one or the other; include as random variable
- Properties of the stimulus (saliency, complexity, size)
  - Counterbalancing



"The boy will eat/move the cake."





LME: Linear Regression

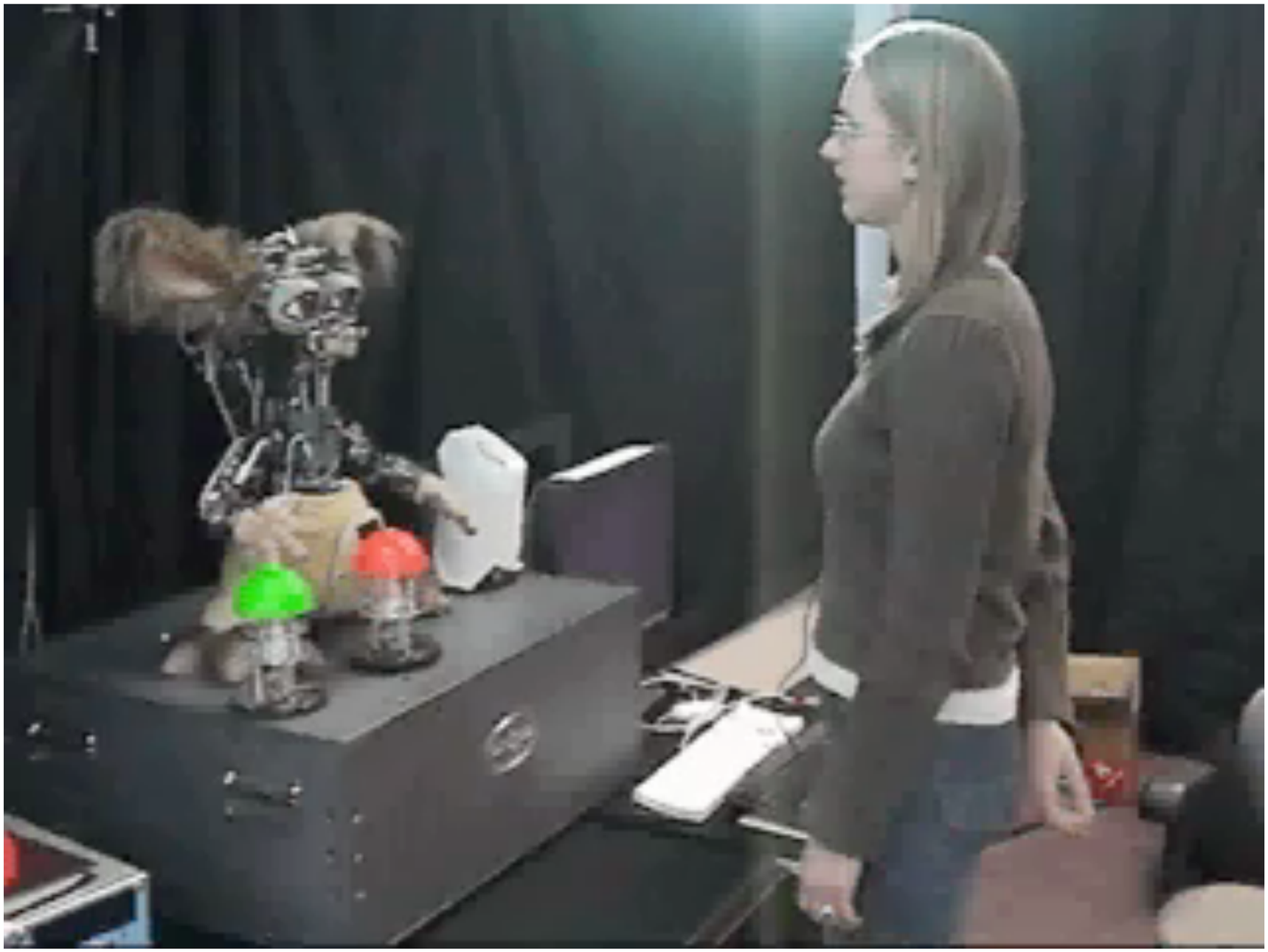
Logistic Regression

Control vs Naturalness

- "Trials" or "Interaction"
  - Do mechanisms persist under different tasks?
  - Do they persist under unusual circumstances?
    - Communicative acts in non-communicative setting, e.g. production & comprehension studied separately
    - Ecological validity
- Amount of data points
  - Qualitative vs quantitative argumentation

Interaction

Interaction



Interaction

- Peter Kiefer, ETH Zurich

Gaze data

- Continuous or discrete?
  - Binning, Proportions
- Fixations, Saccades, blinks
  - Change vs duration
- Independent measures?
  - Within a trial or across trials, Repeated-Measures
- Random variables / subject- and item-variation
  - Average over one or the other; include as random variable
- Properties of the stimulus (saliency, complexity, size)
  - Counterbalancing

DV: When and what is measured?

IV: Gaze as Input Variable?