Comparing Three or More Groups

Vanessa LoBue

Jamil Bhanji

with a little help from Andy Field

Topics

- Comparing several means with the linear model
  - Planned Contrasts/Comparisons
  - Post Hoc Tests
- ANCOVA
- Factorial Design

A Puppy Example

- A puppy therapy RCT in which we randomized people into three groups:
  1. A control group
  2. 15 minutes of puppy therapy
  3. 30 minutes of puppy therapy
- The outcome is happiness (0 = unhappy) to 10 (happy)

A Puppy Example

- Predictions:
  1. Any form of puppy therapy should be better than the control (i.e. higher happiness scores)

  2. A dose-response hypothesis that as exposure time increases (from 15 to 30 minutes) happiness will increase too.

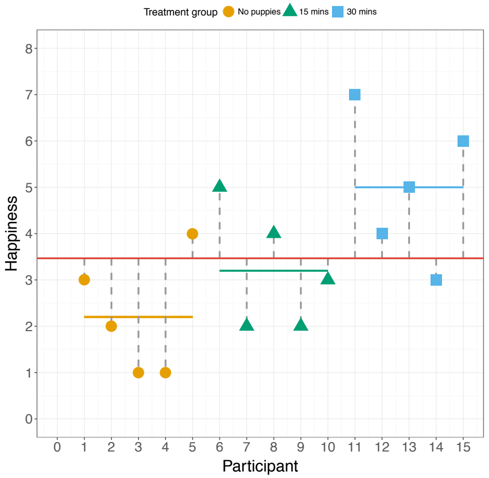
Theory of the *F*-statistic

- We calculate how much variability there is between scores
  - Total Sum of squares (SS_T).
- We then calculate how much of this variability can be explained by the model we fit to the data
  - How much variability is due to the predictor variable/experimental manipulation, Model Sum of Squares (SS_M)...
- … and how much cannot be explained
  - How much variability is due to individual differences in performance, Residual Sum of Squares (SS_R).

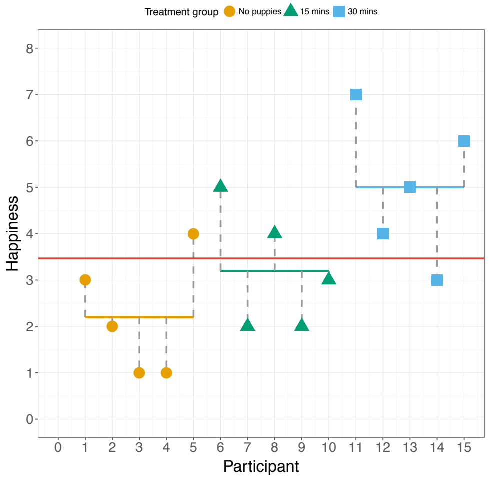
SS_T, SS_M, SS_R

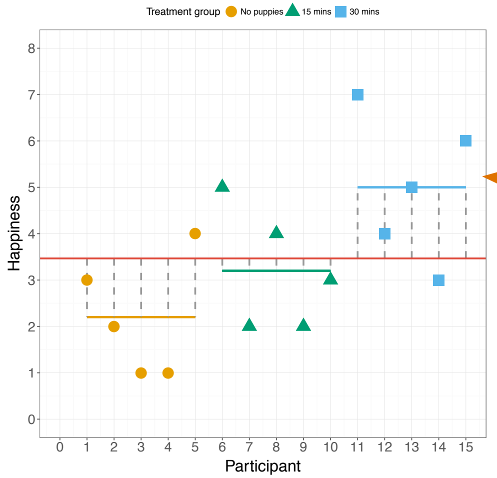
Does this look familiar? (Week 6 - regression, slide 14)

SS_T uses the differences between the observed data and the mean value of Y



SS_R uses the differences between the observed data and the model (group means)





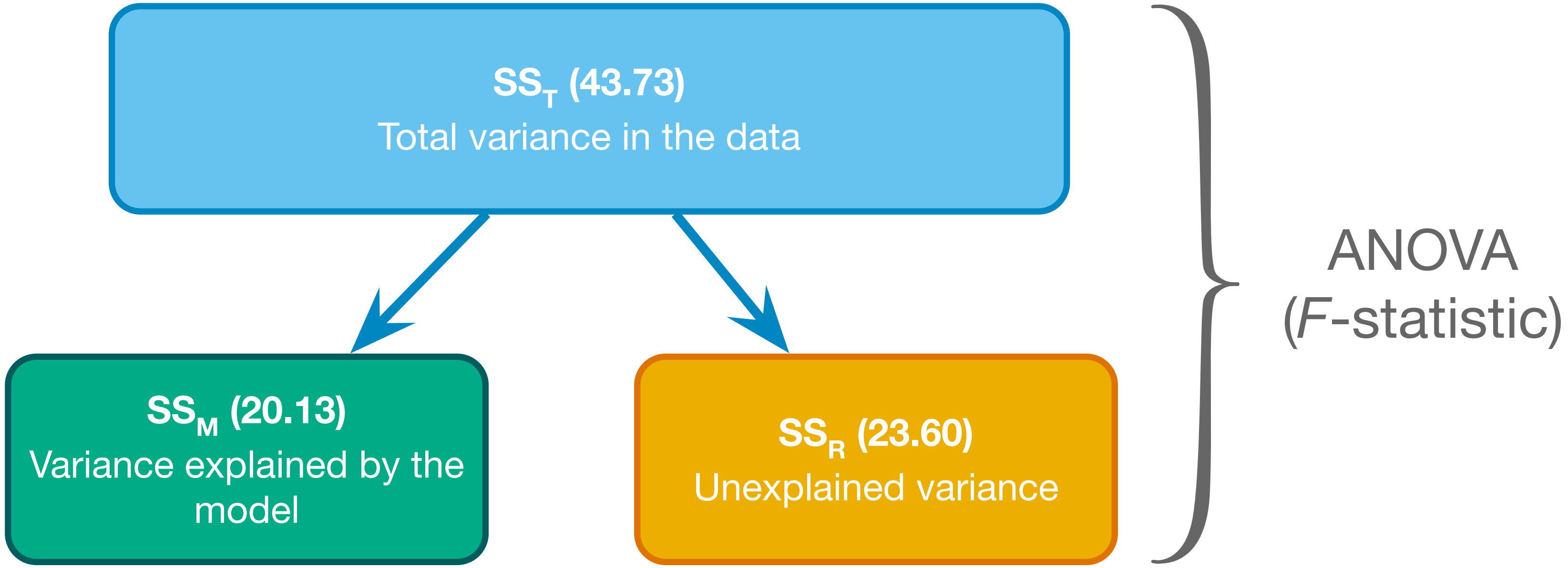
SS_M uses the differences between the mean value of Y and the model (group means)

Theory of the *F*-statistic

- We compare the amount of variability explained by the model (experiment), to the error in the model (individual differences)
  - This ratio is called the *F*-statistic

- If the model explains a lot more variability than it can’t explain, then the experimental manipulation has had a significant effect on the outcome (DV)

Theory of the *F*-statistic



- If the experiment is successful, then the model will explain more variance than it can’t
  - SS_M will be greater than SS_R

Follow-Up Tests

- The *F*-statistic tells us only that the experiment was successful
  - i.e. group means were different
- It does not tell us specifically which group means differ from which
- We need additional tests to find out where the group differences lie
  - planned contrasts
  - post-hoc tests

Follow-Up Tests

- Planned Contrasts
  - Hypothesis driven
  - Planned a priori

- Post Hoc Tests
  - Not Planned (no hypothesis)
  - Compare all pairs of means

Planned Contrasts

- Basic Idea:
  - The variability explained by the Model (experimental manipulation, $\mathbf{SS_M}$) is due to participants being assigned to different groups
  - This variability can be broken down further to test specific hypotheses about which groups might differ
  - We break down the variance according to hypotheses made *a priori* (before the experiment)

Planned Contrasts: Rules

- Independent
  - contrasts must not interfere with each other (they must test unique hypotheses)
- Only 2 Chunks
  - Each contrast should compare only 2 chunks of variation
  - *K*-1
  - You should always end up with one less contrast than the number of groups

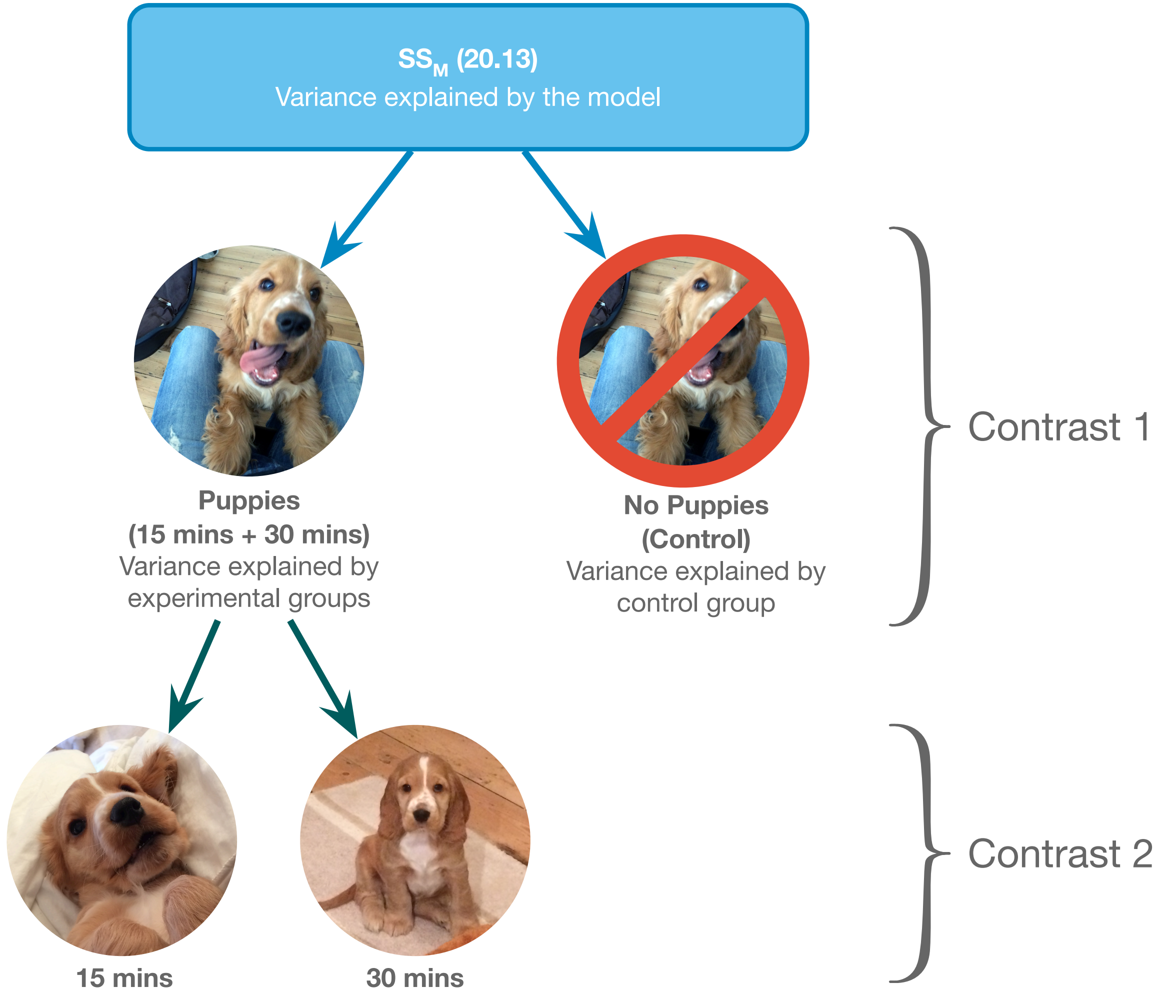
Choosing Contrasts

- Big Hint:
  - In most experiments we usually have one or more control groups
  - The logic of control groups dictates that we expect them to be different to groups that we've manipulated
  - The first contrast will always be to compare any control groups (chunk 1) with any experimental conditions (chunk 2)

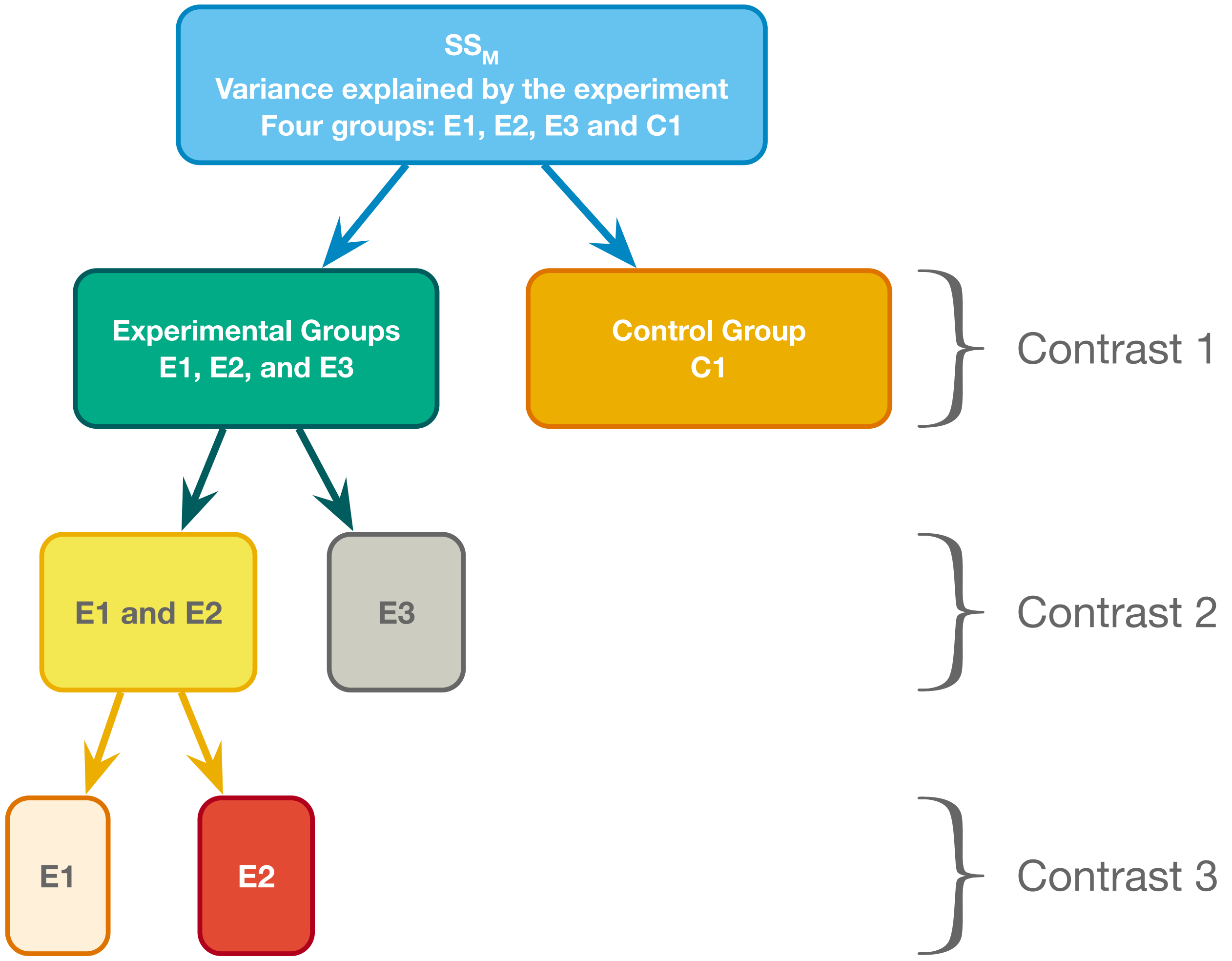
Choosing Contrasts

- Hypothesis 1:
  - People who receive puppy therapy will be happier than those who don't
  - Control $\neq$ (15 mins, 30 mins)
- Hypothesis 2:
  - People who receive a long dose of puppy therapy will be happier than those who receive a shortdose
  - 15 mins $\neq$ 30 mins

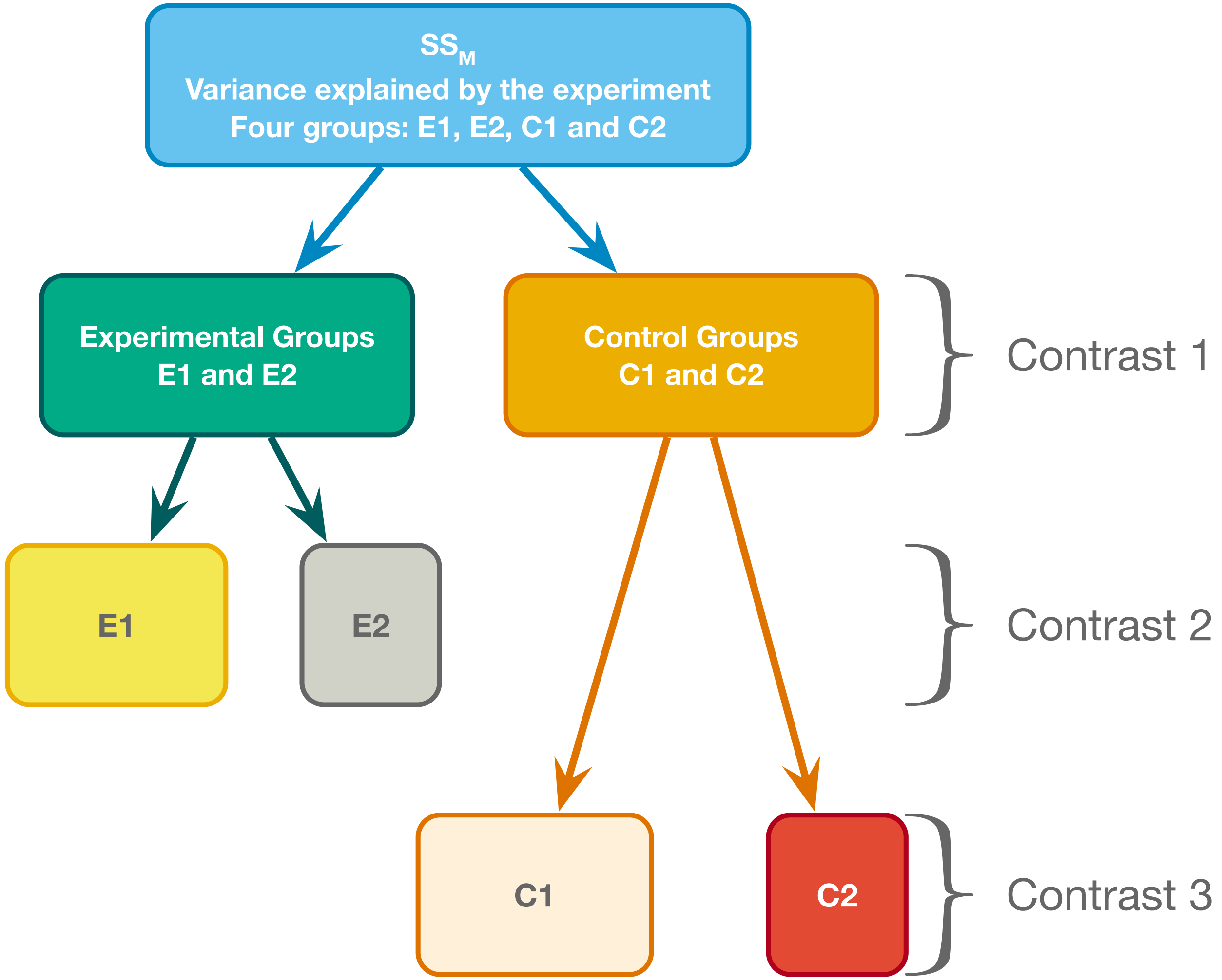
Choosing Contrasts: Example 1



Choosing Contrasts: Example 2



Choosing Contrasts: Example 3



Post hoc Tests

- Compare each mean against all others
- In general terms they use a stricter criterion to accept an effect as significant
  - Hence, control the familywise error rate.
  - Simplest example is the Bonferroni method:

$$P_{Crit} = \frac{\alpha}{k}$$

Post Hoc Tests Recommendations:

- Assumptions met:
  - REGWQ or Tukey HSD
- Safe Option:
  - Bonferroni
- Unequal Sample Sizes:
  - Gabriel’s (small *n*), Hochberg’s GT2 (large *n*)
- Unequal Variances:
  - Games-Howell

ANCOVA

- To test for differences between group means when we know that an extraneous variable affects the outcome variable

- Used to adjust for known extraneous variables

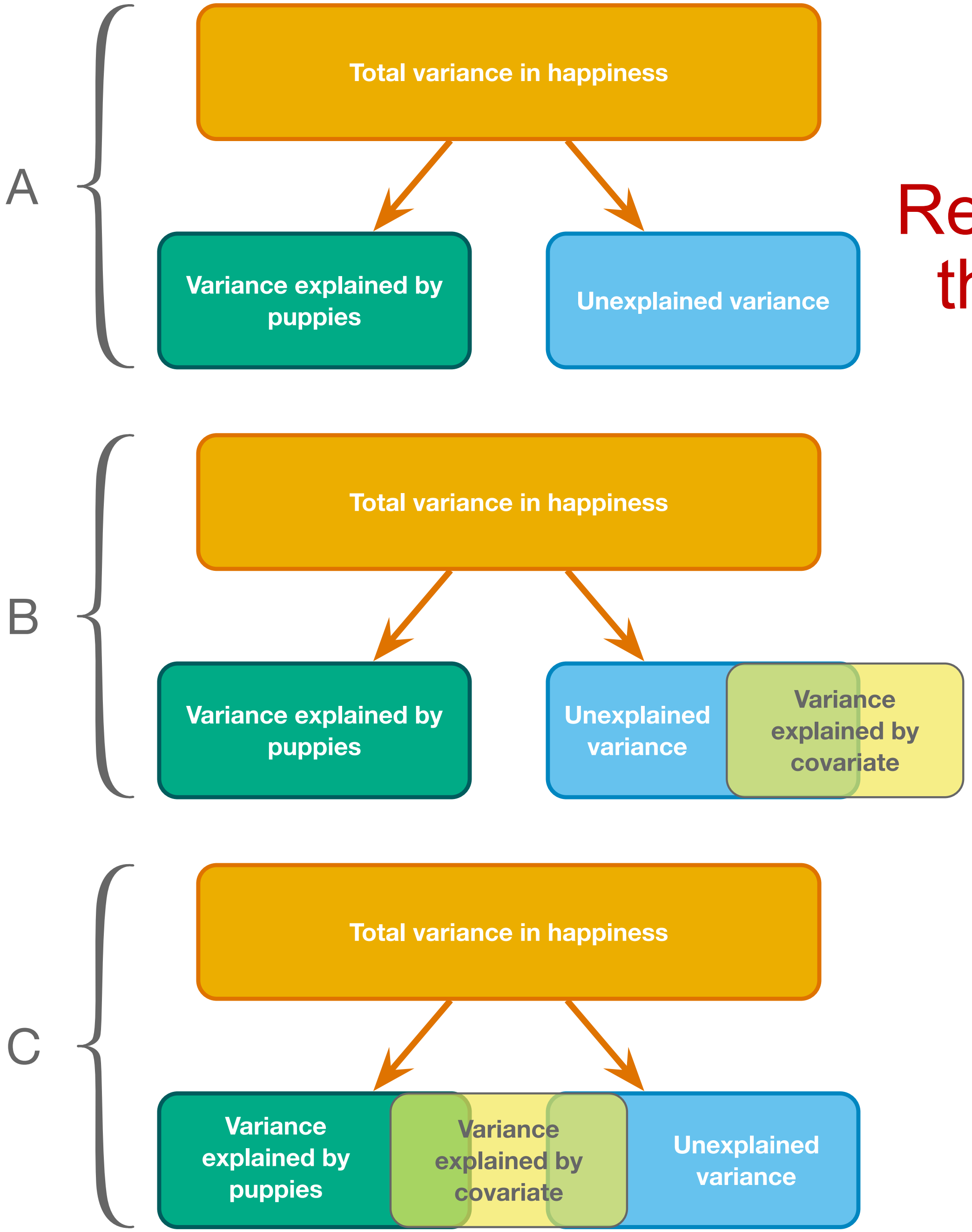
Advantages of ANCOVA

- Reduces Error Variance
  - By explaining some of the unexplained variance (SS_R) the error variance in the model can be reduced
- Greater Experimental Control:
  - By controlling known extraneous variables, we gain greater insight into the effect of the predictor variable(s)

Back to Puppies

- We can replicate the RCT of puppy therapy but also measure love of puppies.
  - Outcome (or DV) = participant's happiness
  - Predictor (or IV) = Dose of puppy therapy (control, 15-mins & 30-mins)
  - Covariate = love of puppies

Relationships between the IV and covariate



Factorial Designs

- Rationale of factorial designs
- Partitioning variance
- Interaction effects
  - Interaction graphs
  - Interpretation

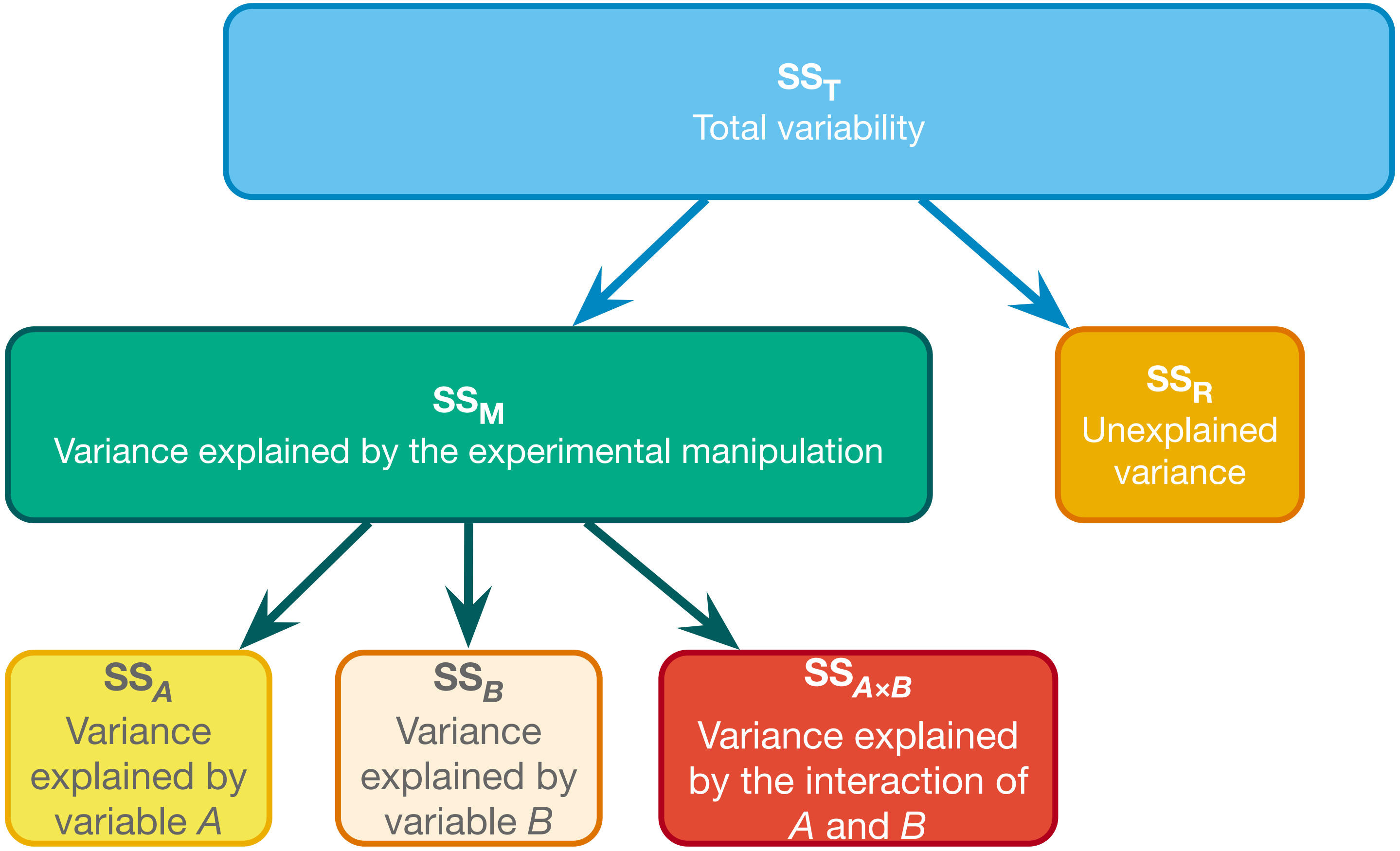
Naming Experimental Designs

- Factorial design
  - More than one independent/predictor variable has been manipulated
- The number of independent/predictor variables manipulated
  - *n*-way = *n* predictors/independent variables, for example:
    - Two-way = 2 independent variables
    - Three-way = 3 independent variables
- The allocation of participants
  - Independent design = different entities in all conditions
  - Repeated measures design = the same entities in all conditions
  - Mixed design = different entities in all conditions of at least one IV, the same entities in all conditions of at least one other IV

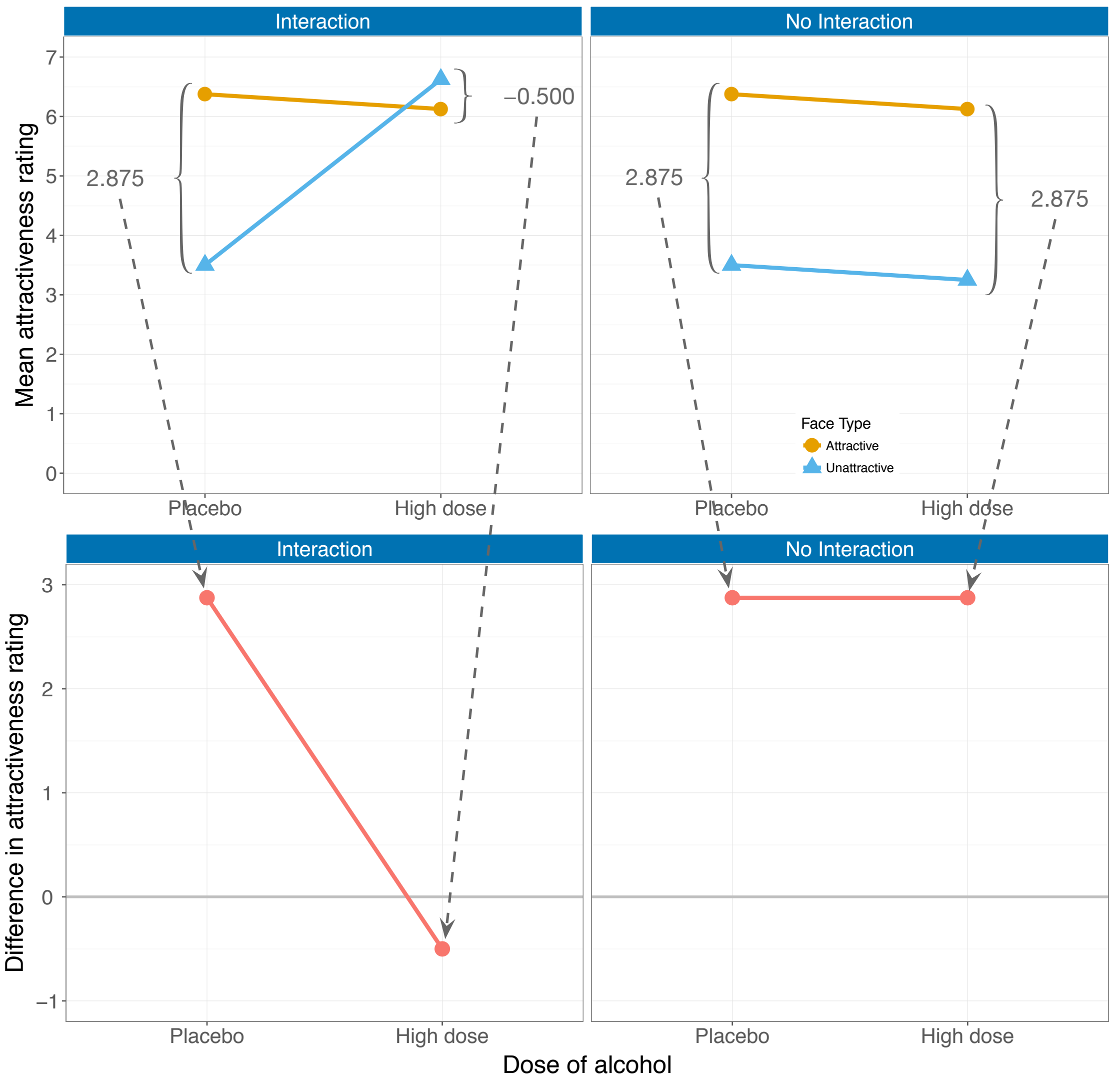
Benefit of Factorial Designs

- We can look at how variables *Interact*.
- Interactions
  - Show how the effects of one predictor might depend on the effects of another
  - Are often more interesting than main effects
  - Think back to ***moderation***.

Partitioning Variance



Interaction Graphs

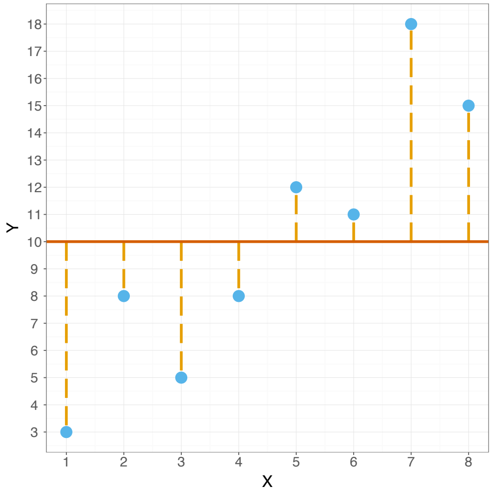


On to the lab activity!
the following slides are extra
(for further clarification if needed)

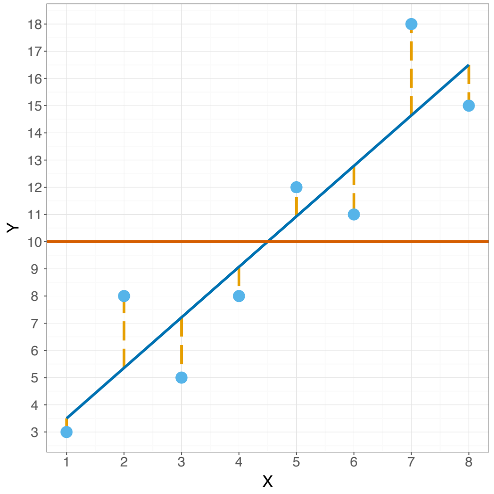
Regression: SS_T, SS_M, SS_R

SS_T is in the ”Total” row in SPSS output

SS_T uses the differences between the observed data and the mean value of *Y*

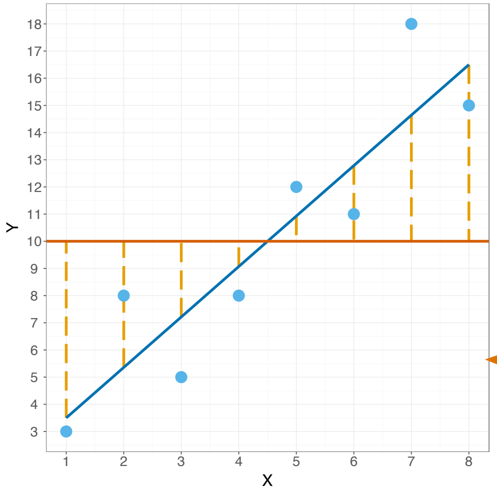


SS_R uses the differences between the observed data and the model



SS_R is in the ”Residual” row in SPSS output

SS_M is in the ”Regression” row in SPSS output



SS_M uses the differences between the mean value of *Y* and the model

Sums of squares

Regression Summary

- SS_T
  - Total variability (variability between scores and the mean).
- SS_R
  - Residual/error variability (variability between the model and the actual data).
- SS_M
  - Model variability (difference in variability between the model and the mean).

Using the regression menus

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20.133	2	10.067	5.119	.025[b]
	Residual	23.600	12	1.967		
	Total	43.733	14			

a. Dependent Variable: Happiness (0-10)

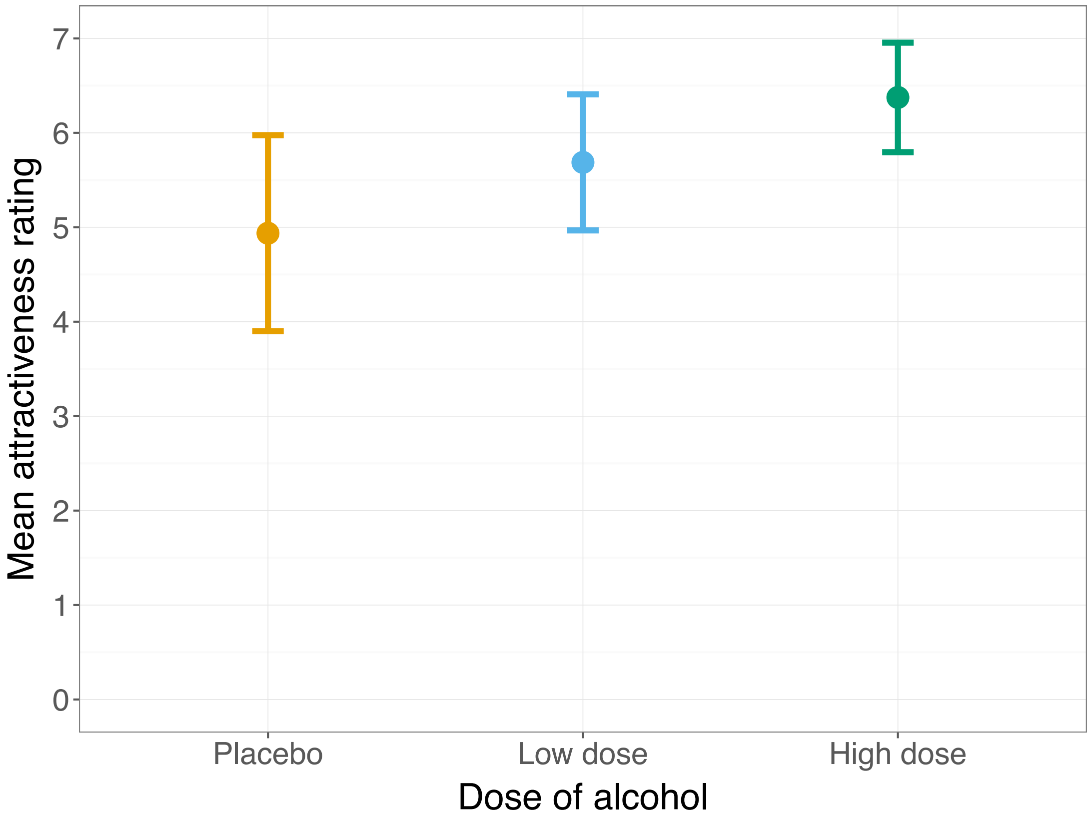
b. Predictors: (Constant), Dummy 2: 15 mins vs. Control, Dummy 1: 30 mins vs. Control

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.200	.627		3.508	.004
	Dummy 1: 30 mins vs. Control	2.800	.887	.773	3.157	.008
	Dummy 2: 15 mins vs. Control	1.000	.887	.276	1.127	.282

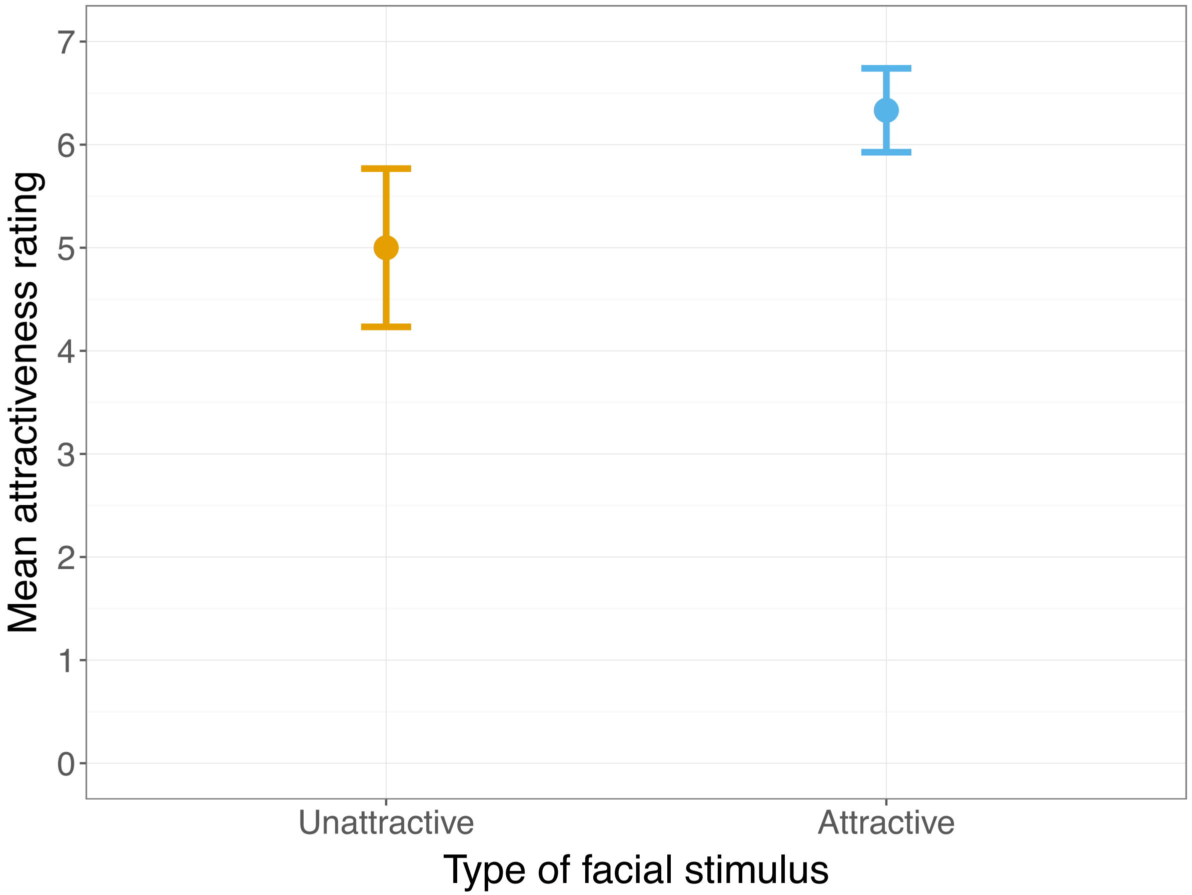
a. Dependent Variable: Happiness (0-10)

Main effect of alcohol



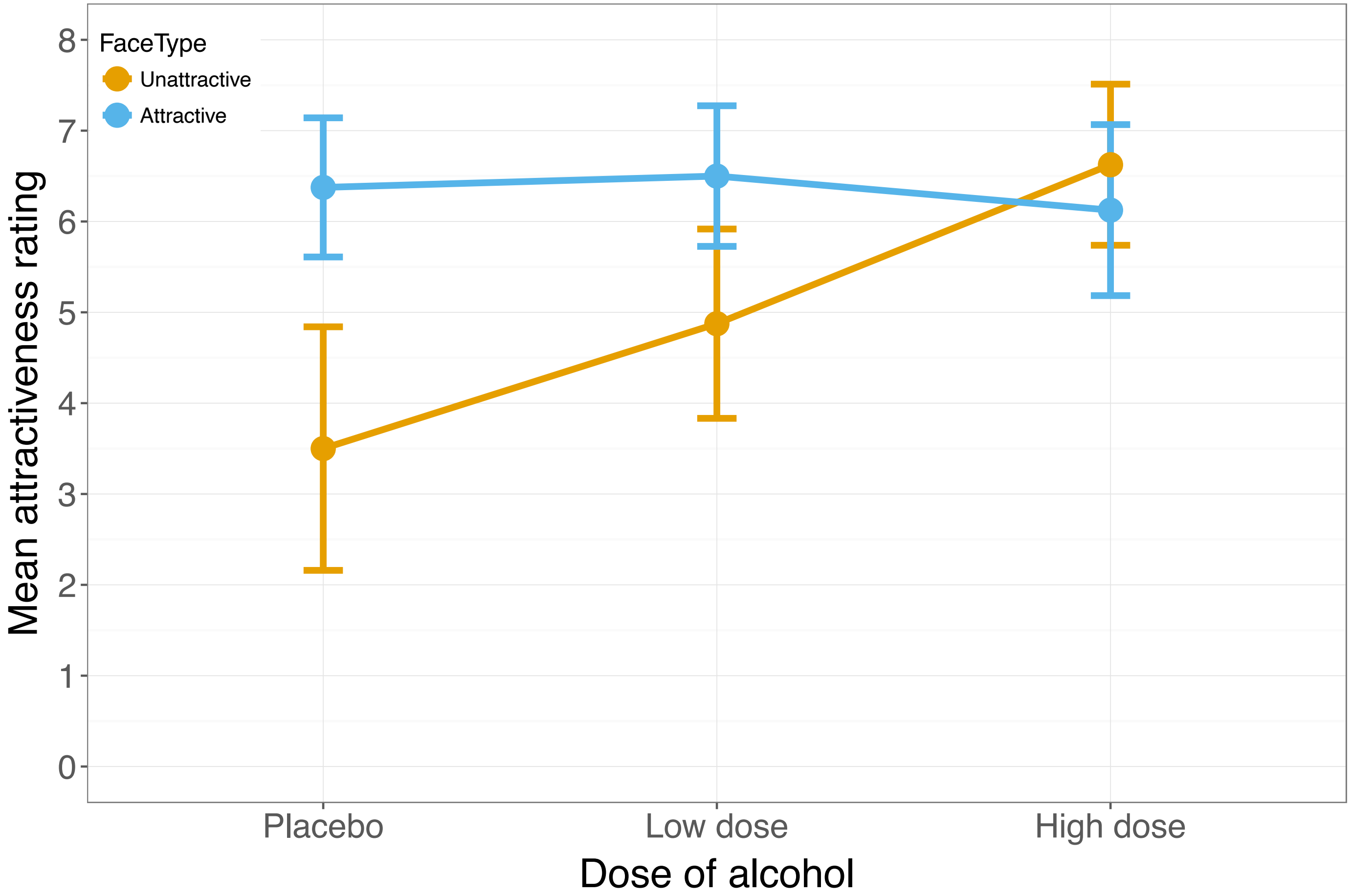
There was a significant main effect of the amount of alcohol consumed on ratings of the attractiveness of faces, $F(2, 42) = 6.04$, $p = 0.005$

Main effect of face type



Attractive faces were rated significantly higher than unattractive faces, $F(1, 42) = 15.58$, $p < 0.001$

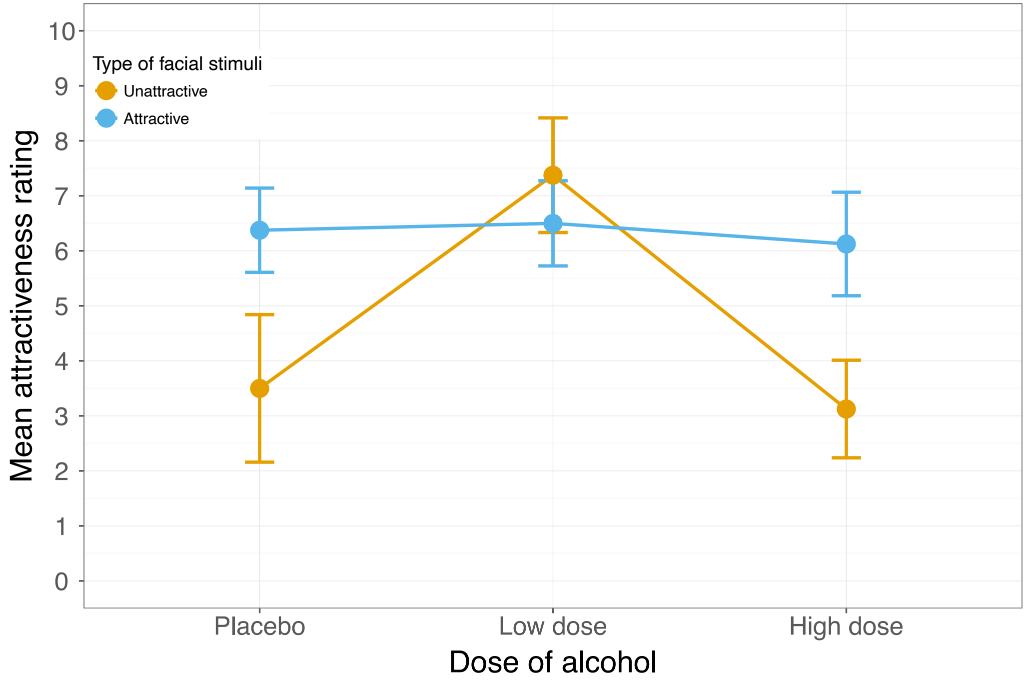
The interaction effect



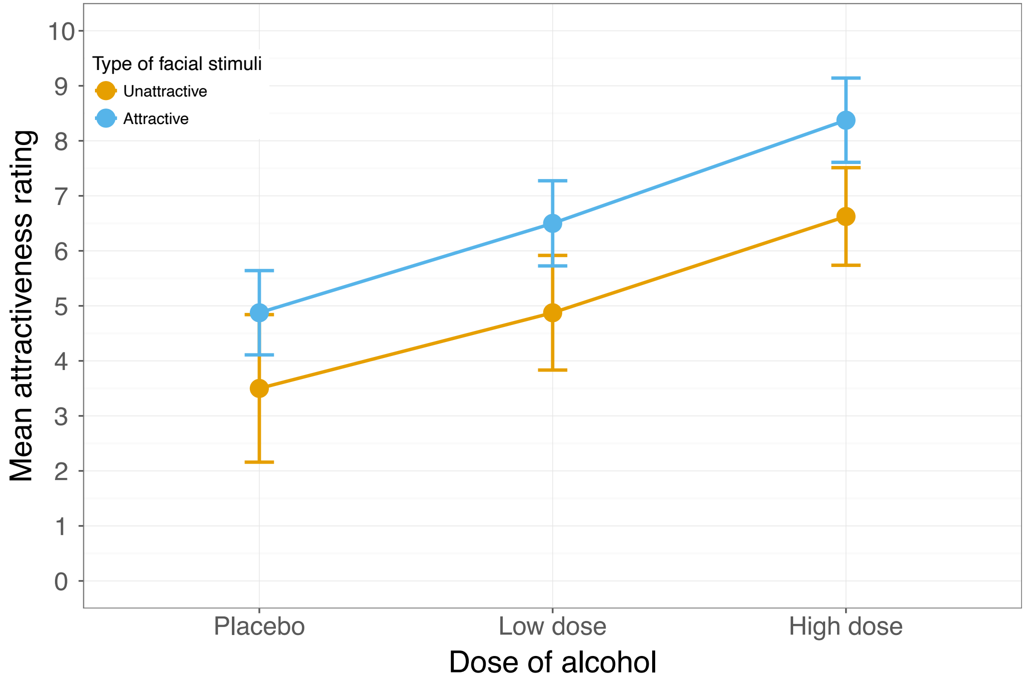
There was a significant interaction between the amount of alcohol consumed and the type of face of the person rated on attractiveness, $F(2, 42) = 8.51$, $p = 0.001$

Is there likely to be a significant interaction effect?

Yes



No



Is there likely to be a significant interaction effect?