DATA ANALYSIS

Week 13: Additional predictors

logistics

- PS6 revisions due TODAY
- PS7 opt-out deadline Apr 23
- PS7 due Apr 30
- class participation:
  - Canvas discussion board posts due Apr 30
  - “practice” questions (10 multiple-choice/true-false) due Apr 24
- LAST DAY to submit any late work: May 13

Week 13: Additional Predictors

- Opt-out of Problem Sets (Deadline 3: After Midterm 2)
  Apr 23 | 1 pts
- Problem Set 7: First Attempt
  Apr 30 | 2.5 pts
- Problem Set 7: Second Attempt
  May 8 | 2.5 pts

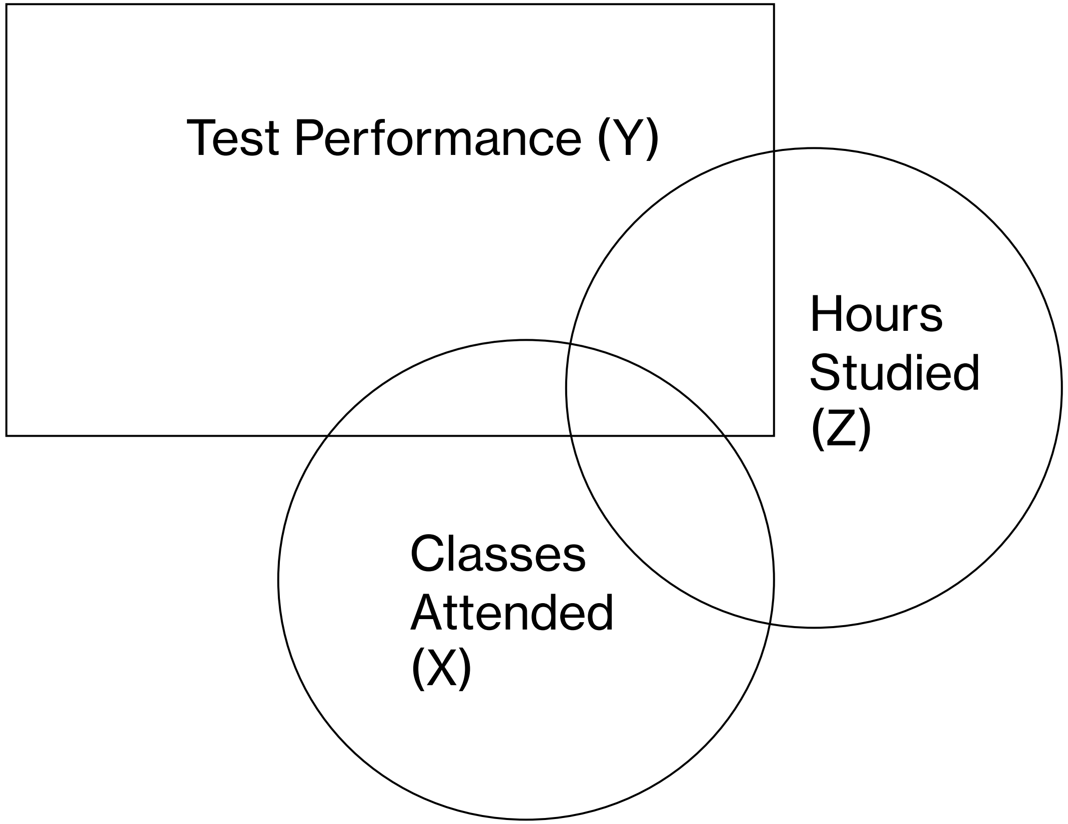
Class Participation

- Data Around Us!
  Apr 30 | 5 pts
- Meme Submission
  1 pts
- Student Practice Questions
  Apr 24 | 2.5 pts

12	F: April 12, 2024	**Exam (Midterm) 2**
13	W: April 17, 2024	W13: Additional Predictors
13	F: April 19, 2024	W13 continued…
14	T: April 23, 2024	**Problem Set Opt-out Deadline 3**
14	W: April 24, 2024	W14: Non-Independent/Miscellaneous Data
14	F: April 26, 2024	W14 continued…
15	T: April 30, 2024	**Problem Set 7 due**
15	W: May 1, 2024	W15: Odds and Ends
15	F: May 3, 2024	**Final Exam**
16	W: May 8, 2024	**Wrapping Up!**

additional predictors = complex models

- often, outcomes/dependent variables depend on not just one IV, but several IVs
- in such situations, modeling the variation in our dependent variable using only one variable leads to an impoverished model: we could do better by examining multiple variables
- data = model + error
  - *one IV*: $Y = a + bX + error$
  - *multiple IVs*: $Y = a + b_1X_1 + b_2X_2 + ... + error$

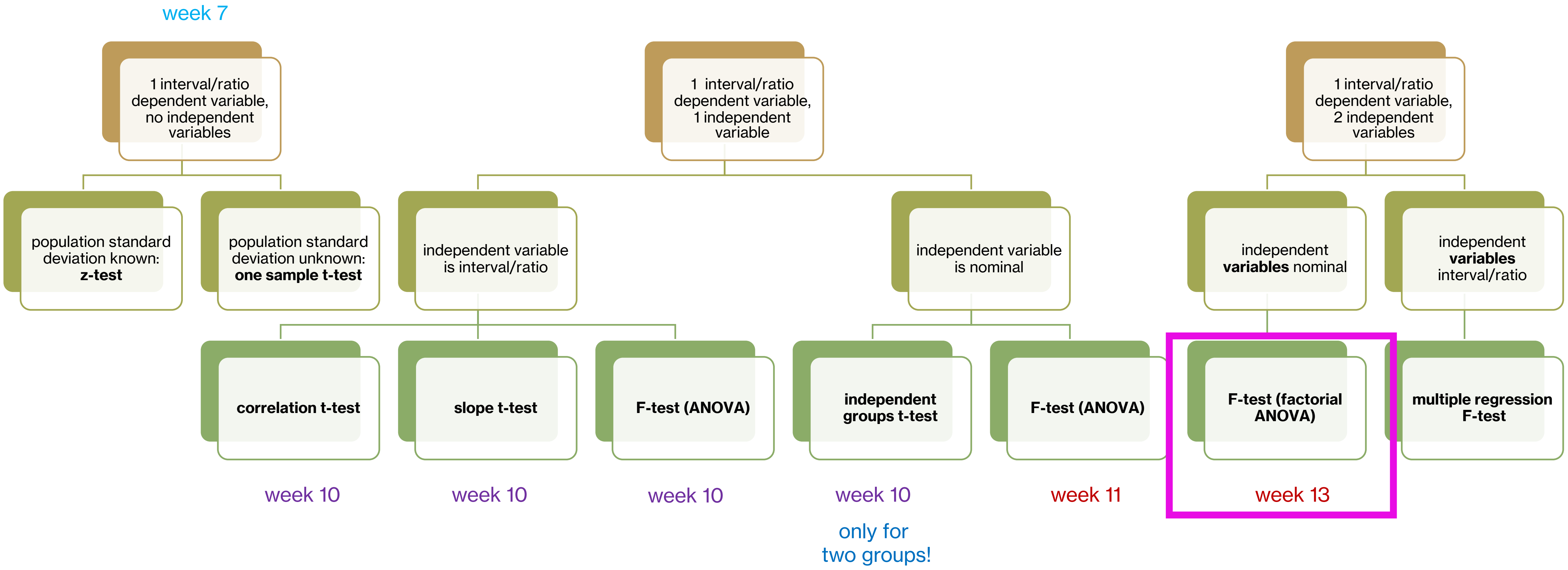


complex models: data types

- for a one DV and one IV situation, we saw how the data could come in different forms
- when more than one IV is involved, several permutations and combinations are possible
  - one DV ~ interval/ratio IV_1 + interval/ratio IV_2
  - one DV ~ interval/ratio IV_1 + nominal IV_2
  - one DV ~ nominal IV_1 + interval/ratio IV_2
  - one DV ~ nominal IV_1 + nominal IV_2
- no fear...general linear models are here!

	one independent variable		
dependent variable	nominal	ordinal	interval/ ratio
nominal			
ordinal			
interval/ratio	F / anova		t / F

hypothesis chart



the tooth growth dataset

- this in-built R dataset contains the "length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid"
- think about the design of this experiment
  - dependent variable?
  - independent variable(s) and their levels?
  - broad research question?

factorial designs

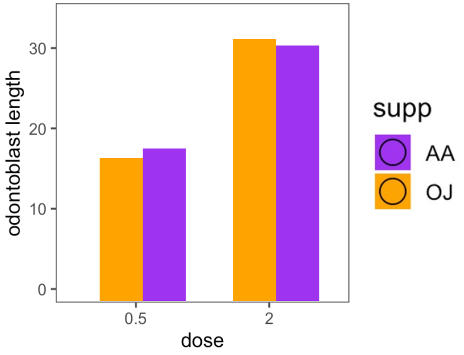
- factorial designs refer to situations where more than one independent variable or "factor" is manipulated in the same experiment (nominal IVs)
- common terminology
    - 2 x 2 factorial design, i.e., two independent variables (number of x's + 1), and each of them had 2 levels
    - 3 x 2 factorial design, i.e., 2 independent variables, one of them had 3 levels, and another had 2 levels
    - 3 x 5 x 4 x 6 factorial design, i.e., you are crazy
- what about our tooth decay design?
    - technically a 3 (dose: 0.5/1/2) x 2 (delivery: OJ, AA) design
    - we will examine a subset of this data that is 2 x 2
    - PS 7 has a problem with a 3 x 2 design! (arousal x task difficulty)

tooth growth dataset: visualization

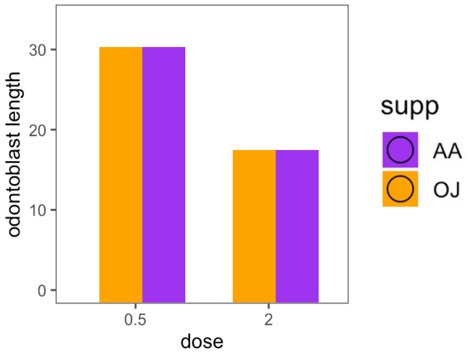
- let’s try to visualize the pattern of **tooth growth** as a function of **dose** and **supplements**
    - **dose**: 0.5 mg and 2 mg
    - **supplements**: OJ and AA
- sketch a possible bar graph of tooth growth based on the research question: is tooth growth impacted by dosage and delivery method of vitamin C?
    - **dose** on x axis
    - **tooth growth** on y axis
    - **supplement** by color

tooth growth dataset: visualization

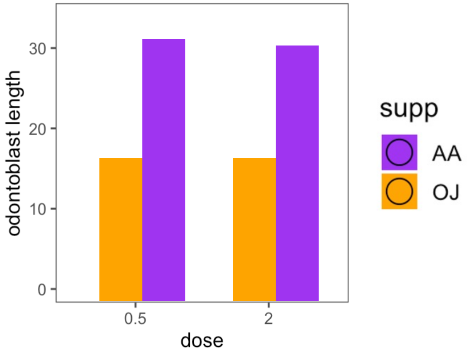
dose matters
supplement does not matter



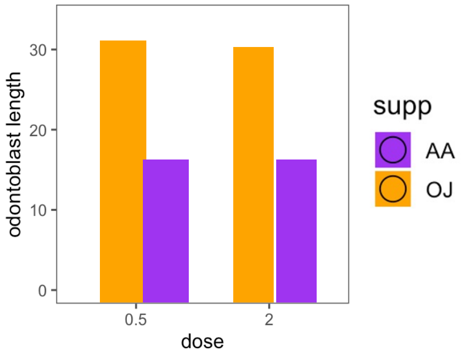
OR



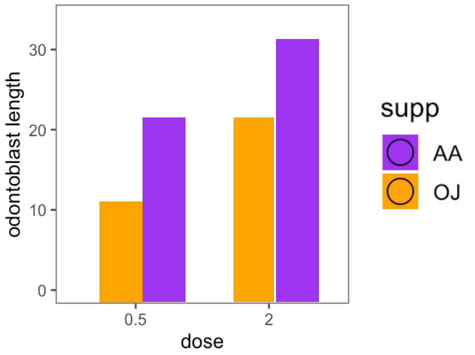
dose does not matter
supplement matters



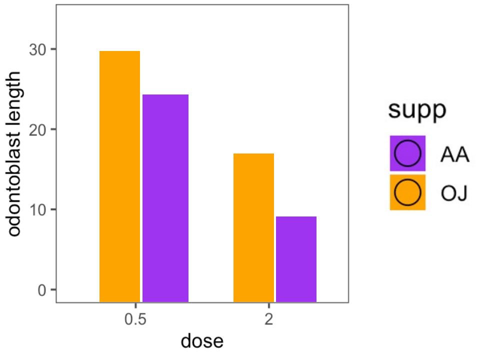
OR



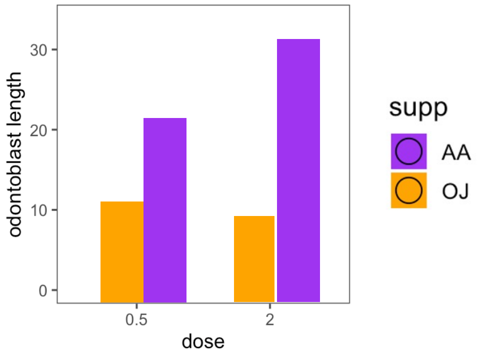
dose matters
supplement matters
dose and supplement
do not influence each other



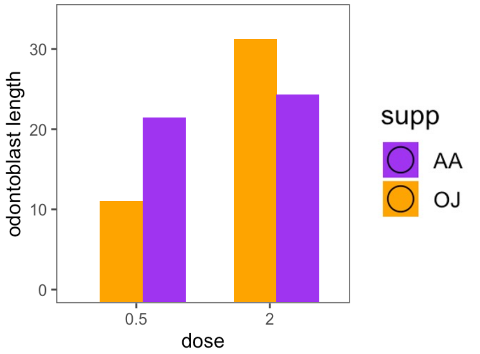
OR



dose matters
supplement matters
dose and supplement
influence each other

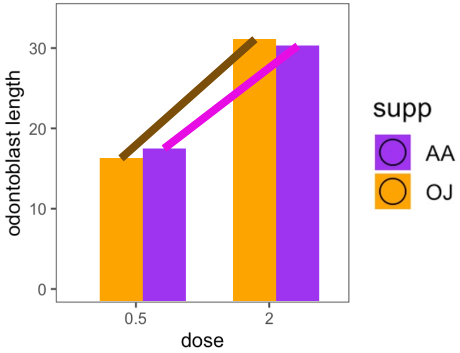


OR

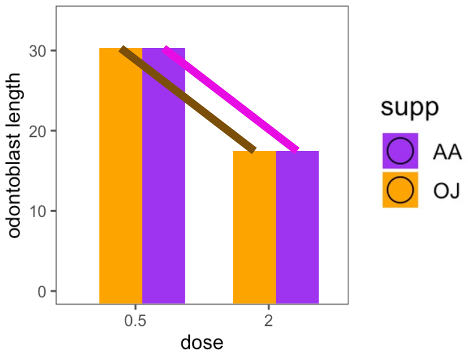


tooth growth dataset: visualization

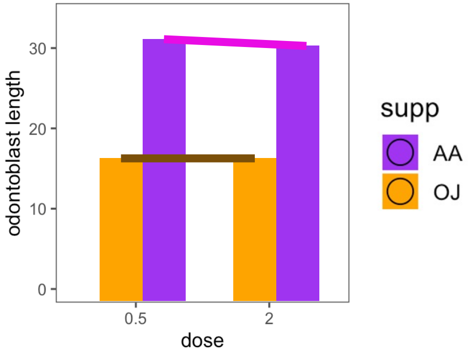
dose matters
supplement does not matter



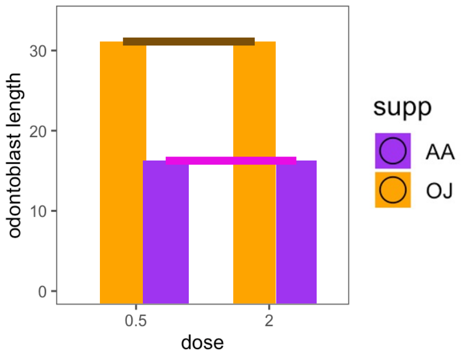
OR



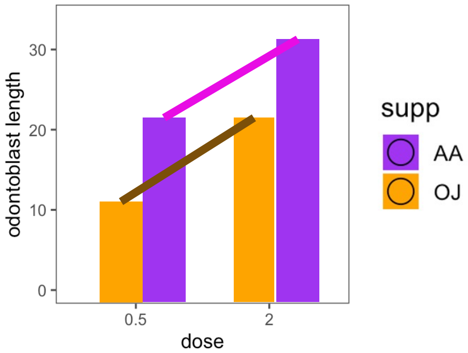
dose does not matter
supplement matters



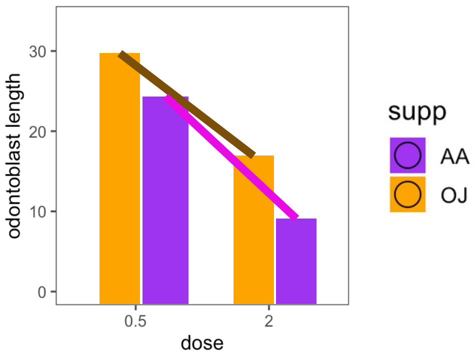
OR



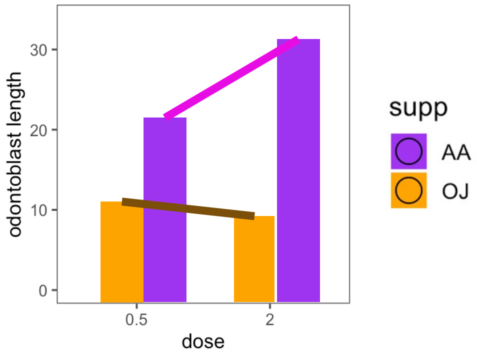
dose matters
supplement matters
dose and supplement
do not influence each other



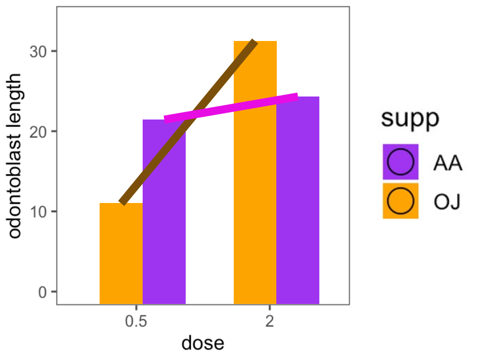
OR



dose matters
supplement matters
dose and supplement
influence each other

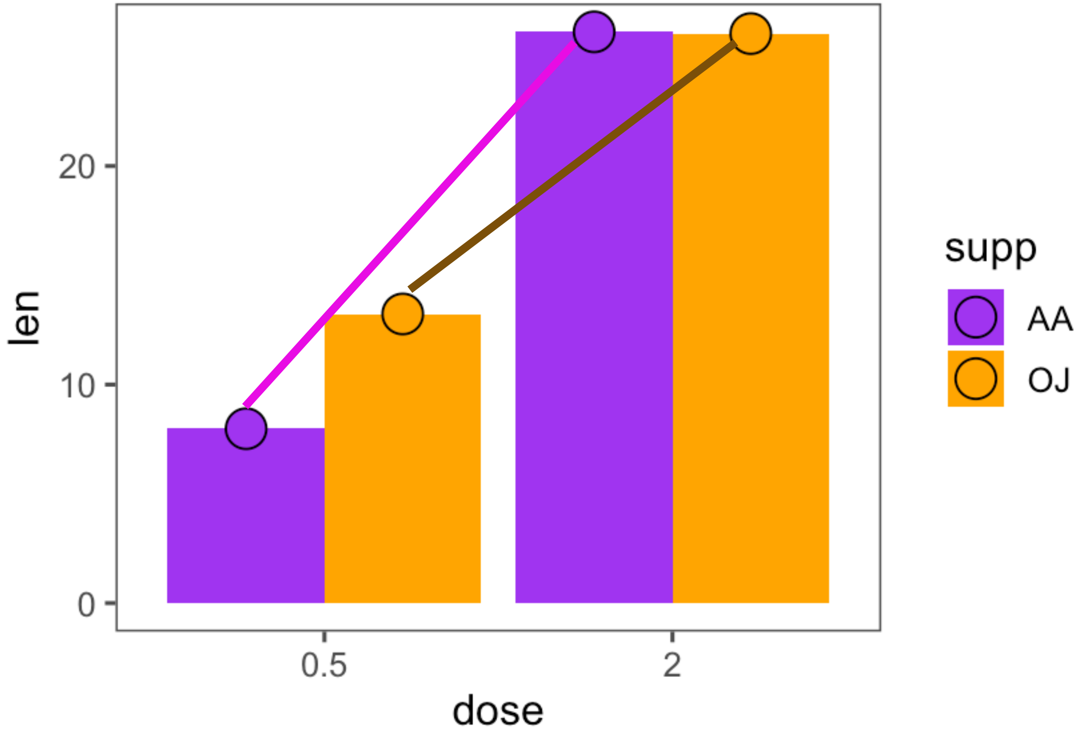


OR



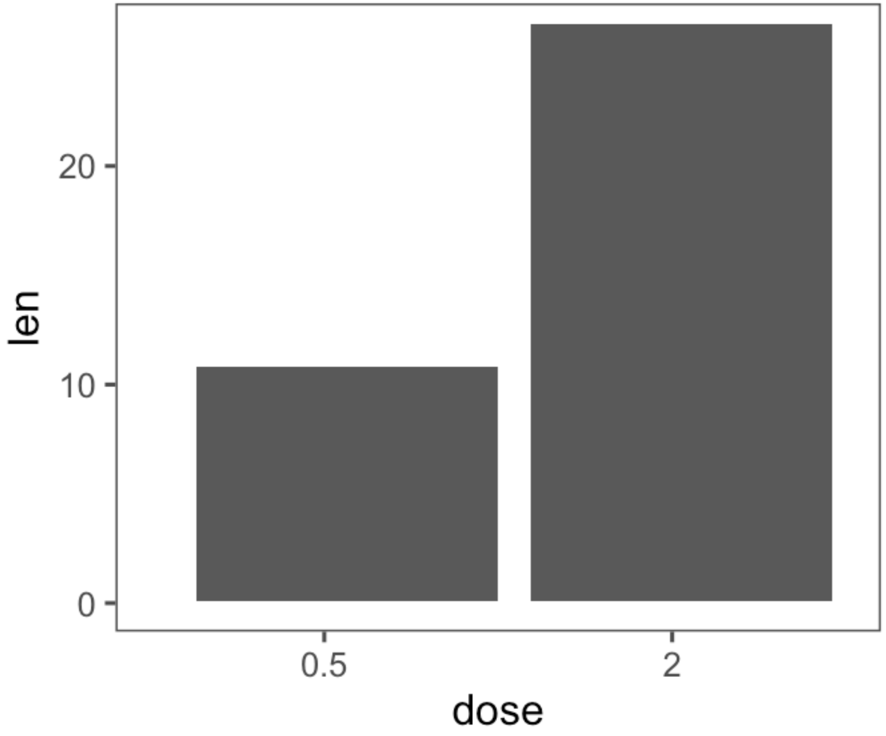
tooth growth dataset: actual pattern

- **dose** matters (0.5 mg << 2 mg)
- **supplement** matters (OJ > AA slightly)
- **dose** and **supplement** <u>influence each other</u>
  - at 0.5 mg, delivery method matters
  - at 2 mg, delivery method stops mattering



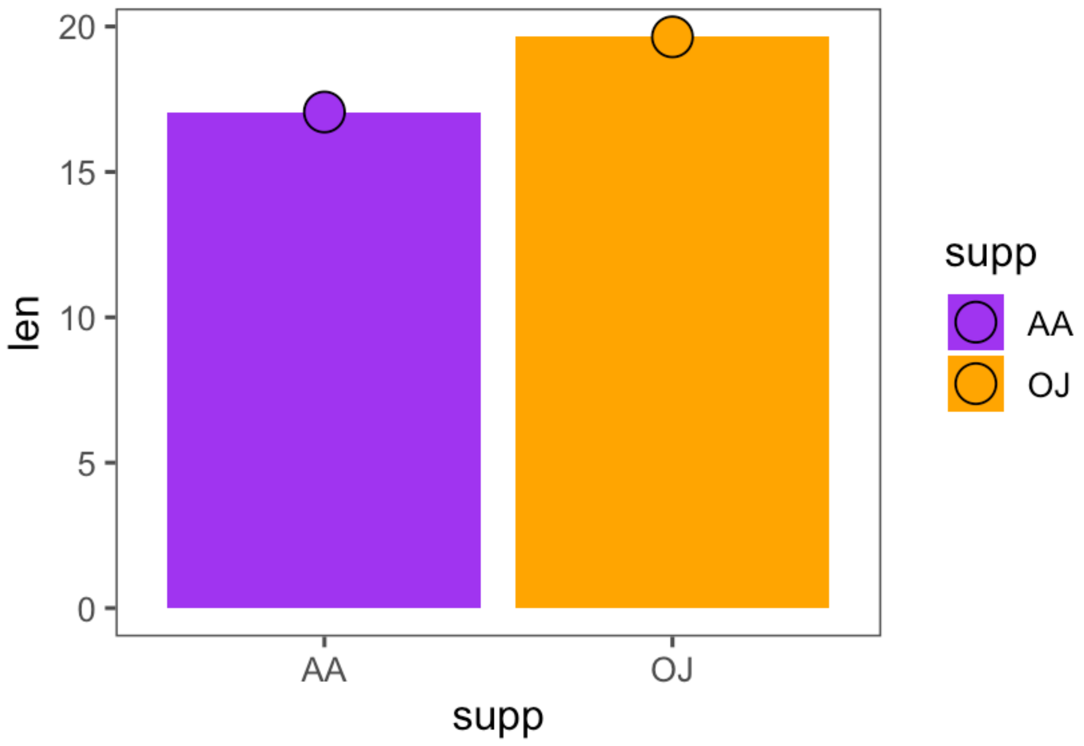
tooth growth dataset: main effects

- **dose** matters (0.5 mg << 2 mg)
  - MAIN effect: the "overall" effect of dose (ignoring delivery method), i.e., difference in tooth growth for 0.5 mg vs. 2 mg
  - $M_{0.5mg} – M_{2mg}$



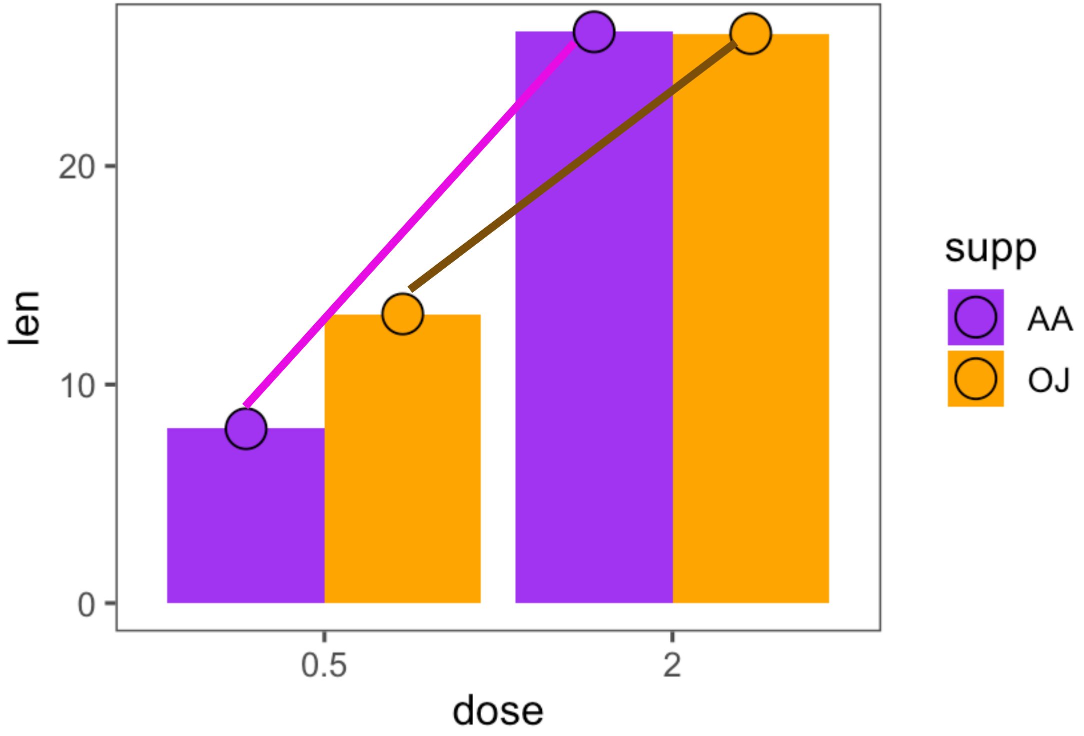
tooth growth dataset: main effects

- **supplement** matters (OJ > AA)
  - MAIN effect: the "overall" effect of supplement (ignoring dose), i.e., difference in tooth growth for OJ vs. AA
  - $M_{OJ} – M_{AA}$



tooth growth dataset: interactions

- **dose** and **supplement** influence each other
  - INTERACTION effect: the difference between differences
    - $OJ_{0.5mg} - OJ_{2mg}$ vs. $AA_{0.5mg} - AA_{2mg}$
- what would the plot look like if there was NO interaction?
  - parallel lines!



main effects and interactions

- main effects represent the “overall” effect of one independent variable when ignoring the influence of other variables
- interactions represent the full relationship between multiple independent variables
- when interactions are present in the model, **main effects need to be qualified**, i.e., you cannot truly understand the influence of that variable in isolation

practice question #1

- For a two-factor experiment with 2 levels of factor A and 3 levels of factor B and n = 10 subjects in each treatment condition, how many participants are in <u>each level of factor B</u>?
  - 10
  - 20
  - 30
  - 60

practice question #2

- A two-factor research study is used to evaluate the effectiveness of a new blood-pressure medication. In this two-factor study, Factor A is medication versus no medication and factor B is male versus female. The medicine is expected to reduce blood pressure for both males and females, but it is expected to have a much greater effect for males. What pattern of results should be obtained if the medication works as predicted?
    - significant main effect for factor A (medication).
    - a significant interaction.
    - a significant main effect for factor A and a significant interaction.
    - none of the above.

practice question #3

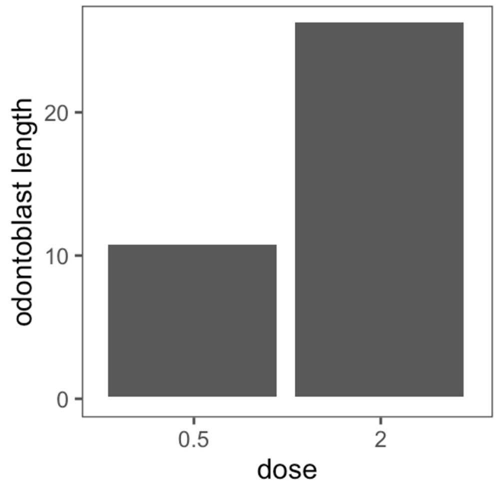
- In a line graph showing the results from a two-factor experiment, the levels of factor A (A1 and A2) are presented on the X-axis and separate lines are used to display the means for B1 and B2. If the points on the line for B1 are consistently 10 points lower than the corresponding point on the line for B2, what pattern of results is indicated?
  - an indication of an overall A-effect
  - an indication of an overall B-effect
  - an indication of a significant interaction
  - no claims can be made

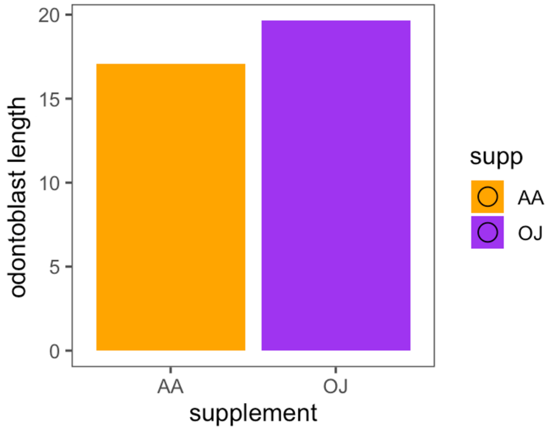
practice question #4

- In a line graph showing the results from a two-factor experiment, the levels of factor A (A1 and A2) are presented on the X-axis and separate lines are used to display the means for B1 and B2. If the points on the line for B1 are consistently **at least** 10 points lower than the corresponding point on the line for B2, what pattern of results is indicated?
  - an indication of an overall A-effect
  - an indication of an overall B-effect
  - an indication of a significant interaction
  - no claims can be made

building a factorial model

- we can start with three simple models
- grand mean model : toothGrowth ~ **grand mean**
- main effect 1: toothGrowth ~ **dose**
  - model = dose means
  - obtain $SS_{dose\_model} = SS_{total} - SS_{Y-\hat{Y}_{dose\_model}}$
- main effect 2: toothGrowth ~ **supp**
  - model = supplement means
  - obtain $SS_{supp\_model} = SS_{total} - SS_{Y-\hat{Y}_{supp\_model}}$





activity: compute the means

- use the tooth growth dataset
- compute all means

supplement	dose=0.5	dose=2
AA	7.98	26.14
OJ	13.23	26.06

AA_overall	17.06
OJ_overall	19.645
dose_0.5	10.605
dose_2	26.1

next time

- **before** class
  - *watch:* Hypothesis Testing (Factorial ANOVA) [33 min]
  - *explore:* Problem Set 7!
  - *post*: Data Around Us OR practice questions (class participation)
- **during** class
  - review for midterm 2!