

Homework Feedback

- Avoid “statistical significant effect”
 - Hypothesis:
Video game increase hippocampal volume (compared to...?)
 - IV: give a general name, the IV is not “control, music training & video game training”
 - Sample size for ANOVA / Regression:
30 → to satisfactorily represent the population
 - Type I error → controlled
 - Conclusion: We do not have enough evidence to conclude...
 - Too long... It’s a template for your exam and in the future
- 

HW: Parent and Family Involvement in Education

cbook_pfi_pu_doc.pdf - Adobe Acrobat Pro DC

File Edit View Window Help

Home Tools Document

Comment

E12. How much do you agree or disagree with the following statement:

Variable Name : SEENJOY Label: E12. Child enjoyment c

Position : 37-38

Format : N2.

Response	code	frequency
1 Strongly agree	1	5,130
2 Agree	2	6,868
3 Disagree	3	1,235
4 Strongly disagree	4	290
-1 Valid Skip	-1	552

8.50 x 11.00 in

Search Comments...

19 Comments

Page 1

Jonas

Highlighted Text

basmid

3/2/2018 10:59 AM

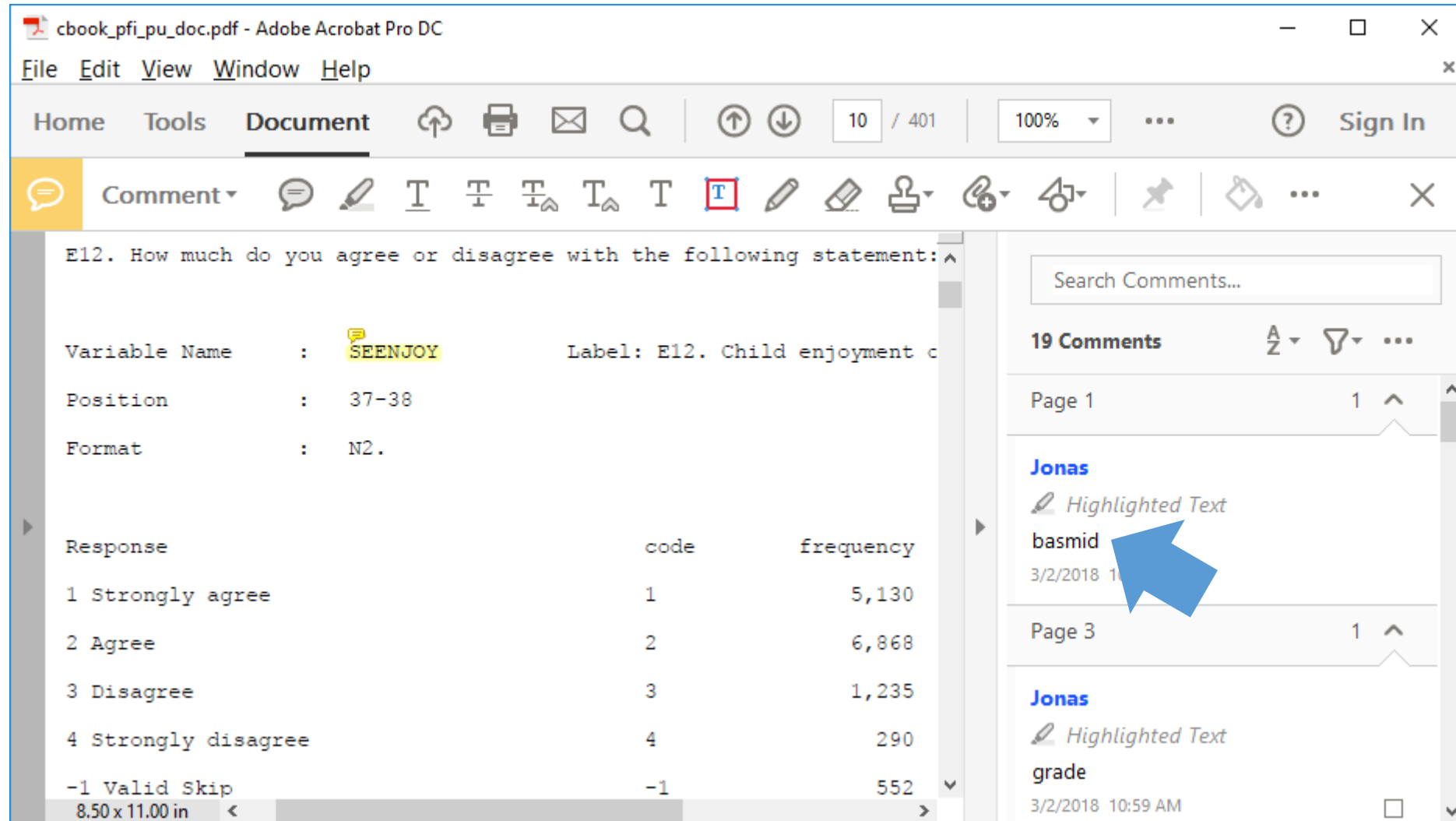
Page 3

Jonas

Highlighted Text

grade

3/2/2018 10:59 AM



Original Questionnaire

National Household Education Survey			
Start Here		You / Person 1	Person 2
The Department of Education is studying the education and training experiences of adults and children. Each household is different, and we need your response so we can send you a survey that is right for your household. ▶ Return this form even if there are only one or two people in the household. ▶ This survey should be filled out by an adult household member living at this address. ▶ Please use a blue or black pen if available.		2. What is his or her first name, initials, or nickname? First names will be used only to ask you questions about the education of a specific person. First name/initials/nickname	First name/initials/nickname
1. How many people live in this household?		3. What is this person's month and year of birth? month / year of birth	month / year of birth
		4. What is this person's sex? ☐ Male ☐ Female	☐ Male ☐ Female

Remove Outliers ± 3 sd of each column mean

```
# remove the outliers
# Hint/optional: use a for-loop
for (i in 1:ncol(eduData)){
  # calculate the column mean
  colMean = NA

  # calculate the column sd
  colSD = NA

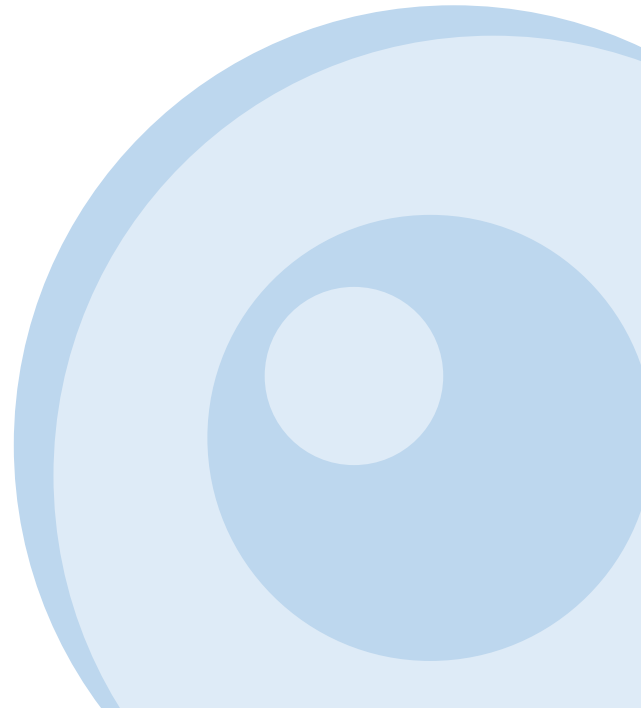
  # retain all values within  $\pm 3$ sd in each column
  eduData = eduData[NA,]
}
```

PSYC 193R:

Data Analysis & Visualization with R

Lecture 16: Two-way ANOVA

Jonas S.-H. Lau [silau -et- ucsd.edu]
psyc193r.ucsd.edu



Second Language (L2) Acquisition

It is believed that adults have more difficulties acquiring a second language (L2) compared to children.

A group of language researchers at UC San Diego would like to examine the claim. In addition to age, they also wanted to study the effect of gender, which is also thought to contribute to L2 acquisition. The researchers recruited a group of English monolingual first graders, and a group of English monolingual undergraduates.

In a school year, these learners were given the same Cantonese 101 class administered by a group of experience teachers. Cantonese (a Chinese dialect) was chosen because it is very distinct from English, and its deep connections with Chinese literature. Written and spoken language abilities were assessed at the end of the course.

L2 Acquisition: The Design

- What was the research question?
- What was the target population?
- What were the Factors(s) and DV(s)?
- How were the IV(s) and DV(s) operationalized?
(also, scale of measurement?)
- Was it a between / within-subjects design?
- What are the expected results?
Comparing the two age groups:
Comparing male vs female:



Factors in ANOVA

Age

{ 6
20 }

Gender

{ Female
Male }

2 levels ×

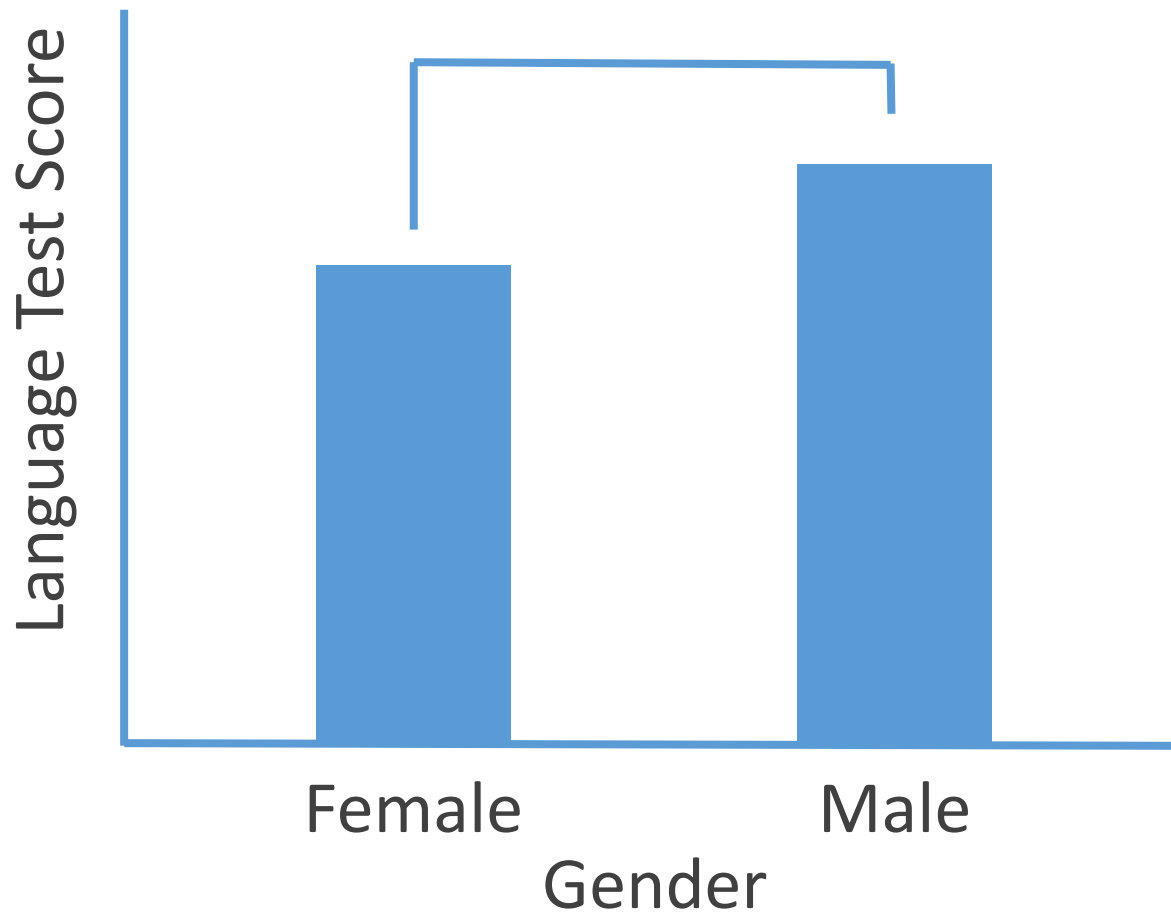
2 levels

Factorial Design



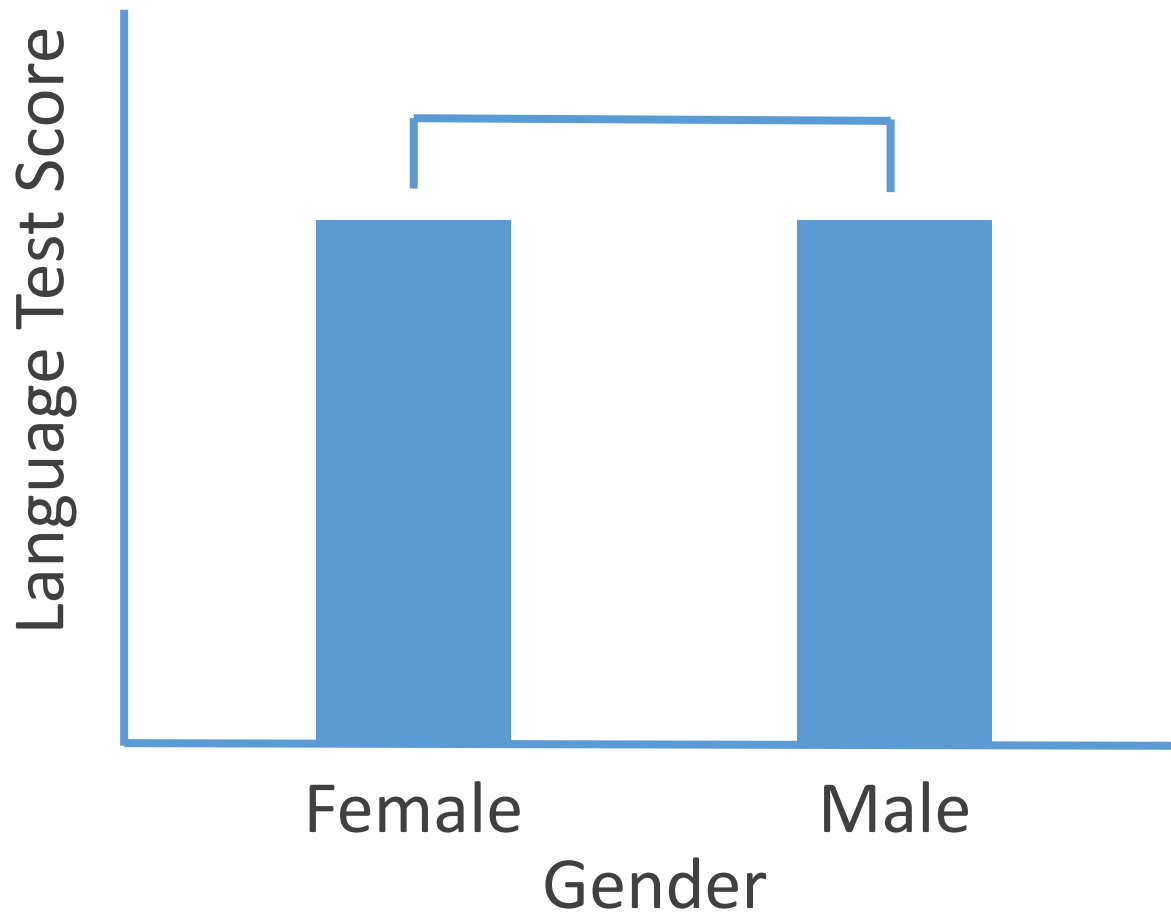
Main Effect of Gender: Hypothetical

- Main effect of gender: Is there a difference in language test scores between the two genders?



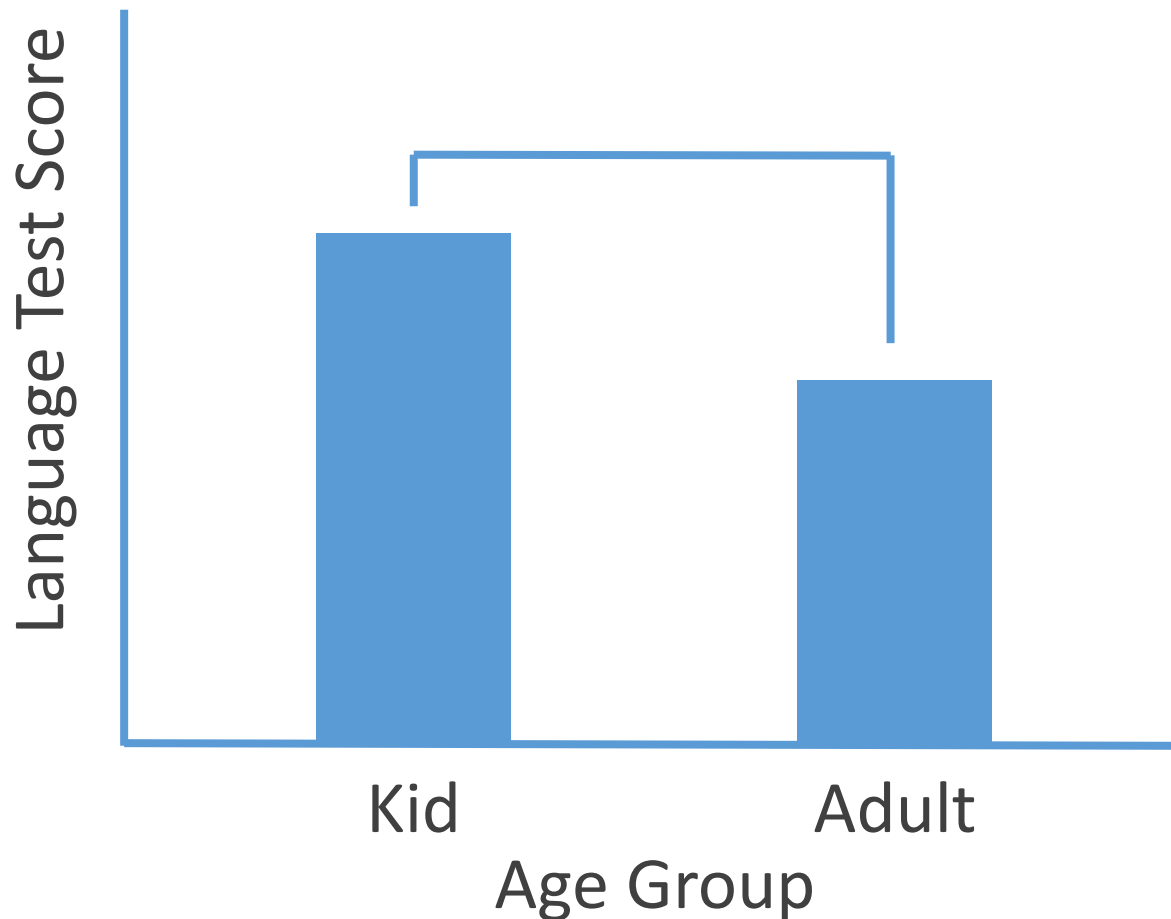
Main Effect of Gender: Hypothetical

- Main effect of gender: Is there a difference in language test scores between the two genders?



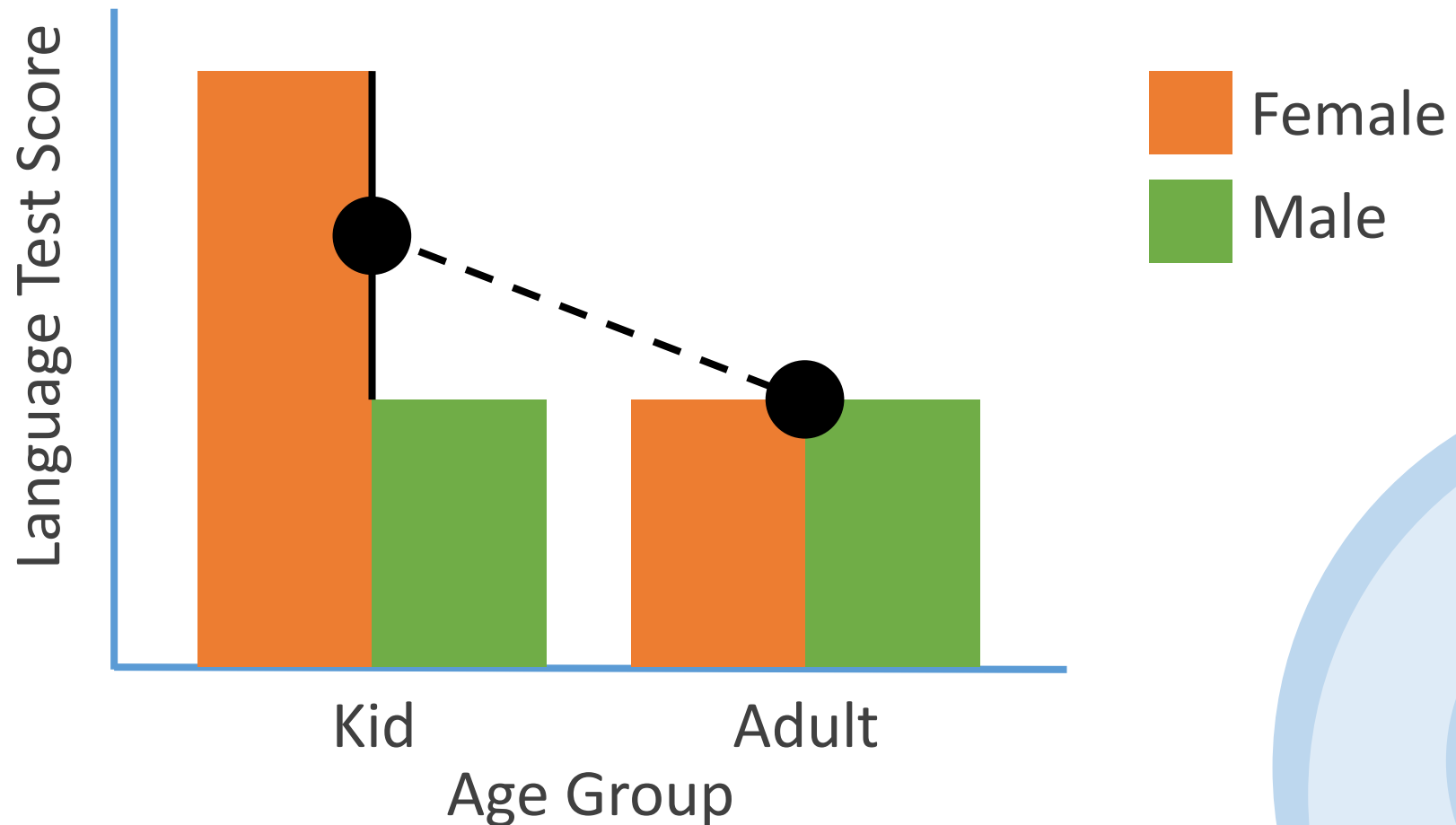
Main Effect of Age: Hypothetical

- Main effect of age: Is there a difference in language test scores between the age groups?



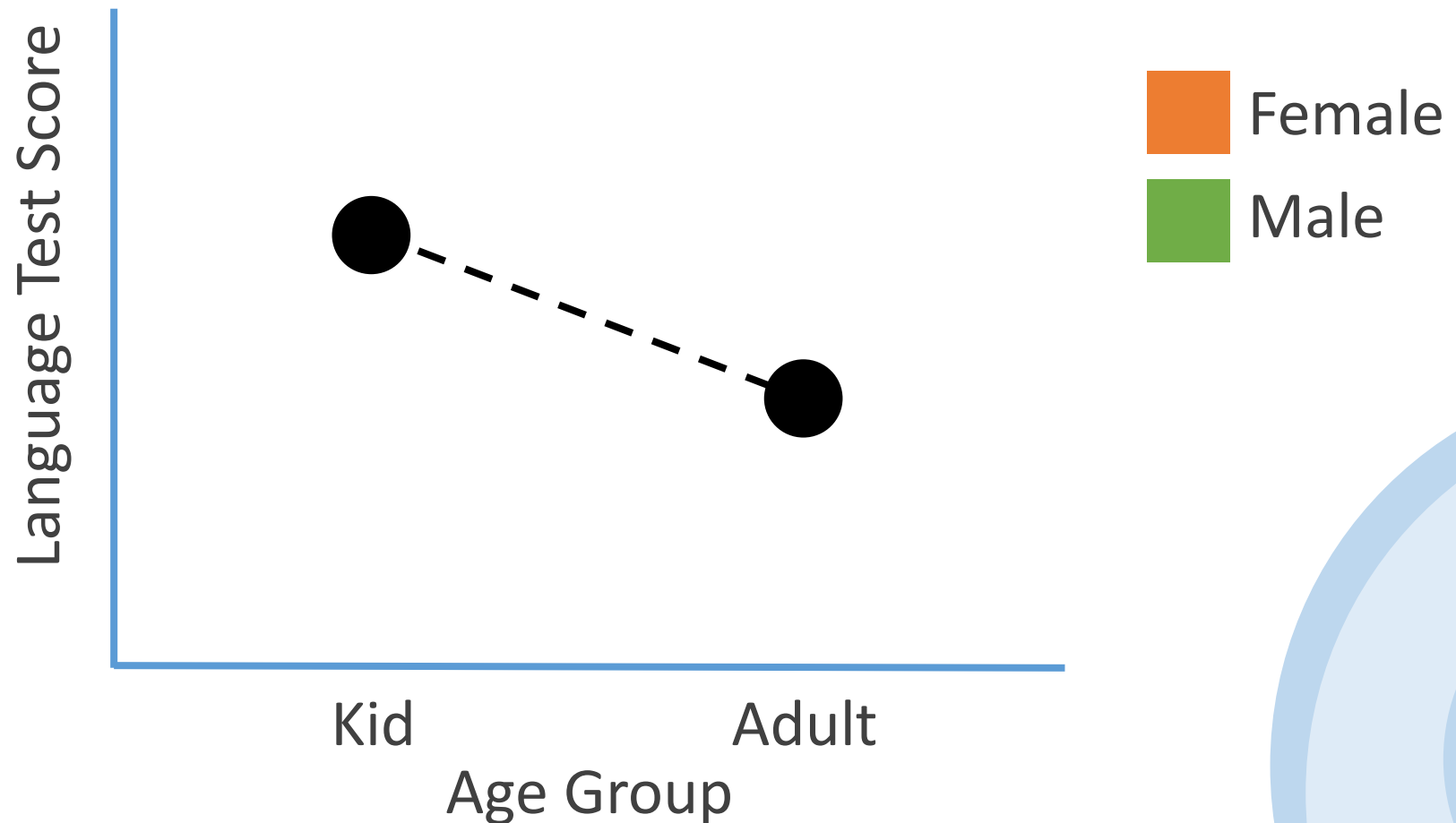
Main Effect of Age: Hypothetical

- Main effect of age: Is there a difference in language test scores between the age groups?



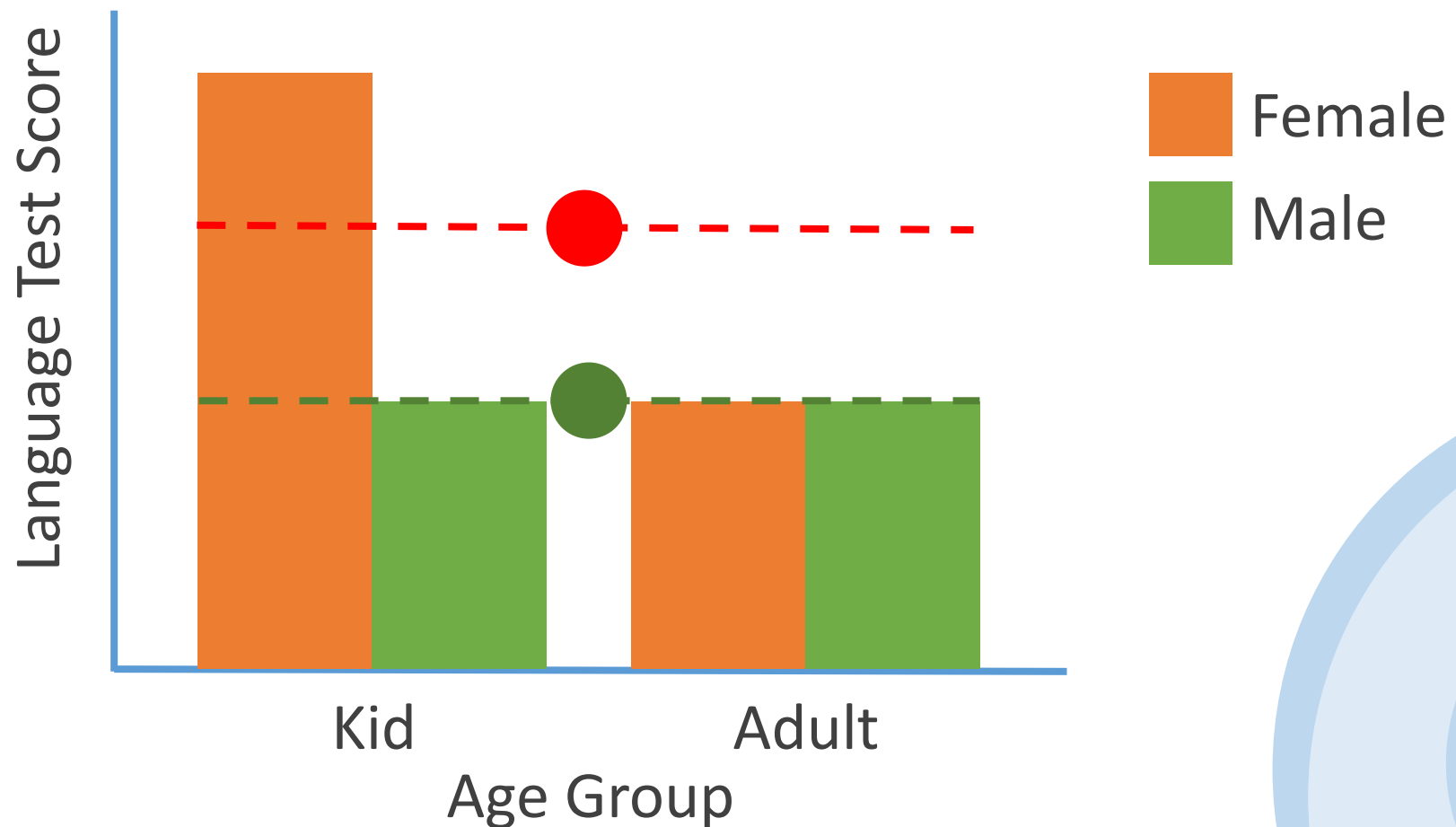
Main Effect of Age: Hypothetical

- Main effect of age: Is there a difference in language test scores between the age groups?



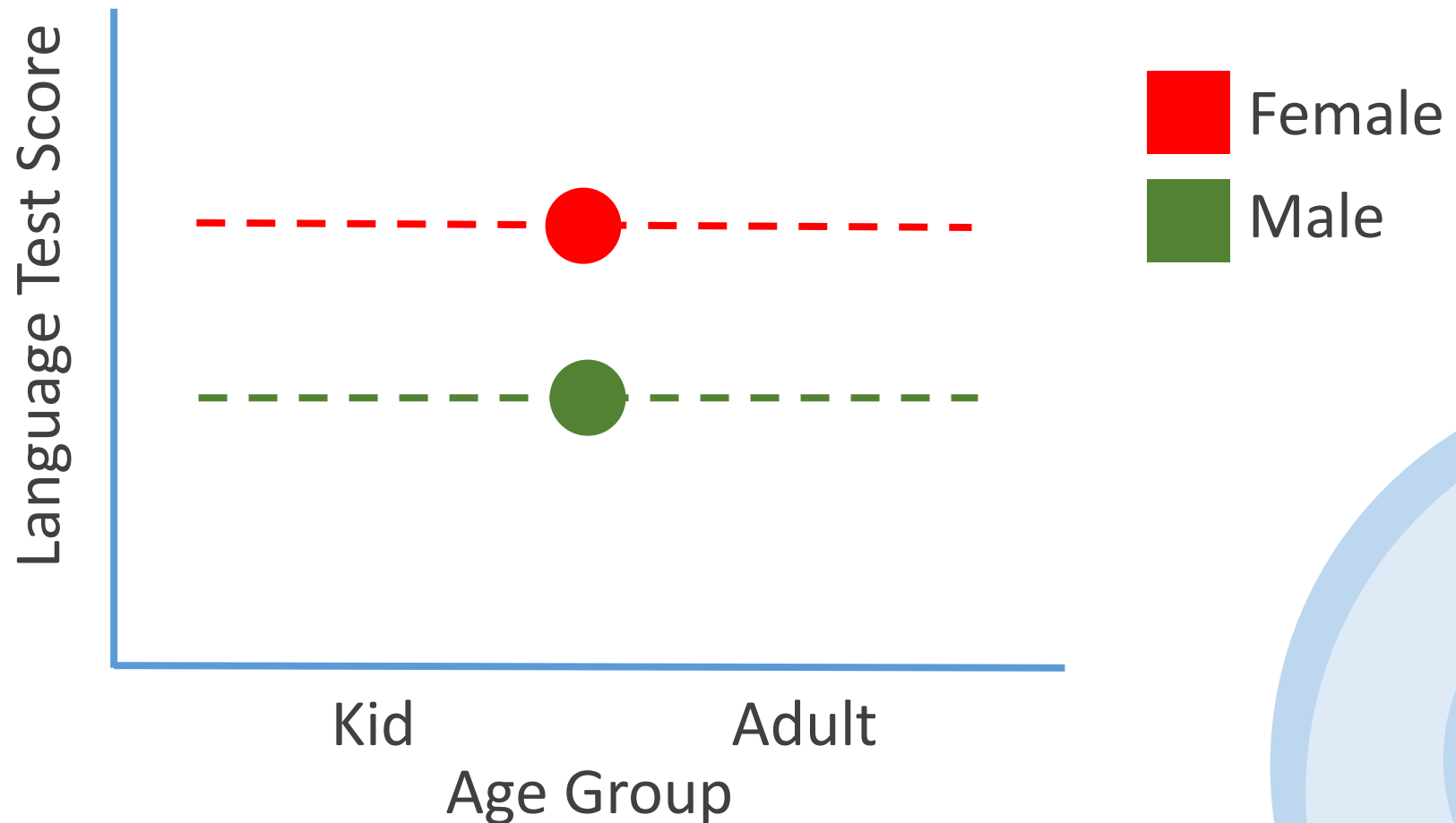
Main Effect of Gender: Hypothetical

- Main effect of gender: Is there a difference in language test scores between the genders?



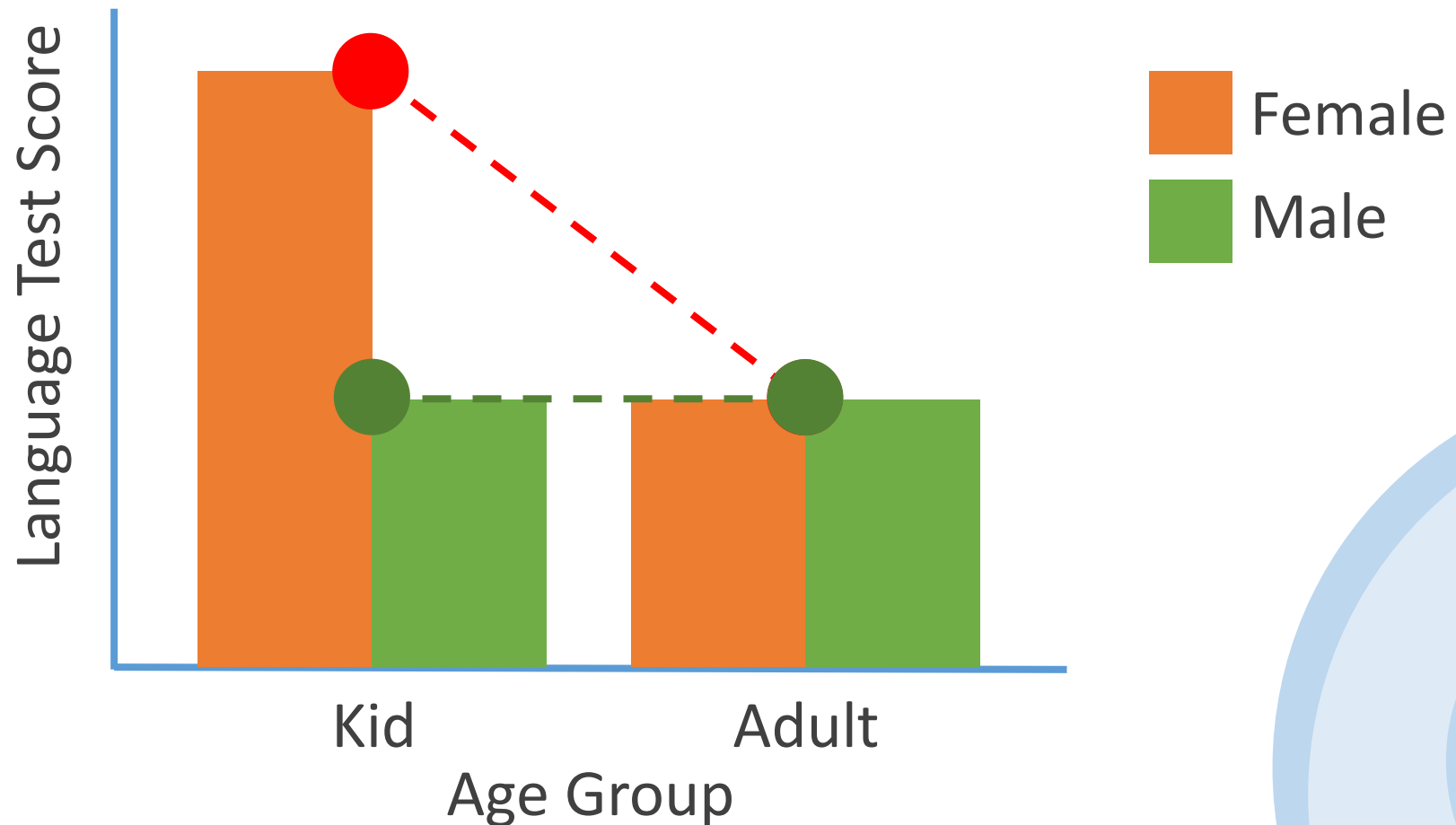
Main Effect of Gender: Hypothetical

- Main effect of gender: Is there a difference in language test scores between the genders?



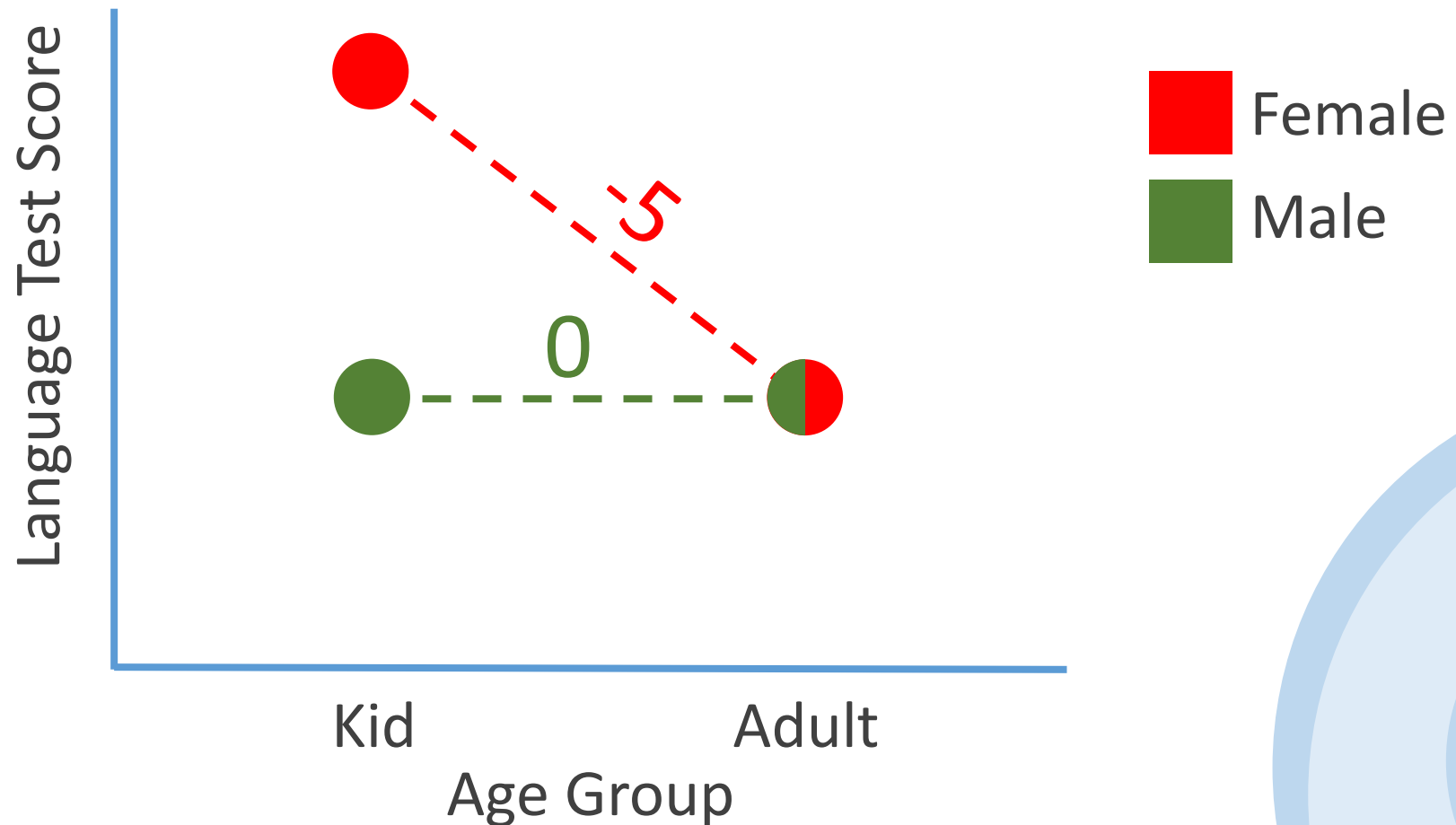
Interaction: Hypothetical

- Interaction between gender and age: Is the difference between the two female groups the same as the difference between the male groups?

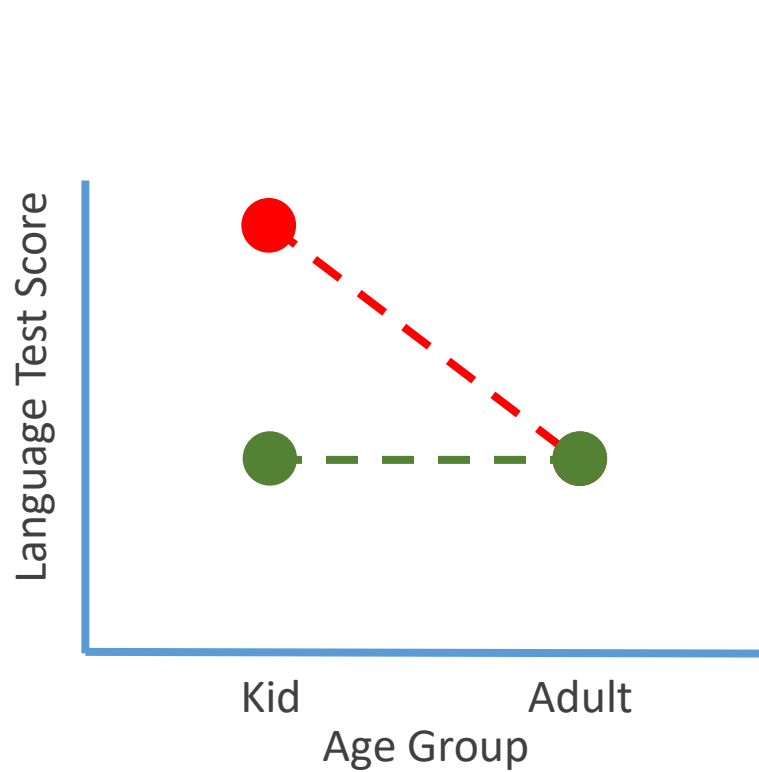


Interaction: Hypothetical

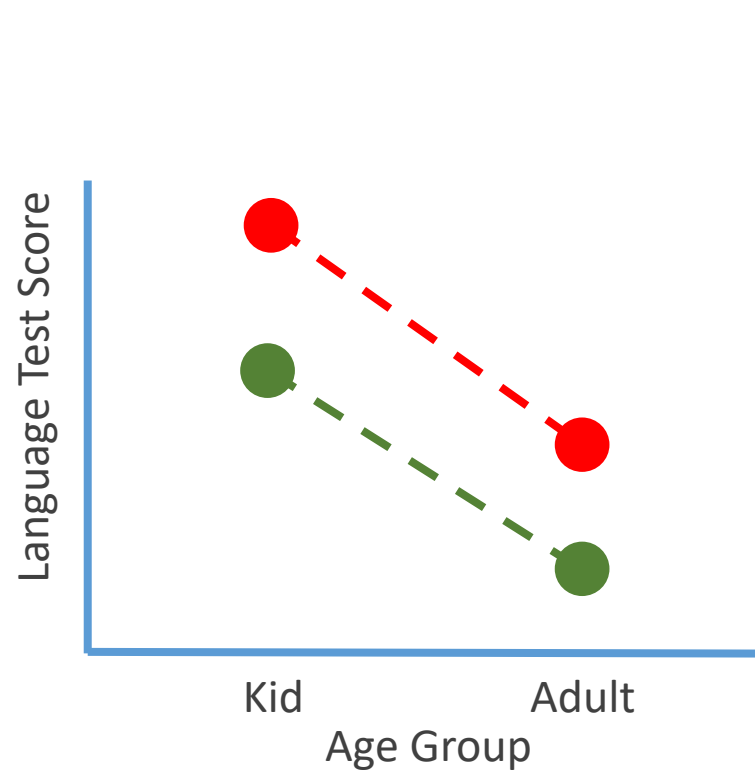
- Interaction between gender and age: Is the difference between the two female groups the same as the difference between the male groups?



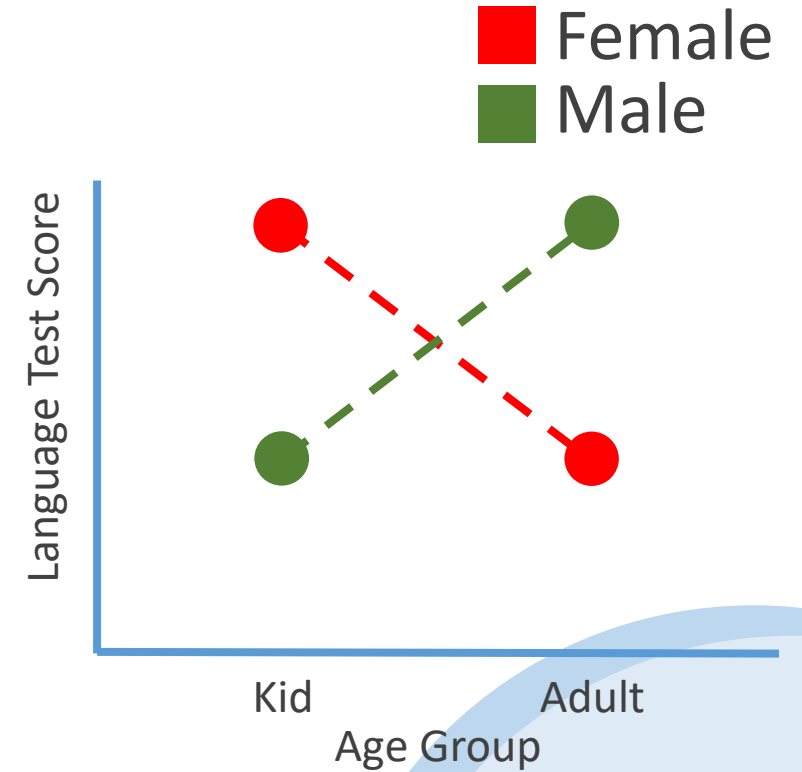
Main Effect / Interaction Patterns



Main effect (gender)
Main effect (age)
Interaction

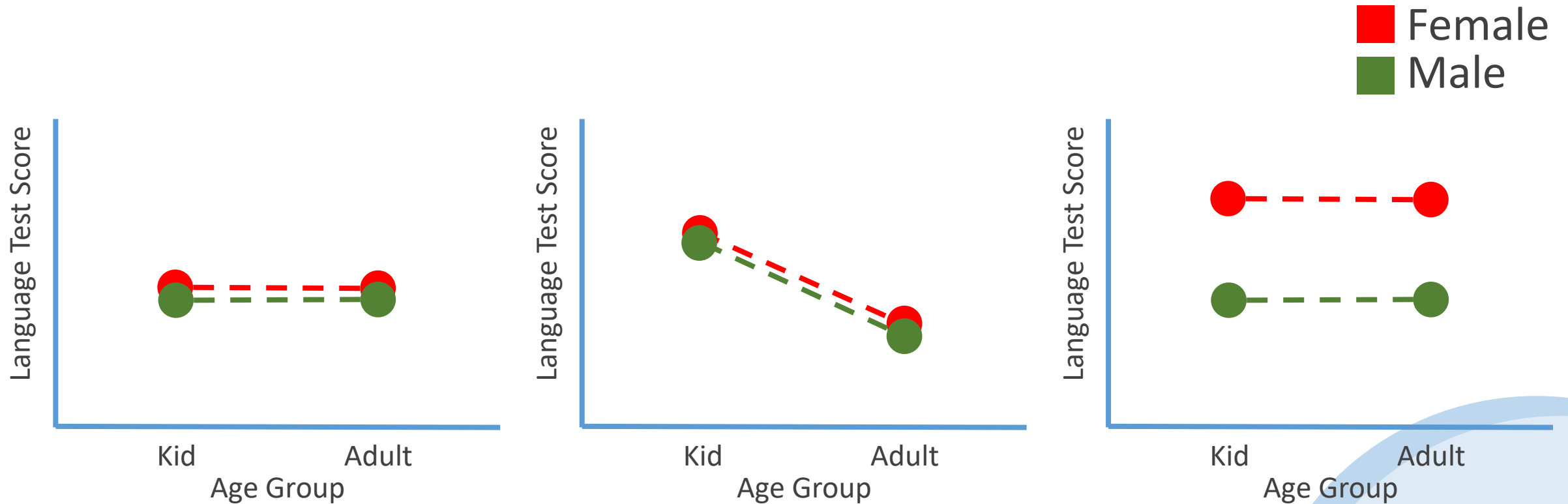


Main effect (gender)
Main effect (age)
Interaction



Main effect (gender)
Main effect (age)
Interaction

Main Effect / Interaction Patterns



Main effect (gender)
Main effect (age)
Interaction

Main effect (gender)
Main effect (age)
Interaction

Main effect (gender)
Main effect (age)
Interaction

Second Language (L2) Acquisition

It is believed that adults have more difficulties acquiring a second language (L2) compared to children.

A group of language researchers at UC San Diego would like to examine the claim. In addition to **age**, they also wanted to study the effect of **gender**, which is also thought to contribute to L2 acquisition. The researchers recruited a group of English monolingual first graders, and a group of English monolingual undergraduates.

In a school year, these learners were given the same Cantonese 101 class administered by a group of experience teachers. Cantonese (a Chinese dialect) was chosen because it is very distinct from English, and its deep connections with Chinese literature. **Written and spoken language abilities** were assessed at the end of the course.

Import Data

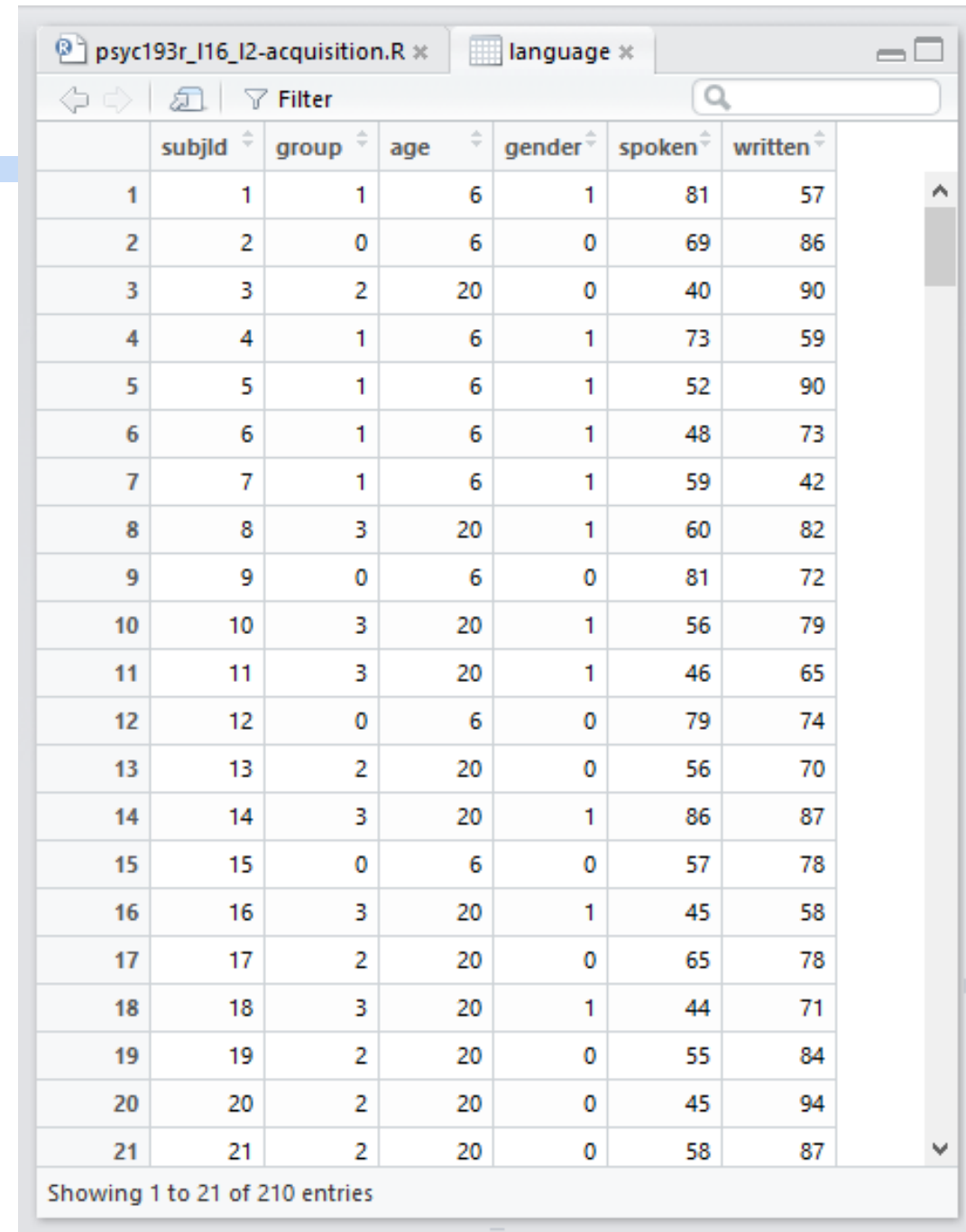
- Download the data file:
https://psyc193r.ucsd.edu/data/psyc193r_language_data.csv
- Tell R where your data file is:
Session → Set Working Directory → Choose Directory
`setwd()`
- Check the *comma separated values* (csv) file with MS Excel to ensure correct format
- Import the file into R as a data frame, saving it to “language” variable
`... = read.csv("file_name.csv")`

Data structure

The data set has a "long" format

Five variables:

- group: 0 – first grader/female, 1 – first grader/male, 2 – undergrad/female, 3 – undergrad/male
- age: 6 (first grader) or 20 (undergrad)
- gender: 0 – female, 1 – male
- spoken: spoken test score
- written: written test score
- score: total test score (to be created)



psyc193r_I16_I2-acquisition.R × language ×

Filter

	subjId	group	age	gender	spoken	written
1	1	1	6	1	81	57
2	2	0	6	0	69	86
3	3	2	20	0	40	90
4	4	1	6	1	73	59
5	5	1	6	1	52	90
6	6	1	6	1	48	73
7	7	1	6	1	59	42
8	8	3	20	1	60	82
9	9	0	6	0	81	72
10	10	3	20	1	56	79
11	11	3	20	1	46	65
12	12	0	6	0	79	74
13	13	2	20	0	56	70
14	14	3	20	1	86	87
15	15	0	6	0	57	78
16	16	3	20	1	45	58
17	17	2	20	0	65	78
18	18	3	20	1	44	71
19	19	2	20	0	55	84
20	20	2	20	0	45	94
21	21	2	20	0	58	87

Showing 1 to 21 of 210 entries

Data Formatting / Processing / Cleaning

- Decide which are the nominal variables and factorize them
 - Also label the groups whenever appropriate

```
language$____ = ...
```

"age" as factor or numeric?

- Coerce ... to numeric variables

```
language$spoken = ...  
language$written = ...
```

Create a total test score column

- It can be the total, or the average of spoken and written test scores

```
language$score = ...
```



Summary statistics

- Use aggregate to find the descriptive statistics for each group

- Means

```
aggregate( ____ ~ ____, data = ____, FUN = ____)
```

- Standard Deviations

```
aggregate( ____ ~ ____, data = ____, FUN = ____)
```

- Numbers of participants

```
aggregate( ____ ~ ____, data = ____, FUN = ____)
```



Checking Assumptions

- Representative of the population: $n = 30$

- Outliers

Plot a histogram for each group

```
# draw a two by two plot grid
par(mfrow=c(NA,NA))
# draw a histogram for each age and gender level
for (i in levels(NA)){
  for (j in levels(NA)){
    hist(language[language$age == i & language$gender == j,]$NA,
          xlab = "score",
          main = paste0(i, " / ", j))
  }
}
# reverse back to a 1 by 1 grid
par(mfrow=c(1,1))
```

Checking Assumptions

- Equal variance assumptions

```
# Bartlett Test  
bartlett.test(NA ~ NA, data = language)
```

Bartlett test of homogeneity of variances

data: score by group

Bartlett's K-squared = 0.46597, df = 3, p-value = 0.9263



Building an aov model

- Choosing between models:
- Main effect models:

```
# age effect only
language.aov = aov(score ~ age, data = language)
# age and gender main effects only
language.aov = aov(score ~ age + gender, data = language)
```

- Interaction only model:

```
# interaction between age and gender only
language.aov = aov(score ~ age:gender, data = language)
```

- Full 2-way ANOVA model:

```
language.aov = aov(score ~ age*gender, data = language)
```

- Read out the summary

```
summary(language.aov)
```

Building an aov model

- Full 2-way ANOVA model:

```
language.aov(score ~ age*gender, data = language)
summary(language.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
age	1	45	44.9	0.425	0.51528	
gender	1	904	904.1	8.552	0.00384	**
age:gender	1	748	747.7	7.074	0.00844	**
Residuals	206	21776	105.7			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Building an aov model

- Full 2-way ANOVA model:

```
language.aov(score ~ age*gender, data = language)
summary(language.aov)
```

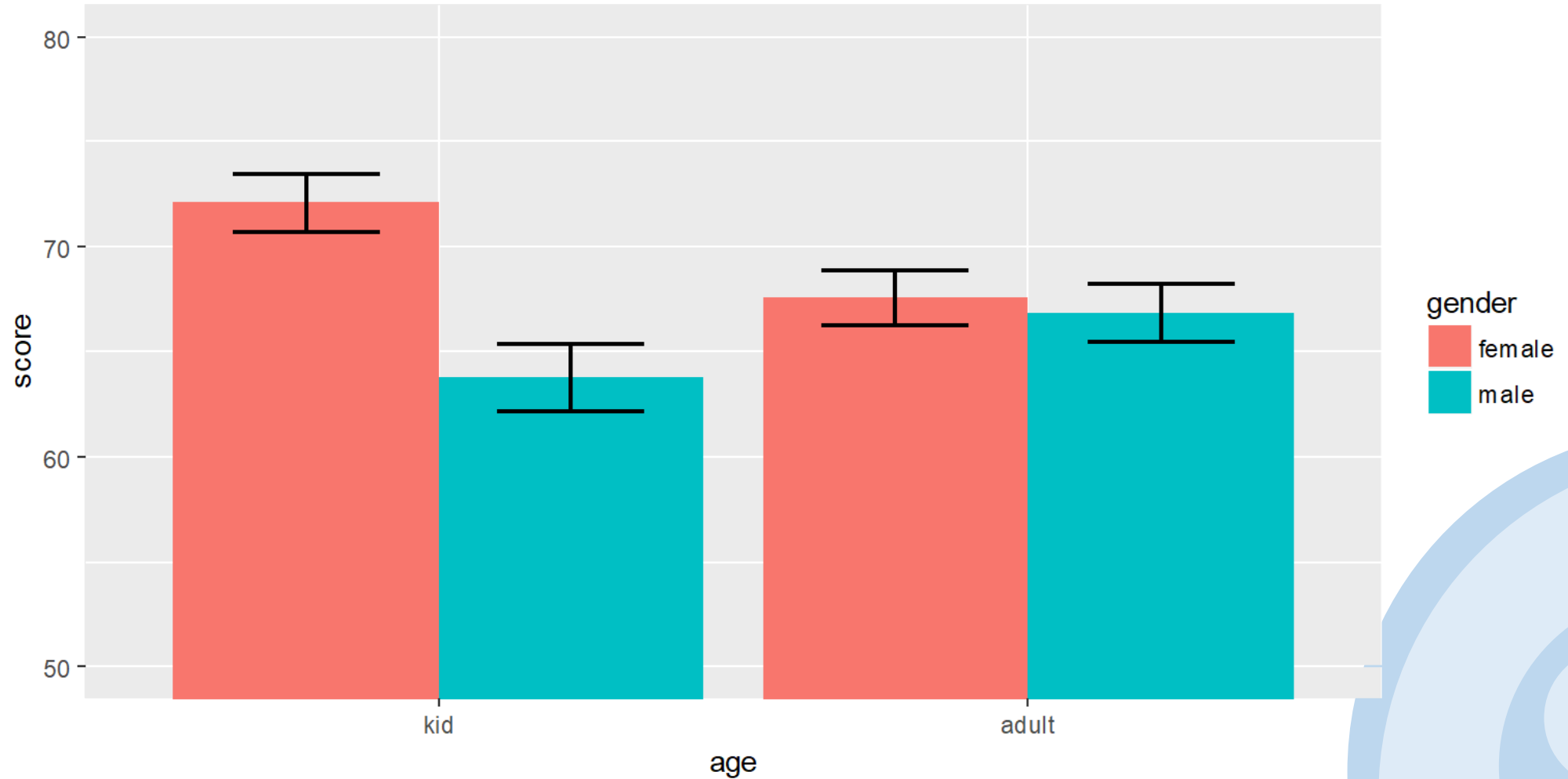
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
age	1	45	44.9	0.425	0.51528	
gender	1	904	904.1	8.552	0.00384	**
age:gender	1	748	747.7	7.074	0.00844	**
Residuals	206	21776	105.7			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Degrees of Freedom:

- main effect: $(n_{\text{group}} - 1)$
- interaction: $(n_{\text{factor}_1} - 1) \times (n_{\text{factor}_2} - 1)$
- residuals: $n_{\text{subjects}} - df_{\text{main_effect}} - df_{\text{interaction}} - 1$

Visualization (next time)



Summarizing the ANOVA test results

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
age	1	45	44.9	0.425	0.51528	
gender	1	904	904.1	8.552	0.00384	**
age:gender	1	748	747.7	7.074	0.00844	**
Residuals	206	21776	105.7			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- **Main effect of age:** We _____ that there is a main effect of age, $F(1,206) = \underline{\hspace{1cm}}$, $p = \underline{\hspace{1cm}}$. We _____ that first graders ($M_{\text{first_grader}} = 68$, $SD_{\text{first_grader}} = 10.9$) are better language learners compared to adults (...), or vice versa.
- **Main effect of gender:** We _____ that there is a main effect of gender, $F(1,206) = \underline{\hspace{1cm}}$, $p = \underline{\hspace{1cm}}$. We _____ that female ($M_{\text{female}} = 70$, $SD_{\text{female}} = 10.3$) are better language learners compared to male (...)
- **Interaction:** We _____ that there _____ an interaction between age and gender, $F(1,206) = \underline{\hspace{1cm}}$, $p = \underline{\hspace{1cm}}$.
(Describe the interaction pattern, if any).

Effect Size (% Variance explained)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
age	1	45	44.9	0.425	0.51528	
gender	1	904	904.1	8.552	0.00384	**
age:gender	1	748	747.7	7.074	0.00844	**
Residuals	206	21776	105.7			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$$\eta_{age}^2 = \frac{SS_{age}}{SS_{total}} = \frac{45 + 904 + 748 + 21776}{(45 + 904 + 748 + 21776)} = \text{---}\%$$