Analysis of Psychological Data

Lab 10. Extending What We Have Learned: Two-Way ANOVA and Correlation

Ihnwhi Heo (iheo2@ucmerced.edu)
Quantitative Methods, Measurement, and Statistics
Website: https://ihnwhiheo.github.io
Office: https://ucmerced.zoom.us/j/2093557522 (Thursday 3:30 - 5:30 pm)

Some announcements

Boost your GPA

Extra credit assignments on April 21 & April 28

Homework 6

Will be published tomorrow

Lab 11 (next Wednesday)

Will be our final lab meeting

What are we going to do?

Recap to give you a big picture

Two-Way ANOVA

Correlation

Q&A session

ANOVA terminology

N-way ANOVA (# of factors)

There are N independent variables (IVs = factors = grouping variables = categorical variables)

N = 1 $\longrightarrow$ One-way ANOVA = ANOVA with 1 IV

N = 2 $\longrightarrow$ Two-way ANOVA = ANOVA with 2 IVs

N = 2 or more $\longrightarrow$ Factorial ANOVA $\longrightarrow$ Start to consider interactions

ANOVA terminology

$A \times B \times C \times ... \times D$ ANOVA (# of levels)

$4 \times 4 \times 4$ ANOVA = ANOVA with 3 IVs, each with 4 levels

3×5 ANOVA = ANOVA with 2 IVs, one with 3 levels and the other with 5 levels

$4 \times 2 \times 5$ ANOVA = ?

Factorial ANOVA hypothesis

Two-way ANOVA

$\longrightarrow$ null hypothesis H_0: Nothing is going on. No effect (no difference)!

H_0: IV1 does not have a significant effect on DV
= Group means on DV across levels in IV1 are not different

H_0: IV2 does not have a significant effect on DV
= Group means on DV across levels in IV2 are not different

H_0: Interaction between IV1 and IV2 does not have a significant effect on DV

Factorial ANOVA hypothesis

Two-way ANOVA
$\longrightarrow$ alternative hypothesis H_1: Something is going on. There is an effect!

H_1: IV1 has a significant effect on DV
= Group means on DV across levels in IV1 are different

H_1: IV2 has a significant effect on DV
= Group means on DV across levels in IV2 are different

H_1: Interaction between IV1 and IV2 has a significant effect on DV

Factorial ANOVA table

Source	SS	df	MS	F	F_{crit}
A	SS_A	J-1	SS_A/df_A	MS_A/MS_{Within}	$F_{crit}(df_A, df_{Within})$
B	SS_B	K-1	SS_B/df_B	MS_B/MS_{Within}	$F_{crit}(df_B, df_{Within})$
AB	SS_{AB}	(J-1)(K-1)	SS_{AB}/df_{AB}	MS_{AB}/MS_{Within}	$F_{crit}(df_{AB}, df_{Within})$
Within	SS_{Within}	N-JK	SS_{Within}/df_{Within}		
Total	SS_{Total}	N-1			

Instead of calculating SS_{AB}, we use...

$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Within}$ In addition to this, also...

$$df_{Total} = df_A + df_B + df_{AB} + df_{Within}$$

Factorial ANOVA table

Source	SS	df	MS	F	F_{crit}
A	54	2	27	13.5	9.55
B	0	1	0	0	10.13
AB	0	2	0	0	9.55
Within	6	3	2		
Total	60	8			

Do you remember how to find the critical F-ratio in the F-table? :)

Factorial ANOVA table

Effect size for factorial ANOVA: Partial eta-squared η_p^2

$$\eta_{pA}^2 = \frac{SS_A}{SS_A + SS_{Within}}$$

$$\eta_{pB}^2 = \frac{SS_B}{SS_B + SS_{Within}}$$

$$\eta_{pAB}^2 = \frac{SS_{AB}}{SS_{AB} + SS_{Within}}$$

Factorial ANOVA table

Source	SS	df	MS	F	F_{crit}
A	54	2	27	13.5	9.55
B	0	1	0	0	10.13
AB	0	2	0	0	9.55
Within	6	3	2		
Total	60	8			

For example,

$$\eta^2_{pA} = \frac{SS_A}{SS_A+SS_{Within}} = \frac{54}{54+6} = \frac{54}{60} = .90$$

Factor A accounts for 90% of the between-subjects variance in DV

Effect

Wait... we have been using a word **effect** a lot...

Null hypothesis (no effect), alternative hypothesis (there is an effect), effect size...

Let me ask you: What is an effect in statistics?

Effect

Intuitively, effect means '**differences**' in statistics!

Null hypothesis $\longrightarrow$ no effect $\longrightarrow$ no difference

Alternative hypothesis $\longrightarrow$ there is an effect $\longrightarrow$ there is a difference

Effect size $\longrightarrow$ How big our differences are

Effect

Main effect and interaction effect are also about 'differences'!

Remember, ANOVA is all about comparing means...

Working example

Two-way ANOVA where two factors are school type and region

Dependent variable: Happiness

School type (UC Merced and Merced College) & Region (California and Massachusetts)

Main effect

Mean differences across the levels of a single factor

An effect of a single factor averaged across levels of the other factors

Whether there are mean differences in happiness between UC Merced and Merced College (or between California and Massachusetts)

Differences in marginal means

N-way ANOVA $\longrightarrow$ There are N main effects

Two-way ANOVA $\longrightarrow$ Two main effects!

Main effect

Mean differences across the levels of a single factor

An effect of a single factor averaged across levels of the other factors

Differences in marginal means

	UC Merced	Merced College	Mean
California	20	10	
Massachusetts	10	8	
Mean			

Main effect

Mean differences across the levels of a single factor

An effect of a single factor averaged across levels of the other factors

Differences in marginal means

	UC Merced	Merced College	Mean
California	20	10	(20+10)/2
Massachusetts	10	8	(10+8)/2
Mean	(20+10)/2	(10+8)/2	(20+10+10+8)/4

Main effect

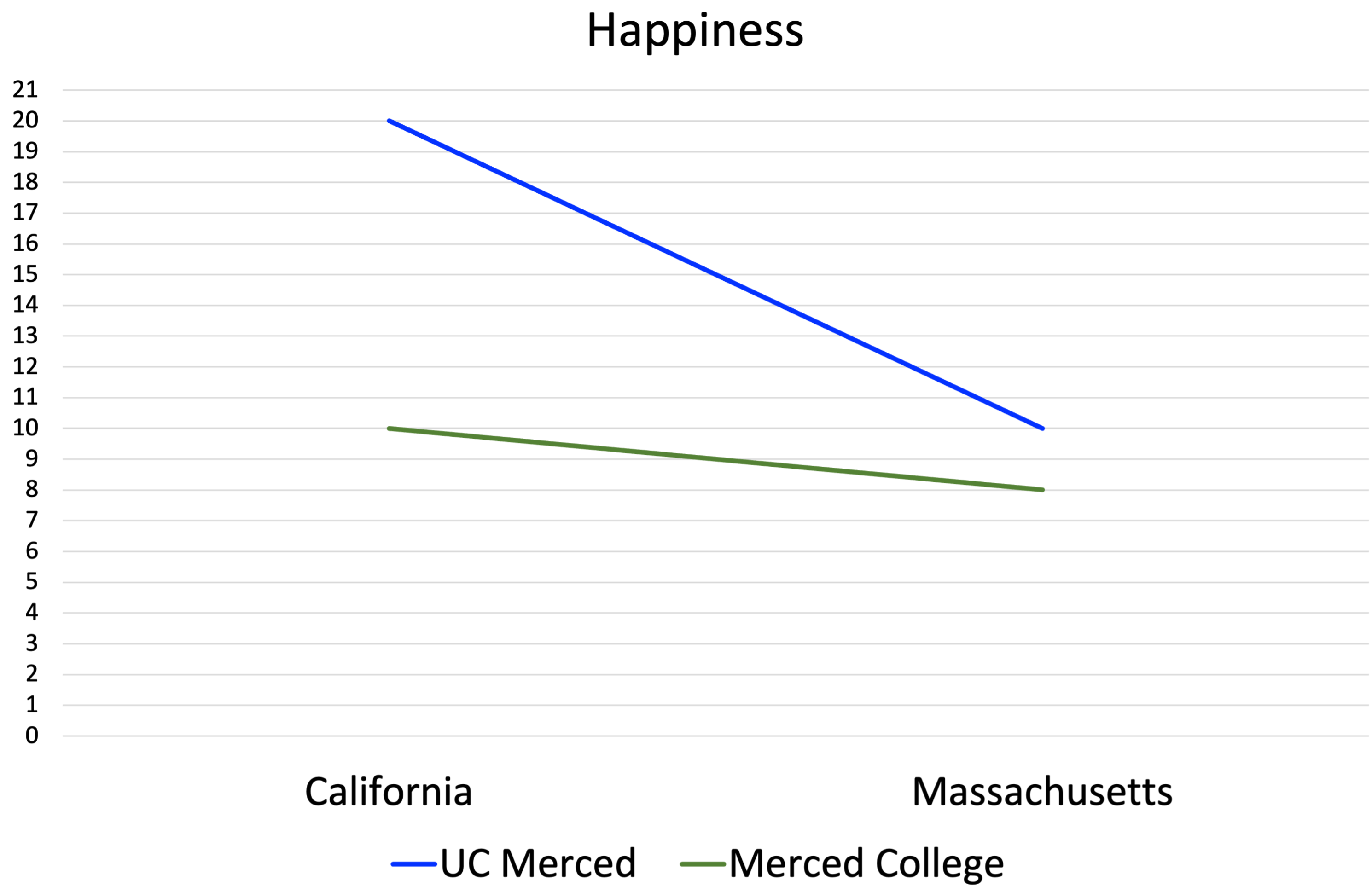
Mean differences across the levels of a single factor

An effect of a single factor averaged across levels of the other factors

Differences in marginal means

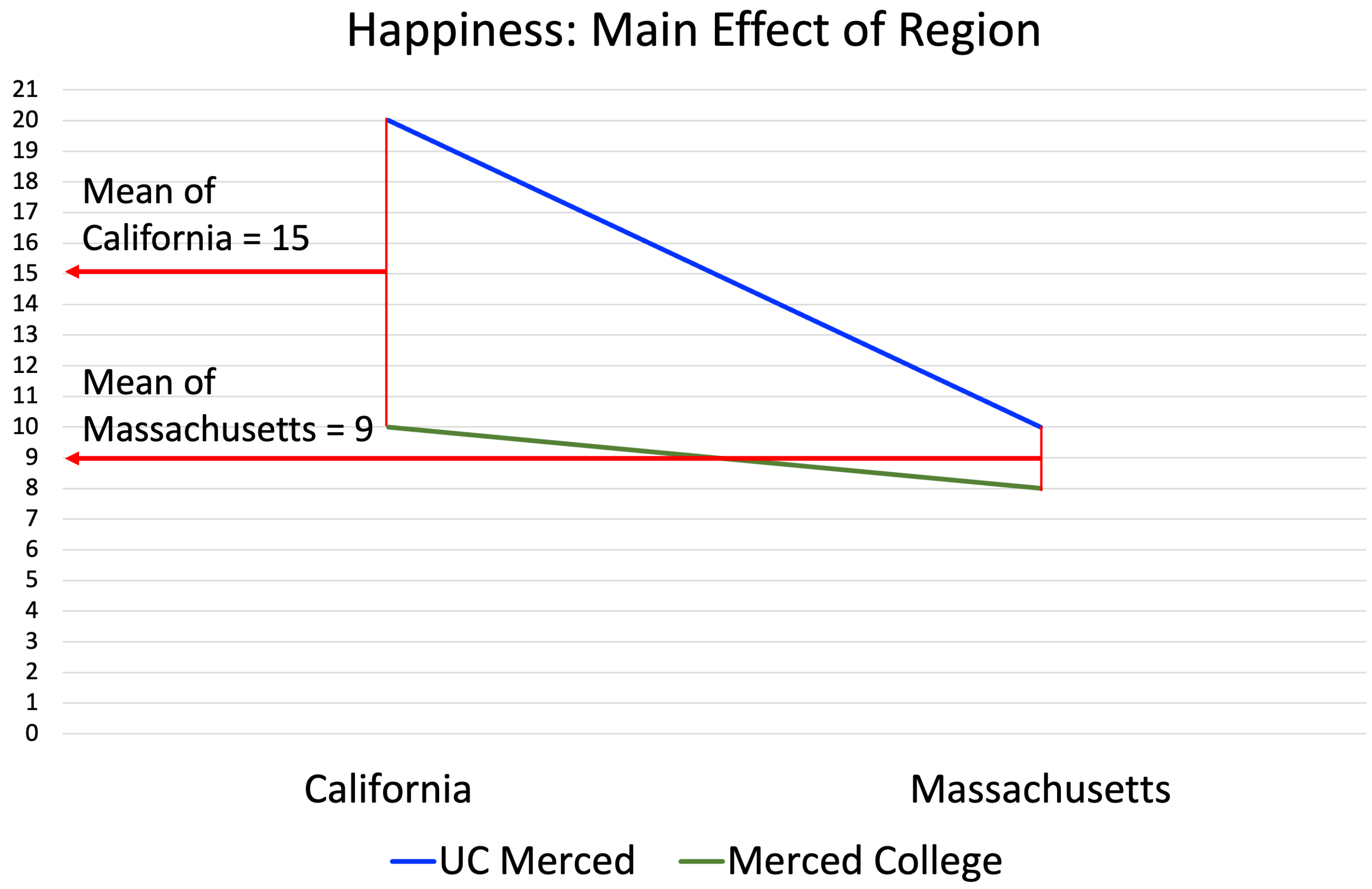
	UC Merced	Merced College	Mean
California	20	10	15
Massachusetts	10	8	9
Mean	15	9	12

Main effect



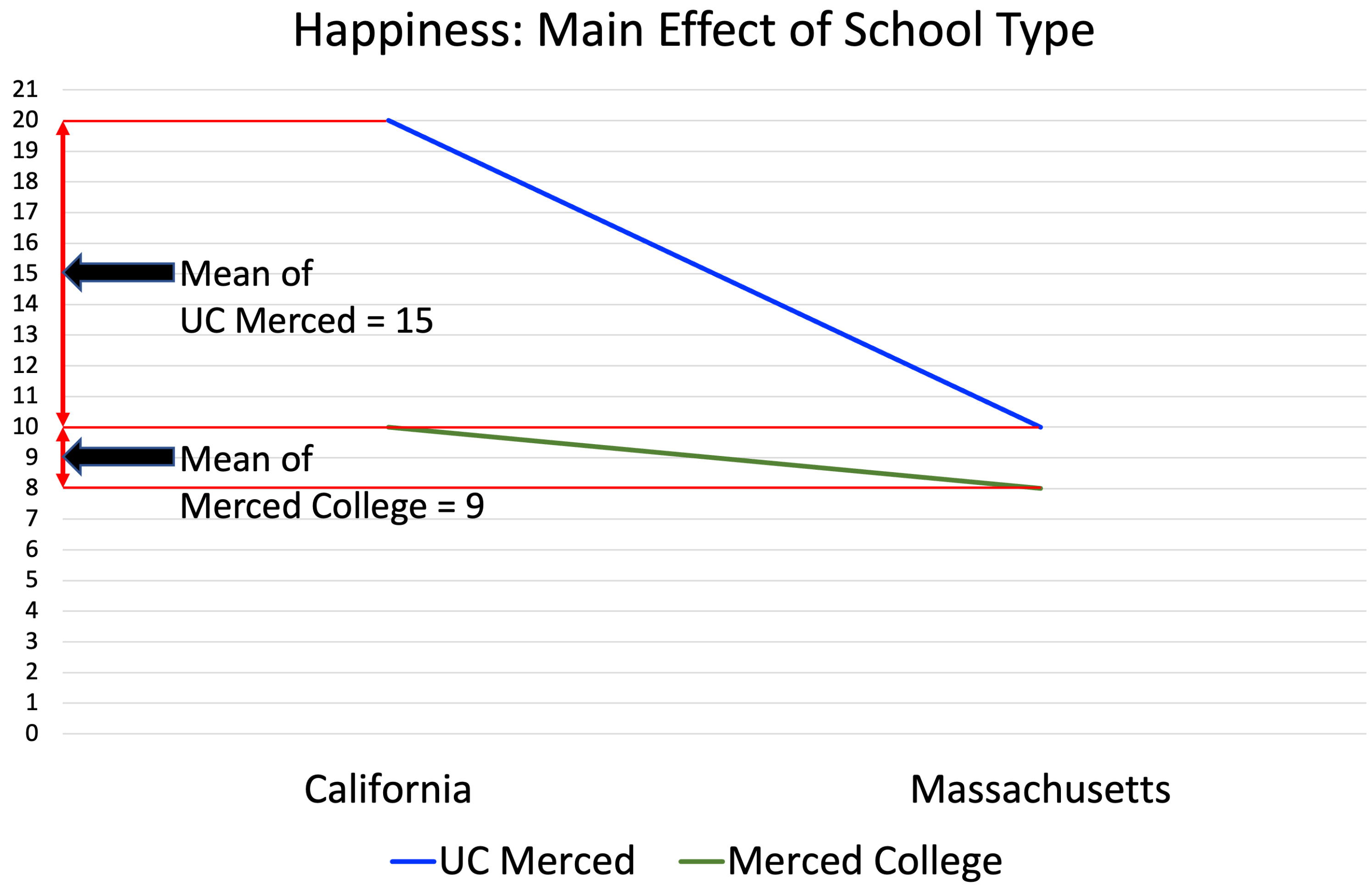
Happiness
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
California
Massachusetts
UC Merced
Merced College

Main effect



Happiness: Main Effect of Region
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Mean of
California = 15
Mean of
Massachusetts = 9
California
Massachusetts
UC Merced
Merced College

Main effect



Happiness: Main Effect of School Type
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Mean of
UC Merced = 15
Mean of
Merced College = 9
California
Massachusetts
UC Merced
Merced College

Interaction effect

If an effect of a single factor differs across levels of the other factors

Mean differences at each level of one factor change across the levels of the other factors

Whether the mean differences in happiness between two levels of school type (or region) change across the levels of region (or school type)

Differences in differences $\longrightarrow$ unparalleled lines

N-way ANOVA $\longrightarrow$ There are $2^N - N - 1$ interaction effects

Two-way ANOVA $\longrightarrow$ One interaction effect!

Interaction effect

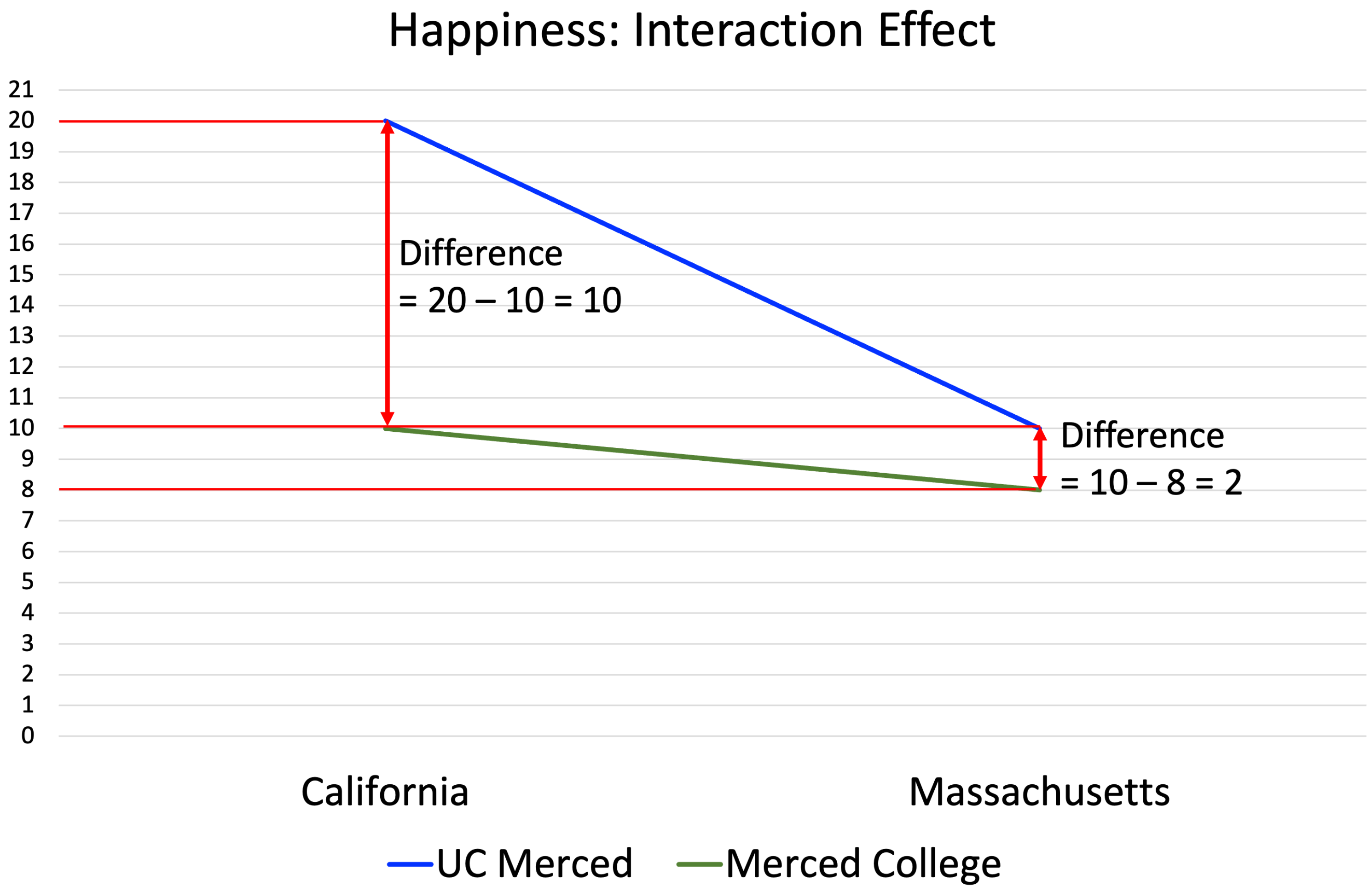
If an effect of a single factor differs across levels of the other factors

Mean differences at each level of one factor change across the levels of the other factors

Differences in differences $\longrightarrow$ unparalleled lines

	UC Merced	Merced College	Mean
California	20	10	15
Massachusetts	10	8	9
Mean	15	9	12

Interaction effect



Note

Whenever we say there are whether main effect or interaction effect, this does not necessarily mean these effects are statistically significant

It is enough for us to suspect there might be

To check statistical significance
$\longrightarrow$ p-value < α $\longrightarrow$ Observed F value > Critical F value

Descanso

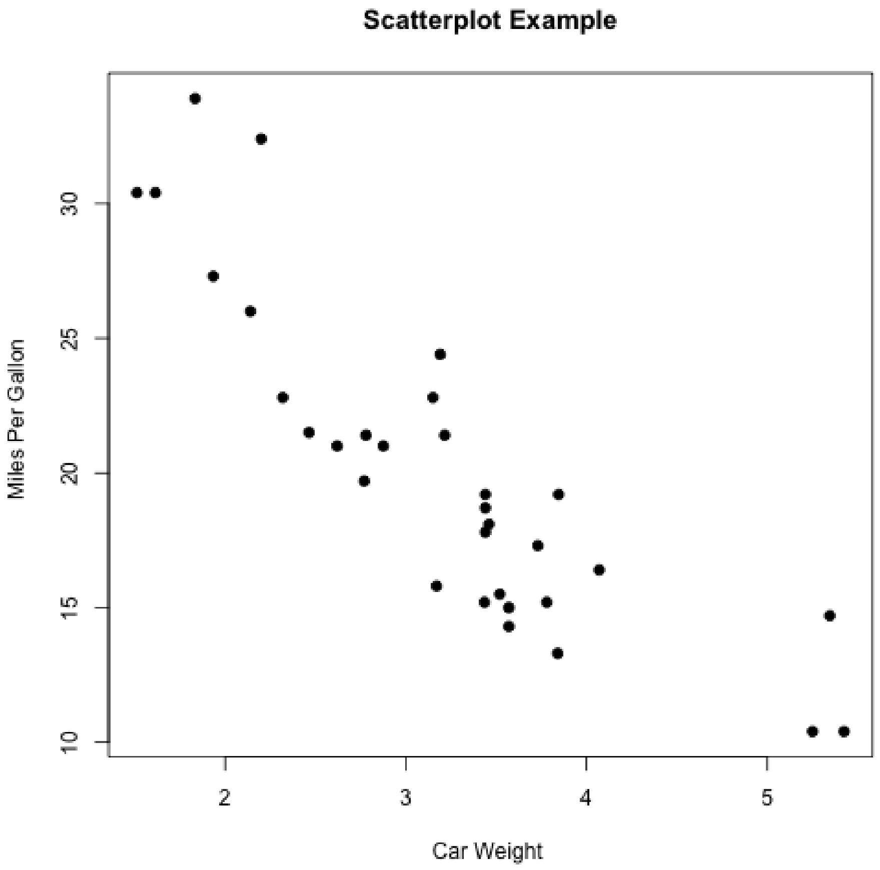
Stretch yourself :)

Correlation

Do you remember scatterplot?

Shows how two variables are related

```
plot(mtcars$wt, mtcars$mpg,
   main="Scatterplot Example",
   xlab="Car Weight ",
   ylab="Miles Per Gallon ", pch=19)
```



Correlation

Goal

To describe a relationship between two variables (i.e., how two variables are related)

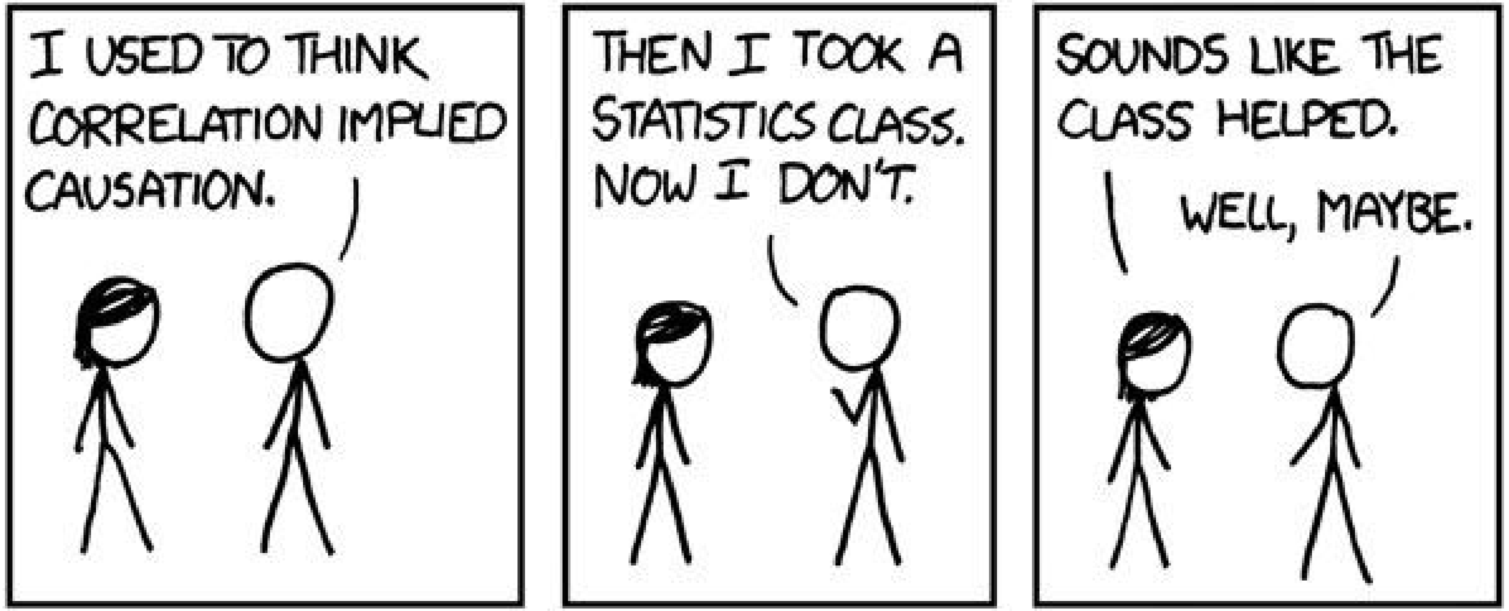
Real-life examples

The longer time you drive, the more gasoline your will need

The taller the person is, the heavier the person is

Correlation

Does not mean causation



XKCD

Correlation

Two key ideas

Direction $\longrightarrow$ positive vs. negative

Strength $\longrightarrow$ perfect vs. strong vs. weak vs. no

Need to know how to interpret

Scatterplot

Correlation coefficient

Correlation

Direction

Positive $\longrightarrow$ As one variable increases, the other increases

Negative $\longrightarrow$ As one variable increases, the other decreases

Strength

How much two variables covary $\longrightarrow$ Degree to which the data fall on a straight line

Correlation

Scatterplot

One variable is on the x-axis whereas the other variable is on the y-axis

Correlation coefficient

Correlation coefficient ragnes from -1 (perfect negative) to 1 (perfect positive)

Correlation

Direction and strength

Correlation

Pearson's correlation coefficient

A statistic that describes the direction and the strength of the linear relationship

$$r = \frac{SS_{XY}}{\sqrt{SS_X SS_Y}}$$

SS_{XY} is the sum of products of deviations for two variables X and Y

SS_X is the sum of squares for one variable X

SS_Y is the sum of squares for the other variable Y

Hypothesis teseting - Correlation

1. State the null and alternative hypothesis
2. Choose your α-level of significance
3. Determine the degrees of freedom
4. Locate the critical r value
5. Calculate Pearson's r
6. Compare observed and critical r value $\longrightarrow$ Reject the null hypothesis if $|r_{obt}| > |r_{crit}|$

Hypothesis teseting - Correlation

State hypotheses

$H_0 : r = 0$
The population correlation coefficient between time and satisfaction is zero.

$H_1 : r \neq 0$
The population correlation coefficient between time and satisfaction is different from zero.

Choose your α-level of significance

Let's use two-tailed test with the α-level of 0.05

Hypothesis teseting - Correlation

Determine the degrees of freedom

$df = n - 2$ where n refers to the sample size

Hypothesis teseting - Correlation

Locate the critical r value from the correlation table

	Level of Significance for One-Tailed Test			
	.05	.025	.01	.005
	Level of Significance for Two-Tailed Test			
$df = n - 2$	.10	.05	.02	.01
1	.988	.997	.9995	.99999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.549	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708

Hypothesis teseting - Correlation

Pearson's correlation coefficient $\longleftarrow$ standardized covariance

Formula is as follows:

$$r = \frac{SS_{XY}}{\sqrt{SS_X SS_Y}}$$

SS_{XY} is the sum of products of deviations for two variables X and Y

SS_X is the sum of squares for one variable X

SS_Y is the sum of squares for the other variable Y

Effect size and R-squared

Effect size (r)

0.1 (small), 0.3 (medium), 0.5 (large)

Coefficient of determination (r^2)

The proportion of Y explained by X

Q&A session

Any questions?

Before you go home...

Lab materials are available at

https://github.com/IhnwhiHeo/PSY010

Any questions or comments?

Office hours or my email

Thanks! Have a good one!