



Analysis of Psychological Data

Lab 11. We Go Together, Even Beyond the Horizon: Correlation, Regression, and Metascience

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

Assignments

Extra credit assignments 2 is due April 29 (Friday)

Homework 6 is due May 3 (Tuesday)

Final exam

Exam 3 is on May 5 (Thursday)

Scantron!



Some announcements

Scantron form

ParScore® STUDENT ENROLLMENT SHEET

INSTRUCTOR: Only write your lab section number or TA's name in this area
CLASS: _____
HOUR/DAY: _____

SCANTRON Recruit Form No. F-288-PAR-L www.scantron.com 800-722-6776

DIRECTIONS

- MAKE DARK MARKS
- ERASE COMPLETELY TO CHANGE
- EX.

ID NUMBER: 100098765

PHONE NUMBER: LEAVE THIS BLANK

LAST NAME: CARLOS FIRST NAME: JOSE M.I.: LEAVE THIS BLANK

ParScore® TEST FORM

ID NUMBER: 100098765 TEST FORM: 001

DIRECTIONS

- MAKE DARK MARKS
- ERASE COMPLETELY TO CHANGE
- EX.

You will have 50 multiple choice questions. Answer them here.

TEST FORM: 001

This information will be shown on the first page of your exam sheet

TEST FORM: 001

1 T F 101 T F 191 T F 281 T F 371 T F 461 T F 551 T F 641 T F 731 T F 821 T F 911 T F 1001 T F 1091 T F 1181 T F 1271 T F 1361 T F 1451 T F 1541 T F 1631 T F 1721 T F 1811 T F 1901 T F 2001 T F 2091 T F 2181 T F 2271 T F 2361 T F 2451 T F 2541 T F 2631 T F 2721 T F 2811 T F 2901 T F 3001 T F 3091 T F 3181 T F 3271 T F 3361 T F 3451 T F 3541 T F 3631 T F 3721 T F 3811 T F 3901 T F 4001 T F 4091 T F 4181 T F 4271 T F 4361 T F 4451 T F 4541 T F 4631 T F 4721 T F 4811 T F 4901 T F 5001 T F

FEED THIS DIRECTION

LEAVE THIS BLANK

SCANTRON EM-F-288-PAR-L-034421 © 2007 Scantron Corporation. ALL RIGHTS RESERVED.



What are we going to do?

Recap to give you a big picture

Correlation

Regression

... and Metascience

Do it together



It's been a long, long semester

**READY
FOR
PSY010!**

**NO MORE
STATS...
PLEASE**





It's been a long, long semester

What you've learned...

Fundamentals (variables, statistic/parameter), descriptive vs. inferential statistics, measurement scale (nominal, ordinal, interval, ratio), data visualization (bar chart, pie chart, histogram, scatterplot), central tendency (mean, median, mode), variability (range, variance, standard deviation), normal distribution, z-score, central limit theorem, standard error, (frequentist) null hypothesis significance testing, significance level, p-value, z-test, test-statistic, effect size (Cohen's d, eta-squared, partial eta-squared, omega-squared), z-table, t-test (one-sample, independent-sample, related-sample), t-table, degrees of freedom, type I and type II error, statistical power, one-way ANOVA (sum of squares, mean square, F-value), F-table, two-way ANOVA, main effect, interaction effect, correlation, correlation table, coefficient of determination, regression, slope, intercept

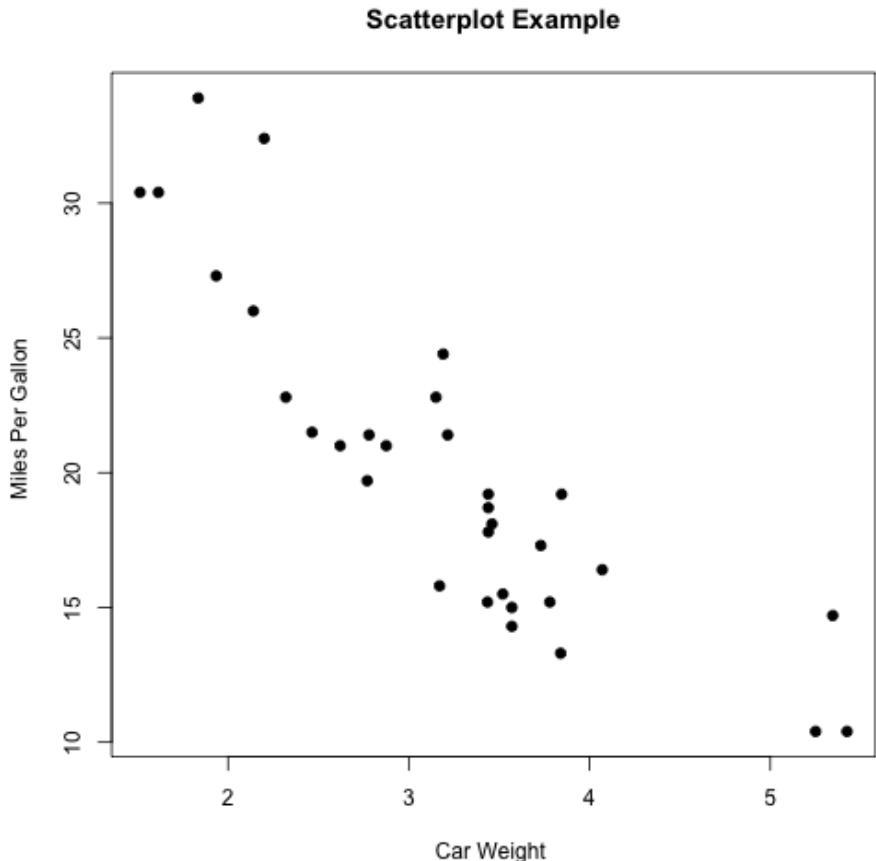


Correlation

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

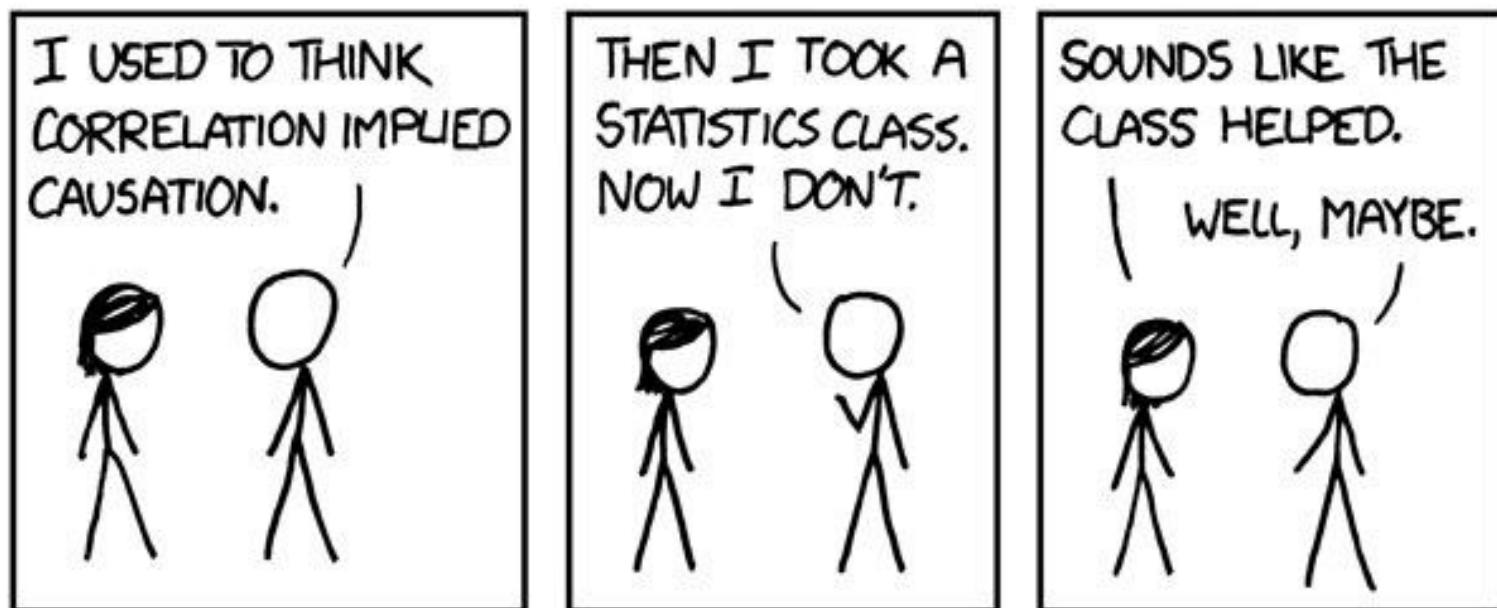
The longer time you drive, the more gasoline your will need

The taller the person is, the heavier the person is



Correlation

Does not mean causation





Correlation

Two key ideas

Direction → positive vs. negative

Strength → perfect vs. strong vs. weak vs. no

Need to know how to interpret

Scatterplot

Correlation coefficient



Correlation

Direction

Positive → As one variable increases, the other increases

Negative → As one variable increases, the other decreases

Strength

How much two variables covary → 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 ranges 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 → Reject the null hypothesis if $|r_{obt}| > |r_{crit}|$



Hypothesis teseting - Correlation

State hypotheses

$$H_0 : r = 0$$

The population correlation coefficient is zero.

$$H_1 : r \neq 0$$

The population correlation coefficient is different from zero.

Choose your α -level of significance

We usually use 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

df = n - 2	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 testing - Correlation

Pearson's correlation coefficient ← 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



Do it together - Correlation





Regression

Goal

To predict values of the dependent variable (Y) from values of the independent variable (X)

Real-life examples

If you drive 5 hours, how many gallons of gasoline you will need

If the person is 6 feet, how heavy the person would be



Regression

Regression equation

$$Y' = a + bX$$

Y' is a predicted value of the dependent variable Y

X is an individual value on the independent variable X

a is an intercept → a value of Y when X is 0

b is a slope → rate of change in Y when X changes by one-unit (*aka.* steepness)



Regression

Slope $\rightarrow b$

$$Y' = a + bX$$

Rate of change in Y when X changes by one-unit (*aka.* steepness of a regression line)

$$b = \frac{SS_{XY}}{SS_X}$$

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



Regression

Intercept $\rightarrow a$

$$Y' = a + bX$$

A value of Y when X is 0

$$a = \bar{Y} - b\bar{X}$$

\bar{Y} is the mean of the dependent variable Y

\bar{X} is the mean of the independent variable X



Hypothesis testing - Regression

State hypotheses

$$H_0 : \beta = 0$$

Regression slope is 0 in the population (no effect)

The independent variable predicts nothing about the dependent variable

$$H_1 : \beta \neq 0$$

Regression slope is different from 0 in the population (there is an effect)

The independent variable predicts something about the dependent variable

β is the population regression slope



Hypothesis testing - Regression

Choose your α -level of significance

We usually use the α -level of 0.05

Compute the test-statistic

z-test or t-test depending on whether you know the population standard deviation



Hypothesis testing - Regression

Make a decision

Refer to a z-table or t-table

Find a critical z-value or critical t-value

Reject the null hypothesis if

p -value < α -level

$$|z_{obt}| > |z_{crit}| \iff |z_{obt}| > |z_{crit}|$$



Do it together - Regression





Metascience

Recall that I kept emphasizing the problems of p-values... but why?



Metascience

The replication crisis in psychology

Only 36% of the significant findings (p-value below 0.05) was replicable

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

SCIENCE sciencemag.org

28 AUGUST 2015 • VOL 349 ISSUE 6251



Metascience

One of the problems pointed out was the p-value

The Earth Is Round ($p < .05$)

Jacob Cohen

After 4 decades of severe criticism, the ritual of null hypothesis significance testing—mechanical dichotomous decisions around a sacred .05 criterion—still persists. This

sure how to test H_0 , chi-square with Yates's (1951) correction or the Fisher exact test, and wonders whether he has enough power. Would you believe it? And would you



Metascience

A movement to redefine or ban the p-value

nature human behaviour

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature human behaviour](#) > [comment](#) > [article](#)

Comment | Published: 01 September 2017

Redefine statistical significance

[Daniel J. Benjamin](#)✉, [James O. Berger](#), ... [Valen E. Johnson](#)✉ [+ Show authors](#)

[Nature Human Behaviour](#) 2, 6–10 (2018) | [Cite this article](#)

139k Accesses | 964 Citations | 854 Altmetric | [Metrics](#)

We propose to change the default *P*-value threshold for statistical significance from 0.05 to 0.005 for claims of new discoveries.



Metascience

Stop using p-values!

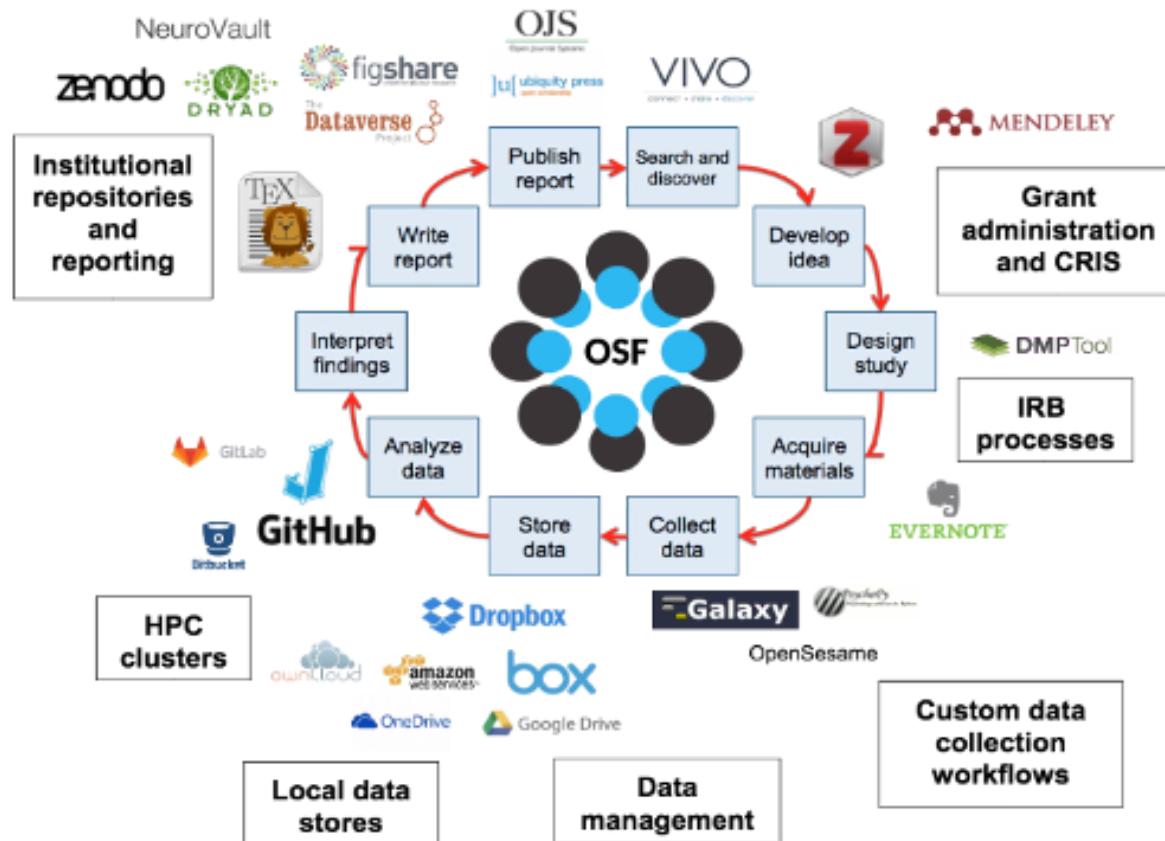
Prof. Dr. Rens van de Schoot (<https://www.youtube.com/watch?v=owb5A8IVSoo>)





Metascience

Open Science Movement





Concluding remark

What I want to convey to you...

Be critical in doing everything in your life

Be transparent in doing science

Enjoy your undergraduate life (jealous of you!)



Before we finish our semester...

Any **FINAL** questions or comments?





Thank you for the great semester :) Cheers!

