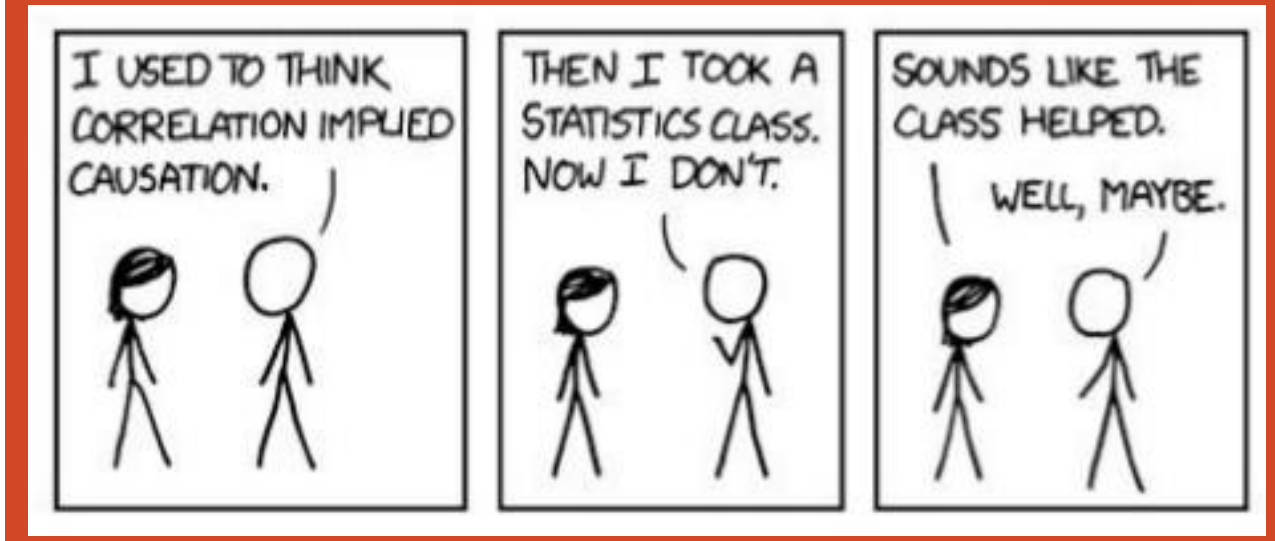


CH. 12 – Correlation

PY 221 Statistics & Research Methods I

Dr. Valenti

Outline for Ch. 12 - Correlation

1. **Review & overview of when to use correlation**
2. What are correlations, conceptually?
   - null and alternative hypotheses
   - calculation of Pearson's *r*, the *correlation coefficient*
   - interpretation of *r*
3. The coefficient of determination, R^2
4. Range restriction
5. Outliers
6. Correlation matrices

REVIEW

- What do ***t*-tests** allow you to do?
  - examine whether the apparent **differences** between **two groups** on an outcome variable are real differences (vs. produced by chance)
  - examine whether the apparent **changes** in people's scores across **two time points** are real changes (vs. produced by chance)

REVIEW

- What does **ANOVA** allow you to do?
  - examine whether apparent **differences** between **three or more groups** on an outcome variable are real differences (vs. produced by chance)
    - *One-way* ANOVA involves one predictor/IV
    - *Two-way* ANOVA involves two predictors/IVs

Comparison of the tests we've learned so far…

TEST	t-test	1-way ANOVA	2-way ANOVA
Type of *predictor*			
Type of *outcome*			
What does test look for?			
Null hyp.			

Comparison of the tests we've learned so far…

TEST	t-test	1-way ANOVA	2-way ANOVA
Type of *predictor*	1 qualitative/categorical, w/only 2 categories /levels (binary)	1 qualitative/categorical, with 3+ categories/levels	2 qualitative/categorical variables, with 2+ levels
Type of *outcome*	quantitative	quantitative	quantitative
What does test look for?			
Null hyp.			

Comparison of the tests we've learned so far…

TEST	t-test	1-way ANOVA	2-way ANOVA
Type of *predictor*	1 qualitative/categorical, w/only 2 categories /levels (binary)	1 qualitative/categorical, with 3+ categories/levels	2 qualitative/categorical variables, with 2+ levels
Type of *outcome*	quantitative	quantitative	quantitative
What does test look for?	Are there differences in means across groups?	Are there differences in means across groups?	Are there differences in means across groups of factor A? of factor B? And, do the differences across groups of A differ depending on groups of B?
Null hyp.	$\mu_1 = \mu_2$ *OR* $\mu_1 - \mu_2 = 0$	$\mu_1 = \mu_2 = \mu_3$ (assuming a 3-level predictor)	*Didn't cover in a formal way.*

Correlation: for when you want to assess whether and how two quantitative variables are related

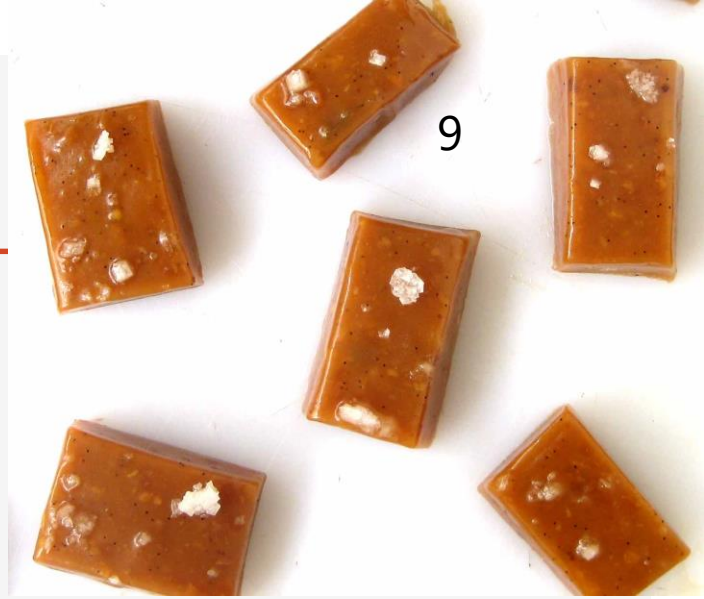
- What does **CORRELATION** allow you to do?
  - examine whether an apparent *linear relationship* between two quantitative variables is real (vs. produced by chance)
- requires pairs of scores for each participant

IMPORTANT NOTE about CORRELATION

*Only **two** variables can be examined in any one analysis*

*Both variables must be **quantitative***

Watch Commercials, Buy Candy?



Imagine that 5 people were asked how many advert-isements (commercials) for a certain candy they saw one week, and then researchers measured how many packets of that candy they purchased the next week . . .

Relevant research question: Is the # of candy commercials a person saw one week *related to* the # of candy packets they purchased the next week?

TABLE 6.1

Participant #	1	2	3	4	5	Mean	S
Adverts Watched	5	4	4	6	8	5.4	1.67
Packets Bought	8	9	10	13	15	11.0	2.92

Research questions that would involve correlation - examples

1. Is amount of Netflix you typically watch (measured in minutes per week) related to how much you exercise (in minutes per week)?
2. Is there a relationship between frequency of reading for pleasure and one's GPA (range 0 to 4)?
3. Do people who drink a lot of alcohol themselves tend to believe that most other people also drink a lot of alcohol, and vice-versa?
4. Is the amount of alcohol you consume associated with how happy you are, when happiness is measured on a 5-point scale?
5. **Write a research question about daily social media usage (i.e., how many minutes/day you spend using social media) that could be answered with a correlational analysis.**

*Only **two** variables can be examined in any one analysis*

*Both variables must be **quantitative***

Outline for Ch. 12 - Correlation

1. Overview of when to use correlation
2. **What are correlations, conceptually?**
   - **null and alternative hypotheses**
   - calculation of Pearson's *r*, the *correlation coefficient*
   - interpretation of *r*
3. The coefficient of determination, R^2
4. Range restriction
5. Outliers
6. Correlation matrices

RECALL FROM CHAPTER 7 on NHST . . .

Step 1: State 2 hypotheses about the unknown population

- **Null hypothesis, H_0**
  - *FOR STATISTICAL TESTS IN GENERAL*: In the unknown population, there is no change, no difference, or no relationship; IV has no effect on DV.

- **Alternative hypothesis, H_1**
  - *IN GENERAL:* In the unknown population, there *is* a change, there *is* a difference, or there *is* a relationship; IV does have an effect on DV.

  - Hypotheses for correlational studies will say something about the predicted **relationship** (or **association**).

Null and alternative hypotheses for candy ad study

Null (words): There is **no relationship** between number of candy advertisements watched and candy packets purchased the next week.

Null (symbols): $H_0: \rho = 0$

rho, the population parameter for *correlation*
(just like μ was the population parameter for the *mean*)

Alternative (words): There is **a relationship** between number of candy advertisements watched and candy packets purchased the next week.

Alternative (symbol): $H_1: \rho \neq 0$

We will continue to use *non-directional* alternative hypotheses

Null and alternative hypotheses – another example

- Suppose an instructor is interested in whether students' quiz scores are related to their paper scores. Do students with higher quiz scores also earn higher paper scores, and vice-versa?
- PRACTICE: state the null hypothesis in words and symbols

Student	Quiz Score Avg (X)	Paper score (Y)
Leslie	42	48
Jennifer	50	39
Ruthann	48	41
Tim	47	29
etc.		

Null and alternative hypotheses for quiz & paper study

Null (words): There is **no relationship** between students' quiz scores and paper scores

Null (symbols): $\boldsymbol{H_0: \rho = 0}$

Alternative (words): There is **a relationship** between students' quiz scores and paper scores

Alternative (symbol): $\boldsymbol{H_1: \rho \neq 0}$

Outline for Ch. 12 - Correlation

1. Overview of when to use correlation
2. **What are correlations, conceptually?**
   - null and alternative hypotheses
   - **calculation of Pearson's *r*, the *correlation coefficient***
   - interpretation of *r*
3. The coefficient of determination, R^2
4. Range restriction
5. Outliers
6. Correlation matrices

Calculating the correlation coefficient, Pearson's *r*

r is a test statistic (like *t, F*) **and** a descriptive statistic (like *x-bar*)

- It's also an effect size!

Big picture view of calculating r

We need to see whether, as one variable increases, the other variable
increases in a consistent way?
decreases in a consistent way? or
stays the same/does nothing consistent?

To calculate *r,* the first step is to calculate the covariance, *cov(x,y)*

Review from Ch. 3

DEVIATION

Find deviation for each score

(x minus *x-bar*)

SS (sum of sq)

Square each deviation, & then sum the squared deviations

VARIANCE (S^2)

Find "average" of the squared deviations

Divide SS by ***N* - 1**

Variance vs. Covariance

- The variance tells us: how much scores of a ***single*** *variable* deviate from that variable's mean
- The covariance tells us: how much scores on *two* variables (X & Y) differ from *their respective means*.
  - For each participant, if their scores on X and Y deviate from X's and Y's means by the *same* amount – and if this is the case across all your Ps – the two variables are likely to be *related*.

Watch Commercials, Buy Candy?

Imagine that 5 people were asked how many advertisements for a certain candy they saw one week, and then researchers measured how many packets of that candy they purchased the next week . . .

TABLE 6.1

Subject	1	2	3	4	5	Mean	S
Adverts Watched	5	4	4	6	8	5.4	1.67
Packets Bought	8	9	10	13	15	11.0	2.92

$$\text{Variance } (s^2) = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{N - 1}$$

$$Cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

The covariance is, broadly speaking, a measure of the **size of the relationship** between X and Y, how much of Y's variation is associated with X's variation