Lecture 9: Statistical Analysis

CS 181W

Fall 2022

Review: Types of studies

- **Interviews:** conversations with individuals
- **Focus groups:** discussions with groups
- **Surveys:** asynchronous questions

- **Experimental Studies:** randomized multi-condition studies
- **Usability Testing:** observations of tool use
- **Cognitive Walkthrough:** expert evaluation

- **Diary Studies:** contemporary record of real-world behavior
- **Observational Studies:** records of behavior in the wild

- Mixed-methods studies

Types of Studies

- **Descriptive investigations:** construct an accurate description of reality
  - typical methods: interviews, surveys, diary studies, observational studies

Types of Data

Quantitative Data

- Things you can measure with numbers
- Must have an order and consistent intervals
- Two types:
  - **Discrete:** things you can count with integers (e.g., how many clicks it takes)
  - **Continuous:** things you can measure precisely with floats (e.g., how long it takes)

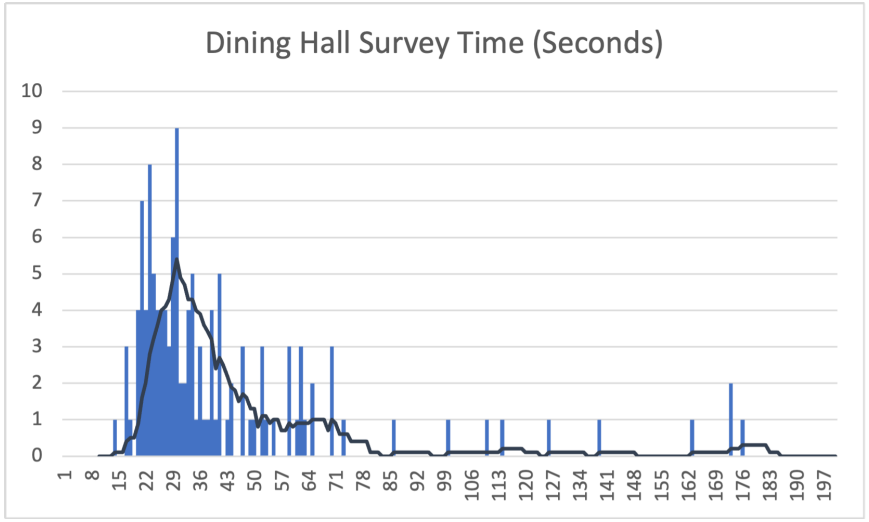
Categorical Data

- Groups/categories
- Either no order or no consistent interval
- Two types:
  - **Nominal:** no order (e.g., favorite dining hall)
  - **Ordinal:** order but no consistent interval (e.g., Likert scale for satisfaction)

Descriptive Data Examples

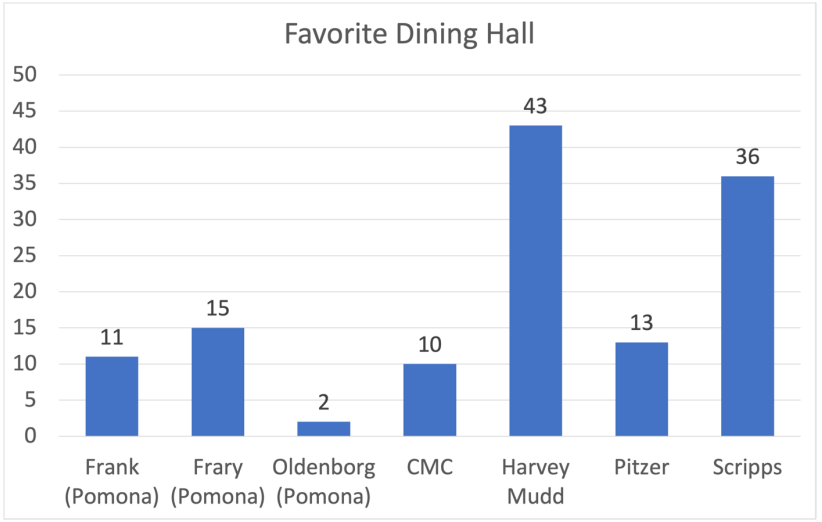
Quantitative Data

- Time taken to complete dining hall survey (continuous)



Categorical Data

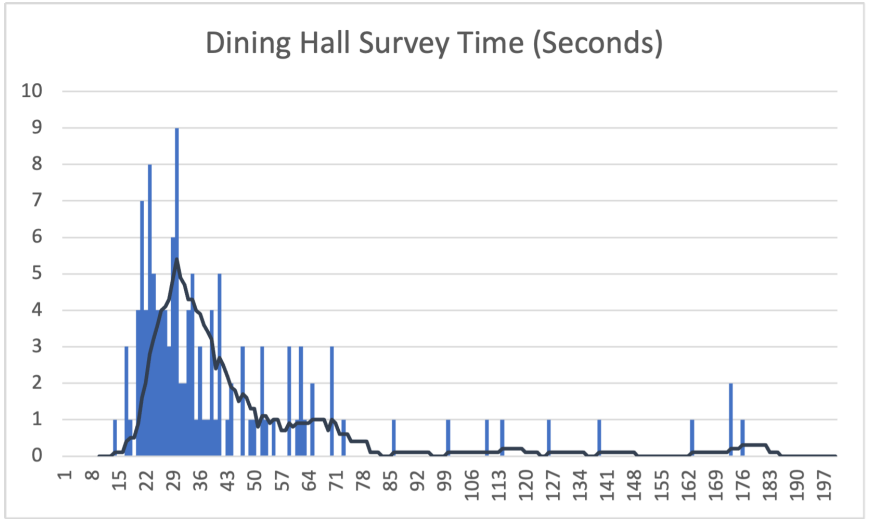
- Favorite dining hall (nominal)



Describing Quantitative Data

Quantitative Data

- Time taken to complete dining hall survey (continuous)



- Visualizations
  - Histogram
  - Smoothed histogram
  - Kernel density estimation
- Measures of central tendency
  - mean
  - median
  - mode
- Measures of spread
  - range
  - variance
  - standard deviation

Types of Studies

- **Descriptive investigations:** construct an accurate description of reality
  - typical methods: interviews, surveys, diary studies, observational studies
- **Relational investigations:** identify relations between multiple factors
  - typical methods: surveys, observational studies
- **Experimental investigations:** identify causal effects between factors
  - experimental studies

Types of Variables

Independent Variables

- factors of interest that might "cause" observed effects
- controlled variables
  - condition (e.g., design of mechanism)
- attributes
  - age, computer experience, education, culture, motivation

Dependent Variables

- outcome or effect that is measured in a study
- e.g., time, success, accuracy, satisfaction, ease

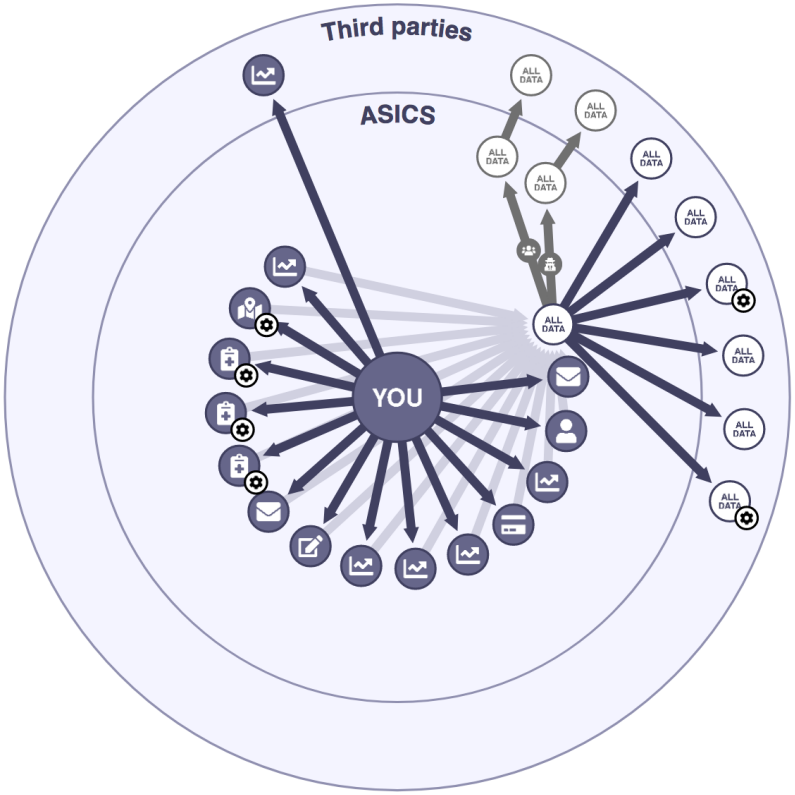
Statistical Tests

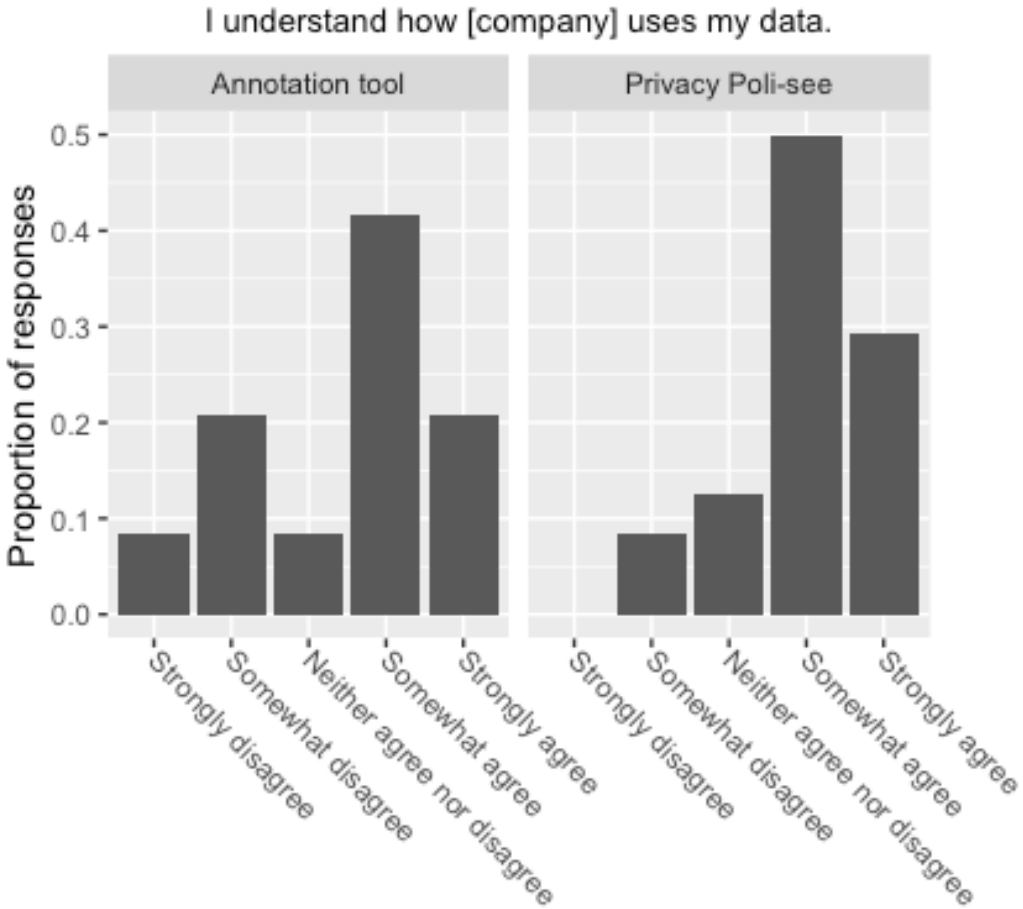
		Dependent Variable	
		Categorical	**Quantitative**
Independent Variable	**Categorical**	Chi-Square Test Fisher's Exact Test	t-Test (paired t-Tests) ANOVA (RM ANOVA)
	Quantitative	Logistic Regression	Correlation Linear Regression

Hypothesis Testing

Categorical Independent Variable

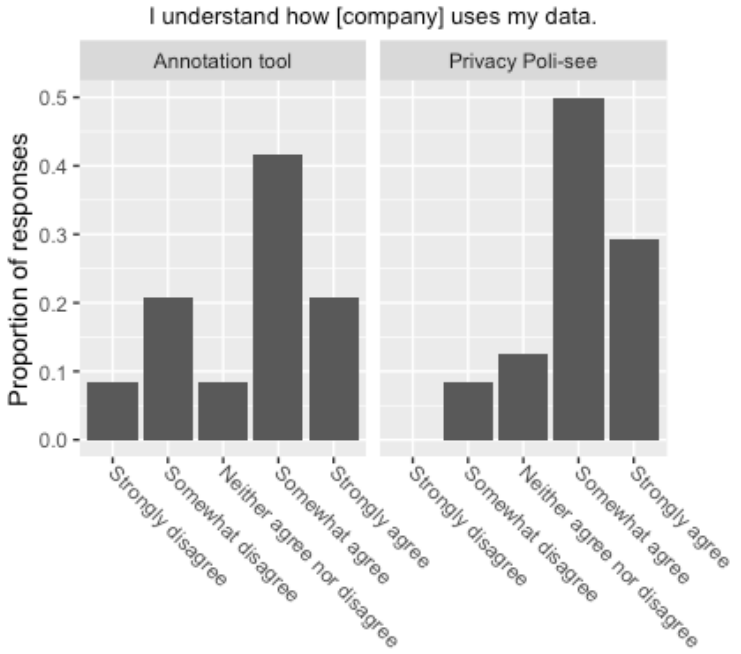
- Different distributions of your dependent variable for each possible value of your (categorical) independent variable





Hypotheses

- **Null hypothesis (H_0)**
  - There is no difference (in distributions)
- **Alternative hypothesis (H_1)**
  - There is a difference



You generally either find evidence in support of the alternative hypothesis or do not find evidence in support of the alternative hypothesis.

Why significance tests?

Consider the following statements:

1. Mike's height is 6'2". Mary's height is 5'8". So Mike is taller than Mary.

2. The average height of three males (Mike, John, and Ted) is 5'5". The average height of three females (Mary, Rose, and Maria) is 5'10". So, on average, females are taller than males.

Why significance tests?

- Since we only sample a subgroup of the population, significance tests calculate confidence that results generalize to the entire population

- p-value is probability of getting the observed distributions if the different samples are drawn from the same distribution (i.e., if the null hypothesis is true)

- $p < .05$ means 95% probability that result applies to the entire population

What statistical tests?

		Dependent Variable	
		Categorical	**Quantitative**
Independent Variable	**Categorical**	Chi-Squared Test Fisher's Exact Test	
	Quantitative		

Hypothesis Testing

Chi-Squared Test

- H_0: Pomona students and off-campus students are equally likely to prefer a Pomona Dining Hall

Favorite	Pomona Dining Hall	Other Dining Hall
Pomona Students	21	80
Other 5C Students	8	23

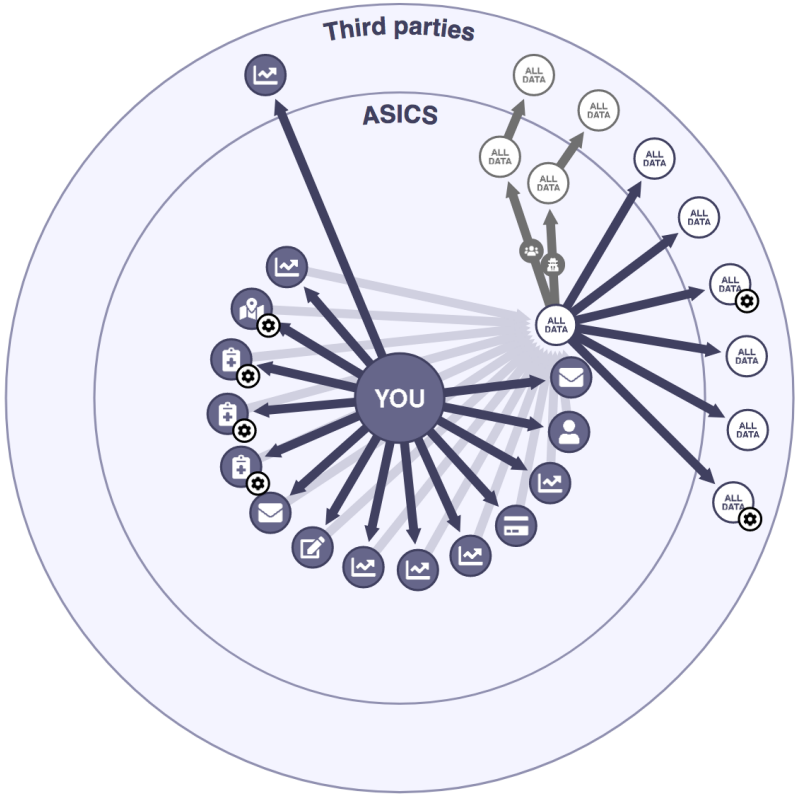
$$p = .732$$

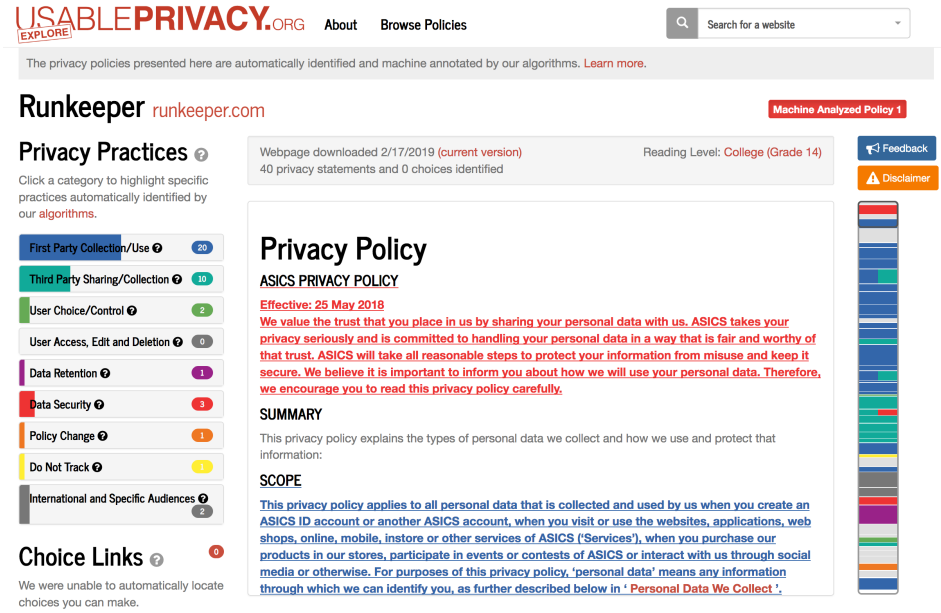
- H_0: Pomona students and off-campus students are equally likely to most frequently eat at a Pomona Dining Hall

Eat Most Often	Pomona Dining Hall	Other Dining Hall
Pomona Students	95	6
Other 5C Students	2	29

$$p < .001$$

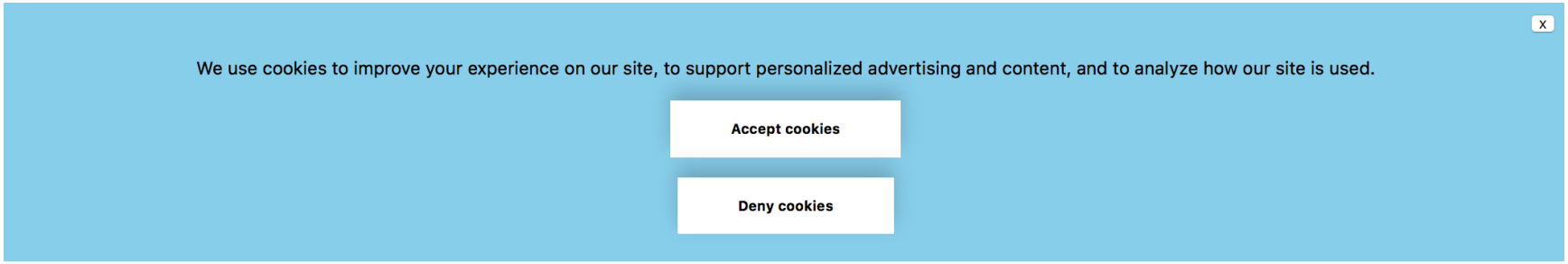
Example: Categorical-Categorical

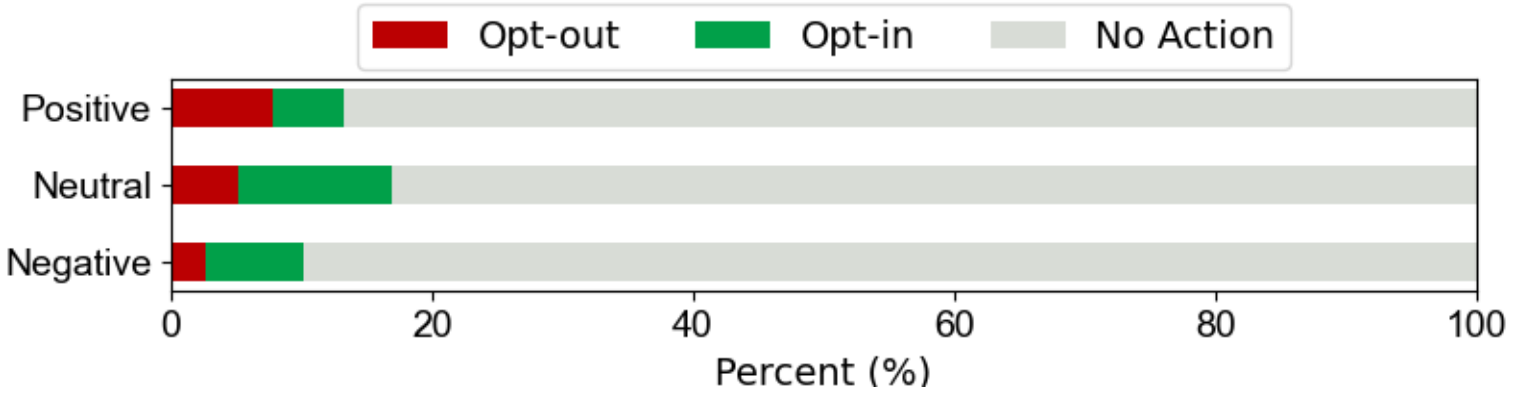




- H_0: There is no difference in in how many people answer factual questions correctly between Polisee and annotated privacy policies

Example: Categorical-Categorical





- H_0: There is no difference in in how many people opt-out of cookies between cookie banners with positive, neutral, and negative framings.

What statistical tests?

		Dependent Variable	
		Categorical	**Quantitative**
Independent Variable	**Categorical**	Chi-Squared Test Fisher's Exact Test	t-Test (paired t-Tests) ANOVA (RM ANOVA)
	Quantitative		

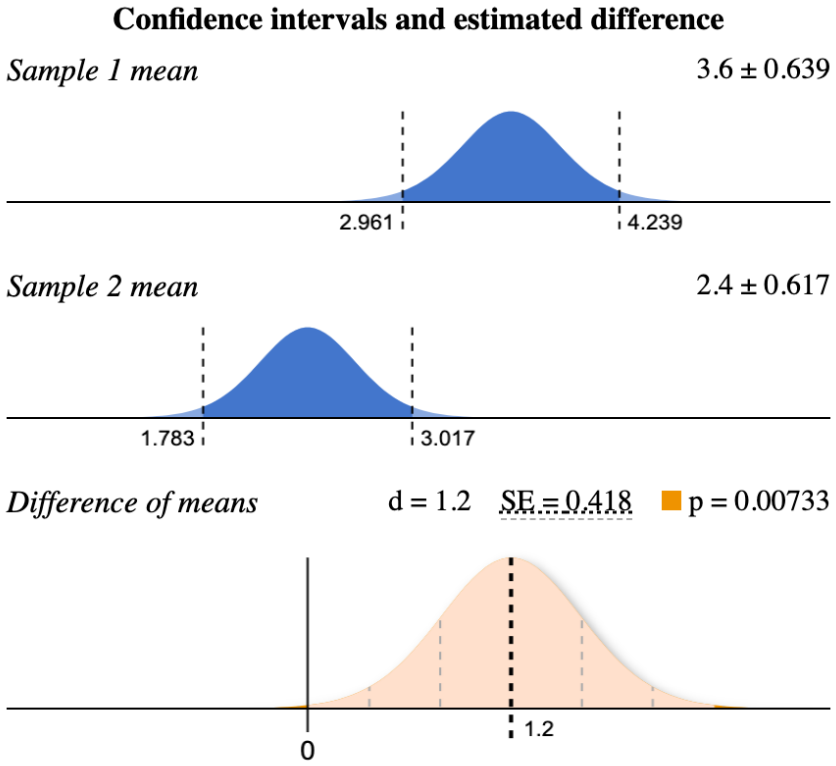
Hypothesis Testing

Between group

Within group

t-test and ANOVA

- H_0: There are no differences in the **means** between groups.
  - 2 groups: T-test
  - 3+ groups: ANOVA
- **Assumptions:**
  - **Simple random sample**
  - **Independent observations**
  - **Normal distribution**
  - **Homogeneity of variance**



*If you don't have a normal distribution, try a Mann-Whitney U test

Example: Categorical-Quantitative(?)

Technical Version

How comfortable would you be using a product or service if the data use policy says

"We use **pixel tags to collect information**"?

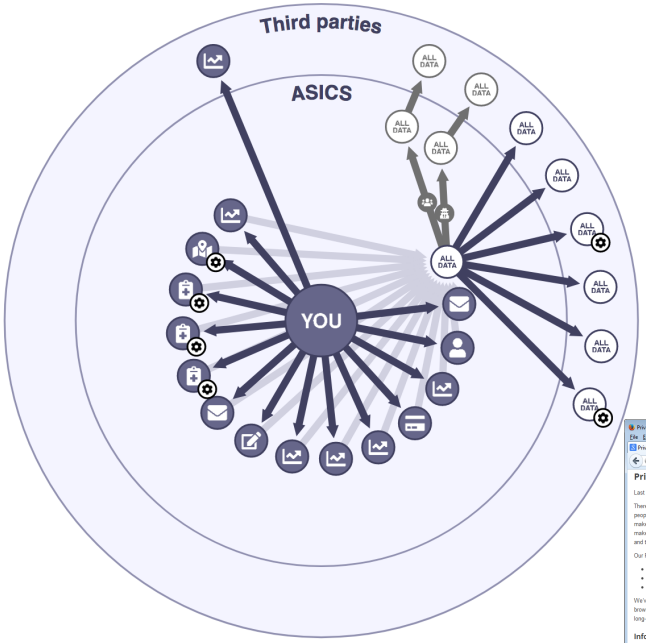
Explanatory Version

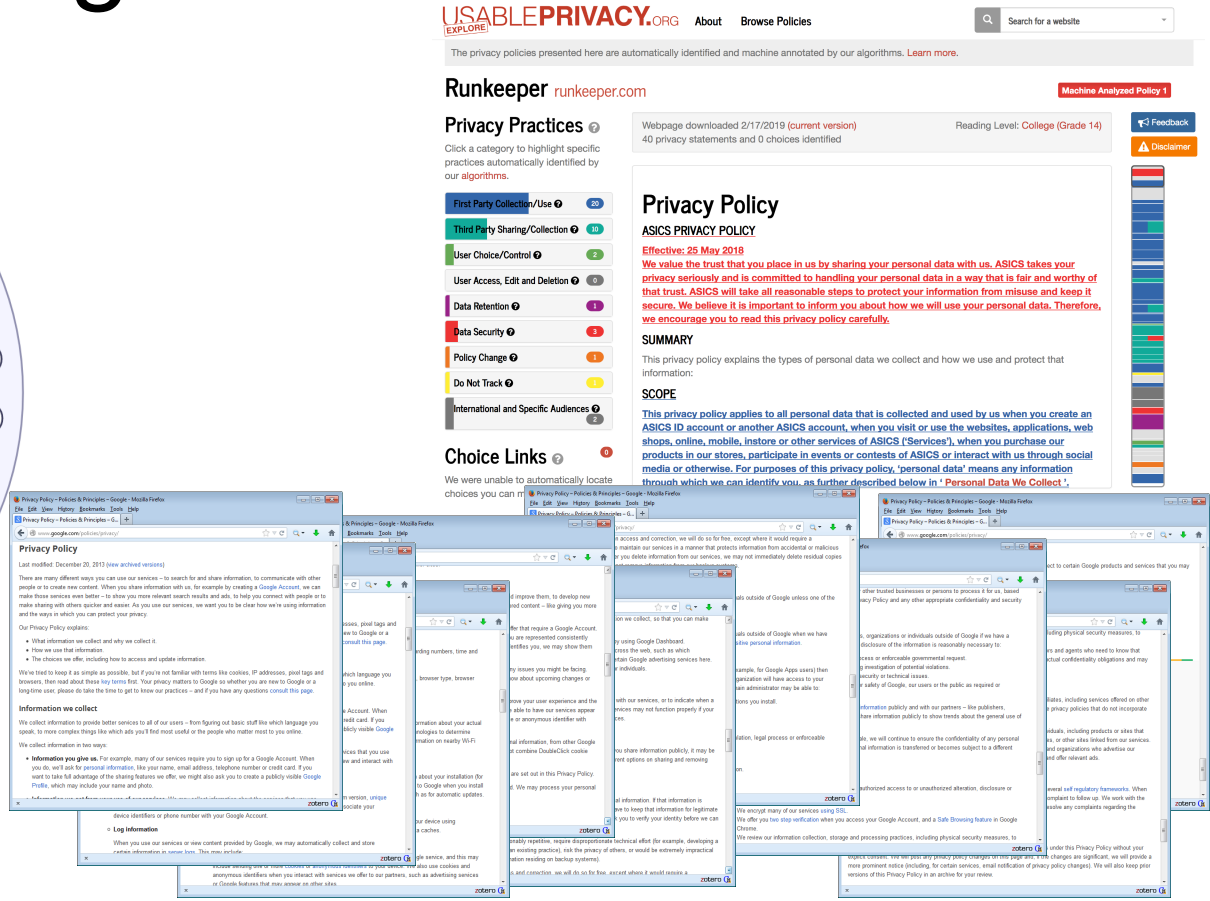
How comfortable would you be using a product or service if the data use policy says

"We **monitor user behaviors across different websites using invisible images** and other website elements"?

- H_0: There is no difference in in comfort level between technical language and explanatory language.

Example: Categorical-Quantitative





- H_0: There is no difference in timing distribution for how long it takes people to answer factual questions between Polisee and annotated privacy policies

Statistical Tests

		Dependent Variable	
		Categorical	**Quantitative**
Independent Variable	**Categorical**	Chi-Square Test Fisher's Exact Test	t-Test (paired t-Tests) ANOVA (RM ANOVA)
	Quantitative	Logistic Regression	Correlation Linear Regression

Relationship Testing

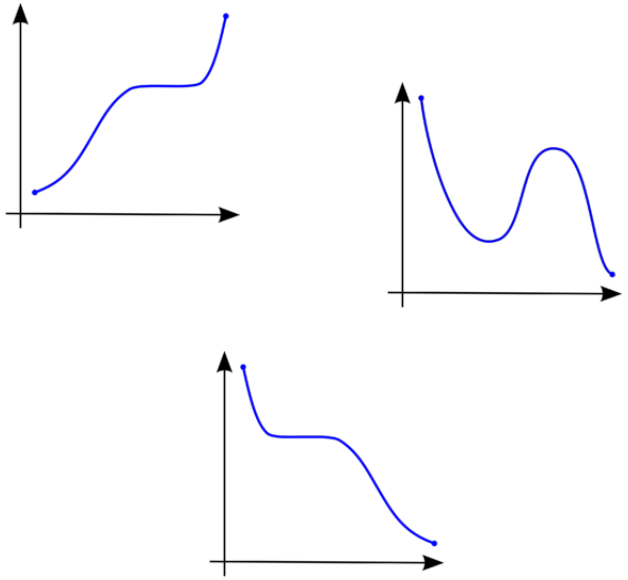
Pearson correlation

- Looks for a linear relationship between variables
  - Is the number of viruses on a unicorn's computer linearly related to how much time it spends watching videos of cats online?

- **Assumptions:**
  - Requires that both variables be normally distributed
  - Linearity
  - Homoscedasticity

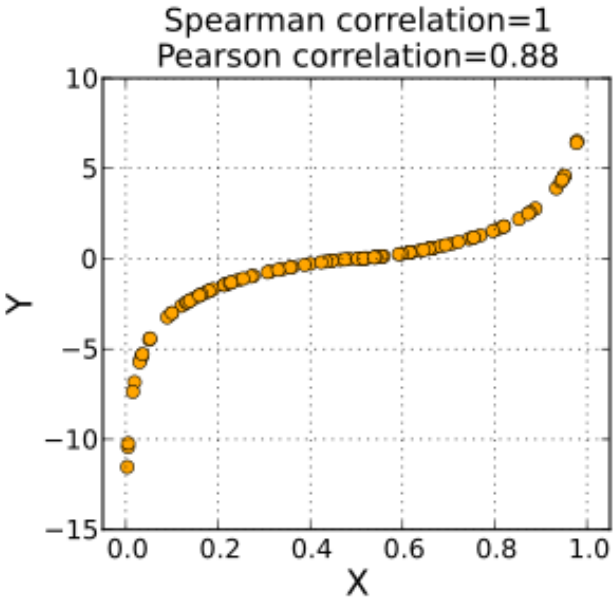
Spearman's rank correlation (Spearman's ρ)

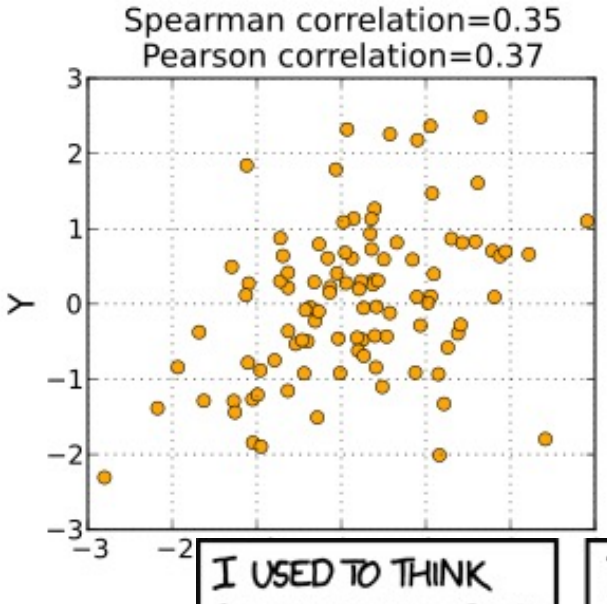
- More generic correlation test
  - Allows for non-linear relationships.
  - Checks for monotonicity
- **Assumption:**
  - Data should be at least ordinal

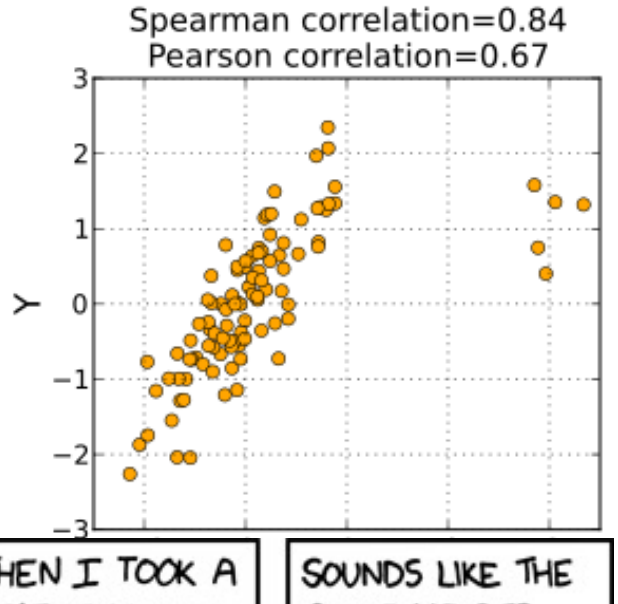


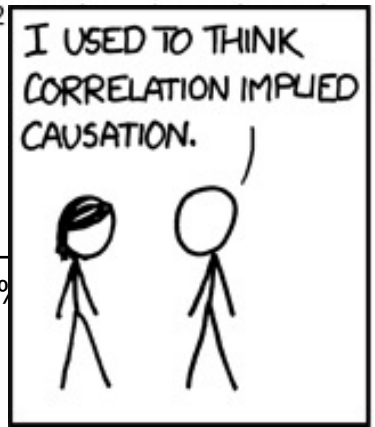
Images from: https://en.wikipedia.org/wiki/Monotonic_function

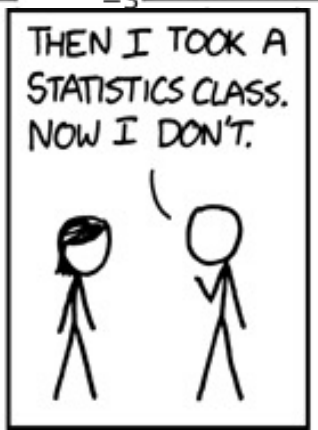
What Does Correlation Look Like?











Images from: https://en.wikipedia.org/wiki/Spearman%

Limitations of Correlation Analysis

- only compares two variables
- doesn't make predictions

Regressions

- model relationship between 1 or more independent variables and a dependent variable

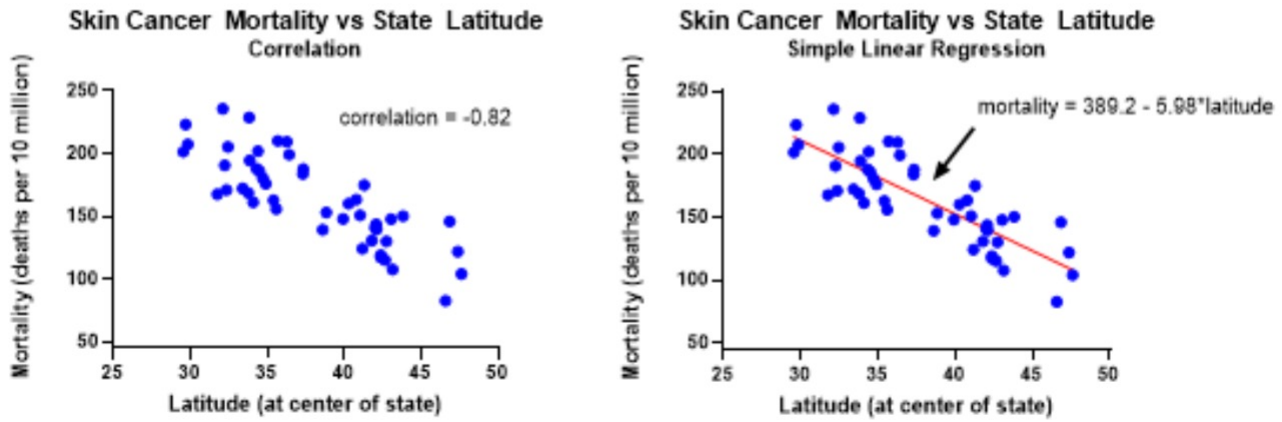
- different types of regression for different types of dependent variable:
  - Continuous outcome: linear regression
  - Binary outcome: logistic regression
  - Ordinal outcome: ordinal / ordered regression

- linear regression defines a linear function to predict the dependent variable

Correlation vs Regression

- There's some overlap in cases where you would use correlation vs cases where you would use regression

Correlation quantifies the direction and strength of the relationship between two numeric variables, X and Y, and always lies between -1.0 and 1.0. Simple linear regression relates X to Y through an equation of the form Y = a + bX.



https://www.graphpad.com/support/faq/what-is-the-difference-between-correlation-and-linear-regression/

Statistical Tests

		Dependent Variable	
		Categorical	**Quantitative**
Independent Variable	**Categorical**	Chi-Square Test Fisher's Exact Test	t-Test (paired t-Tests) ANOVA (RM ANOVA)
	Quantitative	Logistic Regression	Correlation Linear Regression

Exercise: Choosing a statistical test

1. I split unicorns into living in the forest or living in the clouds, and they each indicated whether or not they liked their new living environment. Does the assigned system impact whether or not they liked it?
2. I measured how much time unicorns spend at home. Do unicorns that live in the cloud spend more time at home than those that live in the forest?
3. I measured how long unicorns stay home and their magicalness score (1 to 100). Are these values related to each other?
4. I measured how long unicorns stay home and their magicalness score, age and weight. I'm curious what input factors (if any) impact the output.

“statistically significant effect” ≠ “big effect”

If you have a p-value that is really low, like $p<0.0001$, what that tells you is that it would be quite strange for your data to exist if the null hypothesis is true.

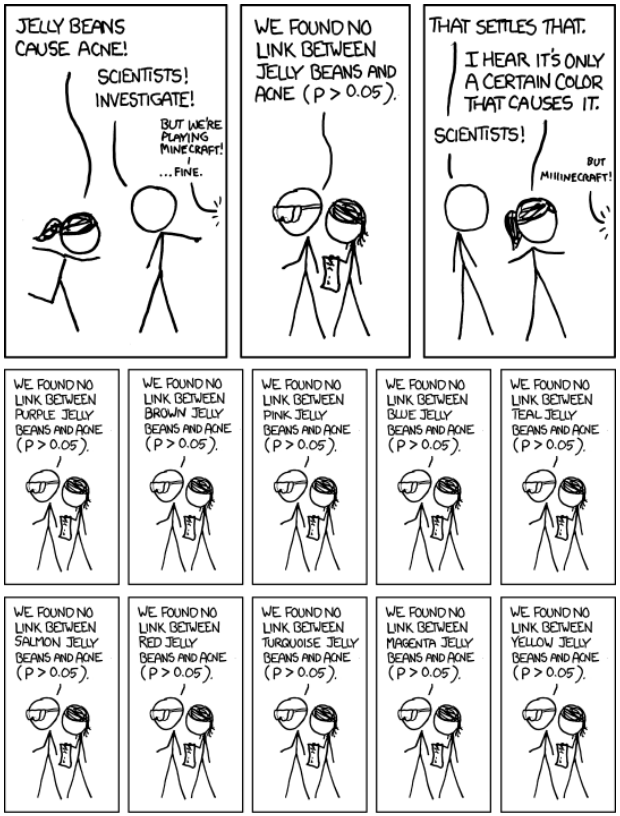
It does NOT tell you whether your effect size is small or large.

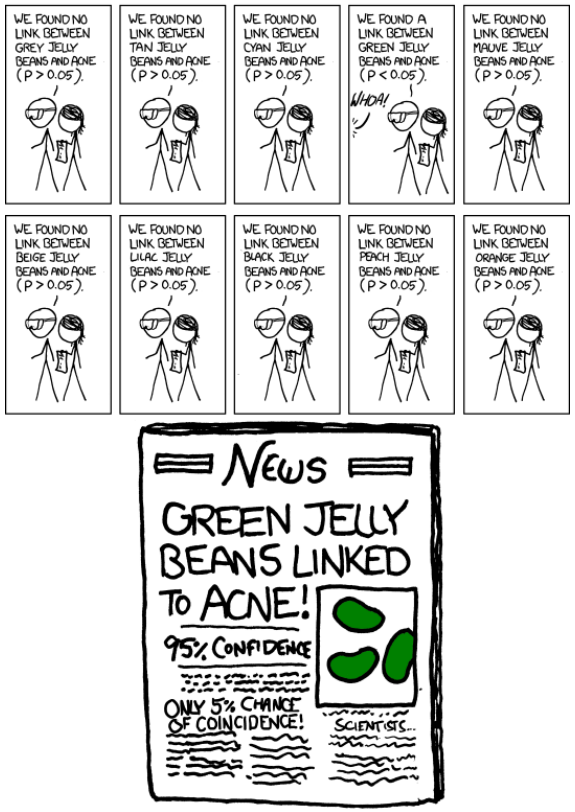
Or whether it’s meaningful or important.

Or whether one thing causes the other.

P Hacking

- Don't ➔





- Correct for **multiple hypothesis testing**!
  - Bonferroni
  - False Discover Rate (FDR) Control
  - Benjamini-Hochberg (BH) Threshold
  - ….

Exercise: Statistical Analysis

Statistical Analysis