07. Surveys; Analyzing Quantitative Data; Privacy on Social Media

Blase Ur, April 12th, 2017
CMSC 23210 / 33210

THE UNIVERSITY OF CHICAGO

Security, Usability, & Privacy Education & Research

Today's class

- Discuss survey design in the context of social networks
- Statistics whirlwind tour

Overall survey considerations

- How do we distribute it?
- How long should it be?
- One-time survey? Longitudinal survey?
- Will you use personalized data?
- What will participants learn?
  - What can we randomize to minimize this?
- Can we randomize the answer choices?

Are all answer options covered?

- With whom do you regularly share posts on social media?
  - Family
  - Friends
- Allow multiple answers?
- Include “other” option (write-in)?
- Do we care about previous use?

Are all answer options covered?

- I connect to Facebook over HTTPS
  - True
  - False
- What about “I don’t know”?

Are we biasing the answer?

- Strangers seeing your Facebook posts would cause you grave privacy concern.
  - Strongly agree
  - Agree
  - Neither agree nor disagree
  - Disagree
  - Strongly disagree

How will responses be distributed?

- For how long have you had Facebook?
  - Less than one day
  - Between one day and one week
  - More than one week

Should we force an answer?

- What gender are you? (* required)

  -Female -Male

- What gender are you?

  -Female -Male -I prefer not to answer

- With what gender do you identify?

  -Female -Male -Non-binary -I prefer to self-describe___ -I prefer not to answer

Likert-scale data?

- Respond to the following statement: Unicorns are amazingly magical.
  - 7: Strongly agree
  - 6: Agree
  - 5: Mildly agree
  - 4: Neutral
  - 3: Mildly disagree
  - 2: Disagree
  - 1: Strongly disagree

Likert-scale data?

- Which of the following best describes how you perceive unicorns to be?
  - 7: Very magical
  - 6: Magical
  - 5: Somewhat magical
  - 4: Neither magical nor unmagical
  - 3: Somewhat unmagical
  - 2: Unmagical
  - 1: Very unmagical

What demographics do we collect?

- Tech expertise, age, domain knowledge, gender, location, employment, etc.
- Don't ask people to self-rate expertise
  - Ask questions with concrete answers
  - e.g., Have you earned a degree in, or held a job in, computer science, IT, or…
  - Include a knowledge test if you want to know about expertise
- Consider why you are collecting this info

Creative survey designs

- Quantifying the invisible audience in social networking sites
  - M.S. Bernstein, E. Bakshy, M. Burke, B. Karrer. Quantifying the invisible audience in social networks. In *Proc. CHI 2013*.
- Compared actual data about what people saw (by working with Facebook) to survey questions about what they expected

Look through survey questions

Statistics!

- The main idea and building blocks
- Major tests you'll see
- Non-independent data

Important Note

- In some cases in discussing stats, we will intentionally be imprecise (and sometimes not technically accurate) about certain concepts. We are trying to give you some intuition for these concepts without extensive formal background.

BUILDING BLOCKS

Statistics

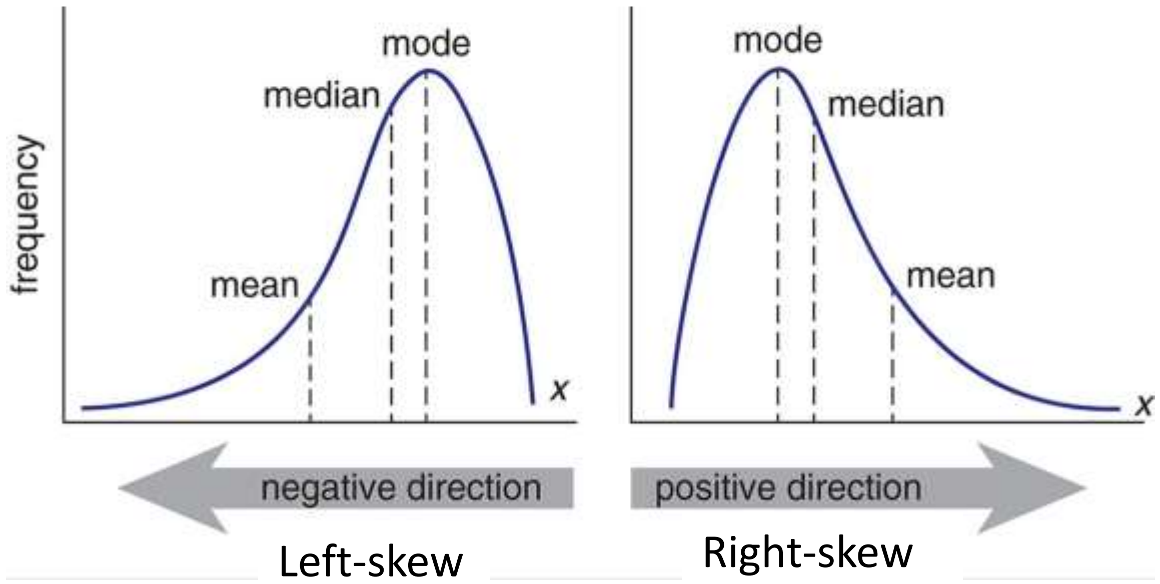
- In general: analyzing and interpreting data
- Statistical hypothesis testing: is it unlikely the data would look like this unless there is actually a difference in real life?
- Statistical correlations: are these things related?

What kind of data do you have?

- Quantitative
  - Discrete
  - Continuous
- Categorical
  - Nominal (no order)
  - Ordinal (ordered)

Exploratory Data Analysis (EDA)

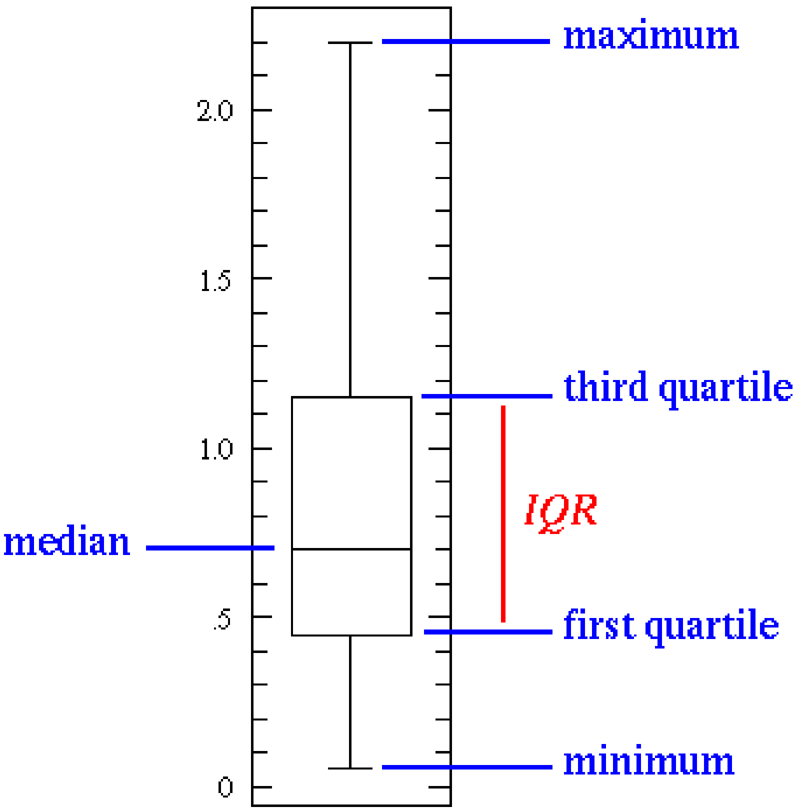
- Shape
  - Kurtosis
  - Skewness
- Center
  - Mean
  - Median
  - Mode



http://schoolbag.info/physics/physics_math/physics_math.files/image790.jpg

EDA Continued

- Spread
  - Standard Deviation
  - Variance
  - Interquartile range



http://www.physics.csbsju.edu/stats/simple.box.defs.gif

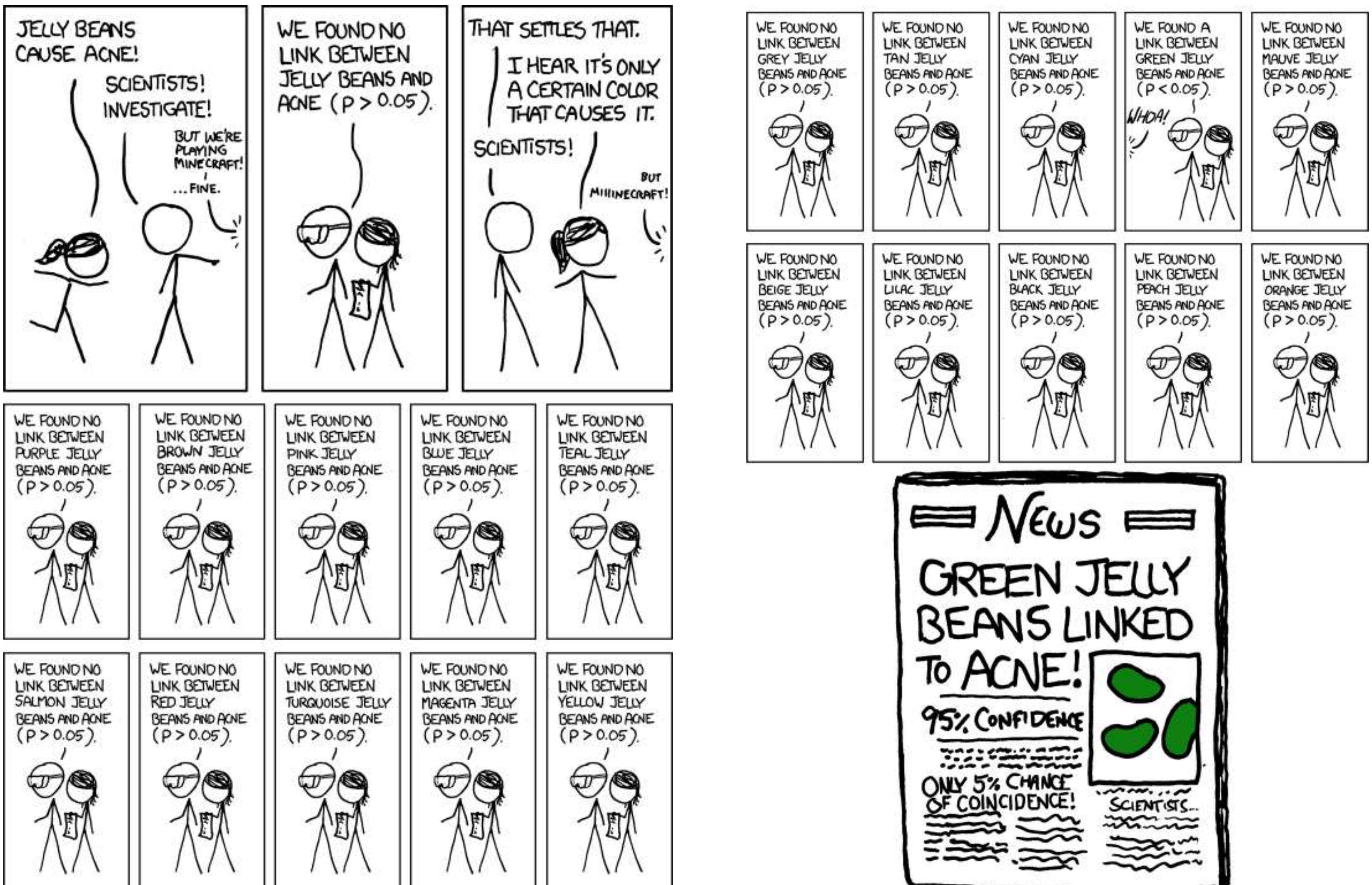
P values (reminder)

- What is the probability that the data would look like this if there's no actual difference?
- Most often, $\alpha = 0.05$
  - If $p < 0.05$, reject null hypothesis; there is a "significant" difference
  - You don't say that something is "more significant" because the p value is lower

Type I Errors

- Type I error (false positive)
  - You would expect this to happen 5% of the time if $\alpha = 0.05$
- What happens if you conduct a lot of statistical tests in one experiment?

Contrasts



https://xkcd.com/882/

Contrasts

- If we determine that the variables are dependent, we may compare conditions
- Planned vs. unplanned **contrasts**
  - You have a limited number of planned contrasts (depending on the DF) for which you don’t need to correct p values.
- Bonferroni correction (multiply p values by the number of tests) is the easiest to calculate but most conservative

Type II Errors

- Type II error (false negative)
  - There is actually a difference, but you didn't see evidence of a difference
- Statistical power is the probability of rejecting the null hypothesis if you should
  - You could do a **power analysis**, but this requires that you estimate the effect size

PICKING THE RIGHT TEST

Not all tests are created equal

Different types of dependent and independent variables?

– Different tests!

Different data distributions?

– Different **assumptions**

→Different tests!!

Parametric vs non-parametric

Which tests are we learning about today?

Focusing on parametric tests!

		Independent Variable	
		Categorical	**Quantitative**
Dependent Variable	**Categorical**	Chi-Squared Test Fisher's Exact Test	Logistic Regression
	Quantitative	t-Test ANOVA	Correlation Linear Regression

DV: CATEGORICAL
IV: CATEGORICAL

(Pearson's) Chi-squared (χ^2) Test

- Examples:
  - Does the gender (male, female) of the unicorn correlate with a unicorn's favorite color?
  - Does the type of food it eats correlate to its privacy concerns?
- H_0: Variable X factors are equally distributed across variable Y factors **(independence)**
- (Not covered today) Goodness of fit: Does the distribution we observed differ from a theoretical distribution?

Contingency tables

- Rows are one variable, columns the other

CreateAnnoying

Counts:

	0	1
0	161	32
1	165	33
2	168	34
3	170	30
4	164	32
5	161	35
6	167	32
7	129	60
8	128	61
9	154	40
10	153	40
11	154	38
12	142	42
13	121	67
14	124	76

Percentages:

	0	1
0	"83.42%"	"16.58%"
1	"83.33%"	"16.67%"
2	"83.17%"	"16.83%"
3	"85%"	"15%"
4	"83.67%"	"16.33%"
5	"82.14%"	"17.86%"
6	"83.92%"	"16.08%"
7	"68.25%"	"31.75%"
8	"67.72%"	"32.28%"
9	"79.38%"	"20.62%"
10	"79.27%"	"20.73%"
11	"80.21%"	"19.79%"
12	"77.17%"	"22.83%"
13	"64.36%"	"35.64%"
14	"62%"	"38%"

- χ^2 = 97.013, df = 14, p = 1.767e-14

Chi-squared (χ^2) Notes

- Use χ^2 if you are testing if one categorical variable (usually the assigned condition or a demographic factor) impacts another categorical variable
  - If you have fewer than 5 data points in a single cell, use Fisher's Exact Test
- Do not use χ^2 if you are testing quantitative outcomes!

What are Likert-scale data?

- Some people treat it as continuous (meh!)
- Other people treat it as ordinal (ok!)
  - You can use Mann-Whitney U / Kruskal-Wallis **(non-parametric)**
- A simple way to compare the data is to “bin” (group) the data into binary “agree” and “not agree” categories (ok!)
  - You can use χ^2 **(parametric)**

DV: CATEGORICAL
IV: QUANTITATIVE

Choosing a numerical test

- Do your data follow a normal (Gaussian) distribution? (You can calculate this!)

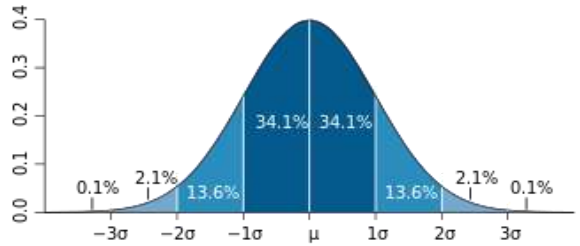


Image from http://www.wikipedia.org

  - If so, use parametric tests. If not, use non-parametric tests
- Does the data set have equal variance?
- Are your data independent?
  - If not, repeated-measures, mixed models, etc.

Independence

- Why might your data not be independent?
  - Non-independent sample (bad!)
  - The inherent design of the experiment (ok!)
- If you have two data points of unicorns' race completion times (before and after some treatment), can you actually do a single test that assumes independence to compare conditions?

Numerical data

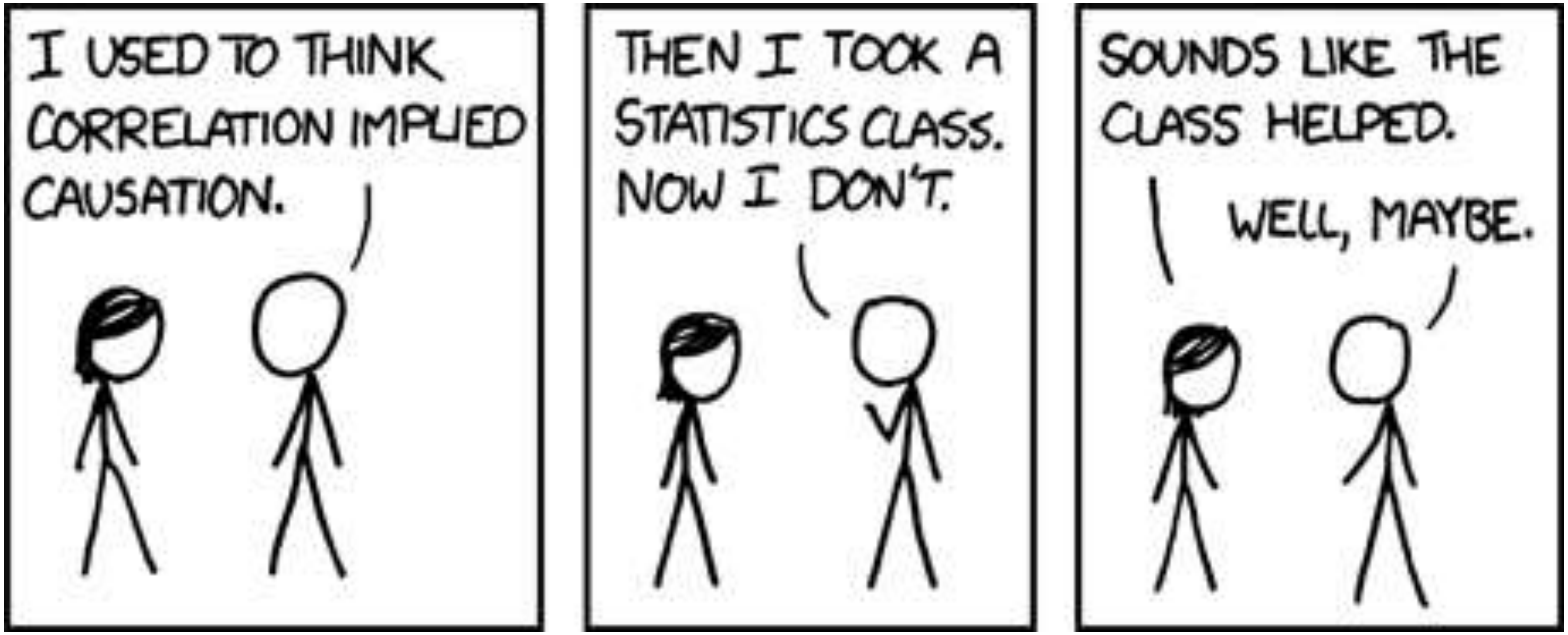
- Are values bigger in one group?
- Normal, continuous data (compare mean):
  – H_0: There are no differences in the means.
    - 2 conditions: t-test
    - 3+ conditions: ANOVA
- Non-normal data / ordinal data:
  – H_0: No group tends to have larger values.
    - 2 conditions: Mann-Whitney U (AKA Wilcoxon rank-sum test)
    - 3+ conditions: Kruskal-Wallis

DV: QUANTITATIVE

Correlation

- Usually less good: Pearson correlation
  - Requires that both variables be normally distributed
  - Only looks for a linear relationship
- Often preferred: Spearman's rank correlation coefficient (Spearman's ρ)
  - Evaluates a relationship's monotonicity
    - always going in the same direction or staying the same

Correlation **DOES NOT** imply causation



Choosing a numerical test

Check the assumptions!

- Equal variance
- Normality
- Independence of errors
- Linearity
- Fixed-x

Regressions

- What is the relationship among variables?
  - Generally one outcome (dependent variable)
  - Often multiple factors (independent variables)
- The type of regression you perform depends on the outcome
  - Binary outcome: logistic regression
  - Ordinal outcome: ordinal / ordered regression
  - Continuous outcome: linear regression

Interactions in a regression

- Normally, outcome = $ax_1 + bx_2 + c + \ldots$
- Interactions account for situations when two variables are not simply additive. Instead, their interaction impacts the outcome
  - e.g., Maybe silver unicorns, and only silver unicorns, get a much larger benefit from eating pop-tarts before a race
- Outcome = $ax_1 + bx_2 + c + d(x_1x_2) + \ldots$

Example regression

- Outcome: completed unicorn race (or not)
- Independent variables:
  - Age
  - Number of prior races
  - Diet: hay or pop-tarts
  - (Indicator variables for color categories)
  - Etc.

WHAT IF THERE IS NO INDEPENDENCE?

Non-independence

- Repeated measures (multiple measurements of the same thing)
  - e.g., before and after measurements of a unicorn’s time to finish a race
- Paired t-test (two samples per participant, two groups)
- Repeated measures ANOVA (more general)
  - Extra assumption! → Sphericity

Non-independence

- For regressions, use a **mixed model**
  - “Random effects” based on hierarchy/group
- Case 1: Many measurements of each unicorn
- Case 2: The unicorns have some other relationship. e.g., there are 100 unicorns each trained by one of 5 trainers. The identity of the trainer might impact a whole class of unicorns’ performance.

What if you have lots of questions?

- If we ask 40 privacy questions on a Likert scale, how do we analyze this survey?
- One technique is to compute a “privacy score” by adding their responses
  - Make sure the scales are the same (e.g., don’t add agreement with “privacy is dumb” and “privacy is smart”… reverse the scale)
  - You should verify that responses to the questions are correlated!

What if you have lots of questions?

- Another option: factor analysis, which evaluates the latent (underlying) factors
  - You specify *N*, a number of factors
  - Puts the questions into N groups based on their relationships
  - You should examine factor loadings (how well each latent factor correlates with a question)
  - Generally, you want questions to load primarily onto a single factor to be confident

Picking a test

- http://webspace.ship.edu/pgmarr/Geo441/Statistical%20Test%20Flow%20Chart.pdf
- http://abacus.bates.edu/~ganderso/biology/resources/statistics.html
- http://med.cmb.ac.lk/SMJ/VOLUME%203%20DOWNLOADS/Page%2033-37%20-%20Choosing%20the%20correct%20statistical%20test%20made%20easy.pdf

What can we conclude statistically

- X varies in a way that's related to Y
  - As the age of a unicorn increases, its max speed decreases
  - Pearson's correlation / Spearman's correlation
- Assignment to X impacts Y (category)
  - Unicorns randomly assigned to eat vegan food (as opposed to non-vegan food) are more likely to be rated as successful (as opposed to unsuccessful)
  - χ^2, Fisher's exact test

What can we conclude statistically

- Assignment to X impacts Y (numerical)
  - Unicorns randomly assigned to eat vegan food (as opposed to non-vegan food) are more likely to take a shorter time to run a race
  - ANOVA, Kruskal-Wallis, etc.
- Lots of factors impact Y (category)
  - Logistic regression
- Lots of factors impact Y (numerical)
  - Regression