HYPOTHESIS TESTING

MPA 630: Data Science for Public Management

November 15, 2018

Fill out your reading report on Learning Suite

PLAN FOR TODAY

Randomness, repetition, and replicability

Why are we even doing this?
(again!)

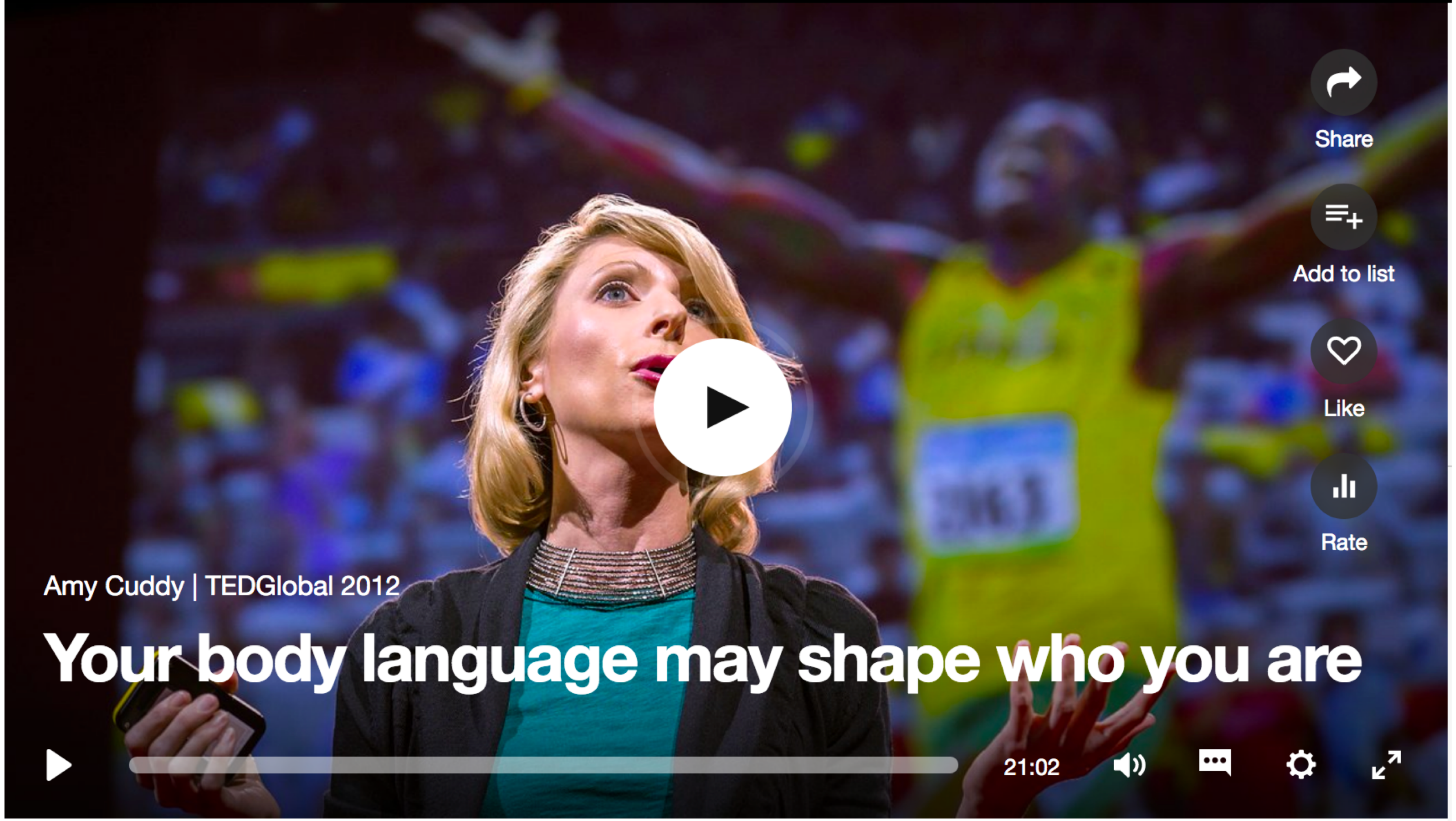
Burdens of proof

How to test any hypothesis

RANDOMNESS, REPETITION, & REPLICABILITY

TED Ideas worth spreading

WATCH DISCOVER AT



Details
About the talk

Transcript
51 languages

Comments (2367)
Join the conversation

Body language affects how others see us, but it may also change how we see ourselves. Social psychologist Amy Cuddy argues that "power posing" — standing in a posture of confidence, even when we don't feel confident — can boost feelings of

49,718,157
views

POWER POSING

Increases individual perception of power

Increases testosterone and decreases cortisol

She made a guess at a population parameter and published it

This is the process of science!

RUH ROH

'Power poses' don't work, eleven new studies suggest

Date: September 11, 2017

Source: Michigan State University

Summary: The claim that holding a 'power pose' can improve your life becar
eral years ago, fueling the second most-watched TED talk ever bu
about the science behind the assertion. Now comes the most def
date suggesting that power poses do not improve your life.

Share:

RELATED TOPICS

FULL STORY

H

HEALTH • SCIENCE

'Power Poses' Don't Actually Work. Try These Confidence-Boosting Strategies Instead

TECH & SCIENCE

'POWER POSES' DON'T REALLY MAKE YOU MORE POWERFUL, NINE MORE STUDIES CONFIRM

BY MEGHAN BARTELS ON 9/13/17 AT 12:25 PM

FEATURE

When the Revolution Came for Amy Cuddy

As a young social psychologist, she played by the rules and won big: an influential study, a viral TED talk, a prestigious job at Harvard. Then, suddenly, the rules changed.

BUT WAIT

70,286 views | Apr 3, 2018, 03:52pm

Power Posing Is Back: Amy Cuddy Successfully Refutes Criticism

EMOTION, METHODS, REPLICATIONS

March 28, 2018

54-study analysis says power posing does affect people's emotions and is worth researching further

Increases individual perception of power

~~**Increases testosterone and decreases cortisol**~~

MORAL OF THE STORY

Randomness is weird

Capturing true population parameters is hard

Replication and repetition are needed to check the net

WHY ARE WE EVEN DOING THIS?

Round 2!

POPULATION PARAMETERS

Key assumption in the flavor of statistics we're doing:

There are true, fixed population parameters out in the world

POPULATION VS. SAMPLE

Proportion	p	$\hat{p}$
Mean	μ	
Difference between proportions	$p_1 - p_2$	
Difference between means	$\mu_1 - \mu_2$	
Intercept	β_0	$\hat{\beta}_0$

δ

WHY WE SIMULATE

Is it accurate?	Is it real?	Is it substantive?
Is $\bar{x}$ an accurate guess of μ?	Is $\bar{x}_2 - \bar{x}_1$ or $\hat{p}_2 - \hat{p}_1$ real? Does it matter?	
Width of confidence interval	Important numbers included in confidence interval	

BURDENS OF PROOF

AMERICAN LEGAL SYSTEM

Accused must be judged

Presumption of innocence

Accuser has burden of proving guilt

Judge/jury decide guilt based on amount of evidence

We never prove innocence;
we try (and fail) to reject innocence

LEGAL EVIDENTIARY STANDARDS

Preponderance of evidence

Clear and convincing evidence

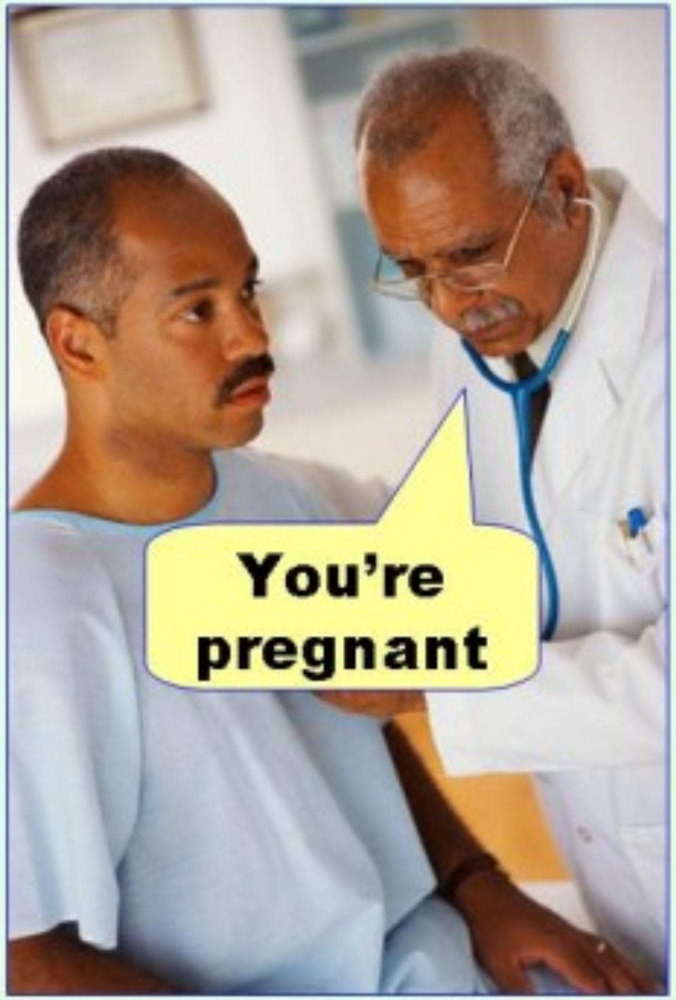
Beyond reasonable doubt

Why do we have these different levels?

We're afraid of locking up an innocent person

		Actual truth	
		Guilty	Not guilty
Jury decision	Guilty	Yay! True positive	Oh no! False positive (I)
	Not guilty	Oh no! False negative (II)	Yay! True negative

Type I error
(false positive)



Type II error
(false negative)

STATISTICAL "LEGAL" SYSTEM

Sample statistic (δ) must be judged

Presumption of no effect (null)

You have burden of proving effect

You decide "guiltiness" of effect based on amount of evidence

We never prove that the null is true; we try (and fail) to reject the null

		Actual truth	
		Yes effect	No effect
Result of hypothesis test	Yes effect	Yay! True positive	Oh no! False positive (I)
	No effect	Oh no! False negative (II)	Yay! True negative

α
0.10
0.05
0.01

STATISTICAL SIGNIFICANCE

There's enough evidence to safely reject the null hypothesis

P-VALUES

The probability of observing an effect at least that large when no effect exists

NOBODY UNDERSTANDS THESE

http://fivethirtyeight.com/pvalue

HOW TO TEST ANY HYPOTHESIS

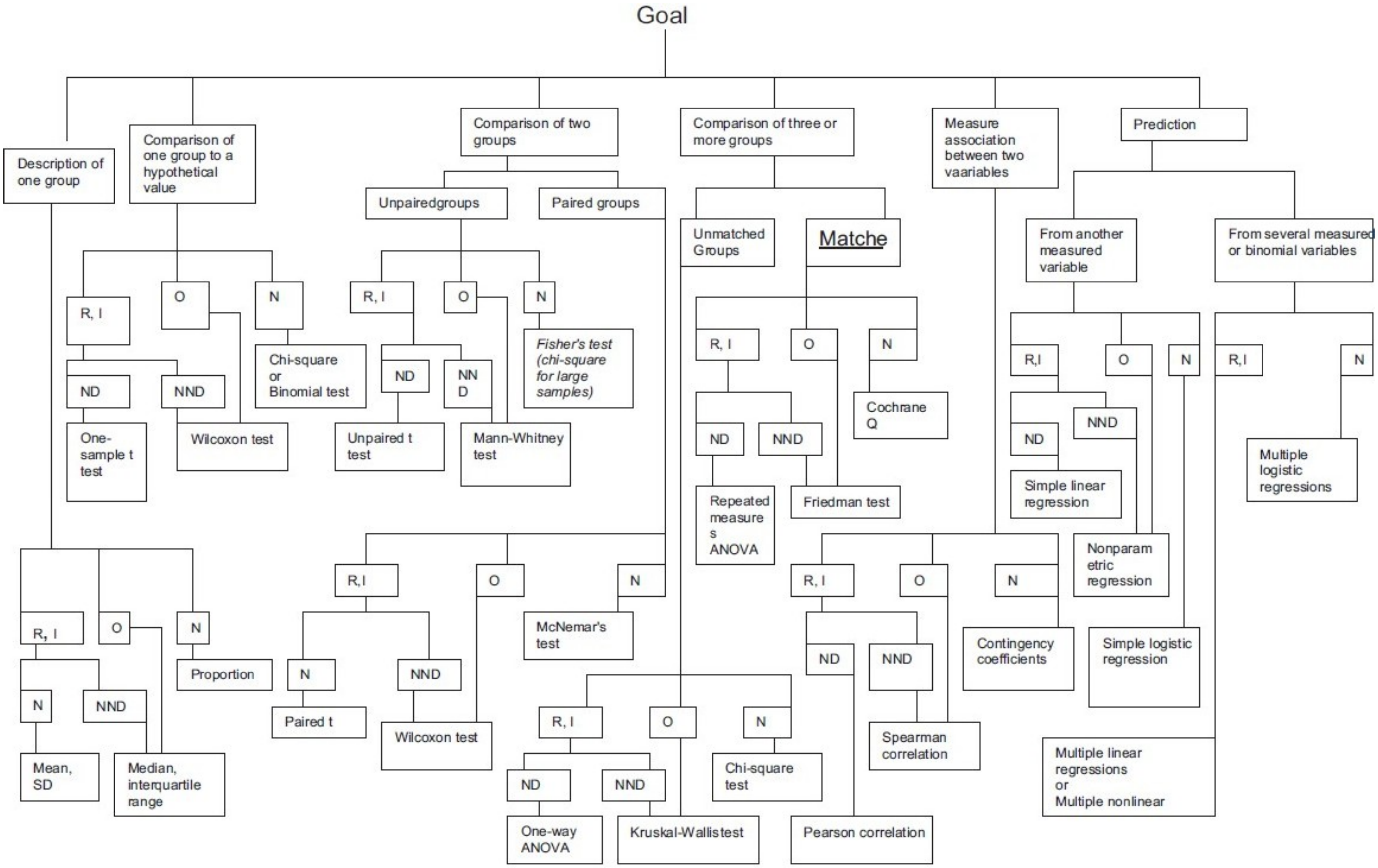
Is your data count data (ie; number of something counted per day, per week, etc), with low number of counts such that the stochasticity in data are not in the Normal regime?

Yes, this is low-count count data

- You are on the wrong page. Go here

No, this either is not count data, or it is high-count count data

- Are you testing if the mean of just one sample is consistent with some value, mu?
- Yes, I am testing if just one mean is consistent with some value
  - Does the sample have at least N~10 measurements used to calculate the mean (ie; the Central Limit Theorem applies)?
  - Yes, there are at least 10 measurements
    1. Calculate the sample mean, bar(X) and standard error on the mean, SE
    2. Calculate the Z statistic
    3. Use pnorm(Z) in R to calculate the p-value.
    4. The p-value tests the null hypothesis that the true mean of the probability distribution underlying the sample is consistent with mu.
    5. Reject the null hypothesis if the p-value is close to 0 or 1 (ie; within 0.05 of 0 or 1)
  - No, there are not at least ~10 measurements
    - Do you think the data are likely consistent with being Normally distributed?
    - Yes, the data are Normally distributed
      1. Calculate the sample mean, bar(X) and standard error on the mean, SE
      2. Calculate the t statistic
      3. Calculate number degrees of freedom df=(N-1)
      4. To calculate the p-value, use pt(t,df) in R
      5. The p-value tests the null hypothesis that the true mean of the probability distribution underlying the sample is consistent with mu
      6. Reject the null hypothesis if the p-value is close to 0 or close to 1 (ie; within 0.05 of 0 or 1)
    - No, the data aren't Normally distributed
      - Beyond the scope of this course because it involves likelihood methods (but note that, wrong or not, usually people just assume that small samples of data are in fact Normally distributed, and go ahead and use the t-test)
- No, I'm testing equality of more than one mean
  - Are you testing if means of <u>two</u> samples are consistent with being equal?
  - Yes, I am testing just two means

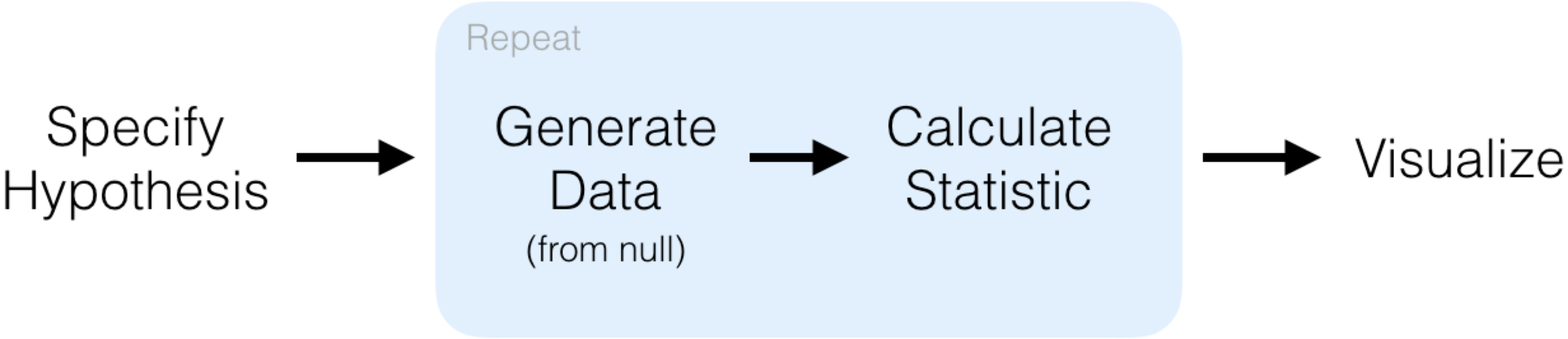


R, I = Ratio and Interval data O= Ordinal data N = Nominal data

N = Normal distribution NND = Non normal distribution

Degrees of freedom	α									
	0.995	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.262	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

SIMULATIONS AND HYPOTHESES




```
specify(response) %>%

    hypothesize(null)    %>%    generate(reps)    %>%    calculate(stat)    %>%    visualize()
```

Find δ

The sample statistic: diff in means, mean, diff in props, etc.

Invent world where δ is null

Simulate what the world would look like if there was no effect.

Look at δ in the null world

Is it big and extraordinary, or is it a normal thing?

Calculate probability that δ could exist in the null world

This is your p-value!

Decide!