nature International weekly journal of science

NATURE | RESEARCH HIGHLIGHTS: SOCIAL SELECTION

Psychology journal bans *P* values

Test for reliability of results 'too easy to pass', say editors.

Chris Woolston

26 February 2015 | Clarified: 09 March 2015

A controversial statistical test has finally met its end, at least in one journal. Earlier this month, the editors of *Basic and Applied Social Psychology* (*BASP*) announced that the journal would no longer publish papers containing *P* values because the statistics were too often used to support lower-quality research [1].

nature International weekly journal of science

NATURE | NEWS

First results from psychology's largest reproducibility test

Crowd-sourced effort raises nuanced questions about what counts as replication.

Monya Baker

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

News & Comment > News > 2019 > May > Article

NATURE | NEWS

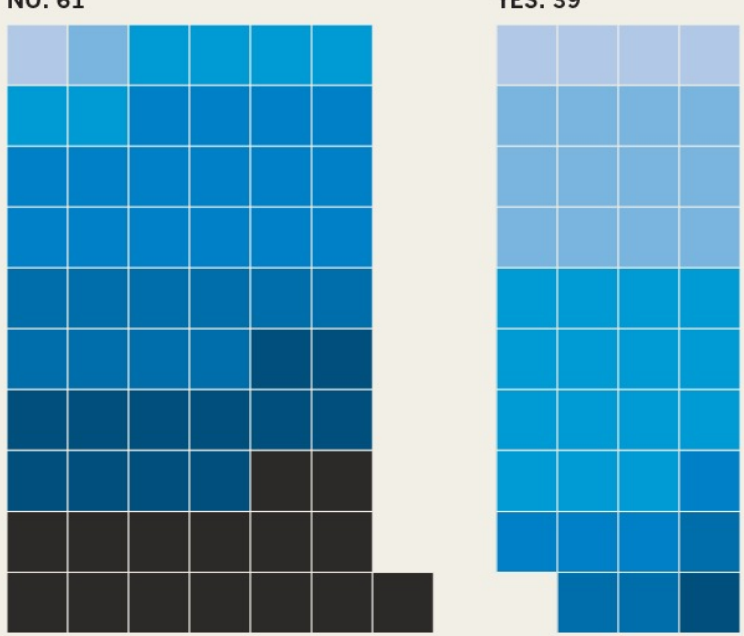
Psychology's reproducibility problem is exaggerated – say psychologists

Reanalysis of last year's enormous replication study argues that there is no need to be so pessimistic.

RELIABILITY TEST

An effort to reproduce 100 psychology findings found that only 39 held up.* But some of the 61 non-replications reported similar findings to those of their original papers.

Did replicate match original's results?



* based on criteria set at the start of each study

nature International weekly journal of science

NATURE | NEWS FEATURE

1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016 | Corrected: 28 July 2016

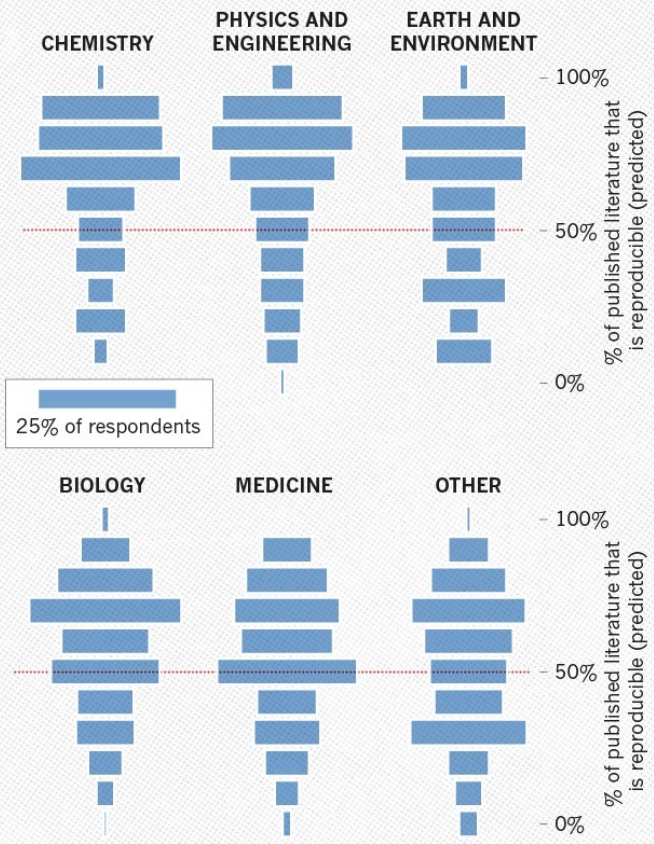
IS THER

No, there is
cri

38%
Yes, a slight
crisis

HOW MUCH PUBLISHED WORK IN YOUR FIELD IS REPRODUCIBLE?

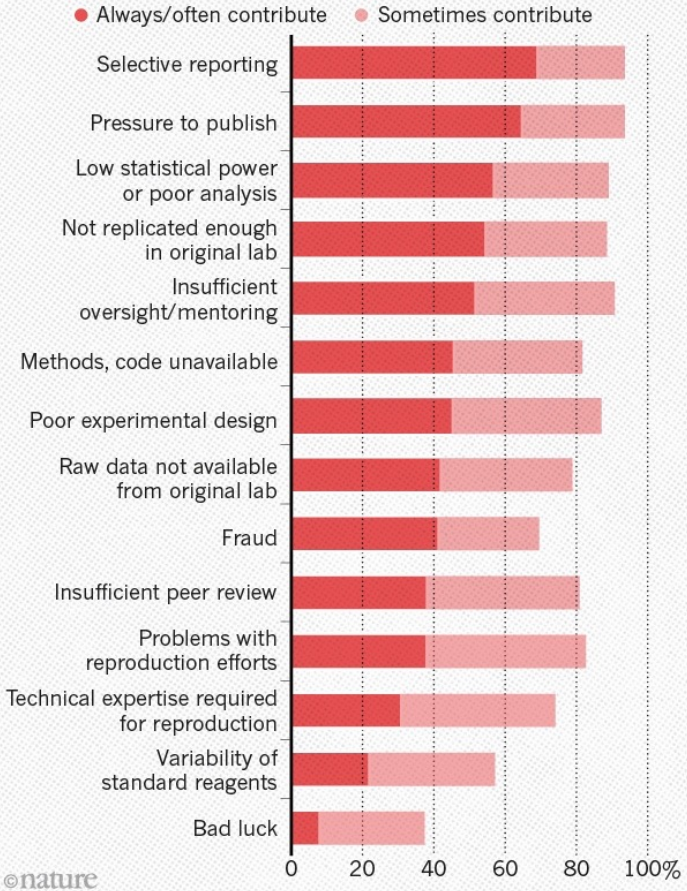
Physicists and chemists were most confident in the literature.



Number of respondents from each discipline:
Biology **703**, Chemistry **106**, Earth and environmental **95**, Medicine **203**, Physics and engineering **236**, Other **233**

WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



©nature

WHAT FACTORS COULD BOOST REPRODUCIBILITY?

Respondents were positive about most proposed improvements but emphasized training in particular.

©nature

nature
International weekly journal of science

NATURE | COMMENT

Statistics: *P* values are just the tip of the iceberg

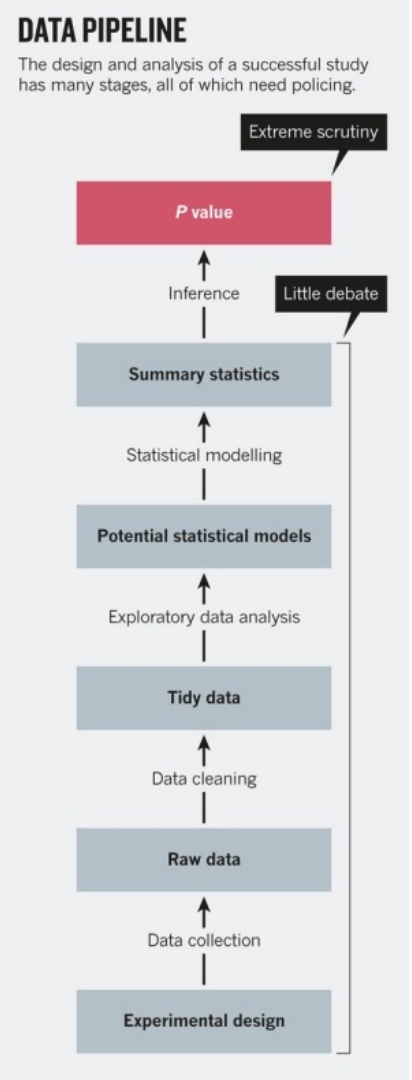
Jeffrey T. Leek & Roger D. Peng

28 April 2015

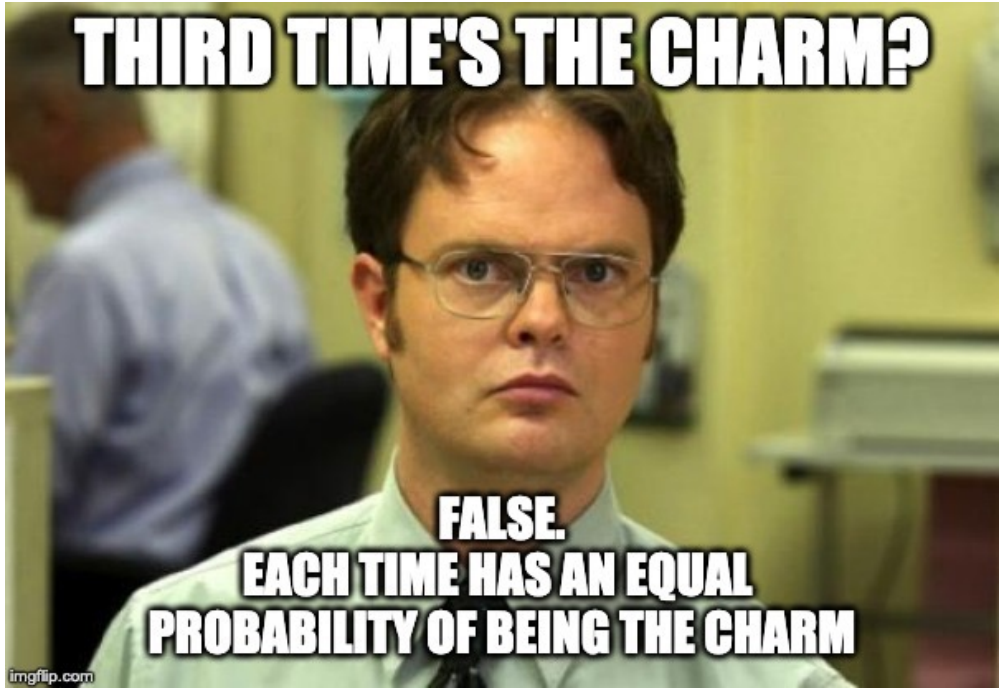
Ridding science of shoddy statistics will require scrutiny of every step, not merely the last one, say Jeffrey T. Leek and Roger D. Peng.

"Arguing about the P value is like focusing on a single misspelling, rather than on the faulty logic of a sentence."

"But education is not enough. Data analysis is taught through an apprenticeship model, and different disciplines develop their own analysis subcultures."

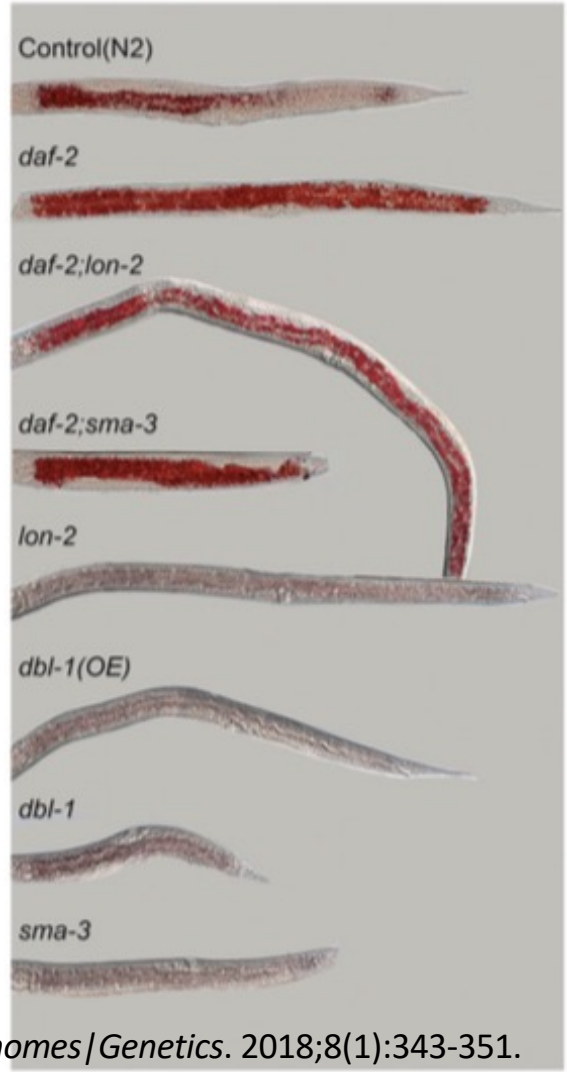


Abrasive Statistical Apprenticeship



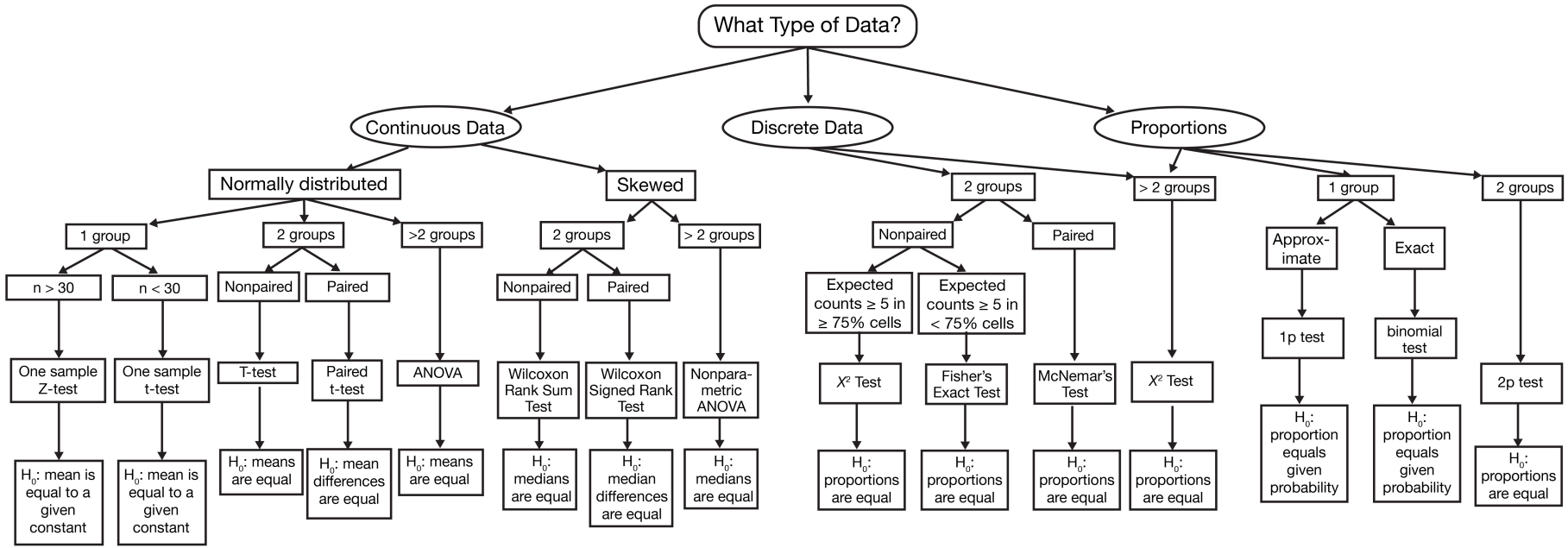
How to formulate a test

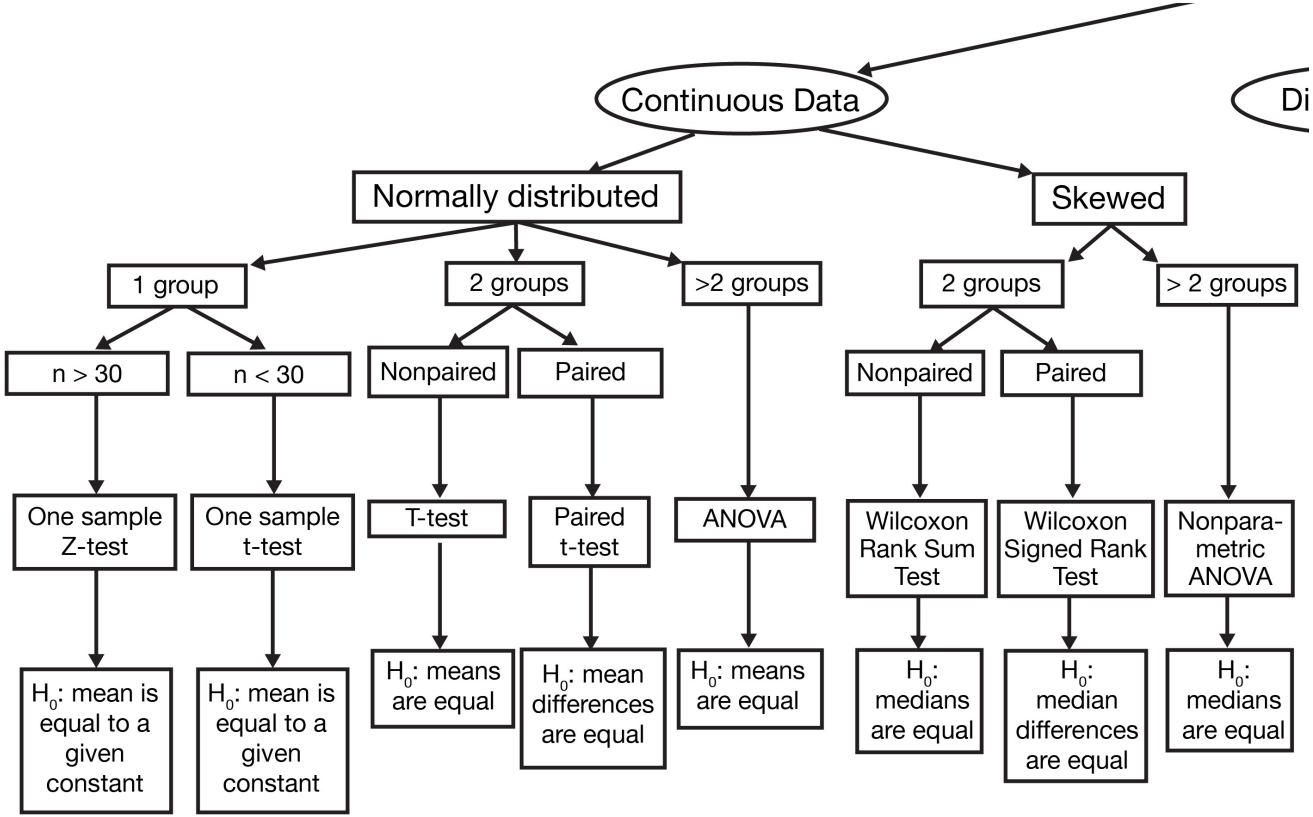
- What comparison am I making/testing?
  - A group of observations versus a given value
  - Ex. Are my worm lengths > [some standard worm length]?
  - Two groups compared to each other
  - Ex. Does group A have longer worms than group B?
  - Multiple groups
  - Ex. Are there any groups (A, B, C, …) with different worm lengths?
  - Find evidence to reject your **null hypothesis of NO DIFFERENCE**
- What type of data do I have?
  - Consult a flowchart

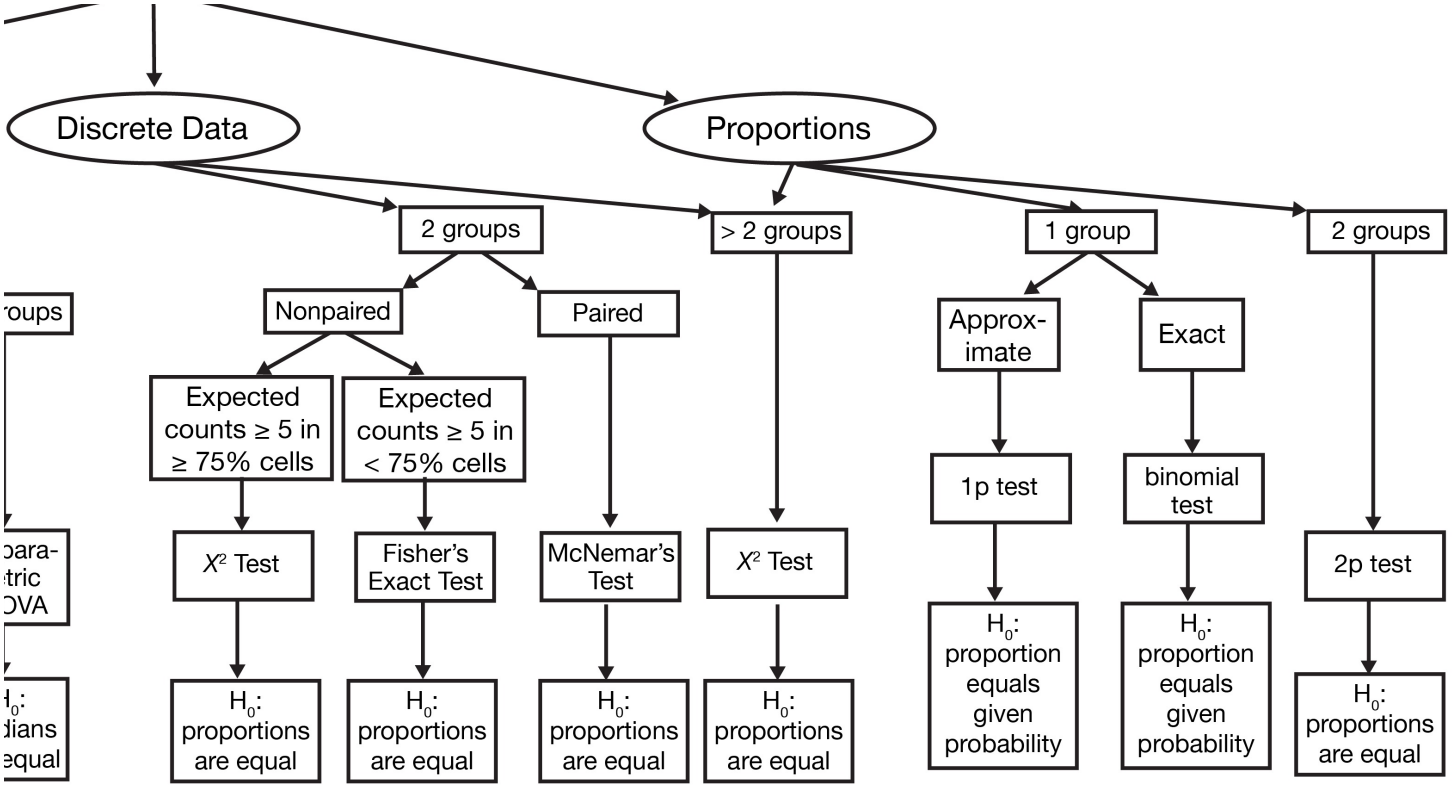


Clark JF, *et al. G3: Genes|Genomes|Genetics*. 2018;8(1):343-351.

Flow chart: which test statistic should you use?




Continuous Data
Di
Normally distributed
Skewed
1 group
2 groups
>2 groups
2 groups
> 2 groups
n > 30
n < 30
Nonpaired
Paired
Nonpaired
Paired
One sample Z-test
One sample t-test
T-test
Paired t-test
ANOVA
Wilcoxon Rank Sum Test
Wilcoxon Signed Rank Test
Nonpara-metric ANOVA
H0: mean is equal to a given constant
H0: mean is equal to a given constant
H0: means are equal
H0: mean differences are equal
H0: means are equal
H0: medians are equal
H0: median differences are equal
H0: medians are equal


Discrete Data
Proportions
2 groups
> 2 groups
1 group
2 groups
roups
Nonpaired
Paired
Approx-imate
Exact
Expected counts ≥ 5 in ≥ 75% cells
Expected counts ≥ 5 in < 75% cells
1p test
binomial test
para-tric OVA
X^2 Test
Fisher's Exact Test
McNemar's Test
X^2 Test
2p test
H_0: dians equal
H_0: proportions are equal
H_0: proportions are equal
H_0: proportions are equal
H_0: proportions are equal
H_0: proportion equals given probability
H_0: proportion equals given probability
H_0: proportions are equal

Tests are built in to statistical software like R

- T-test: `t.test`
- Chi-sq test: `chisq.test`
- ANOVA: `anova`
- Proportion test: `prop.test, binom.test`
- And many, many more: `fisher.test, wilcox.test, kruskal.test`

Testing Non-normal groups for large sample sizes

- Use the t-test
- Because: The Central Limit Theorem
- Statquest YouTube series with Josh Starmer: https://www.youtube.com/watch?v=YAlJCEDH2uY