L18: Lasso + Regularized Regression

Jeff M. Phillips

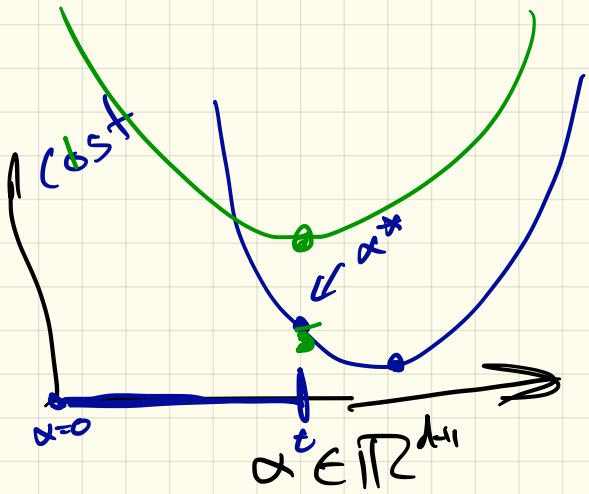
March 14, 2018

$$\rightarrow \underset{\alpha}{\operatorname{argmin}} \|X\alpha - y\|_2 + s\|\alpha\|_1 \qquad \text{(Lasso)}$$

equivalent to

$$\rightarrow \underset{\alpha}{\operatorname{argmin}} \|X\alpha - y\|_2 \quad \text{such that } \|\alpha\|_1 \leq t$$

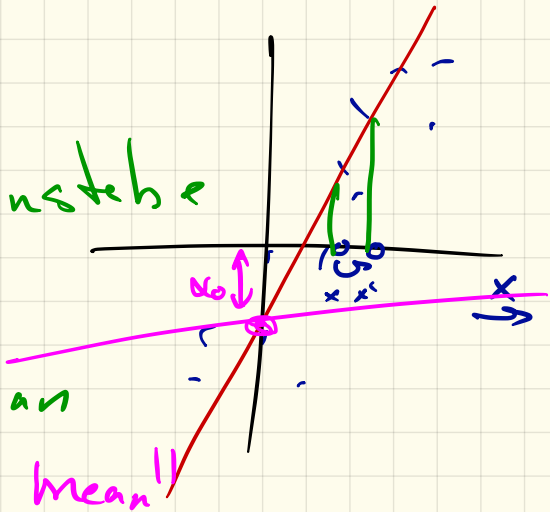
where t depends on X, y, s



as s increases
t decreases

- Why regularization

+ high slope is unstable



+ biasing towards mean

"regression to the mean"

Input X, y $x_i \in X$ $x_i \in (x_0, x_1, \dots x_d)$

$y = (y_1, y_2 \dots y_n)$ $y_i \in \mathbb{R}$ fix $x_0 = 1$

find $f_\alpha(x) \to \mathbb{R}$

$\alpha = (\alpha_0, \alpha_1 \dots \alpha_d)$ $f_\alpha(x) = \sum_{i=0}^{d} \alpha_i x_i = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots$

$$\alpha^* = \underset{\alpha}{\operatorname{argmin}} \| X\alpha - y \|_2$$

$$= \underset{\alpha}{\operatorname{argmin}} \sum_{x_j \in X} (\langle x_j, \alpha \rangle - y_j)^2$$

$$\sum (f_\alpha(x_j) - y_j)^2$$

$$\alpha^* = (X^T X)^{-1} X^T y$$

regularization

$+ \gamma \|\alpha\|_{(1),2,3}$

1-norm (lasso)

sparsity in $\alpha \Rightarrow$ many $\alpha_j = 0$

2-norm Ridge Reg

$\alpha_\gamma^* = (X^T X + \gamma^2 I)^{-1} X^T y$

Orthogonal Matching Pursuit (OMP)

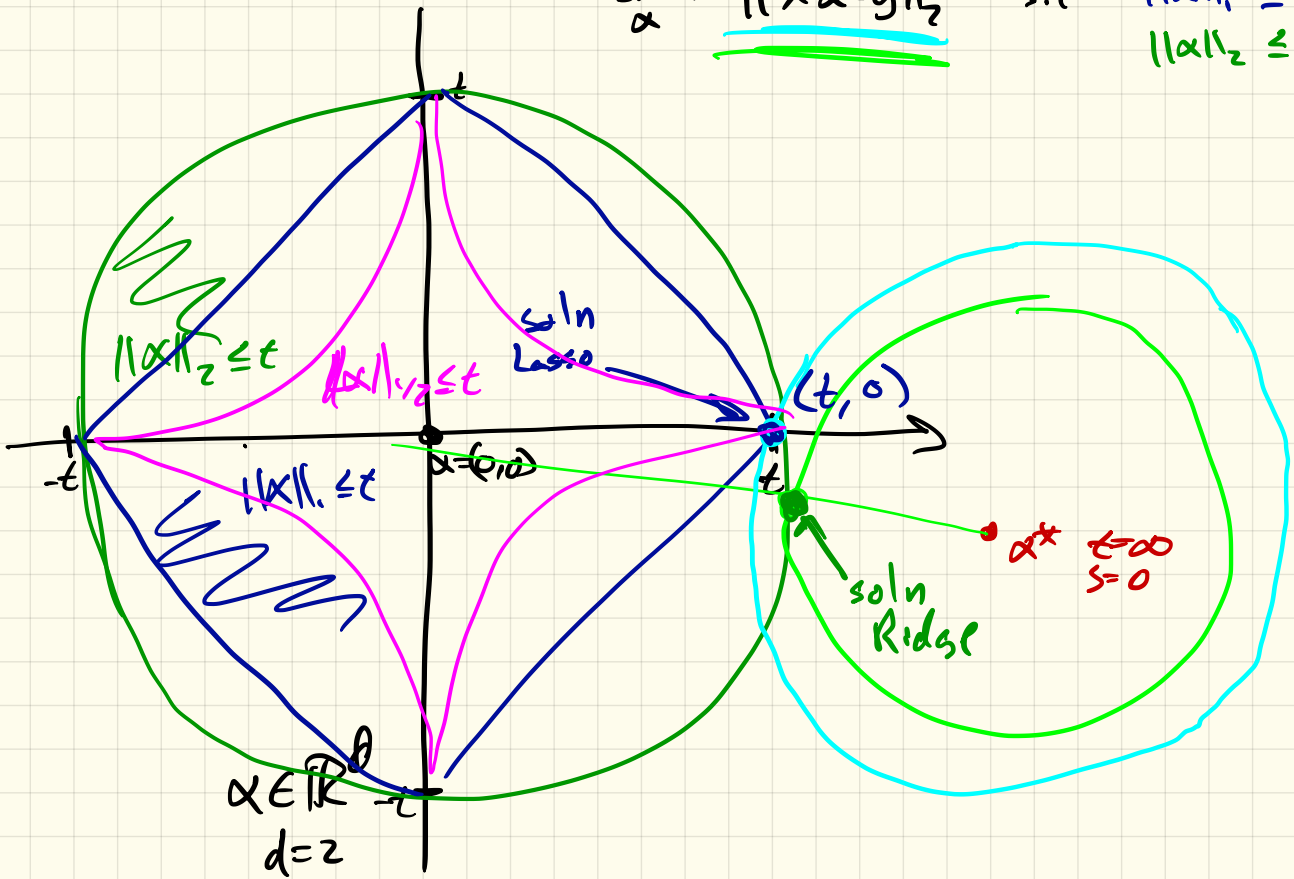
Find $\alpha^* = \arg\min_{\alpha \in \mathbb{R}^d} \|X\alpha - y\|_2 + s\|\alpha\|_1$
Forward Subset Selection:

Orthogonal Matching Pursuit

Set $r = y$.
for $i = 1$ **to** t **do**
  Set $X_j = \arg\max_{X_{j'} \in X} |\langle r, X_{j'} \rangle|$.
  Set $\alpha_j = \arg\min_\alpha \|r - X_j\alpha\| + s|\alpha|$.
  Set $r = r - X_j\alpha_j$.
Return $\hat{S}$ where $\hat{s}_j = \gamma_j$ (or 0).

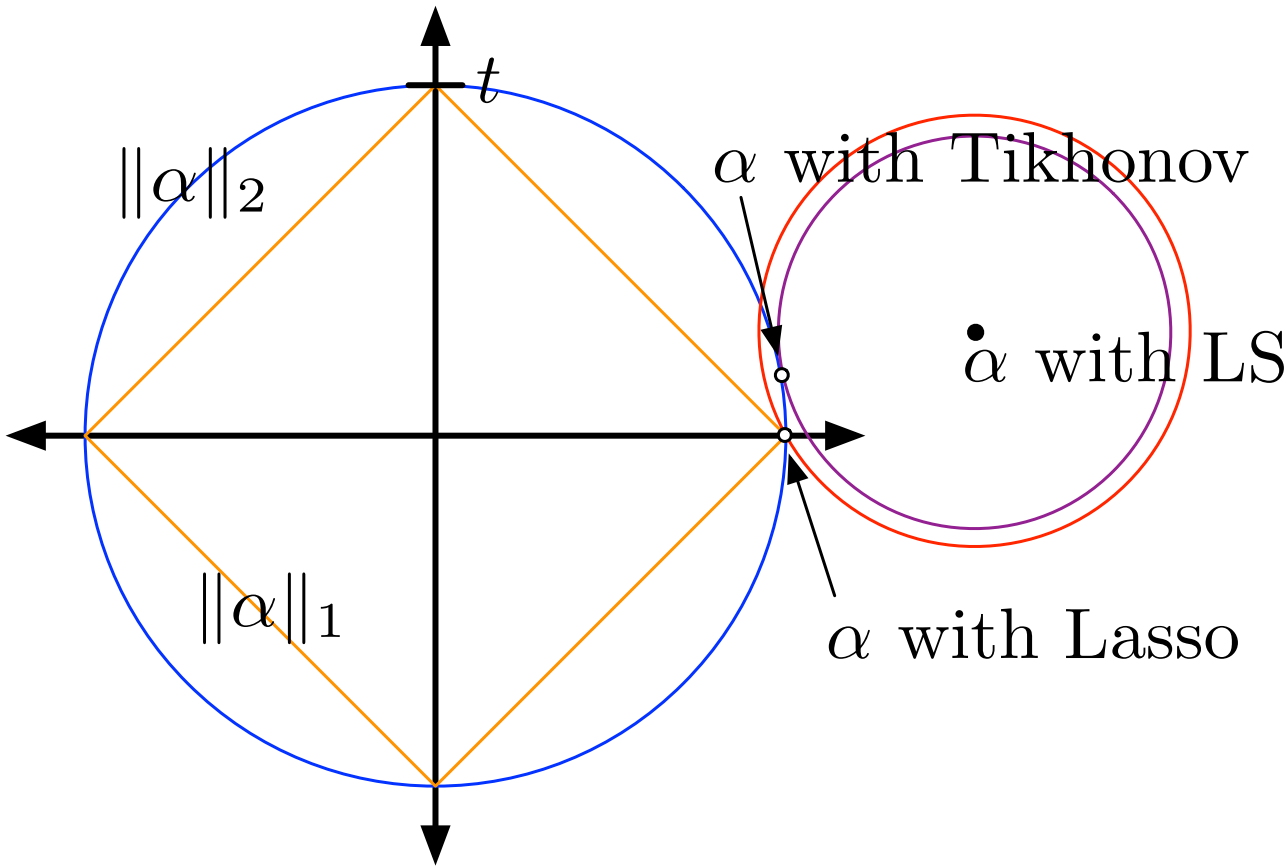
$\underset{\alpha}{\operatorname{argmin}} \|X\alpha - y\|_2$ s.t $\|\alpha\|_1 \le t$

$\|\alpha\|_2 \le t$



Lasso Illustration

Find $\alpha^* = \arg\min_{\alpha \in \mathbb{R}^d} \|X\alpha - y\|_2 + s\|\alpha\|_1$



Least Angle Regression

↳ solves Lasso exactly.

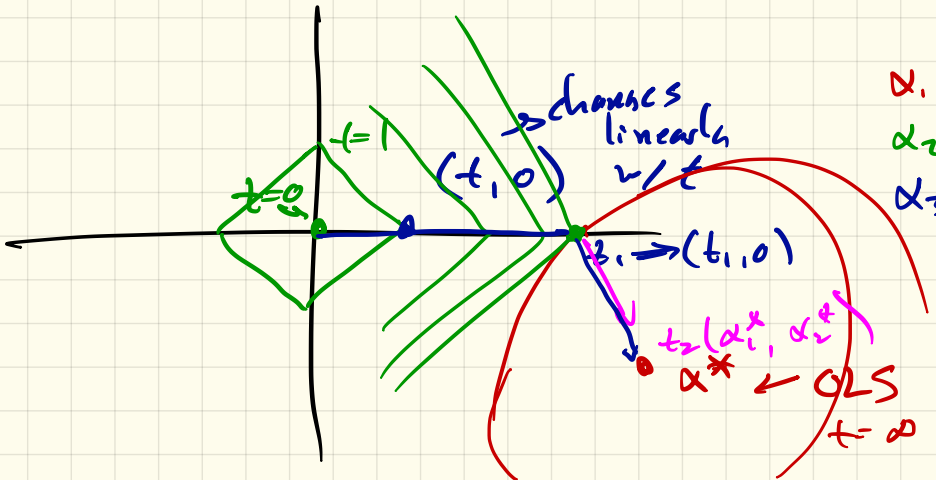
for all values of s (or t)

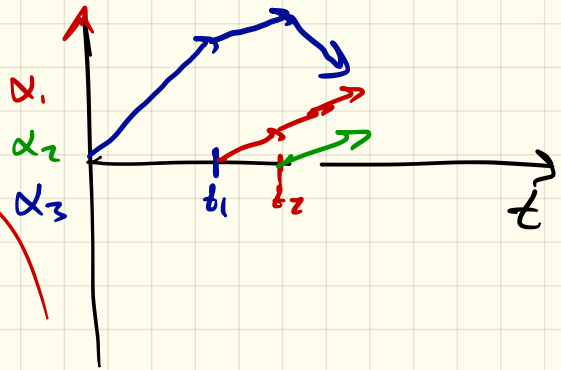
$\|X\alpha - y\|_2$ s.t. $\|\alpha\|_1 \le t$

Starts w/ $t = 0$

↳ adds if

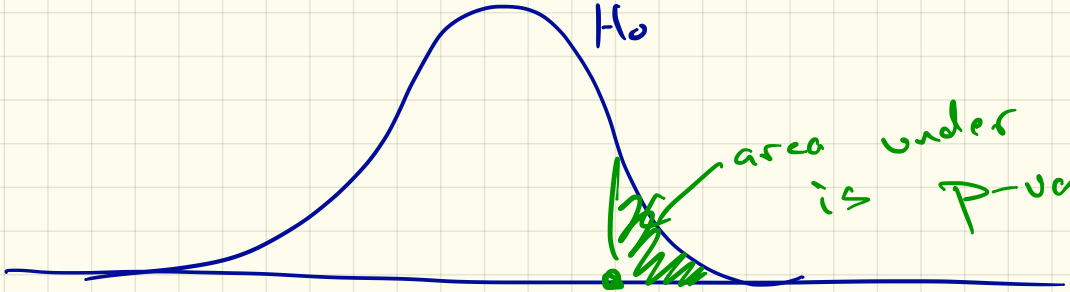
increases t while maintaining optimal α_t





P-values

Null Hypothesis H_0 Distribution.



often if p-value ≤ 0.05
be then "significant".

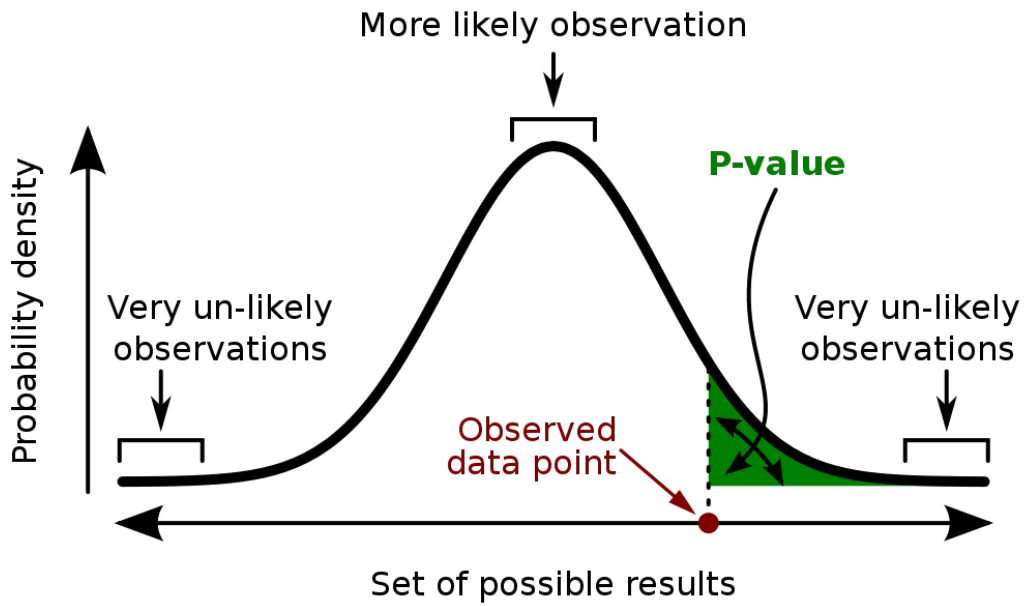
hypothesis ≡ α

Important:

Pr (observation | hypothesis) ≠ Pr (hypothesis | observation)

The probability of observing a result given that some hypothesis is true is *not equivalent* to the probability that a hypothesis is true given that some result has been observed.

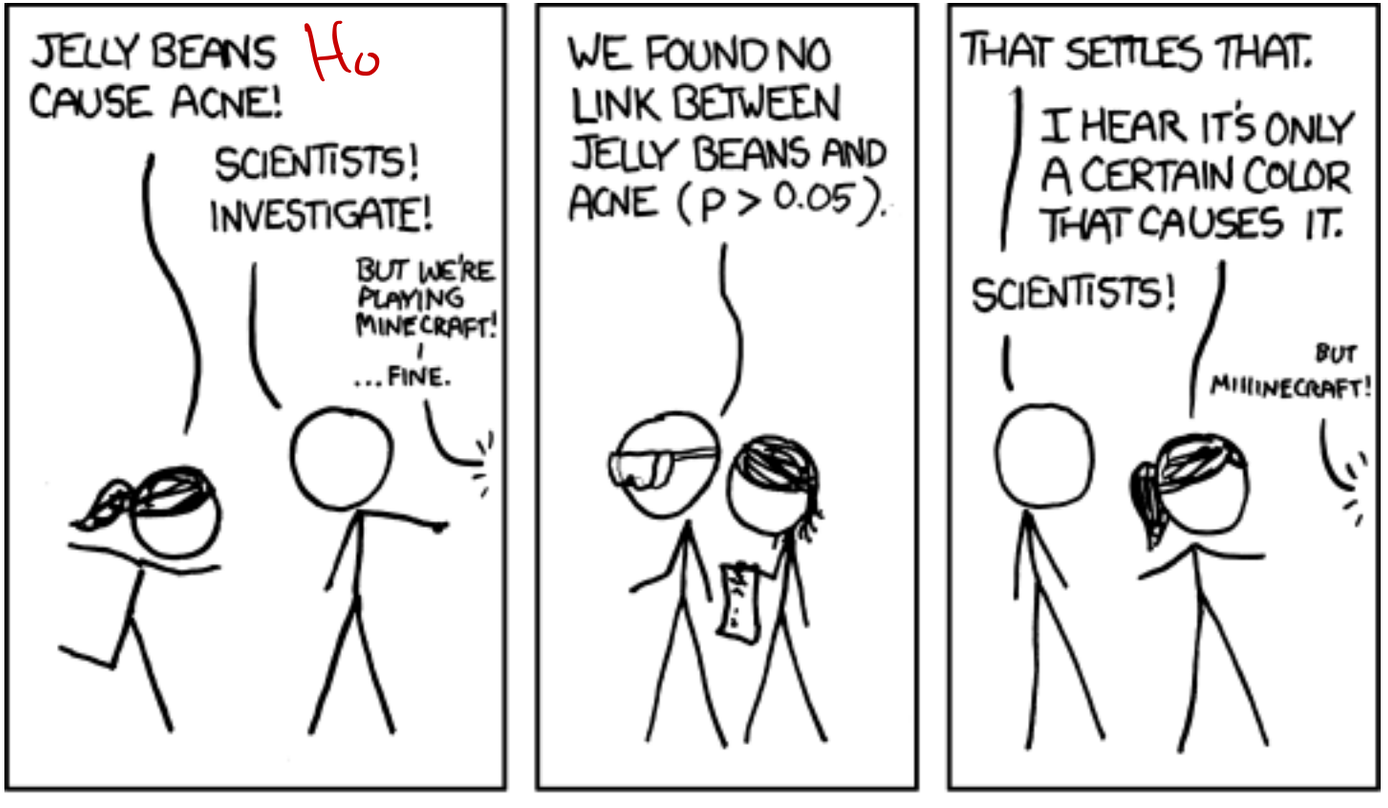
Using the p-value as a "score" is committing an egregious logical error:
the transposed conditional fallacy.



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

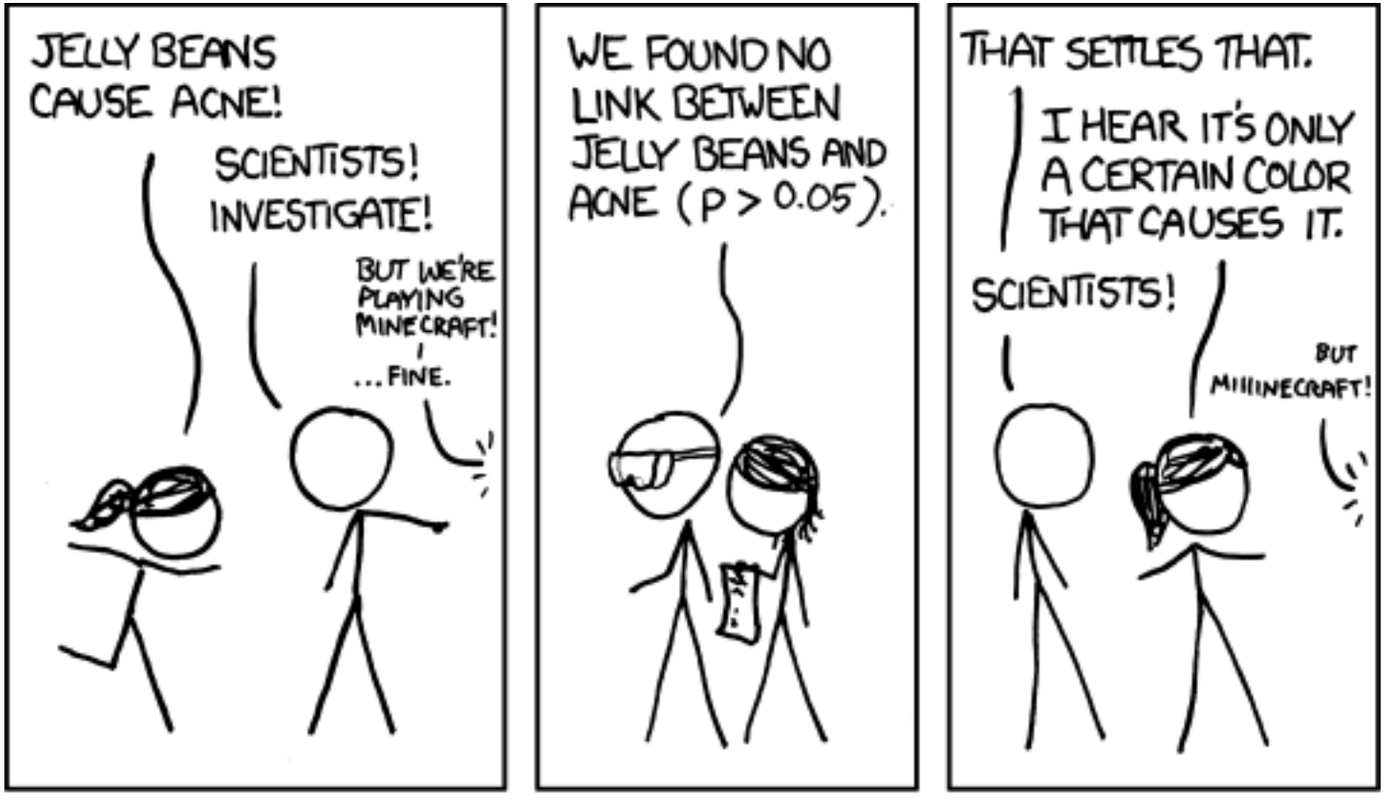
1. Multiple Hypothesis Testing

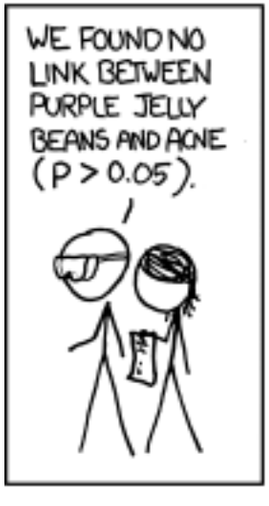
https://xkcd.com/882/



1. Multiple Hypothesis Testing

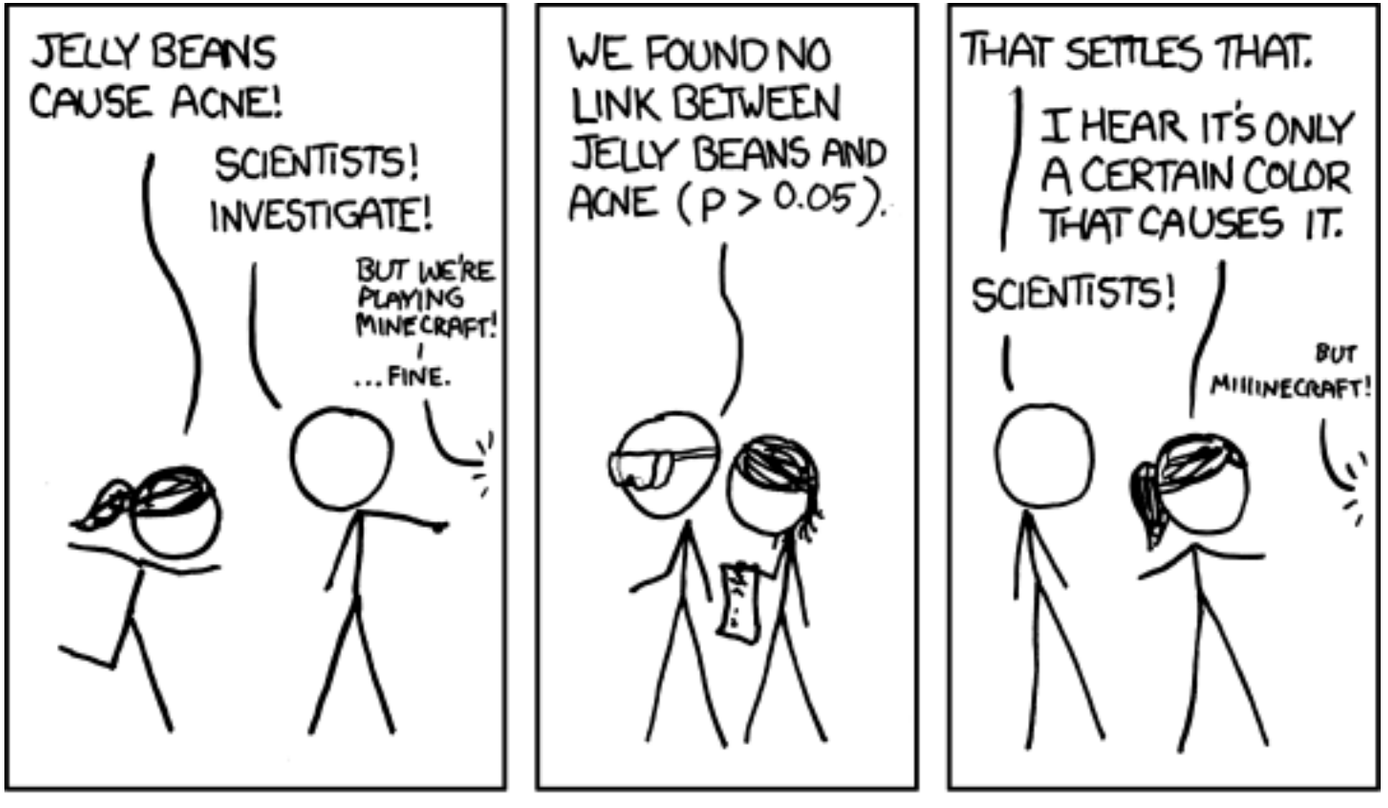
https://xkcd.com/882/

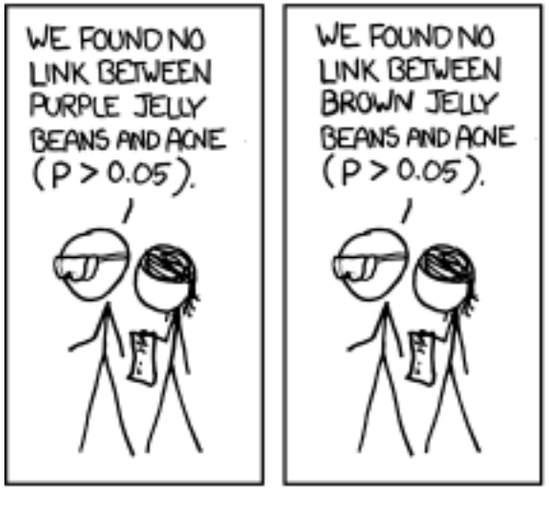




1. Multiple Hypothesis Testing

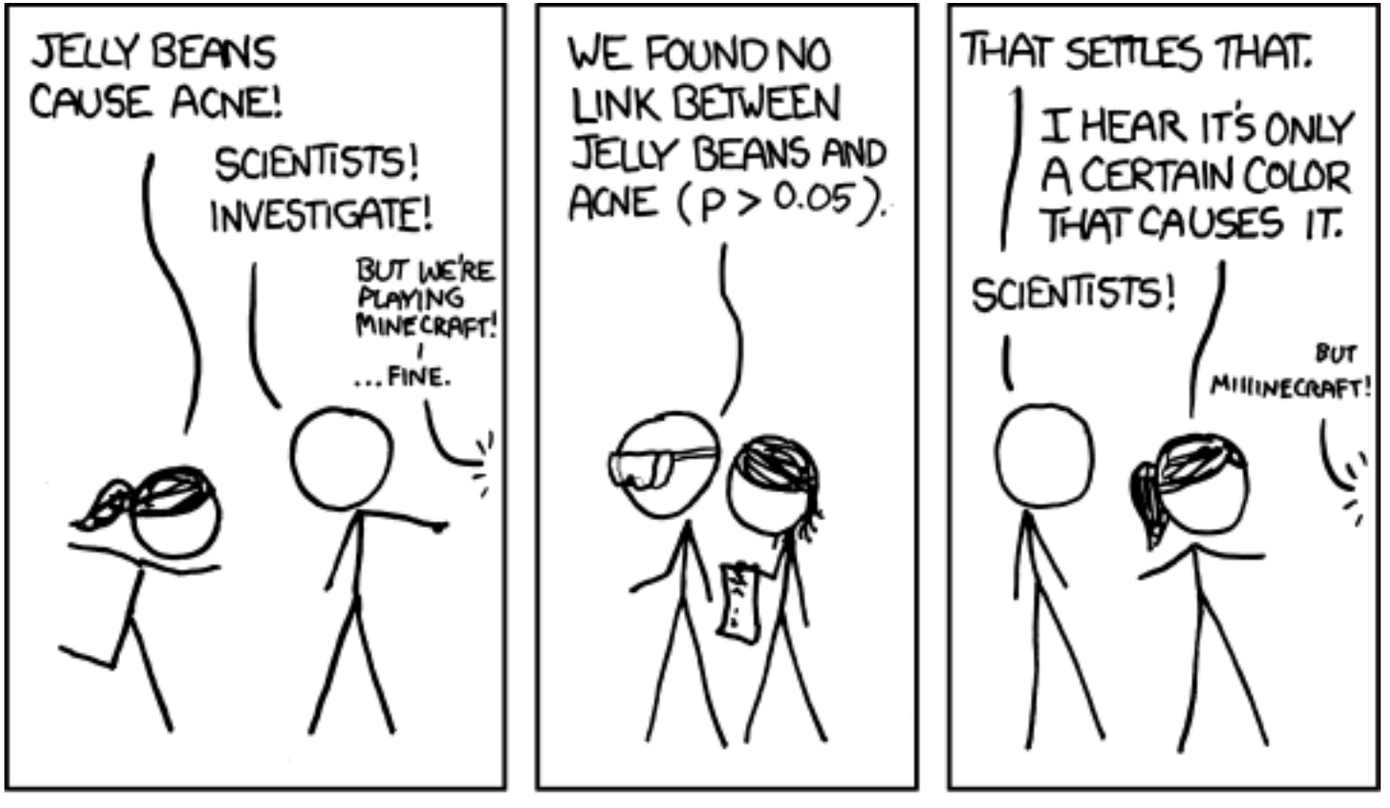
https://xkcd.com/882/

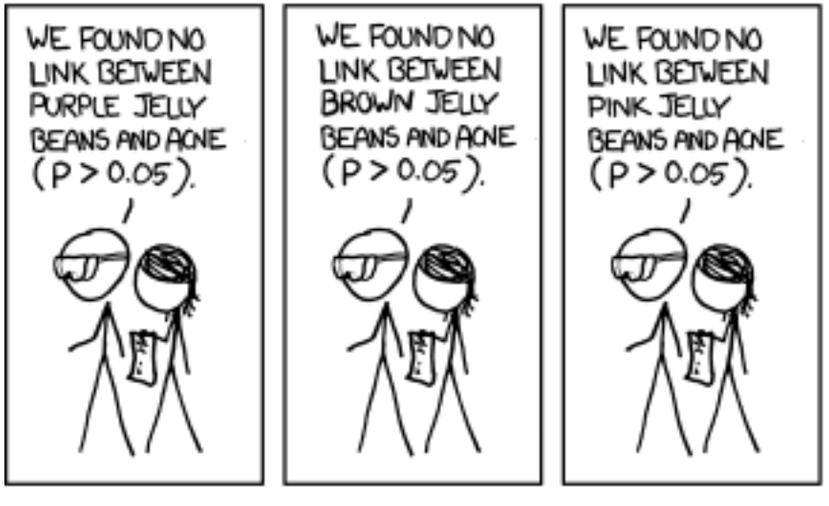




1. Multiple Hypothesis Testing

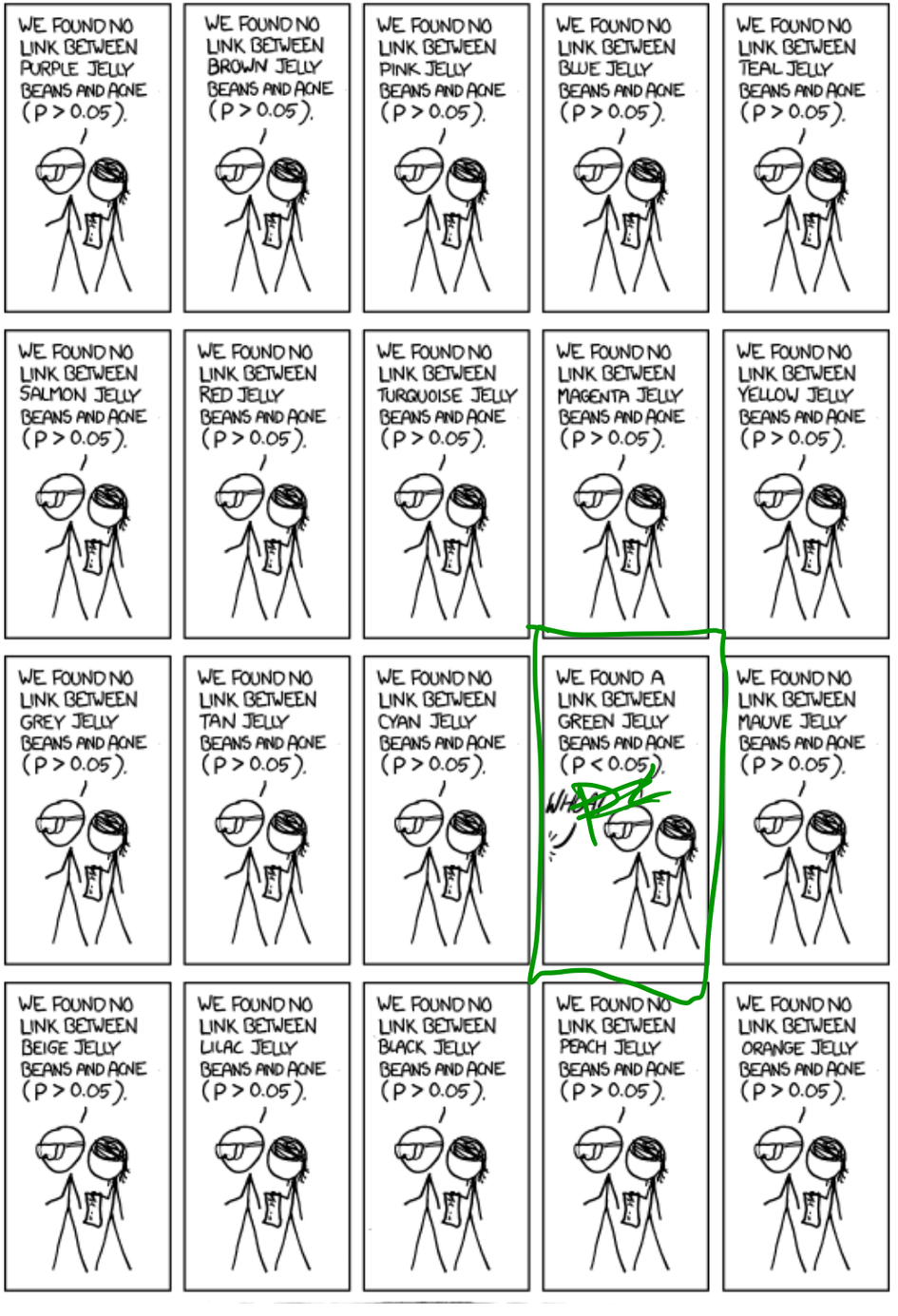
https://xkcd.com/882/





1. Multiple Hypothesis Testing

https://xkcd.com/882/

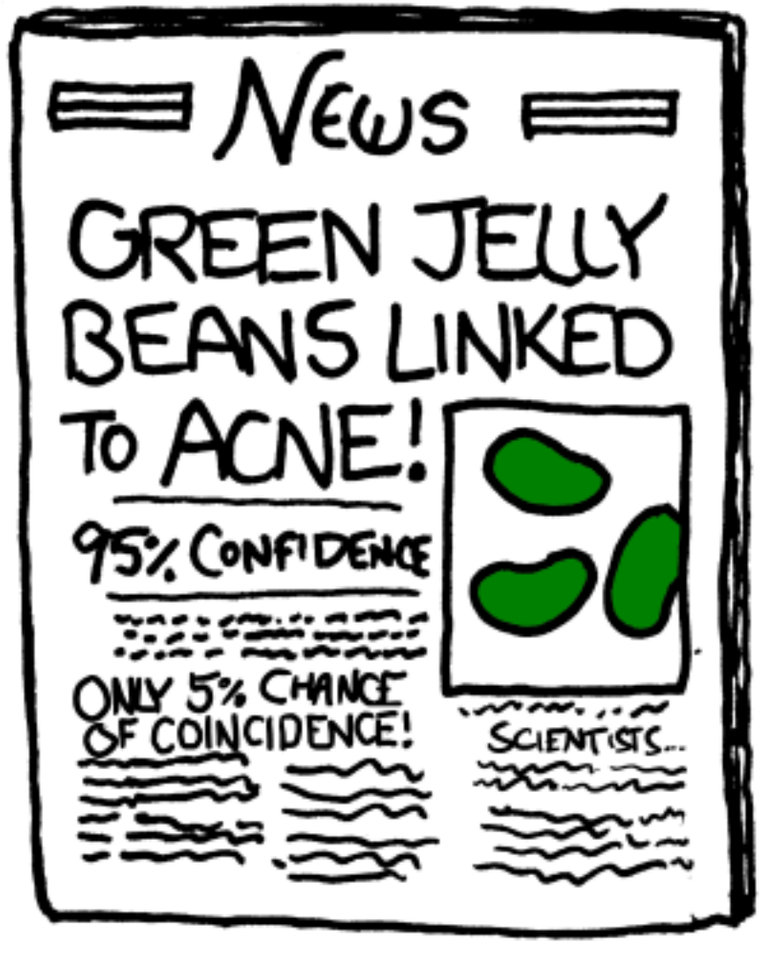


green jelly beans!

$P \leq 0.05$

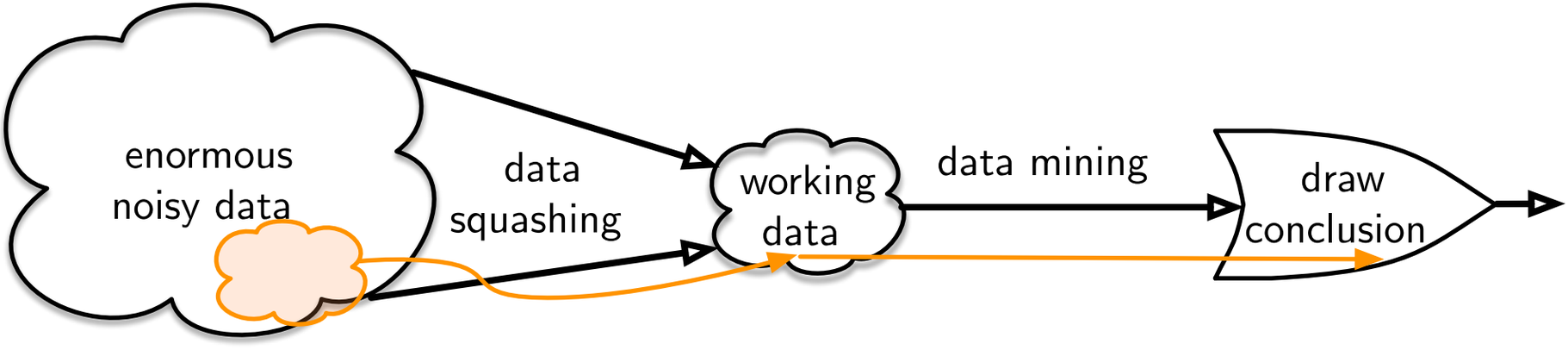
1. Multiple Hypothesis Testing

https://xkcd.com/882/

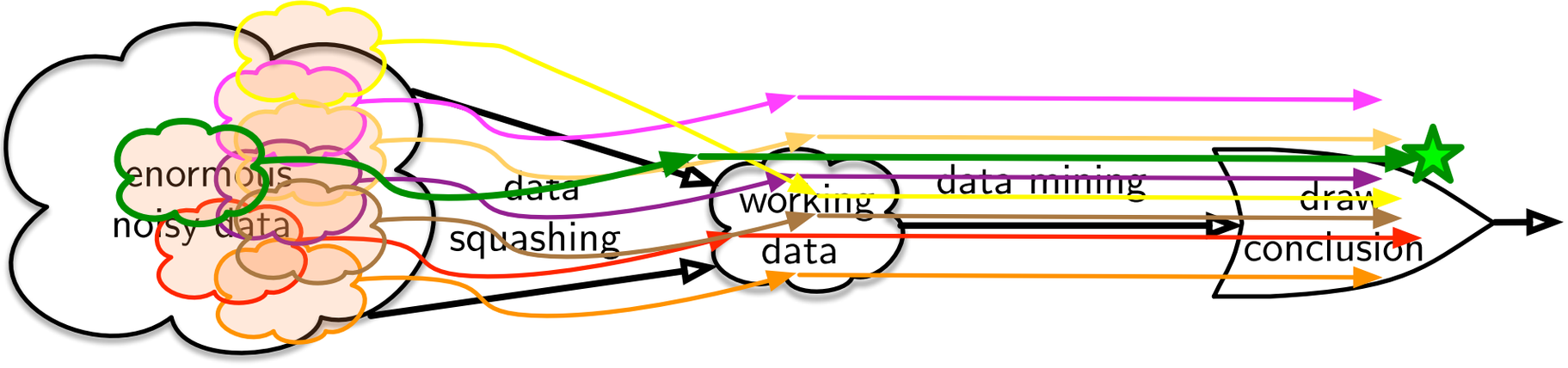


$\frac{1}{20} = 0.05$

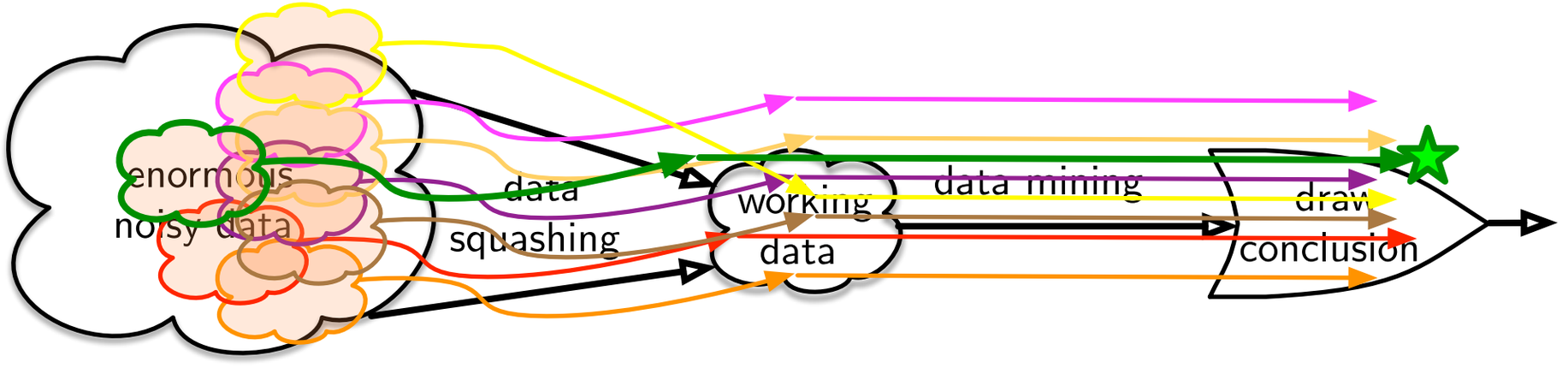
1. Multiple Hypothesis Testing



1. Multiple Hypothesis Testing



1. Multiple Hypothesis Testing



Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis PLOS 2:8, 2005

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the

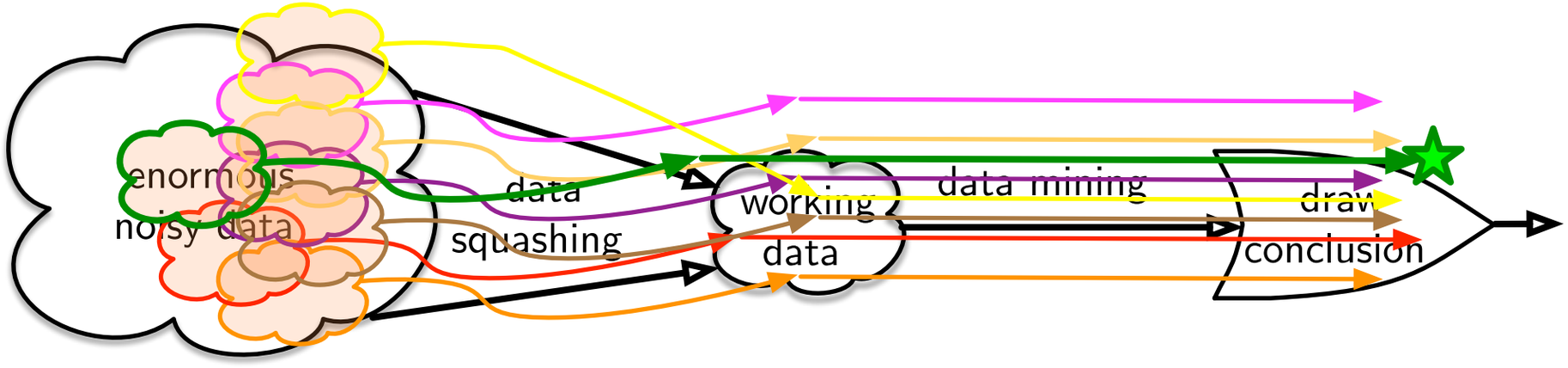
factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11]

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands

1. Multiple Hypothesis Testing



Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis PLOS 2:8, 2005

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–

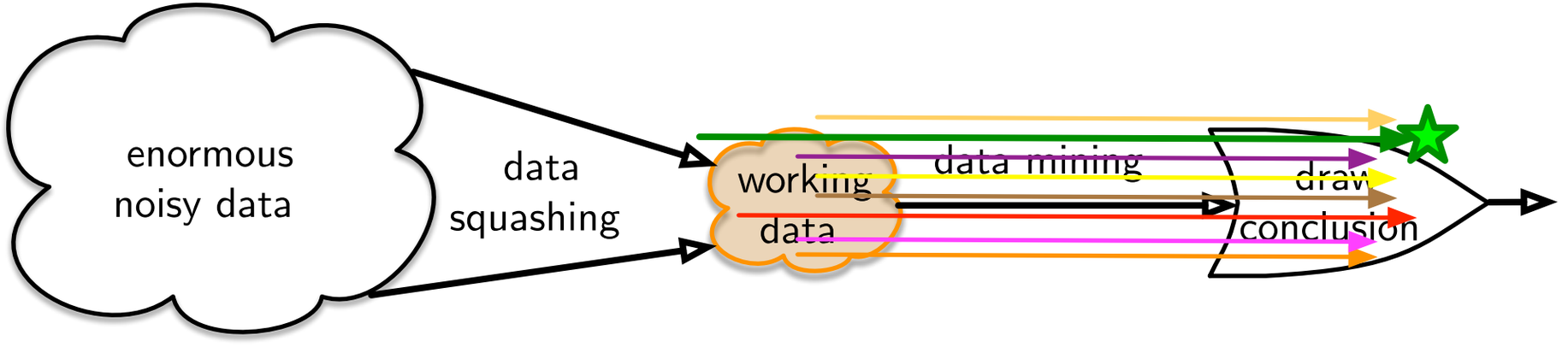
is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships

Bonferroni Correction?

20 hypotheses

$P \leq \frac{0.05}{20}$

1. Multiple Hypothesis Testing



Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis PLOS 2:8, 2005

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11]

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands

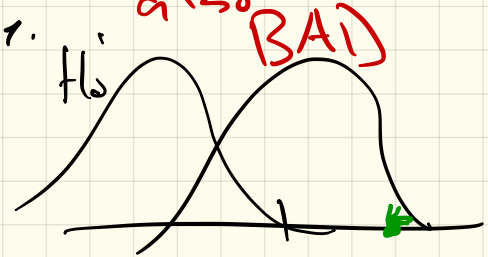
Bonferroni Correction?

Gelman + Loken 2013

1. Exactly / Reasonable H_0 (Fisher) ok

2. Pre state / Pre-register hypothesis, parameters ok

3. Fix test ahead of time. Use data to set parameters



4. p-hacking → Search parameter space for low p-value. BAD