New Statistics

Statistics with R

Basel R Bootcamp

April 2019

New Statistics

New Statistics = Old Statistics

Better studies

- Informative designs
- Sample size planning
- Registrations
- No -hacking
- Full reporting
- Replication

Better statistics

- Reporting uncertainty
- Lowering
- Bayesian statistics

from amazon.com

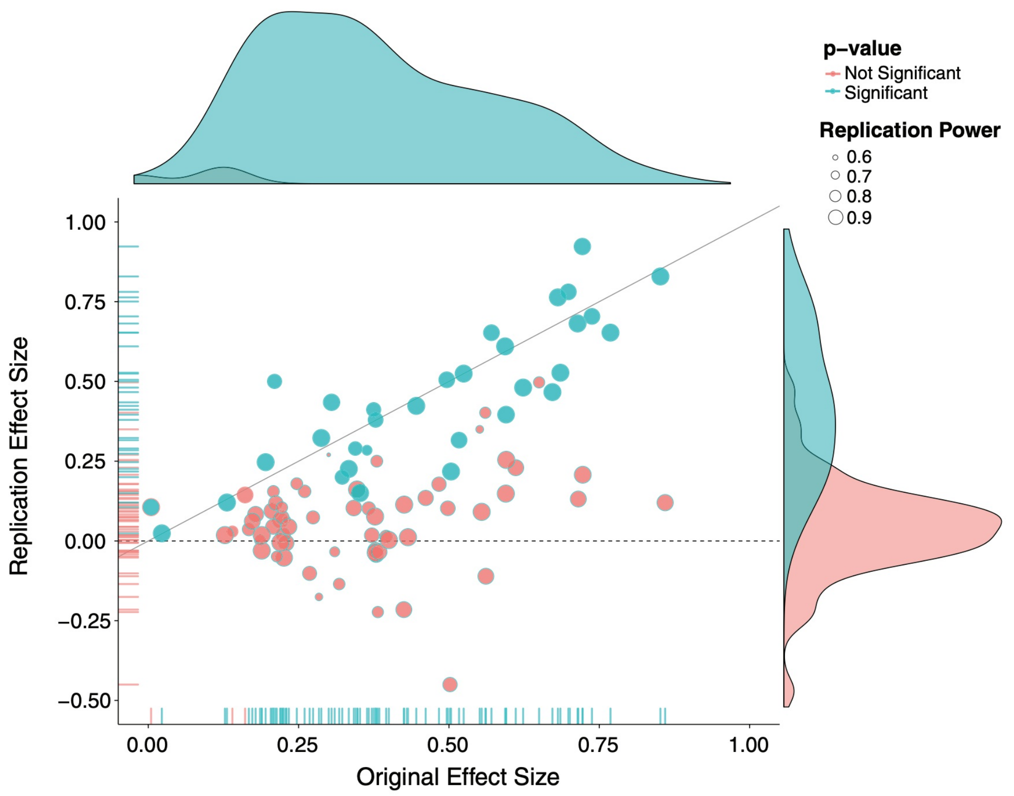
Replication crises

There are various **replication crises rumbling** in diverse academic fields.

Large-scale replication attempt in **Psychology** found that **only 36% were replicable**.

The low replicability is the result of **Questionable Research Practices**.

Similar assessments in **Medicine**, **Economics**, **Marketing**, **Social sciences**.



from **Open Science Collaboration**

Hallmarks of good studies

1 Informative designs

2 **Sample size planning**

3 Registrations

4 **No p-hacking**

5 Full reporting

6 Replication

adapted from Someecards.com

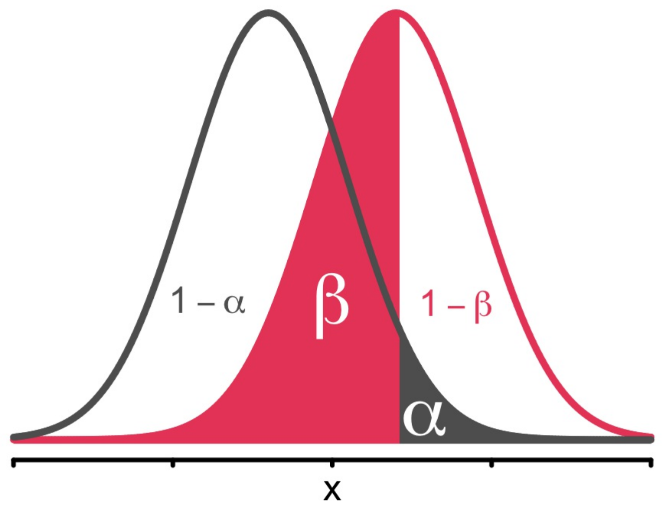
Sample size planning

Sample size should be planned such that there is **sufficient power to detect an effect**, if it is present.

Increasing sample size means **narrower sampling distributions** and **smaller decision errors**.

	Effect present	Effect absent
Significant result	1-β	α
Non-significant result	β	1-α

Small N

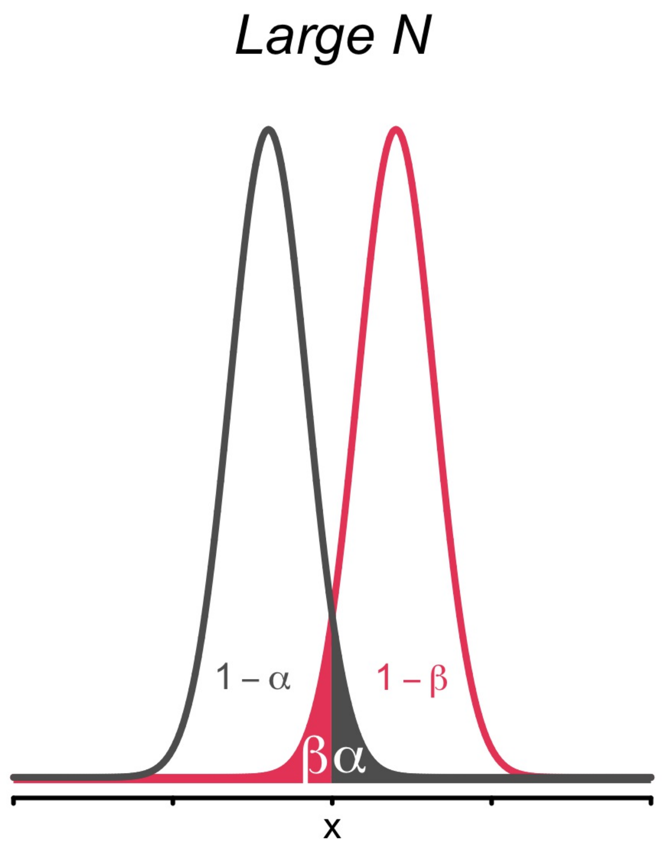


Sample size planning

Sample size should be planned such that there is **sufficient power to detect an effect**, if it is present.

Increasing sample size means **narrower sampling distributions** and **smaller decision errors**.

	Effect present	Effect absent
Significant result	1-β	α
Non-significant result	β	1-α



Sample size planning in R

Sample size should be planned such that there is **sufficient power to detect an effect**, if it is present.

Increasing sample size means **narrower sampling distributions** and **smaller decision errors**.

	Effect present	Effect absent
Significant result	$1\text{-}\beta$	α
Non-significant result	β	$1\text{-}\alpha$

```
# Load pwr package for pwr.t.test
library(pwr)

# N for large effect in t-test
pwr.t.test(sig.level = .05,
           power = .95,
           d = .8) # large effect
```

```
##
##      Two-sample t test power calculation
##
##               n = 41.59
##               d = 0.8
##       sig.level = 0.05
##           power = 0.95
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Sample size planning in R

Sample size should be planned such that there is **sufficient power to detect an effect**, if it is present.

Increasing sample size means **narrower sampling distributions** and **smaller decision errors**.

	Effect present	Effect absent
Significant result	$1-\beta$	α
Non-significant result	β	$1-\alpha$

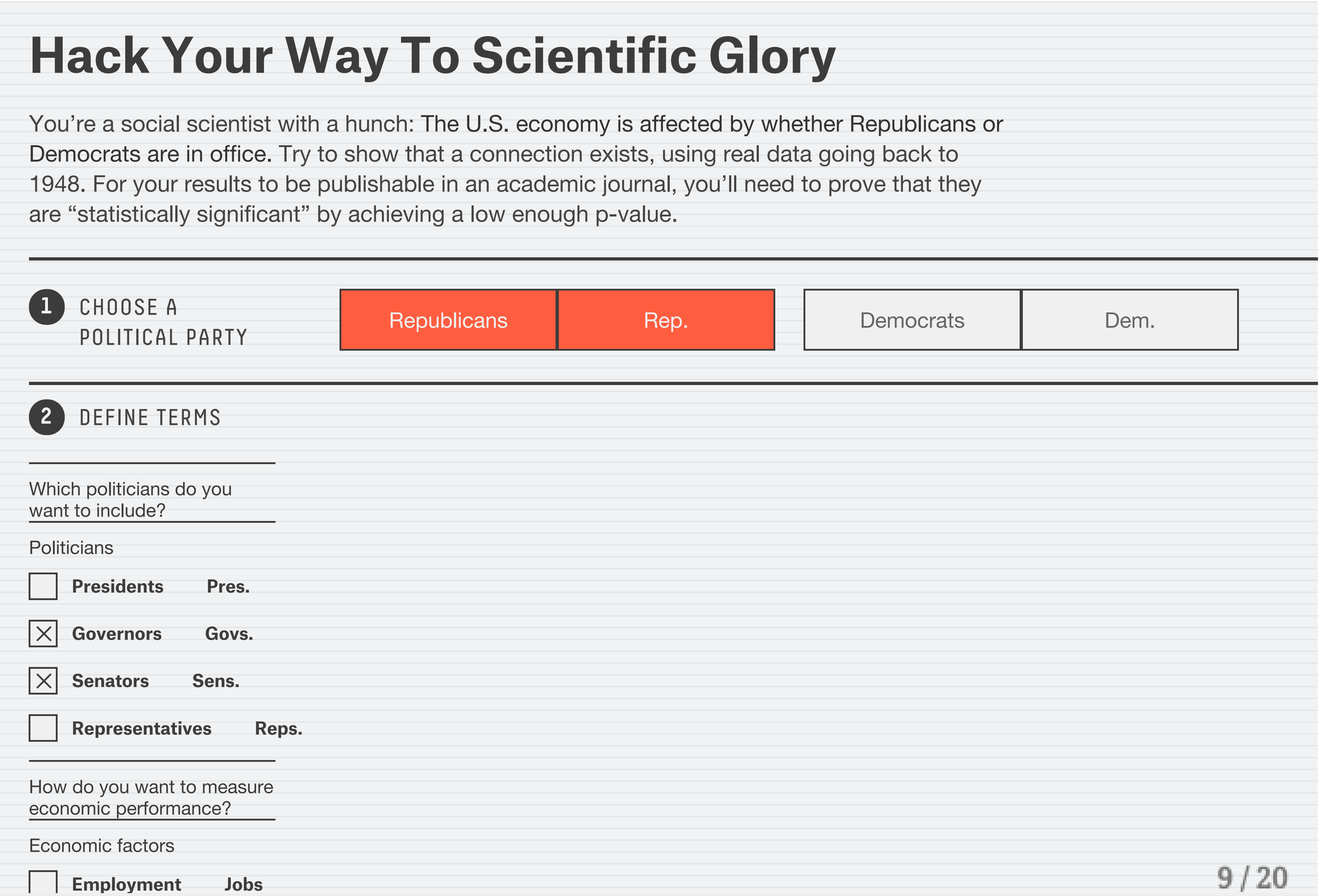
```
# Load pwr package for pwr.t.test
library(pwr)

# N for small effect in t-test
pwr.t.test(sig.level = .05,
           power = .95,
           d = .2) # small effect
```

```
##
##      Two-sample t test power calculation
##
##               n = 650.7
##               d = 0.2
##       sig.level = 0.05
##           power = 0.95
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

p-hacking

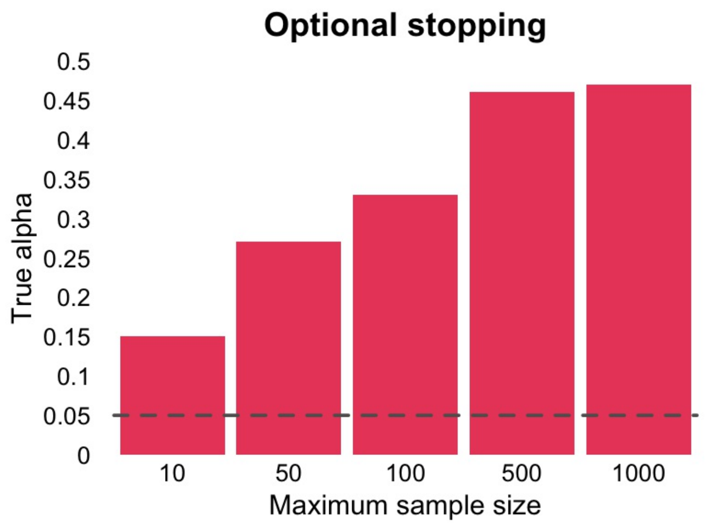
Ronald Coase

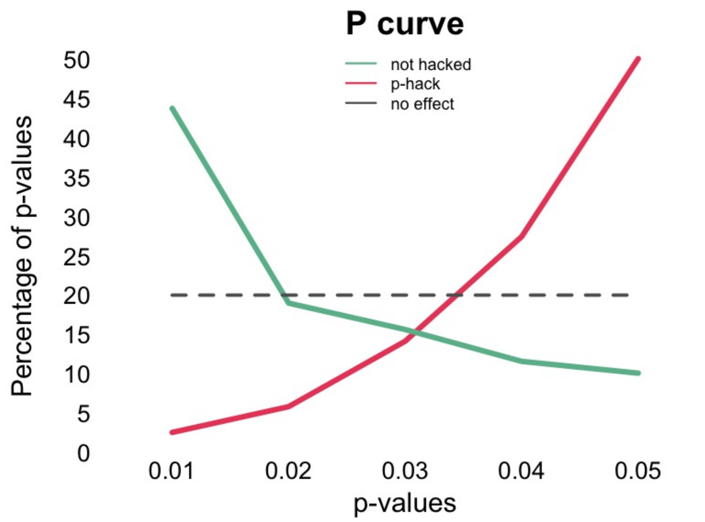
from fivethirtyeight.com



p-hacking

> **p-hacking is the misuse of data analysis** to find patterns in data that can be presented as statistically significant when in fact there is no real underlying effect.
>
> Wikipedia





Optional Stopping - Stopping data collection once significance is reached.

HARKing - Hypothesizing After the Data are Known. Committed when presenting any non-planned analyses, including the introduction of covariates.

Dos and Don'ts

Columbus looking for India, from history.com

Villemard vision for 2000, from sadanduseless.com

Everything.

Make predictions.
Predetermine sample size.
Predetermine analysis plan.
Register.

Present your result as confirmatory.

Engage in non-planned analyses.

New Statistics

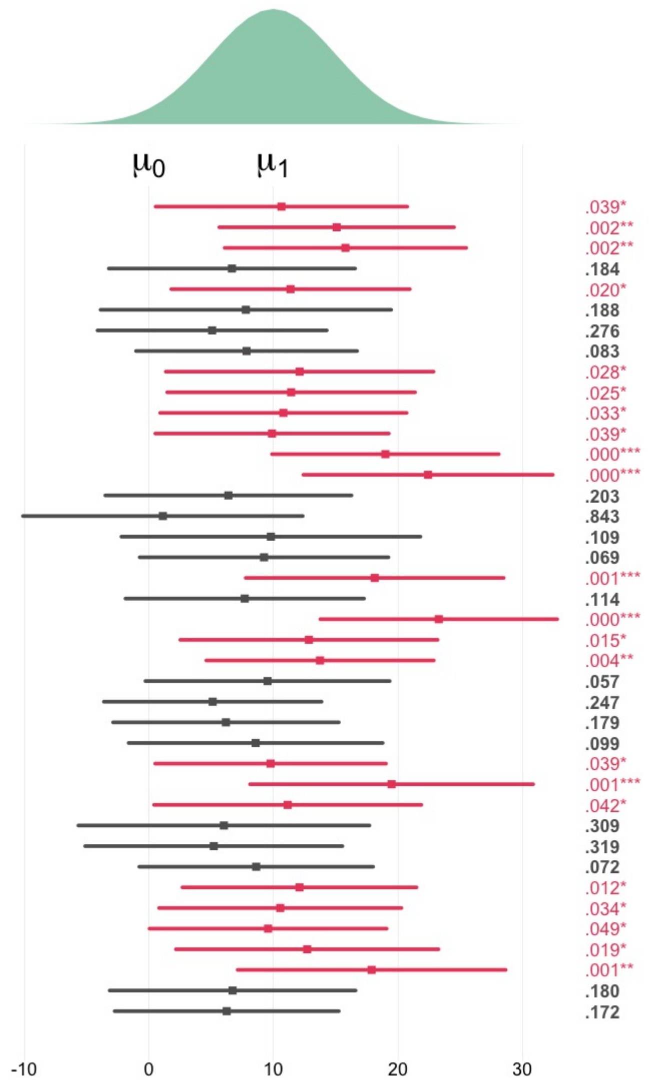
Problems with -values

- They are **difficult to interpret** as measures of evidence.
- They are used for **arbitrary binarization** via comparison to α.

New statistics

- Focus on **estimation**: rather than .
- Communication of **uncertainty**: rather than .

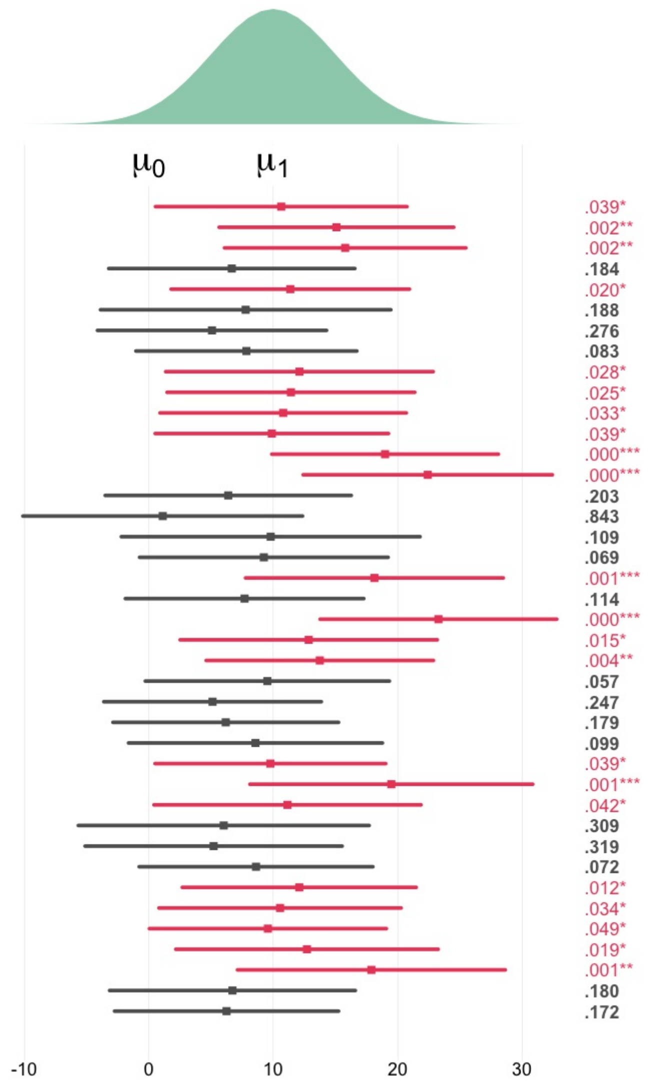
$$Confidence\,Interval\,(CI) = \bar{x} \pm t_{1-\alpha}\sigma_{\bar{x}}$$



Confidence interval

Confidence intervals essentially are **rearranged significance tests**.

t-test	$\dfrac{\lvert\bar{x}\rvert}{\sigma_{\bar{x}}} > t_{1-\alpha}$
Step 1	$\lvert\bar{x}\rvert > t_{1-\alpha}\sigma_{\bar{x}}$
Step2	$\lvert\bar{x}\rvert - t_{1-\alpha}\sigma_{\bar{x}} > 0$
CI	$\bar{x} - t_{1-\alpha}\sigma_{\bar{x}} > 0$ $\bar{x} + t_{1-\alpha}\sigma_{\bar{x}} < 0$



Confidence interval

Confidence intervals essentially are **rearranged significance tests**.

t-test	$\frac{\lvert\bar{x}\rvert}{\sigma_{\bar{x}}} > t_{1-\alpha}$
Step 1	$\lvert\bar{x}\rvert > t_{1-\alpha}\sigma_{\bar{x}}$
Step2	$\lvert\bar{x}\rvert - t_{1-\alpha}\sigma_{\bar{x}} > 0$
CI	$\bar{x} - t_{1-\alpha}\sigma_{\bar{x}} > 0$ $\bar{x} + t_{1-\alpha}\sigma_{\bar{x}} < 0$

```r
# standard confidence interval
N     <- length(t_1)
Delta <- mean(t_1 - t_2)
SE    <- sd(t_1 - t_2) / sqrt(N)
Delta + SE * qt(.95, N - 1) * c(-1, 1)
```

```
## [1] -1.1480 -0.7421
```

Confidence interval

Confidence intervals essentially are **rearranged significance tests**.

t-test	$\dfrac{\lvert\bar{x}\rvert}{\sigma_{\bar{x}}} > t_{1-\alpha}$
Step 1	$\lvert\bar{x}\rvert > t_{1-\alpha}\sigma_{\bar{x}}$
Step2	$\lvert\bar{x}\rvert - t_{1-\alpha}\sigma_{\bar{x}} > 0$
CI	$\bar{x} - t_{1-\alpha}\sigma_{\bar{x}} > 0$ $\bar{x} + t_{1-\alpha}\sigma_{\bar{x}} < 0$

```r
# bootstrapped confidence interval
bf <- function(x,ind) {
  sum(x[ind])/length(x[ind])
  }
boot_res <- boot(t_1 - t_2, bf, 1000)
boot.ci(boot_res)
```

```
## Warning in boot.ci(boot_res): bootstrap
## variances needed for studentized intervals

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot_res)
##
## Intervals :
## Level      Normal              Basic
## 95%   (-1.1777, -0.7113 )   (-1.1896, -0.7136 )
```

Bayesian statistics

Bayesian statistics extends classic (Frequentist) statistics by a **prior** distribution, specifying the prior **probability of the hypotheses** before seeing the data.

The prior permits calculation of a **true -value**, the **posterior probability**.

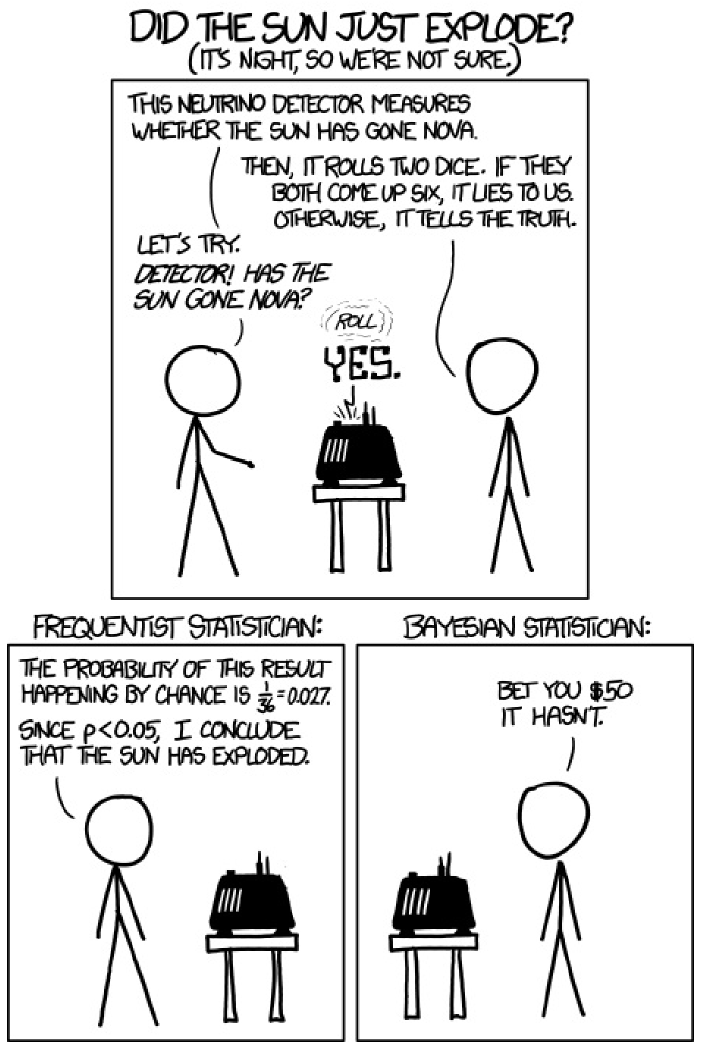
Posterior
Probability of the hypothesis given the data: *A real p-value*

Likelihood
Or the classic *p*-value

Prior
Initial belief in the hypothesis

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}$$

Normalization
Sum of numerators
$p(D) = \sum p(D|H_i)p(H_i))$



from xkcd.com

Bayesian statistics

For two reasons Bayesian statistics has long been rejected...

1 **Computational needs** to calculate $p(D)$

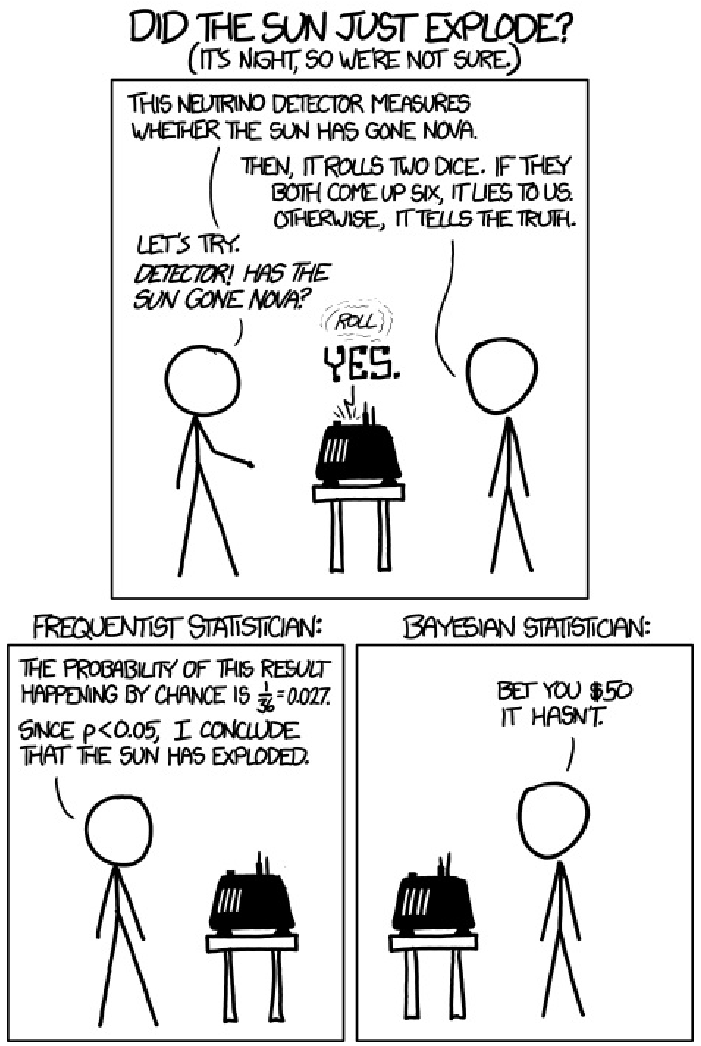
2 **Subjectivity** in choosing the prior

However,...

1 Computers have become **fast enough**

2 Statistical inference is **necessarily subjective**

3 The influence of priors can be limited through **ignorance priors**



from xkcd.com

Bayesian statistics

The rstanarm and BayesFactor packages make Bayesian stats very easy to use.

Function	Package	Description
stan_glm, stan_glmer	rstanarm	Bayesian (mixed) regression
ttestBF, anovaBF	BayesFactor	Standard h-tests
lmBF	BayesFactor	Bayesian (mixed) regressions

```
# Baysian stats with rstanarm
library(rstanarm)

stan_glm(formula = income ~ height,
         data = baselers)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      income ~ height
##  observations: 300
##  predictors:   2
## ------
##             Median MAD_SD
## (Intercept) 9292.9 2188.2
## height       -10.9   12.8
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 2635.5  107.2
##
## Sample avg. posterior predictive distribution of y:
##          Median MAD_SD
## mean_PPD 7439.5  211.7
##
## ------
## * For help interpreting the printed output see ?print.st
## * For info on the priors used see ?prior_summary.stanreg
```

Bayesian statistics

The `rstanarm` and `BayesFactor` packages make Bayesian stats very easy to use.

Function	Package	Description
stan_glm, stan_glmer	rstanarm	Bayesian (mixed) regression
ttestBF, anovaBF	BayesFactor	Standard h-tests
lmBF	BayesFactor	Bayesian (mixed) regressions

```r
# Baysian stats with BayesFactor
library(BayesFactor)

lmBF(formula = income ~ height,
     data = baselers)
```

```
## Bayes factor analysis
## --------------
## [1] height : 0.1813 ±0%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

Practical