

P-values

Bert van der Veen

Department of Mathematical Sciences, NTNU

Motivation

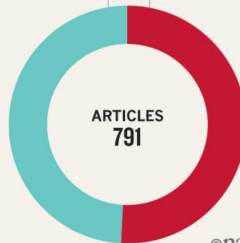
WRONG INTERPRETATIONS

An analysis of 791 articles across 5 journals* found that around half mistakenly assume non-significance means no effect.

*Data taken from: P. Schatz *et al.* *Arch. Clin. Neuropsychol.* **20**, 1053–1059 (2005); F. Fidler *et al.* *Conserv. Biol.* **20**, 1539–1544 (2006); R. Hoekstra *et al.* *Psychon. Bull. Rev.* **13**, 1033–1037 (2006); F. Bernardi *et al.* *Eur. Sociol. Rev.* **33**, 1–15 (2017).

Appropriately interpreted
49%

Wrongly interpreted
51%



©nature

Source: V. Amrhein *et al.*

Recap

Null hypothesis: There is no interaction of Time and Site

Fit M_0 :

```
nmodel <- glm(cbind(grahami, opalinus)~Time+Site,
               data = lizards, family="binomial")
```

Alternative hypothesis: There is an interaction of Time and Site

Fit M_1 :

```
amodel <- glm(cbind(grahami, opalinus)~Time*Site,
               data = lizards, family="binomial")
```

Lizards: LRT

```
(Lambda <- 2*(logLik(amodel)-logLik(nmodel)))
```

```
## 'log Lik.' 1.098943 (df=6)
```

```
k <- attr(logLik(amodel),"df")-attr(logLik(nmodel),"df")  
pchisq(Lambda, k, lower.tail=FALSE)
```

```
## 'log Lik.' 0.5772548 (df=6)
```

P-values: what are they?

We want to calculate some test-statistic, say Λ

The p-value is correctly interpreted as:

The probability to get this test-statistic or more extreme, assuming the null hypothesis is true

Unpacking:

- ▶ It is a probability
- ▶ Related to a **test statistic**
- ▶ (Usually) requires knowledge of its distribution

P-values: how to interpret them?

$P < 0.05$: This result is probably not due to chance $P > 0.05$:
This result is likely to due chance

- ▶ P-values from a single study will change!
- ▶ P-values are computed assuming the null hypothesis is true
- ▶ P-values are a function of **effect size** and **sample size**

Given enough data, an effect near zero will be statistically significant

P-values: how not to interpret them

- ▶ Probability of the null hypothesis
- ▶ Probability of wrongly rejecting the null hypothesis
- ▶ $P > 0.05$ as lack of effect
- ▶ Importance of the effect
- ▶ Anything else?


P-values are not a reflection of scientific importance

P-values as measure of evidence

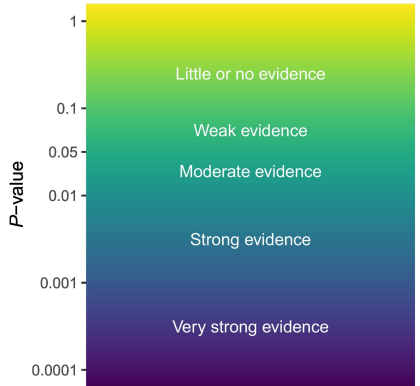
Trends in
Ecology & Evolution

Opinion

Rewriting results sections in the language
of evidence

Stefanie Muff ^{1,2,*}, Erlend B. Nilsen,^{2,3,4} Robert B. O'Hara,^{1,2} and
Chloé R. Nater^{2,3}

 Cell
OPEN ACCESS



Trends in Ecology & Evolution

Errors

Null hypothesis is true	Accurate	Too optimistic
Alternative hypothesis is true	Too conservative	Accurate
	Null hypothesis is accepted	Alternative hypothesis is accepted

Type 1 error

A false positive: mistakenly rejecting the null hypothesis.

- ▶ Controlled by α
- ▶ We choose $\alpha = 0.05$ usually

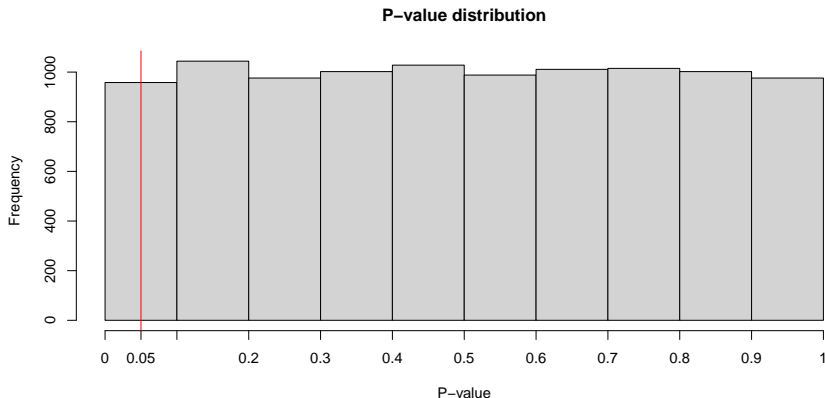
Type 2 error

A false negative: mistakenly accepting the null hypothesis.

- ▶ β
- ▶ Not controlled explicitly
- ▶ Implicitly controlled by the sample size

Distribution of p-values

Under the null hypothesis:



When the alternative hypothesis is true, the distribution is clustered near zero.

Example: Lizards

Fit M_0 :

```
nmodel <- glm(cbind(grahami, opalinus)~Time+Site,  
              data = lizards, family="binomial")
```

Alternative hypothesis: Diameter matters.

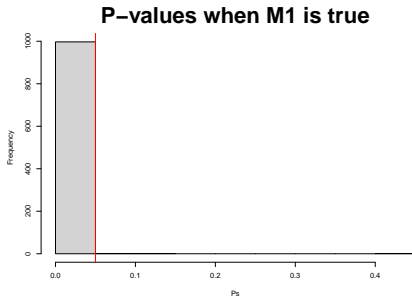
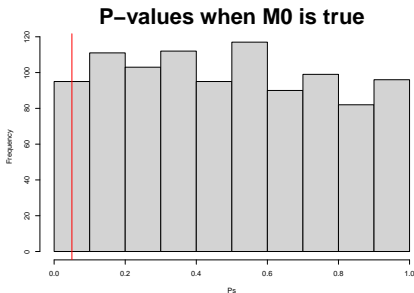
Fit M_1 :

```
amodel <- update(nmodel, formula = .~Time+Site+Diameter)
```

P-values by simulation

```
Ps <- NULL
for(i in 1:1000){
  ynew <- as.matrix(stats::simulate(nmodel))
  nmodel2 <- glm(ynew~Time+Site,
                 data = lizards, family="binomial")
  amodel2 <- update(nmodel2, formula = .~Time+Site+Diameter)
  # Store P-value
  (Lambda <- 2*(logLik(amodel2)-logLik(nmodel2)))
  k <- attr(logLik(amodel2),"df")-attr(logLik(nmodel2),"df")
  Ps <- c(Ps, pchisq(Lambda, k, lower.tail=FALSE))
}
```

P-values by simulation



P-values larger than 0.05: 0.95

P-values larger than 0.05: 0.003

To conclude

- ▶ 5% of the time we will wrongly accept the alternative hypothesis
- ▶ Thus **your** statistically significant result may be due to chance
- ▶ ?% of the time we will wrongly accept the null hypothesis
- ▶ “Highly significant” does not exist. Everything $p < 0.05$ is statistically significant
- ▶ Rejection of the alternative hypothesis does not mean “no effect”

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	} HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	
0.049	} SIGNIFICANT
0.050	} OH CRAP. REDO CALCULATIONS.
0.051	} ON THE EDGE OF SIGNIFICANCE
0.06	} HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P < 0.10$ LEVEL
0.07	
0.08	
0.09	
0.099	} HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥ 0.1	

Figure 1: <http://xkcd.com/>