Décision statistique et falsification en science:

Aller au-delà de l'hypothèse nulle!

Dominic Beaulieu-Prévost
Laboratoire des rêves
Département de psychologie
Université de Montréal

Statistical decision and falsification in science :

Going beyond the null hypothesis!

Dominic Beaulieu-Prévost
Laboratoire des rêves
Département de psychologie
Université de Montréal

The importance of statistical decisions in science

One of the main job of a researcher is the empirical evaluation of hypotheses.

And a majority of studies will base their decision to support or not a hypothesis on the statistical significance of their results.

In a way, the statistical procedure IS the decision process.

The model of statistical decision

State the competing hypotheses

$H_0 : \mu=0$

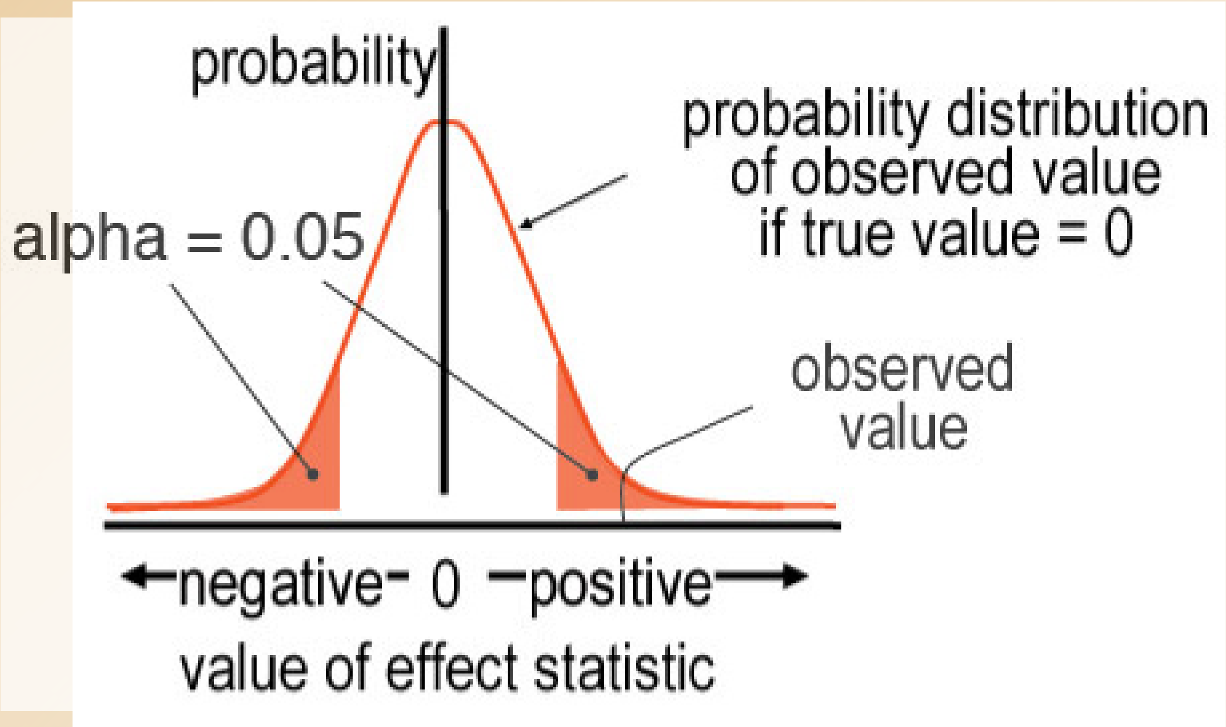
$H_1 : \mu\neq0$

Choose a probabilist decision criterion (alpha)

$\alpha=0.05$

The model of statistical decision

Compare the data to the model (H_0) and decide to reject or not H_0 by comparing the *p* value to the alpha.



The model of statistical decision

Take into account the risk of type II error through an evaluation of the statistical power

	H_0 is true	H_0 is false
Reject H_0	Type I	Correct
Accept H_0	Correct	Type II

But what can we get from a statistical test of significance?

1. The probability that H_0 is false?
2. The probability of doing a type 1 error if we reject H_0?
3. The probability that an experimental replication produces statistically significant results (by calculating 1-*p*)?
4. The probability that the decision (reject H_0 or not) is correct?
5. The probability that the data are the result of chance?
6. The probability to obtain results as extreme as these if H_0 is true?

But what can we get from a statistical test of significance?

Differentiating p(D/H) and p(H/D)

Analogy:

What is the probability of being dead if I was hung ?

And what is the probability of having been hung if I'm dead?

Major problems with the model

1. The relation to N
2. The logical improbability of H_0
3. The lack of plausibility of H_0
4. Consequences...

Major problems with the model

The relation to N (sample size)

p is related both to the effect size and to the sample size

Consequence 1

For a specific effect size, *p* is an index of sample size

Consequence 2

You will ALWAYS have a statistically significant result if your sample size is big enough

Major problems with the model

<u>The logical improbability of H_0</u>

The non-equivalence of H_0 and H_1

("point" vs "interval" hypotheses)

The improbability of a point hypothesis on a continuous scale

The precision of an hypothesis limits its logical probability of being true

<u>Consequence</u>

H_0 is ALWAYS false ($1/\infty$), H_1 is ALWAYS true and the probability of a type 1 error is nil ($1/\infty$)

Major problems with the model

The lack of plausibility of H_0

The notion of ambient correlational noise (Lykken, 1968) or "crud factor" (Meehl, 1990)

Major problems with the model

Consequences...

We try to reject a hypothesis that we already know is false.

It's like trying to boost our confidence in the "scientific" hypothesis by killing a dead man...

Is there a way out !?

Going back to basics:

1) The falsification principle

2) What are you trying to do?

Constructing a model? (parameters estimation)

- or -

Evaluating a model? (hypothesis testing)

Simple solutions !!

Parameters estimation

- Standard parameters estimation
- Impact estimation

Hypothesis testing

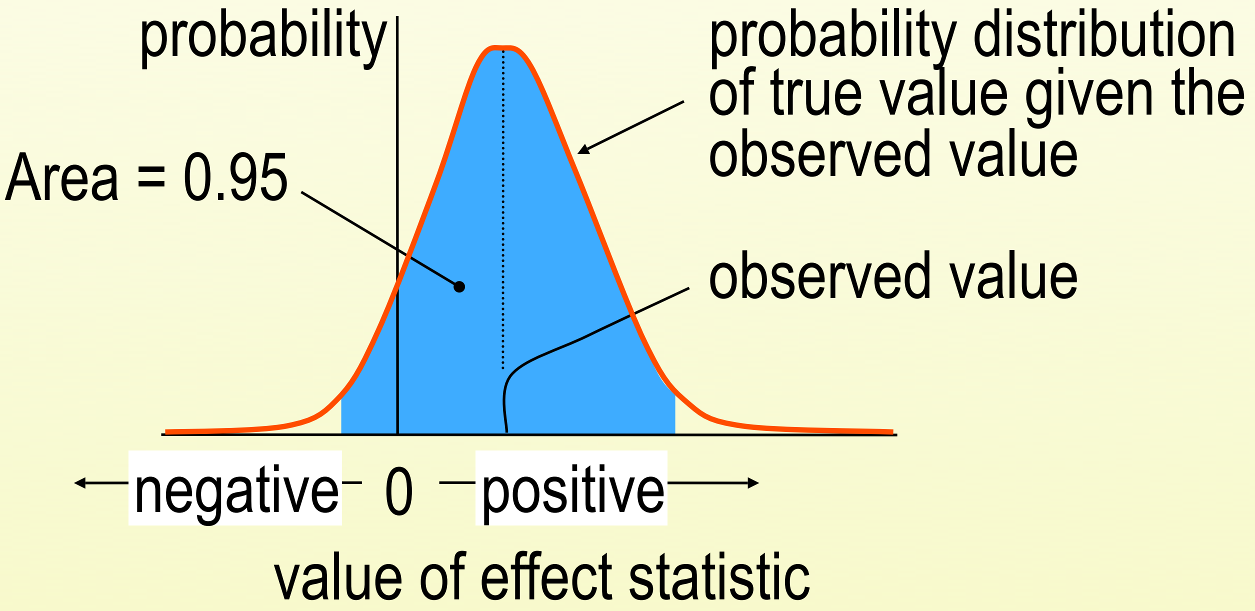
- Irrelevance hypothesis
- Comparison of relevant hypotheses

Parameters estimation

Standard parameters estimation

Confidence intervals for :

- raw scores
- mean differences
- correlations
- proportions

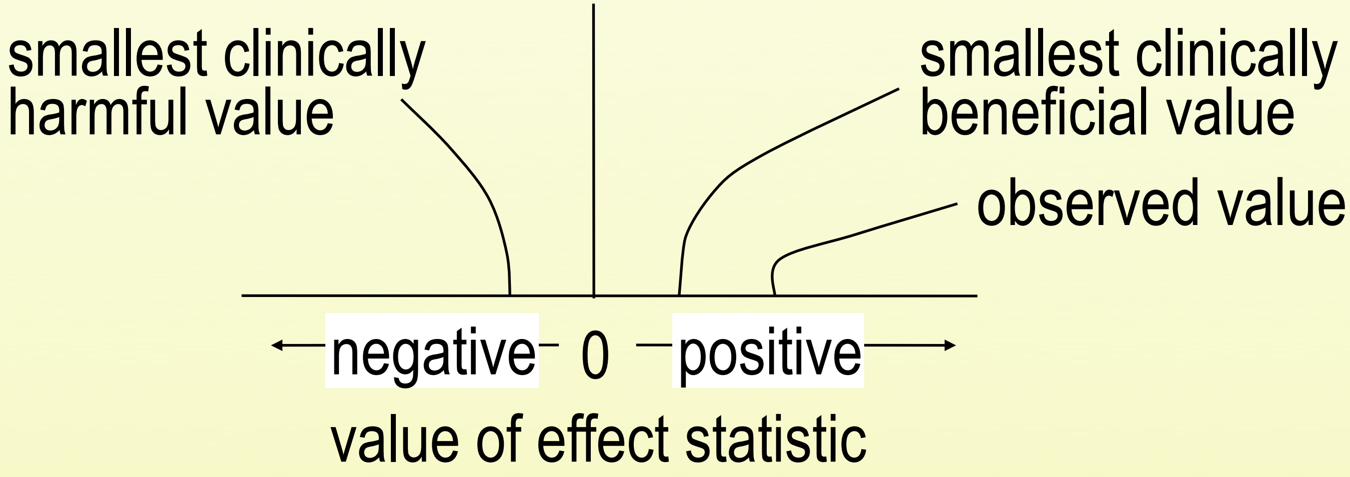


It specifies the precision of the estimation

Parameters estimation

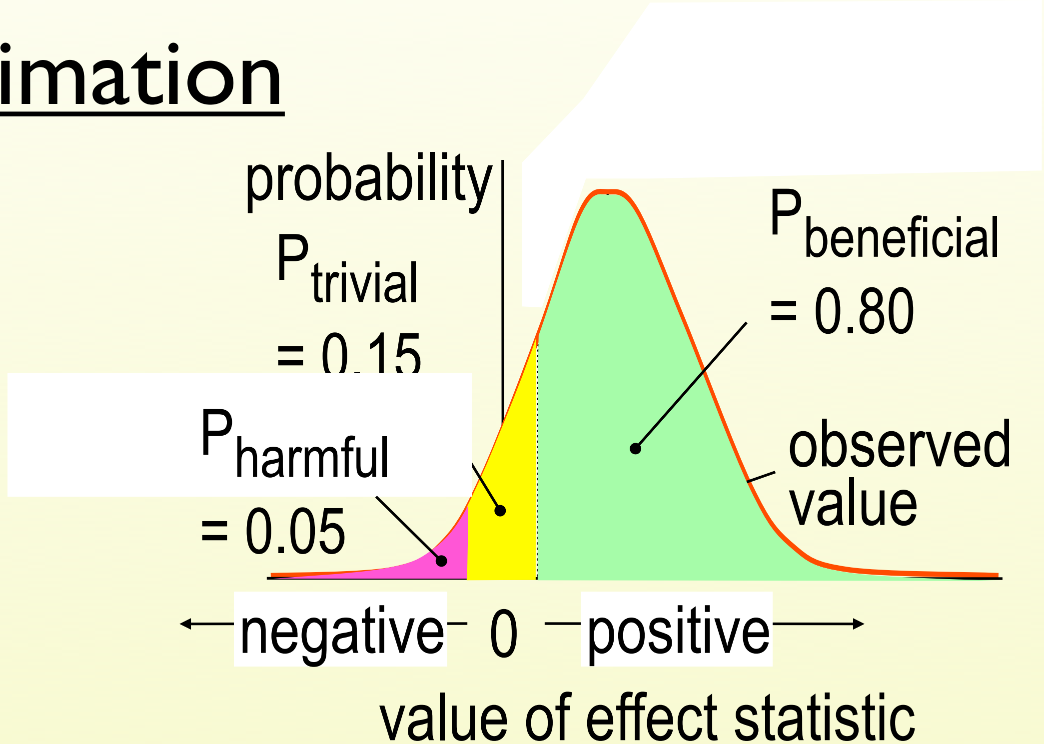
Impact estimation

Evalute the smallest beneficial effect possible and the smallest harmful effect (usually equal in size and opposite in sign)



Parameters estimation

Impact estimation



It defines the probabilities that the effect could be beneficial, trivial or harmful.

Hypothesis testing

Testing against the irrelevance hypothesis

The maximum level of irrelevance has to be quantified according to:

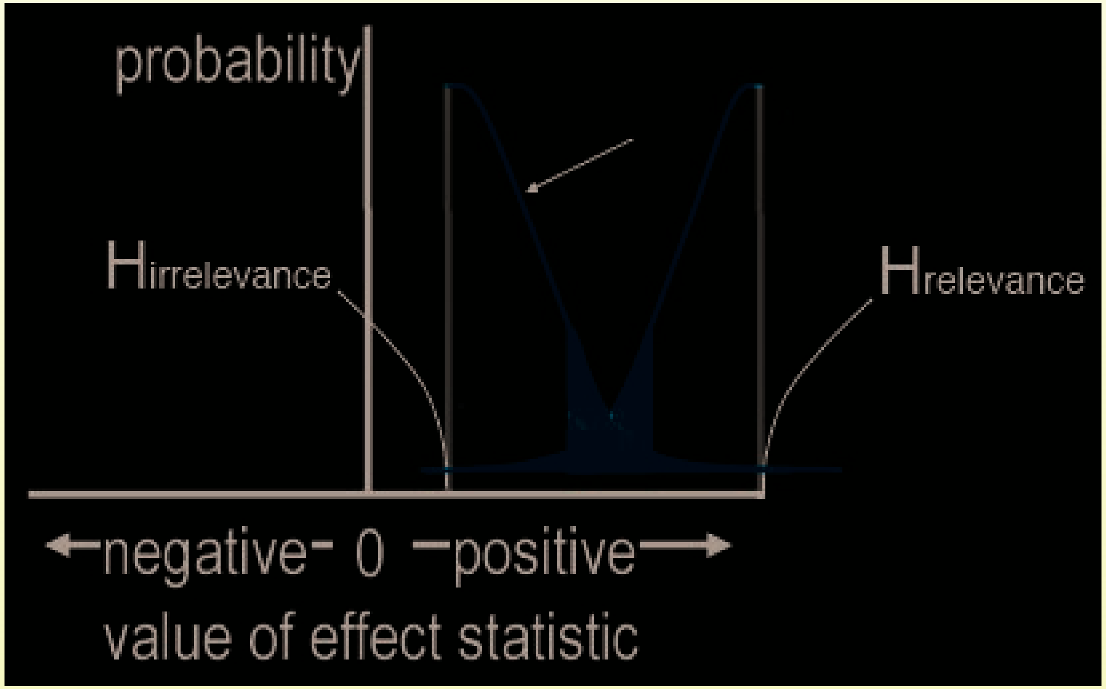
- The theoretical importance of the effect
- The potential cost of an intervention
- And minimally, the sensitivity of the scale!

Hypothesis testing

Testing against the irrelevance hypothesis

$H_{irrelevance}$: $r \leq 0.10$ (i.e. 1% of explained variance)

$H_{relevance}$: $r \geq 0.30$ (i.e. 9% of explained variance)



Hypothesis testing

Comparison of relevant hypotheses

Between treatments A and B, can we evaluate which one is better?

The importance of defining the smallest significant difference

Define H_1 and H_2 and evaluate their "explanatory success"

$H_1 : \mu \geq$ smallest significant difference

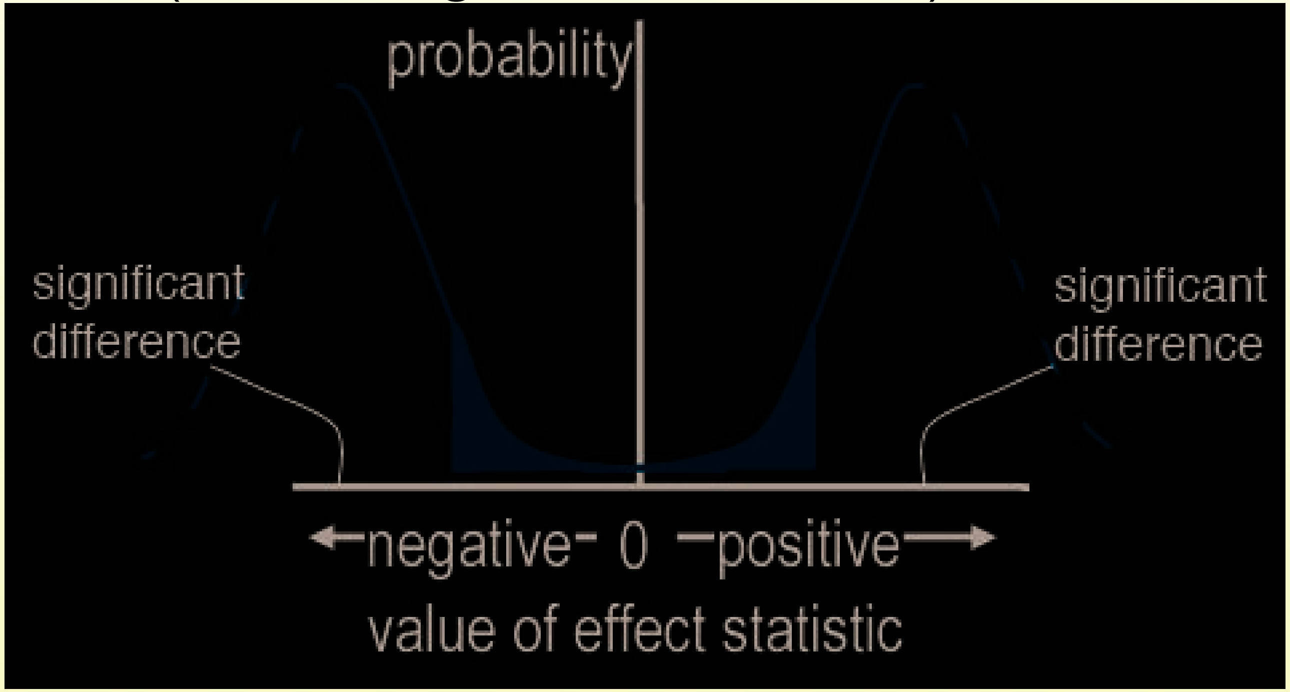
$H_2 : \mu \leq$ -(smallest significant difference)

Hypothesis testing

Comparison of relevant hypotheses

H_1 : u ≥ smallest significant difference

H_2 : u ≤ -(smallest significant difference)



Hypothesis testing

Other options...

Relative likelihood ratio

Test of equivalence

...

In summary...

Focus on :

- the effect size
- the precision of the estimation (C.I.)

Remember the dead man and operationalise your hypotheses instead of testing a useless H_0

Understand what you're measuring

Decide what is relevant

... and play with your data !