

The Concept of the Null Hypothesis

NRC 601

Research Concepts in Natural Resources

Department of Natural Resources Conservation
University of Massachusetts Amherst

Fall 2009

Null:

- definition in Webster's dictionary (1993).
- from the Latin nullus, ne- "not" + ullus "any".
- 1. having no binding force: invalid; 2. amounting to nothing: nil; 3. having no value.
- hypothesis is singular; hypotheses is plural.

Null Hypothesis:

- a statement of no difference.
- a starting point of a scientific investigation.
- a way of **simplifying** our approach and conforming to the concept of **falsification**.

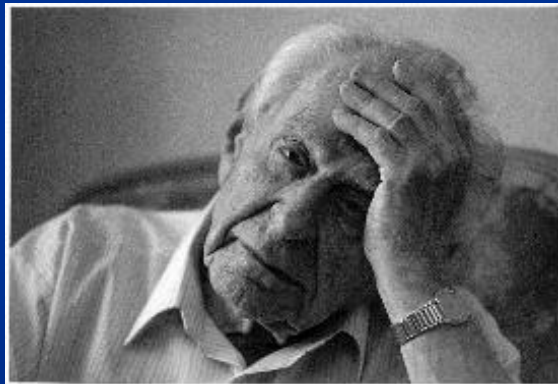


Francis Bacon:

Promote negative instances . . .

Try to disprove an idea . . .

Cannot prove beyond a doubt



Karl Popper:

*Our belief in any natural law cannot have
a safer basis than our unsuccessful
attempts to refute it.*



Charles Romesburg:

Use the H-D method or else . . .

Sir William of Ockham (c 1285-1349)

- English scholar and Franciscan monk
- taught at Oxford University 1310-1324
- an empiricist; questioned current philosophies and the power of the pope – charged with heresy because of his Master's thesis
- thought to have died of the bubonic plague



Believed that unnecessary complexity was vain and insulting to God . . .

Occam's razor: "What can be done with fewer [assumptions] is done in vain with more."

aka **The Principle of Parsimony**

Null hypotheses do two things:

- (a) Restructure our question into the falsification framework, as proposed by Bacon, Popper, et al.;
- (b) Try to account for patterns in the data in the simplest way possible, which usually means . . .

 . . . that we attribute variation in the data to randomness or measurement error.

For example:

- We want to know if snowmobiles result in increase stress in wolves.
- We compare the hormone levels in blood of wolves exposed to snowmobiles to hormone levels in blood of wolves not exposed to snowmobiles.
- The means of the 2 groups are not likely to be exactly equivalent . . . The question is, is the difference enough to be due to some external force (i.e., snowmobiles) other than random variation alone?

The null hypothesis is paired with an alternative:

- (a) The null states that there is “no pattern”; i.e., no difference between (2) or among (>2) groups, or no relationship between two continuous variables.
- (b) The alternative states that a pattern does exist; i.e., there are distinct differences between or among groups, or a clear relationship exists between two continuous variables.
- (c) If the alternative is the case, then you must ask how such patterns relate to the scientific hypothesis you are testing.

Statistical Hypotheses versus Scientific Hypotheses:

- (a) The statistical hypothesis tests for a pattern, and helps you decide, based on the test statistic, if there is no pattern (NULL) or there is a pattern (ALTERNATIVE) at some specific level of probability.

- (b) A scientific hypothesis is your candidate explanation for how or why that pattern exists . . .
 - [. . . keeping in mind that we can, and should, have multiple competing hypotheses.]

Intermediate Disturbance Hypothesis = biodiversity will be greatest where disturbance is intermediate, and lower at the two extremes of very low disturbance and very high disturbance.

Does this apply to human development? Hypothesize that biodiversity of dragonflies will be highest at intermediate development levels.

$$H_o: B_{\text{HIGH}} = B_{\text{MED}} = B_{\text{LOW}}$$

$$H_{A1}: B_{\text{HIGH}} \neq B_{\text{MED}} \neq B_{\text{LOW}}$$

$$H_{A2}: B_{\text{HIGH}} < B_{\text{MED}} > B_{\text{LOW}}$$

A sidebar:

The “Intermediate Disturbance Hypothesis” might more accurately be described as a prediction, i.e., we predict that biodiversity will be greatest where disturbance is intermediate, and lower at the two extremes of very low disturbance and very high disturbance.

Our hypotheses, i.e., our candidate explanations for this phenomenon, could be the following . . .

In areas with intermediate levels of disturbance, there is . . .

- more structural complexity, and thus more niches;
- more diversity in food resources;
- more competition among predators;
- less opportunity for habitat specialists . . . Etc.

The procedure:

- Set up H_0 (the NULL) and H_A (the ALTERNATIVE).
- Conduct an appropriate statistical test and produce the *test statistic*, which is the numerical result of the test, and an associated *probability value* (or *P-value*).
- If the test statistic is sufficiently large, and the associated *P-value* sufficiently small, you reject the null of no difference and conclude that, in our case, the diversity of dragonflies is different among the 3 levels of human disturbance.

Thus, we rejected H_0 in favor of H_A .

If our test statistic was low and P was high, we would NOT *accept* the null,

. . . we would *fail to reject* the null.

We would state that there is no evidence that level of disturbance is related to diversity of species in our study,

. . . but we would not be able to say that this is never the case.

***P*-values:**

Traditionally, we pre-select some level of alpha (i.e., P) to indicate “significance”

Usually, $P \leq 0.05$. . . But in field biology,
we sometimes use $P \leq 0.10$

Now, if we reject the null: $H_0: B_{\text{HIGH}} = B_{\text{MED}} = B_{\text{LOW}}$

and simply report, biodiversity differed among areas with different levels of disturbance ($P \leq 0.05$) – not very informative.

But if we report means and variances (SE) for each area, and the test statistic, degrees of freedom, and exact P -value – much more informative.

***P*-values, more:**

The lower the *P*-value, the more confidence you can be:

$P \leq 0.01$, you might say the difference is highly significant

$P \leq 0.05$, you might say the difference is significant

$P \leq 0.10$, you might say the difference is suggestive
(although I would probably say significant)

What you would want to report (and *not* report):

$P = 0.04$ --- the exact *P*-value.

$P = 0.04$ or maybe $P \leq 0.039$ --- limit significant digits.

$P = 0.04$, not $P = .04$ --- no naked decimals.

And always include the means and some measure of variability,
and the test statistic and degrees of freedom.

***P*-values, still more:**

Our primary goal is to reject the null, but we don't want to say there is a difference (i.e., reject H_0) if there really is not:

H_0 : $B_{\text{HIGH}} = B_{\text{MED}} = B_{\text{LOW}}$ is really true (all are equal) but
we say they are not equal

If we did this, we would be making a TYPE I ERROR.

Some researchers refer to this as “spurious results.”

This hurts our credibility as good scientists, so we set alpha low (e.g., ≤ 0.05) to avoid this particular problem.

***P*-values, and still more:**

However, if there really is a difference, we don't want to miss it:

$H_0: B_{\text{HIGH}} = B_{\text{MED}} = B_{\text{LOW}}$ is really not true (all are not equal), but we say they are equal

This would be a TYPE II ERROR, which could be bad if some treatment (e.g., a pollutant, clearcutting) truly hurts an endangered species.

Studies with low sample sizes (n) are especially prone to Type II errors . . . *How might you try to counteract this?*