A brief glance at hypothesis testing

Sundar Srinivasan
IRyA-UNAM

DAWGI Meeting
2024.05.29

Hypothesis testing

"Do the data provide **sufficient evidence** to conclude that we **must depart from our original assumption** concerning the state of Nature?"

– J. C. Watkins, *An Introduction to the Science of Statistics*

Hypothesis: a statement concerning the state of Nature. Can be tested using data.

Simple (complete description of underlying distribution) e.g., "the errors are Gaussian with mean 0, std 1".

Composite (underlying population distribution unclear) e.g., "the mean is not 0".

Two-tailed/non-directional e.g., "$-5 \leq M \leq 5$", "$|g| > 6.5$" or one-tailed/directional e.g., "$M \geq 5$", "$g < 6.5$".

Involves two statements: one is the **default belief**, typically stating that the current observation is not "out of the ordinary" ("**null hypothesis**"); the other states that the data is inconsistent with the default belief ("**alternate hypothesis**").

We either conclude that the data is inconsistent with the "default" (i.e., we reject the null hypothesis), or that it is, indeed, consistent (i.e., the null hypothesis cannot be rejected).

General procedure

1. What is the question you need answered? Frame it well!
2. Define null (H_0) and alternate (H_a) hypotheses based on this question.
3. Decide/choose statistic.
4. Decide threshold/tolerance.
5. Under H_0, compute probability distribution of statistic.
6. Compute observed value of statistic.
7. Compute *p*-value (probability of statistic being “more extreme” than the value observed).
8. If *p*-value < tolerance, reject H_0. Else, unable to reject H_0.

Formulating the right question

What kinds of questions can be answered using hypothesis tests?

(from Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*)

~~What is the straight line fit for y vs. x?~~ → Does y increase with x?

~~What is the strength of the effect?~~ → Is the effect present?

~~What is the value of the correlation?~~ → Is there a correlation?

~~What are the values of a and b?~~ → Do a and b have the same value?

~~Are the properties of these two samples identical?~~ →

Are the means equal?

Are the samples drawn from the same distribution?

The *p*-value

The probability that, **given the null hypothesis**, a value equal to (or more extreme than) the observed data is obtained.

Example: coin tosses; H_0 = coin is fair, H_a = coin is biased

Observation: 8 heads in 10 tosses

Statistic: number of heads

Distribution: if coin is fair, number of heads follows the **binomial distribution.**

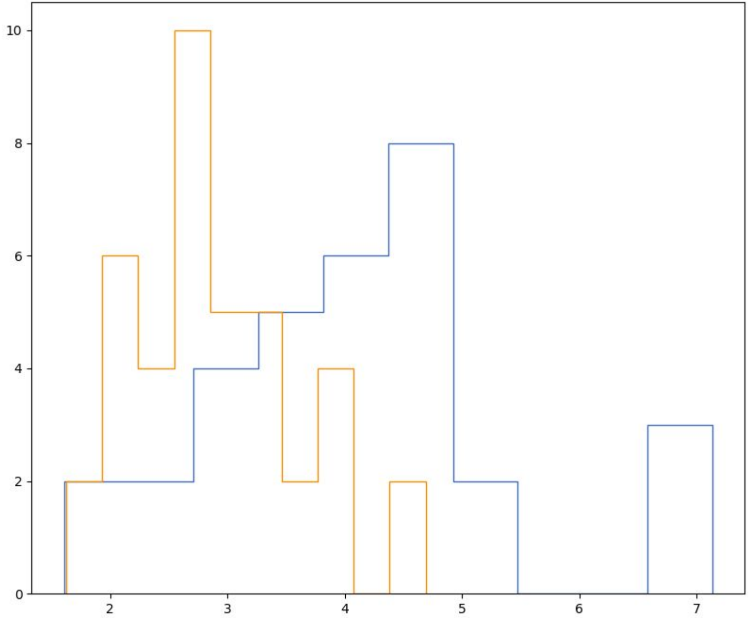
Threshold: 5%

Using the binomial distribution, p-value = P(9 H) = ${}^{10}C_8\ (0.5)^8\ (1\text{-}0.5)^2 = 45\ (0.5)^{10} \approx 0.044$

p-value < threshold ⇒ H_0 rejected.

Example: comparing two samples

Researcher: “I want to prove that these two samples have different properties, so I fit them with Gaussians and showed that their means are well separated”



This is bad because (for example)

(a) it assumes that the samples are drawn from Gaussians (**model-dependent**)

(b) a bad Gaussian fit in this case means a larger standard deviation and therefore LESS evidence for a statistically significant difference in the means. Difficulty in convincing referee of your (by-eye) conclusion.

Example: comparing two samples

Better questions can be framed and tested with the raw data (no need for a histogram, either!):

“Are the two samples drawn from Gaussian distributions with the same means?” **Assumes a model**!

→ perform a 2-sample t-test assuming samples are independent (can also drop this assumption).

```
t_statistic, p_value = scipy.stats.ttest_ind(data1, data2)
    Performing 2-sample t test for unequal variances (Welch's t-test)
    Null hypothesis: the means of the two samples are equal.
    The estimated p-value is 0.000028.
    The null hypothesis is rejected!
```

The probability that the relative difference between the sample means is more extreme than the observed value of `t_statistic` is much smaller than the threshold, so it is extremely unlikely that the data were drawn from distributions with the same means.

Example: comparing two samples

Better questions can be framed and tested with the raw data (no need for a histogram, either!):

“Are the two samples drawn from the same distribution”? Does not assume a model! **Nonparametric test.**

→ perform a 2-sample Kolmogorov-Smirnov (or Anderson-Darling) test.

```
ks_statistic, p_value = scipy.stats.ks_2samp(data1, data2)
     Performing 2-sample Kolmogorov-Smirnov test
     Null hypothesis: both samples are drawn from the same distribution.
     The estimated p-value is 0.000070.
     The null hypothesis is rejected!
```

We are able to reject the hypothesis that the two samples are drawn from the same distribution (**regardless of what shape that distribution has**, which is the model-free part of the solution).

Code and follow up

Raw data and code for this example available on the DAWGI Github!

https://github.com/dawg-at-irya/hypothesis-testing/tree/main

Have data, want to frame the right question(s)? Come talk to me!