Statistical Test

A **test of significance** is a formal procedure for comparing observed data with a claim (also called a hypothesis) whose truth we want to assess.

- The claim is a statement about a parameter, like the population proportion p or the population mean μ.
- We express the results of a significance test in terms of a probability that measures how well the data and the claim agree.

Set up the Hypotheses

The claim tested by a statistical test is called the **null hypothesis (H_0).** The test is designed to assess the strength of the evidence against the null hypothesis. Often the null hypothesis is a statement of "no effect" or "no difference in the true means."

The claim about the population that we are trying to find evidence for is the **alternative hypothesis (H_a).** The alternative is **one-sided** if it states that a parameter is *larger* or *smaller than* the null hypothesis value. It is **two-sided** if it states that the parameter is *different from* the null value (it could be either smaller or larger).

Test Statistic

Many tests of significance are based on a statistic that estimates the parameter that appears in the hypotheses. When H_0 is true, we expect the estimate to take a value near the parameter value specified in H_0.

Values of the estimate far from the parameter value specified by H_0 give evidence against H_0.

A **test statistic** calculated from the sample data measures how far the data diverge from what we would expect if the null hypothesis H_0 were true. So often the TS looks like a Normal variable defined as

$$z = \frac{\text{estimate - hypothesized value}}{\text{standard deviation of the estimate}}$$

Large values of the statistic show that the data are not consistent with the hypothesized value assumed by H_0.These large values will be in the tails of the Normal curve…

P-Value

The null hypothesis H_0 states the claim that we are seeking evidence *against*. The number (probability) that measures the strength of the evidence against a null hypothesis is called a ***P*-value**.

The probability, computed assuming H_0 is true, that the statistic would take a value as extreme as or more extreme than the one we actually observed in our Test Statistic is called the ***P*-value** of the test. The smaller the *P*-value, the stronger the evidence against H_0 provided by the data.

- Small *P*-values are evidence against H_0 because they say that the observed TS result is unlikely to occur when H_0 is true.
- Large *P*-values fail to give convincing evidence against H_0 because they say that the observed TS result is likely to occur when H_0 is true.

Statistical Significance

The final step in performing a significance test is to draw a conclusion about the competing claims you were testing. We will make one of two decisions based on the strength of the evidence against the null hypothesis (and in favor of the alternative hypothesis)—**reject H_0 or fail to reject H_0.**

- If our sample result is too unlikely to have happened by chance assuming H_0 is true, then we'll reject H_0.
- Otherwise, we will fail to reject H_0.

Note: A fail-to-reject H_0 decision in a significance test doesn't mean that H_0 is true. For that reason, you should never "accept H_0" or use language implying that you believe H_0 is true.

So:

P-value small → reject H_0 → conclude H_a (in context!)

P-value large → fail to reject H_0 → cannot conclude H_a (in context!)

Statistical Significance

There is no rule for how small a *P*-value we should require in order to reject H_0. But we can compare the *P*-value with a fixed value that we regard as a decisive, cut-off value, called the **significance level.** We write it as α, the Greek letter alpha. When our *P*-value is less than or equal to the chosen α, we say that the result is **statistically significant.**

If the *P*-value is smaller than or equal to alpha, we say that the data is **statistically significant at level α and we reject the null hypothesis.** The quantity α is called the **significance level** or the **level of significance.**

P-value ≤ α → data is significant → reject H_0 → conclude H_a (in context!)

P-value > α → data is not significant → fail to reject H_0 → cannot conclude H_a (in context!)

Five Steps of Significance Testing

1. If appropriate, define the parameters being tested about in words (context!).

2. State the null and alternative **hypotheses** in terms of these parameters.

3. Calculate the value of the **test statistic, assuming the null hypothesis is <u>true</u>.**

4. Find the ***P*-value** for the observed data – always computed <u>assuming the null hypothesis is true.</u>

5. State a **conclusion in the context of the problem!**

We will learn the details of many tests of significance in the following chapters. The proper test statistic is determined by the hypotheses and the data collection design.