MTL108
Hypothesis Testing

Rahul Singh

Unlike point or interval estimation, which aim to find the explicit value of unknown parameters, hypothesis testing uses sample data to evaluate the validity of a specific claim (or hypothesis) regarding those parameters.

Key Definitions

- **Statistical Hypothesis:** A formal statement or claim about a set of parameters within a population distribution.
- **The Decision Rule:** A statistical procedure evaluates if the observed random sample is mathematically consistent with the hypothesis.
  - If consistent, the hypothesis is informally **"accepted"**.
  - If inconsistent (meaning the data would be highly unlikely to occur if the hypothesis were true), it is **"rejected"**.

The Philosophy of "Acceptance"

A critical conceptual point in testing is that "accepting" a hypothesis does *not* mean we are claiming it is absolutely true. Rather, acceptance simply implies that the observed data does not provide sufficient statistical evidence to contradict it.

Illustrative Examples

- **Industrial Quality Control:** A construction firm tests a sample of $n = 10$ cables to verify a manufacturer's claim that the true average breaking strength is $\mu \geq 7,000$ psi.
- **Normal Population Mean:** Suppose a population follows $N(\theta, 1)$ and we test the hypothesis $\theta < 1$ using a sample of $n = 10$.
  - If the observed sample mean is $\bar{x} = 1.25$, while it does not actively support $\theta < 1$, it is not extreme enough to be deemed completely inconsistent, leading to "acceptance" (i.e., a failure to reject).
  - Conversely, if $\bar{x} = 3$, observing such a large sample mean is so statistically improbable under the assumption that $\theta < 1$ that the hypothesis is deemed strictly inconsistent and is rejected.

Statistical Hypothesis

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with parameter $\theta \in \Theta$, where Θ is the entire parameter space. We partition Θ into two disjoint subsets, Θ_0 and Θ_1, such that $\Theta_0 \cup \Theta_1 = \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$.

- **Null Hypothesis (H_0):** The claim that $\theta \in \Theta_0$. This is the baseline assumption, representing the status quo, "no effect," or "no difference." It is the hypothesis we aim to challenge.
- **Alternative Hypothesis (H_1 or H_a):** The claim that $\theta \in \Theta_1$. This is the research hypothesis, representing the new theory, "effect," or "difference" we are trying to statistically prove.

Definition 1 (Simple and composite hypotheses)**.** A hypothesis is **simple** if it perfectly specifies the underlying distribution (e.g., $H_0 : \mu = 50$). It is **composite** if it does not (e.g., $H_1 : \mu > 50$).

- **Simple Hypothesis:** A hypothesis that, when true, completely specifies the population distribution. Example: $H_0 : \theta = 1$ (for a normal distribution with known variance).
- **Composite Hypothesis:** A hypothesis that does not completely specify the distribution. Example: $H_0 : \theta \leq 1$.

The Decision Matrix, Errors, and Power

When we conduct a test, our mathematical algorithm issues a binary decision: "Reject H_0" or "Fail to Reject H_0". Because this decision is based on a random sample, it is subject to sampling error. There are two distinct ways our algorithm can be mathematically wrong.

Type I and Type II Errors

- **Type I Error (False Positive):** Rejecting H_0 when H_0 is actually true.
- **Type II Error (False Negative):** Failing to reject H_0 when H_1 is actually true.

We quantify the probabilities of these errors using conditional probability statements. Let C denote the **Critical Region** (or Rejection Region), the subset of the sample space where, if our data falls within it, we reject H_0.

- The probability of a Type I error is usually denoted by α, also known as the **Significance Level** or the size of the test, i.e.,

$$\alpha = P(\text{Type I Error}) = P(X \in C \mid H_0 \text{ is true}) = P(X \in C \mid \theta \in \Theta_0).$$

- The probability of a Type II error is denoted by β, i.e.,

$$\beta = P(\text{Type II Error}) = P(X \notin C \mid H_1 \text{ is true}) = P(X \notin C \mid \theta \in \Theta_1)$$

The Power Function

In statistical testing, we want to maximize our ability to detect a true effect. The **Power** of a test is the probability of correctly rejecting a false Null Hypothesis. It is the complement of the Type II error, that is,

$$\text{Power} = 1 - \beta = P(\text{Reject } H_0 \mid H_1 \text{ is true}) = P(X \in C \mid \theta \in \Theta_1).$$

More formally, we define the **Power Function**, $\pi(\theta)$, as the probability of rejecting H_0 for any given value of θ in the entire parameter space:

$$\pi(\theta) = P_\theta(X \in C)$$

Ideally, an optimal test has $\pi(\theta) \leq \alpha$ for all $\theta \in \Theta_0$ (strict Type I error control), and $\pi(\theta)$ as close to 1 as possible for all $\theta \in \Theta_1$.

General Approach to Developing a Test

For testing $H_0 : \theta \in \Omega_0$ at a significance level α:

1. Determine a point estimator for θ, say $d(\mathbf{X})$.
2. The hypothesis is rejected if $d(\mathbf{X})$ falls "far away" from the region Ω_0.
3. To rigorously define how "far away" is required to reject H_0, you must determine the probability distribution of $d(\mathbf{X})$ assuming H_0 is true.
4. Use this distribution to establish the critical region C that meets the required significance level α.

Testing mean of normal population with known variance

Suppose that $X_1, \ldots, X_n$ is a sample of size n from a normal distribution having an unknown mean μ and a known variance σ^2 and suppose we are interested in testing the null hypothesis

$$H_0 : \mu = \mu_0$$

against the alternative hypothesis

$$H_1 : \mu \neq \mu_0$$

where μ_0 is some specified constant.

Since $\bar{X} = \sum_{i=1}^{n} X_i/n$ is a natural point estimator of μ, it seems reasonable to accept H_0 if $\bar{X}$ is not too far from μ_0. That is, the critical region of the test would be of the form

$$C = \left\{X_1, \ldots, X_n : |\bar{X} - \mu_0| > c\right\} \tag{1}$$

for some suitably chosen value c.

If we desire that the test has significance level α, then we must determine the critical value c in Equation (1) that will make the type I error equal to α. That is, c must be such that

$$P_{\mu_0}\{|\bar{X} - \mu_0| > c\} = \alpha \tag{2}$$

where we write P_{μ_0} to mean that the preceding probability is to be computed under the assumption that $\mu = \mu_0$. However, when $\mu = \mu_0$, $\bar{X}$ will be normally distributed with mean μ_0 and variance σ^2/n and so Z, defined by

$$Z \equiv \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$$

will have a standard normal distribution. Now Equation (2) is equivalent to

$$P\left\{|Z| > \frac{c\sqrt{n}}{\sigma}\right\} = \alpha$$

or, equivalently,

$$2P\left\{Z > \frac{c\sqrt{n}}{\sigma}\right\} = \alpha$$

where Z is a standard normal random variable. However, we know that

$$P\{Z > z_{\alpha/2}\} = \alpha/2$$

and so

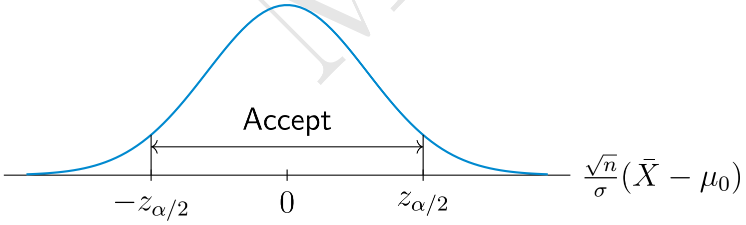
$$\frac{c\sqrt{n}}{\sigma} = z_{\alpha/2}$$

or

$$c = \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$$

Thus, the significance level α test is to reject H_0 if $|\bar{X} - \mu_0| > z_{\alpha/2}\sigma/\sqrt{n}$ and accept otherwise; or, equivalently, to

$$\begin{aligned} &\text{reject} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| > z_{\alpha/2} \\ &\text{accept} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| \leq z_{\alpha/2} \end{aligned} \tag{8.3.3}$$



Acceptance and rejection regions

This can be pictorially represented as shown in the figure above, where we have superimposed the standard normal density function [which is the density of the test statistic $\sqrt{n}(\bar{X} - \mu_0)/\sigma$ when H_0 is true].

Test Statistics

Definition 2. To map high-dimensional sample data into a one-dimensional decision boundary, we use a **Test Statistic**, $T(X)$.

Hypothesis Testing for a Normal Population Mean

Suppose we have a random sample $X_1, X_2, \dots, X_n$ drawn from a normal population $N(\mu, \sigma^2)$. We wish to test a specific claim about the unknown population mean, μ, using a predetermined significance level α (the probability of a Type I error).

Every hypothesis test involves a **Null Hypothesis** (H_0), representing the baseline claim, and an **Alternative Hypothesis** (H_1), representing the research claim we are trying to prove. We evaluate these claims using a **Test Statistic**, whose sampling distribution is known perfectly under the assumption that H_0 is true.

The **Rejection Region** (or Critical Region) is the set of values for the test statistic that are so extreme (unlikely) under H_0 that they provide sufficient mathematical evidence to reject H_0 in favor of H_1.

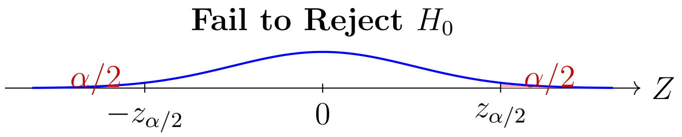
Case 1: Variance (σ^2) is Known

When the true population variance σ^2 is known, the sample mean $\bar{X}$ is normally distributed exactly as $\bar{X} \sim N(\mu, \sigma^2/n)$. Under the null hypothesis $H_0 : \mu = \mu_0$, we standardize $\bar{X}$ to create our **Test Statistic**:

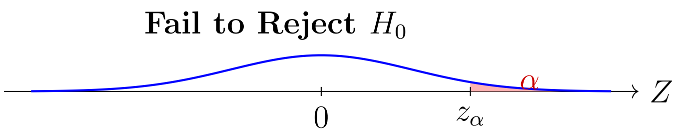
$$Z_{test} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Because we assumed H_0 is true, this statistic follows a Standard Normal distribution: $Z_{test} \sim N(0, 1)$.

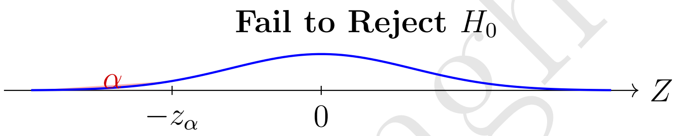
Two-Sided Test
Hypotheses: $H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0$
We reject H_0 if the sample mean is significantly larger *or* significantly smaller than μ_0. We split the significance level α equally into both tails of the standard normal distribution ($\alpha/2$ in each tail).
Rejection Rule: Reject H_0 if $|Z_{test}| > z_{\alpha/2}$.
p-value: $2 \cdot P(Z > |Z_{test}|)$



One-Sided (Right-Tailed) Test
Hypotheses: $H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0$
We reject H_0 only if the sample mean is significantly *larger* than μ_0. The entire significance level α is placed in the right tail.
Rejection Rule: Reject H_0 if $Z_{test} > z_{\alpha}$.
p-value: $P(Z > Z_{test})$



One-Sided (Left-Tailed) Test
Hypotheses: $H_0 : \mu \geq \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0$
We reject H_0 only if the sample mean is significantly *smaller* than μ_0. The entire significance level α is placed in the left tail.
Rejection Rule: Reject H_0 if $Z_{test} < -z_\alpha$.
p-value: $P(Z < Z_{test})$



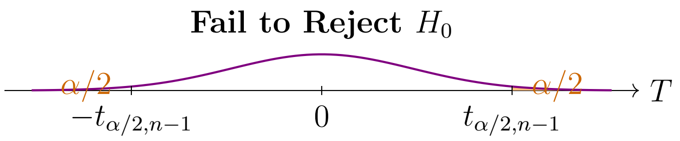
Case 2: Variance (σ^2) is Unknown

In real-world scenarios, the true population variance σ^2 is almost never known. We must estimate it using the sample standard deviation, $S = \sqrt{\frac{1}{n-1} \sum (X_i - \bar{X})^2}$.

Because we are introducing a second random variable (S) into our formula, our test statistic no longer follows a perfect Standard Normal distribution. The extra uncertainty creates heavier tails. Under the null hypothesis $H_0 : \mu = \mu_0$, our **Test Statistic** follows a Student's t-distribution with $n-1$ degrees of freedom:

$$T_{test} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

The testing logic remains exactly identical to the Z-test, but we use critical values derived from the t-distribution table ($t_{\alpha,n-1}$) instead of the Z-table (z_α).
Two-Sided Test
Hypotheses: $H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0$
Rejection Rule: Reject H_0 if $|T_{test}| > t_{\alpha/2,n-1}$.
p-value: $2 \cdot P(T_{n-1} > |T_{test}|)$



One-Sided (Right-Tailed) Test
Hypotheses: $H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0$
Rejection Rule: Reject H_0 if $T_{test} > t_{\alpha,n-1}$.
p-value: $P(T_{n-1} > T_{test})$
One-Sided (Left-Tailed) Test
Hypotheses: $H_0 : \mu \geq \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0$
Rejection Rule: Reject H_0 if $T_{test} < -t_{\alpha,n-1}$.
p-value: $P(T_{n-1} < T_{test})$

Mathematical Derivations of Test Properties

Derivation of the Rejection Region (α)

The critical values that define our rejection regions are derived algebraically from the formal probability definition of a Type I error. By definition, α is the probability of rejecting H_0 when the null hypothesis is perfectly true.

Taking the right-tailed test with known variance ($H_0 : \mu \leq \mu_0$) as an example, we set the probability of a false rejection to exactly α,

$$\alpha = P(\text{Reject } H_0 \mid \mu = \mu_0)$$

Let the rejection rule be in the form $\bar{X} > c$. We must isolate c,

$$\alpha = P(\bar{X} > c \mid \mu = \mu_0)$$

We standardize the random variable $\bar{X}$ using the null parameters,

$$\alpha = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c - \mu_0}{\sigma/\sqrt{n}}\right)$$

Because the left side is now exactly Z, which follows a Standard Normal distribution under H_0, we have

$$\alpha = P\left(Z > \frac{c - \mu_0}{\sigma/\sqrt{n}}\right)$$

By the definition of the normal quantile function, the specific constant that leaves an area of α in the right tail is z_α. Therefore,

$$z_\alpha = \frac{c - \mu_0}{\sigma/\sqrt{n}}$$

Solving for the critical sample mean c,

$$c = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

This demonstrates mathematically why the rejection region is equivalent to checking if our test statistic $Z_{test} > z_\alpha$.

Derivation of Type II Error (β) and Power

While α is controlled by the researcher, the Type II error (β) depends on the *true* underlying mean, μ_1, assuming the alternative hypothesis is true.

Using the same right-tailed test, the Type II error occurs when we fail to reject H_0 (meaning $\bar{X} \leq c$) even though the true mean is $\mu_1 > \mu_0$.

$$\beta = P(\text{Fail to Reject } H_0 \mid \mu = \mu_1)$$

Substituting our derived critical value c,

$$\beta = P\left(\bar{X} \leq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \;\middle|\; \mu = \mu_1\right)$$

To calculate this probability, we must standardize $\bar{X}$ around the *true* mean, μ_1,

$$\beta = P\left(\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} \leq \frac{\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} - \mu_1}{\sigma/\sqrt{n}}\right)$$

$$\beta = P\left(Z \leq z_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right)$$

The **Power** of the statistical test at the specific alternative μ_1 is the complement of β:

$$\text{Power}(\mu_1) = 1 - \beta = P\left(Z > z_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right)$$

This function mathematically proves that the statistical power of our test increases as the true mean (μ_1) shifts further away from the hypothesized mean (μ_0), or as we collect a larger sample size (n).

Approaches

In modern statistics, we use a hybrid of two historically distinct frameworks. It is mathematically and philosophically vital to distinguish between Neyman-Pearson Hypothesis Testing and R.A. Fisher's Pure Significance Testing.

The Neyman-Pearson Lemma

How do we know we are using the "best" test? For testing a simple null hypothesis ($H_0 : \theta = \theta_0$) against a simple alternative ($H_1 : \theta = \theta_1$), the Neyman-Pearson Lemma provides the mathematical blueprint for constructing the **Most Powerful (MP) Test**—the test that maximizes Power $(1-\beta)$ for a fixed significance level α.

Theorem 1 (Neyman-Pearson Lemma, without proof)**.** *A test with rejection region R is the Most Powerful test of size α if there exists a constant $k > 0$ such that*

$$\frac{L(\theta_0 \mid x)}{L(\theta_1 \mid x)} \leq k \quad \textit{for } x \in C$$

$$\frac{L(\theta_0 \mid x)}{L(\theta_1 \mid x)} > k \quad \textit{for } x \notin C$$

Where $L(\theta \mid x)$ is the likelihood function. In plain terms, we reject H_0 if the sample data is significantly more likely to have been generated by θ_1 than by θ_0.

Testing proportions

Many real-world experiments result in binary outcomes: pass/fail, defective/non-defective, or heads/tails. Let $X_1, X_2, \ldots, X_n$ be a random sample of independent Bernoulli trials, where $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$. The parameter p represents the true, unknown population proportion.

We wish to test a baseline claim about this proportion against a specific alternative. To apply the Neyman-Pearson (NP) Lemma, we initially restrict ourselves to a **simple versus simple** hypothesis test. Assume we suspect the true proportion is higher than the baseline. We want to test

$$H_0 : p = p_0 \quad \text{(Baseline / Null) against}$$

$$H_1 : p = p_1 \quad \text{(Alternative, where } p_1 > p_0\text{)}$$

The Neyman-Pearson Test

To find the Most Powerful (MP) test of size α, the Neyman-Pearson Lemma instructs us to evaluate the Likelihood Ratio. We reject H_0 if the data is sufficiently more likely under H_1 than under H_0.

Step 1: Construct the Likelihood Function

The joint probability mass function (PMF) for n independent Bernoulli trials is

$$L(p \mid \mathbf{x}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

Step 2: Form the NP test

The NP Lemma dictates that the rejection region C is given by

$$\frac{L(p_1 \mid \mathbf{x})}{L(p_0 \mid \mathbf{x})} \geq k$$

Substituting our likelihood functions:

$$\frac{p_1^{\sum x_i}(1-p_1)^{n-\sum x_i}}{p_0^{\sum x_i}(1-p_0)^{n-\sum x_i}} \geq k$$

Group the terms raised to the power of $\sum x_i$ and the remaining constants

$$\left(\frac{p_1}{p_0}\right)^{\sum x_i}\left(\frac{1-p_1}{1-p_0}\right)^{n-\sum x_i} \geq k \implies \left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)^{\sum x_i}\left(\frac{1-p_1}{1-p_0}\right)^{n} \geq k.$$

Take the natural logarithm of both sides to bring down the exponents

$$\left(\sum_{i=1}^{n} x_i\right)\ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right) + n\ln\left(\frac{1-p_1}{1-p_0}\right) \geq \ln(k)$$

We want to isolate our sample data, $\sum x_i$. Let $Y = \sum x_i$, which represents the total number of “successes” in our sample. Then, the NP test is

$$Y \cdot \ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right) \geq \ln(k) - n\ln\left(\frac{1-p_1}{1-p_0}\right).$$

Gathering all constants on the right side into a single new constant c, we get NP test as

$$Y = \sum_{i=1}^{n} X_i \geq c.$$

Conclusion of the Lemma:

The NP Lemma mathematically proves that the Most Powerful test for a proportion simply involves counting the total number of successes (Y). We reject H_0 if the total number of successes meets or exceeds some critical value c. The critical value c is decided based on the level of significance α.

The Exact Binomial Test

To complete the test, we must find the exact value of c such that our Type I error probability is bounded by α. Because $Y = \sum X_i$ is the sum of n independent Bernoulli trials, Y follows a Binomial distribution: $Y \sim \text{Binomial}(n, p_0)$ under H_0.

We need to find an integer c such that

$$P(Y \geq c \mid p = p_0) \leq \alpha$$

$$\sum_{y=c}^{n} \binom{n}{y} p_0^y (1-p_0)^{n-y} \leq \alpha$$

Remark on Discreteness: Because the Binomial distribution is discrete, we usually cannot achieve a Type I error of *exactly* α (e.g., exactly 0.05). In practice, statisticians choose the smallest integer c that keeps the probability strictly less than or equal to α to remain conservative.

Extension to the Normal Approximation (via CLT)

Calculating exact Binomial probabilities using factorials is computationally expensive and tedious for large sample sizes (e.g., $n = 1000$). Fortunately, the **Central Limit Theorem (CLT)** provides an elegant, highly accurate asymptotic approximation.

Step 1: Applying the Central Limit Theorem

Let $\hat{p} = \frac{Y}{n} = \frac{1}{n}\sum X_i$ be the sample proportion. The expected value is $E[\hat{p}] = p$, and the variance is $\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$. By the Central Limit Theorem, for large n, the sampling distribution of $\hat{p}$ is approximately a Normal distribution, precisely,

$$\hat{p} \overset{approx}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$$

Step 2: Constructing the Test Statistic

Under the Null Hypothesis ($H_0 : p = p_0$), we standardize $\hat{p}$ to create a Z-statistic,

$$Z = \frac{\hat{p} - E[\hat{p}]}{\sqrt{\text{Var}(\hat{p})}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

Because we assume H_0 is true, this test statistic asymptotically follows a Standard Normal distribution: $Z \sim N(0, 1)$.

Step 3: The Asymptotic Decision Rule

Recall from our NP Lemma derivation that the optimal rejection region is $Y \geq c$, which algebraically translates to $\hat{p} \geq c^*$. In the standardized Normal space, this directly corresponds to rejecting H_0 when our Z-statistic falls into the upper tail of the Standard Normal distribution.

For a fixed significance level α:

- **Right-Tailed Test ($H_1 : p > p_0$):** Reject H_0 if $Z > z_\alpha$.
- **Left-Tailed Test ($H_1 : p < p_0$):** Reject H_0 if $Z < -z_\alpha$.
- **Two-Sided Test ($H_1 : p \neq p_0$):** Reject H_0 if $|Z| > z_{\alpha/2}$.

The Fisher Approach (Pure p-value Testing)

In Fisher's pure significance testing, there is no formal Alternative Hypothesis (H_1), no Type II Error (β), no Power, and no strict pre-determined α cutoff. Instead, the p-value is treated as a **continuous measure of evidence** against the Null Hypothesis. The smaller the p-value, the stronger the statistical evidence that the baseline assumption (H_0) is incompatible with the observed data. Rather than blindly issuing a "Reject" or "Accept" command, a statistician using the pure p-value approach reports the exact probability and interprets the weight of the evidence.

Why this matters in practice? Imagine two different clinical trials testing a new drug. Trial A yields a p-value of 0.049. Trial B yields a p-value of 0.0001. Under a strict Neyman-Pearson framework with $\alpha = 0.05$, both trials result in the exact same binary conclusion: "Reject H_0." However, under a pure p-value framework, we recognize that Trial B provides vastly superior, nearly incontrovertible mathematical evidence against the null compared to Trial A, which barely crossed the arbitrary threshold. Pure testing demands that we report the exact p-value to convey the exact magnitude of the evidence.

Remark 1. The p-value is possibly the most misinterpreted metric in applied statistics. It is *not* the probability that H_0 is true.

Definition 3. The p-value is the exact conditional probability of observing a test statistic as extreme as, or more extreme than, the one actually calculated from the sample, assuming the Null Hypothesis is true.

The Decision Rule:

- If p-value $\leq \alpha \implies$ Reject H_0. (The data is statistically incompatible with the baseline assumption).
- If p-value $> \alpha \implies$ Fail to Reject H_0.

Example 1 (Access to clean fuel, NFHS-data)**.** Suppose a policy aims to achieve a proportion of access to clean fuel of $p_0 = 0.60$. A sample of $n = 1000$ households reveals exactly 560 have access ($\hat{p} = 0.56$).

- $H_0 : p = 0.60$ (The state meets the federal target).
- $H_1 : p < 0.60$ (The state is falling statistically short of the target).

Using the Normal approximation for sample proportions:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.56 - 0.60}{\sqrt{\frac{0.60(0.40)}{1000}}} = \frac{-0.04}{0.0155} \approx -2.58$$

Probability Statement:

$$p\text{-value} = P(Z \leq -2.58 \mid p = 0.60) \approx 0.0049$$

Because the p-value is 0.0049 (less than the standard $\alpha = 0.05$), we reject H_0.

Remark 2 (Optional)**.** Read more about p-values at `https://journals.lww.com/ijom/fulltext/2023/04260/the__p__value__the_primary_alphabet_of_research.41.aspx`.

References

[1] Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability.* Chapman and Hall/CRC.

[2] Ross, Sheldon M. (2020). *Introduction to probability and statistics for engineers and scientists.* Academic press.

Disclaimer

This lecture note is prepared solely for teaching and academic purposes. Some parts of the material, including definitions, examples, and explanations, have been adapted or reproduced from the references. These notes are not intended for commercial distribution or publication, and all rights remain with the respective copyright holders.