

Introduction to Statistical Method

Hypothesis Testing

A second major statistical method for gaining information on a probability.

The goal is to reject or fail to reject statements (hypotheses) based on statistical data.

Hypothesis Definition

A statement about a population parameter θ . The hypothesis will compare θ to a null value denoted θ_0 .

Fisher's Null Hypothesis Test

This hypothesis will be denoted by H_0 and is null hypothesis.

Three forms: $H_0 : \theta = \theta_0$ or $H_0 : \theta \leq \theta_0$ or $H_0 : \theta \geq \theta_0$.

A hypothesis test is based on rejecting a hypothesis.

One-Tailed Test

The test of a hypothesis of the form $H_0 : \theta \leq \theta_0$ or $H_0 : \theta \geq \theta_0$ is said to be one-tailed tests.

P-Value for a One-Tailed Test

- Apply an example first

We want to find evidence that a new car design has a mean mileage greater than 26 mpg. Therefore, we set up the null hypothesis: $H_0 : \mu \leq 26$.

The goal is to reject the null hypothesis.

- Example explanation

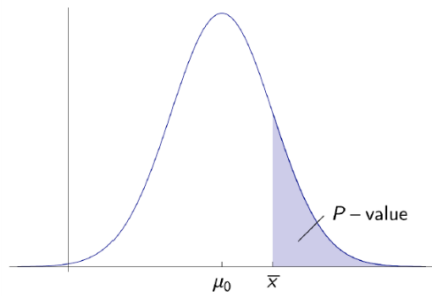
We take a random sample and calculate \bar{X} , if it is much greater than 26, then there is reason to believe that H_0 is false.

Take a random sample of size n and find the value \bar{x} for the sample mean.

The probability of obtaining the measured value of \bar{x} or a larger result if H_0 is true is the **significance** or **P-Value** of the test.

(TM还是说中文吧：就是说，猜测 $\mu \leq 26$ ，然后我们得到了样本的平均的观测值 \bar{x} ，根据样本个数和标准差，我们可以对样本的平均值 \bar{X} 得到基于 $\mu \leq 26$ 的 test，然后可以放宽到 $\mu = 26$ ，然后可以用 $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ 算出 P 的值来判断这个样本是否符合这个假设)

$$P[\bar{X} \geq \bar{x} | \mu \leq 26] \leq P[\bar{X} \geq \bar{x} | \mu = 26]$$



$H_0: \mu \leq \mu_0$ shows the case if $\mu = \mu_0$, the curve shift left if $\mu < \mu_0$

The shaded area shows the probability of obtaining $\bar{X} \geq \bar{x}$ if $\mu = \mu_0$.

- The P-value is therefore an upper bound of the probability of obtaining the data if H_0 is true.
- $P = P[D|H_0]$ if D represents the statistical data, we will reject H_0 if it is small.

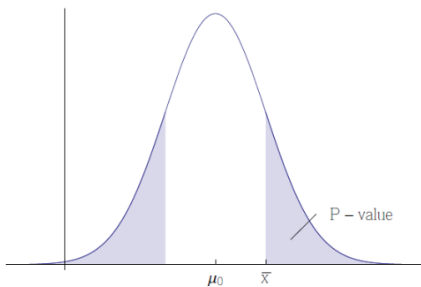
So either:

- fail to reject H_0 at P level of significance
- reject the H_0 at P level of significance

The statistic on which the P is based is **test statistic**.

Two-Tailed Test

If we are testing a hypothesis of the form $H_0: \theta = \theta_0$, we say we are performing a two-tailed test.



$H_0: \mu = \mu_0$ The P is twice the value of one-tailed test.

Does a Small P-Value Provide Evidence that H_0 is False

Since we know the fact that the $P = P[D|H_0]$, but some researcher want $P[H_0|D]$.

We can derive the fact from the Bayes's theorem:

$P[D|H_0] = P[D \cap H_0]/P[H_0]$, then we can derive $P[H_0|D] = P[D \cap H_0]/P[D]$.

$$\begin{aligned} P[H_0|D] &= \frac{P[D \cap H_0]}{P[D]} = \frac{P[D|H_0] \cdot P[H_0]}{P[D]} = \frac{P[D|H_0] \cdot P[H_0]}{P[D|H_0] \cdot P[H_0] + P[D|\neg H_0] \cdot (1 - P[H_0])} \\ &= \frac{P[D|H_0]}{P[D|H_0] + P[D|\neg H_0] \cdot \left(\frac{1-P[H_0]}{P[H_0]}\right)} \end{aligned}$$

Is Hypothesis Testing Logical?

Since we get the $P[H_0|D]$ representation, we can let it be close to 1 depending on $P[H_0]$.

Hence, it is possible that: given H_0 and the data is very unlikely, but given the data H_0 is very likely.

- In the classic argument:
If P then Q ; not Q therefore not P
- In hypothesis testing, we want to argue that
If P then Q ; Q is unlikely therefore P is unlikely
Actually this is wrong.

Bayesian & Frequentist Statistics

Bayesian

Claim to understand the **logical inconsistencies** and intend to compensate for them with **prior and posterior probability** distributions.

Theoretically true, difficult to implement in practice.

Frequentist

Mainly ignore the problems mentioned here or claim that they are not relevant in their specific research.

Neyman-Pearson Decision Theory

Two competing hypothesis: H_0, H_1 .

Seek to reject H_0 to accept H_1 .

- H_0 is **null hypothesis**
- H_1 is **research hypothesis** or **alternative hypothesis**.

So there are four possible outcomes of the decision-making process:

- We reject H_0 when H_0 is untrue.
- **Type I Error**: We reject H_0 even though H_0 is true.
- **Type II Error**: We fail to reject H_0 even though H_0 is untrue.
- We fail to reject H_0 when H_0 true.

Type I and Type II error should be as small as possible.

Power, Type I & II Error Probabilities

$$\alpha = P[\text{Type I Error}] = P[\text{reject } H_0 \mid H_0 \text{ true}] = P[\text{accept } H_1 \mid H_1 \text{ false}]$$

$$\beta = P[\text{Type II Error}] = P[\text{fail to reject } H_0 \mid H_0 \text{ false}]$$

$$\text{Power} = 1 - \beta$$

The power shows how likely our experiment is successful.

- By requiring strong evidence before rejecting H_0 (a value of the test statistic that is very different from its null value), α can be made small.
- The range of values for the test statistic that causes us to reject H_0 is **critical region**.

We choose the critical region in such a way to make α small.

- The more evidence we require to reject H_0 (the smaller the critical region is), the harder it is to actually reject H_0 in the first place.

The power decreases with β becomes larger.

- For given H_0 and H_1 , β can be controlled by increasing the sample size.

Example of Neyman-Pearson Decision Theory

The mean is supposed to be $\mu_0 = 40$ and the standard deviation is $\sigma = 2$.

Two hypothesis $H_0 : \mu = 40, H_1 : |\mu - 40| \geq 1$

Then the sample size is $n = 25$.

The probability of committing Type I Error is $\alpha \leq 5\%$. (derive from H_0 and H_1 ?)

So we apply the test statistic $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ and let $-z_{\alpha/2} \leq Z \leq z_{\alpha/2}$ to get probability of $1 - \alpha$ and $P[|Z| > z_{\alpha/2}] = \alpha$.

The $\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq 1.96$ is the critical region, with this critical region there are 5% of Type I Error.

If the sample mean is $\bar{x} = 40.9$, then the test statistic is $z = 2.25 > 1.96$.

Type II Error

$|Z| = \left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq z_{\alpha/2} = 2.575$ when $\alpha = 1\%$. So with $z = 2.25 < 2.575$, there is a Type II Error.

Assume null hypothesis $H_0 : \mu = \mu_0$ and the true value is $\mu = \mu_0 + \delta, \delta \in \mathbb{R} \setminus \{0\}$.

$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ is the test statistic actually follows a normal distribution with unit variance and mean $\delta\sqrt{n}/\sigma$.

Review that if $-z_{\alpha/2} \leq Z \leq z_{\alpha/2}$ then we cannot reject H_0 , deriving the Type II Error.

$$\beta = P[|Z| \leq z_{\alpha/2}]$$

Remark

In order for the statistical procedure to be valid, **a critical region must be fixed before any data obtained.**