5 Hypothesis Testing

ENSY SILVER[1]

Monday 17th August, 2020

[1]Thanks to my family, my friend and freedom.

1 Introduction

We have already introduced the estimation of parameters, that is, suppose we have known the model, we estimate parameters from the experimental data.

However, there exist situations that we need to estimate hypothesis, which is the content we will introduce here. A court psychiatrist, for example, may be called upon to pronounce an accused murderer either "sane" or "insane".

The process of dichotomizing the possible conclusions of an experiment and then using the theory of probability to choose one option over the other is known as **hypothesis testing**. The two competing propositions are called the **null hypothesis** (written H_0) and the **alternative hypothesis** (written H_1).

2 The decision rule

Now we give an example, if we have a batch of data with average μ_0, which derived from the same distribution, and we estimate whether an additional condition will change the distribution, particularly, the mean. We assume that we have known the experimental data y with the additional condition, and we assume the the distribution with the additional condition has mean μ.

Applying the central limit theorem, we get that $\frac{\overline{y}-\mu_0}{\sigma/\sqrt{n}}$ is a random variable with standard distribution. Moreover, n is the number of sample outcome y, σ^2 is the variance of y. So we have the theorem.

Theorem 2.1. *If the null hypothesis $H_0 : \mu = \mu_0$ is rejected using a 0.05 decision rule, the difference between $\overline{y}$ and μ_0 is said to be statistically significant.*

2.1 Expressing decision rules in terms of Z ratios

Theorem 2.2. *Any function of the observed data whose numerical value dictates whether H_0 is accepted or rejected is called a **test statistic**. The set of values for the test statistic that result in the null hypothesis being rejected is called the critical region and is denoted C. The particular point in C that separates the rejection region from the acceptance region is called the **critical value**.*

Theorem 2.3. *The probability that the test statistic lies in the critical region when H_0 is true is called the level of significance and is denoted α.*

2.2 Testing $H_0 : \mu = \mu_0(\sigma$ known)

Theorem 2.4. *Let $y_1, y_2, \cdots, y_n$ be a random sample of size n from a normal distribution where σ is known. Let $z = \frac{\overline{y}-\mu_0}{\sigma/\sqrt{n}}$.*

1. *To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ at the α level of significance, reject H_0 if $z \geq z_\alpha$.*

2. *To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$ at the α level of significance, reject H_0 if $z \leq z_\alpha$.*

3. *To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ at the α level of significance, reject H_0 if $z \geq z_{\alpha/2}$ or $z \geq z_{\alpha/2}$.*

2.3 The P-value

There are two general ways to quantify the amount of evidence against H_0 that is contained in a given set of data. The first involves the *level of significance* concept introduced in theorem 2.3. Using that format, the experimenter selects a value for α (usually 0.05 or 0.01) *before any data are collected.* Once α is specified, a corresponding critical region can be identified. If the test statistic falls in the critical region, we reject H_0 at the α level of significance. Another strategy is to calculate a P-value.

Theorem 2.5. *The P-value associated with an observed test statistic is the probability of getting a value for that test statistic as extreme as or more extreme than what was actually observed (relative to H_1) given that H_0 is true.*

Remark. *The difference between the two methods is that when we use level of significance, we determine a range in Z's distribution and state that H_0 is false if the data locate in the range; conversely, the method of P-value calculates the probability of the data locate as extreme as or more extreme than the observed data.*

2.4 Testing binomial data

We can apply theorem 2.3 to test hypothesis, but we need to ensure that n is big enough to meet 3σ rule. 3σ requires that the random variable $z = \frac{\overline{y}-\mu_0}{\sigma/\sqrt{n}}$ varies in the range $[-R, R]$, where $R > 3$.

$$\begin{aligned}
\frac{\overline{y}-\mu_0}{\sigma/\sqrt{n}} &> 3 \\
\overline{y}-p_0 &> \frac{3\sigma}{\sqrt{n}} \\
\overline{y}-p_0 &> 3\sqrt{p_0(1-p_0)/n} \\
n\overline{y} &> np_0 + 3\sqrt{np_0(1-p_0)} \\
n &> np_0 + 3\sqrt{np_0(1-p_0)}
\end{aligned}$$

$$\begin{aligned}
\frac{\overline{y}-\mu_0}{\sigma/\sqrt{n}} &< -3 \\
\overline{y}-p_0 &< -\frac{3\sigma}{\sqrt{n}} \\
\overline{y}-p_0 &< -3\sqrt{p_0(1-p_0)/n} \\
n\overline{y} &< np_0-3\sqrt{np_0(1-p_0)} \\
0 &< np_0+3\sqrt{np_0(1-p_0)}
\end{aligned}$$

So, if $n > np_0 + 3\sqrt{np_0(1-p_0)}$ and $0 < np_0 + 3\sqrt{np_0(1-p_0)}$, the theorem 2.4 can be applied to test hypothesis.

3 Type I and type II erroes

First, we give the definitions about type I and type II errors.

$$P(\text{Type I error}) = P(\text{Reject } H_0 | H_0 \text{ is true})$$

$$P(\text{Type II error}) = P(\text{Fail to reject } H_0 | H_1 \text{ is true})$$

The problem is, we have not known the distribution with additional conditions, so we only make hypothesis testing from the current distribution. Apparently, it is possible that H_1 is true and we do not reject H_0. The P(Type II error) depends on the new distribution.

We give an example, suppose we have a normal distribution with $\mu = 25$, we test if a normal distribution has $\mu > 25$ with additional conditions. Now, we have samples from the new distribution, and test

$$\begin{aligned}
H_0 : &\quad \mu = 25 \\
H_1 : &\quad \mu > 25
\end{aligned}$$

at the α level of significance. So

$$P(\text{Type II error}) = P(z < z_\alpha | z \text{ is a sample from distribution with } \mu > 25)$$

We find that as μ gets larger, the P(Type II error) gets smaller. We define the function with μ as a variable, and $1-P$(Type II error) as a dependent variable. $1 - P$(Type II error) is called a **power**, and the function is called a **power curve**, whose value is an integral of a normal distribution. For convenience, we denote the power by $1-\beta$.

3.1 Factors that influence the power of a test

If other factors remain stable, *increasing α decreases β and increases the power.*

If other factors remain stable, *increasing* σ *decreases* β *and increases the power.*

If other factors remain stable, *increasing* n *decreases* β *and increases the power.*

3.2 Nonnormal data

When the data size is too samll, z may not converges to a normal distribution. In this case, we can test hypothesis directly.

4 Generalized likelihood ratio

In the next several chapters we will be studying some of the particular hypothesis tests that statisticians most often use in dealing with real-world problems. All of these have the same conceptual heritage a fundamental notion known as the **generalized likelihood ratio**, or GLR.

As a starting point in answering those questions, it will be necessary to define two parameter spaces, ω and Ω. In general, ω is the set of unknown parameter values admissible under H_0. The second parameter space, Ω, is the set of all possible values of all unknown parameters.

Let $y_1, y_2, \cdots, y_n$ be a random sample from $f_Y(y;\theta)$. The generalized likelihood ratio, λ, is defined to be

$$\lambda = \frac{\max_\omega L(\theta)}{\max_\Omega L(\theta)} = \frac{L(\omega_e)}{L(\Omega_e)}$$

A generalized likelihood ratio test (GLRT) is one that rejects H_0 whenever

$$0 < \lambda \leq \lambda^*$$

where λ^* is chosen so that

$$P(0 < \Lambda \leq \lambda^* | H_0 \text{ is true}) = \alpha$$