Chapter 6

Hypothesis testing

6.1 Introduction to testing

A statistical hypothesis test is a method of statistical inference using data from a scientific study. In statistics, a result is called statistically significant if it has been predicted as unlikely to have occurred by chance alone, according to a pre-determined threshold probability, the significance level. The phrase "test of significance" was coined by statistician Ronald Fisher ([Fisher(1925)]). These tests are used in determining what outcomes of a study would lead to a rejection of the null hypothesis for a pre-specified level of significance; this can help to decide whether results contain enough information to cast doubt on conventional wisdom, given that conventional wisdom has been used to establish the null hypothesis.

6.2 Key elements in testing

Types of Hypotheses

- Null hypothesis (H_0), status quo
- Alternative hypothesis (H_1 or H_A)– what we want to demonstrate.

Example 6.2.1 (Clinical trial)**.** *Let* $\tau = P(\textit{Success for surgical procedure})$.

$$H_0 : \tau \leq 0.2, vs \quad H_1 : \tau > 0.2.$$

Both H_0 and H_1 are composite in the sense of each comprising more than one distribution.

Example 6.2.2 (Incidence of West Nile Vrius). *In 2002 there were several conformed cases of West Nile Virus in the New York metropolitan area. Let* τ *= the number of West Nile Virus in New York Metropolitan area per million residents*

$$H_0 : \tau = 55, vs \quad H_a : \tau \neq 55.$$

Here H_0 is simple; H_1 is composite.
Test statistic: The test statistic is a function of the data;

Rejection region: If the statistic falls within a specified range of values, the researcher rejects the null hypothesis. The range of values that leads the researcher to reject the null hypothesis is called the region of rejection. The rejection region is denoted as $\mathcal{R}$.

Types of Errors: An hypothesis test is a data driven rule involving a test statistic $T(\boldsymbol{Y})$ where $\boldsymbol{Y}$ is our data vector. We reject H_0 when $T(\boldsymbol{Y}) \in \mathcal{R}$ and fail-to-reject H_0 if $T(\boldsymbol{Y}) \notin \mathcal{R}$. There are two possible errors:

- Type I error: α, or significance level of size
  – Simple: $\alpha = P(\text{Reject } H_0 | H_0 \text{ true})$;
  – Composite: $\alpha = \max_{P_0 \in H_0}\{P(\text{Reject } P_0 | P_0 \text{ true})\}$;
- Type II error: β,
  – Simple: $\beta = P(\text{Do not reject } H_0 | H_0 \text{ not true})$;
- Power: $1 - \beta = P(\text{Reject } H_0$ — Alternative values of the parameter).

Example 6.2.3. *In a clinical trial, there are* $n = 25$ *subject. Let* Y *be the number of patients with successful procedure and assume that* $Y \sim Bin(n, \tau)$. *Consider a set of simple hypotheses:*

$$H_0 : \tau = 0.2, vs \quad H_a : \tau = 0.5.$$

Set the rejection rule as $\mathcal{R} = \{Y : Y > 8\}$. What is the type I and type II errors?

P-value: the p-value is the probability of obtaining a test statistic result at least as extreme or as close to the one that was actually observed, assuming that the null hypothesis is true.

In the above example, what is the p-value if we observe $Y = 10$?

Steps in constructing a hypothesis:

1. Set up the hypotheses.
2. Determine the desired Type I error rate.
3. Determine a test statistic
4. Find the distribution of the test statistic under the null hypothesis.
5. find the rejection region such that the Type I error rate is satisfied.
6. Possible determine power under the alternative.
7. If data are available, evaluate the test and make a decision. Determine a p-value.

6.3 Generalized Likelihood Ratio Tests

In this experiment, it is easy to find a test statistic. What if the test statistic is not intuitive obvious? Or if we have several candidates, how do we choose the best test statistic and the rejection region?

We want to test the composite hypothesis where

$$H_0 : \theta \in \Theta_0, vs \quad H_a : \theta \in \Theta_1.$$

where Θ_0 and Θ_1 denote subsets of the parameter space. The GLRT reject H_0, when

$$LR = \frac{\max_{\theta\in\Theta_1} L(\theta)}{\max_{\theta\in\Theta_0} L(\theta)} > k.$$

Or equivalently,

$$\Lambda = \frac{\max_{\theta\in\Theta_0\cup\Theta_1} L(\theta)}{\max_{\theta\in\Theta_0} L(\theta)} > k^*.$$

The next problem is that we need to find the distribution of Λ to setup the rejection region and calculate the p-values. Sometimes we can find an exact distribution of Λ; sometimes, we can only work with the asymptotic approximation.

Example 6.3.1 (Test for the mean with known variance)**.** *Suppose that hourly wages in the petroleum industry in Texas are normally distributed with a mean of \$17.60 and a standard deviation of \$1.30. A large company in this industry randomly sampled 50 of its workers, determining that their hourly wage was \$17.30. Stating your assumptions, can we conclude that this company's average hourly wage is different from that of the entire industry?*

Some results:

1. Simple null, $H_0 : \tau = \tau_0 vs \quad H_a : \tau \neq \tau_0$, then

$$2 \log \Lambda | H_0 \sim \chi^2_1$$

2. Nested null and alternative: $H_0 : \theta \in \Theta_{p-r} vs \quad H_a : \theta \in \Theta_p$, where Θ_{p-r} is a subset of Θ_p, then

$$2 \log \Lambda | H_0 \sim \chi^2_r.$$

Example 6.3.2 (Seasonal Changes in Teen Suicide)**.** *In 2001, teen suicide was the 3rd leading cause of death among young adults and adolescents 15 to 24 years of age, following unintentional injuries and homicide. The research question is whether there exists seasonal variation in suicide rate.*

The data is available at
http://www.familyfirstaid.org/parenting/emotional/teen-suicide/

The null hypothesis is that the suicide rate is constant across months. The alternative is that there is seasonal variation in suicide rate.

Model: Let Y_i is the number of suicides in the i-th month ($i = 1, 2, \cdots, 12$). Let θ_i be the daily suicide rate in the i-th month and d_i be the number of days in the i-th month. Assume that $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n) \sim Multinomial(n, \boldsymbol{\pi})$ where $\boldsymbol{\pi} = (d_1\theta_1, \cdots, d_{12}\theta_{12})$.

Let $\boldsymbol{Y}^T = (Y_1, \cdots, Y_n)$ be a vector denoting the number of times that an independent observation falls in to the i-th category, $i = 1, 2, \cdots, n$ in a series of n trials, where the probability of falling into the i-th category is π_i, $\sum_i \pi_i = 1$. Then

$$P(Y_1 = y_1, \cdots, Y_n = y_n) = \frac{n!}{y_1! \cdots y_n!} \prod_{i=1}^{n} \pi_i^{y_i}.$$

Note that in this model, the constraint on $\boldsymbol{\pi}$ is $\sum \pi_i = 1$ and the constraint on $\boldsymbol{Y}$ is $\sum y_i = n$. Note that this distribution is an extension of the binomial, and the MLE for each π_i is $\hat{\pi}_i = \frac{Y_i}{n}$.

Hypothesis:

Test statistic:

Null distribution:

Rejection region:

P-value:

Conclusion: