Packet 6: Hypothesis Testing

Chap 8.3 Test about Proportions

Given a hypothesis testing problem, we need to design a test such that α and β are balanced. We'll start our exploration of hypothesis tests by focusing on population proportions.

Example 1: Let p be the proportion of failures of the production line at a manufacturing company, and it was 0.06. A new procedure was implemented, and we wanted to know whether it could reduce the failure rate.

Every time we perform a hypothesis test, this is the basic procedure that we will follow:

1. We'll make an initial assumption about the population parameter, H0;
2. We'll collect evidence in the form of data;
3. Based on the available evidence (data), calculate a test statistic;
4. Decide whether to "reject" or "not reject" our initial assumption.

Let's try to make this outlined procedure more concrete with the above example.

The new procedure was tested on 200 products, the number of failure is $Y \sim$ Binomial (200, p). To determine whether the new procedure reduces the failure rate, we set

$$H_0 : p = 0.06 \text{ v.s. } H_1 : p < 0.06$$

Decision rule: reject H_0 if we observe the number of failures, $y \leq 7$.

Type 1 error:

$\alpha = P(Y \leq 7; p = 0.06) = \sum_{k=0}^{7} P(Y = k; p = 0.06) = \sum_{k=0}^{7} \frac{200!}{k!(200-k)!} 0.06^k 0.94^{200-k} = 0.09.$

Suppose that the true failure rate under new procedure is 0.03.

Type 2 error:

$\beta = P(Y > 7; p = 0.03) = 1 - \sum_{k=0}^{7} P(Y = k; p = 0.03) = 1 - \sum_{k=0}^{7} \frac{200!}{k!(200-k)!} 0.03^k 0.97^{200-k} = 0.256.$

As the cutoff decreases,

Significance level, Test statistic, rejection region, and critical value:

Type I error is often more serious (like the court example), we design a decision rule such that α is controlled (e.g., $\alpha = 0.05$). The predetermined α is called **the significance level of the test**. On the other hand, β is hard to obtain unless knowing true value of θ.

Let $X_1, \ldots, X_n$ be a random sample from a distribution with parameter θ. Let $T = h(X_1, \ldots, X_n)$ be a statistic and let R be a subset of the real line. Suppose we choose to "*reject* H_0 *if* $T \in R$". Then T is called a **test statistic** and R is called the **critical region** or **rejection region** of the test.

Classically, a test statistic T is based on the MLE or a function of the MLE of the parameter, e.g., $\bar{X}$ for testing some value of μ in a Normal case, or $\hat{p}$ for testing values of p in a Bernoulli or Binomial case.

Note that it is not convenient to check Type I Error under all possible cutoff values. So we consider the normal approximation, which is appropriate when sample size is large.

A rejection region is usually in the form of $T \geq c$, $T \leq c$, or $|T| \geq c$. The direction of the inequality is set by the direction of H_1. The constant c is called the **critical value**, which is determined by the significance level $\alpha = P(T \in \text{Reject Region} \mid H_0 \text{ is true})$.

Critical Value Approach:

Find a threshold value for the test statistic (critical value) so that if our observed test statistic is more extreme than the critical value, then we reject the null hypothesis.

We choose the test statistic that has a known distribution and define critical value such that type I error is controlled.

Example 1: Let us standardize the MLE, $\hat{p}$, like what we did for C.I. derivations.

If we want to control $\alpha = 0.05$, we need to find the critical value z^* such that the probability of getting more extreme value of Z is 0.05.

What if we test at a new significance level 0.1? $H_0 : p = 0.06$ v.s. $H_1 : p < 0.06$

What if we test against a new alternative $H_1 : p > 0.06$?

What if we test against a new alternative $H_1 : p \neq 0.06$?

Let's summarize the procedure in terms of performing a hypothesis test for a population proportion using the critical value approach:

1. State the null hypothesis H_0 and the alternative hypothesis H_1.
2. Calculate the test statistic: $Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$.
3. Determine the critical region.

4. Make a decision. Determine if the test statistic falls in the critical region. If it does, reject the null hypothesis. If it does not, do not reject the null hypothesis.

P-Value Approach:

Instead of specify a significance level and determine a rejection region, we may report the significance using probability values (p-**value**).

Definition: p-**value** is the probability of observing the test statistic calculated from data, or a more extreme one in the direction of the alternative, given the null hypothesis is true.

p-value is not the probability that H_0 is true.

We may replace step 3 and 4 of the critical region procedure with

3. Determine the p-value.

4. Make a decision. Determine if the p-value is less than α. If it does, reject the null hypothesis. If it does not, do not reject the null hypothesis.

Confidence Interval Approach:

1. State the null hypothesis H_0 and the alternative hypothesis H_1.
2. Calculate the confidence interval (one sided or two sided depending on H_1).

3. Make a decision. Determine if the confidence interval covers the value in H_0. If it does not, reject the null hypothesis. If it does, do not reject the null hypothesis.

Comparing two proportions:

Suppose there are two independent populations: $Y_1 \sim \text{Bin}(n_1, p_1)$ and $Y_2 \sim \text{Bin}(n_2, p_2)$.

By central limit theorem, we have

$$Y_1/n_1 \sim N(p_1, \frac{p_1(1-p_1)}{n_1}),$$

$$Y_2/n_2 \sim N(p_2, \frac{p_2(1-p_2)}{n_2}),$$

Under H_0: $p_1 = p_2 = p$.

Example 2: Time magazine reported the result of a telephone poll of 800 adult Americans. The question posed of the Americans who were surveyed was: "Should the federal tax on cigarettes be raised to pay for health care reform?" The results of the survey were:

Non-Smokers	Smokers
$n_1 = 605$	$n_2 = 195$
$y_1 = 351$ said "yes"	$y_2 = 41$ said "yes"
$\widehat{p}_1 = \frac{351}{605} = 0.58$	$\widehat{p}_2 = \frac{41}{195} = 0.21$

Is there sufficient evidence at the $\alpha = 0.05$ level, say, to conclude that the two populations – smokers and nonsmokers – differ significantly with respect to their opinions?