Hypothesis Test

Yanrong Yang

Research School of Finance, Actuarial Studies and Statistics
The Australian National University

May 8th, 2017

Content

Hypothesis Testing

Definition (Hypothesis)

A **hypothesis** is a statement about a population parameter.

The goal of a **hypothesis test** is to decide, based on a sample from the population, which of two complementary hypotheses is true.

Definition (Two Complementary Hypotheses)

The two complementary hypotheses in a hypothesis testing problem are called the **null hypothesis** and the **alternative hypothesis**. They are denoted by H_0 and H_1 (or H_a) respectively.

Hypothesis Testing

Definition (Hypothesis Test)

A **hypothesis testing procedure** or **hypothesis test** is a rule that specifies:

1. for which sample values the decision is made to accept H_0 as true;
2. for which values H_0 is rejected and H_1 is accepted as true.

The subset of the sample space for which H_0 will be rejected is called the **rejection region** (RR) or **critical region**. The complement of the rejection region is called the **acceptance region** (AR).

A hypothesis test is specified by a **test statistic** to determine AR or RR.

An Intuitive Example

Question: We have a bent coin and there is interest in p, the probability of heads coming up on a single toss.

Analysis: The following hypothesis test is used to decide whether a bent coin is fair or not.

1. **The null hypothesis** $H_0 : p = 1/2$. This states that the coin is fair.
2. **The alternative hypothesis** $H_a : p \neq 1/2$. This states that the coin is unfair.
3. **Statistical Experiment**: toss the coin $n = 10$ times and record Y, the number of heads that come up.
4. **Test Statistic**: $Y \sim Bin(10, 0.5)$ if H_0 is true.
5. **Acceptance Region**: if $3 \leq Y \leq 7$ then accept H_0; otherwise reject H_0.

Two Criteria to Evaluate Hypothesis Test

Definition (Type I and Type II Errors)

A **Type I error** occurs if H_0 is rejected when H_0 is true.
The probability of a Type I error is denoted by α, and may also be called the **significance level (SL)** of the test.
A **Type II error** occurs if H_a is rejected when H_a is true.
The probability of a Type II error is denoted by β.

Relation between Two Type Errors

Consider the following hypothesis test

1. $H_0 : \theta \in \Theta_0$ vs $H_a : \theta \in \Theta_0^c$.
2. Test Statistic: $\mathbf{X}$.

Then

$$P_\theta(\mathbf{X} \in RR) = \begin{cases} P(Type\ I\ error) & \theta \in \Theta_0; \\ 1 - P(Type\ II\ error) & \theta \in \Theta_0^c. \end{cases} \tag{1}$$

Example 2

Question: Find α and β for the test in the last example.

Analysis:

1. As $Y \sim Bin(10, 0.5)$, $\alpha = P(Reject\ H_0|H_0\ true) = P(Y \in \{0,1,2,8,9,10\}) = 2P(Y \le 2) = 2 \times 0.055 = 0.11$.
2. As $Y \sim Bin(10, p)$ with $p \neq 0.5$, $\beta = P(Reject\ H_a|H_a true) = P(Do\ not\ reject\ H_0|H_a true) = P(Y \notin \{0,1,2,8,9,10\}) = P(3 \le Y \le 7) = \sum_{y=3}^{7} \binom{10}{y} p^y (1-p)^{10-y}$, $0 \le p \le 1$ with $p \neq 0.5$.

Example 3

Question: Consider Example 2 again with the rejection region changed into $\{0, 1, 9, 10\}$. Find α and β under this case.

Analysis:

1. As $Y \sim Bin(10, 0.5)$, $\alpha = P(Reject\ H_0|H_0\ true) = P(Y \in \{0, 1, 9, 10\}) = 2P(Y \leq 1) = 2 \times 0.011 = 0.022$.
2. As $Y \sim Bin(10, p)$ with $p \neq 0.5$,
$\beta = P(Reject\ H_a|H_a true) = P(Do\ not\ reject\ H_0|H_a true) = P(Y \notin \{0, 1, 9, 10\}) = P(2 \leq Y \leq 8) = \sum_{y=2}^{8} \binom{10}{y} p^y (1-p)^{10-y}$,
$0 \leq p \leq 1$ with $p \neq 0.5$.

Another Criterion: The power

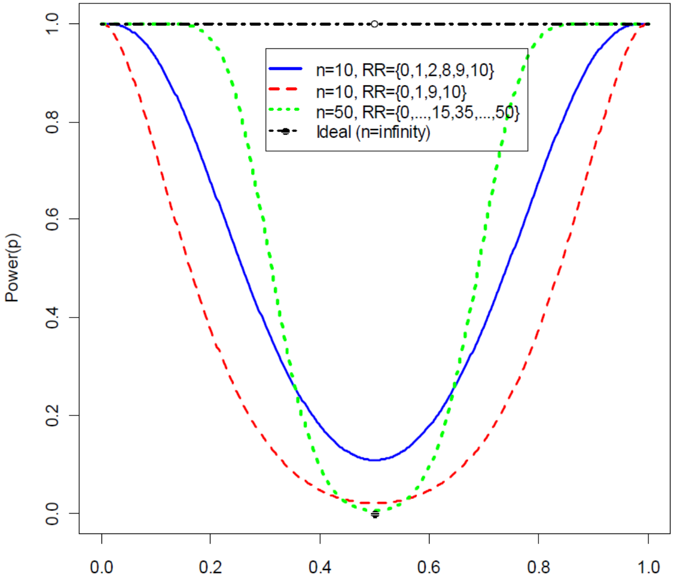
The power of a test combines all the information provided by α and β.

Definition (The power function)

The power function denoted by $Power(\theta)$ is the probability of rejecting the null hypothesis when the true value of the parameter under consideration is θ.

$$Power(\theta) = P_\theta(\mathbf{X} \in RR) = \begin{cases} P(Type\ I\ error) & \theta \in \Theta_0; \\ 1 - P(Type\ II\ error) & \theta \in \Theta_0^c. \end{cases}$$

Power of the Example



Example 4

Question: A coin is to be tossed $n = 100$ times to test whether or not it is fair. Determine a rejection region for Y, the number of heads which come up, so that the probability of making a Type I error is 5%.

Analysis:

1. As $Y \sim Bin(100, 0.5)$, $0.05 = P(Y \in RR)$.
2. The RR should be of the form $RR = \{0, \ldots, 50 - k, 50 + k, \ldots, 100\}$.
3. $0.025 = P(Y \geq 50 + k) = P\left(\frac{Y-50}{5} \geq \frac{50+k-50}{5}\right) \approx P(U > k/5)$, $U \sim N(0, 1)$.
4. By normal table, $1.96 = k/5$, and then $k = 9.8$.

Some Standard Hypothesis Tests

1. Z-test for a binomial proportion
2. Z-test for the difference between two binomial proportions
3. Z-test for a normal mean
4. t-test for a normal mean
5. Z-test for the mean of a distribution
6. t-test for the difference between two normal means
7. Two variants of the t-test for the difference between two normal means

Z-test for a Binomial Proportion

1. $Y \sim Bin(n, p)$ with large n $(n > 30)$.
2. $\hat{p} = \frac{Y}{n}$.
3. H_0: $p = p_0$ vs H_a: $p \neq p_0$.
4. Test Statistic: $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \dot{\sim} N(0, 1)$ if H_0 is true.
5. Rejection Region: $|Z| > z_{\alpha/2}$.

Example 5

Question: A die was tossed 1000 times and 196 sixes came up. Conduct an appropriate hypothesis test at the 1% level to decide whether or not the die is fair.

Analysis:

1. $H_0 : p = 1/6$ vs $H_a : p \neq 1/6$.
2. Test Statistic $Z = \frac{\hat{p}-1/6}{\sqrt{\frac{(1/6)(1-1/6)}{1000}}} \dot{\sim} N(0,1)$ under H_0.
3. Rejection Region: $|Z| > z_{0.005} = 2.576$.
4. $\hat{p} = y/n = 0.196$ and $z = \frac{0.196-1/6}{\sqrt{\frac{(1/6)(1-1/6)}{1000}}} = 2.489$.
5. Since $|2.489| < 2.576$, conclusion is that the die is fair.

Z-test for Difference between Two Bionomial Proportions

1. $X \sim Bin(n, p)$, $Y \sim Bin(m, q)$, $X \perp Y$, n and m are both large, $\hat{p} = X/n$, $\hat{q} = Y/m$.
2. $H_0 : p - q = \delta$ vs $H_a : p - q = \delta$.
3. Test Statistic: $Z = \frac{\hat{p}-\hat{q}-\delta}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{\hat{q}(1-\hat{q})}{m}}} \dot{\sim} N(0, 1)$ under H_0.
4. Rejection Region: $|Z| > z_{\alpha/2}$.

Remark

As $\delta = 0$ under H_0, $p = q$. Then we propose the estimator $\hat{r} = \frac{X+Y}{m+n}$. By replacing $\hat{p}$ and $\hat{q}$ with $\hat{r}$ in the TS above, we have a new TS $Z = \frac{\hat{p}-\hat{q}}{\sqrt{\hat{r}(1-\hat{r})\left(\frac{1}{n}+\frac{1}{m}\right)}}$.

Example 6

Question: 100 people were sampled in Sydney and 63 of these were found to be Liberals. 200 people were sampled from Melbourne and 102 of these were found to be Liberals. We are interested in whether the proportion of Liberals in Sydney is the same as the proportion of Liberals in Melbourne. Carry out an appropriate hypothesis test at the 5% level.

Analysis:

1. $H_0 : p - q = 0$ vs $H_a : p - q \neq 0$.
2. TS: $Z = \frac{\hat{p}-\hat{q}}{\sqrt{\hat{r}(1-\hat{r})\left(\frac{1}{n}+\frac{1}{m}\right)}} \dot{\sim} N(0,1)$.
3. RR: $|Z| > z_{\alpha/2}$.
4. Calculation: $\hat{p} = 63/100 = 0.63$, $\hat{q} = 102/200 = 0.51$, $\hat{r} = (63+102)/(100+200) = 0.55$, $\alpha = 0.05$, $z_{\alpha/2} = z_{0.025} = 1.96$. Then the value of TS is $z = 1.97$ and RR is $|Z| > 1.96$. So our conclusion is to reject H_0 and say that at the 5% level of statistical significance, the proportions of Liberals in the two cities are different.

Z-test for a Normal Mean

1. $Y_1, \ldots, Y_n \sim i.i.dN(\mu, \sigma^2)$ with σ^2 being known.
2. $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$.
3. TS: $Z = \frac{\bar{Y} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1)$ under H_0.
4. RR: $|Z| > z_{\alpha/2}$.

Example 7

Question: A bottling machine dispenses volumes that are normally distributed with a variance of 16 square ml. 8 bottles were sampled and their average volume was 941.6 ml. Test whether the machine dispenses 950 ml on average at the 5% level.

Analysis:

1. $H_0 : \mu = 950$ vs $H_a : \mu \neq 950$.
2. TS: $Z = \frac{\bar{Y} - 950}{4/\sqrt{8}} \sim N(0, 1)$ under H_0.
3. RR: $|Z| > z_{\alpha/2} = z_{0.025} = 1.96$.
4. $z = \frac{941.6 - 950}{4/\sqrt{8}} = -5.94 \in RR$. So we should reject H_0 and conclude that the machine does not dispense 950 ml on average.

t-test for a Normal Mean

1. $Y_1, \ldots, Y_n \sim i.i.dN(\mu, \sigma^2)$ with σ^2 being unknown.
2. $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$.
3. TS: $Z = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$ under H_0.
4. RR: $|Z| > t_{\alpha/2}(n-1)$.

Example 8

Question: A bottling machine dispenses volumes that are normally distributed with unknown variance. 8 bottles were sampled and their average volume was 941.6 ml. Also the sample variance of the 8 volumes was 15.2. Test whether the machine dispenses 950 ml on average at the 5% level.

Analysis:

1. $H_0 : \mu = 950$ vs $H_a : \mu \neq 950$.
2. TS: $T = \frac{\bar{Y}-950}{S/\sqrt{8}} \sim t(7)$ under H_0.
3. RR: $|T| > t_{0.025}(7) = 2.365$.
4. $t = \frac{941.6-950}{\sqrt{15.2}/\sqrt{8}} = -6.09 \in RR$. So we should reject H_0 and conclude that the machine does not dispense 950 ml on average.

Z-test for the Mean of a distribution

1. $Y_1, \ldots, Y_n \sim i.i.d(\mu, \sigma^2)$ with σ^2 being unknown.
2. $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$.
3. TS: $Z = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \dot{\sim} N(0, 1)$ under H_0.
4. RR: $|Z| > z_{\alpha/2}$.

t-test for the Difference between Two Normal Means

1. $X_1, \ldots, X_n \sim i.i.dN(\mu_1, \sigma_1^2)$ and $Y_1, \ldots, Y_m \sim i.i.d.N(\mu_2, \sigma_2^2)$. $X_i \perp Y_j$. $\sigma_1^2 = \sigma_2^2$ are both unknown.
2. $H_0 : \mu_1 - \mu_2 = \delta \quad$ vs $\quad H_a : \mu_1 - \mu_2 \neq \delta$.
3. TS: $T = \frac{(\bar{X} - \bar{Y}) - \delta}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n + m - 2)$ with $S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$.
4. RR: $|T| > t_{\alpha/2}(n + m - 2)$.

Example 9

Question: 6 bottles filled by a bottling machine today have volumes $1.80, 1.90, 1.75, 1.81, 1.74, 1.82$. 4 bottles filled by the same machine yesterday have volumes $1.61, 1.79, 1.78, 1.83$. Using a significance level 2% test that the mean volumes were the same today and yesterday. Assume that volumes are normal with the same variance on both days.

Analysis:

1. $H_0 : \mu_1 - \mu_2 = 0$ vs $H_a : \mu_1 - \mu_2 \neq 0$.
2. TS: $T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{6} + \frac{1}{4}}} \sim t(8)$.
3. RR: $|T| > t_{0.01}(8) = 2.896$.
4. $t = 1.05 \notin RR$ and H_0 is not rejected. We conclude that there is no difference between the mean volumes dispensed today and yesterday.

Two Variants of t-test for the Difference between Two Normal Means

Suppose that σ_1^2 and σ_2^2 are known.

1. TS: $Z = \frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1)$.
2. RR: $|Z| > z_{\alpha/2}$.

Suppose that σ_1^2 and σ_2^2 are unknown, the underlying distributions are not normal, and n and m are large.

1. TS: $Z = \frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}} \dot{\sim} N(0,1)$.
2. RR: $|Z| > z_{\alpha/2}$.

Example 10

Question: 450 people were sampled in Sydney and their incomes were determined. The mean and standard deviation of the sample incomes were 12054 and 1501. 875 people were sampled in Melbourne and their incomes were determined. The mean and standard deviation of the sample incomes were 9043 and 989. Test at the 2% level whether the mean income in Sydney is exactly $\$2000$ more than the mean income in Melbourne.

Analysis:

1. $H_0 : \mu_1 - \mu_2 = 2000$ vs $H_a : \mu_1 - \mu_2 \neq 2000$.
2. TS: $Z = \frac{\bar{X} - \bar{Y} - 2000}{\sqrt{\frac{S_1^2}{450} + \frac{S_2^2}{875}}} \dot{\sim} N(0, 1)$.
3. RR: $|Z| > z_{0.01} = 2.326$.
4. $z = 12.9 \in RR$. So H_0 should be rejected and we conclude that the difference is not $\$2000$.

Another Criterion: p-value

Definition (p-value)

The p-value or attained significance level, is the smallest level of significance α for which the observed data indicate that the null hypothesis H_0 should be rejected.

Rejection region and p-value are two different criteria to be utilized for determining the conclusion.

Example 11

Question: A random sample of 300 widgets produced in a factory weights a total of 565.8 kg and the sample standard deviation of the weights of the 300 widgets is 7.6 kg. We are interested in whether the average weight of all widgets is 1 kg. Carry out an hypothesis test at the 5% level and report the associated p-value.

Analysis:

1. $H_0 : \mu = 1 \quad$ vs $\quad H_a : \mu \neq 1$.
2. TS: $Z = \frac{\bar{Y}-1}{S/\sqrt{300}} \dot{\sim} N(0,1)$.
3. RR: $|Z| > 1.96$. $z = 2.02 \in RR$. So H_0 should be rejected. We conclude that the average weight of all widgets is not exactly 1 kg.
4. $p - value = P(|Z| > z) = 2P(Z > 2.02) = 0.0435$. Since $0.0435 < 0.05 = \alpha$, we should reject H_0 and get the same conclusion.

One-sided Tests

One-sided test includes

1. **lower-tail test**: lower-tail RR $Z < -z_\alpha$;
2. **upper-tail test**: upper-tail RR $Z > z_\alpha$.

Example 12

Question: It has been alleged that the proportion of defective bolts produced by a certain factory is greater than 10%. A sample of 1000 bolts was taken, and 118 of these were found to be defective. Conduct an appropriate hypothesis test at the 5% level and calculate the associated p-value.

Analysis:

1. $H_0 : p = 0.1 \quad \text{vs} \quad H_a : p > 0.1.$
2. TS: $Z = \frac{\hat{p}-0.1}{\sqrt{p(1-p)/n}} \dot{\sim} N(0,1).$
3. RR: $Z > z_{0.05} = 1.645.$
4. $z = 1.90 \in RR$ and H_0 is rejected. We conclude that the allegation is true.
5. $p-value = P(Z > 1.90) = 0.0287 < 0.05.$ The same conclusion can be derived.

Example 13

Question: There is concern that a chocolate factory is producing chocolate bars which are on average lighter than the advertised 100 grams. 8 bars were sampled from the production line and weighted. The sample weights had a mean of 96.7 and a standard deviation of 2.9. Conduct a hypothesis test at the 1% level and report the p-value. State any assumptions made.

Analysis:

1. $H_0 : \mu = 100$ vs $H_a : \mu < 100$.
2. TS: $T = \frac{\bar{Y}-100}{S/\sqrt{8}} \sim t(7)$.
3. RR: $T < -t_{0.01}(7) = -2.998$.
4. $t = -3.22 \in RR$ and H_0 should be rejected. We conclude that the factory is making underweight chocolate bars.

Summary

1. The elements of a hypothesis test: hypotheses, test statistics, significance level, rejection region, p-value, conclusion;
2. criteria to evaluate a hypothesis test: type I error, type II error and the power function;
3. some standard test statistic for classical hypothesis tests.