Mathematical Statistics II

Ch.8 Tests of Statistical Hypotheses

Jungsoon Choi

jungsoonchoi@hanyang.ac.kr

Table of Contents

- Tests of the Equality of Two Means
- Tests about Proportions

Ch8.2 Tests of the Equality of Two Means

Tests of the Equality of Two Means

Suppose that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Here, we assume that $\sigma_X^2 = \sigma_Y^2$. We are interested in

$$H_0 : \mu_X = \mu_Y \quad vs \quad H_1 : \mu_X < \mu_Y$$

In computing C.I for the difference in means,

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n + m - 2)$$

where S_p^2 is the pooled estimator for σ^2

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n + m - 2}$$

Under H_0 is true, the test statistic is

$$T = \frac{(\overline{X} - \overline{Y})}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2)$$

Thus, the rejection region with α is

$$T \leq -t_\alpha(n+m-2)$$

$X_i \sim N(\mu_X, \sigma_X^2)$, $\quad Y_j \sim N(\mu_Y, \sigma_Y^2)$ $\quad j = 1, \cdots, m$

$\sigma_X^2 = \sigma_Y^2 \overset{let}{=:} \sigma^2$

$H_0 : \mu_X = \mu_Y$ vs $H_1 : \mu_X < \mu_Y$

$\Leftrightarrow H_0 : \mu_X - \mu_Y = 0$ vs $H_1 : \mu_X - \mu_Y < 0$

Reject H_0 if $\underline{\bar{X} - \bar{Y} < C_\alpha}$ s.t $C_\alpha < 0$

$\underline{\alpha = P(\bar{X} - \bar{Y} < C_\alpha \mid \mu_X = \mu_Y)}.$

Since $\bar{X} \sim N(\mu_X, \frac{\sigma_X^2}{n})$, $\bar{Y} \sim N(\mu_Y, \frac{\sigma_Y^2}{m})$.

$\bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, \sigma^2(\frac{1}{n} + \frac{1}{m}))$.

$\underline{\alpha = P[\bar{X} - \bar{Y} < C_\alpha \mid (\bar{X} - \bar{Y}) \sim N(0, \sigma^2(\frac{1}{n} + \frac{1}{m}))]}$

$= P[T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} < \underbrace{\frac{C_\alpha}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}}_{-t_\alpha(n+m-2).}]$ $\quad T \sim t(n+m-2).$

$C_\alpha = S_p\sqrt{\frac{1}{n} + \frac{1}{m}} \times (-t_\alpha(n+m-2))$

$\therefore$ Rejection region is i) $\bar{X} - \bar{Y} < -t_\alpha(n+m-2) S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$

or

test statistic

ii) $T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} < -t_\alpha(n+m-2)$

$p\text{-value} = P(T < \frac{\bar{x} - \bar{y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}).$

Example 8.2-1

Suppose that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Here, we assume that $\sigma_X^2 = \sigma_Y^2$. We are interested in

$$H_0 : \mu_X = \mu_Y \quad vs \quad H_1 : \mu_X < \mu_Y$$

We have $\overline{x} = 1.03$, $s_X^2 = 0.24$ from 11 samples and $\overline{y} = 1.66$, $s_Y^2 = 0.35$ from 13 samples. Compute the test statistic and make the conclusion under $\alpha = 0.05$. Also, compute the p-value.

재밌기. ~

Suppose that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, where $\sigma_X^2,\ \sigma_Y^2$ are known. We are interested in

$$H_0 : \mu_X = \mu_Y \quad vs \quad H_1 : \mu_X < \mu_Y$$

Under H_0 is true, the test statistic is

$$Z = \frac{(\overline{X} - \overline{Y})}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1)$$

If $\sigma_X^2,\ \sigma_Y^2$ is unknown but $n,\ m$ are large,

$$Z = \frac{(\overline{X} - \overline{Y})}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim N(0,1)$$

Thus, the rejection region with α is

$$Z \leq -Z_\alpha$$

$X_i \sim N(\mu_X, \sigma_X^2)$, $Y_j \sim N(\mu_Y, \sigma_Y^2)$ $j = 1, \cdots, m$

σ_X^2, σ_Y^2 known.

$H_0: \mu_X = \mu_Y$ vs $H_1: \mu_X < \mu_Y$

$\Leftrightarrow$ $H_0: \mu_X - \mu_Y = 0$ vs $H_1: \mu_X - \mu_Y < 0$

Reject H_0 if $\bar{X} - \bar{Y} < C_\alpha$ s.t $C_\alpha < 0$

$\alpha = P(\bar{X} - \bar{Y} < C_\alpha \mid \mu_X = \mu_Y)$.

Since $\bar{X} \sim N(\mu_X, \frac{\sigma_X^2}{n})$, $\bar{Y} \sim N(\mu_Y, \frac{\sigma_Y^2}{m})$.

$\bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})$.

$$\alpha = P[\bar{X} - \bar{Y} < C_\alpha \mid (\bar{X} - \bar{Y}) \sim N(0, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})]$$

$$= P\left[\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = Z < \frac{C_\alpha}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}\right] \quad Z \sim N(0,1).$$

If σ_X^2, σ_Y^2 unknown, but n, m large.

$$= P\left[\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} = Z < \frac{C_\alpha}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}\right] \quad Z \sim N(0,1).$$

Example 8.2-3

Suppose that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. We are interested in

$$H_0 : \mu_X = \mu_Y \quad vs \quad H_1 : \mu_X < \mu_Y$$

We have $\overline{x} = 6.701,\ s_X = 0.108$ from 50 samples and $\overline{y} = 6.841,\ s_Y = 0.155$ from 40 samples. Compute the test statistic and make the conclusion under $\alpha = 0.01$.

Ch8.3 Tests about Proportions

Testing hypothesis about proportion

Example

In an electronic company, the error rate is $p = 0.06$. They proposed a new method to improve the performance. To test the new method, 200 items from the new method were made. If the error rate from the new method is less than 0.06, then we suggest that the new method improves the performance.

Let Y be the number of error items among 200 samples.

- Construct the null hypothesis and alternative hypothesis.
- Find the rejection region given the $\alpha = 0.05$.

$H_0 : p = 0.06 \quad vs \quad H_1 : p < 0.06.$

p = error rate.

Y : # of error items among 200 samples

$\sim Bin(200, p).$

As $n \to \infty$ $\hat{p} = \frac{Y}{200} \sim N(p, \frac{p(1-p)}{n})$

Reject H_0 if $\hat{p} = \frac{Y}{n} < c_\alpha$

$$\alpha = P(\tfrac{Y}{n} = \hat{p} < c_\alpha \mid p = 0.06)$$

$$= P[\hat{p} < c_\alpha \mid \hat{p} \sim N(0.06, \tfrac{0.06 \times 0.94}{200})]$$

$$= P\left(\frac{\hat{p} - 0.06}{\sqrt{\frac{0.06 \times 0.94}{200}}} = Z < \frac{c_\alpha - 0.06}{\sqrt{\frac{0.06 \cdot 0.94}{200}}}\right]$$

$Z \sim N(0,1).$

$-Z_{0.05} = -1.64$

$$\alpha = 0.05 \Rightarrow \frac{c_{0.05} - 0.06}{\sqrt{\quad}} = -1.64 \Rightarrow c_{0.05} = \_\_\_\_$$