Comparing two Treatments

Hypothesis Tests with Two Means

Prof Hans Georg Schaathun

Høgskolen i Ålesund

21st March 2014

Example Problem

Claim An average Engineering student drink more beer on Saturdays than an average Business student.

- We compare two means
  - average of engineering students μ_E
  - average of business students μ_B
- We can form hypotheses
  - $H_0 : \mu_E = \mu_B$
  - $H_1 : \mu_E > \mu_B$
- Poll two populations
  - $E_1, E_2, E_3, \ldots, E_n$
  - $B_1, B_2, B_3, \ldots, B_m$
- Two sample means to use: $\bar{E}$ and $\bar{B}$

In this video, we assume known σ_E and known σ_B.

Reducing the number of variables

- $H_0 : \mu_E = \mu_B$
- Two stochastic variables $\bar{E}$ and $\bar{B}$
- How do we deal with two variables?

1. Define one key quantity $\delta = \mu_E - \mu_B$
2. $H_0 := \delta = 0$ where
3. One stochastic variable $D = \bar{E} - \bar{B}$
4. Note $E(D) = \delta$

Reducing the number of variables

- $H_0 : \mu_E = \mu_B$
- Two stochastic variables $\bar{E}$ and $\bar{B}$
- How do we deal with two variables?

1. Define one key quantity $\delta = \mu_E - \mu_B$
2. $H_0 := \delta = 0$ where
3. One stochastic variable $D = \bar{E} - \bar{B}$
4. Note $E(D) = \delta$

Probability Distribution

We need to know the probability distribution of $D = \bar{E} - \bar{B}$ *under* H_0

- Normal Distribution if either
  1. Large samples — the Central Limit Theorem applies
  2. The populations (E and B) are normally distributed
- $D \sim N(0, \sigma_D)$ under H_0

Finding σ_D

The variance of a sum

$$D = \bar{E} - \bar{B}, \tag{1}$$

$$\operatorname{var}(D) = \operatorname{var}(\bar{E}) + \operatorname{var}(\bar{B}), \tag{2}$$

$$\operatorname{var}(D) = \frac{\sigma_E^2}{n_E} + \frac{\sigma_B^2}{n_B} \tag{3}$$

$$\sigma_D = \sqrt{\frac{\sigma_E^2}{n_E} + \frac{\sigma_B^2}{n_B}} \tag{4}$$

Normalising

$$Z = \frac{D}{\sigma_D} \tag{5}$$

$$Z = \frac{\bar{E} - \bar{B}}{\sqrt{\sigma_E^2/n_E + \sigma_B^2/n_B}} \tag{6}$$

- Z has Standard Normal Distribution
- We can proceed as we did with a single mean
- (one-sided) reject if $Z > z_\alpha$
- (two-sided) reject if $|Z| > z_{\alpha/2}$

HØGSKOLEN I ÅLESUND Aalesund University College

More generally

- $H_0 : \mu_E - \mu_B = \delta_0$

$$Z = \frac{\bar{E} - \bar{B} - \delta_0}{\sqrt{\sigma_E^2/n_E + \sigma_B^2/n_B}}$$

- Z has Standard Normal Distribution
- We can proceed as we did with a single mean
- (one-sided) reject if $Z > z_\alpha$
- (two-sided) reject if $|Z| > z_{\alpha/2}$

Summary

- Comparing two treatments is a common problem
- Null hypothesis: $H_0 : \mu_X - \mu_Y = 0$
  - more generally $H_0 : \mu_X - \mu_Y = \delta$
- Observable statistic: $D = \bar{X} - \bar{Y}$

$$Z = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\sigma_X^2/n_X + \sigma_Y^2/n_Y}}$$

- (one-sided) reject if $Z > z_\alpha$
- (two-sided) reject if $|Z| > z_{\alpha/2}$