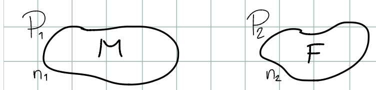
Resumé



$\hat{P}_1 - \hat{P}_2$ - statistic

$E[\hat{P}_1 - \hat{P}_2] = P_1 - P_2$ - unbiased estimater

$Var[\hat{P}_1 - \hat{P}_2] = \frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}$

$\hat{P}_1 - \hat{P}_2$ - is approximately Normal for large n_1 and n_2 due to the CLT.

Assumption

- Random sampling
- Independent samples   arbitrary rule of thumb
- $x_1, n_1 - x_1, x_2, n_2 - x_2 \geq 5$

$\hat{P}_1 - \hat{P}_2 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$, two-sided $100(1-\alpha)\%$

Ex

GP ordered a survey. 1181 adults answered the survey conducted by Svensha Vegandforeningen. 747 men and 434 women answered.
276 men and 195 women said that they wanted V-food.
Find a 90% CI for $P_1 - P_2$.

$n = 1181$

$n_1 = 747$, $x_1 = 276$

$n_2 = 434$ $x_2 = 195$

$\alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow Z_{\frac{\alpha}{2}} = 1.645$

$\hat{P}_1 = \frac{276}{747} = 0.369$

$\hat{P}_2 = \frac{195}{434} = 0.449$

$0.369 - 0.449 \pm 1.645\sqrt{\frac{0.369(1-0.369)}{747} + \frac{0.449(1-0.449)}{434}} \Rightarrow$ 90% CI : $(-0.129, -0.031)$

Comparing two means

Ex

AAUP wanted to compare the mean sallery between public and private universities.
The survey included 35 faculty members from private unis and 30 from public.
The samples are given in thousands of USD/year

Sample 1 (Private)	Sample 2 (Public)
87.3, 75.9, 108.8, 83.9, 56.6, 99.2, 54.9...	49.9, 105.57, 116.1, 40.3, 123.1, 79.3...

Population 1 - All faculty in private uni.

Population 2 - All faculty in public uni.

μ_1 - Mean salery for Population 1.

μ_2 - Mean salery for Population 2.

1) $H_0 : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

2) Test statistic $\overline{X}_1 - \overline{X}_2$, $\overline{X}_x$ is the sample mean of population x.

$\overline{x}_1 = \frac{3086.8}{35} = 88.19$

$\overline{x}_2 = \frac{2195.4}{30} = 73.18$

1. $E[\overline{X}_1 - \overline{X}_2] = E[\overline{X}_1] - E[\overline{X}_2] = \mu_1 - \mu_2$
2. $Var[\overline{X}_1 - \overline{X}_2] = Var[\overline{X}_1] + Var[\overline{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ Population
3. $\overline{X}_1 - \overline{X}_2$ is approximately normal: $N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ ~ approx $N(0,1)$

$\sigma_1^2 = \sigma_2^2 \Rightarrow Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ ~ approx $N(0,1)$

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\underbrace{n_1+n_2-2}_{\text{makes it unbiased}}}$

$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$ ~ Pooled sample deviation

$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ ~ t-distribution, df: n_1+n_2-2

Pooled t-test

Assumptions

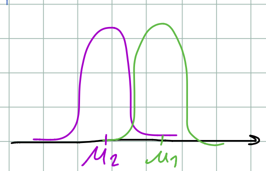
1. Random sampling
2. Independent samples
3. Normal populations, or large sample
4. Equal populations and deviations

Steps

1. Formulate H_0 and H_A.
   $H_0: \mu_1 = \mu_2$
   $H_A: \mu_1 \neq \mu_2$ ~ 2-sided
   $\mu_1 \gtrless \mu_2$ ~ 1-sided
2. Decide α
3. $T = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$ if $H_0: \mu_1 - \mu_2 = 0$
4. Critical values
   $\pm t_{(\frac{\alpha}{2}, n_1+n_2-2)}$ ~ 2-sided
   $-t_{(\frac{\alpha}{2}, n_1+n_2-2)}$ ~ 1-sided left tail
   $+t_{(\frac{\alpha}{2}, n_1+n_2-2)}$ ~ 1-sided right tail
5. Find when t (the value of test statistics) lies
6. Interpret the result

Back to uni Ex

$S_1 = 26.21$ $S_2 = 23.95$



So far: $S_1 = 26.21$ $S_2 = 23.95$
$x_1 = 88.19$ $x_2 = 73.18$
$n_1 = 35$ $n_2 = 30$

2. $\alpha = 0.05$

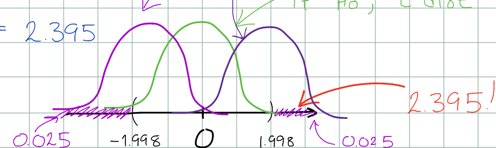
if H_A it will be moved away from 0.

3. $S_p = 25.19$

if H_0, t-dist - goes towards normal around 0.

$t = \frac{88.19 - 73.18}{25.19\sqrt{\frac{1}{35} + \frac{1}{30}}} = 2.395$

4. $t_{\frac{\alpha}{2}, 63} = 1.998$



5. $2.399 \notin (-1.998, 1.998)$

6. Accept H_1 at $\alpha = 0.05$

$100(1-\alpha)\%$

$100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$, given $\sigma_1^2 = \sigma_2^2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \cdot S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Example, continues

$$(88.19 - 73.18) \pm 1.998 \sqrt{\frac{1}{35} + \frac{1}{30}} = (2.49, 27.53) \sim 95\% \text{ CI}$$

The difference between salaries is somewhere between 2 490 and 27 530 USD.