Large-Sample Tests of Hypotheses (Part 2)

Outline

- Large-sample test of hypothesis for the difference between two population means
- Large-sample test of hypothesis for a binomial proportion
- Large-sample test of hypothesis for the difference between two binomial proportions

Large-sample test of hypothesis for the difference between two population means

Assumptions. Given two *large* samples (n_1>30 and n_2>30) *randomly and independently drawn* from two populations.

(1) Null hypothesis H_0: $\mu_1 - \mu_2 = D_0$

(2) Alternative hypothesis H_a: $\mu_1 - \mu_2 > D_0$, or Ha: $\mu_1 - \mu_2 < D_0$ (one-tailed test)

H_a: $\mu_1 - \mu_2 \neq D_0$ (two-tailed test)

(3) Test statistic

$$z = \frac{(\mu_1 - \mu_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx \frac{(\mu_1 - \mu_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

(4) Rejection region

$z > z_\alpha$ for H_a: $\mu_1 - \mu_2 > D_0$

$z < -z_\alpha$ for H_a: $\mu_1 - \mu_2 < D_0$

$z > z_{\alpha/2}$ or $z < -z_{\frac{\alpha}{2}}$ for H_a: $\mu_1 - \mu_2 \neq D_0$

OR $\mathrm{pValue} < \alpha$ (for any H_a)

Large-sample test of hypothesis for the difference between two population means

Example. A 2018 survey from 100 randomly selected foreign tourists in Hanoi and 100 randomly selected tourists in HCMC revealed that the average stay in Hanoi and in HCMC were 3.4 days and 3.67 days respectively. The standard deviation of the Hanoi sample is 1.2 days and of the HCMC sample is 1 day. Was there enough evidence to conclude that the average length of stay of foreign tourists are different between the 2 cities? Use $\alpha = .05$

Calculate the 95% confidence intervals of the difference between 2 population means.

Large-sample test of hypothesis for a binomial proportion

Assumptions. Given a *large* number of n identical trials *randomly drawn* from a binomial population, i.e. $\mathrm{n}\hat{p} > 5$ and $\mathrm{n}(1-\hat{p}) > 5$

- (1) Null hypothesis $\mathrm{H_0}$: $p = p_0$
- (2) Alternative hypothesis $\mathrm{H_a}$: $p < p_0$ or $\mathrm{H_a}$: $p > p_0$ (one-tailed test)

  $\mathrm{H_a}$: $p \neq p_0$ (two-tailed test)
- (3) Test statistic $z = (\hat{p} - p_0)/\sqrt{p_0(1-p_0)/n}$

  where $\hat{p}$ is sample proportion
- (4) Rejection region: $z > z_\alpha$ for $\mathrm{H_a}$: $p > p_0$

  $z < -z_\alpha$ for $\mathrm{H_a}$: $p < p_0$

  $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$ for $\mathrm{H_a}$: $p \neq p_0$

OR $\mathrm{pValue} < \alpha$ (for any $\mathrm{H_a}$)

Large-sample test of hypothesis for a binomial proportion

Example. A survey in the first half of 2019 observed that out of 500 random visitors to HCMC, 123 were foreigners. Is this evidence sufficient to conclude that the proportion of foreign visitors to HCMC has increased from 2018, which was approximately 20.4%? Use $\alpha = .05$

Large-sample test of hypothesis for the difference between 2 binomial proportions

Assumption. Given two samples *independently and randomly* drawn from two binomial populations, and that each sample has *large* number trials, i.e. $n_1\widehat{p_1}$, $n_2\widehat{p_2}$, $n_1(1-\widehat{p_1})$, and $n_2(1-\widehat{p_2})$ larger than 5.

(1) Null hypothesis — $H_0: p_1 - p_2 = D_0$

(2) Alternative hypothesis — $H_a: p_1 - p_2 > D_0$, or $H_a: p_1 - p_2 < D_0$ (one-tailed test)

$H_a: p_1 - p_2 \neq D_0$ (two-tailed test)

(3) Test statistic — $z = (\widehat{p_1} - \widehat{p_2} - D_0) \Big/ \sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$

where $\widehat{p_1}$ and $\widehat{p_2}$ are proportion of sample 1 and sample 2 respectively

(4) Rejection region — $z > z_\alpha$ for $H_a: p_1 - p_2 > D_0$

$z < -z_\alpha$ for $H_a: p_1 - p_2 < D_0$

$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$ for $H_a: p_1 - p_2 \neq D_0$

OR pValue $< \alpha$ (for any H_a)

Large-sample test of hypothesis for the difference between 2 binomial proportions

Example. A 2018 survey observed that 102 out of random 500 tourists visiting HCMC were foreigners. The same survey observed that only 98 out of random 500 tourists visiting Hanoi were foreigners. Is there enough evident to conclude that HCMC has a higher proportion of foreign visitors compared to Hanoi? Use $\alpha = .05$

Calculate the 95% confidence interval of the difference between the proportions of foreign visitors of the 2 cities.