BT1101 Introduction to Business Analytics

Hands-on Tutorial 05

① Point estimation
② Interval estimation
③ hypothesis Testing

Mathematical Statistics :

samples $\xrightarrow{\text{inference}}$ population X

known: it has a normal distribution

$N(\mu, \sigma^2)$ unknown: parameters μ, σ^2

Height of a person in the world.

population mean : μ

population variance : σ^2

before observed : $(X_1, X_2, \cdots, X_n)$ ← n random variables i.i.d

after observed : $(x_1, x_2, \cdots, x_n)$ ← n fixed constants

Point Estimation : samples $(X_1, \cdots, X_n)$ i.i.d with mean μ and variance σ^2

Unknown Parameters	Estimator	
population mean μ	☆ Sample Mean	$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$
population variance σ^2	☆ sample variance	$S^2 := \frac{1}{\underline{n-1}}\sum_{i=1}^{n}(X_i - \underset{\uparrow \text{ sample mean}}{\bar{X}})^2$
population proportion p	Sample proportion	$\bar{p} = \frac{\sum_{i=1}^{n} g_i}{n}$, $g_i = 1_{\{\text{the ith person is a girl}\}}$

why choose these estimators ? unbiased and consistent!
→ More in ST2132.

Rmk : Estimators are random variables! ☆very important!

∴ ☆☆Sample mean $\bar{X}$, Sample variance S^2, ☆sample proportion $\bar{p}$ — all have mean, variance and probability distribution!

☆ CLT : n is large, sample mean $\bar{X}$ is approx. normally distributed. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Must be tested!

NOTE : You Must state why $\bar{X}$ is normally distributed in your exam!

Interval Estimation : Give me a range instead of a single number!

Goal : We have 95% confidence that the population mean will be in the interval [50, 60]

General form : sample mean $\bar{X}$ ± margin of error

Summary :

① population mean with known σ : $\bar{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

② population mean with unknown σ : $\bar{X} \pm t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}}$

③ population proportion : $\bar{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ $\qquad Var(\bar{p}) = \frac{\bar{p}(1-\bar{p})}{n}$

Question: How to construct a confidence interval for unknown parameter?

Answer: Suppose we construct a random variable $W(X_1, X_2, \cdots, X_n, \theta)$ whose distribution is known and we can find proper constants c_1 and c_2 such that

unknown parameter

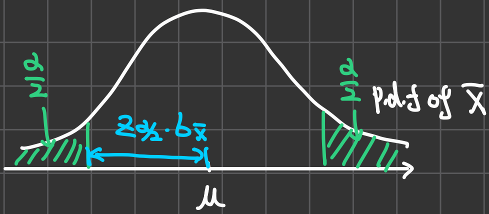
$$\gamma \leq P(c_1 \leq W(X_1, X_2, \cdots, X_n, \underline{\theta}) \leq c_2)$$

$$= P(S_1(c_1, X_1, X_2, \cdots, X_n) \leq \theta \leq S_2(c_2, X_1, X_2, \cdots, X_n)),$$

Then the interval $[S_1, S_2]$ is the confidence interval for the unknown θ with level γ

We call $W(X_1, X_2, \cdots, X_n, \theta)$ the pivotal.

Example: σ known. Want the CI of population mean μ



Choose pivotal: $\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

$$\therefore 1-\alpha \leq P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right)$$

$$= P\left(\mu - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \mu + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

$\therefore$ We have $1-\alpha$ confidence that the population mean is in the interval $[\mu - z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}}, \mu + z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}}]$

General form: $\bar{X} \pm z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}}$

Example: σ unknown. Want μ. See you in ST2132!

pivotal: $U := \dfrac{\sqrt{n} \cdot (\bar{X}_n - \mu)}{\sigma'} \sim t(n-1)$ with $\sigma' := \left[\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2\right]^{\frac{1}{2}}$

Hypothesis Testing:

Example 1. [Metro EMS]

40 samples of the response time of medical emergencies.

we calculated the sample mean = 13.25 minutes. population s.d. is known. $\sigma = 3.2$ min.

$\Rightarrow$ test statistic: $Z = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Hypothesis testing: With a 0.05 level of significance, to determine whether the service goal of 12 minutes or less is being achived.

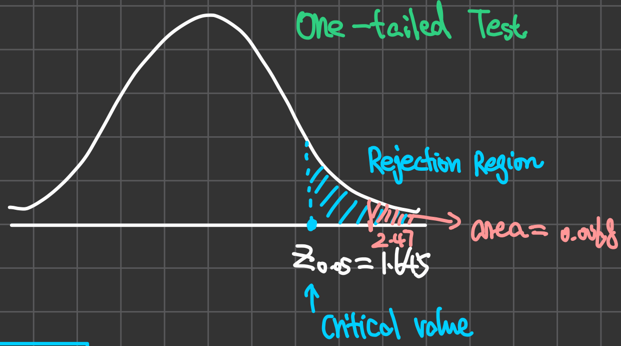
Step 1: Develop the Alternative Hypothesis

$\bar{x} = 13.25$ obviously longer than the target 12 minutes.

$\therefore \begin{cases} H_0: \mu \le 12 \\ H_A: \mu > 12 \end{cases}$

H_A should follow the obvious situation $\Rightarrow H_A: \mu > 12 \Rightarrow H_0: \mu \le 12$

One-failed Test



Step 2: Specify the level of significance $\alpha = 0.05$

Step 3: Compute the value of the test Statistic

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{13.25 - 12}{3.2/\sqrt{40}} = 2.47$$

$\boxed{\text{P-Value Approach}}$

Step 4: Compute the p-value

for $z = 2.47$, cumulative probability $= 0.9932$

$\therefore$ p-value $= 1 - 0.9932 = 0.0068$

Step 5: Determine whether to reject H_0

p-value $= 0.0068 \le \alpha = 0.05 \Rightarrow$ reject H_0.

$\boxed{\text{Critical Value Approach}}$

Step 4: Determine the critical value and rejection rule

For $\alpha = 0.05$, $z_{0.05} = 1.645$

$\therefore$ Reject H_0 if $z \ge 1.645$

Step 5: Determine whether to reject H_0.

Since $z = 2.47 > 1.645 \Rightarrow$ reject H_0

Critical values

NOTE: ① One-tailed: z_α or $-z_\alpha$

Two-tailed: $z_{\frac{\alpha}{2}}$ and $-z_{\frac{\alpha}{2}}$

② test statistics for μ with σ unknown

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

sample s.d.

test statistic for population proportion:

$$z = \frac{\bar{p} - p_0}{\sigma_p} \quad \text{with } \sigma_p = \sqrt{\frac{p_0(1-p_0)}{n}}$$

assuming $np \ge 5$ and $n(1-p) \ge 5$

$$\bar{p} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$$