GE2262 Business Statistics
Topic 7 Confidence Interval Estimation and Hypothesis Testing for Population Proportion

Lecturer: Dr. Iris Yeung

Room : LAU-7239

Tel No.: 34428566

E-mail: msiris@cityu.edu.hk

Outline

- Confidence Interval Estimate for Population Proportion
- Sample Size Determination for Estimating Population Proportion
- Hypothesis Testing for Population Proportion

Reference
Levine, D.M., Krehbiel, T.C. and Berenson, M.L., *Business Statistics: A First Course*, Pearson Education Ltd, Chapter 7 & 8 & 9

Part One

- Confidence Interval Estimate for Population Proportion
- Sample Size Determination for Estimating Population Proportion
- Hypothesis Testing for Population Proportion

Sampling Distribution of the Sample Proportion

Topic 3 Bernoulli/Binomial Experiment

- Four conditions:
  - The experiment is repeated *n* times (*n* trials).
  - Each trial has only two possible outcomes (denoted as success S and failure F).
  - The probability of success, denoted by *p*, is the same for each trial.
    - The probability of failure for each trial is equal to *q=1-p*.
  - The trials are independent (the outcome of a trial does not depend on the outcomes of previous trials).
- We are interested in the number of successes *X* observed in *n* trials → *X~BIN(n,p) where X can be 0,1,2,…,n.*
- If *np≥5 and nq≥5*, we can approximate a binomial distribution X ~ Bin (n, p) by a normal distribution $X \sim N(\mu, \sigma^2)$ with $\mu=np$, $\sigma^2=npq$

Topic 4 Sampling Distribution of Sample Proportion

- *p* is the probability of success or the population proportion of success
- We are interested in the sample proportion of success $\hat{p}$
- If *n* is large such that *np≥5, nq≥5*, the sampling distribution of the sample proportion is approximately normal with mean=*p*, variance = *pq/n*

$$\hat{p} \sim N(p, \frac{pq}{n}) \quad \Rightarrow Z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$

$$\hat{p} = \frac{X}{n} = \frac{\text{number of successes in } n \text{ trials}}{n \text{ trials}}$$

Two Types of Inferential Statistics

- Estimation (Topic 5)
  - Estimate the unknown population parameter
  - Example
    - We want to estimate the proportion of customers being satisfied with bank service
- Hypothesis Testing (Topic 6)
  - Test whether a hypothesis (claim or statement) about the population parameter holds or not
  - Example: suppose a bank manager claims that the proportion of customers being satisfied with their service is at least 0.9. We want to estimate whether the manager's claim holds or not

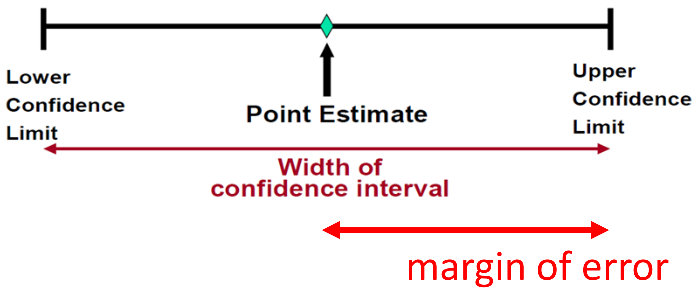
Measure	Population parameter	Sample statistic	Lecture
Mean	μ	$\bar{X}$	Topic 5 (estimation) Topic 6 (hypothesis testing)
Proportion	p	$\hat{p}$	Topic 7 (estimation and hypothesis testing)

Point and Interval Estimation for Population Parameter

- Point estimation - use the value of a sample statistic (a single number) to estimate unknown population parameter

	Population Parameter	Sample Statistic
Mean	μ	$\bar{X}$
Proportion	p	$\hat{p}$

- Confidence interval estimation - use a range (or an interval) of numbers to estimate unknown population parameter and state the level of confidence
  - Confidence interval = point estimate ± margin of error
  - The level of confidence is $100(1-\alpha)\%$. Most common confidence levels are: 90% (α=0.10) , 95% (α=0.05) , and 99% (α=0.01). Note that it can never be 100% confident



100(1-α)% Confidence Interval for Population Proportion *p*

If $np \geq 5,\ nq \geq 5,\ \hat{p} \sim N(p, \frac{pq}{n}) \Rightarrow Z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$

$$P\left(-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \leq z_{\alpha/2}\right) = 1-\alpha \quad \Rightarrow P\left(-z_{\alpha/2}\sqrt{\frac{pq}{n}} \leq \hat{p}-p \leq z_{\alpha/2}\sqrt{\frac{pq}{n}}\right) = 1-\alpha$$

$$\Rightarrow P\left(z_{\alpha/2}\sqrt{\frac{pq}{n}} \geq p-\hat{p} \geq -z_{\alpha/2}\sqrt{\frac{pq}{n}}\right) = 1-\alpha \quad \Rightarrow P\left(\hat{p}-z_{\alpha/2}\sqrt{\frac{pq}{n}} \leq p \leq \hat{p}+z_{\alpha/2}\sqrt{\frac{pq}{n}}\right) = 1-\alpha$$

- As the population proportion *p* is unknown, the standard deviation of sample proportion is estimated by $\sqrt{\frac{\hat{p}\hat{q}}{n}}$
- 100(1-α)% confidence interval for p is

$$\hat{p}-z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \text{ or}$$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Standard Error of sample proportion

Sampling Error E

100(1-α)% Confidence Interval for Population Proportion *p*

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \text{ or}$$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Special cases
  - If $\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < 0$,

    we replace the lower bound of confidence interval by 0
  - If $\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} > 1$,

    we replace the upper bound of confidence interval by 1

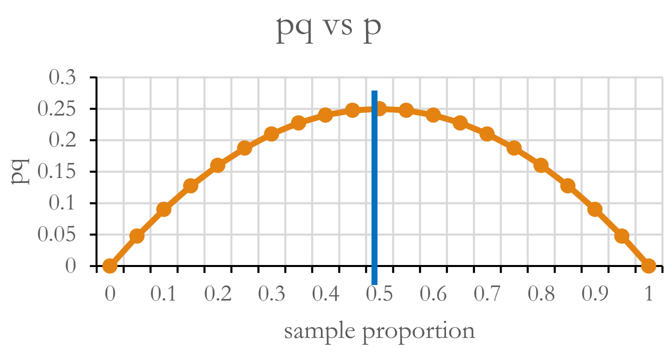
Factors Affecting Interval Width (Precision)

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \text{ to } \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Sample size
  - $n \uparrow \rightarrow \sqrt{\frac{\hat{p}\hat{q}}{n}} \downarrow \rightarrow$ width of interval $\downarrow$
- Level of confidence
  - Measured by $100(1-\alpha)\%$
  - $(1-\alpha)\uparrow \rightarrow |Z\text{-value}| \uparrow \rightarrow$ width of interval $\uparrow$

Confidence level	Confidence coefficient 1-α	$Z_{\alpha/2}$ value
90%	0.90	1.645
95%	0.95	1.96
99%	0.99	2.575

- Sample proportion $\hat{p}$
  - If $\hat{p}$ increases from 0 to 0.5, then $\hat{p}\hat{q}$ increases from 0 to 0.25, leading to a wider interval
  - If $\hat{p}$ further increases from 0.5 to 1, then $\hat{p}\hat{q}$ drops from 0.25 to 0, leading to a narrower interval



Example 1: Confidence Interval for Population Proportion

Among the 200 depositors you randomly selected, 95 of them have RMB deposit account at the bank. Find 95% confidence interval for the population proportion of depositors having RMB deposit account at the bank.

For these data, n=200, $\hat{p} = \frac{95}{200} = 0.475$

As $n\hat{p} = 95 > 5, n(1-\hat{p}) = 105 > 5$

→ The sampling distribution of $\hat{p}$ follows Normal distribution approximately

95% confidence interval for population proportion p is:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.475 \pm 1.96\sqrt{\frac{0.475(1-0.475)}{200}}$$

$$= (0.406, 0.544)$$

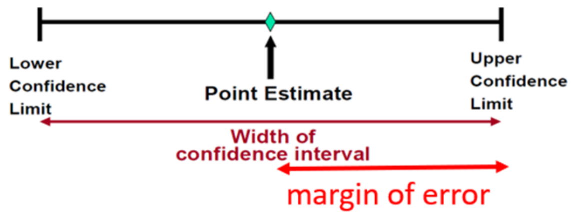
Confidence level	Confidence coefficient 1-α	$Z_{\alpha/2}$ value
90%	0.90	1.645
95%	0.95	1.96
99%	0.99	2.575

We are 95% confident that the population proportion of depositors having RMB deposit account is between 0.406 and 0.544

Part Two

- Confidence Interval Estimate for Population Proportion
- Sample Size Determination for Estimating Population Proportion
- Hypothesis Testing for Population Proportion

Determining Sample Size for Estimating Population Proportion



- What sample size is needed to be $100(1-\alpha)\%$ confident of being correct to within $\pm E$?
- Sampling error (or margin of error)

$$E = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

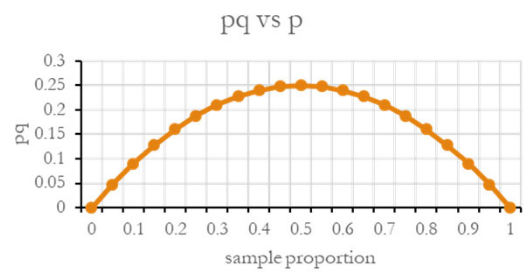
$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{pq}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right) = 1-\alpha$$

- Solving the equation for n gives

$$n = \frac{(z_{\alpha/2})^2 p(1-p)}{E^2}$$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- *p* is unknown. What can we do?
  - Method 1: Use prior information about *p*
  - Method 2: Use *p* = 0.5. When *p*=0.5, $p(1-p) = 0.25$, which is the largest value. It can fulfill the margin of error requirement for any true value of *p*
- If the computed n is not an integer, round it up to nearest integer



Example 2: Determining Sample Size for Population Proportion

- A product manager wants to estimate the proportion of customers who are likely to purchase a new product to within $\pm$ 0.04 with 95% confidence. What is the minimum sample size does he need?

Method 1: A pilot sample of size 100 was selected and the sample proportion was 0.62.

$$n = \frac{(z_{\alpha/2})^2 \hat{p}(1-\hat{p})}{E^2} = \frac{(1.96)^2 0.62(1-0.62)}{0.04^2}$$

$$= 565.68 \cong 566$$

Round Up

Confidence level	Confidence coefficient 1-α	$Z_{\alpha/2}$ value
90%	0.90	1.645
95%	0.95	1.96
99%	0.99	2.575

Method 2: Use $p = 0.5$

$$n = \frac{(z_{\alpha/2})^2 p(1-p)}{E^2} = \frac{(1.96)^2 0.5(1-0.5)}{0.04^2}$$

$$= 600.25 \cong 601$$

Round Up

Part Three

- Confidence Interval Estimate for Population Proportion
- Sample Size Determination for Estimating Population Proportion
- Hypothesis Testing for Population Proportion

Five-Step Hypothesis Testing Procedure

Step 1: State the null and alternative hypotheses

Step 2: Determine the test statistic (Z or t)

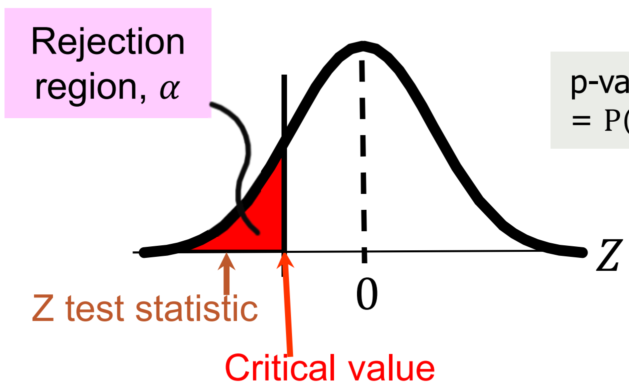
Step 3: Determine the rejection region based on the significance level

Step 4: Compute the value of the test statistic

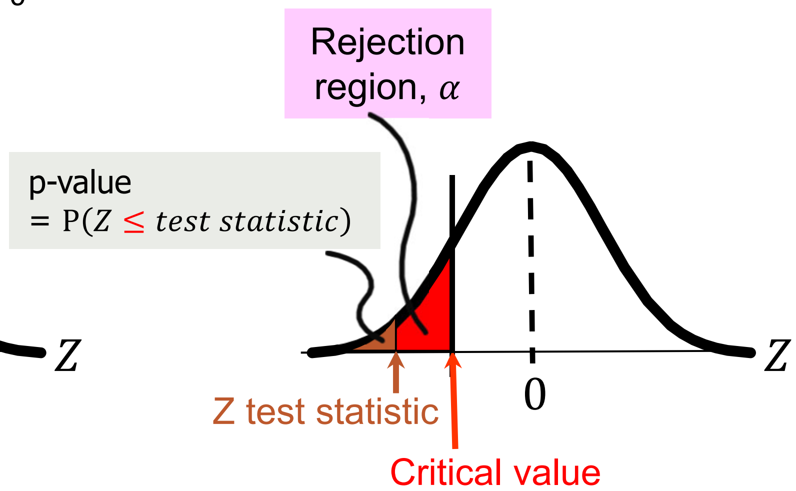
Step 5: Make statistical decision (Do not reject H_0 , Reject H_0) and give a conclusion in terms of the original problem

Critical Value and p-value Approach to Hypothesis Testing

- Critical value approach -- Compare the test statistic (Z or t) with critical value
  - If the test statistic falls in the rejection region, reject H_0
  - Otherwise, do not reject H_0
- *p* value approach -- Compare the p-value with the level of significance α
  - The p-value is the probability of obtaining a test statistic as extreme or more extreme ($\leq$ or $\geq$) than the observed test statistic value given H_0 is true
  - If p-value < α, reject H_0
  - Otherwise, do not reject H_0



Reject H_0 if Z test statistic falls in rejection region



Reject H_0 if p-value < α

Hypothesis Testing for Population Proportion *p*

- Steps of conducting hypothesis testing
  - **Step 1**: State the null and alternative hypothesis

Lower tail	Upper tail	Two tail
$H_0: p \geq p_o$ $H_1: p < p_o$	$H_0: p \leq p_o$ $H_1: p > p_o$	$H_0: p = p_o$ $H_1: p \neq p_o$

  - **Steps 2-3**: Determine test statistic, critical value(s) and rejection region based on α
  - **Step 4**: Collect sample data and compute test statistic and *p*-value assuming that H_0 is true
  - **Step 5**: Make statistical decision and conclusion

If H_0 is true $(p = p_0)$, $np_0 \geq 5$, $n(1 - p_0) \geq 5$,

$$\hat{p} \simeq N\left[p_0, \frac{p_0(1-p_0)}{n}\right]$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

reject H_0 if

Lower tail: $Z < -z_\alpha$

Upper tail: $Z > z_\alpha$

Two tail: $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$

Example 3: Hypothesis Testing for Population Proportion

- A bank had the business objective of serving 80% of the customers within 5 minutes upon their arrival. Of the 45 randomly selected customers, 39 are served within 5 minutes upon their arrival. Test the claim of the bank at 5% level of significance

Let p = population proportion of customers to be served within 5 mins upon their arrival

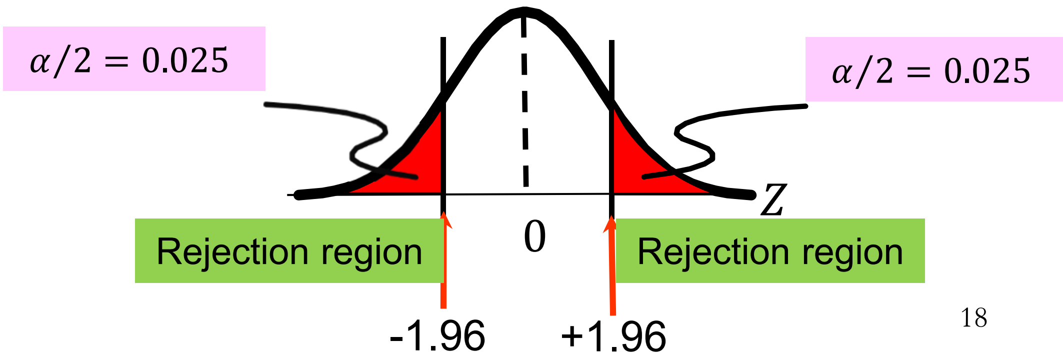
$H_0: p = 0.80$ (Step 1)
$H_1: p \neq 0.80$

(Steps 2-3)
$np_0 = 45 * 0.8 = 36 > 5$
$n(1 - p_0) = 45 * 0.2 = 9 > 5$
$\rightarrow$ sample proportion is approximately normally distributed

At $\alpha = 0.05$
Critical Value = ± 1.96
Reject H_0 if $Z < -1.96$ or $Z > +1.96$

Step 4: $$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{39}{45} - 0.8}{\sqrt{\frac{0.8(1-0.8)}{45}}} = 1.118$$

Step 5: As Z=1.118 < 1.96, do not reject H_0 at $\alpha = 0.05$.
There is insufficient evidence that the true proportion of customers to be served within 5 minutes upon their arrival is not 80%

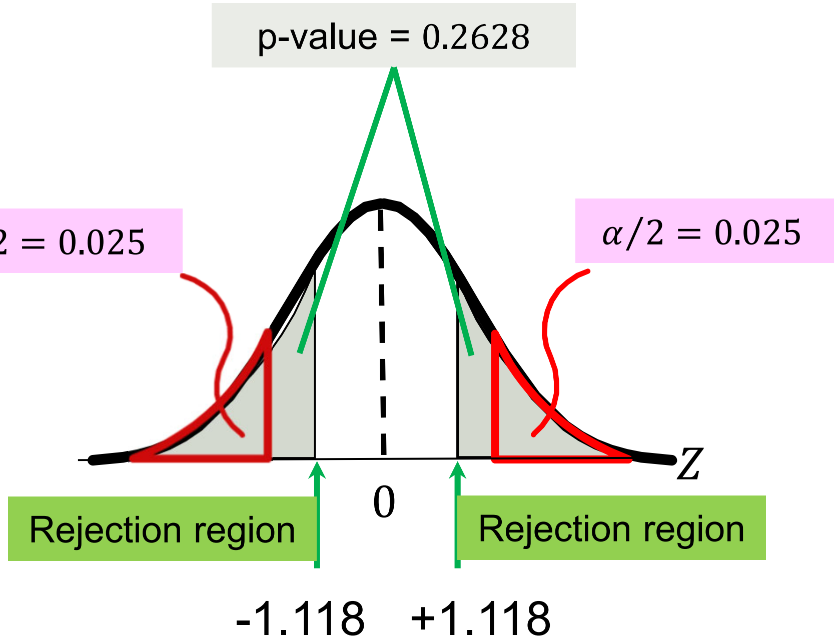


Example 3: Hypothesis Testing for Population Proportion

$H_0: p = 0.80$
$H_1: p \neq 0.80$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{39}{45} - 0.8}{\sqrt{\frac{0.8(1-0.8)}{45}}} = 1.118$$

p-value
$= P(Z \leq -1.118) + P(Z \geq 1.118)$
$= 2 \times P(Z \leq -1.118)$
$= 2 \times 0.1314$
$= 0.2628$



As p-value > α, do not reject H_0

There is insufficient evidence that the true proportion of customers to be served within 5 minutes upon their arrival is not 80%

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379

Connection of Two Tail Tests to Confidence Intervals

- If the hypothesized mean p_0 is in the CI → do not reject H_0
- If the hypothesized mean p_0 is not in the CI → reject H_0

Example	$\boldsymbol{p_0}$	$\boldsymbol{\alpha}$	P-value	Decision	$\mathbf{100(1-\alpha)\%}$ Confidence Interval (CI)
3	**0.80**	**0.05**	**0.2628**	**Do not reject H_0**	$\hat{p} = 39/45 = 0.8667$ 95% Confidence interval for p: $\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ $= 0.8667 \pm 1.96\sqrt{\dfrac{0.8667(1-0.8667)}{45}}$ $= (0.767, 0.966)$