Summary---Topic 7: Inference for the Proportion

Population Proportion(π) and Sample Proportion(p)

Number of success, Y ~ B(n, π),
→ Mean of binomial random variable: E[Y]=nπ
→ Variance of binomial random variable: Var(Y)=nπ (1- π)

- Population Proportion,

$$\pi = \frac{\text{no. of success in population (Y)}}{\text{population size (N)}}$$

- Sample Proportion,

$$p = \frac{\text{no. of success in sample (Y)}}{\text{sample size (n)}}$$

Sampling distribution, p

- Mean of sample proportion: $\mathrm{E}(p) = \mu_p = \pi$
- Variance of sample proportion: $\mathrm{Var}(p) = \sigma_p^2 = \frac{n\pi\,(1-\pi)}{n^2} = \frac{\pi\,(1-\pi)}{n}$

- If n ≥ 30, nπ ≥ 5, n(1-π) ≥ 5, then

$$p \sim \mathrm{N}(\mu_p, \sigma_p^2) = \mathrm{N}\left(\pi, \sqrt{\frac{\pi\,(1-\pi)}{n}}^{\,2}\right) \text{ approximately}$$

➔Standardization of sample proportion distribution

$$\mathrm{Z} = \frac{p - \mu_p}{\sigma_p} = \frac{p - \pi}{\sqrt{\frac{\pi\,(1-\pi)}{n}}}$$

Confidence Interval Estimation

Conditions: n ≥ 30, np ≥ 5, n(1- p) ≥ 5
∴Sampling distribution of *p* is **approximately** normal.

❑ Confidence interval of π: $p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

Since $\pi \in [0,1]$,

✓ If the lower boundary is lower than 0, replace the lower boundary by 0;

✓ If the upper boundary is higher than 1, replace the upper boundary by 1.

Determining Sample Size(n)

- Standard error= $\sqrt{\frac{\pi(1-\pi)}{n}}\,(\pi\ is\ known)\ or \sqrt{\frac{p(1-p)}{n}}\,(\pi\ is\ unknown) \geq 0$
- Sampling error $E = Z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\,(\pi\ is\ known)\ or\ Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\,(\pi\ is\ unknown)$

➔ $n = \frac{(Z_{\alpha/2})^2\pi(1-\pi)}{E^2}\,(\pi\ is\ known)\ or\ n = \frac{(Z_{\alpha/2})^2 p(1-p)}{E^2}\,(\pi\ is\ unknown)$

- If both π and p are unavailable, use π = 0.5 to calculate the sample size.
- If n is not an integer, remember to round it up to the nearest integer (not round down, as rounding down the sample size will cause the sampling error larger than the desired sampling error).

Hypothesis Test for Population Proportion----Critical Value Approach

Step 1: Define Hypotheses

Two-tail test	Lower-tail test	Upper-tail test
$H_0: \pi = \pi_0$ $H_1: \pi \neq \pi_0$	$H_0: \pi \geq \pi_0$ $H_1: \pi < \pi_0$	$H_0: \pi \leq \pi_0$ $H_1: \pi > \pi_0$

Step 2: Check Conditions and Determine Rejection Region

Conditions: 1) Number of successes follows binomial distribution;
2) $n \geq 30$, $np \geq 5$, $n(1- p) \geq 5$

∴Sampling distribution of *p* is approximately normal.

Reject H_0 if

Two-tail test	Lower-tail test	Upper-tail test
$Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$	$Z < -Z_{\alpha}$	$Z > Z_{\alpha}$

Step 3: Compute Test Statistic: $Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 (1-\pi_0)}{n}}}$

Step 4: Make Statistical Decision

✓If the Z test statistic value falls in the rejection region, reject H_0→ There is sufficient evidence that the H_1 is true;

✓Otherwise, do not reject H_0→ There is insufficient evidence that the H_1 is true.

P-value Approach

Step 1: Define Hypotheses

- Same as the one under the critical value approach

Step 2: Check Conditions and Choose Rejection Region

- Reject H_0 if p-value < α

Step 3: Compute Test Statistic and p-value

- Compute the test statistic $= \dfrac{p-\pi_0}{\sqrt{\dfrac{\pi_0\,(1-\pi_0)}{n}}}$

- Obtain the p-value

Two-tail test	Lower-tail test	Upper-tail test
$P(Z \leq -\lvert\text{test statistic}\rvert) + P(Z \geq \lvert\text{test statistic}\rvert)$	$P(Z \leq \text{test statistic})$	$P(Z \geq \text{test statistic})$

Step 4: Make statistical decision

✓ If p-value < α, then reject H_0 → There is sufficient evidence that the H_1 is true.

✓ if p-value ≥ α, then do not reject H_0 → There is insufficient evidence that the H_1 is true.

Exercises and Solutions

Q1. The following data represent the responses (Y for yes and N for no) from a sample of 40 college students to the question "Do you currently own shares in any stocks?"

N	N	Y	N	N	Y	N	N	N	Y	N	N	Y	N	N	Y	N	N	N	Y
N	N	N	N	N	N	N	N	N	N	Y	N	N	N	Y	N	N	N	N	N

a) Find the sample proportion of college students who own shares.

b) Find the standard error of the sample proportion of college students who own shares.

Solution:

a) Sample Proportion,

$$p = \frac{\text{no. of students who own shares of stock}}{\text{sample size (n)}} = \frac{8}{40} = 0.2.$$

b) Standard error (π is unknown)= $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.2(1-0.2)}{40}} = 0.0632.$

Q2. You plan to conduct a marketing experiment in which students are to taste

To compute non-standard normal probabilities(Topic 3):

1. Do standardization: from the non-standard normal distribution to standard normal distribution: $\boldsymbol{Z} = \frac{\boldsymbol{X}-\boldsymbol{\mu}}{\boldsymbol{\sigma}}$.
2. Check the Standard Normal Table.

a) What is the probability that the sample will have between 50% and 60% of the identifications correct?

Solution:

a) Let *p* be the sample proportion of students which can distinguish the brand. We are asked to compute **$P(0.5<p<0.6)=?$** Hence we need find its distribution first.

Check the conditions: If n ≥ 30, nπ ≥ 5, n(1-π) ≥ 5, then

$$p \sim \mathrm{N}\left(\mu_p, \sigma_p^2\right) = \mathrm{N}\left(\pi, \sqrt{\frac{\pi\,(1-\pi)}{n}}^{\,2}\right)$$

Since an individual has no ability to distinguish between the two brands, the proportion of students which can distinguish the brand = 0.5 (i.e., π=0.5).

Since n=200>30, nπ=100= n(1-π)>5, we have $p \sim \mathrm{N}\left(\mu_p, \sigma_p^2\right) = \mathrm{N}\left(0.5, \frac{0.5\,(1-0.5)}{200}\right)$

$$\mathrm{P}(0.5 < p < 0.6) = \mathrm{P}\left(\frac{0.5-\mu_p}{\sigma_p} < \frac{p-\mu_p}{\sigma_p} < \frac{0.6-\mu_p}{\sigma_p}\right) = \mathrm{P}\left(\frac{0.5-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{200}}} < Z < \frac{0.6-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{200}}}\right)$$

$$= \mathrm{P}(0 < Z < 2.83) = \mathrm{P}(Z < 2.83) - \mathrm{P}(Z < 0) = 0.9977 - 0.5 = 0.4977.$$

Q2. b) The probability is 90% that the sample percentage is contained within what symmetrical limits of the population percentage?

n=200; $\pi = 0.5$; $p \sim \mathrm{N}(\mu_p, \sigma_p^2) = \mathrm{N}\left(0.5, \frac{0.5\,(1-0.5)}{200}\right)$

Solution b)

90% of the values are between two values → $\mathrm{P}(\mathrm{a} < p < \mathrm{b}) = 0.9$

symmetrically distributed around the mean→

$$\begin{cases} \mathrm{P}(p < a) = 0.05 \\ \mathrm{P}(p < b) = 0.95 \end{cases}$$

Do the standardization, we have

$$\begin{cases} \mathrm{P}\left(\frac{p-\mu_p}{\sigma_p} < \frac{a-\mu_p}{\sigma_p}\right) = 0.05 \\ \mathrm{P}\left(\frac{p-\mu_p}{\sigma_p} < \frac{b-100}{\sigma_p}\right) = 0.95 \end{cases} \Rightarrow \begin{cases} \mathrm{P}\left(Z < \frac{a-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{200}}}\right) = 0.05 \\ \mathrm{P}\left(Z < \frac{b-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{200}}}\right) = 0.95 \end{cases}$$

$$\Rightarrow \begin{cases} \frac{a-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{200}}} = -1.645 \\ \frac{b-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{200}}} = 1.645 \end{cases} \Rightarrow \begin{cases} a = -1.645 \times \sqrt{\frac{0.5\,(1-0.5)}{200}} + 0.5 = 0.4418 \\ b = 1.645 \times \sqrt{\frac{0.5\,(1-0.5)}{200}} + 0.5 = 0.5582 \end{cases}$$

Q2. c) What is the probability that the sample percentage of correct identifications is greater than 65%?

n=200; $\pi = 0.5$; $p \sim \mathrm{N}(\mu_p, \sigma_p^2) = \mathrm{N}\left(0.5, \frac{0.5\,(1-0.5)}{\mathbf{200}}\right)$

Solution c)

$$\mathrm{P}(p > 0.65) = \mathrm{P}\left(\frac{p-\mu_p}{\sigma_p} > \frac{0.65-\mu_p}{\sigma_p}\right) = \mathrm{P}\left(Z > \frac{0.65-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{200}}}\right)$$

$$= \mathrm{P}(Z > 4.24) = 1 - \mathrm{P}(Z < 4.24) = 1 - 0.9999888 \approx 0.$$

Q2. d) Which is more likely to occur – more than 60% correct identifications in the sample of 200 or more than 55% correct identifications in a sample of 1,000? Explain.

n=1000; $\pi = 0.5$; $p \sim \mathrm{N}(\mu_p, \sigma_p^2) = \mathrm{N}\left(0.5, \frac{0.5\,(1-0.5)}{\mathbf{1000}}\right)$

Solution d)

When n=200 → $\mathrm{P}(p > 0.6) = \mathrm{P}\left(\frac{p-\mu_p}{\sigma_p} > \frac{0.6-\mu_p}{\sigma_p}\right) = \mathrm{P}\left(Z > \frac{0.6-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{200}}}\right)$

$$= \mathrm{P}(Z > 2.83) = 1 - \mathrm{P}(Z < 2.83) = 1 - 0.9977 = 0.0023.$$

When n=1000 → $\mathrm{P}(p > 0.55) = \mathrm{P}\left(\frac{p-\mu_p}{\sigma_p} > \frac{0.55-\mu_p}{\sigma_p}\right) = \mathrm{P}\left(Z > \frac{0.55-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{1000}}}\right)$

$$= \mathrm{P}(Z > 3.16) = 1 - \mathrm{P}(Z < 3.16) = 1 - 0.99921 = 0.00079.$$

More than 60% correct identification in a sample of 200 is more likely to occur than more than 55% correct in a sample of 1000.

Q3. In a survey conducted for American Express, 27% of small business owners indicated that they never check in with the office when on vacation. The sample size used in the study was not disclosed.

a) Suppose that the survey was based on 500 small business owners. Construct a 95% confidence interval estimate for the population proportion of small business owners who never check in with the office when on vacation.

Solution a): *p*=0.27

Check the conditions: n =500 > 30, np =500*0.27=135>5, n(1-p)=365> 5, then the sampling distribution of *p* is approximately normal.

95% confidence interval for π:

$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.27 \pm Z_{(1-0.95)/2}\sqrt{\frac{0.27(1-0.27)}{500}} = [0.2311, 0.3089]$$

We are 95% confident that the population proportion of small business owners who never check in with the office when on vacation is estimated to be between 0.2311 and 0.3089.

Q3. In a survey conducted for American Express, 27% of small business owners indicated that they never check in with the office when on vacation. The sample size used in the study was not disclosed.

b) Suppose that the survey was based on 1,000 small business owners. Construct a 95% confidence interval estimate for the population proportion of small business owners who never check in with the office when on vacation.

c) Discuss the effect of sample size on the confidence interval estimate.

Solution b):

Check the condition: n =1000 > 30, np =1000*0.27=270>5, n(1-p)=730> 5, hence the sampling distribution of *p* is approximately normal.

95% confidence interval for π:

$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.27 \pm Z_{(1-0.95)/2}\sqrt{\frac{0.27(1-0.27)}{1000}} = [0.2425{,}0.2975].$$

We are 95% confident that the population proportion of small business owners who never check in with the office when on vacation is estimated to be between 0.2425 and 0.2975.

Solution c):

The larger the sample size, (the sampling error↓) the narrower is the confidence interval holding everything constant.

Q4*. A study of 658 CEOs conducted by the Conference Board reported that 250 stated that their company's greatest concern was sustained and steady top-line growth.

a) Construct a 95% confidence interval for the proportion of CEOs whose greatest concern was sustained and steady top-line growth.

b) To conduct a follow-up study to estimate the population proportion of CEOs whose greatest concern was sustained and steady top-line growth to within 0.01 with 95% confidence, how many CEOs would you survey?

Solution a): Let *p* be the sample proportion of CEOs whose greatest concern was sustained and steady top-line growth. Then *p*=250/658=0.3799.

Check the condition: n =658 > 30, np =658*0.3799=250>5, n(1-p)=408> 5, then the sampling distribution of *p* is approximately normal.

95% confidence interval for π:

$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.3799 \pm Z_{(1-0.95)/2}\sqrt{\frac{0.3799(1-0.3799)}{658}} = [0.3429, 0.4170].$$

We are 95% confident that the population proportion of CEOs whose greatest concern was sustained and steady top-line growth is estimated to be between 0.3429 and 0.4170.

Solution b): sampling error E=0.01

$$n = \frac{\left(Z_{\alpha/2}\right)^2 p(1-p)}{\mathrm{E}^2} = \frac{{Z_{0.05/2}}^2 \times 0.3799 \times (1-0.3799)}{0.01^2} = \frac{1.96^2 \times (0.3799) \times (1-0.3799)}{0.01^2}$$
$$= 9049.92 \approx 9050 \text{ (round up)}$$

Q5. One of the biggest issues facing e-retailers is the ability to reduce the proportion of customers who cancel their transactions after they have selected their products. It has been estimated that about half of prospective customers cancel their transactions after they have selected their products. Suppose that a company changed its web site so that customers could use a single-page checkout process rather than multiple pages. A sample of 500 customers who had selected their products was provided with the new checkout system. Of these 500 customers, 210 cancelled their transactions after they had selected their products.

a) At the 0.01 level of significance, is there evidence that the population proportion of customers who select products and then cancel their transaction is less than 0.50 with the new system?

Solution:

a) Let π be the population proportion of customers who select products and then cancel their transaction.

$$H_0: \pi \geq 0.5$$
$$H_1: \pi < 0.5$$

Since $n = 500 > 30; np = 500 \times \left(\frac{210}{500}\right) = 500 \times 0.42 = 210 > 5;\ n(1-p) = 290 > 5$, $p \sim N$ approximately and we should use Z-test (lower tail).

At $\alpha = 0.01$, Reject H_0 if $Z < -Z_{0.01} = -2.33$

Compute the Z-statistic $Z = \frac{p-\pi_0}{\sqrt{\frac{\pi_0\,(1-\pi_0)}{n}}} = \frac{0.42-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{500}}} = -3.577$ ➔ reject H_0

There is sufficient evidence to conclude that the proportion of customers who selected products and then cancelled their transaction is less than 0.50 with the new system.

Q5. b) Suppose that a sample of n=100 customers (instead of n=500 customers) were provided with the new checkout system and that 42 of those customers cancelled their transactions after they had selected their products. At the 0.01 level of significance, is there evidence that the population proportion of customers who select products and then cancel their transaction is less than 0.50 with the new system?

c) Compare the results of (a) and (b) and discuss the effect that sample size has on the outcome, and, in general, in hypothesis testing.

Solution b): *n=100; p=42/100=0.42*

Since $n = 100 > 30; np = 42 > 5; \; n(1-p) = 58 > 5, \;\; p \sim N$ approximately and we should use Z-test (lower tail).

At $\alpha = 0.01$, Reject H_0 if $Z < -Z_{0.01} = -2.33$

Compute the Z-statistic $Z = \dfrac{p-\pi_0}{\sqrt{\dfrac{\pi_0\,(1-\pi_0)}{n}}} = \dfrac{0.42-0.5}{\sqrt{\dfrac{0.5\,(1-0.5)}{100}}} = -1.6 \Rightarrow$ do not reject H_0

There is insufficient evidence to conclude that the proportion of customers who selected products and then cancelled their transaction is less than 0.50 with the new system.

Solution c): The larger the sample size, the smaller is the standard error. Even though the sample proportion is the same value at 0.42 in (a) and (b), the test statistic is more negative while the p-value is smaller in (a) compared to (b) because of the larger sample size in (a).

Q6. a) Grant, Inc., a manufacturer of women's dress blouses, knows that its brand is carried in 19 percent of the women's clothing stores in Hong Kong. Grant recently sampled 85 women's clothing stores in mainland China and found that 14 percent of the stores carried the brand.

(i)* At the 0.05 level of significance, is there evidence that Grant has poorer distribution in mainland China than it does in Hong Kong?

(ii)* Interpret the decision you made in (i) in the situation being examined.

Solution:

(i) Let π be the population proportion of stores carried the brand

$$H_0: \pi \geq 0.19$$
$$H_1: \pi < 0.19$$

Since $n = 85 > 30; np = 85 \times 0.14 = 11.9 > 5;\ n(1-p) = 73.1 > 5,\ \ p \sim N$ approximately and we should use Z-test (lower tail).

At $\alpha = 0.05$, Reject H_0 if $Z < -Z_{0.05} = -1.645$

Compute the Z-statistic $\mathrm{Z} = \dfrac{p-\pi_0}{\sqrt{\dfrac{\pi_0\,(1-\pi_0)}{n}}} = \dfrac{0.14-0.19}{\sqrt{\dfrac{0.19(1-0.19)}{85}}} = -1.1751$ ➔do not reject H_0

(ii)There is insufficient evidence that Grant has poorer distribution in Mainland China than it does in Hong Kong.

Q6. a)

(iii) What is the p-value in (i)? Do you make the same decision as in (i) at α = 0.05 if you use the p-value approach?

(iv) Suppose that the sample size n = 85 is fixed and further suppose that the penalty of committing type II is serious, which α value (0.05 or 0.10) do you choose for the hypothesis testing? Briefly explain your choice.

Solution (iii)

p-value $= \mathrm{P(Z} \leq -1.1751) = 0.1190 > \alpha = 0.05$ ➔ do not reject H_0

We do not reject H_0 and making the same decision as in (i).

Solution (iv)

Choose $\alpha = 0.1$

For a fixed sample size, a larger value of α would correspond to a smaller value of β, that can decrease the penalty of committing type II error.

Q6. b) From past records, a charity has found that 42% of donors in a year will denote again in the next year. A random sample of 300 donors from last year was taken.

(i)* What is the standard error of the sample proportion who will donate again this year?

(ii)* What is the probability that the sample proportion is between 0.40 and 0.45?

(iii) Without doing the calculations, state in which of the following ranges the sample proportion is more likely to lie: 0.39 to 0.41, 0.41 to 0.43, 0.43 to 0.45.

(iv)* Interpret the answer obtained in (ii).

Solution i): standard error $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} (\pi \textit{ is known}) = \sqrt{\frac{0.42(1-0.42)}{300}} = 0.0285$

Solution ii):

Since $n = 300 > 30; np = 300 \times 0.42 = 126 > 5;\ n(1-p) = 174 > 5,\ \ p \sim N$ approximately.

$$\mathrm{P}(0.4 < p < 0.45) = \mathrm{P}\left(\frac{0.4-\mu_p}{\sigma_p} < \frac{p-\mu_p}{\sigma_p} < \frac{0.45-\mu_p}{\sigma_p}\right) = \mathrm{P}\left(\frac{0.4-0.42}{0.0285} < Z < \frac{0.45-0.42}{0.0285}\right)$$
$$= \mathrm{P}(-0.7 < Z < 1.05) = \mathrm{P}(Z < 1.05) - \mathrm{P}(Z < -0.7)$$
$$= 0.8531 - 0.242 = 0.6111.$$

Solution iii): The range of 0.41 to 0.43 is more likely to lie because this range contains the population proportion that is 0.42.

Solution iv): In the sample size 300, 61.11% of sample will be expected to have the sample proportions between 0.4 and 0.45.

Q7. a)*In order to estimate the unemployment rate of Hong Kong, a random sample of 8500 people was selected in 2002 and 618 people were found to be unemployed. Find a 95% confidence interval for the unemployment rate of Hong Kong in 2002. Give your answer to the fourth decimal place.

Solution:

a) Let *p* be the proportion of unemployment rate of Hong Kong in 2002, then $p = \frac{618}{8500}$.

Since $n = 8500 > 30; np = 618 > 5;\ n(1-p) = 7882 > 5,\ \ p \sim N$ approximately.

95% confidence interval for π:

$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.0727 \pm Z_{(1-0.95)/2}\sqrt{\frac{\frac{618}{8500}\times(1-\frac{618}{8500})}{8500}} = [0.0672{,}0.0782].$$

We are 95% confident that the population proportion of Hong Kong unemployment rate is estimated to be between 0.0672 and 0.0782.

Q7. b) If you want to be 95% confident of estimating the unemployment rate of Hong Kong in part (a) to within ±0.2%, what sample size is needed?

Solution b): the sampling error E=0.2%=0.002

- Sampling error $\mathrm{E} = Z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\,(\pi\ is\ known)$

  $or\ \boldsymbol{E} = \boldsymbol{Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\,(\pi\ is\ unknown)}$

$$n = \frac{\left(Z_{\alpha/2}\right)^2 p(1-p)}{\mathrm{E}^2} = \frac{Z_{0.05/2}{}^2 \times \frac{618}{8500} \times (1-\frac{618}{8500})}{0.002^2} = \frac{1.96^2 \times \frac{618}{8500} \times (1-\frac{618}{8500})}{0.002^2}$$

$= 64749.91544 \approx 64750$ (round up)

Q7. c)* According to the report given by the Census and Statistics Department of Hong Kong, the actual unemployment rate of Hong Kong in 2002 is 7.3%. In a survey of 620 people in Shatin, 34 people were found to be unemployed. Is there evidence that the Shatin unemployment rate was lower than the Hong Kong unemployment rate at the 0.05 level of significance?

Solution c):

Let π be the proportion of unemployment rate of Shatin in 2002

$$H_0: \pi \geq 0.073$$
$$H_1: \pi < 0.073$$

Since $n = 620 > 30; np = 34 > 5;\ n(1-p) = 586 > 5,\ \ p \sim N$ approximately and we should use Z-test (lower tail).

At $\alpha = 0.05$, Reject H_0 if $Z < -Z_{0.05} = -1.645$

Compute the Z-statistic $\mathrm{Z} = \dfrac{p-\pi_0}{\sqrt{\dfrac{\pi_0\,(1-\pi_0)}{n}}} = \dfrac{\frac{34}{620}-0.073}{\sqrt{\dfrac{0.073(1-0.073)}{620}}} = -1.738$ ➔ reject H_0

There is sufficient evidence that the unemployment in Shatin is lower than Hong Kong in 2002.