Tutorial 10

CB2200 Business Statistics, YANG Yihang

November 17, 2020

Intended Learning Outcomes

After this tutorial, you may know how to

- identify the sample proportion distribution,
- construct confidence interval of the mean of proportion,
- implement hypothesis test based on the population 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,

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

Recap

- 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

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

➔Standardization of sample proportion distribution

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

Question 1

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.

Question 1

Q1

a) $$\text{Sample proportion, p} = \frac{\text{no. of students who own shares of stock}}{\text{sample size(n)}}$$

$$= 8/40 = 0.2$$

b) Standard error of the proportion, $\sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.2(1-0.2)}{40}} = 0.0632$

Question 2

Q2

You plan to conduct a marketing experiment in which students are to taste one of two different brands of soft drink. Their task is to correctly identify the brand tasted. You select a random sample of 200 students and assume that the students have no ability to distinguish between the two brands. (Hint: If an individual has no ability to distinguish between the two soft drinks, then each brand is equally likely to be selected.)

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

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

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

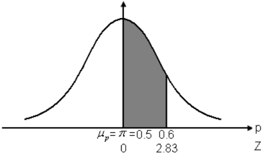
d) Which is more likely to occur – more than 60% correct identifications in the sample of 200 or more than 55% correct identifications in a sample of 1,000? Explain.

Q2

a) $\because$ An individual has no ability to distinguish between the two brands.
$\therefore$ π = proportion of students which can distinguish the brand = 0.5

n = 200 > 30, nπ=200(0.5)= 100 > 5, n(1-π)=200(1-0.5)=100 > 5
So sampling distribution of p is approximately normal

$$p \sim N(\pi, \frac{\pi(1-\pi)}{n}) \Rightarrow p \sim N(0.5, \frac{0.5(1-0.5)}{200})$$



$$P(0.5 < p < 0.6) = P(\frac{0.5-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} < Z < \frac{0.6-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}})$$

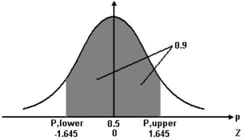
$$= P(\frac{0.5-0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}} < Z < \frac{0.6-0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}})$$

$$= P(0 < Z < 2.83)$$

$$= 0.9977 - 0.5$$

$$= 0.4977$$

b) $P(p_{lower} < p < p_{upper}) = 0.90$



$$P\left(\frac{p_{lower} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}} < Z < \frac{p_{upper} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}}\right) = 0.90$$

For symmetric distribution of probability (0.9) on both sides,

$$P\left(Z < \frac{p_{lower} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}}\right) = 0.05 \text{ and } P\left(Z < \frac{p_{upper} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}}\right) = 0.95$$

Since $P(Z < -1.645) = 0.05$ and $P(Z < 1.645) = 0.95$,

$$\begin{cases} \dfrac{p_{lower} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}} = -1.645 \\ \dfrac{p_{upper} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}} = 1.645 \end{cases} \qquad \begin{cases} p_{lower} = -1.645\left(\sqrt{\frac{0.5(1-0.5)}{200}}\right) + 0.5 = 0.4418 \\ p_{upper} = 1.645\left(\sqrt{\frac{0.5(1-0.5)}{200}}\right) + 0.5 = 0.5582 \end{cases}$$

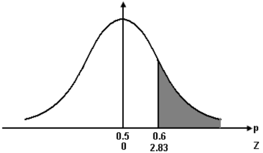
Hence, 90% of the sample proportion will be between 0.4418 and 0.5582.

Question 2

c) $P(p > 0.65) = P(Z > \dfrac{0.65 - 0.5}{\sqrt{\dfrac{0.5(1-0.5)}{200}}}) = P(Z > 4.24) \approx 0$

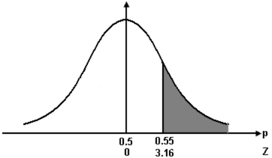
d) Case 1: $P(p > 0.6) = P(Z > \dfrac{0.6 - 0.5}{\sqrt{\dfrac{0.5(1-0.5)}{200}}})$

$= P(Z > 2.83)$
$= 1\text{- } 0.9977$
$= 0.0023$



Case 2: $P(p > 0.55) = P(Z > \dfrac{0.55 - 0.5}{\sqrt{\dfrac{0.5(1-0.5)}{1000}}})$

$= P(Z > 3.16)$
$= 1\text{- } 0.99921$
$= 0.00079$



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.

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.

Question 3

Q3

a) n=500>30, np=500(0.27) =135 >5, n(1-p)=500(1-0.27)=365>5

$\therefore$ Sampling distribution of p is approximately normal.

$p = 0.27$, $\alpha = 0.05$, $Z_{0.05/2} = 1.96$

95% confidence interval for π:

$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.27 \pm 1.96\sqrt{\frac{0.27(1-0.27)}{500}} = [0.2311, 0.3089]$$

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.

b) n=1000>30, np=1000(0.27) =270 >5, n(1-p)=1000(1-0.27)=730 >5

$\therefore$ sampling distribution of p is approximately normal.

$p = 0.27$, $\alpha = 0.05$, $Z_{0.05/2} = 1.96$

95% confidence interval for π:

$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.27 \pm 1.96\sqrt{\frac{0.27(1-0.27)}{1000}} = [0.2425, 0.2975]$$

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.

c) The larger the sample size, 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?

Question 4

Q4

a) n=658>30, np=658(0.3799) =250 >5, n(1-p)= 658(1-0. 3799)=408 >5

$\therefore$ sampling distribution of p is approximately normal.

$$p = \frac{X}{n} = \frac{250}{658} = 0.3799, \ \alpha = 0.05, \ Z_{0.05/2} = 1.96$$

95% confidence interval for π:

$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.3799 \pm 1.96\sqrt{\frac{0.3799(1-0.3799)}{658}} = [0.3429, 0.4170]$$

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.

b) Replace π by p

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

Q5

A bank recently has conducted a survey to explore differences in banking behavior between customers with e-banking account and customer without e-banking account. Suppose the survey was based on 200 respondents who have e-banking account and 400 respondents who don't have e-banking account. The numbers of respondents who visited the bank in person last month are as follows:

a)

	Service	With e-banking account	Without e-banking account
Visited the bank in person last month	Yes	102	288
	No	98	112
	Total	200	400

Suppose the management of the bank plans to conduct another survey. What sample size is needed to estimate the population proportion of customers without e-banking account that visited the bank in person last month to within ± 0.03 with 95% confidence? What sample size is needed if the information from the previous survey is not available?

Question 5

Q5

$$n = \frac{Z_{\frac{\alpha}{2}}^{\;2}\, p(1-p)}{E^2} = \frac{(1.96)^2 (\frac{288}{400})(\frac{112}{400})}{(0.03)^2} = 860.5 \cong 861$$

If information is not available, let p=0.5

$$n' = \frac{(1.96)^2 (0.5)(0.5)}{(0.03)^2} = 1067.1 \cong 1068$$

Question 6

Q6

The manager of Pizza Delight claims that at least 95% of its orders are delivered within 15 minutes of the time the order is place. A random sample of 150 orders revealed that 138 were delivered within the promised time.

a) Test the manager's claim at the 0.1 level of significance by using a four-step p-value approach to hypothesis testing.

b) Use the p-value obtained in part (a) to set up a decision rule for testing the manager's claim at other significance levels.

Question 6

Q6

a) $H_0: \pi \geq 0.95$

$H_1: \pi < 0.95$

$$Z = \frac{0.92 - 0.95}{\sqrt{\frac{(0.95)(0.05)}{150}}} = -1.69$$

$\therefore$ p-value=$P(Z \leq -1.69)$=0.0455

Since p-value=0.0455 $< 0.1 = \alpha$
Ho is rejected, and we conclude that there is not sufficient evidence to support that manager's claim. i.e. Less than 95% of Pizza Delight's orders are delivered within 15 minutes of the time when order is placed at the 0.1 level of significance.

b) The decision Rule is given as follows:
Ho is rejected if the given significance level $\alpha > 0.0455$;
Ho is not rejected if the given significance level $\alpha \leq 0.0455$

Q7

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?

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.

Question 7

Q7

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: $\pi < 0.5$

$\because$ n =500> 30, np=210>5, n(1-p)=290>5

$\therefore$ Sampling distribution of p is approximately normal

Reject H_0 if $Z < -2.33$

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{210/500 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{500}}} = -3.577$$

Since Z =-3.577 < -2.33

Reject H_0 at $\alpha = 0.01$

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.

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

H_0: $\pi \geq 0.5$

H_1: $\pi < 0.5$

$\because$ n =100> 30, np=42>5, n(1-p)=58>5

$\therefore$ Sampling distribution of p is approximately normal

$\therefore$ Z-test (Lower Tail)

With $\alpha = 0.01$, $-Z_{0.01} = -2.33$

Reject H_0 if $Z < -2.33$

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{42/100 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}} = -1.6$$

Since Z =-1.6 > -2.33

Do not Reject H_0 at $\alpha = 0.01$

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.

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).