Tutorial 8

CB2200 Business Statistics, YANG Yihang

November 4, 2019

Intended Learning Outcomes

After this tutorial, you may know how to

- Calculate appropriate sample size for a specific confidence interval.
- Construct a hypothesis test.
- Compute p-value and interpret it.

Q4

An advertising agency that serves a major radio station wants to estimate the mean amount of time that the station's audience spends listening to the radio on a daily basis. From past studies, the standard deviation is estimated as 45 minutes. Assume the population has the normal distribution.

a) What sample size is needed if the agency wants to be 90% confident of being correct to within $\pm$ 10 minutes?

b) If 99% confidence is desired, how many listeners need to be selected?

Q4

a) Since population distribution is normal, the sampling distribution of $\bar{X}$ is also normal.
As σ is known, <u>Z distribution</u> is used

$\sigma = 45, \quad \alpha = 1 - 0.9 = 0.1, \quad Z_{0.1/2} = 1.645$

"Correct to within ± 10 minutes" means sampling error E = 10.

$$n = \frac{(1.645)^2 (45)^2}{(10)^2} = 54.797 \approx 55 \ (\textit{round up})$$

$\therefore$ The agency needs at least 55 observations in order to be 90% confident that the error of estimation is within ± 10 minutes.

b) $$n = \frac{(2.575)^2 (45)^2}{(10)^2} = 134.2702 \approx 135 \ (\textit{round up})$$

$\therefore$ The agency needs at least 135 observations in order to be 99% confident that the error of estimation is within ± 10 minutes.

Q5

The personal director of a large corporation wants to study absenteeism among clerical workers at the corporation's central office during the year. A random sample of 25 clerical workers reveals the following:

Absenteeism: $\overline{X}$=9.7 days, s=4.0 days

a) Set up a 95% confidence interval estimate of the average number of absences for clerical workers. Give a practical interpretation of the interval obtained.
b) What assumption must hold in order to perform the estimation in (a)?
c) If the personnel director also wants to take a survey in a branch office, what sample size is needed if the director wants to be 95% confident of being correct to within ±1.5 days and the population standard deviation is assumed to be 4.5 days?

Q5

a) Assume that population is normal,

$\because$ n = 25 < 30, σ unknown, use t-distribution

$$95\%\text{ CI } = \overline{X} \pm t_{\frac{\alpha}{2},n-1} \frac{s}{\sqrt{n}} = 9.7 \pm (2.0639)\frac{4}{\sqrt{25}} = [8.0489,\ 11.3511]$$

We are 95% confident that the average number of absences is between 8.049 and 11.351 days.

b) $\because$ n = 25 < 30

$\therefore$ Assume that the number of absences is normally distributed, so that the sampling distribution of $\overline{X}$ is normal

c) $\alpha = 0.05$,

$$n = \frac{Z_{\frac{\alpha}{2}}^{2}\sigma^{2}}{E^{2}} = \frac{(1.96)^{2}(4.5)^{2}}{(1.5)^{2}} = 34.57 \approx 35$$

Q8

From past experience, the numbers of vitamin supplements sold per day in a health food store well fit a normal distribution with variance 9. The number of vitamin supplements sold per day in a sample of 11 days is obtained:

19	19	20	20	20	22	23	25	26	27	30

a) Construct a 95% confidence interval for population average number of vitamin supplements sold per day in the store.

b) Someone suggests constructing the confidence interval by t-distribution. Do you agree? Explain your opinion.

c) The data analyst found that there is a data-entry mistake. The last observation should be smaller than 30. How will this affect the confidence interval?

d) If the company would like to limit the sampling error within ± 0.5, what is the least sample size needed for a 90% confidence interval?

Topic 5-Question 8

Q8

a) Since the population is a normal distribution with known standard deviation (=3), normal distribution is used to construct the confidence interval.
For 95% confidence, α=0.05, $Z_{\alpha/2}$=1.96
$\overline{X} = 22.8182$ $\sigma = \sqrt{9}$=3
95%C.I. = [22.8182 – 1.96*3/sqrt(11) , 22.8182 + 1.96*3/sqrt(11)]
= [22.8182 – 1.7729, 22.8182+1.7729]
= [21.0453, 24.5911]
We are 95% confident that the population average number of vitamin supplements sold per day in the store is between 21.0453 and 24.5911

b) No, although the sample size is small, the sampling distribution follows normal distribution as the population distribution is normal. In addition, the population variance (standard deviation) is given, normal distribution, instead of t-distribution, should be used to construct the confidence interval.

c) The new observation is smaller, the new sample mean will be smaller as well. The confidence interval will shift to left with center at the new sample mean. However, the width of the interval will not be changed as the width of the confidence interval is related to level of confidence (α), population standard deviation (σ) and sample size (n). While these three values keep unchanged, the width of interval will not change.

d) 90% C.I. $Z_{\alpha/2}$=1.645
E = 0.5, σ=3
Sample size, n >= $(Z_{\alpha/2} * \sigma / E)^2 = (1.645*3/0.5)^2 = (9.87)^2$ =97.4169 $\cong$ 98 (round up)

Recap

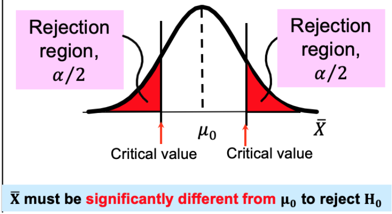
Step 1: Define hypotheses

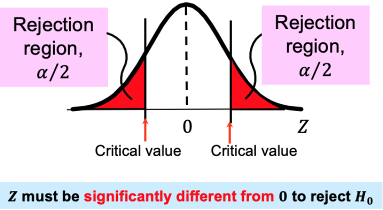
Step 2: Collect the data and identify the rejection region(s)

Step 3: Compute test statistic

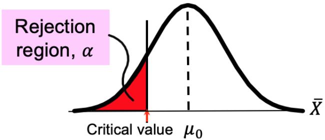
Step 4: Make statistical decision

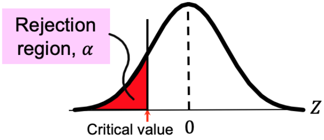
- For **two-tail** test: $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$





- For **lower-tail** test: $H_0: \mu \geq \mu_0$ against $H_1: \mu < \mu_0$

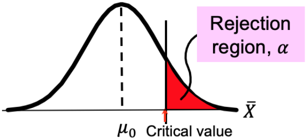


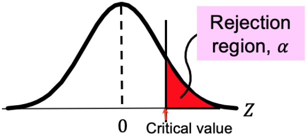


$\bar{X}$ **must be significantly smaller than** μ_0 **to reject** H_0

Z **must be significantly smaller than 0 to reject** H_0

- For **upper-tail** test: $H_0: \mu \leq \mu_0$ against $H_1: \mu > \mu_0$





$\bar{X}$ **must be significantly larger than** μ_0 **to reject** H_0

Z **must be significantly larger than 0 to reject** H_0

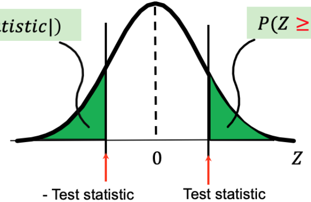
15

- For **two-tail** test: $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$

p-value = $P(Z \leq -|test\ statistic|) + P(Z \geq |test\ statistic|)$

$P(Z \leq -|test\ statistic|)$

$P(Z \geq |test\ statistic|)$



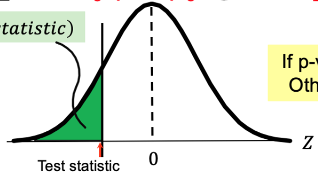
If p-value $< \alpha$, then reject H_0
Otherwise, do not reject H_0

Recap

Cont'd

- For **lower-tail** test: $H_0: \mu \geq \mu_0$ against $H_1: \mu < \mu_0$

p-value = $P(Z \leq test\ statistic)$

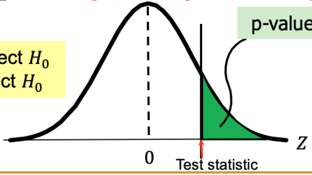


If p-value $< \alpha$, then reject H_0
Otherwise, do not reject H_0

- For **upper-tail** test: $H_0: \mu \leq \mu_0$ against $H_1: \mu > \mu_0$

p-value = $P(Z \geq test\ statistic)$

If p-value $< \alpha$, then reject H_0
Otherwise, do not reject H_0



Q1

Do students at your school study more, less, or about the same as at other business schools? Business Week reported that at the top 50 business schools, students studied an average of 14.6 hours. Set up a hypothesis test to try to prove that the mean number of hours studied at your school is different from the 14.6 hour benchmark reported by Business Week.

a) State the null and alternative hypotheses.
b) What is a Type I error for your test?
c) What is a Type II error for your test?

Q1

a) H_0: $\mu = 14.6$ hour

H_1: $\mu \neq 14.6$ hour

b) A Type I error is the mistake of **concluding** that the mean number of hours studied at your school is **different** from the 14.6 hour benchmark reported by Business Week when **in fact** it is not any different.
(Type I Error occurs if reject the H_0 when it is true)

c) A Type II error is the mistake of **not concluding** that the mean number of hours studied at your school is **different** from the 14.6 hour benchmark reported by Business Week when it is **in fact** different.
(Type II Error occurs if do not reject the H_0 when it is false)

Q2

ATMs must be stocked with enough cash to satisfy customers making withdrawals over an entire weekend. But if too much cash is unnecessarily kept in the ATMs, the bank is forgoing the opportunity of investing the money and earning interest. Suppose that at a particular branch the population mean amount of money withdrawn from ATMs per customer transaction over the weekend is \$160 with a population standard deviation of \$30.

a) If a random sample of 36 customer transactions is examined and the sample mean withdrawal is \$148, is there evidence to believe that the population average withdrawal is less than \$160? (Use a 0.05 level of significance.)

b) Compute the p-value and interpret its meaning.

Q2

a)

$H_0: \mu \geq 160$

$H_1: \mu < 160$

Let μ be the population mean of withdrawal

Population distribution unknown, $\because$ n = 36> 30, by Central Limit Theorem, the sampling distribution of $\overline{X}$ is approximately normal

σ known (σ=30) $\therefore$ Z test can be used (Lower-tail test)

At $\alpha = 0.05$, Critical value = -Z = -1.645

Reject H_0 if Z < -1.645

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{148 - 160}{30 / \sqrt{36}} = -2.4$$

Since Z = -2.4 < -1.645, Reject H_0 at $\alpha = 0.05$

There is sufficient evidence that the population mean amount of money withdrawn from ATMs per customer transaction is less than $160.

b) p -value = $P(Z \leq -2.4) = 0.0082$
Probability of obtaining a test statistic -2.4 or less is 0.0082, given H_0 is true.