Introductory Statistics for Economics

ECON1013: LAB 3

Duong Trinh

University of Glasgow

Feb 2024

Intro

- Duong Trinh
  - PhD Student in Economics (Bayesian Microeconometrics)
  - Email: Duong.Trinh@glasgow.ac.uk

- ECON1013-LB04
  - Monday 1-2 pm
  - 3 sessions (29-Jan, 12-Feb, 26-Feb)
- ECON1013-LB05
  - Tuesday 12-1 pm
  - 3 sessions (30-Jan, 13-Feb, 27-Feb)
- ECON1013-LB06
  - Tuesday 1-2 pm
  - 3 sessions (30-Jan, 13-Feb, 27-Feb)

Record Attendance

Setup

- Step 1: Download Lab materials from **Moodle** page $\rightarrow$ Extract the folder in PC.
- Step 2: Log in **Microsoft onedrive** using your student account https://onedrive.live.com/login/ and upload the folder above.
- Step 3: Launch the **Excel** online https://www.office.com/launch/excel?auth=2, which we will use for all lab sessions.

Exercise 1. Confidence intervals.

Exercise 1. Confidence intervals.

- Data set: `testscores.xls`
- About: A sample ($n = 200$) of student test scores in Math and English
  - Minimal text score is 0 and maximal test score is 100.

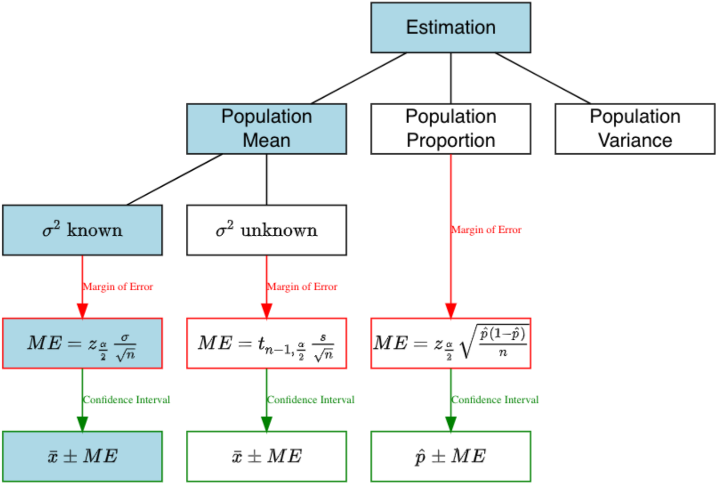
Part 1. Confidence interval for the mean (σ known).

We know that the population standard deviation σ_μ for variable "English" is equal to 4.6.

1. Find the sample mean for the variable "English".
2. Find the standard error of the sample mean for the variable "English".
3. Find the margin of error at the 95% confidence level.
4. Find the 95% confidence interval for the mean.

Hint: $SE = 0.33$, $ME = 0.64$.

Where we currently are



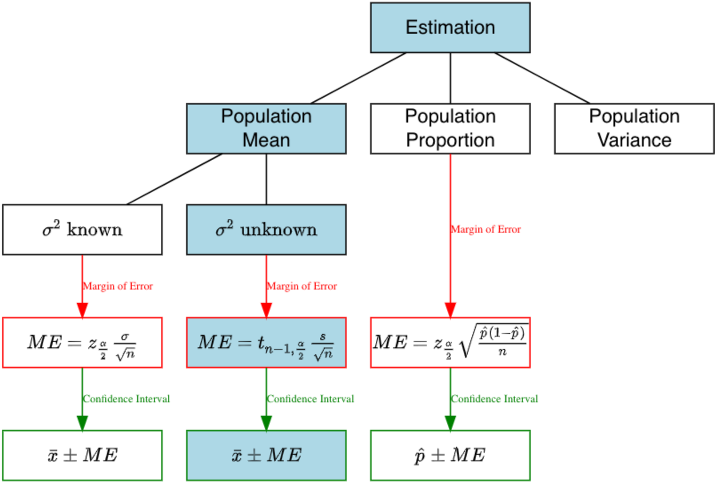
Part 2. Confidence interval for the mean (σ unknown).

We do not know the population standard deviation σ_μ for the variable "Math".

1. Find the sample mean of variable "Math".
2. Find the sample standard deviation of variable "Math", denote as s.
3. Find (an estimate) for the standard error of the sample mean using s.
4. Find the margin of error at the 95% confidence level.
5. Find the 95% confidence interval for the mean.

Hint: $\widehat{SE} = 0.93$, $ME = 1.83$.

Where we currently are



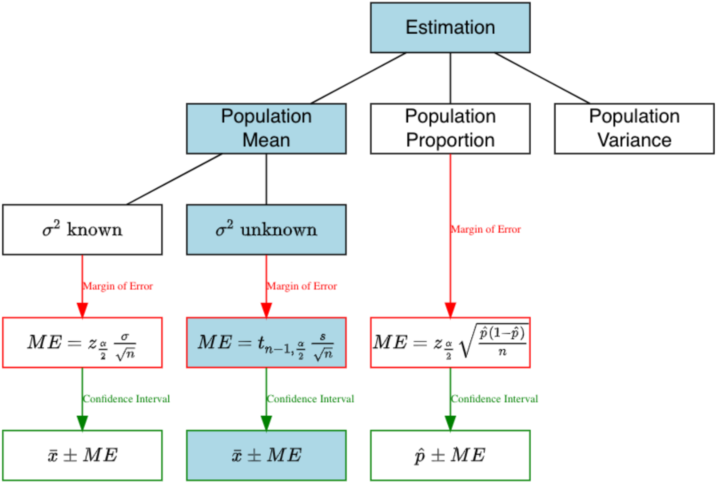
Exercise 2. Confidence intervals.

Part 1. Confidence intervals.

Behave “as if” we did not know that the population mean is 2 and population variance is 1.

For each of the 10 samples, construct the 90% confidence interval for the sample mean.

Where we currently are



Part 2. Coverage.

- Now, we use again our knowledge about the true population mean. Using this information, fill in the blue row, indicating whether the confidence interval contains the true population mean (2).
- Update the table a few times (for example, by refreshing the website) to see what happens. How often is the true population mean contained in the confidence interval?

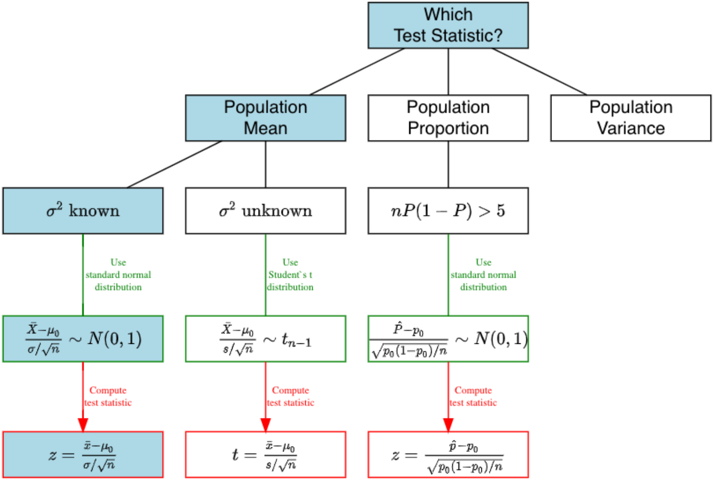
Exercise 3. Hypothesis testing.

Exercise 3. Hypothesis testing.

- Data set: `testscores.xls`
  - About: A sample ($n = 200$) of student test scores in Math and English, drawn from a larger population.
  - We know that the population standard deviation σ_μ for variable "English" is equal to 4.6.

- We want to test the following hypothesis at the 5% significance level:
  - H_0 : The population mean for variable "English" is equal to 73.5.
  - H_1 : The population mean for variable "English" is different from 73.5. Run the test using either the "critical values approach" or the "p-value approach" depending on what you prefer.

Where we currently are



Tests of the Mean of a Normal Distribution Sigma Known

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion

Tests of the Mean of a Normal Distribution Sigma Known

Denote μ the true population mean of English score.

- Null hypothesis H_0:
  H_0: The population mean of "English" is equal to 73.5.
  H_0: $\mu = 73.5$

- Alternative hypothesis H_1:
  H_1: The population mean of "English" is different from 73.5.
  H_1: $\mu \neq 73.5$

- Test statistic
- Decision rule
- Conclusion

Tests of the Mean of a Normal Distribution Sigma Known

Denote μ the true population mean of English score.

- Null hypothesis H_0:
  H_0: The population mean of "English" is equal to 73.5.
  H_0: $\mu = 73.5$

- Alternative hypothesis H_1:
  H_1: The population mean of "English" is different from 73.5.
  H_1: $\mu \neq 73.5$
  $\Rightarrow$ This is a two-tailed test.

- Test statistic
- Decision rule
- Conclusion

Test statistic

Given the sample size and the sampling scheme, the sample average is asymptotically normally distributed:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

We compute the z-score of the observation:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{73.74 - 73.5}{4.6/\sqrt{200}} \approx 0.738$$

Decision Rule

- We are conducting a two-tailed test (*look again* H_1)
- The significance level $\alpha = 0.05$
- Is the decision rule based on **critical values** or **p-value**?

Decision Rule - (1) Critical values approach

For two-tailed test, reject H_0 if

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_{\frac{\alpha}{2}} \text{ or } z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\frac{\alpha}{2}}$$

- Compute critical value:
  $z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = z_{0.025} = 1.96$
  (`Excel` function: $=$ NORM.INV$(0.975, 0, 1)$).

- Compare test statistic to critical value:
  Notice that $z \approx 0.738$, which is greater than -1.96 $(-z_{0.025})$ and less than 1.96 $(z_{0.025})$
  $\Rightarrow$ We DO NOT reject H_0 at $\alpha = 0.05$.

Decision Rule - (2) P-value approach

P-value corresponds to the probability of finding something more extreme than the observed result, under the assumption that the null hypothesis is true.

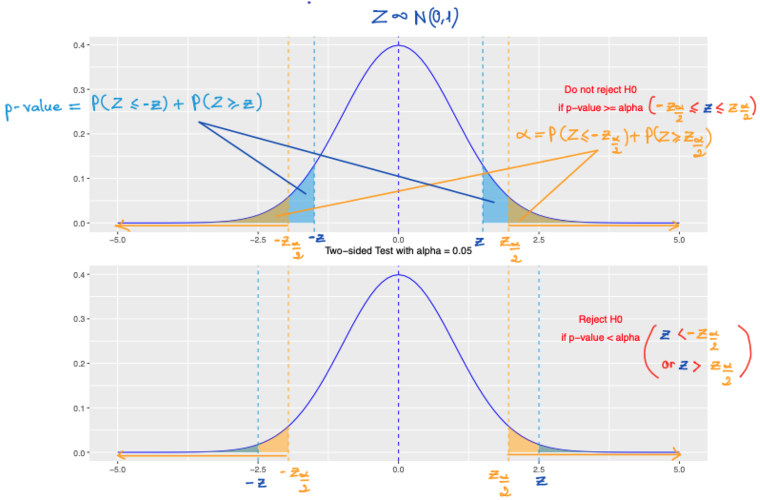
For two-tailed test, reject H_0 if $p - value \leq \alpha$.

- Compute p-value associated with the test statistics $z \approx 0.738$:

$$p - value = \Pr(Z \leq -0.738) + \Pr(Z \geq 0.738) \approx 2 \times 0.23 = 0.46$$

  (`Excel` function: $= \text{NORM.DIST}(-0.738, 0, 1, \text{TRUE})$)
- Compare test statistic to the chosen α: The p-value (0.46) is larger than $\alpha = 0.05$. $\Rightarrow$ We DO NOT reject H_0 at $\alpha = 0.05$ (the same as Critical values approach).

Decision Rule - Two Equivalent Approaches

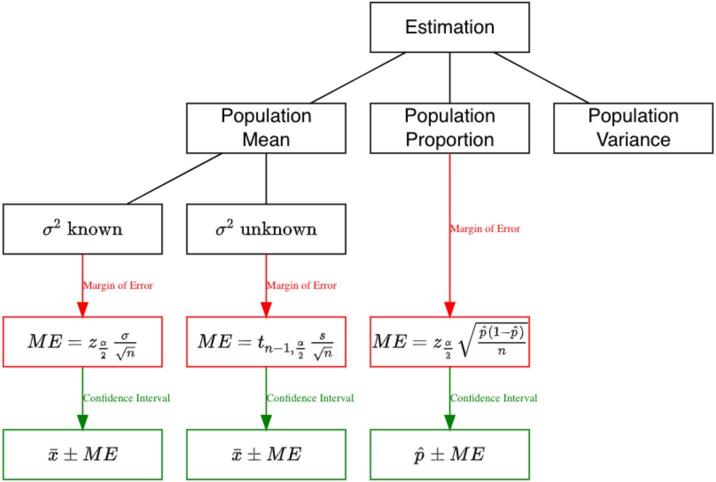


Conclusion

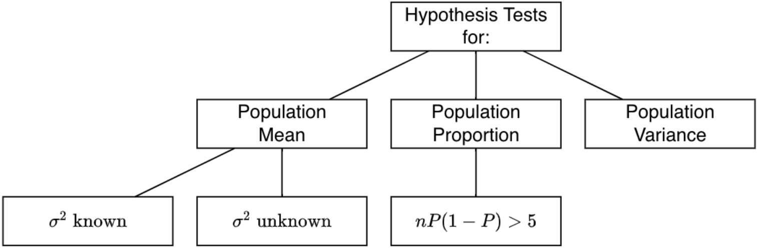
We maintain the null hypothesis. We do not reject the claim that the population mean is equal to 73.5.

BRIEF REVIEW

Estimation


Estimation
Population Mean
Population Proportion
Population Variance
σ^2 known
σ^2 unknown
Margin of Error
Margin of Error
Margin of Error
$ME = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
$ME = t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$
$ME = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Confidence Interval
Confidence Interval
Confidence Interval
$\bar{x} \pm ME$
$\bar{x} \pm ME$
$\hat{p} \pm ME$

Hypothesis Testing

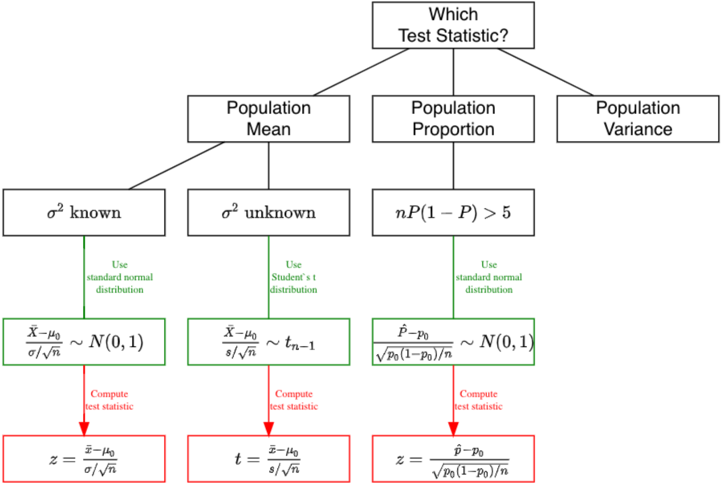


Hypothesis Testing

General procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion

Hypothesis Testing



Tests of the Mean of a Normal Distribution Sigma Known

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion

Tests of the Mean of a Normal Distribution Sigma Known

Procedure includes 5 steps:

- Null hypothesis H_0:

  $$H_0 : \mu = \mu_0$$

  where μ_0 is a hypothesized value.
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion

Tests of the Mean of a Normal Distribution Sigma Known

Procedure includes 5 steps:

- Null hypothesis H_0:

$$H_0 : \mu = \mu_0$$

  where μ_0 is a hypothesized value.

- Alternative hypothesis H_1:

Test	H_1
Two-sided	$\mu \neq \mu_0$
Lower-tail	$\mu < \mu_0$
Upper-tail	$\mu > \mu_0$

- Test statistic
- Decision rule
- Conclusion

Tests of the Mean of a Normal Distribution Sigma Known

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule:
  - Is this a two-sided test or an one-sided (lower-tail/upper-tail) test?
    $\Longrightarrow$ Look again H_1.
  - What is the **significance level** α?
    $\Longrightarrow$ Usually chosen to be 0.01, 0.05 or 0.10.
  - Is the decision rule based on **critical values** or **p-value**?
    $\Longrightarrow$ Distinguish...
- Conclusion

Decision Rule - Two Equivalent Approaches

Approach 1: **Critical-value Test**

Test	H_1	Reject H_0 if
Two-sided	$\mu \neq \mu_0$	$z^s < -z_{\frac{\alpha}{2}}$ or $z^s > z_{\frac{\alpha}{2}}$
Lower-tail	$\mu < \mu_0$	$z^s < -z_{\alpha}$
Upper-tail	$\mu > \mu_0$	$z^s > z_{\alpha}$

Approach 2: **p-value Test**

Test	H_1	p-value	Reject H_0 if
Two-sided	$\mu \neq \mu_0$	sum probabilities to the right of $\lvert z^s \rvert$ and to the left of $-\lvert z^s \rvert$	p-value $\leq \alpha$
Lower-tail	$\mu < \mu_0$	probability to the left of z^s	p-value $\leq \alpha$
Upper-tail	$\mu > \mu_0$	probability to the right of z^s	p-value $\leq \alpha$

*Note: p-value is probability of obtaining a test statistic more extreme ($\leq$ or $\geq$) than the observed sample value given H_0 is true.

Tests of the Mean of a Normal Distribution Sigma Known

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion:
  - Do you *reject* or or *fail to reject* the null hypothesis at the significance level α?

EXCEL NOTES

Relevant functions (I)

Launch the **Excel** online

https://www.office.com/launch/excel?auth=2

NORM.INV() To return the inverse of the normal cumulative distribution for the specified mean and standard deviation (*real number*).

```
= NORM.INV(probability,mean,standard_dev)
```

T.INV() To return the t-value of the Student's t-distribution as a function of the probability and the degrees of freedom (*real number*).

```
= T.INV(probability,degrees_freedom)
```

NORMSDIST() To return the standard normal cumulative distribution (*probability*).

```
= NORMSDIST(z)
```

Relevant functions (II)

Launch the **Excel** online

https://www.office.com/launch/excel?auth=2

NORM.INV(RAND()) To draws a random variable from the normal distribution with the specified mean and standard deviation (*real number*).

```
= NORM.INV(RAND(),mean,standard_dev)
```