L12

Basic Statistics for Economists

Spring 2020

Department of Statistics

Stockholm University

Today

NCT sections 9.1-5

– ***Hypothesis testing***, hypothesis tests, significance level

- How to test the mean $\boldsymbol{\mu}$
- How to test proportion $\boldsymbol{P}$

To make decisions

- A new lion cage has been constructed for a zoo
- We need to determinee if we can display the lion in the new cage
- This depends on whether it is secure, or not…
  - *two hypotheses: ”it is secure” vs. ”it is not secure”*
- So we test it and it is secure!
- But is it really secure?
- This depends on the design of the test…
  - *The probability that the cage passed the test even though it is not secure? And the other way around?*
- Was the outcome of the test random?

Stockholm University

Hypothesis testing

- **Two complementary hypotheses,** i.e. one is the complement of the other
  - We either accept the first hypothesis…
  - or we reject it and accept its complement

- You start from a **null hypothesis** denoted $\boldsymbol{H_0}$
  - This is what you first assume to be true; it is the assumption that the test might reject

- The null hypothesis is tested against an **alternative hypothesis** denoted $\boldsymbol{H_A}$ or $\boldsymbol{H_1}$
  - This is the compliment of the null hypothesis

Hypothesis tests, examples

A. NCT p. 348 there are **two** statements about a mean $\boldsymbol{\mu}$:

$$\boldsymbol{H_0: \mu \leq 16} \quad \text{or} \quad \boldsymbol{H_1: \mu > 16}$$

$$\boldsymbol{H_0: \mu = 16}$$

Composite hypotheses – pertain to a **range** of values (intervals). This is most common.

B. Another example with **two** possible values of proportion $\boldsymbol{P}$:

$$H_0: P = 0{,}5 \quad \text{mot} \quad H_1: P = 0{,}6$$

Simple hypoteses – pertain to only two values of P

Stockholm University

Test variable

- Every test has a **test variable** (random variable)

- Typically based on the statistic used as **point estimate** of the **parameter** that we test
  - E.g. for hypotheses regarding $\mu \Rightarrow$ use $\bar{X}$
  - E.g. for hypotheses regarding $P \Rightarrow$ use $\hat{p}$

- We know the **distribution** of the test variable **given that H_0 is true.** We **assume** that the **null hypothesis is true** when we create our test.

Stockholm University

The test variable is a random variable

- Since the test variable is a random variable, we cannot be 100% confident about the outcome.

- In other words:
  - If the lion cage passes the test, it could have been a random even. It could still be that it is not secure.
  - Or, if the lion cage does <u>not</u> pass the test, this may also be due to randomness. It could be that the cage really is secure.

- We want to **minimize the risk** (probability) of making the **wrong decision**.

Stockholm University

Example

- Somebody claims that $\boldsymbol{\mu = 10}$ $\leftarrow \mathbf{10} = \boldsymbol{\mu_0}$
- You draw a sample and observe a sample mean of $\bar{X} = \bar{x}$
- If $\boldsymbol{\mu = 10}$, how likely is it to observe:
  - $\bar{X} > 11$ $\quad 0,292$
  - $\bar{X} > 13$ $\quad 0,050$
  - $\bar{X} > 16$ $\quad 0,0005$
  - $\bar{X} > 20$ $\quad 0$
  - $\bar{X} < 7$ $\quad 0,050$

The probabilities depend on $\boldsymbol{\sigma^2}$ and $\boldsymbol{n}$. If we assume X_i $\mathbf{iid}$ $N(10, 10^2)$ and $\boldsymbol{n} = \mathbf{30}$, we get these probabilities:

Which observed values of $\bar{x}$ are extreme enough that we **no longer believe** in the statement/hypothesis that $\boldsymbol{\mu = 10}$?

Decision

We define a **critical area** that we can denote C

We **observe** a value of the test **test variable** and if the value

1. **Is in the area C, we reject H_0**
2. **Is not in C, we cannot reject H_0** (we accept)

The critical area is determined by

- the distribution of the test variable
- the alternative hypothesis ← How?
- **the significance level α** ← What?

Stockholm University

The distribution of the test variable

- **Assumption:** observations X_i iid $N(\mu, \sigma^2)$ where σ^2 is known.
- **Null hypotesis:** $H_0: \mu = \mu_0$
- **Test variable** = statistic that estimates μ: $\bar{X}$
- **Distribution given H_0:** $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$
- **Standardized distribution given H_0:** $\boxed{Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)}$

NOTE! We start with the assumption that H_0 is true. This is why we use the value μ_0 when we pick the distribution of the test variable.

The alternative hypothesis

For tests of **one parameter**

$$H_0: \mu = \mu_0 \qquad \text{where } \mu_0 \text{ is some parameter}$$

The alternative hypotesis determines the type of test:

- **Double-sided** test

  $$H_1: \mu \neq \mu_0$$

- **One-sided** test

  $$H_1: \mu < \mu_0 \qquad \text{or} \qquad H_1: \mu > \mu_0$$

Significance level α

- Given that H_0 is true, what is the probability that we observe a standardized test variable $Z > 1.6449$?
  - **Use table 1 or 2**

- **Assuming that** H_0 is true, we get $Z \sim N(0,1)$ and hence $P(Z \leq 1.6449) = 0.95$.

- Suppose that we decide to **reject** H_0 if we observe Z **outside** the interval $(-\infty, 1.6449)$

- The probability to reject H_0 when H_0 is true is $1 - 0.95 = 0.05$

- **Significance level = 5 % =** α

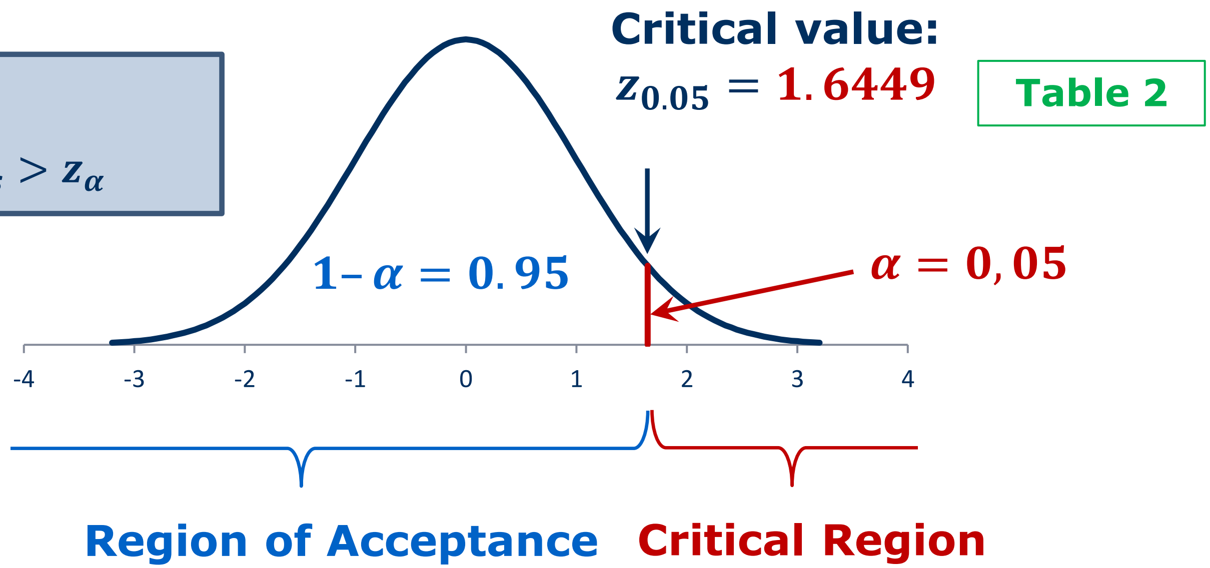
Stockholm University

Critical area, one-sided test 1

$H_0: \mu = \mu_0$ against $H_1: \mu > \mu_0$ **(greater than)**

Distribution given H_0: $Z \sim N(0,1)$

Decision rule:
Reject H_0 if $z_{obs} > z_\alpha$

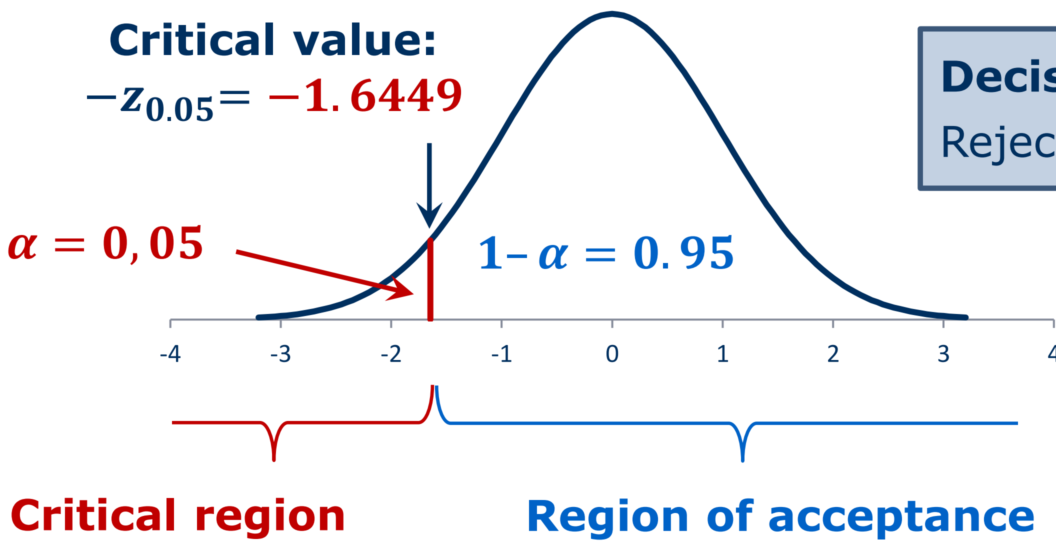


Stockholm University

Critical area, one-sided test 2

$H_0: \mu = \mu_0$ mot $H_1: \mu < \mu_0$ **(less than)**

Distribution under H_0: $Z \sim N(0,1)$



Decision rule:
Reject H_0 if $z_{obs} < -z_{\alpha}$

NOTE! The negative of the table value!

Stockholm University

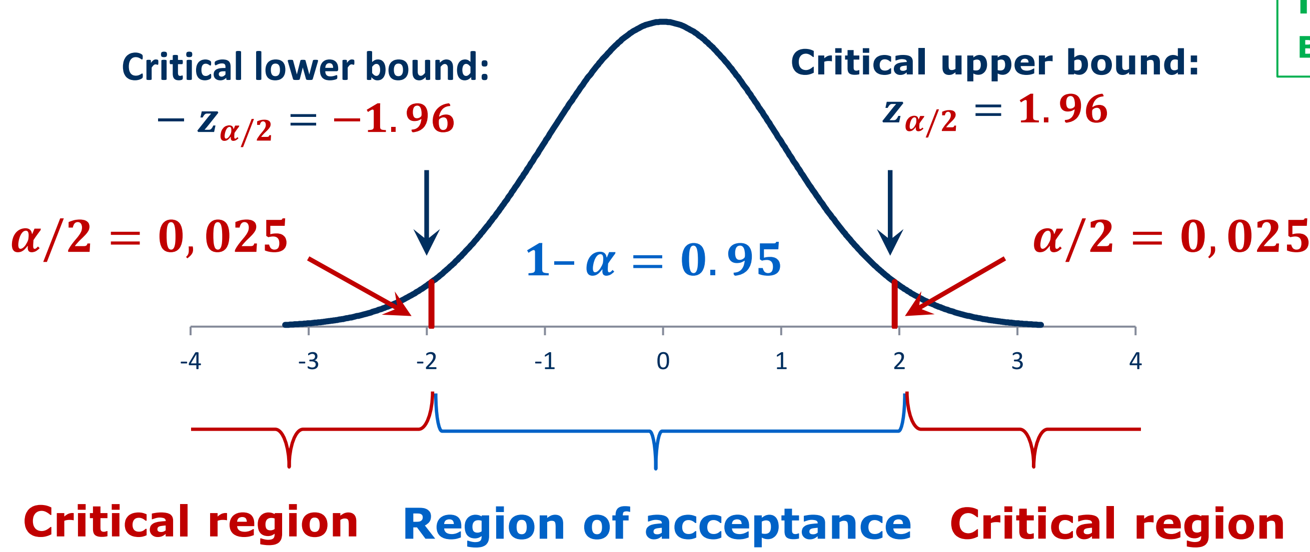
Critical Region, double-sided test

$H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$ (≠ means not equal to)

Decision rule:
Reject H_0 if
$|z_{obs}| > z_{\alpha/2}$

Distribution under H_0: $Z \sim N(0, 1)$

Note! Divide by two: $\alpha/2$



Stockholm University

Types of Errors

- $\boldsymbol{H_0}$ can be true or false

Consequence	Situation/State	
Decision	H_0 true	H_0 false
Accept H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

- $\boldsymbol{H_0}$ can be accepted or rejected
- The decision can be correct or incorrect

Significance level α and power

Significance level or **risk of type I error**

- Significance level = P(Type I) = P(reject H_0 | H_0 true) = α
  - Want this to be **small**

You set α yourself!
Most common: $\alpha = 0,05$

The power of the test (sv. styrka)

- Power = P(reject H_0 | H_0 false) = $1 - \beta$
  - Want this to be **large**, close to 1

This value is harder to determine.

- P(Type II) = P(not reject H_0 | H_0 false) = β
  - Want this to be **small**

Stockholm University

Significance and power, error types

Consequence & probability	Actual situation	
Decision	$\boldsymbol{H_0}$ true	$\boldsymbol{H_0}$ false
Accept $\boldsymbol{H_0}$	Correct decision $(1-\alpha)$	Type II error (β)
Reject $\boldsymbol{H_0}$	Type I error (α)	Correct decision $(1-\beta)$

- Probability of Type I Error = significance level
- Probability $1-\beta$ = power

Hypothesis testing for μ

- Hypothesis testing for μ is closely related to calculating confidence interval for μ. The same principles hold true .
- You start with the statistic $\bar{X}$ when you test μ. The starting point is the **distribution** after **standardizing**.

Different cases:

- Large or small samples, $n < 30$ or $n \geq 30$
- Normally distributed or not normally distributed observations X_i
- Known or unknown σ^2

Stockholm University

Summary – which should we use

- When should you use z and when should you use t?

Sample size n	Distribution	Variance σ^2	
		known	unknown, use s^2
Large	normal	Z	Z
	not normal	**CLT** $\Rightarrow Z$	**CLT** $\Rightarrow Z$
Small	normal	Z	t
	not normal	**Not included**	**Not included**

- In the cases marker blue above, where the sample is large and the variance unknown, one could argue that t should be used instead of z. This gives more conservative results, i.e. slightly smaller risk α; using t adjusts for the slightly larger uncertainty that follows from the estimation of the variance.

Exercise/example 1

- Normally distributed X_i with known variance $\sigma^2 = 25$, $n = 100$ large.
- Hypotheses: $H_0: \mu = 50$ vs. $H_1: \mu > 50$ (One-sided, greater than)
- Significance level: set to $\alpha = 1\,\%$
- Test variable: $$Z = \frac{\bar{X} - 50}{\sigma/\sqrt{n}} = \frac{\bar{X} - 50}{0,5} = 2 \cdot (\bar{X} - 50) \sim N(0,1)$$
- Critical value and decision rule:
  - Reject H_0 if $z_{obs} > z_{0,01} = 2,3263$ (Table 2)
- Alt: $z_{obs} = 2(\bar{x}_{obs} - 50) > 2,3263 \iff \bar{x}_{obs} > 51,163$

Stockholm University

Exercise/example 2

- Not normally distributed X_i with unknown variance σ^2, where $s^2 = 25$, and with large sample size, $n = 100$.
- Hypotheses: $H_0: \mu = 50$ vs. $H_1: \mu < 50$ (NOTE! LESS THAN!)
- Significance level: set to $\alpha = 1\,\%$

APPROXIMATE: CLT

- Test variable:

$$Z = \frac{\bar{X} - 50}{S/\sqrt{n}} = \frac{\bar{X} - 50}{0,5} = 2 \cdot (\bar{X} - 50) \sim N(0,1)$$

- Critical value and decision rule:
  - Reject H_0 if $z_{obs} < -z_{0,01} = -2,3263$ (Remember: less than!)
- Alt: $z_{obs} = 2(\bar{x}_{obs} - 50) < -2,3263 \;\Leftrightarrow\; \bar{x}_{obs} < 48,837$

Stockholm University

Exercise/example 3

- Normally distributed X_i with unknown variance σ^2, where $s^2 = 25$, and with small sample size, $n = 16$.
- Hypotheses: $H_0: \mu = 50$ vs. $H_1: \mu \neq 50$ (**Note**! Double-sided!)
- Significance level: set to $\alpha = 5\,\%$ and $\alpha/2 = 2,5\,\%$
- Test variable:

$n - 1 = 15$ Deg. of Freedom

$$t_{n-1} = \frac{\bar{X} - 50}{S/\sqrt{n}} = \frac{\bar{X} - 50}{1,25} = 0,8 \cdot (\bar{X} - 50) \sim t(15)$$

- Critical values and decision rule:
  - Reject H_0 if $|t_{obs}| > t_{15;0,025} = 2,131$ (Table 3)
- Alt: Reject H_0 if $\bar{x}_{obs} < 47,336$ or if $\bar{x}_{obs} > 52.664$

Stockholm University

Exercise/example 4

- Not normally distributed $\boldsymbol{X_i}$ with unknown variance $\boldsymbol{\sigma^2}$, where $\boldsymbol{s^2 = 25}$, and where the sample size is small, $\boldsymbol{n = 16}$.
- Hypotheses: $\boldsymbol{H_0: \mu = 50}$ vs. $\boldsymbol{H_1: \mu > 50}$ (One-sided!)
- significance level: set to $\boldsymbol{\alpha = 5}\ \%$
- Test variable: **NOT PART OF THE COURSE!**

Test of proportion P

- Independent observations from a **Bernoulli distribution** i.e. the observations take the values **0** or **1** where $P(X_i = 1) = P$
- The sum $\sum X_i$ = the number of 1's among n observered
- The sum $\sum X_i$ is then **binomially distributed** $Bin(n, p)$
- $\bar{X} = \sum X_i / n = \hat{p}$ is a unbiased estimate of p
- If n **is large,** the **CLT applies**, i.e. under H_0 $\hat{p}$ is approximately normally distributed:

$$\hat{p} \rightarrow N\left(p; \frac{p(1-p)}{n}\right)$$

Test for proportion P, cont.

- Null hypothesis: $H_0 : P = p_0$
- Expected value: $\mu_{\hat{p}} = E(Y/n) = P$
- Variance: $\sigma_{\hat{p}}^2 = Var(Y/n) = P(1-P)/n$
- But P is unknown and then the variance is unknown!
- Use the value from the null hypotesis P_0 and standardize:

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

Exercise 5: exercise 3 from PE3

Exercise 3.

- After a marketing campaign, it is claimed that at least 20% of the target group know about a particular product.
- A sample $\boldsymbol{n = 1000}$ is drawn. Among the sample, $\boldsymbol{X = 150}$ people know about the product.
- Estimate of the proportion: $\boldsymbol{\hat{p} = Y/n = 150/n = 0,15}$
- Is it still reasonable to claim that $\boldsymbol{P = 0,20}$?
- Test!

Stockholm University

Exercise, cont.

Assumed, $\boldsymbol{p \geq 0,20}$.

- Hypoteser:

$$H_0: p = 0,20$$

$$H_1: p < 0,20$$

The opposite of what is claimed; the opposite of ”at least 20 % know about the product.”

- significance level: set to $\alpha = 0,05$ (for instance)
- Test variable: $$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \sim N(0,1) \qquad p_0 = 0,20$$
- Decision rule and critical value:

Reject H_0 if $z_{obs} < z_{krit} = -z_{\alpha} = -z_{0,05} = -1,6449$

Stockholm University

Exercise, cont.

- Calculations:

$$z_{obs} = \frac{0{,}15 - 0{,}20}{\sqrt{0{,}20 \cdot (1 - 0{,}20)/1000}} = -\frac{0{,}05}{\sqrt{0{,}00016}} = \mathbf{-3,953} < \mathbf{-1,6449}$$

- Decision and conclusion/interpretation:

Since $\boldsymbol{z_{obs}} = \mathbf{-3,95} < \mathbf{-1,6449}$ **we reject $\boldsymbol{H_0}$.**

The observed proportion $\hat{\boldsymbol{p}} = \mathbf{15}\,\%$ **is significantly less than** $\mathbf{20}\,\%$ at a $\alpha = 5\,\%$ significance level.

The claim that the true proportion **is less than 20%** is **statistically significant** at the $5\,\%$ level**.**

Summary

- A parameter is to be examined/tested
  - E.g. μ or P
- Two opposite claims about the parameter: H_0 and H_1
- Start with what is known about the distribution
  - iid, normal distribution, known variance, sample size, CLT
- Identify suitable test variable, Z or t_{n-1}
- Determine α (choose yourself if not stated, 5 % common)
- Find actual bounds, e.g. z_α or $t_{n-1,\alpha/2}$

And interpret!

- Calculate the test variable and make the right decision: reject or keep H_0

Stockholm University

Summary, cont.

- **If the test is**
  - One-sided – use z_α or $t_{n-1;\alpha}$ respectively
  - Double-sided – use $z_{\alpha/2}$ or $t_{n-1;\alpha/2}$ respectively
- **The distribution of the population**
  - Normally distributed gives exact results
  - not normally distributed give **approximate** results thanks to CLT (large sample)
- What should be null hypothesis vs. alternative hypothesis?

  Often we can consider
  - $H_0 =$ what is claimed, the established ”truth”, what you want to prove false
  - $H_1 =$ the opposite of H_0

Often, this is obvious, but not always!

Stockholm University

For next time

Think about the example with the lion cage. Which statement would you pick as null hypothesis?

H_0: ”The cage is secure”

or

H_0: ”The cage is not secure”

Think about the error types, I and II, and their probabilites.

Read

- ***p*-values** (NCT p. 354-356) and a little about power (ch. 9.5)
- Ch. 10 – how to compare and **test <u>two</u> parameters**

Stockholm University