Unit 4

推理

Inference: Confidence Intervals and Hypothesis Tests

ESIGELEC

Instructor: Federico Perea

1

Introduction

- In some populations it is impossible to compute the value of their parameters (mean, variance, …)
- Normally, this happens because the population is too large, or even its size is unknown.
- In such cases you *estimate* the value of these parameters.
  - Punctual estimations.
  - Confidence intervals.
- You may want to answer questions about the value of these parameters.
  - Hypothesis tests.
- Example: what is the average height of the humans? Is the average height of Dutch people greater than the average height of French people?

2

Confidence intervals

- A **confidence interval** is a range (or an interval) that is *likely* to contain a population parameter (mean, standard deviation,…) under study.
- The degree of confidence is the probability that the parameter is contained in the confidence interval, often expressed as the equivalent percentage value, and denoted by 1-α. In this unit you will use 1-α=0.95, although other values are possible (0.9 and 0.99 are also used frequently).
- α is called the level of significance.
- If you increase the confidence, the interval becomes larger
- (170-180) contains the true (unknown) average height of all ESIGELEC students with probability 0.95. (160-190) with probability 0.99.
- If you want a higher probability (confidence) the interval becomes larger!

Confidence interval about the mean of normal populations: example

Certain company wants to study the average income per family in a certain well-off neighborhood in order to propose the price of a new product. The gross net income of ten families (in thousands of Euros) in this district is:(125,135,145,130,120,145,125,130,150,145). Assuming that the monthly income follows a **Normal distribution**, obtain a 95% confidence interval for the average net income (μ) in this city in the following cases:

a) When the population standard deviation is unknown

b) When the population standard deviation is σ=10.

In this example:

Population: the families of the studied neighborhood.
Sample: The ten families analyzed.

hope the confidence interval as small as possible

In the example you have $\bar{x} = 135$, and S=10.54

a) The interval with σ unknown is :

$$\bar{x} \pm t_{n-1}^{\alpha/2} \frac{S}{\sqrt{n}} \Rightarrow 135 \pm 2.262 \frac{10.54}{\sqrt{10}} \Rightarrow [127.46, 142.54]$$

b) When σ is known : increase n for a smaller c.l

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \Rightarrow 135 \pm 1.96 \frac{10}{\sqrt{10}} \Rightarrow [128.8, 141.19]$$

- The confidence interval is the average plus/minus the value obtained before. In the example:

	lower limit	upper limit
sigma = 10	128,80205	141,19795
sigma unknown	127,459476	142,540524

Which interval is larger?
More info, smaller interval (more accurate,better)

- If the population standard deviation (σ) is unknown: in R use function *t.test(income)*.
- If the population standard deviation is known: in R use function *z.test*, under package *TeachingDemos: z.test(income,sd=10)*

5

Values of α and confidence intervals

normally the confidence is fix given by clients

α (level of significance)	1-α (confidence)	Lower limit	Upper limit
0,01	0,99	126,4139023	143,586098
0,05	0,95	128,4667867	141,533213
0,1	0,9	129,5171546	140,482845
0,25	0,75	131,1655021	138,834498
0,5	0,5	132,7517008	137,248299
0,75	0,25	133,9378688	136,062131
0,9	0,1	134,5811288	135,418871
0,95	0,05	134,7909774	135,209023
0,99	0,01	134,9582218	135,041778

6

Confidence interval about the proportion: example

- A bank wants to estimate the proportion of clients that use their credit card more than once per week. Since there are too many clients, a sample of *n=200* clients was analyzed, 120 of which turned out to use their credit card more than once per week. Find a 95% confidence interval for this proportion.
- Solution: The 100(1-α)% confidence interval for the population proportion *P* is: $\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$

  Since in this example $\bar{p} = \frac{120}{200} = 0.6$ and α=0.05, we get:

  $$0.60 \pm 1.96\sqrt{\frac{0.60(1-0.60)}{200}} = \left[0.5321, \quad 0.668\right] \Rightarrow \left[53.21\%, \quad 66.80\%\right]$$
- $In\ R\colon prop.test(120{,}200, conf.level = 0.95)$

Example

- In order to test the effectiveness of an advertisement campaign, the expenses on a certain product (assumed normally distributed) was checked for 8 different consumers, both before they watched the campaign and after. The results were:

Consumer	1	2	3	4	5	6	7	8
Before	26.2	20.3	25.4	19.6	21.5	28.3	23.7	24.0
After	24.1	21.3	23.7	18.0	20.1	25.8	22.4	21.4

- The differences for this data and the C.I. are:

Consumer	1	2	3	4	5	6	7	8
Difference	2.1	-1.0	1.7	1.6	1.4	2.5	1.3	2.6

$t.test(Difference) =$ (0.5815285, 2.4684715)

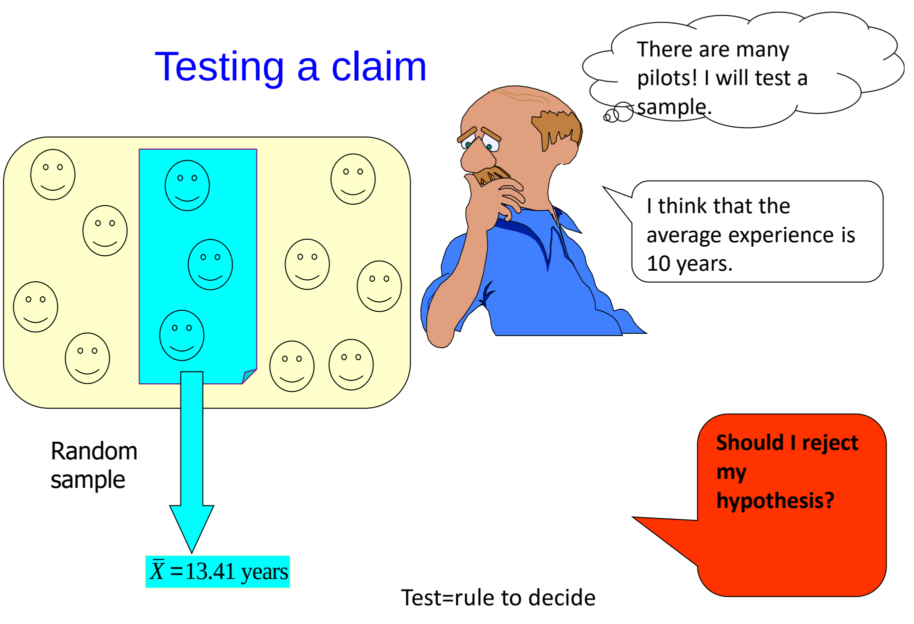
- Since the C.I. is all positive, we can state that the advertisement campaign had a negative effect (the expenses have decreased!)

Hypothesis tests

- Also called tests of hypotheses.
- In statistics, a hypothesis is a statement that something is true.
  - An airline industry representative claims that the mean experience of commercial pilots is 10 years.
- From sample data, we will decide whether or not the claim is true.
  - A sample of 40 randomly selected commercial pilots has a mean experience of 13.41 years and a standard deviation of 8.28 years. Is this data enough to state that the representative's claim is not true?

9

Testing a claim



10

Null vs. Alternative

- Null hypothesis, denoted by H_0:
  It is a statement about the value condition of equality. For the mean, we will state one of these conditions:
  $$\mu = \mu_0; \mu \geq \mu_0; \mu \leq \mu_0$$
- Alternative hypothesis, denoted by H_1
  The statement that must be true if the null hypothesis is false. We will state one of these conditions:
  Two-tail: $\mu \neq \mu_0$ One-tail: $\mu > \mu_0 \; or \; \mu < \mu_0$
- In the commercial pilot example:
  $$H_0: \mu = 10 \; ; H_1: \mu \neq 10$$

11

Testing hypothesis: notation

- Type I error: the mistake of rejecting the null hypothesis when it is true, which is done with probability α (usually 0.05).
- Type II error: the mistake of failing to reject the null hypothesis when it is false, which is done with probability β.
- When α increases β decreases, and vice versa.
- If we want to decrease both, the sample size must be increased.
- In general α+β≠1.
- Accepting the claim does not mean that it is true, it only means that we do not have enough evidence to reject it (WEAK CONCLUSION).
- Rejecting the claim is a stronger statement: we do have evidence to reject the claim (STRONG CONCLUSION).

12

How to accept/reject a test

- Build the *statistic* of the test.
- Build the *acceptance region* of the test
- If statistic is in acceptance region: Do not reject the null hypothesis
- If statistic is not in acceptance region: reject the null hypothesis.

OR (as we will do in this course)

- Compute the $p - value$: probability of obtaining a result equal to *or more extreme than what was actually observed*, assuming the null hypothesis is true.
- If $p - value < \alpha \Rightarrow$ Reject the null hypothesis and accept the alternative
- $p - value \geq \alpha \Rightarrow$ Do not reject the null hypothesis (accept!)

13

TEST ABOUT THE MEAN OF A NORMAL POPULATION

Example:

The speed of n=9 different cars on a certain point of a rural road (normally distributed) was: 69,60,80,85,68,74,60,86,92. Can we admit that the average speed in this point is 70? Use the standard significance level $\alpha = 0.05$.

$H_0: \mu = \mu_0; H_1: \mu \neq \mu_0$ (two-tailed hypothesis)

SOL: t.$test(speed, mu = 10,).p - value = 0.2406$. Do not reject the null hypothesis, the average speed could be considered 70

70,不是10

Example: Is the average speed in this point less than 70?

$H_0: \mu \geq 70; H_1: \mu < 70$ (left-tailed hypothesis)

alternative default value is "two_sided" means 不等号

SOL: *t.test(speed,mu=70, alternative="less"). P-value = 0,8797*. Do not reject the null hypothesis (do not accept the alternative hypothesis). You cannot state that the average speed is less than 70.

there is not enough evidence to reject Ho so that we can't state...

Example: Is the average speed in this point more than 70?

$H_0: \mu \leq 70; H_1: \mu > 70$ (right-tailed hypothesis)

SOL: *t.test(speed,mu=70, alternative="greater"). P-value = 0,1203*. Do not reject the null hypothesis (do not accept the alternative hypothesis). You cannot state that the average speed is more than 70.

In this example, you never get a strong conclusion. Why? The sample size is too small. How to fix that? Analyze more cars !!

14

TEST ABOUT PROPORTIONS.

- Two-tailed alternative hypothesis (H_0: $P=P_0$ H_1: $P\neq P_0$)
- Right-tailed alternative hypothesis. (H_0: $P\leq P_0$ H_1: $P>P_0$)
- Left-tailed alternative hypothesis. (H_0: $P\geq P_0$ H_1: $P<P_0$)

15

Example

- A certain production process typically produces 5% of defective units. You want to check whether this process works as usual or not. For this aim you verify 20 randomly chosen units, out of which two are defective. With a significance level of α=0.05, what can you conclude?
- *Solution:*

The test is: H_o: $P=P_o$; H_1: $P\neq P_o$, with $P_o=0.05$.

P-value = prop.test(2,20,p=0.05,alternative="two.sided") =0.6

→ **you do not reject** *H_o and conclude that there was not a change (increase) in the proportion of defective units.* you may accept p=0.05

What if the sample size was 200, and 20 of them were defective? p-value=0.002055, Strong conclusion there is evident to reject H0, you may accept p!=0.05

What if the question was "p>0.05"?

16

EXAMPLE

- Due to the crisis, you want to check whether or not there has been an increase in the bankruptcy risk in the technological sector (number of companies in risk of bankruptcy divided by the total number of technological companies), which usually is less or equal to 0.02. For this aim, 200 randomly chosen companies have been analysed, out of which 6 were in risk of bankruptcy. With a significance level of 5%, what can you conclude? population: companies in tech sector

p=proportion of companies in risk if p>0.02 Ho: p<=0.02, H1: p>0.02

Solution

The test is: H_0*: P=*P_0*=0.02, against* H_1*: P>0.02*

P-value=prop.test(6,200,p=0.02,alternative = "greater") =0.2243>0.05 weak conclusion

Since 0.030< 0.037 you conclude that there is not enough evidence to reject the null hypothesis, so, there is not enough evidence to accept that the bankruptcy risk is greater than 0.02. But the average proportion is 6/200 = 0.03 … What if you had 60 past-due loans out of 2000? Same average proportion, but same weak conclusion?

17

Testing hypothesis about two means

- It is usually important to compare the means of two populations
- You need to assume that either you have sample data from two normal populations or that the two sample sizes are greater than 30.

- H_0: $\mu_1-\mu_2 = 0$; H_1: $\mu_1-\mu_2 \neq 0$

Do not reject H_0 if p-value is greater than the significance level. Reject if p-value is less than the significance level.

- There are three cases:
  - σ_1, σ_2 known (z.test),
  - σ_1, σ_2 unknown but equal (t.test, var.equal = TRUE),
  - σ_1, σ_2 unknown (t.test, var.equal = FALSE).

18

EXAMPLE

- In order to compare the effectiveness in public institutions of two countries, there is data available about the needed extra hours in 9 different offices in each country along one month. In country A: 200, 230, 205, 185, 190, 300, 250, 245, 208. In country B: 190, 220, 200, 180, 190, 260, 240, 241, 200. Assuming that the number of extra hours follow a normal distribution with equal variance in both countries, and taking significance level equal to 0.05, what can you conclude about the efficiency of both countries' institutions?

Solution: sigma is same? Ho: sigmaA=sigmaB,,, H1: sigma! != sigmaB

H_0:μ_1=μ_2, H_1:$\mu_1 \neq \mu_2$ NOTEBOOK

Since the p-value: *t.test(country.A, country.B, var.equal = TRUE)= p-value=0,516,* greater than the significance level, there is not enough evidence to reject that both countries' offices need the same extra hours.

You need to specify that both variances are equal. The equality of variances can be tested with var.test(), of package *stats*.

TEST ABOUT THE MEAN DIFFERENCE OF PAIRED SAMPLES

- You proceed analogously as you did for the confidence intervals.
- Consider the following example, in which data are taken from a sample of size 9, checking computer performance (measured in number of tasks) before and after a certain new operating system has been installed:

computer	1	2	3	4	5	6	7	8	9
BEFORE	200	230	205	185	190	300	250	245	208
AFTER	190	220	200	180	190	260	240	241	200

- You want to know whether or not this new O.S. made a difference (that is, whether or the average number of tasks has changed)

1 population, 2 R.V : called paired data/ paired sample

mu1=mu2?

EXAMPLE

- In these cases, in which data are paired (before-after) we can use the t-test of paired data. If μ_d *is defined as the mean difference of both populations (the elements before, and the elements after), the test is:*

H_0:μ_d=0, H_1:$\mu_d \neq 0$, and we do not reject H_0 if p-value is more than the significance level:
p-value = *t.test(BEFORE, AFTER, paired = TRUE) = t.test(BEFORE-AFTER) = p-value*

= 0,03028

Since p-value < 0.05, you reject H_0 *and admit that the new operating system makes a difference in the number o tasks done by the computers.*

Exercise

A company training program is designed to reduce the number of defects created by employees on an assembly line for engines, which is assumed to follow a normal distribution. Sample results (obtained on randomly selected days) for randomly selected employees are:

Employee	1	2	3	4	5	6	7	8	9
Defects before program	24	24	27	19	31	29	33	20	26
Defects after program	17	22	9	12	16	21	15	15	19

At the 0.05 significance level, test the claim that the training program had a positive effect on the average number of defects created by employees. At the same significance level, can you admit that the average number of defects decreased 10 units?

SOLUTION:$\mathrm{H}_0: \mu_D = 0;\ H_1: \mu_D > 0.\ \mathrm{p-value} = \mathrm{t.test(dif, mu = 0, alternative = "greater")} = 0.0005.$
Reject the null hypothesis, accept the program had a positive effect.
$H_0: \mu_D = 10\ ; H_1: \mu_D \neq 10.\ p-value = t.test(dif, mu = 10) = 0{,}8681.$ Do not reject the null hypothesis
you can accept that the average number of defects decreased 10 units.

Exercise

The amount of steel present in a certain type of chips is known to follow a normal distribution with mean 35g and standard deviation 1g. After a modification in the production process, it is suspected that such average amount of steel has increased. In order to study such modification, a sample of 10 products has been analyzed, with the following results (in grams):

Unit	1	2	3	4	5	6	7	8	9	10
Steel	35.5	34.0	36.8	37.0	35.8	37.5	34.5	38.0	37.5	36.0

Can you state that the new production process has increased the average amount of steel in the product? Use a 5% level of significance.

SOLUTION: $H_0 : \mu \leq 35$; $H_1 : \mu > 35, p - value =$
$z.test(steel, mu = 35, stdev\ = \ 1, alternative\ = \ \text{greater}) = 0{,}00003.$
Reject the null hypothesis, and accept that the average has increased.

» 23