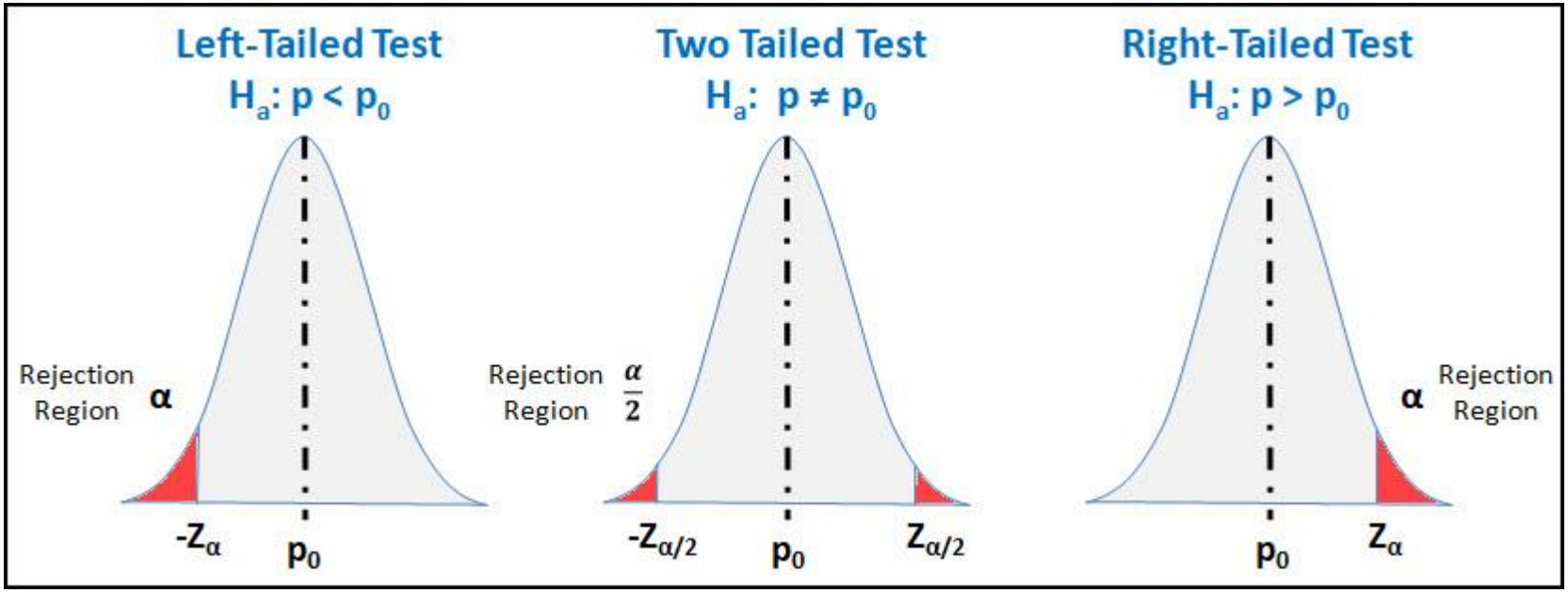
Class 12: Hypothesis tests



Objective

Show how to formally assess whether a sample provides evidence in favour of a particular experimental idea.

Recommended reading

To understand the motivation, look at this link on the relation between hypothesis tests and criminal trials

What is a hypothesis test?

A hypothesis is an affirmation about the population.

The hypothesis is parametric if it refers to the value taken by a population parameter.

For example, a parametric hypothesis is: “the population mean is positive” ($\mu > 0$).

A hypothesis test is a statistical technique for judging whether or not the data provide evidence to confirm a hypothesis.

Example

Given some of the recent policy decisions taken by the Albert Rivera and Cs, it is natural to think that their popularity amongst students might have gone down over the last year.

We recorded the difference between the ratings now and those given 2 years ago by 10 students. The results are:

-2, -0.4, -0.7, -2, +0.4, -2.2, +1.3, -1.2, -1.1, -2.3

Most of the data are negative but do these data provide sufficient evidence that the true mean rating of Rivera in the student population has reduced?

The sample mean of these data is: $\bar{x}$= -1,02.

Does this reflect a real decrease in popularity or is it just due to random chance?

The elements of a hypothesis test

The hypothesis that you want to find evidence for is called the alternative or experimental hypothesis. This is denoted by H_1. In the example:

$$H_1 : \mu < 0$$

The contrary hypothesis to H_1 is called the null hypothesis. This is denoted by H_0. In the example:

$$H_0 : \mu = 0$$

As we want to see whether the mean grade really has gone down, we test:

$$H_0: \mu = 0 \quad \text{vs} \quad H_1: \mu < 0$$

The basic approach to carrying out the test is as follows:

1. Suppose that H_0 is true, $\mu = 0$.

2. Are the data ($\bar{x} = -1.02$) unlikely to have occurred if H_0 is true?

3. If the data are unlikely, this provides evidence against H_0 and in favour of H_1.

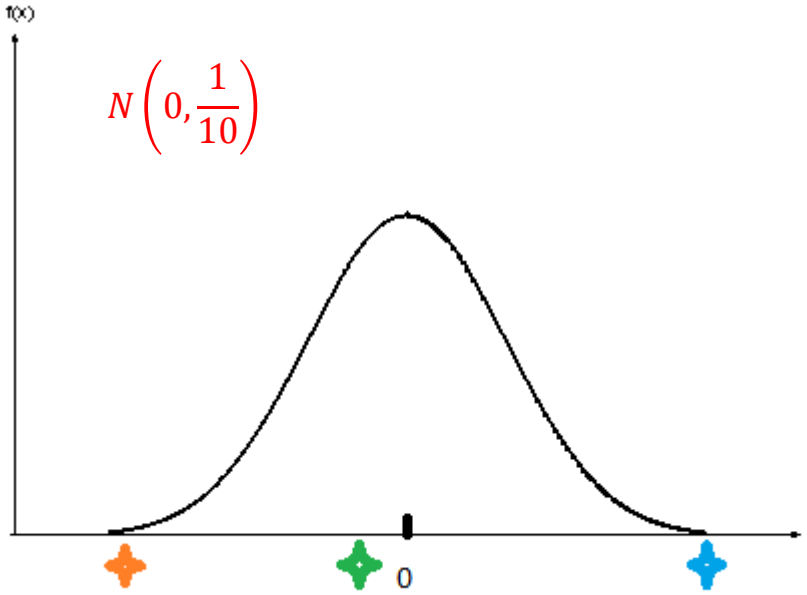
   To carry out the previous analysis we need to study the values that we would expect $\bar{x}$ to take if H_0 really was true (and H_1 false).

   To simplify things, assume that the population is normal and the population variance is known to be equal to 1.

Remember that $\bar{X}$ has a normal distribution with mean μ and variance $\frac{\sigma^2}{N}$.

We know that N = 10 and σ = 1 and if H_0 is true, then μ = 0 so in this case, $\bar{X}$ has a normal distribution with mean 0 and variance $\frac{1}{10}$.

Which values of $\bar{X}$ would provide evidence that H_0 is false and H_1 is true?



Unlikely under H_0 but even more unlikely if H_1 is true.

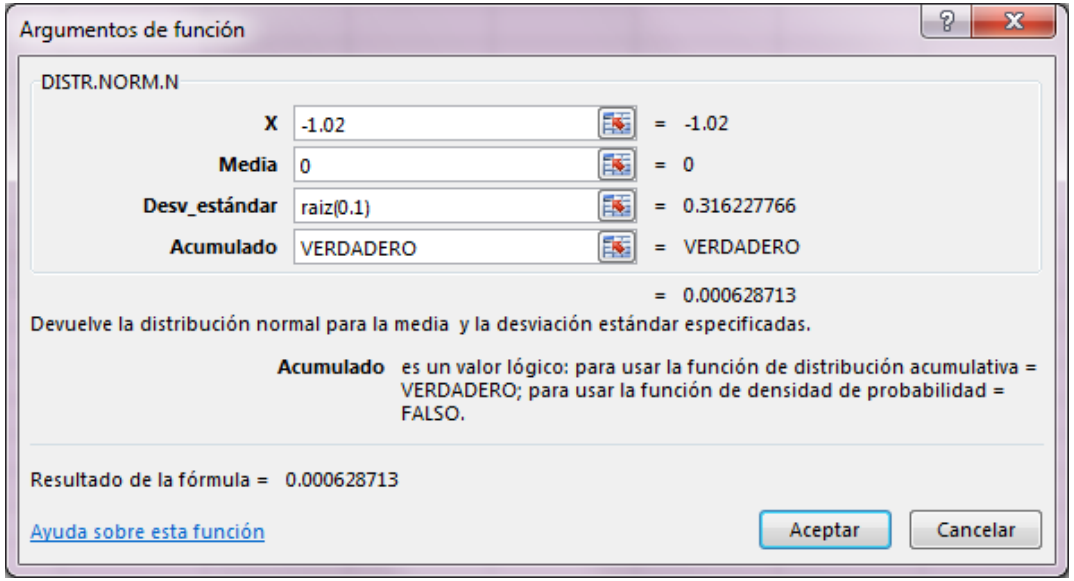
Not unlikely if H_0 is true.

Unlikely under H_0 and more likely if H_1 is true.

The actual data mean is $\bar{x}$ = -1.02. What is the chance of observing such a low value if H_0 is true?

The chance of seeing such a small value is very low.

This would suggest strong evidence against H_0 and in favor of H_1.

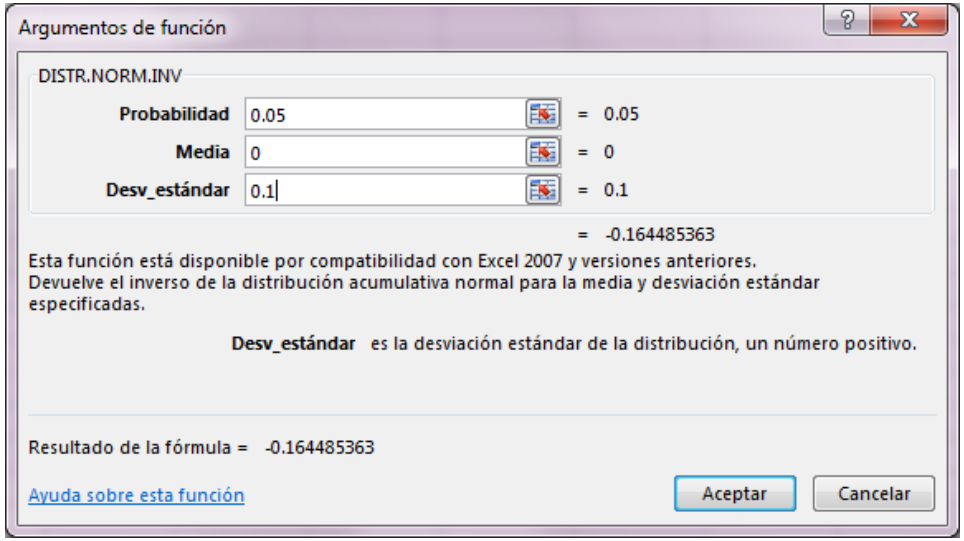


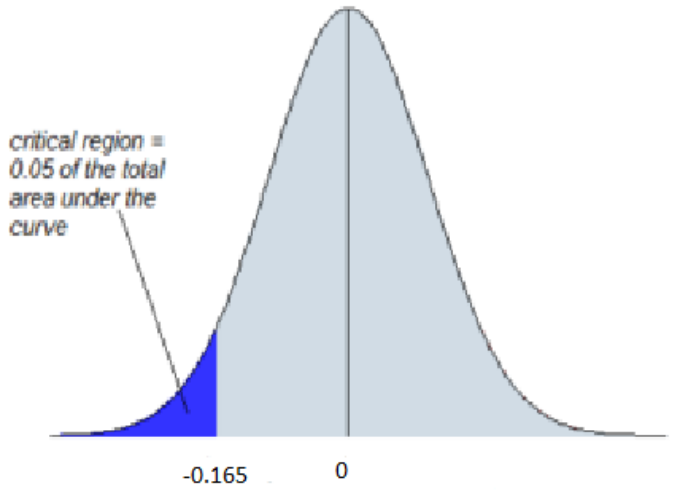
Types of error in an hypothesis test

	H_0 is true	H_1 is true
Don't reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Which of the 2 errors is more serious?

The significance level and the critical region





We can control the type I error by fixing (a priori) the significance level:

$$\alpha = P(\text{reject } H_0 | H_0 \text{ is true})$$

Typical values for α are 0.1 or 0.05 or 0.01.

Given the significance level, the critical region or rejection region is the set of values of the statistic such that we reject H_0 in favor of H_1.

In our example, we would reject H_0 whenever $\bar{x} < 0.165$. This is the critical value.

The p-value

For small values of α, it is harder to reject the null hypothesis.

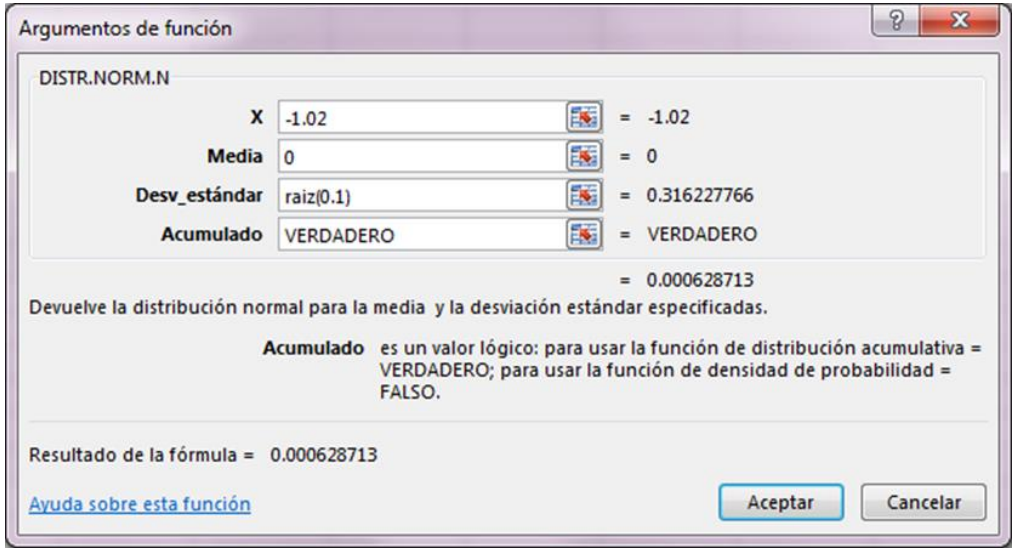
The minimum value of α for which H_0 would be rejected is called the p-value.

The p-value is interpreted as a measure of the statistical evidence in favour of H_1 (or against H_0) given by the data: When the p-value is small, there is strong evidence in favour of H_1.

For fixed α, if the p-value is $< \alpha$ we reject H_0.

p = 0.00063. Very strong evidence against H_0 and in favour of H_1.

We would reject H_0 at a 5%, 1% or even 0.01% significance level.



Example: doing a hypothesis test formally

The table from the CIS barometer of July 2018 gives the estimated mean ratings of various political leaders.

Pedro Sánchez gets one of the highest ratings (4.04) in the barometer, but is there any evidence (at a 10% significance level) that his true mean rating is above 4?

	Media	Desviación típica	(N)
Íñigo Alli	2,56	2,48	(99)
Joan Baldoví Roda	4,05	2,95	(309)
Marian Beitialarrangoitia	3,16	3,01	(110)
Carles Campuzano	3,39	2,59	(213)
Aitor Esteban	3,83	2,54	(240)
Miguel Anxo Fernández Bello	3,25	2,60	(80)
Alberto Garzón	3,98	2,86	(1.624)
Pablo Iglesias	2,96	2,77	(2.289)
Lucía Martín González	2,90	2,38	(63)
Isidro Martínez Oblanca	2,29	2,18	(56)
Ana María Oramas	3,82	2,79	(244)
Pedro Quevedo	3,25	2,55	(159)
Mariano Rajoy	2,83	3,02	(2.374)
Albert Rivera	3,35	2,85	(2.252)
Pedro Sánchez	4,04	2,75	(2.310)
Joan Tardà	2,59	2,89	(982)

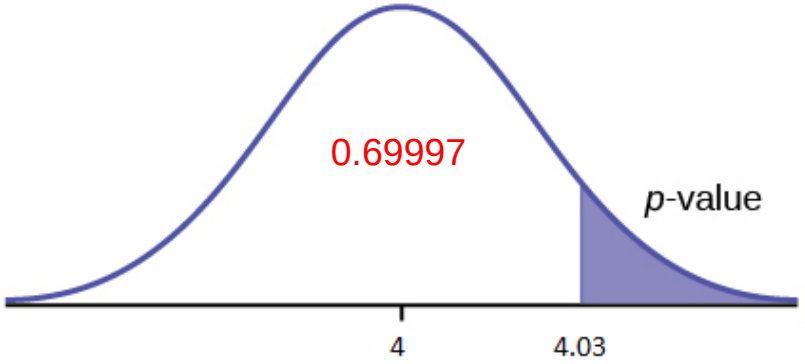
1. μ = true mean rating of Pedro Sánchez.
2. Hypotheses: H_0: $\mu = 4$ (null hypothesis), H_1: $\mu > 4$ (experimental hypothesis).
3. Data: N = 2310, $\bar{x} = 4.04$, $\frac{\sigma^2}{N} = \frac{2.75^2}{2310}$, $\alpha = 0.01$.
4. Distribution of $\bar{X}$ under H_0 is normal with mean 4 and standard deviation $\frac{2.75}{\sqrt{2310}}$.

In reality this is the sample variance, s^2, and not the population value, σ^2, but for a large sample (> 30), this does not matter.

Method I: comparing the p-value with α

5. Draw the p-value.

   Evidence against H_0 and in favor of H_1 from large values of the sample mean.

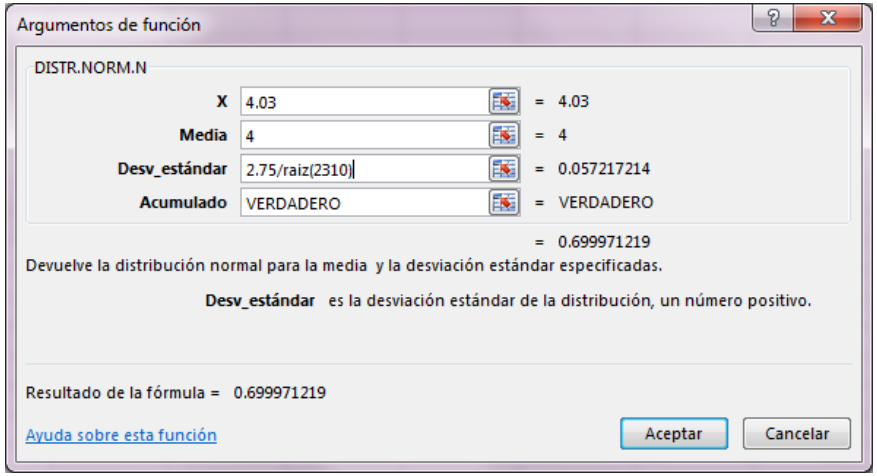


6. Calculate the p-value.

   p-value = 1 – 0.69997 ≈ 0.3.

7. Compare with the significance level.

   0.3 > 0.1.



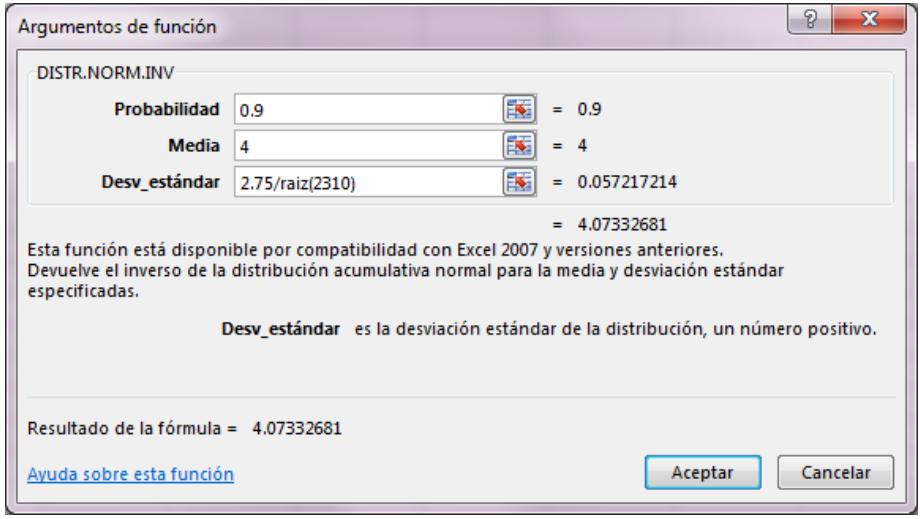
8. Formal conclusion: there is no evidence to reject H_0 at a 10% significance level.

9. Conclusions in real words: there is no evidence that the true mean rating of Pedro Sánchez is above 4.

Method II: is $\bar{x}$ in the rejection region?

5. Calculate the critical value.

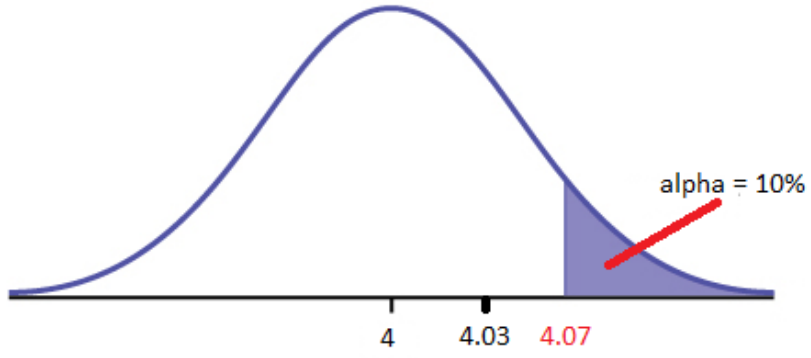
   10% probability on the right means 90% on the left so the critical value is 4.07



6. Draw the rejection region

5. Is $\bar{x}$ in the rejection region?

   4.03 < 4.07 so it is not.



6. Formal conclusion: there is no evidence to reject H_0 at a 10% significance level.

7. Conclusions in real words: there is no evidence that the true mean rating of Pedro Sánchez is above 4.

Two sided tests and confidence intervals

In the previous examples, the tests have been unilateral or one sided (H_1: $\mu < 0$ and H_1: $\mu > 4$). Sometimes we want to test if μ is different from some specified value.

Is there any evidence (at a 5% significance level) that the true mean rating of Albert Rivera is different from 3?

	Media	Desviación típica	(N)
Íñigo Alli	2,56	2,48	(99)
Joan Baldoví Roda	4,05	2,95	(309)
Marian Beitialarrangoitia	3,16	3,01	(110)
Carles Campuzano	3,39	2,59	(213)
Aitor Esteban	3,83	2,54	(240)
Miguel Anxo Fernández Bello	3,25	2,60	(80)
Alberto Garzón	3,98	2,86	(1.624)
Pablo Iglesias	2,96	2,77	(2.289)
Lucía Martín González	2,90	2,38	(63)
Isidro Martínez Oblanca	2,29	2,18	(56)
Ana María Oramas	3,82	2,79	(244)
Pedro Quevedo	3,25	2,55	(159)
Mariano Rajoy	2,83	3,02	(2.374)
Albert Rivera	3,35	2,85	(2.252)
Pedro Sánchez	4,04	2,75	(2.310)
Joan Tardà	2,59	2,89	(982)

1. μ = true mean rating of Albert Rivera.

2. Hypotheses: H_0: $\mu = 3$ (null hypothesis), H_1: $\mu \neq 3$ (experimental hypothesis).

3. Data: N = 2252, $\bar{x} = 3.35$, $\frac{\sigma^2}{N} = \frac{2.85^2}{2252}$, $\alpha = 0.01$.

4. Distribution of $\bar{X}$ under H_0 is normal with mean 3 and standard deviation $\frac{2.85}{\sqrt{2252}}$.

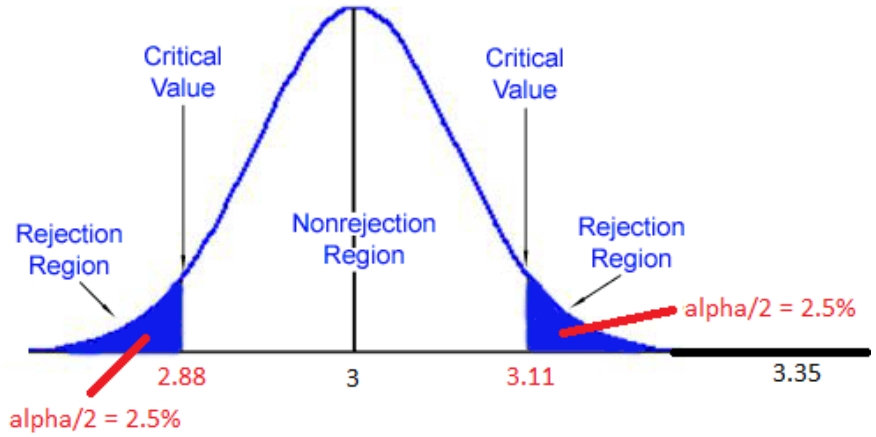
Now we will use Method II to calculate the rejection region. In this case, note that sample means much higher or lower than 3 will give evidence against H_0.

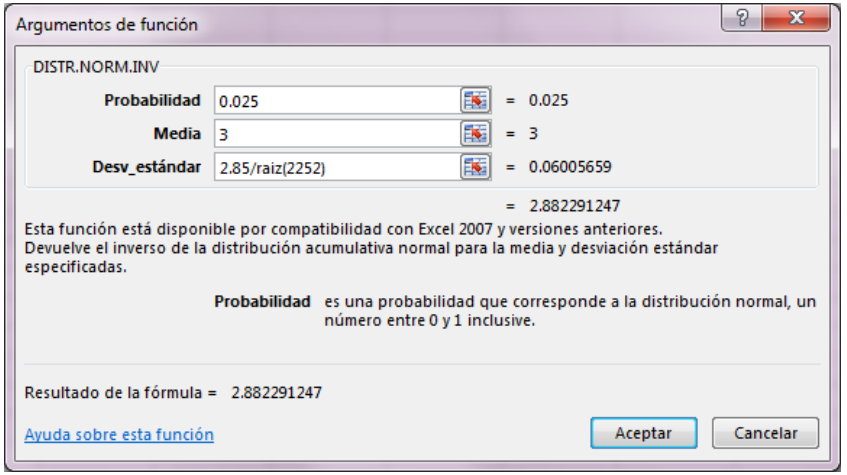
Method II: is $\bar{x}$ in the rejection region

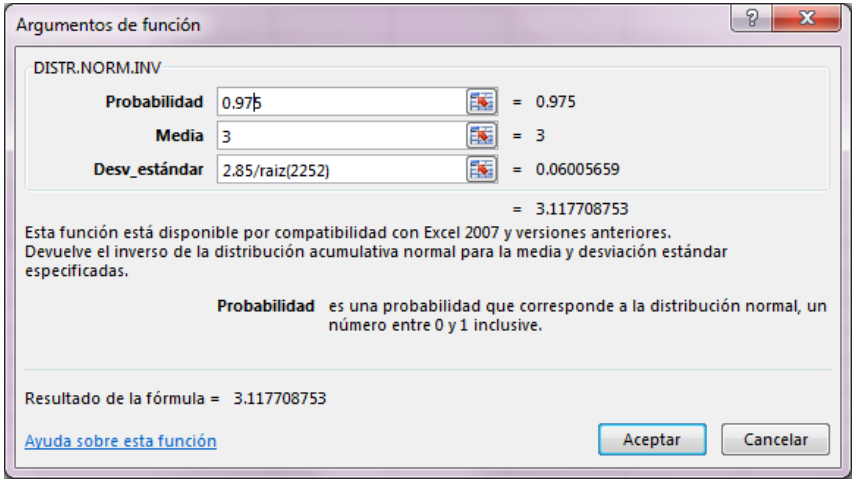
5. Calculate the critical value.

   5% in total is 2.5% on each side

6. Draw the rejection region.







7.-9. 3.35 is in the rejection region so we reject H_0 in favor of H_1 at a 5% significance level. There is evidence that Rivera's true mean rating is different from 3.

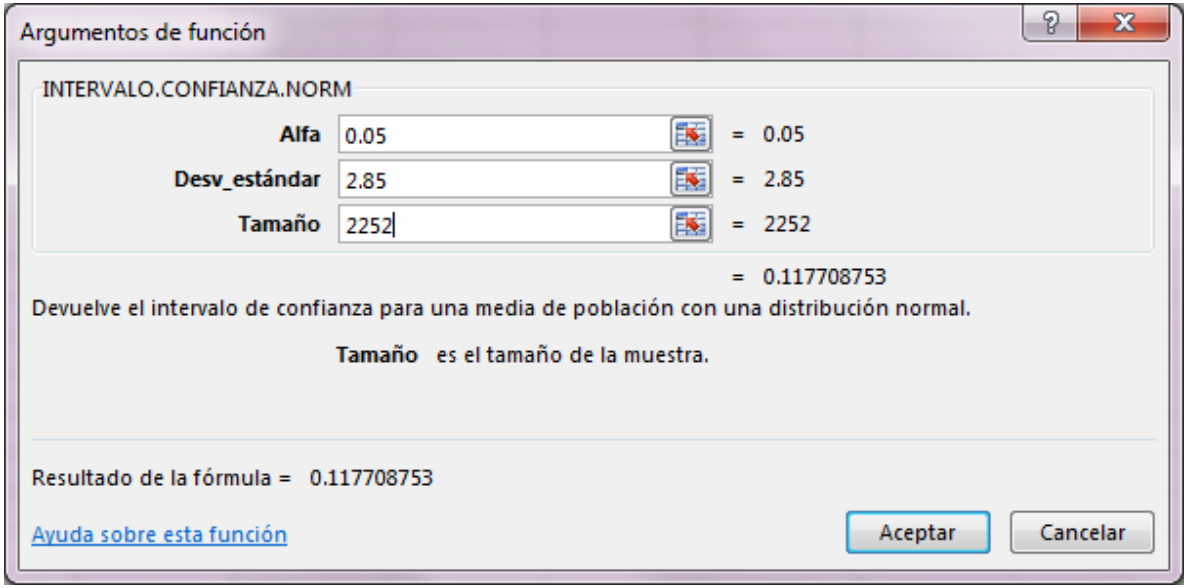
Confidence intervals and two sided hypothesis tests

Suppose that we calculate a 95% = (100 – 5)% confidence interval for the true mean rating of Albert Rivera.

The interval is 3.35 ± 0.12:

(3.23, 3.47).

μ = 3 is outside the interval!



The results of a two sided test with significance level α coincide with the results of a 100(1-α)% confidence interval: we reject H_0: $\mu = \mu_0$ in favor of H_1: $\mu \neq \mu_0$ if μ_0 is outside the interval and do not reject H_0 if μ_0 is inside the interval.

We do not need to do a hypothesis test!

Tests for a proportion

For a large sample, the sample proportion $\hat{p}$ is approximately normal distributed with mean p and variance $\frac{p(1-p)}{N}$ where p is the true population proportion.

Suppose we wish to test the null hypothesis H_0: $p = p_0$.

If H_0 is true, then $\hat{p}$ is normal with mean p_0 and variance $\frac{p_0(1-p_0)}{N}$.

Then we can carry out a test in the same way as previously.

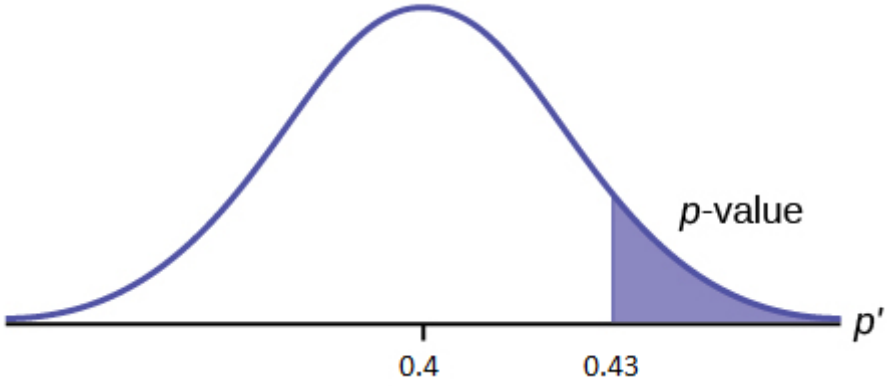
Example

In the last elections, 40% of Madrileños voted PSOE. In a recent study of 100 people, 43 said they would vote PSOE at the next election.

Is there any evidence (at a 5% significance level) that the true proportion of PSOE voters in Madrid has increased from 40%?

1. p = true proportion of PSOE voters in Madrid.
2. Hypotheses: H_0: p = 0.4 (null hypothesis), H_1: p > 0.4 (experimental hypothesis).
3. Data: N = 100, $\hat{p} = \frac{43}{100} = 0.43$, $p_0 = 0.4$, $\alpha = 0.05$.
4. Distribution of $\hat{p}$ under H_0 is normal with mean 0.4 and variance $\frac{0.4(1-0.4)}{100} = 0.0024$.

5. Draw the p-value (or critical region).



6. Calculate p-value.

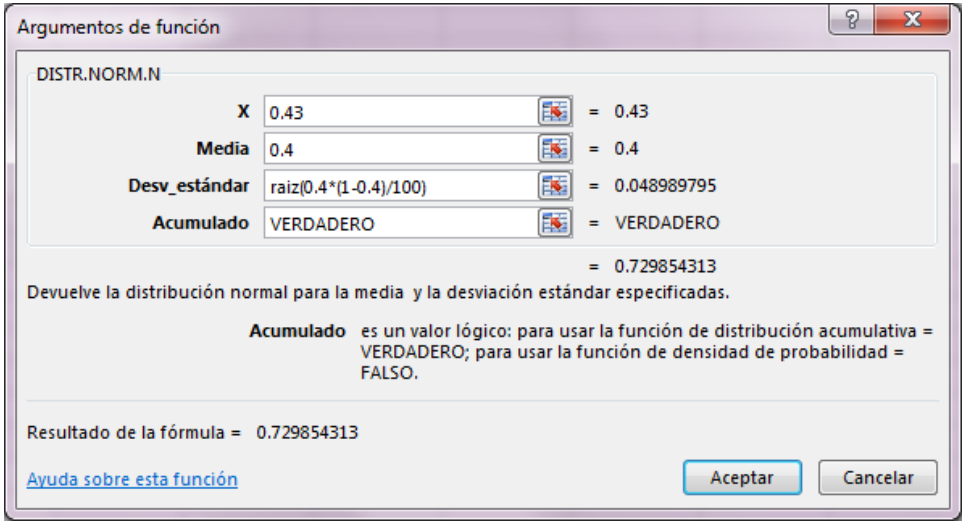
   p-value = 1 – 0.7299 = 0.2701.

7. Compare with the significance level.

   0.2701 > 0.05.

8. Conclusion: There is no evidence to reject H_0 at a 5% significance level.

9. In real words: There is no real evidence that the true proportion of PSOE voters now is higher than 40%.



Example

The following table comes from the CIS barometer of 2011.

PREGUNTA 2
Y, ¿cree Ud. que la situación económica actual del país es mejor, igual o peor que hace un año?

	%	(N)
Mejor	5.3	(130)
Igual	35.1	(865)
Peor	57.6	(1418)
N.S.	1.7	(42)
N.C.	0.3	(8)
TOTAL	100.0	(2463)

More than 50% of the people surveyed thought that the situation got worse in 2011, but is there any real evidence that the true proportion of Spaniards who think this is different to 50%?

Carry out the test at a 5% significance level.

What if we calculated a confidence interval? Is 50% inside?

Example

The following news item was reported in The Daily Telegraph online on 8th May 2010.

General Election 2010: half of voters want proportional representation

Almost half of all voters believe Britain should conduct future general elections under proportional representation, a new poll has found.

The ICM survey for The Sunday Telegraph revealed that 48 per cent backed PR – a key demand of the Liberal Democrats. Some 39 per cent favoured sticking with the current "first past the post system" for electing MPs.
The public was split when asked how they wanted Britain to be governed after Thursday's general election resulted in a hung parliament, with the Conservatives, on 306 seats, the largest party.
Some 33 per cent wanted a coalition government between the Tories and the Liberal Democrats, while 32 per cent thought Nick Clegg's party should team up with Labour.
Just 18 per cent favoured a minority Tory government.
…
*ICM Research interviewed a random sample of 532 adults aged 18+ by telephone on 8 May 2010.

Is there any evidence that less than 50% of UK voters are in favour of PR. Use a 5% significance level.

Example

The following is taken from *Electrometro.com: La web de encuestas electorales en España*.

The PSdG could renew its coalition with BNG in A Coruña (Antena 3)

Lunes 9 Mayo 2011

According to the results of the survey carried out by TNS-Demoscopia for Antena 3 and Onda Cero, the **PP** will get **38.7%** of the votes in **A Coruña**, which will give them **12-13 councilmen** as opposed to the 10 they have at the moment. On the other hand, the **PSdG** will lose 5.6 point with respect to the previous elections and will obtain **29,4%** of the votes which will give them **9 or 10 councilmen.** The **BNG** will obtain **5 or 6 councilmen** by getting **17.7%** of the votes, 3 points less than four years ago.
FICHA TÉCNICA: 500 interviews carried out on **3rd and 4th of May** by **TNS-Demoscopia** for **Antena 3** and **Onda Cero**.

Test whether there is any evidence that BNG will receive less than 20% of the votes. Use a 5% significance level.