Chapter 5.5: Hypothesis Tests

1. What is a hypothesis test?
2. The elements of a test: null and alternative hypotheses, types of error, significance level, critical region
3. Tests for the mean of a normal population
4. Tests for a proportion
5. Two sample problems
6. Tests for independence

5.5.1: 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 decision taken by the Minister of Education, it is natural to think that his popularity rating might have gone down over the last two years.

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 Wert 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?

5.5.2: 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:

$H_0 : \mu = 0$ vs $H_1 : \mu < 0$

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 $$\frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \sim N(0,1)$$

If H_0 is true, then $$\frac{\bar{X}}{1/\sqrt{N}} \sim N(0,1)$$

To see if the sample mean is compatible with μ = 0, calculate

$$\frac{\bar{x}}{1/\sqrt{N}} = -1.02 \times \sqrt{10} \approx -3.225$$

and compare this value with the standard normal distribution.

A value as low as -3,2255 is fairly unlikely given a standard normal distribution N(0, 1), (from the normal tables P(Z < -3.2255) < 0.001), so the data are giving quite a lot of evidence against H_0 and in favour 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 α = P(reject H_0|H_0 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 os the set of values of the statistic such that we reject H_0.

if α = 0,05, we reject H_0 if

$$z = \frac{\bar{x}}{1/\sqrt{10}} < -1.64$$

That is, we reject H_0 if the sample mean is below -0,52.

Setting α = 0,025 we reject H_0 if $\bar{x}$ < -0,62.

The p-value

For small values of α, it is harder to reject the nulll 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-valor is small, there is strong evidence in favour of H_1.

In the example, z = -3,2255 implies that the p-value = 0,00063.
There is a lot of evidence in favour of H_0 and against H_1.

5.5.3: Tests for the mean of a normal population (known variance)

One sided tests

H_0	H_1	Rejection region
$\mu = \mu_0$	$\mu < \mu_0$	$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < z(\alpha)$
$\mu = \mu_0$	$\mu > \mu_0$	$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z(1-\alpha)$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left\lvert\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right\rvert > z(1-\alpha/2)$

Two sided test

Calculation in Excel

	A	B	C	D	E	F	G	H
1	Estudiante	Diferencia						
2	1	-2		n	10			
3	2	-0,4		sigma	1			
4	3	-0,7						
5	4	-2		μ_0	0		$H_0: \mu=0$	
6	5	0,4						
7	6	-2,2		alfa	0,05			
8	7	1,3		z(alfa)	-1,64485	=DISTR.NORM.ESTAND.INV(0,05)	$H_1: \mu<0$	
9	8	-1,2						
10	9	-1,1		z	-3,22552	=(B12-E5)/(E3/raiz(E2))	Cálculo del estadístico	
11	10	-2,3						
12	Media	-1,02		p-valor	0,000629	=DISTR.NORM.ESTAND(E10)	Cálculo del p-valor	

We reject the null hypothesis in favour of the alternative. There is lots of evidence that Wert has grown less popular over the last two years.

We have done the test with tables (and without Excel) as well. It isn't too tough!

Faster calculation with Excel

We can use the function prueba.z(data; μ_0 ; σ).

- The result is the p-value for the test with the alternative hypothesis $H_1: \mu > \mu_0$
- To test $H_1: \mu < \mu_0$ use 1 – prueba.z(…) to get the p-valor.
- For the two sided test, $H_1: \mu < \mu_0$, the p-value is: 2*min(prueba.z(…),1-prueba.z(…))

In the example: calculate 1 – prueba.z(B2:B11;0;1) = 0,00062871

5.5.4: Tests for a proportion

One sided tests

H_0	H_1	Rejection region
$p = p_0$	$p < p_0$	$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < z(\alpha)$
$p = p_0$	$p > p_0$	$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z(1-\alpha)$
$p = p_0$	$p \neq p_0$	$\left\lvert \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right\rvert > z(1-\alpha/2)$

Two sided test

Example

In the last elections, 40% of Madrileños voted PSOE. In a recent study of 100 personas, 37 said they would vote PSOE at the next election.

Calculate a 95% confidence interval for the proportion of people who say they will vote PSOE now.

Is there any evidence that this is different from 0,4? Use a 5% significance level.

Computation in Excel

First calculate the confidence interval.

	A	B	C
1	n	100	
2	x	37	
3	p (estimado)	0,37	
4			
5		0,0946	=intervalo.confianza(0,05;raiz(B3*(1-B3));B1)
6		0,2754	=B3-B5
7		0,4646	=B3+B5

The value 0,4 is inside the interval.

Let's do the test formally as well.

Let p be the true proportion of PSOE voters.
Specify the hypotheses:

H_0: p = 0,4
H_1: p ≠ 0,4

	A	B	C	D	E	F
1	n	100				
2	x	37				
3	p (estimado)	0,37				
4						
5	alfa	0,05				
6	alfa/2	0,025				
7	1-alfa/2	0,975				
8	valor crítico	1,96	=DISTR.NORM.ESTAND.INV(B7)	Calculando el valor crítico		
9						
10	p0	0,4		Valor bajo H_0		
11	z	-0,6124	=(B3-B10)/raiz(B10*(1-B10)/B1)	Estadístico del contraste		
12	|z|	0,6124		en el caso bilateral		

We on't reject the null hypothesis. There is no evidence that p is different from 0,4.

Looking at the interval gives the same conclusion.
Would the same thing apply for tests for a population mean?

Example: (Exam question)

The following data come from the last CIS barometer. The ratings are assumed to come from normal distributions with standard deviations as in the table.

Rosa Diez is the highest rated but has not passed in the sample. Is there any evidence that her true mean rating in Spain is below 5?

Carry out the test at a 5% significance level.

	Media	Desviación típica	(N)
Enrique Álvarez Sostres	2.72	2.38	(133)
Joan Baldoví Roda	3.06	2.76	(104)
Uxue Barkos	4.27	2.74	(302)
Alfred Bosch	3.69	2.78	(211)
Rosa Díez	4.33	2.50	(1594)
Josep A. Durán i Lleida	2.63	2.40	(1454)
Josu Erkoreka	2.85	2.53	(491)
Mikel Errekondo	2.50	2.67	(224)
Francisco Jorquera	3.01	2.48	(216)
Cayo Lara	3.88	2.62	(1379)
Ana María Oramas	3.43	2.50	(170)
Alfredo Pérez Rubalcaba	3.40	2.57	(2314)
Mariano Rajoy	2.81	2.69	(2372)
Carlos Salvador	2.28	2.25	(96)

Ejemplo: (Pregunta de Examen)

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?

Ejemplo: (Pregunta de Examen)

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: (Exam question)

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.

Additional Material

Tests for a normal mean (unknown variance)

One sided tests

H_0	H_1	Rejection region
$\mu = \mu_0$	$\mu < \mu_0$	$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} < t_{n-1}(\alpha)$
$\mu = \mu_0$	$\mu > \mu_0$	$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1}(1-\alpha)$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left\lvert\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right\rvert > t_{n-1}(1-\alpha/2)$

Two sided test

Computation in Excel

In the Wert example, assume the variance is unknown.

	A	B	C	D	E	F	G
1	Estudiante	Diferencia					
2	1	-2					
3	2	-0,4		n	10		
4	3	-0,7		s	1,19610479		
5	4	-2					
6	5	0,4		μ_0	0		H_0: $\mu = 0$
7	6	-2,2					
8	7	1,3		alpha	0,05		
9	8	-1,2		t_{n-1}(alpha)	0,06429815	DISTR.T.INV(0,95;9)	H_1: $\mu < \mu_0$
10	9	-1,1					
11	10	-2,3		t	-2,69668949	(B12-E6)/(E4/raiz(E3))	calculando el estadístico
12	Media	-1,02		p-valor	0,01226316	DISTR.T(-E11;9;1)	calculando el p-valor
13							

We still reject H_0, but it is tougher to do the calculation.

5.5.5: Two sample problems

Suppose we want to test the difference between two normal means.

Consider 4 different situations.

1. Paired samples
2. Two samples with known variances
3. Unknown, equal variances
4. Unknown, unequal variances

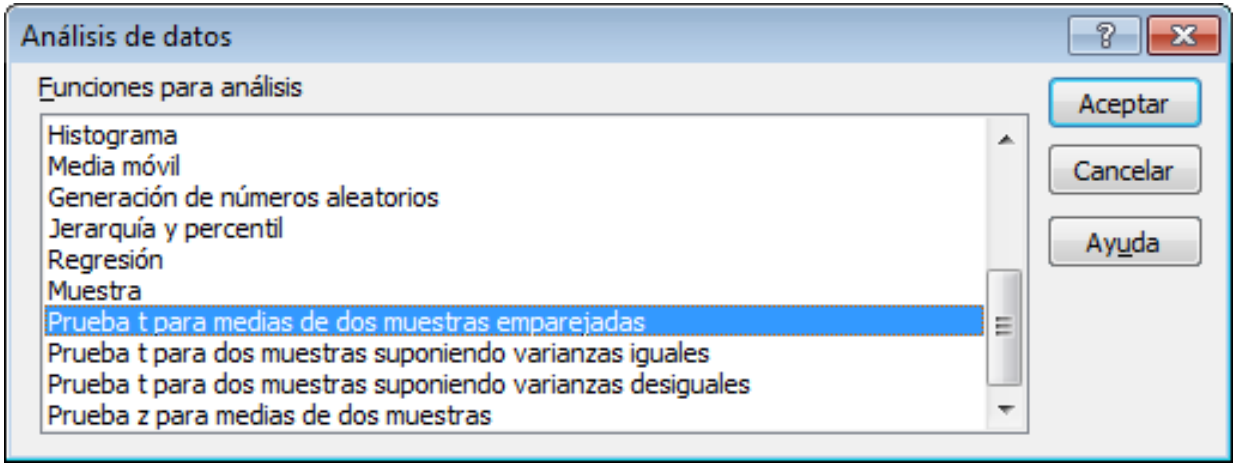
Paired samples

Return to the bankers example. Suppose we want to see is bankers' mean wages went up in 2013

Banker	Year 2012	2013
1	1300	1200
2	1100	1000
3	1200	1500
4	900	800
5	800	750
6	2000	2400
7	1100	1000
8	1500	1600
9	700	700
10	500	600

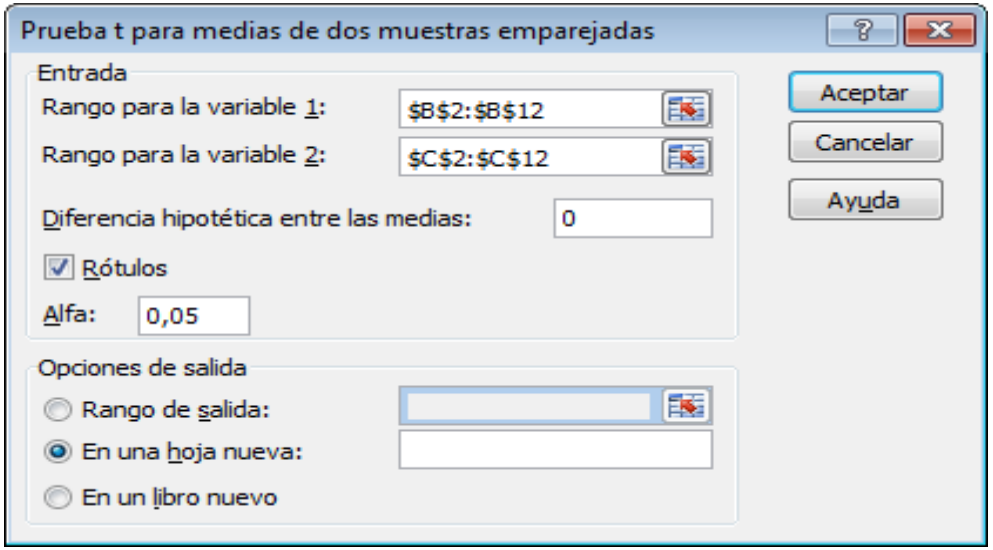
This is easy in Excel.

Usa a t test for two paired samples



Fix $H_0: \mu_{2012} - \mu_{2013} = 0$

and $H_1: \mu_{2012} - \mu_{2013} < 0$



The t statistic is based on the difference of the two sample means
If H_0 is false, we would expect the statistic to be negative.

Prueba t para medias de dos muestras emparejadas

	2012	*2013*
Media	1110,00	1155,00
Varianza	185444,44	302472,22
Observaciones	10,00	10,00
Coeficiente de correlación de Pearson	0,962	
Diferencia hipotética de las medias	0,000	
Grados de libertad	9,000	
Estadístico t	-0,790	
P(T<=t) una cola	0,225	
Valor crítico de t (una cola)	1,833	
P(T<=t) dos colas	0,450	
Valor crítico de t (dos colas)	2,262	

The probability of a low value given H_0 is 22,5%.

Watch out! The critical value is -1,833. (Excel just gives the value associated with H_1: difference > 0)

Therefore we don't reject H_0. There isn't enough evidence to suggest that bankers' mean wages have gone up.