Institute of Psychiatry, Psychology and Neuroscience

KING'S College LONDON

Dr Silia Vitoratou

Department: Biostatistics and Health Informatics

Module Title: Introduction to Statistics

Session Title: Hypothesis testing in SPSS

Topic title: Confidence and significance (II)

Topic materials:
Silia Vitoratou

Contributions:
Zahra Abdulla

Improvements:
Nick Beckley-Hoelscher
Kim Goldsmith
Sabine Landau

Learning Outcomes

- To understand the idea of hypothesis testing in science
- To understand the null and the alternative hypotheses

Hypothesis Testing

We will be repeating this procedure for all tests that we will learn in this course!

Step 1: Create the **null** and the **alternative** hypothesis for the population parameter.

Step 2: **Sample** from the population and compute the correct **statistic** to **estimate** the parameter.

Step 3: Create the **sampling distribution** for this statistic, under the null.

Step 4: Find the **rejection area**.

Step 5: Check if your **sampled** value **falls** in the rejection area.

Equality of Means: The One Sample t-test

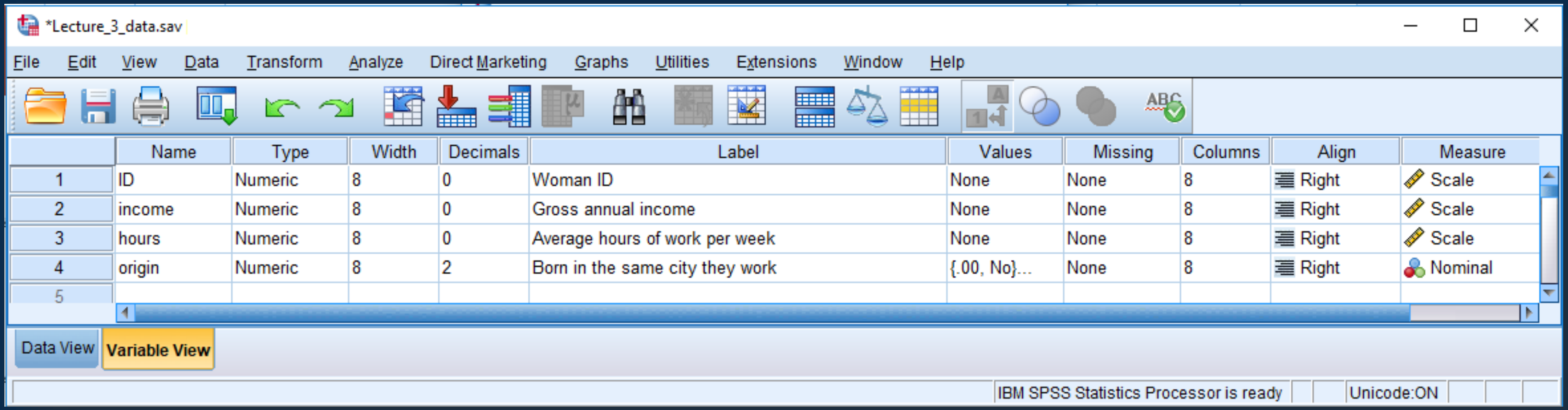
Hypotheses	Suitable test	Decision
H_0: is equal H_a: not equal	*test statistic*	p-value>0.05 do not reject the H_0 p-value≤0.05 reject the H_0

Hypotheses	One sample t-test
H_0: $\mu=\mu_0$ H_a: $\mu\neq\mu_0$	$t=\frac{\bar{x}-\mu_0}{s.e.}$, df=n-1 $s.e.=\sqrt{s^2/n}$

Is the population mean (μ) equal to a certain value (μ_0)?

SPSS Slide: ‘how to’

If you haven’t done already, you can download the data that we are going to use during the lecture. The dataset is the **lecture_3_data.sav**.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	ID	Numeric	8	0	Woman ID	None	None	8	Right	Scale
2	income	Numeric	8	0	Gross annual income	None	None	8	Right	Scale
3	hours	Numeric	8	0	Average hours of work per week	None	None	8	Right	Scale
4	origin	Numeric	8	2	Born in the same city they work	{.00, No}...	None	8	Right	Nominal
5										

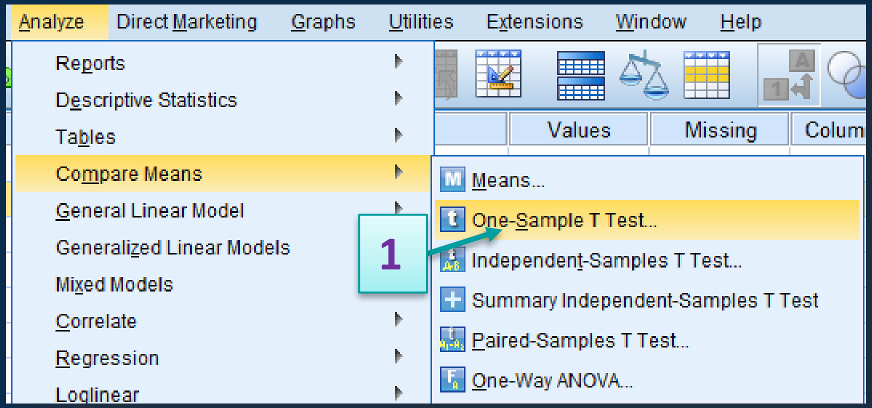
The dataset contains data from 100 women working in a particular industry sector, with respect to

- income: their gross annual income,
- hours: the hours they work each day
- origin: whether the women work in the same city they were born or not

Before anything else, take a minute to think what type of variables you have, compute the descriptive indices and ‘clean the data’.

SPSS Slide: ‘how to’

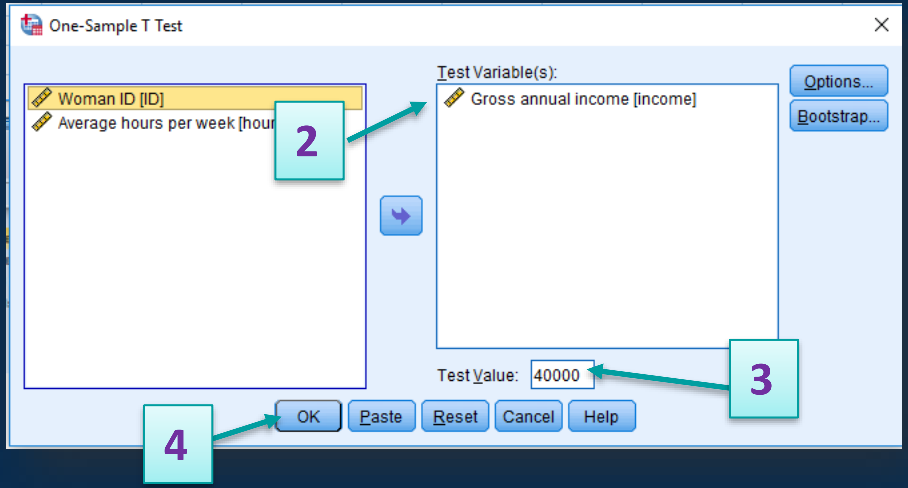
Analyse -> Compare means -> ‘One sample t-test’



Add the variable of interest in the ‘Test Variables’ box (Weight1)

Add in the known test value of interest

Click on ‘OK’



Output and Interpretation

SPSS prints a table with descriptive statistics and one with the one sample t-test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Gross annual salary	100	35151.85	2011.201	201.120
	N	$\bar{x}$	s	se

mean, sd, and se estimated by the sample

Output and Interpretation

SPSS prints a table with descriptive statistics and one with the one sample t-test

Population mean if the null is TRUE

One-Sample Test

Test Value = 40000 → μ_0

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Gross annual salary	-24.106	99	.000	-4848.150	-5247.22	-4449.08

We tested if the difference is zero, rather than if the values are equal.

Now this is for the difference.

H_0: μ=£40000 ⟷ μ-£40000=0

H_α: μ≠£40000 ⟷ μ-£40000 ≠ 0

Output and Interpretation

SPSS prints a table with descriptive statistics and one with the one sample t-test

One-Sample Test

	Test Value = 40000					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Gross annual salary	-24.106	99	.000	-4848.150	-5247.22	-4449.08

Test statistic

$$t = \frac{\bar{x} - \mu_0}{s.e.}$$

Degrees of freedom: n-1

p-value: $p<0.001$

If we reject the null hypothesis, there is less than 1 chance in a 1000 to be wrong.

If the null hypothesis is true, there would be less than 1 chance in a 1000 to sample the value we did. Thus we feel confident enough to reject the null hypothesis based on our data.

Output and Interpretation

SPSS prints a table with descriptive statistics and one with the one sample t-test

One-Sample Test

	Test Value = 40000					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Gross annual salary	-24.106	99	.000	-4848.150	-5247.22	-4449.08

We infer, the difference between men's and women's incomes is ‘statistically’ significant.

That is, womens' income is statistically different than that of men.

Based on our sample, the expected mean difference in the income between women and men is -£4848.2 (95% CI: [-5247.2,-4449.1]). This difference is statistically significant (t=-24.106, df=99, p<0.001).

Equality of Proportions: The One Sample χ^2-test

Hypotheses	**Suitable test**	**Decision**
H_0: means are equal H_a: means are not equal	• *test statistic* and • *degrees of freedom*	p-value>0.05 go with H_0 p-value≤0.05 go with H_a

Hypotheses

H_0: $\mu=\mu_0$

H_a: $\mu\neq\mu_0$

One sample t-test

$t = \frac{\bar{x}-\mu_0}{s.e.}$, df=N-1

$s.e. = \sqrt{s^2/N}$

Is the population mean (μ) equal to a certain value (μ_0)?

Hypotheses

H_0: $\pi=\pi_0$

H_a: $\pi\neq\pi_0$

One sample χ^2-test

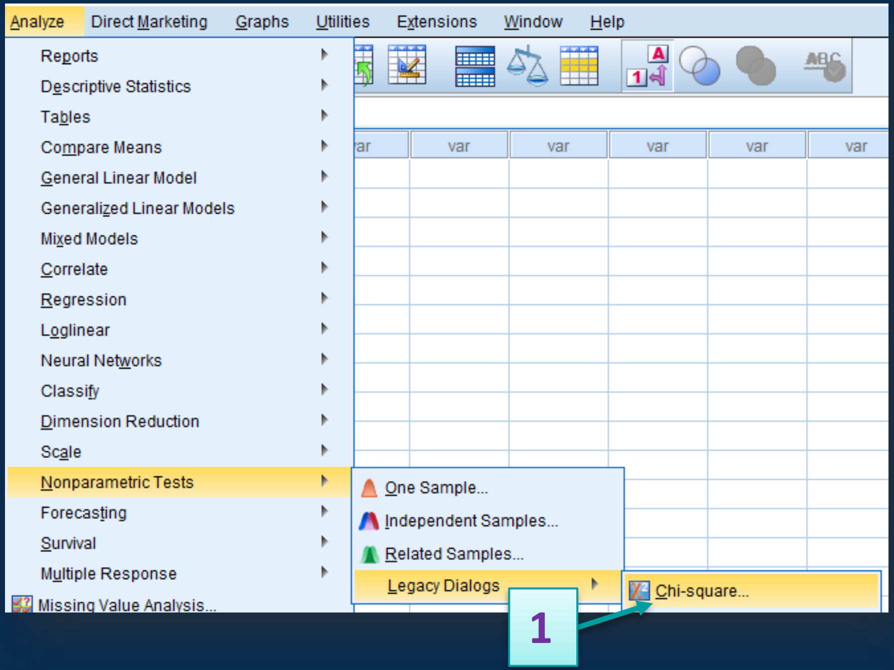
$\chi^2 = \sum\frac{(\mathrm{O}-\mathrm{E})^2}{\mathrm{E}}$, df=c-1

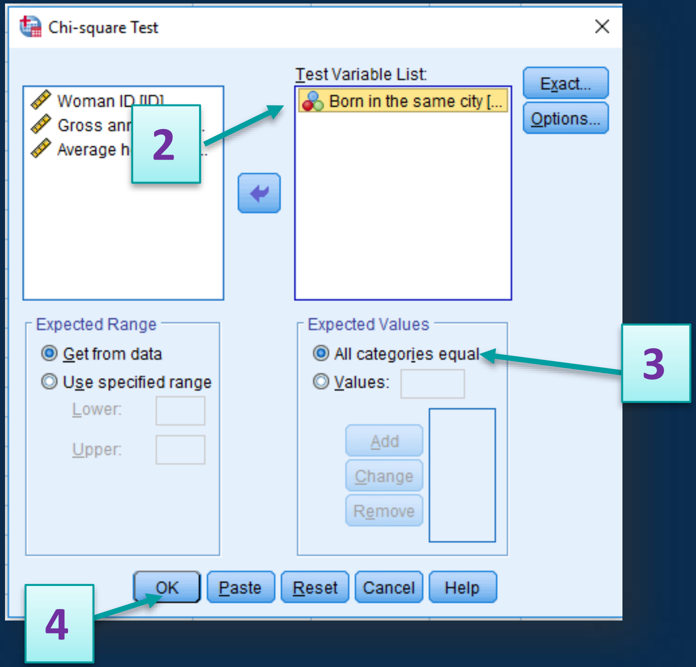
Is the population proportion (π) equal to a certain value (π_0)?

SPSS Slide: ‘how to’

To test if among the women working in this sector the proportion of the women being born in the same city is statistically different than the proportion of the women born in other cities (50%-50%) we will use the ‘one sample χ2-test’.

Analyse -> non parametric tests-> ‘Chi-square’





H_0: $\pi=0.5$

H_a: $\pi\neq0.5$

Equality of Proportions: The One Sample χ^2-test

SPSS prints a table with descriptive statistics and one with the one sample χ^2 -test

Born in the same city

	Observed N	Expected N	Residual
No	81	49.5	31.5
Yes	18	49.5	-31.5
Total	99		

How things are in our sample

How things should be if the null was correct (π=0.5)

Difference $(O - E)$

H_0: π=0.5
H_a: $\pi \neq 0.5$

Test Statistics

	Born in the same city
Chi-Square	40.091[a]
df	1
Asymp. Sig.	.000

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 49.5.

Test statistic $\sum \frac{(O-E)^2}{E}$

Degrees of freedom: c-1

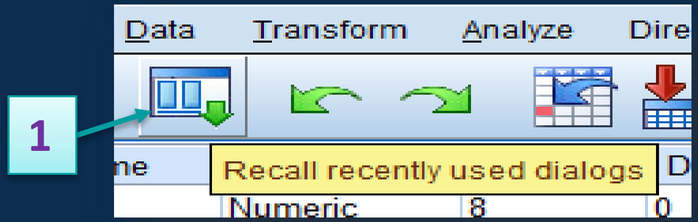
p-value: p<0.001

Based on our sample, among the women working in this industry, the proportion of the women being born in the same city is statistically different than the proportion of the women born in other cities (χ^2=40.091, df=1, p<0.001).

SPSS Slide: ‘how to’

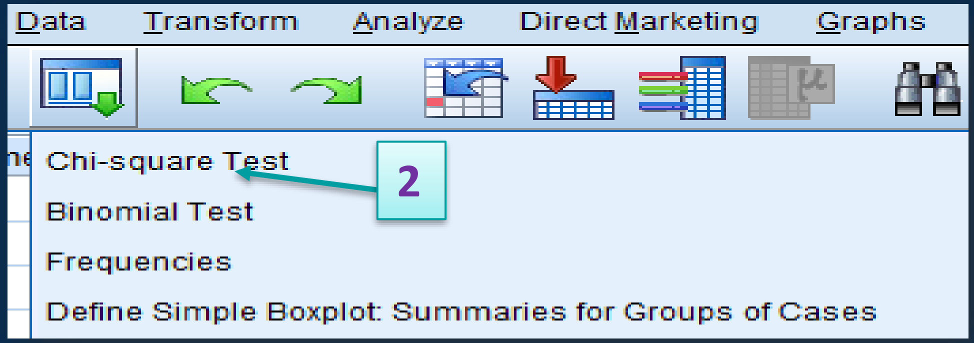
Say now that the proportion of women who work in companies in cities different than those they were born (π) is believed to be 0.2 (20%). Do our data provide evidence against this null hypothesis?

On the command bar press the recall button



H_0: $\pi=0.2$

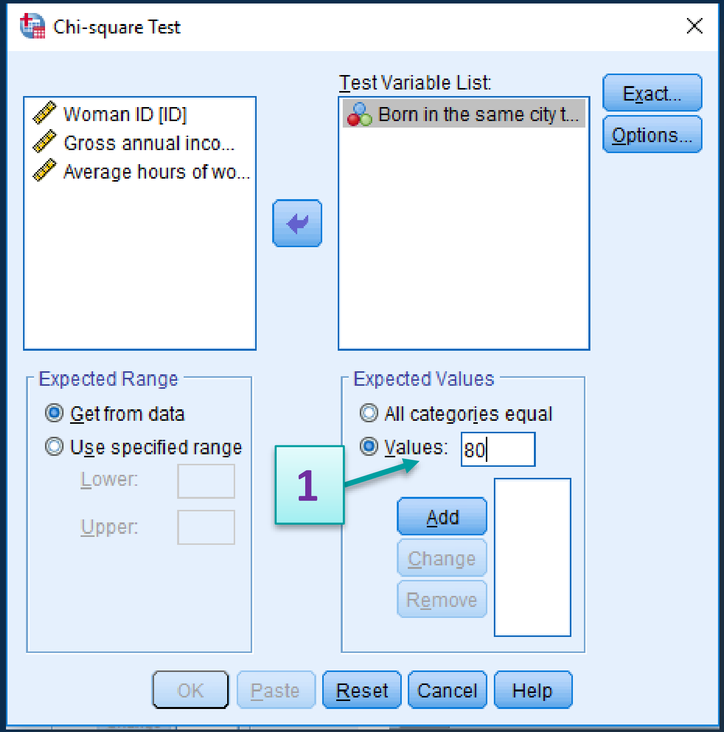
H_a: $\pi\neq0.2$

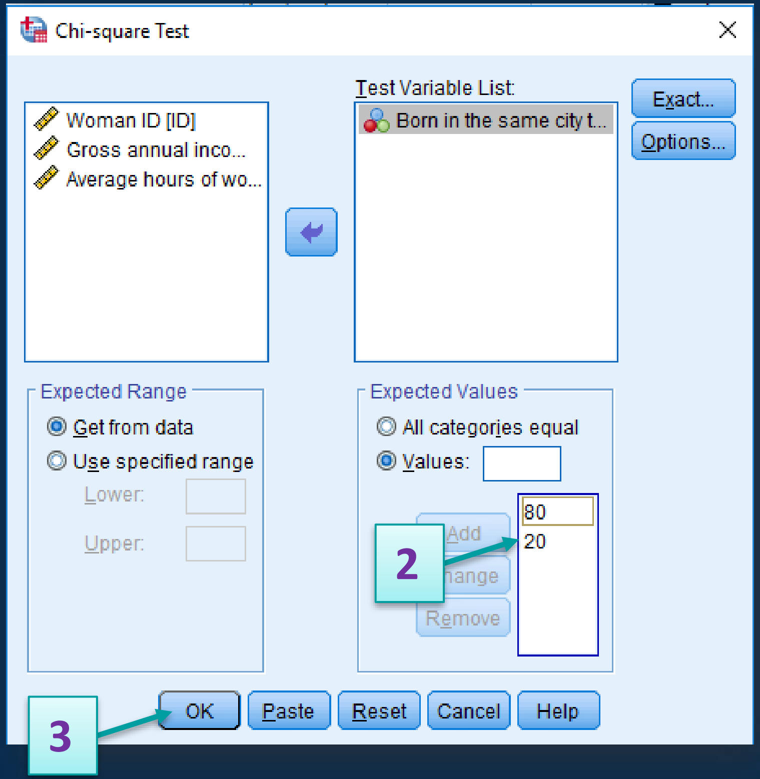


SPSS Slide: 'how to'

Say now that the proportion of women who work in companies in cities different than those they were born (π) is believed to be 0.2 (20%). Do our data provide evidence against this null hypothesis?

In the expected values, type 80% and then 20%





H_0: π=0.2

H_a: π≠0.2

SPSS wants you to add test values for all categories, in order (0, 1, 2, 3…..)

Equality of Proportions: The One Sample χ^2-test

SPSS prints a table with descriptive statistics and one with the one sample χ^2 -test

Born in the same city they work

	Observed N	Expected N	Residual
No	81	79.2	1.8
Yes	18	19.8	-1.8
Total	99		

How things are in our sample

How things should be if the null was correct (π=0.2)

Difference $(\mathrm{O}-\mathrm{E})$

H_0: $\pi=0.2$

H_a: $\pi\neq0.2$

Test Statistics

	Born in the same city they work
Chi-Square	.205[a]
df	1
Asymp. Sig.	.651

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 19.8.

Test statistic $\sum\frac{(\mathrm{O}-\mathrm{E})^2}{\mathrm{E}}$

Degrees of freedom: c-1

p-value: p=0.651

In our sample, the proportion of the women who work in a company in the same city they were born was NOT statistically different than 0.2 (χ^2=0.205, df=1, p=0.651). There were no sufficient evidence to reject the null hypothesis.

Thank you

Please contact <u>your module leader</u> or <u>the course lecturer of your programme</u>, or visit the module's <u>forum</u> for any questions you may have.

If you have comments on the materials (spotted typos or missing points) please contact Dr Vitoratou:

Silia Vitoratou, PhD
Psychometrics & Measurement Lab,
Department of Biostatistics and Health Informatics
IoPPN, King's College London, SE5 8AF, London, UK
silia.vitoratou@kcl.ac.uk

For any other comments or remarks on the module structure, please contact one of the three module leaders of the Biostatistics and Health Informatics department:

Zahra Abdula: **zahra.abdulla@kcl.ac.uk**

Raquel Iniesta: **raquel.iniesta@kcl.ac.uk**

Silia Vitoratou: **silia.vitoratou@kcl.ac.uk**