Institute of Psychiatry, Psychology and Neuroscience

08/2020

Zahra Abdulla

Department: Biostatistics and Health Informatics

Module Title: Introduction to Statistics

Session Title: Risks and Odds

Topic title: Binary Logistic Regression

After working through this session, you should be able to understand:

- Describe odds and odds ratios
- Understand why and when odds and odds ratios can be used
- How to interpret an odds ratio for a binary outcome
- Why risk ratios cannot always be used as measures of relative risk
- How to compare categorical variables across groups using chi-square, fisher's exact test and odds ratios

Some Scenarios

- Are clients with high scores on a personality test more likely to respond to psychotherapy than are clients with low scores?
- Do children have a better chance of surviving a severe illness than do adults?
- Does gender defined at birth increase the risk of being depressed?

What kind of methods would give us an answer to these questions?

Why look at odds?

- We use parametric analysis techniques when we are comparing between groups on a **continuous outcome variable** (e.g. weight differences between gender defined at birth, or cholesterol levels before and after an intervention.

- We use linear regression to look at the change in a **continuous outcome variable** (e.g., mental health dimension)

But…

- What if we don't have a continuous outcome variable?

- What about disease outcomes?
  - Develop disease vs. not develop disease?
  - Depressed vs. not depressed?
  - Death vs. survival?

- These variables would typically be measured as binary and so we need to use a different techniques (**odds ratios and logistic regression**) to examine them.

Odds

In health care, the odds describes the ratio of the number of people with the event to the number without.

Odds of 10-1 at the bookmakers ...

- ... means: the probability that the outcome will not happen is 10 times the probability that it will

Or odds of developing a disorder…

- odds of disorder A = the probability that disorder A **does** happen versus the probability that disorder A **does not** happen

- Other examples:
  - an odds of 0.01 is often written as 1:100,
  - odds of 0.33 as 1:3, and
  - odds of 3 as 3:1

Risk

Risk describes the probability with which a health outcome (usually an adverse event) will occur.
Risk is commonly expressed as a decimal number between 0 and 1

A new drug reduced cancer incidence by 50%

- In absolute terms, the new drug reduced cancer incidence from 2 in 1000 to 1 in 1000.

Relative risk …

- is the probability of an adverse outcome in an exposure group versus its likelihood in an unexposed group. This statistic indicates whether exposure corresponds to increases, decreases, or no change in the probability of the adverse outcome.

- The exposed group has 0.6 times the risk of the outcome (or 40% less risk of the outcome) compared to the unexposed group.

- More examples
  - when the risk is 0.1, about 10 people out of every 100 will have the event;
  - when the risk is 0.5, about 50 people out of every 100 will have the event.
  - In a sample of 1000 people, these numbers are 100 and 500 respectively.

How would we calculate the odds and risk?

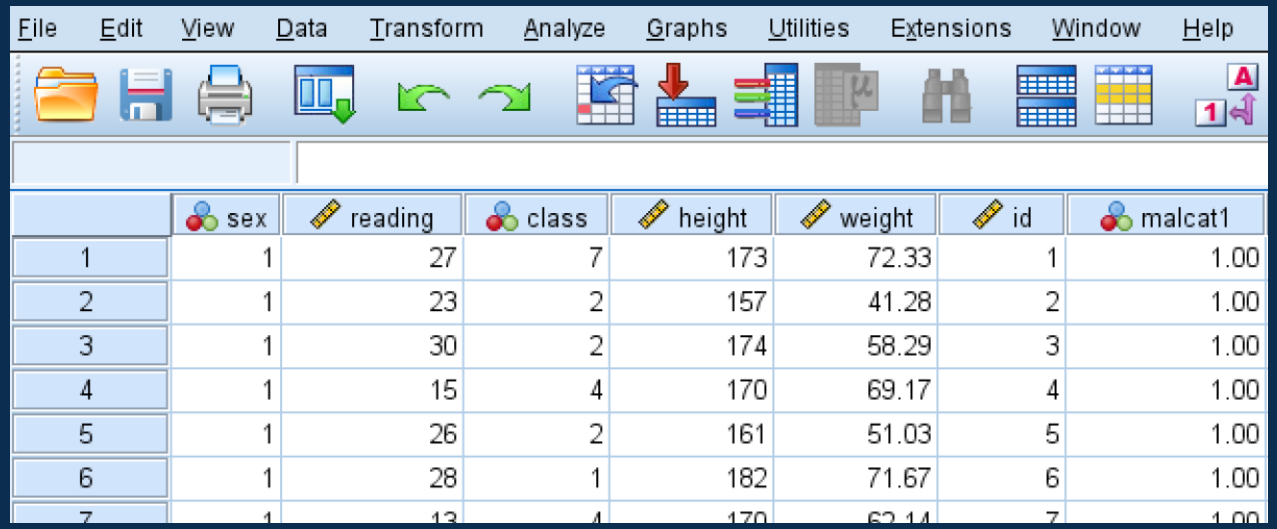
- Could use a **contingency table**

abuse	psychosis	no psychosis	total
exposed +	127	275	**402**
exposed -	187	1,081	**1,268**
total	**314**	**1,356**	**1,670**

- A contingency table summarizes the frequency distribution of each of two categorical variables as well as the **association between two categorical variables**
- Each cell contains the frequency at which combination of its row and column categories occurred
- A contingency table allows us to check
  - How each of the potential explanatory variables are related to the dependent variable, one by one
  - That categories for explanatory variables are large enough (suggest at least 5 cases per category)
  - How many missing cases there are for each variable

SPSS Slide

Download the data that we are going to use during the lecture. The dataset is the **lecture_10_data.sav**.



	sex	reading	class	height	weight	id	malcat1
1	1	27	7	173	72.33	1	1.00
2	1	23	2	157	41.28	2	1.00
3	1	30	2	174	58.29	3	1.00
4	1	15	4	170	69.17	4	1.00
5	1	26	2	161	51.03	5	1.00
6	1	28	1	182	71.67	6	1.00
7	1	13	4	170	62.14	7	1.00

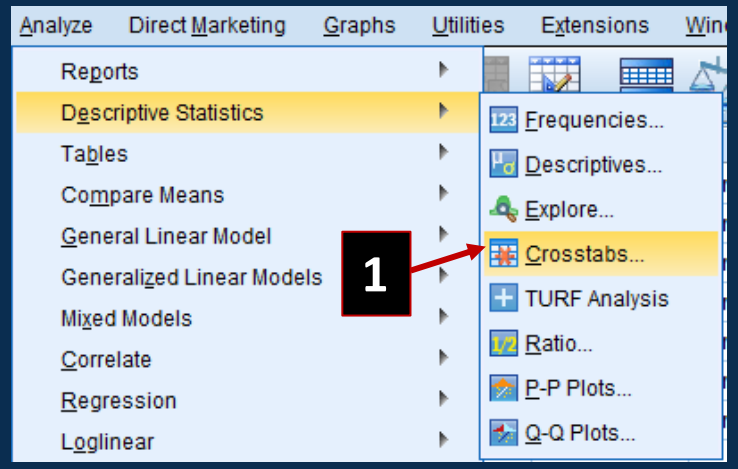
The dataset contains data from 42 babies, with respect to their
Specific body measurements at birth : headcircumf, length, weight (lbs)
Gestation: Gestational age at birth
Information about the baby's mother: smoker, motherage, mnocig, mheight, mppwgt
Information about the baby's father: fage, fedyrs, fnocig, fheight
lowbwt: Low birthweight Baby 0 = No, 1 = Yes
Mage35: 0=under 35, 1=Over 35

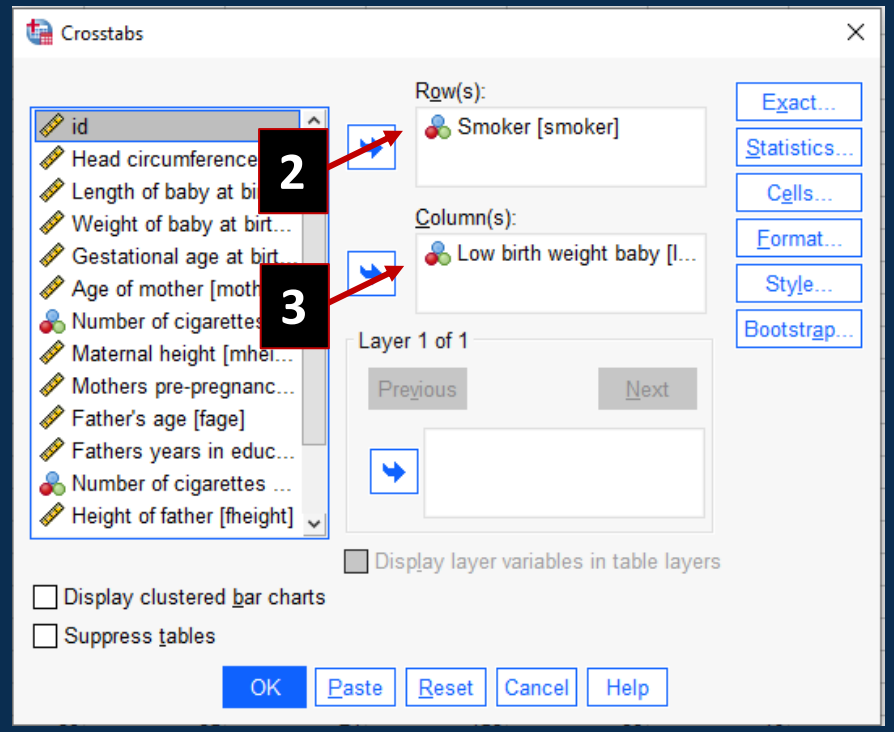
SPSS Slide: ‘how to’

The next question is: Are the proportions of low weight babies different from mothers who smoked through pregnancy compared to those who did not smoke through pregnancy?

Step 1: Create a contingency table

Analyse -> Descriptive Statistics-> Crosstabs

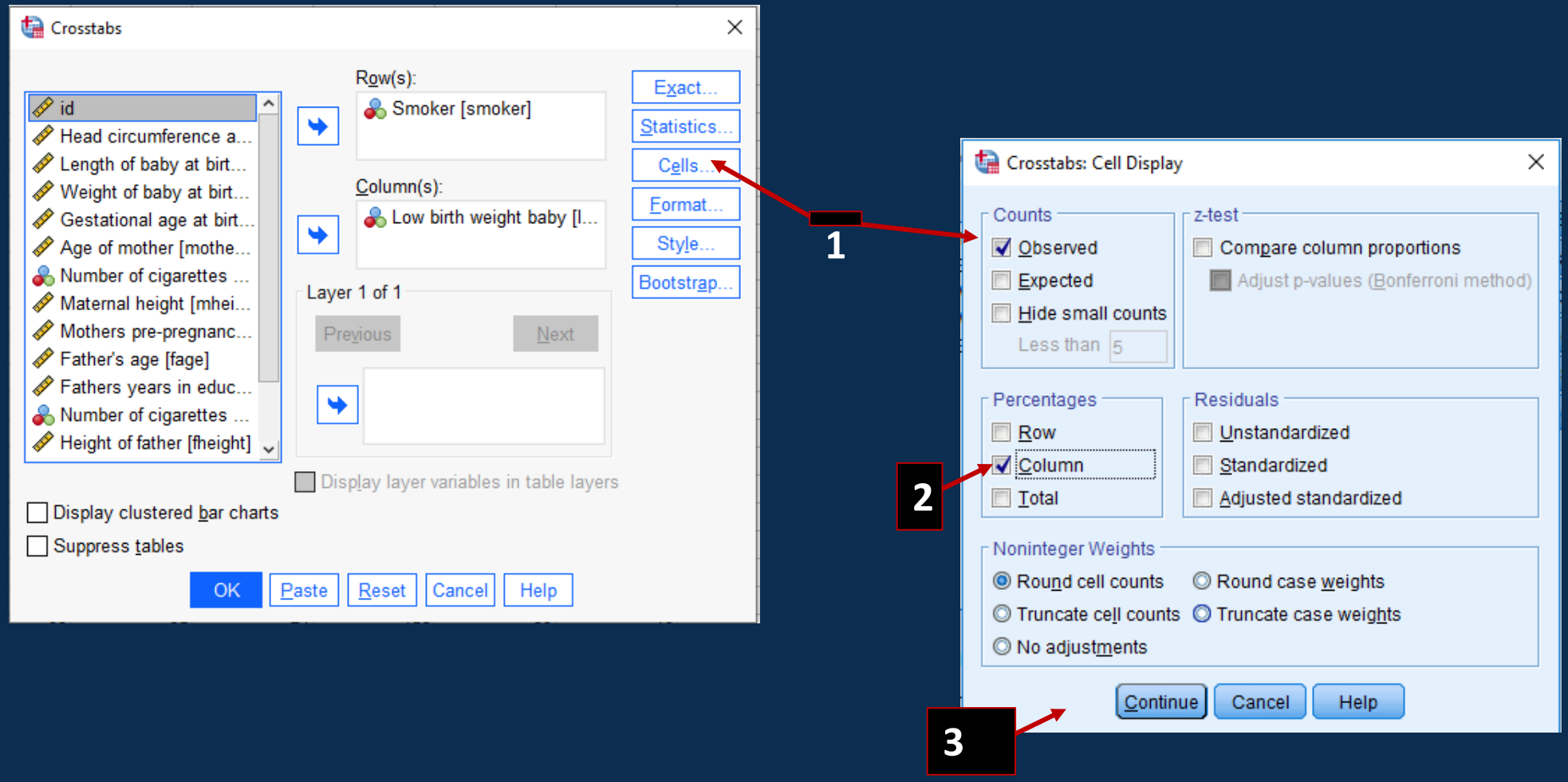




Add the variable of interest (our outcome) (Low Birth Wgt) in to the ‘columns box’.
Add the second variable interest (Smoker) in the ‘rows box’

SPSS Slide: ‘how to’

Step 1: Choose the most appropriate ‘Percentages’



Output and Interpretations

Percentages: Row ☑ Column ☐ Total ☐

Smoker * Low birth weight baby Crosstabulation

			Low birth weight baby		
			No	Yes	Total
Smoker	Non-smoker	Count	15	5	20
		% within Smoker	75.0%	25.0%	100.0%
	Smoker	Count	9	13	22
		% within Smoker	40.9%	59.1%	100.0%
Total		Count	24	18	42
		% within Smoker	57.1%	42.9%	100.0%

Among those babies who were low birth weight, the proportion of those whose mothers smoked during pregnancy was higher than the proportion of whose mothers did not smoke during pregnancy (59.1% versus 25.0%, respectively).

Percentages: Row ☐ Column ☑ Total ☐

Smoker * Low birth weight baby Crosstabulation

			Low birth weight baby		
			No	Yes	Total
Smoker	Non-smoker	Count	15	5	20
		% within Low birth weight baby	62.5%	27.8%	47.6%
	Smoker	Count	9	13	22
		% within Low birth weight baby	37.5%	72.2%	52.4%
Total		Count	24	18	42
		% within Low birth weight baby	100.0%	100.0%	100.0%

Among those mothers who smoked during pregnancy there was a higher proportion who had a baby of low birthweight compared to a baby of normal birthweight (72.2% versus 37.5%, respectively).

NEVER compare percentages which add up to 100%!

Pearson's chi-square test

When to use

To test if, according to the current data, the proportions in the population of babies being born of low-birth-weight changes based on mothers smoking status during pregnancy

Hypotheses:

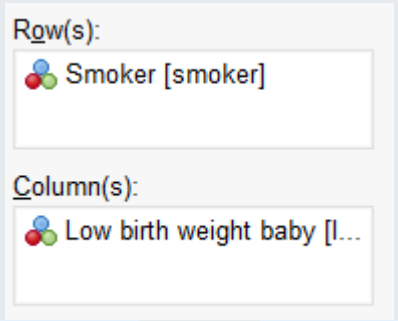
H_0: there is no association between the mother's smoking status and baby's birth weight
H_a: there is an association between the mother's smoking status and baby's birth weight

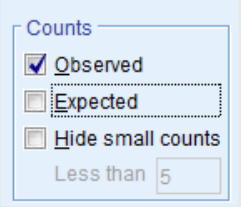
Assumptions:

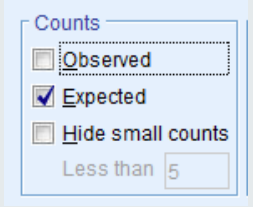
- The observations are randomly and independently drawn
- The number of cells with expected frequencies less than 5, are less than 20%
- The minimum expected frequency is at the very least 1.
- The observations are not paired

Output and Interpretations

Computations: 'Pearson's chi-square test'.







Smoker * Low birth weight baby Crosstabulation

Count

		Low birth weight baby		
		No	Yes	Total
Smoker	Non-smoker	15	5	20
	Smoker	9	13	22
Total		24	18	42

Smoker * Low birth weight baby Crosstabulation

Expected Count

		Low birth weight baby		
		No	Yes	Total
Smoker	Non-smoker	11.4	8.6	20.0
	Smoker	12.6	9.4	22.0
Total		24.0	18.0	42.0

$$\sum \frac{(O-E)^2}{E} = \frac{(15-11.4)^2}{11.4} + \frac{(5-8.6)^2}{8.6} + \frac{(9-12.6)^2}{12.6} + \frac{(13-9.4)^2}{9.4} = 5.05$$

Output and Interpretation Slide

Smoker * Low birth weight baby Crosstabulation

			Low birth weight baby		
			No	Yes	Total
Smoker	Non-smoker	Count	15	5	20
		% within Low birth weight baby	62.5%	27.8%	47.6%
	Smoker	Count	9	13	22
		% within Low birth weight baby	37.5%	72.2%	52.4%
Total		Count	24	18	42
		% within Low birth weight baby	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	4.972[a]	1	.026	.033	.027	
Continuity Correction[b]	3.677	1	.055			
Likelihood Ratio	5.104	1	.024	.033	.027	
Fisher's Exact Test				.033	.027	
Linear-by-Linear Association	4.853[c]	1	.028	.033	.027	.022
N of Valid Cases	42					

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.57.

b. Computed only for a 2x2 table

c. The standardized statistic is 2.203.

Among those mothers who smoked during pregnancy there was a higher proportion who had a baby of low birthweight compared to a baby of normal birthweight (72.2% versus 37.5%, respectively). This difference was statistically significant (Pearson χ^2=4.972, df=1, p=0.026).

Therefore, we conclude that mothers who smoked during pregnancy tend to have babies born with low birthweight than women who did not smoke during pregnancy. The variables 'Smoker' and 'Lowbirthwgt' are associated.

Quantifying the risk

Compare risk of smoking between babies classed as having a low birthweight and those who are not

Smoker * Low birth weight baby Crosstabulation

Count

		Low birth weight baby		
		No	Yes	Total
Smoker	Non-smoker	15	5	20
	Smoker	9	13	22
Total		24	18	42

The risk of an outcome is the number of times the outcome of interest occurs divided by the total number of possible outcomes.

For example: In the above study out of 22 mothers who smoked during pregnancy, there were 13 babies who were born with low birthweight. So, we get the risk of smoking during pregnancy and having a low birthweight baby by the following calculation:

Risk = 13 ÷ 22 = 0·59

Calculating the Risk Ratio (Relative Risk)

If we want to compare the effects of smoking during pregnancy and not smoking during pregnancy we could calculate the risk of having a baby of low birth weight for each group:

Smoker * Low birth weight baby Crosstabulation

Count

		Low birth weight baby		
		No	Yes	Total
Smoker	Non-smoker	15	5	20
	Smoker	9	13	22
Total		24	18	42

Risk of having baby of low birth weight in smokers = **13 ÷ 22 = 0·59**

Risk of having baby of low birth weight in non-smokers = **5 ÷ 20 = 0·25**

We can compare the risk for each of the groups using the risk ratio.

(Risk when smoker) ÷ (Risk when non-smoker)=
0·59 ÷ 0·25 = 2.36

So, the risk of having a low birthweight baby when the mother smoked through pregnancy is 2.36 times that of when the mother did not smoke during pregnancy.

Interpreting the Risk

Relative Risk = 1: The risk ratio equals one when the numerator and denominator are equal.

- This equivalence occurs when the probability of the event occurring in the exposure group equals the likelihood of it happening in the unexposed group.
- E.g: There is no association between mothers smoking status and a baby being born with a low birth weight.

Relative Risk > 1: The numerator is greater than the denominator in the risk ratio.

- Therefore, the event's probability is greater in the exposed group than in the unexposed group.
- E.g. If the RR = 1.4, the smoking status corresponds to a 40% greater probability of a mother having a child with low birthweight.

Relative Risk < 1: The numerator is less than the denominator in the risk ratio.

- Consequently, the probability of the event is lower for the exposed group than for the unexposed group.
- E.g. If the RR = 0.4, the smoking status corresponds to a 60% lower probability of a mother having a child with low birthweight.

Calculating the Odds

The odds in favour of a particular outcome is the number of times the outcome occurs divided by the number of times it doesn't occur.

If we want to compare the effects of smoking and
Not smoking during pregnancy we could calculate the odds
for each group:

Smoker * Low birth weight baby Crosstabulation

Count

		Low birth weight baby		
		No	Yes	Total
Smoker	Non-smoker	15	5	20
	Smoker	9	13	22
Total		24	18	42

Odds for having a baby of low birthweight when mother is a smoker = **13 ÷ 9 = 1.44**

Odds for having a baby of low birthweight when mother is a non-smoker = **5 ÷ 15 = 0.33**

We can compare the odds using the odds ratio. The odds ratio for having a baby of low birthweight when mother smokes during pregnancy compare to a mother who did not smoke during pregnancy

(Odds when smoker) ÷ (Odds when non-smoker) = $1.44\dot{4} \div 0.33\dot{3} = 4.33$

- So the odds of having a low birthweight baby when the mother smoked during pregnancy is about 4.36 times larger than the odds for mothers who did not smoke during pregnancy.

Interpreting the odds

OR = 1

Odds of 1 mean the outcome occurs at the same rate in both groups

- Exposure does not affect the odds of outcome
- E.g There is no difference in the odds of low birth rate between smokers and non-smokers.

OR < 1

Odds of less than 1 mean the outcome occurs less often in the first group than the second group

- Indicates that the exposure is associated with a decreased risk of developing the disease
- E.g if the odds ratio = 0.339 then the odds of a non-smoker having a low birth weight baby is a third (33.9%) of smokers.

OR >1

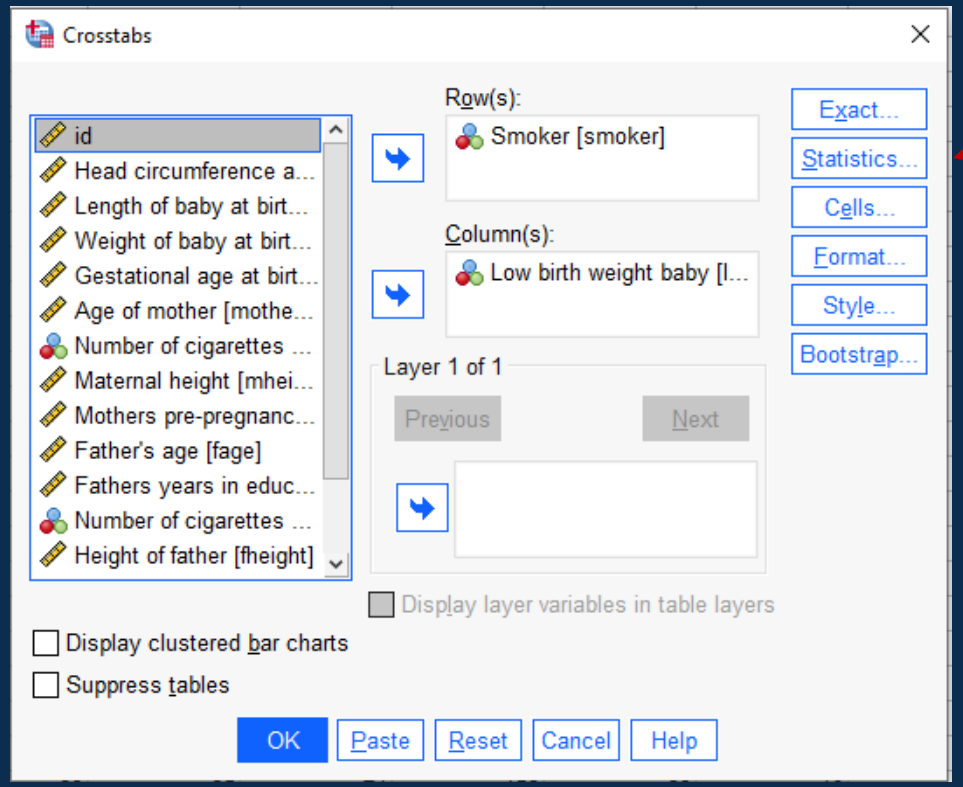
Odds of less than 1 mean the outcome occurs more often in the first group than the second group

- Indicates that the exposure is associated with an increased risk of developing the disease
- E.g. if the odds ratio = 1.5 then the odds of smokers having a low birth weight baby is 1.5 times that of the odds of non-smokers.

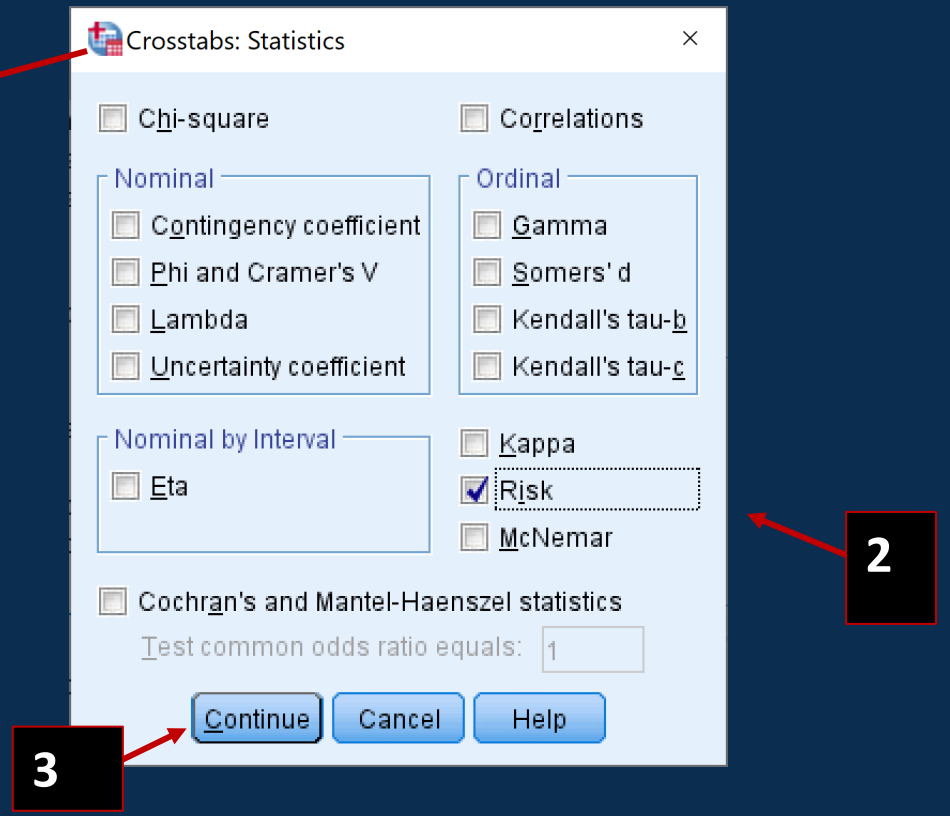
SPSS Slide: ‘how to’

Calculate the risk of having a baby of low birthweight if mothers smoked during pregnancy versus if they did not.

Step 1: Create a contingency table



Step2: Click on ‘**Statistics**’ and Choose **“Risk”**



Output and Interpretation

Risk Estimate

	Value	95% Confidence Interval Lower	Upper
Odds Ratio for Smoker (Non-smoker / Smoker)	4.333	1.156	16.248
For cohort Low birth weight baby = No	1.833	1.045	3.217
For cohort Low birth weight baby = Yes	.423	.184	.975
N of Valid Cases	42		

Odds Ratio

Risk Ratio (Relative Risk)

The risk ratio given here is 0.423, it is the risk of having a low birthweight baby when the mother did not smoke through pregnancy. To understand the risk of a mother who smoked, we take the reciprocal 1/0.423 = 2.36. So, the risk of a mother who smokes to have a low-birth-weight baby is 2.36 times that of a non-smoker.

Risk ratios and Odds ratios

Case-control studies

- The risk ratio cannot be used in a case-control study, the odds ratio can be used. Risk ratios cannot be used in studies where selection of subjects is based on the outcome.

Rare outcomes

- When an outcome is rare the risk ratio and odds ratio will be approximately equal.

Clinical practitioners often prefer the risk ratio due to its more direct interpretation. Statisticians tend to prefer the odds ratio as it applies to a wide range of study designs, allowing comparison between different studies and meta-analysis based on many studies. It also forms the basis of **logistic regression.**

Knowledge Check

Q1: In the paper Caries prevalence in northern Scotland before and 5 years after, water defluoridation (Stephen et al., 1987, BDJ 163: 324-326) the researchers studied two groups of children in Wick; one group whilst the water was fluoridated and one group after defluoridation.

	Caries		
Water Type	**Yes**	**No**	**Total**
Fluoridated	77	29	106
Non-fluoridated	95	31	126
Total	172	60	232

Calculate the Risk Ratio and the Odds Ratio for the risk of Caries considering the water is Fluoridated.

Knowledge Check

Q2: Suppose we conducted a randomised trial to investigate the effect of citalopram on depression. A group of patients who are at risk for depression are randomly assigned to either a placebo or citalopram. At the end of one year, the number of patients suffering with depression is recorded.

	Depression		
Group	Yes +	No -	Total
Placebo	20	80	100
Citalopram	15	135	150
Total	35	215	250

Calculate the Risk Ratio and the Odds Ratio for the risk of depression for placebo versus citalopram.

Q1: Knowledge Check Solutions

Risk Ratio

If we want to compare the effects of fluoridated and non-fluoridated water we could calculate the risk of having caries for each group:

Risk of having caries when water is fluoridated = 77 ÷ 106 = 0·73

Risk of having caries when water is not fluoridated = 95 ÷ 126 = 0·75

We can compare the risk for each of the groups using the risk ratio. The risk ratio for being caries free when water is fluoridated compared to when it is not fluoridated is:

(Risk when fluoridated) ÷ (Risk when not fluoridated) = 0·73 ÷ 0·75 = 0·96

So, the risk of having caries when the water is fluoridated is only 0·96 that of when the water is not fluoridated.

Q1: Knowledge Check Solutions

Odds Ratio

If we want to compare the effects of fluoridated and non-fluoridated water we could calculate the odds for each group:

Odds for having caries when water is fluoridated = 77 ÷ 29 = 2·66

Odds for having caries when water is not fluoridated = 95 ÷ 31 = 3·06

We can compare the odds using the odds ratio. The odds ratio for having caries when water is fluoridated compared to when it is not fluoridated is:

(Odds when fluoridated) ÷ (Odds when not fluoridated) = 2·66 ÷ 3·06 = 0·87

So, the odds of having caries when the water is fluoridated are about 90% those of when the water is not fluoridated.

Q2: Knowledge Check Solutions

Risk Ratio

If we want to compare the effects of Citalopram and placebo we could calculate the risk of having depression for each group:

Risk of having depression when on Citalopram = 15 ÷ 150 = 0·1

Risk of having depression when on placebo = 20 ÷ 100 = 0·2

We can compare the risk for each of the groups using the risk ratio. The risk ratio for being depression free when on Citalopram compared to placebo is:

(Risk when on Citalopram) ÷ (Risk when on placebo) = 0·1 ÷ 0·2 = 0·5

So the risk of having depression halves when the subject is on Citalopram compared to placebo.

Q2: Knowledge Check Solutions

Odds Ratio

If we want to compare the effects of Citalopram and placebo we could calculate the odds of having depression for each group:

Odds for having depression when on Citalopram = 15 ÷ 135 = 0.11

Odds for having depression when on placebo = 20 ÷ 80 = 0.25

We can compare the odds using the odds ratio. The odds ratio for having depression when on Citalopram compared to placebois:

(Odds when citalopram) ÷ (Odds when on placebo) = 0.11 ÷ 0.25 = 0·44

So the odds of having depression when on Citalopram are about 45% of those of on placebo. Or another interpretation (1/odds ratio). The odds of having depression when on placebo is 2.3 times the odds of those on citalopram

References

Altman, D.G., 1991. Practical statistics for medical research, pp. 49-50, 250-253, 259-271. Chapman and Hall, London.

Bowers, D., 1997. Statistics further from scratch : for health care professionals, pp. 156-157, 162-163. John Wiley & Sons, Chichester.

Field, Andy. Discovering statistics using IBM SPSS statistics. Sage, 2013. (Chapter 19)

Agresti, Alan. Categorical data analysis. John Wiley & Sons, 2014.

Thank you

Contact details/for more information:

Zahra Abdulla

Zahra.abdulla@kcl.ac.uk

Department of Biostatistics and Health Informatics (BHI)

IoPPN