Logistic regression for pharmacoepidemiology and DUR

Associate Professor Lisa Pont
FISPE

University of Technology Sydney

At the end of this session you will be able to

- Describe when logistic regression is used in pharmacoepidemiology and DUR
- Understand the data assumptions for logistic regression
- Be able to calculate unadjusted effect measures using logistic regression in SPSS
- Interpret SPSS logistic regression output
- Adjust for covariates in logistic regression using SPSS
- Know how to test for assumptions in SPSS
- Interpret SPSS logistic regression output adjusted for covariates

---

Logistic Regression in Pharmacoepi and DUR

- Regression model used with binary outcomes
  - Binary outcome= 2 levels
  - Eg: Yes/No, Died/Survived, Male/female
- Predictive versus casual modelling
- Used to estimates odds (or risk) ratios
  - Crude
  - Adjusted for confounders

Logistic regression versus linear regression

Linear regression	Logistic regression
• y=mx + b	log(odds of disease)= a + (b x exposure)
• Normal distribution of predictor variables	a (intercept) = log(odds of disease in the unexposed) - similar to constant term in linear regression. b (slope) = log(odds ratio) - similar to coefficient term in linear regression.

Logistic regression assumptions

1. binary logistic regression requires the dependent variable to be **binary**
2. The observations are **independent** of each other.
3. There to be **little or no multicollinearity** among the independent variables.
4. linearity of independent variables and log odds.

Our dataset

The dataset used in this session is hypothetical data based on the sample datasets provided with SPSS.

A cross sectional study was conducted in a large teaching hospital to determine if use of thrombolytics or clot busting drugs, was associated with mortality among patients coming to the emergency department with a myocardial infarction (heart attack). Data from all patients presetting to ED with a heart attach was collected over a 6 month period.

Using SPSS: Groupwork Activity 1 & 2

1. Define your question

Based on the study aim: Are thrombolytic medicines associated with decreased mortality in heart attack

What is the outcome

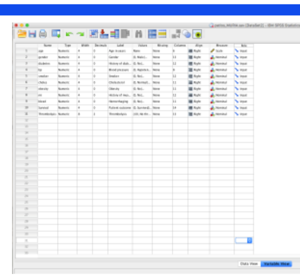
What is the exposure

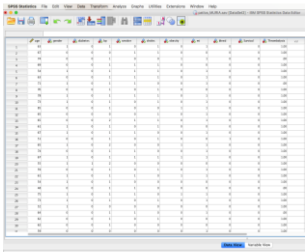
What potential confounders or effect modifiers should you consider?

2. Understanding your data

Open the dataset **patlos_MURIA**

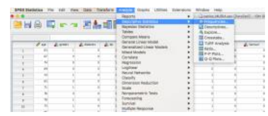
- How many variables are in the dataset?
- How many patients are included in the dataset?
- What is the name of the exposure variable?
- What is the name of the outcome variable?



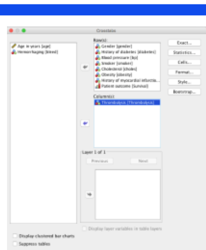


Describing your study population

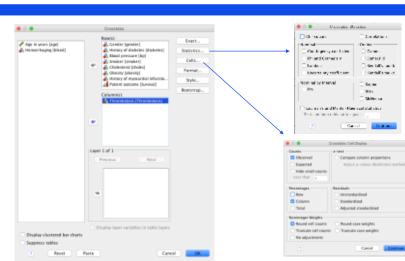
SPSS Command	
Frequency	Counts and proportions of all values within a variable (categorical data)
Descriptives	Count, max, min, mean and standard deviation (continuous data)
Explore	Summary measures of a continuous variable against categories of another variable
Crosstabs	Cross tabulates counts/ proportions in one variable against another



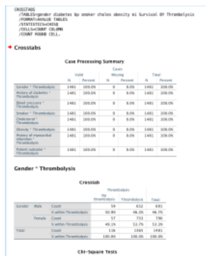
ispe International Society for Pharmacoepidemiology UTS University of Technology Sydney



ispe International Society for Pharmacoepidemiology UTS University of Technology Sydney



ispe International Society for Pharmacoepidemiology UTS University of Technology Sydney



International Society for Pharmacoepidemiology

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Using SPSS: Group activity 3

Characteristic	Exposed participants (N=)	Unexposed participants (N=)
Age		
Gender		
Past MI		
Diabetes		
Hypertension		
Cholesterol		
Obesity		
Smoker		

International Society for Pharmacoepidemiology

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Logistic regression in SPSS

Here we will use the **patlos_MURIA** dataset to explore if thrombolytic medicines are associated with bleeding in patient presenting to the ED with a MI

OUTCOME?

EXPOSURE?

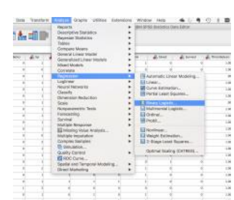
International Society for Pharmacoepidemiology

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Logistic regression in SPSS

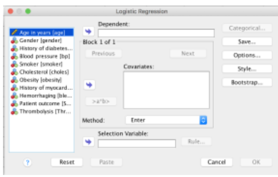
Regression

>Binary Logistic



International Society for Pharmacoepidemiology

UTS UNIVERSITY OF TECHNOLOGY SYDNEY



Dependent=OUTCOME

Covariates =EXPOSURE and CONFOUNDFERS

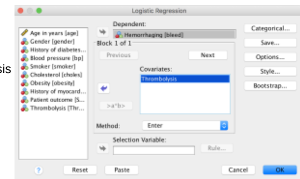
Selection variable =used to identify a subset of cases (not generally relevant for causal modelling)

ispe International Society for Pharmacoepidemiology UTS

Is thrombolysis associated with bleeding?

Outcome=Bleed

Exposure=thrombolysis



ispe International Society for Pharmacoepidemiology UTS

Choice of method

Enter. A procedure for variable selection in which all variables in a block are entered in a single step.

The significance values in your output are based on fitting a single model. Therefore, the significance values are generally invalid when a stepwise method is used.

ispe International Society for Pharmacoepidemiology UTS

Categorical covariates

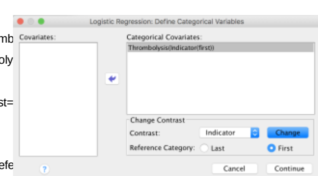
0=no thromb

1=thromboly

First lowest=

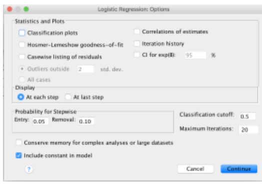
Last

Highest=refe



ispe International Society for Pharmacoepidemiology UTS

Options



- 95% CI around estimates

International Society for Pharmacoepidemiology

UTS

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES bleed
  /METHOD=ENTER Thrombolysis
  /CONTRAST (Thrombolysis)=Indicator(1)
  /PRINT=GOODFIT CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

International Society for Pharmacoepidemiology

UTS

Interpreting SPSS Output

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	1481	100.0
	Missing Cases	0	.0
	Total	1481	100.0
Unselected Cases		0	.0
Total		1481	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Thrombolysis	No thrombolysis	116	.000
	Thrombolysis	1365	1.000

International Society for Pharmacoepidemiology

UTS

Block 0: Beginning Block

Classification Table[a,b]

Observed			Predicted Bleed No	Predicted Bleed Yes	Percentage Correct
Step 0	Bleed	No	1470	0	100.0
		Yes	11	0	.0
	Overall Percentage				99.3

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-4.895	.303	261.627	1	.000	.007

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Thrombolysis(1)	.024	1	.876
	Overall Statistics		.024	1	.876

International Society for Pharmacoepidemiology

UTS

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.023	1	.879
	Block	.023	1	.879
	Model	.023	1	.879

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	129.752[a]	.000	.000

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	.000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Bleed = No		Bleed = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	1355	1355.000	10	10.000	1365
	2	115	115.000	1	1.000	116

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Thrombolysis(1)	-.164	1.053	.024	1	.876	.849	.108	6.688
	Constant	-4.745	1.004	22.320	1	.000	.009		

a. Variable(s) entered on step 1: Thrombolysis.

B=log(OR)
ExpB=OR

OR=0.849
P=0.876
95% CI=0.108-6.688

ispe International Society for Pharmacoepidemiology
UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Is thrombolysis associated with increased bleeding in MI

OR=0.849 (95% CI: 0.108-6.688), p=0.849

There is no evidence that thrombolysis is associated with mortality. Thrombolysis was associated with a 15% reduction in the odds of bleeding but the 95% confidence interval indicates that the true value of lies between a 90% reduction and a 660% increase

ispe International Society for Pharmacoepidemiology
UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Using SPSS Groupwork activity 4

Using logistic regression, what is the crude odds ratio for the effect of thrombolysis on mortality?

ispe International Society for Pharmacoepidemiology
UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Using SPSS: groupwork activity 5 Confounding

Age, gender, previous MI and smoking status are all considered possible confounders.

- Do any or all of them meet the criteria to be a confounder?

Hint: use crosstab to explore the relationship between potential confounder and exposure and potential confounder and outcome.

For age you will need to use a different test

Using SPSS: Groupwork activity 6

Using logistic regression, calculate the odds ratio for the association between thrombolysis and mortality adjusted for each potential confounder in turn.

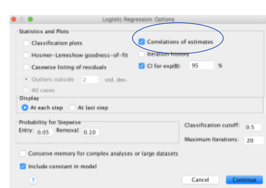
	OR (95% CI)
Unadjusted	
Age	
Gender	
Smoking	
Past MI	

Using SPSS-Groupwork activity 7

What is your final model?

- Include age and gender as a priori confounders as well as any other confounders from question 6.

Assumptions-multicolinearity



Testing multicolinearity

Correlation Matrix

		Constant	Thrombolysis (1)	Gender(1)	Age in years	History of myocardial infarction(1)
Step 1	Constant	1.000	-.271	-.124	-.895	-.111
	Thrombolysis(1)	-.271	1.000	-.047	-.043	-.050
	Gender(1)	-.124	-.047	1.000	-.004	-.009
	Age in years	-.895	-.043	-.004	1.000	-.137
	History of myocardial infarction(1)	-.111	-.050	-.009	-.137	1.000

Final model

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Thrombolysis(1)	-.896	.255	12.330	1	.000	.408	.248	.673
	History of myocardial infarction(1)	1.010	.216	21.804	1	.000	2.746	1.797	4.197
	Age in years	.008	.010	.747	1	.387	1.008	.989	1.028
	Gender(1)	.165	.172	.916	1	.339	1.179	.841	1.653
	Constant	-2.689	.669	16.138	1	.000	.068		

a. Variable(s) entered on step 1: Thrombolysis, History of myocardial infarction, Age in years, Gender.

After adjusting for age, gender and past MI, Thrombolysis is associated with a 59% reduction (95% CI: 0.248-0.673) in the odds of mortality in patients presenting to the emergency department with a MI.

Note: This is hypothetical data for the purposes of this workshop and does not represent data.

ispe International Society for Pharmacoepidemiology
UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Logistic regression-Sample size

A general guideline is that you need a minimum of **10 cases** with the (least frequent) **outcome for each independent variable** in your model.

For example, if you have 5 independent variables and the expected probability of your least frequent outcome is .10, then you would need a minimum sample size of 500 (10*5 / .10).

ispe International Society for Pharmacoepidemiology
UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Thank you