

Introduktion til Logistisk Regression

Morten Frydenberg, Afd. f. Biostatistik

© 2005 Institut for Biostatistik, Århus Universitet

MPH 1. studieår Specialmodul 4

Cand. San. uddannelsen 1. studieår

3. gang:

- Interaktion/effektmodifikation mellem to binære variable.
- Interaktion/effektmodifikation mellem kategoriske forklarende variable med mere to kategorier.
- Interaktion/effektmodifikation mellem en binær og en kontinuert variabel.

Sammenhæng mellem SYS160 og CHD

Fra 1. dag:

	OR	Estimator ln(OR)	Se(ln(OR))
Mænd	2.5586	0.9395	0.2182
Kvinder	3.3552	1.2105	0.2177
Fælles	2.9308	1.0753	0.1537

Under antagelse af :

$$OR_{\text{SYS160 for MÆND}} = OR_{\text{SYS160 for KVINDER}}$$

Ingen **interaktion!**= Ingen **effektmodifikation!**

Er dette en rimelig antagelse ?

Back to basics: Sammenligning af to odds ratioerBetragt forholdet : $ROR = \frac{OR_M}{OR_K}$ **Relativ Odds Ratio**

$$OR_M = ROR \cdot OR_K$$

Hvor meget større er OR for mænd

$$\widehat{ROR} = \frac{\widehat{OR}_M}{\widehat{OR}_K} = \frac{2.5586}{3.3552} = 0.7626$$

Bedste gæt: OR for mænd er 0.76 gange den for kvinder !

Hvad er usikkerheden på dette estimat ?

Hvad er usikkerheden på : $\ln(\widehat{ROR}) = \ln(\widehat{OR}_M) - \ln(\widehat{OR}_K)$?Usikkerheden på : $\ln(\widehat{ROR}) = \ln(\widehat{OR}_M) - \ln(\widehat{OR}_K)$?

$$se(\ln(\widehat{ROR})) = \sqrt{se(\ln(\widehat{OR}_M))^2 + se(\ln(\widehat{OR}_K))^2}$$

$$= \sqrt{0.2182^2 + 0.2177^2}$$

$$= 0.3082$$

95% sikkerhedsinterval for $\ln(ROR)$:

$$\ln(0.7626) \pm 1.96 \cdot 0.3082$$

$$-0.8752 ; 0.3331$$

95% sikkerhedsinterval for ROR :

$$\exp(-0.8752) ; \exp(0.3331)$$

$$0.4168 ; 1.3952 \quad \leftarrow \text{Indeholder 1.0}$$

Resumé: $ROR = \frac{OR_M}{OR_K}$, $\widehat{ROR} = 0.7626$, CI: 0.4168, 1.3952Til beregninger: $\ln(\widehat{ROR}) = -0.2710$, $se(\ln(\widehat{ROR})) = 0.3082$ **Hypotese** : Ingen interaktion/effektmodifikation

$$\Leftrightarrow OR_M = OR_K \Leftrightarrow ROR = 1 \Leftrightarrow \ln(ROR) = 0$$

Estimat Hypotese

$$\text{Test: } z = \frac{\ln(\widehat{ROR}) - 0}{se(\ln(\widehat{ROR}))} = \frac{-0.2710}{0.3082} = -0.88$$

Usikkerhed

$$0.88^2 = 0.77$$

P-værdi: Slå op i en standard normal fordeling : $p = 0.3791$ Fra cc: Test of homogeneity (M-H $\chi^2(1) = 0.77$ $\text{Pr} > \chi^2 = 0.3791$)Resumé: $ROR = \frac{OR_M}{OR_K}$, $\widehat{ROR} = 0.7626$, CI: 0.4168, 1.3952Hypotese: $ROR = 1$ $p = 38\%$ **Konklusion**

- Data er forligelig med hypotesen om ingen effektmodifikation !
- Vi har et rimeligt bredt sikkerhedsinterval og dermed lille viden om problemet !

2 spørgsmål fra salen:

- Kan dette laves vha. logistisk regression ?
- Hvordan får man STATA til at gøre arbejdet ?

Estimation af ROR og test af ingen interaktion vha. log. reg.

1. gang så vi på modellen **uden interaktion**:

$$\ln(odds(SYS160, SEX)) = \beta_0 + \beta_1 \cdot SYS160 + \beta_2 \cdot MALE$$

Lad nu *SYS160M* være 1 for mænd med SBP>160
0 for alle andre

Modellen **med interaktion** kan skrives:

$$\ln(odds(SYS160, SEX)) = \beta_0 + \beta_1 \cdot SYS160 + \beta_2 \cdot MALE + \beta_3 \cdot SYS160M$$
$$\ln(odds(0, k)) = \beta_0 + \beta_1 \cdot 0 + \beta_2 \cdot 0 + \beta_3 \cdot 0 = \beta_0$$
$$\ln(odds(1, k)) = \beta_0 + \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 = \beta_0 + \beta_1$$
$$\ln(odds(0, m)) = \beta_0 + \beta_1 \cdot 0 + \beta_2 \cdot 1 + \beta_3 \cdot 0 = \beta_0 + \beta_2$$
$$\ln(odds(1, m)) = \beta_0 + \beta_1 \cdot 1 + \beta_2 \cdot 1 + \beta_3 \cdot 1 = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

$$\ln(odds(0, k)) = \beta_0 \Rightarrow \beta_0 = \ln(odds(0, k))$$
$$\ln(odds(1, k)) = \beta_0 + \beta_1$$
$$\ln(odds(0, m)) = \beta_0 + \beta_2$$
$$\ln(odds(1, m)) = \beta_0 + \beta_1 + \beta_2 + \beta_3$$
$$\beta_1 = \ln(odds(1, k)) - \ln(odds(0, k)) = \ln(OR_K)$$
$$\beta_1 + \beta_3 = \ln(odds(1, m)) - \ln(odds(0, m)) = \ln(OR_M)$$
$$\beta_2 = \ln(odds(0, m)) - \ln(odds(0, k)) = \ln(OR_{K\emptyset n \text{ for } SBP < 160})$$

Dvs.

$$\beta_1 = \ln(OR_K)$$
$$\beta_2 = \ln(OR_{K\emptyset n \text{ for } SBP < 160})$$
$$\beta_3 = \ln(OR_M) - \ln(OR_K) = \ln(ROR)$$

ln(odds) i referencegruppen
kvinder uden højt blodtryk.

$$\ln(ROR) = \ln\left(\frac{OR_{SYS160 \text{ for M\ae}nd}}{OR_{SYS160 \text{ for Kvinder}}}\right)$$
$$\ln(odds(SYS160, SEX)) = \beta_0 + \beta_1 \cdot SYS160 + \beta_2 \cdot MALE + \beta_3 \cdot SYS160M$$
$$\ln(OR_K) = \ln(OR_{SYS160 \text{ for KVINDER}})$$
$$\ln(OR_{K\emptyset n \text{ for } SBP < 160})$$

Bemærk:

Både koefficienten til *SYS160* og *MALE* har ændret betydning !

Fortolkning afhænger af hvad der ellers er med i modellen !!

STATA

```
char sex[omit]2
xi:logit chdever i.sys160 i.sex i.sex*i.sys160
eller
xi:logit chdever i.sys160 i.sex i.sex*i.sys160, or
på kort form
xi:logit chdever i.sex*i.sys160
eller
xi:logit chdever i.sex*i.sys160, or
```

Start på output :

i.sex	_Isex_1-2	(naturally coded; _Isex_2 omitted)
i.sys160	_Isys160_0-1	(naturally coded; _Isys160_0 omitted)
i.sex*i.sys160	_IsexXsys_#_#	(coded as above)

Iteration 0: log likelihood = -675.62283
Iteration 4: log likelihood = -638.29101

Logit estimates

	Number of obs	=	1363
	LR chi2(3)	=	74.66
	Prob > chi2	=	0.0000
	Pseudo R2	=	0.0553

Log likelihood = -638.29101

chdever	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_Isex_1	.96847	.1819571	5.32	0.000	.61184 1.3251
_Isys160_1	1.2105	.2175571	5.56	0.000	.78410 1.636916
IsexXsys~1	-.2710	.3081473	-0.88	0.379	-.87501 .3328987
_cons	2.2321	.1488151	-15.00	0.000	-2.5238 1.94049

$\ln(ROR)$ $se(\ln(ROR))$ test CI

ROR

chdever	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
_Isys160_1	3.3552	.42994	5.56	0.000	2.19045 5.139293
_Isex_1	2.6339	.41259	5.32	0.000	1.843823 3.762561
IsexXsys~1	.76257	.34904	-0.88	0.379	.4168551 1.395006

Test for ingen interaktion/effektmodifikation:

I outputtet et Wald test: $z = -0.88$ $p = 38\%$

Dette kan også findes således: `testparm _IsexX*`

```
( 1) _IsexXsys_1_1 = 0
      chi2( 1) = 0.77
      Prob > chi2 = 0.3791
```

Et alternativt test : Likelihood ratio test:

```
quietly xi: logit chdever i.sex*i.sys160
estimates store model1
quietly xi: logit chdever i.sex i.sys160
estimates store model2
lrtest model1 model2
```

likelihood-ratio test	LR chi2(1)	=	0.78
(Assumption: model2 nested in model1)	Prob > chi2	=	0.3786

Ingen forskel

Hvis fokus er på køn: 13

$\ln(\text{odds}(0,k)) = \beta_0 \Rightarrow \beta_0 = \ln(\text{odds}(0,k))$

$\ln(\text{odds}(1,k)) = \beta_0 + \beta_1$

$\ln(\text{odds}(0,m)) = \beta_0 + \beta_2$

$\ln(\text{odds}(1,m)) = \beta_0 + \beta_1 + \beta_2 + \beta_3$

$\beta_1 = \ln(\text{odds}(1,k)) - \ln(\text{odds}(0,k)) = \ln(OR_K)$

$\beta_2 + \beta_3 = \ln(\text{odds}(1,m)) - \ln(\text{odds}(1,k)) = \ln(OR_{Køn \text{ for } SBP > 160})$

$\beta_2 = \ln(\text{odds}(0,m)) - \ln(\text{odds}(0,k)) = \ln(OR_{Køn \text{ for } SBP < 160})$

Dvs. $\beta_1 = \ln(OR_K)$

$\beta_2 = \ln(OR_{Køn \text{ for } SBP < 160})$

$\beta_3 = \ln(OR_{Køn \text{ for } SBP > 160}) - \ln(OR_{Køn \text{ for } SBP < 160}) = \ln(ROR_{Køn})$

Hvis fokus er på køn: 14

$\ln(ROR) = \ln\left(\frac{OR_{Køn \text{ for } SBP > 160}}{OR_{Køn \text{ for } SBP < 160}}\right)$

$\ln(\text{odds}(SYS160, SEX)) = \beta_0 + \beta_1 \cdot SYS160 + \beta_2 \cdot MALE + \beta_3 \cdot SYS160M$

$\ln(OR_K) = \ln(OR_{SYS160 \text{ for } KVINDER})$

$\ln(OR_{Køn \text{ for } SBP < 160})$

Bemærk:

Både koefficienten til **SYS160** og **MALE** har ændret betydning !

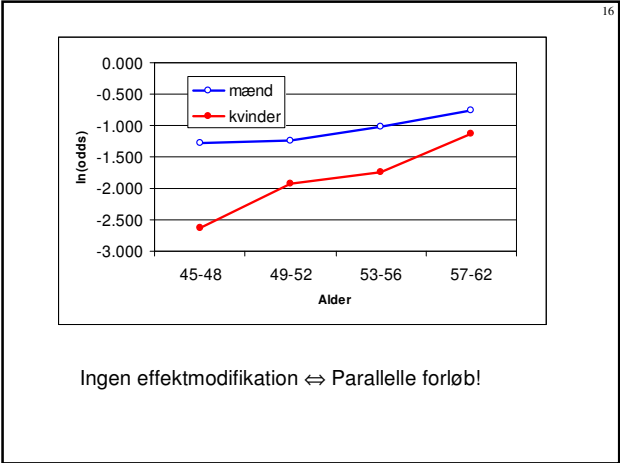
Fortolkning afhænger af hvad der ellers er med i modellen !!

Interaktion mellem SEX og AGE4 15

2 niveauer 4 niveauer

I alt $8 = 2 \cdot 4$ forskellige typer personer:

Alder	Antal		Alder	Odds for CHD	
	Mænd	Kvinder		Mænd	Kvinder
45-48	180	179	45-48	0.2766	0.0719
49-52	156	203	49-52	0.2893	0.1469
53-56	144	174	53-56	0.3535	0.1757
57-62	163	164	57-62	0.4685	0.3226



`char sex[omit] 2`
`char age4[omit] 0` 17

Alder	Odds for CHD		Ref.	Alder	OR i forhold til reference	
	Mænd	Kvinder			Mænd	Kvinder
45-48	0.2766	0.0719		45-48	3.8493	1.0000
49-52	0.2893	0.1469		49-52	4.0255	2.0443
53-56	0.3535	0.1757		53-56	4.9890	2.4448
57-62	0.4685	0.3226		57-62	6.5195	4.4892

Alder	OR indenfor aldersgruppe		$ROR_{Køn}$
	Mænd	Kvinder	
45-48	3.8493	1.0000	1.0000
49-52	1.9692	1.0000	0.5117
53-56	2.0406	1.0000	0.5303
57-62	1.4523	1.0000	0.3774

Ingen effektmodifikation hvis disse er tæt på 1

STATA 18

```
xi:logit chdever i.age4 i.sex i.sex*i.age4
eller
xi:logit chdever i.age4 i.sex i.sex*i.age4,or
på kort form
xi:logit chdever i.sex*i.age4
eller
xi:logit chdever i.sex*i.age4,or
```

Start på output :

Referencer	
i.sex	_isex_1-2 (naturally coded; _isex_2 omitted)
i.age4	_lage4_0-3 (naturally coded; _lage4_0 omitted)
i.sex*i.age4	_isexXage_# (coded as above)

Iteration 0: log likelihood = -675.62283
Iteration 4: log likelihood = -648.48265

Logit estimates

Number of obs	=	1363
LR chi2(7)	=	54.28
Prob > chi2	=	0.0000
Pseudo R2	=	0.0402

Log likelihood = -648.48265

