Epidemiology VHM 811 Course
Fall 2008, Atlantic Veterinary College, PEI
Henrik Stryhn

Index of Lecture 10b

(revised: change on page 1)

Practical Information

Major news:

- next session on Wednesday next week (because Tuesday is Remembrance Day),
- suggested preparation for Wednesday: home assignment 2007 (link from webpage), Questions 1–4.

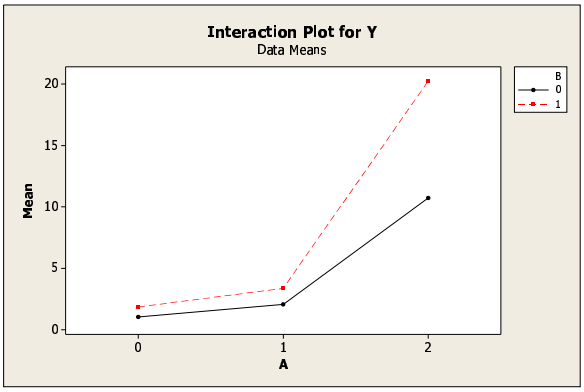
Today's session will be split in two parts

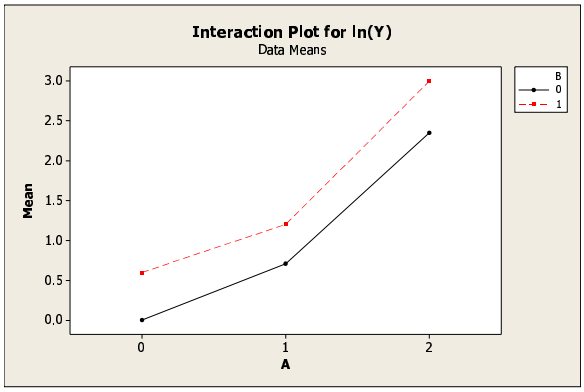
- first part, here and probably about 1.5h:
  - questions for material from Tuesday?,
  - some lecturing:
    - one extra slide on confounding (from L10a),
    - interaction (definition, detection and handling),
  - review of prepared Problem 13.7,
- second part, in the computer lab (225N):
  - individual work on problems,
  - more time for questions and discussion.

Statistical Interaction

Interaction — a pattern in the effects of two variables (factors) possible in most multivariable statistical models:

- the combined effect of two factors is not predictable from the isolated effect of each of them when examined separately,
- effect measure modification[1]: the effect of the first factor depends on the level of the second factor, (or vice versa),
- common notation: $A*B \sim$ interaction between A and B,
- non-additive effect of two factors,
- non-parallel curves in an interaction plot:
(simulated data, 10 replications per group)





Statistical interaction depends on:

- the scale of analysis (compare left/right figures above),
- the type of statistical model.

[1] Rothman & Greenland (1998) argue that this term is preferable to interaction when referring to statistical interaction, in order to distinguish it from other types of interaction.

Assessing Interaction for Risk

Stratification approach includes statistical testing for homogeneity of risk measures across strata $\sim$ no interaction $E * Z$, e.g. a "Wald-type" test for the odds-ratio:

$$X^2_{\text{hom}} = \sum_j \frac{[\ln(\text{OR}_j) - \ln(\text{OR}_{\text{MH}})]^2}{\widehat{\text{Var}}(\ln(\text{OR}_j))} \approx \chi^2\text{-distribution}(J-1)$$

under the null hypothesis of same (population) OR's in all J strata, and where $\widehat{\text{Var}}(\ln(\text{OR}_j))$ is computed as the sum of inverse cell counts in the table for stratum j.[2]

Key message: in evidence of interaction/heterogeneity, report stratum-specific measures instead of M-H estimate!

Stata homogeneity tests, depending on risk measure:

- OR: Wald-type test (default), Breslow-Day test (`bd`), Tarone test (`tarone`, modification of B-D test),[3]
- RR: Wald-type test,
- RD: no test available; for two strata, a test may be constructed from RD estimates and standard errors,
- tests may suffer from low power if the tables for each stratum are sparse $\Rightarrow$ some authors recommend to increase the significance level (e.g. to 0.10).

[2] In formulae, the estimated variance is: $(1/a_{1j}) + (1/a_{0j}) + (1/b_{1j}) + (1/b_{0j})$.

[3] In a multivariable approach, one may also use the Wald or likelihood-ratio test for the interaction term in a logistic regression model; all tests are fairly similar.

Interactions for FEEDLOT Data

Addition to previous analysis — assessment of interaction between BRSV (Mh) and province/IBR in their effect on BRD (Examples 13.6-13.7):

Variable	Stratum	OR	X^2	P-value
province	Alberta	2.75	0.566	
	Ontario	1.51	0.900	
	combined	2.19	1.466	0.23
IBR	+	2.60	0.779	
	−	1.40	0.945	
	combined	1.97	1.724	0.19

Comments:

- X^2_{hom} calculations for demonstration only (should be done by computer),
- both X^2_{hom} test statistics are clearly non-significant $\Rightarrow$ no evidence (and hardly any indication) of interactions involving BRSV,
- OR's are not too different across strata, and all in the same direction,
- conclusion: use of OR_{MH} (or crude OR) justified in both settings.

Interaction for Risk — Example 13.9

`Nocardia` data (for VER):

- subset of real dataset on Nocardia mastitis in Nova Scotia dairy herds collected in 1989,
- case-control study design with 54 case and control herds,
- purpose of study: evaluate the association between exposure variables and case-control status of the herds,
- information collected by herd questionnaires:

Variable	Description	Values
casecont	herd status for Nocardia mastitis	0/1 (control/case)
dcpct	percent of dry cows treated	0–100 %
dneo	use of dry-cow tx product w. Neomycin	0/1 (no/yes)
dclox	use of dry-cow tx product w. Cloxacillin	0/1 (no/yes)

Results for Neomycin and Cloxacillin: (E & Z)

Cloxa-cillin	Case-control	Neomycin 1	Neomycin 0	OR for `dneo`	Homogen. for `dneo`
1	1	5	3		
1	0	10	9	1.50	$X^2_{\text{hom}} = 6.44$
0	1	44	2		$(P = 0.011)$
0	0	15	20	29.3	
OR for `dclox`		0.17	3.33		
Homogen. for `dclox`		$X^2_{\text{hom}} = 6.41$ (P = 0.011)			

- clear interaction ⇒ M-H estimates should not be used!

Interaction in Biology and Epidemiology

Biological interaction:

- interdependent operation of factors to produce (or prevent) an effect,
- sufficient cause model offers framework for interpreting interaction, e.g.,[4]
  - ∗ effect occurs in the presence of both factors but not with only one factor present (synergism[5]),
  - ∗ effect occurs in the presence of one factor and the absence of another factor, but not when both factors are present (antagonism[5]).

Interactions for risk — different concepts arise for different risk measures: (see Appendix for details)

$$\begin{aligned} &\text{additive interaction} && : && \mathrm{RD}_{11} \neq \mathrm{RD}_{10} + \mathrm{RD}_{01}, \\ &\text{multiplicative interaction I} && : && \mathrm{RR}_{11} \neq \mathrm{RR}_{10} \cdot \mathrm{RR}_{01}, \\ &\text{multiplicative interaction II} && : && \mathrm{OR}_{11} \neq \mathrm{OR}_{10} \cdot \mathrm{OR}_{01}, \end{aligned}$$

and "no interaction" can exist only at one scale at a time.

- ∗ *additive scale*: used for public-health impact assessment,
- ∗ *multiplicative scale*: used for multivariable modelling.

[4] For an alternative, counterfactual framework, see Rothman & Greenland (1998), Chapter 18.

[5] According to Michael Thrusfield, the terms synergism and antagonism are best reserved for biological interactions, or interactions with a causal interpretation; Thrusfield (2005), *Veterinary Epidemiology*, 3rd ed.

Scales for Interaction — Example 13.8

Demonstration example for interactions at different scales:

- assume two dichotomous predictors (Mh and BRSV) for disease (BRD),
- assume set effects of each predictor separately (top of table), and consider different scenarios for their combined effect (bottom of table):

Mh	BRSV	BRD[a]	Risk	RD	Type[b]	RR	Type[b]	OR	Type[b]
0	0	1	.001	–	–	–	–	–	–
0	1	20	.020	.19	–	20	–	20.4	–
1	0	10	.010	.009	–	10	–	10.1	–
1	1	100							
1	1	29							
1	1	200							
1	1	171							

[a] BRD measured in cases per 1000 animals

[b] Type of combined effect:
S=synergistic, A=antagonistic, 0=no interaction

- as expected, "no interaction" does not happen simultaneously at different scales.

Stata Do-file

```
* Example 13.1
insheet using "eg13_01.csv"
cs brd mh [fw=n], or by(brsv)

* Diabetes trial
use http://www.stata-press.com/data/r10/ugdp, clear
cs case exposed [fw=pop], by(age)

* feedlot data, Example 13.6
use feedlot.dta, clear
rename phcysc mh
cc brd mh, by(ibrsc)
table brd mh province
cc brd mh, by(ibrsc) bd tarone

* Nocardia data, Example 13.9
use nocardia, clear
cc casecont dneo, by(dclox)
cc casecont dclox, by(dneo)
```

Note: the `Nocardia` dataset at the VER website has a pre-defined label `yesno`, without which you cannot use the `cc` command; solutions:

- redefine the label:
  `label def yesno 0 no 1 yes`
- remove the label from the two variables:
  `label values dneo`
  `label values dclox`

Appendix: Calculations for Risk Interactions

Notation for two dichotomous predictors:

- denote by p_{ij} the risk for predictor set (i, j), where $i, j = 0, 1$,
- denote by $\text{RD}_{ij} = p_{ij} - p_{00}$ the risk difference relative to baseline $(0, 0)$,
- denote by $\text{RR}_{ij} = p_{ij}/p_{00}$ the relative risk relative to baseline $(0, 0)$,
- denote by OR_{ij} the odds-ratio relative to baseline $(0, 0)$.

An additive model for risk assumes

$$\begin{aligned} p_{11} - p_{10} &= p_{01} - p_{00}, \quad \text{or} \\ p_{11} - p_{00} &= p_{10} - p_{00} + p_{01} - p_{00}, \quad \text{or} \\ \text{RD}_{11} &= \text{RD}_{10} + \text{RD}_{01}. \end{aligned}$$

A multiplicative model for risk assumes

$$\begin{aligned} p_{11}/p_{10} &= p_{01}/p_{00}, \quad \text{or} \\ p_{11}/p_{00} &= p_{10}/p_{00} \cdot p_{01}/p_{00}, \quad \text{or} \\ \text{RR}_{11} &= \text{RR}_{10} \cdot \text{RR}_{01}, \quad \text{or} \\ \ln(p_{11}) - \ln(p_{00}) &= \ln(p_{10}) - \ln(p_{00}) + \ln(p_{01}) - \ln(p_{00}), \end{aligned}$$

which is an additive model for $\ln(p)$.

Similarly, a multiplicative model for odds of risk $(p/(1-p))$, or an additive model for log-odds of risk, is equivalent to

$$\text{OR}_{11} = \text{OR}_{10} \cdot \text{OR}_{01}.$$