

Index of Lecture 10b

Page	Title
1	Practical information
2	Statistical interaction
3	Assessing interaction for risk
4	Interactions for <code>feedlot</code> data
5	Interaction for risk — Example 13.10
6	Interaction in biology and epidemiology
7	Scales for interaction — Example 13.9
8	Stata do-file
9	Appendix: calculations for risk interactions

PRACTICAL INFORMATION

Major news:

- suggested preparation for next session: home assignment 2007 (link from 2007 webpage), Questions 1–4,
- home assignment to be posted on Tuesday next week.

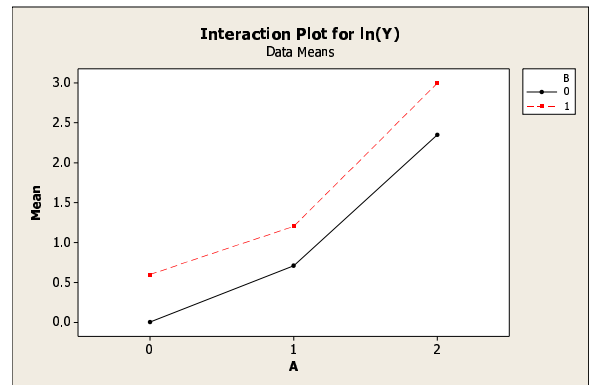
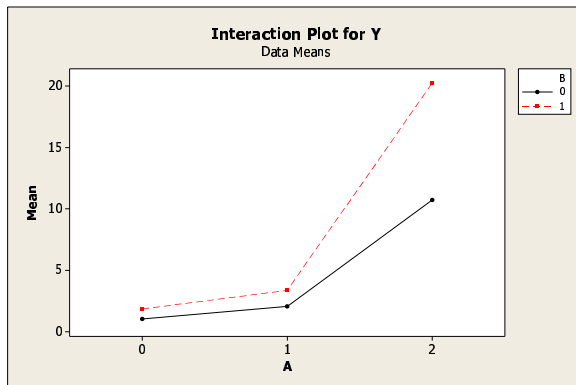
Today's session will be split in two parts

- first part, here and probably 1–1.5 hours:
 - * questions for material from Tuesday?,
 - * some lecturing:
 - interaction (definition, detection and handling),
 - skipped slide? (L10a–13; association created by control),
 - * review of prepared Problem 13.7,
- second part, in the computer lab (225N):
 - * individual work on problems,
note: solution file available (not perfect!),
 - * 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 modification¹: 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 *et al* (2008) 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:

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

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 – 0.15).

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 Mh and province/IBR in their effect on BRD (Examples 13.7-8):

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 Mh,
- 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 13.10

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

- clear interaction \Rightarrow 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 (synergism⁵),
 - * effect occurs in the presence of one factor and the absence of another factor, but not when both factors are present (antagonism⁵).

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

additive interaction : $RD_{11} \neq RD_{10} + RD_{01}$,

multiplicative interaction I : $RR_{11} \neq RR_{10} \cdot RR_{01}$,

multiplicative interaction II : $OR_{11} \neq OR_{10} \cdot OR_{01}$,

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 *et al* (2008), Chapter 5.

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 13.9

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	.019	—	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, Examples 13.7 and 13.8

```
use feedlot.dta, clear  
rename phcysc mh  
cc brd mh, by(province)  
cc brd mh, by(ibrsc)  
table brd mh province  
cc brd mh, by(ibrsc) bd tarone
```

* Nocardia data, Example 13.10

```
use nocardia, clear  
cc casecont dneo, by(dclox)  
cc casecont dclox, by(dneo)
```

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 $RD_{ij} = p_{ij} - p_{00}$ the risk difference relative to baseline $(0, 0)$,
- denote by $RR_{ij} = p_{ij}/p_{00}$ 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

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

A multiplicative model for risk assumes

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

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

$$OR_{11} = OR_{10} \cdot OR_{01}.$$