



# Summarizing Data

Nick Martin, Hermine Maes

TC21

March 2008



# Files to Copy to your Computer

- Faculty/Maes/a21/maes/univariate
  - ozbmi.rec
  - ozbmi.dat
  - ozbmifysat(eqmv).mx
  - Data\_Summary.ppt



# Change Preferences

## ■ Windows

- Tools > Folder Options > View:

- Unselect! Hide extensions for known file types

## ■ Mx

- Job Options > Output: Select Text

- Host Options > Memory: Change to 20000

- Fonts > Script Font: Courier New

- To resize windows: Window >Tile



# Estimating Means & Variances

- Any statistical package (SPSS, SAS, R)
  - R session on Friday
- By Maximum Likelihood
  - Height of multinormal probability density function:  
$$-\left|2\pi\sum\right|^{-n/2} e^{-.5((x_i - \mu) \Sigma^{-1} (x_i - \mu)')}$$
  - Raw data
    - Make use of all available data
    - Get unbiased estimates if missing data are missing at random



# Practical Example

- Dataset: NH&MRC Twin Register
- 1981 questionnaire
- BMI: weight/ height squared
- Young cohort: 18-30 years
- N MZFY: 534, DZFY: 328

# Raw Dataset ozbmi.rec

fam agecat age zyg part wt1 wt2 ht1 ht2 htwt1 htwt2 bmi1 bmi2

115	1	0.21	1	2	58	57	1.7000	1.7000	20.0692	19.7232	20.9943	20.8726
121	1	0.24	1	2	54	53	1.6299	1.6299	20.3244	19.9481	21.0828	20.9519
158	1	0.21	1	2	55	50	1.6499	1.6799	20.2020	17.7154	21.0405	20.1210
172	1	0.21	1	2	66	76	1.5698	1.6499	26.7759	27.9155	23.0125	23.3043
182	1	0.19	1	2	50	48	1.6099	1.6299	19.2894	18.0662	20.7169	20.2583
199	1	0.26	1	2	60	60	1.5999	1.5698	23.4375	24.3418	22.0804	22.3454
221	1	0.23	1	2	65	65	1.7500	1.7698	21.2245	20.7476	21.3861	21.2270
239	1	0.29	1	2	40	39	1.5598	1.5298	16.4366	16.6603	19.5966	19.6912
246	1	0.24	1	2	60	57	1.7598	1.7698	19.3698	18.1940	20.7460	20.3076
251	1	0.28	1	2	76	64	1.7000	1.7300	26.2976	21.3839	22.8863	21.4385
535	1	0.22	1	2	57	55	1.5798	1.5999	22.8329	21.4844	21.8974	21.4713
582	1	0.26	1	2	62	61	1.6799	1.6399	21.9671	22.6800	21.6268	21.8504
587	1	0.28	1	2	60	58	1.6299	1.6399	22.5827	21.5645	21.8203	21.4974
588	1	0.19	1	2	55	55	1.5698	1.5698	22.3133	22.3133	21.7363	21.7363
621	1	0.21	1	2	46	49	1.5698	1.5498	18.6620	20.3954	20.4854	21.1072
630	1	0.27	1	1	54	55	1.7698	.	17.2364	.	19.9292	.
631	1	0.28	1	2	58	60	1.6899	1.6899	20.3074	21.0077	21.0769	21.3142
632	1	0.25	1	2	54	57	1.6399	1.6399	20.0773	21.1927	20.9971	21.3756

# Dat File: ozbmi.dat

- #include ozbmi.dat
- Data NInput=13
  - Rectangular File=ozbmi.rec
  - Labels fam agecat age zyg part wt1 wt2 ht1 ht2 htwt1 htwt2 bmi1 bmi2

	MZF	MZM	DZF	DZM	DZFM
zyg	1	2	3	4	5
	Young	Old(er)		Single	Pair
agecat	1	2	part	1	2



Title		
Data	Calculation	Constraint
Rectangular/Ordinal.. Labels .. / Select .. Definition ..		
	Begin Matrices; e.g. A Full 1 2 End Matrices;	
	Matrix .. / Start .. Specify .. / Free .. Bound .. etc.	
	Begin Algebra; e.g. B=A*A; End Algebra;	
Means matrix formula Covariances formula		Constraint formula
	Options	
	End	





# Mx input script: ozbmiyfsat.mx

- Calls ozbmi.dat
  - Calls ozbmi.rec
- Classical Twin Study (CLT)
  - MZ twins reared together
  - DZ twins reared together
  
  - > 2 group analysis



# Parameters of the model

- Univariate analysis
  - Two variables: bmi1 bmi2
- MZ group
  - 2 means, 2 variances, 1 covariance
- DZ group
  - 2 means, 2 variances, 1 covariance
- Number of parameters: 10

# ! Estimate means and variances - Saturated model

## ! OZ BMI data - young females

Number of groups in Mx script

Replace every occurrence of nvar2 with the number 2

- `#NGroups 2`
- `#define nvar2 2`

Data NInput=13  
Rectangular File=ozbmi.rec  
Labels fam agecat age zyg part wt1 wt2 ht1 ht2 htwt1 htwt2 bmi1 bmi2

- Title 1: MZ data
- `#include ozbmi.dat`
- `Select if zyg =1`
- `Select if agecat =1`
- `Select bmi1 bmi2 ;`
- `Begin Matrices;`
- `M Full 1 nvar2 Free`
- `X Symm nvar2 nvar2 Free`
- `End Matrices;`
- `Start 20 M 1 1 - M 1 nvar2`
- `Start 1 X 1 1 X 2 2`
- `Means M;`
- `Covariance X;`
- `Option RSiduals`
- `End`

- Title 2: DZ data
- `#include ozbmi.dat`
- `Select if zyg =3`
- `Select if agecat =1`
- `Select bmi1 bmi2 ;`
- `Begin Matrices;`
- `M Full 1 nvar2 Free`
- `X Symm nvar2 nvar2 Free`
- `End Matrices;`
- `Start 20 M 1 1 - M 1 nvar2`
- `Start 1 X 1 1 X 2 2`
- `Means M;`
- `Covariance X;`
- `Option RSiduals`
- `End`

[ozbmiyfsat.mx](#)

# ! Estimate means and variances - Saturated model

## ! OZ BMI data - young females

- #NGroups 2
- #define nvar2 2

Data NInput=13  
 Rectangular File=ozbmi.rec  
 Labels fam agecat age zyg part wt1 wt2 ht1 ht2 htwt1 htwt2 bmi1 bmi2

### ■ Title 1: MZ data

- #include ozbmi.dat
- Select if zyg =1
- Select if agecat =1
- Select bmi1 bmi2 ;
- Begin Matrices;
- M Full 1 nvar2 Free
- X Symm nvar2 nvar2 Free
- End Matrices;
- Start 20 M 1 1 - M 1 nvar2
- Start 1 X 1 1 X 2 2
- Means M;
- Covariance X;
- Option RSiduals
- End

Conditional Select

Select Variables for Analysis

Declare Matrices

Mean Vector  $\begin{bmatrix} m1 & m2 \end{bmatrix}$

Covariance Matrix  $\begin{bmatrix} v1 & cov \\ cov & v2 \end{bmatrix}$

Start Values for Means and Variances

Model Statement for Means and Variances

Print Observed and Expected Means and Variances



# Mx Output: ozbmiyfsat.mxo

- Input Mx script
  - Error Messages
  - Notes
  - Warnings
- Mx Parameter Specification
  - 0 for fixed parameters
  - Non-zero distinct number for each free parameter
- Mx Parameter Estimates
  - Goodness-of-fit Statistics

# Goodness-of-fit Indices

Number of parameters	np		10
Observed statistics	os		1777
-2 times log-likelihood of data	-2ll		4055.935
Degrees of freedom	df	=os-np	1767
Akaike's Information Criterion	AIC	=-2ll-2df	521.935
Bayesian Information Criterion	BIC	=.5(-2lnL-df*ln(N))	-3996.552
Sample Size Adjusted BIC	saBIC	=.5(-2lnL-df*ln(N+2/24))	-1190.671
Deviance Information Criterion	DIC	=.5(-2lnL-df*ln(N/2π))	-2372.788

# Estimates for BMI yf

		T1	T2		T1	T2
Saturated model						
mean	MZ	21.34	21.35	DZ	21.45	21.46
cov	T1	.73		T1	.77	
	T2	.59	.79	T2	.24	.82



# Tests

- Saturated model
- Equality of means
  - By birth order:  $m_1 = m_2$  ?
  - By zygosity:  $m_{1MZ} = m_{1DZ} = m_{2MZ} = m_{2DZ}$  ?
- Equality of variances
  - By birth order:  $v_1 = v_2$  ?
  - By zygosity:  $v_{1MZ} = v_{1DZ} = v_{2MZ} = v_{2DZ}$  ?



# Equality Tests

- Main Script

Indicates to Mx that you want to fit submodels which will follow,  
Has to be before the End Statement of the Last Group of your Main Script

- Last Group

Indicates that this Model is the Saturated Model

- .....

- Option Multiple Issat

- End

Saves Script as Binary File,  
Including Data, Model Specification and Parameter Estimates,  
for easy future recall

- Save ozbmisat.mxs

- ! equate means and variances

- Equate M 1 1 1 M 1 1 2 M 2 1 1 M 2 1 2

- Equate X 1 1 1 X 1 2 2 X 2 1 1 X 2 2 2

- End

Equate Matrix Elements, specified by Matrix Name, Group Number, Row Number, Column Number

# Specific Equality Tests

- `Get_ ozbmisat.mxs`

Get Binary File,  
Including Data, Model Specification and Parameter Estimates,  
of previously saved Script

- `! equate means within zygoty groups`

- `Equate M 1 1 1 M 1 1 2`

- `Equate M 2 1 1 M 2 1 2`

- `End`

Submodel, just requires Changes compared to Full Script,  
Must have **End** Statement

- `! equate means across zygoty groups`

- `Equate M 1 1 1 M 1 1 2 M 2 1 1 M 2 1 2`

- `End`

- `! equate variances within zygoty groups`

- `Equate X 1 1 1 X 1 2 2`

- `Equate X 2 1 1 X 2 2 2`

- `End`

- `! equate variances across zygoty groups`

- `Equate X 1 1 1 X 1 2 2 X 2 1 1 X 2 2 2`

- `End`

# Multiple Fit Parameters

	MZ (group 1)					DZ (group 2)					par
	m1	m2	v1	cov	v2	m3	m4	v3	cov	v4	
Full	1	2	3	4	5	6	7	8	9	10	10
Save filename.mxs: 10 free parameters											10
I	1	1	3	4	3	1	1	3	9	3	4
Get filename.mxs: back to 10 free parameters											10
II	1	1	3	4	5	6	6	8	9	10	8
III	1	1	3	4	5	1	1	8	9	10	7
IV	1	1	3	4	3	1	1	8	9	8	5
V	1	1	3	4	3	1	1	3	9	3	4



# Assignment

- Fit saturated model and submodels
  - Summarize goodness-of-fit results
    - See table below
  - Record parameter estimates
    - See table below

# Goodness-of-fit for BMI yf

	-2LL	df	np	diff chi <sup>2</sup>	df	p
saturated						
equate m/v						
equate m1=m2						
equate means						
equate v1=v2						
equate vars						

# Estimates for BMI yf

		T1	T2		T1	T2
Saturated model						
mean	MZ			DZ		
cov	T1			T1		
	T2			T2		
Equated means and variances						
mean	MZ			DZ		
cov	T1			T1		
	T2			T2		

# Goodness-of-fit for BMI yf

	-2LL	df	np	diff chi <sup>2</sup>	df	p
saturated	4055.93	1767	10			
eq m/v	4063.45	1773	4	7.515	6	.276
eq m1=m2	4056.00	1769	8	0.066	2	.967
eq means	4060.15	1770	7	4.216	3	.239
eq v1=v2	4063.09	1772	5	7.161	5	.209
eq vars	4063.45	1773	4	7.515	6	.276

# Estimates for BMI yf

		T1	T2		T1	T2
Saturated model						
mean	MZ	21.34	21.35	DZ	21.45	21.46
cov	T1	.73		T1	.77	
	T2	.59	.79	T2	.24	.82
Equated means and variances						
mean	MZ	21.39	21.39	DZ	21.39	21.39
cov	T1	.80		T1	.80	
	T2	.59	.80	T2	.24	.80





# Mx Filename Conventions

- Input Mx script: filename.mx
  - Part of Mx script describing data file: filename.dat
  - Rectangular data file: filename.rec
  - Ordinal data files: filename.ord
- Output Mx script: filename.mxo
- Binary Save Files: filename.mxs