



Summarizing Data

Hermine Maes

TC19

March 2006



Files to Copy to your Computer

- Faculty/Maes/tc19/maes/univariate
 - ozbmi.rec
 - ozbmi.dat
 - ozbmifysat(eqmv).mx
 - Data_Summary.ppt



Practical Example

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

- Univariate - height of normal pdf

- $\phi(\mathbf{x}) =$

- $(2\pi\sigma^2)^{-0.5} e^{-0.5((\mathbf{x}_i - \mu)^2)/\sigma^2}$

- Multivariate - height of multinormal pdf

- $|2\pi\Sigma|^{-n/2} e^{-0.5((\mathbf{x}_i - \mu)\Sigma^{-1}(\mathbf{x}_i - \mu)')}$

$$\text{RM} = -k \log(2\pi) + \log|\Sigma| + (\mathbf{x}_i - \mu_j)' \Sigma^{-1} (\mathbf{x}_i - \mu_j)$$



Raw Dataset ozbmi2.rec

```
5 1 0.21 1 2 58 57 1.7 1.7 20.0692 19.7232 20.9943 20.8726
1 1 0.24 1 2 54 53 1.6299 1.6299 20.3244 19.9481 21.0828 20.9519
3 1 0.21 1 2 55 50 1.6499 1.6799 20.202 17.7154 21.0405 20.121
2 1 0.21 1 2 66 76 1.5698 1.6499 26.7759 27.9155 23.0125 23.3043
2 1 0.19 1 2 50 48 1.6099 1.6299 19.2894 18.0662 20.7169 20.2583
9 1 0.26 1 2 60 60 1.5999 1.5698 23.4375 24.3418 22.0804 22.3454
1 1 0.23 1 2 65 65 1.75 1.7698 21.2245 20.7476 21.3861 21.227
9 1 0.29 1 2 40 39 1.5598 1.5298 16.4366 16.6603 19.5966 19.6912
3 1 0.24 1 2 60 57 1.7598 1.7698 19.3698 18.194 20.746 20.3076
1 1 0.28 1 2 76 64 1.7 1.73 26.2976 21.3839 22.8863 21.4385
5 1 0.22 1 2 57 55 1.5798 1.5999 22.8329 21.4844 21.8974 21.4713
2 1 0.26 1 2 62 61 1.6799 1.6399 21.9671 22.68 21.6268 21.8504
7 1 0.28 1 2 60 58 1.6299 1.6399 22.5827 21.5645 21.8203 21.4974
3 1 0.19 1 2 55 55 1.5698 1.5698 22.3133 22.3133 21.7363 21.7363
1 1 0.21 1 2 46 49 1.5698 1.5498 18.662 20.3954 20.4854 21.1072
9 1 0.27 1 1 54 55 1.7698 . 17.2364 . 19.9292 .
1 1 0.28 1 2 58 60 1.6899 1.6899 20.3074 21.0077 21.0769 21.3142
2 1 0.25 1 2 54 57 1.6399 1.6399 20.0773 21.1927 20.9971 21.3756
```

Dat File: ozbmi2.dat

- #include ozbmi2.dat

- Data NInput=13

- Rectangular File=ozbmi2.rec

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

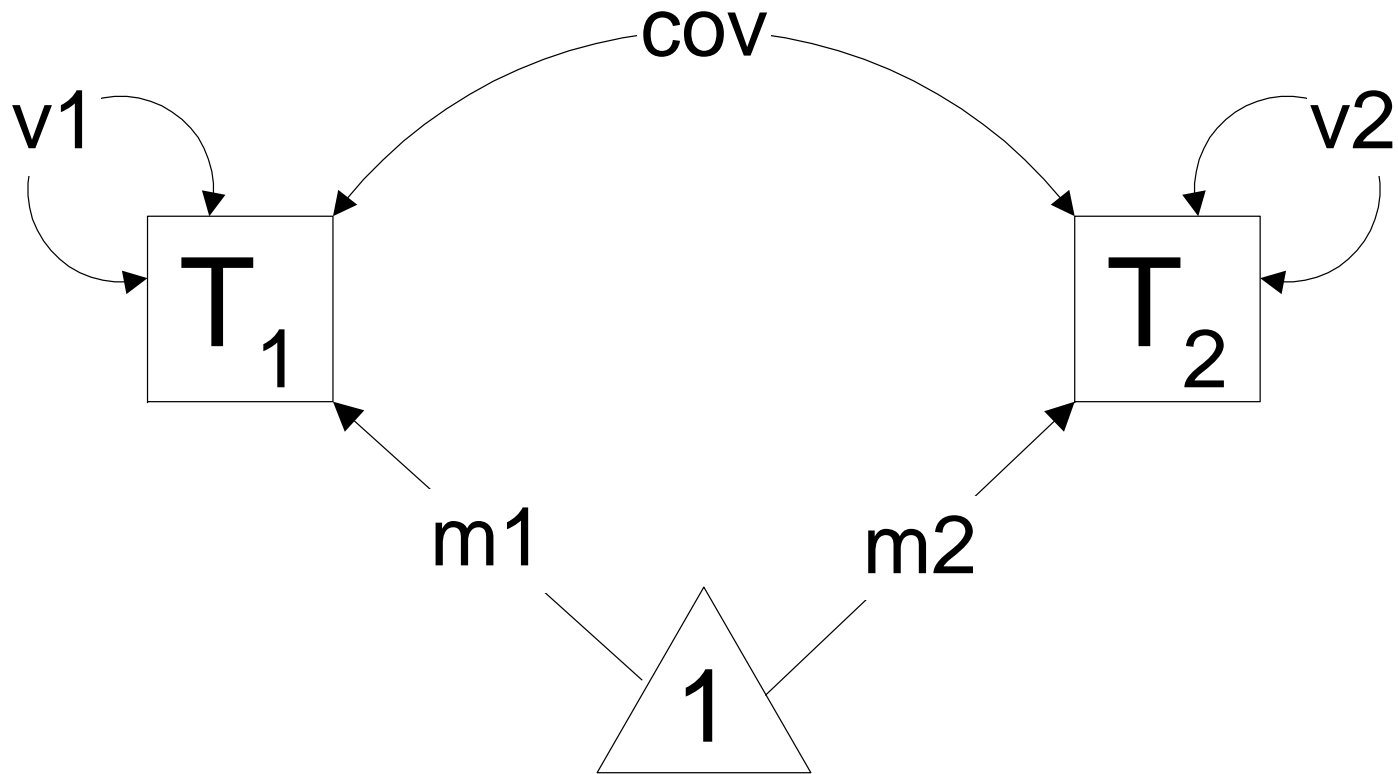
	MZF	MZM	DZF	DZM	DZFM
zyg	1	2	3	4	5
	Young	Old(er)		Single	Pair
agecat	1	2	part	1	2



Estimating Variance by ML

- MxGui Interface
- Mx Script Language

Saturated Model





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

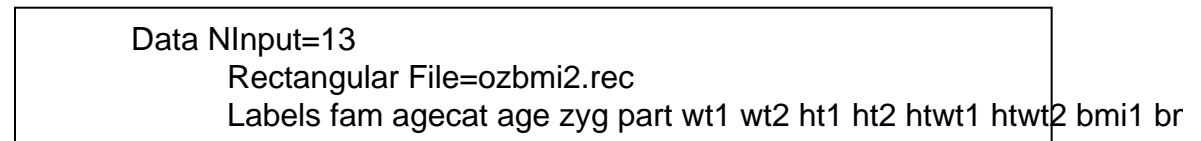
- Calls ozbmi2.dat
 - Calls ozbmi2.rec
- Classical Twin Study (CLT)
 - MZ twins reared together
 - DZ twins reared together

 - > 2 group analysis

! Estimate means and variances - Saturated model

! OZ BMI data - young females

- #NGroups 2
- #define nvar2 2



- Title 1: MZ data
- #include ozbmi2.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 ozbmi2.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

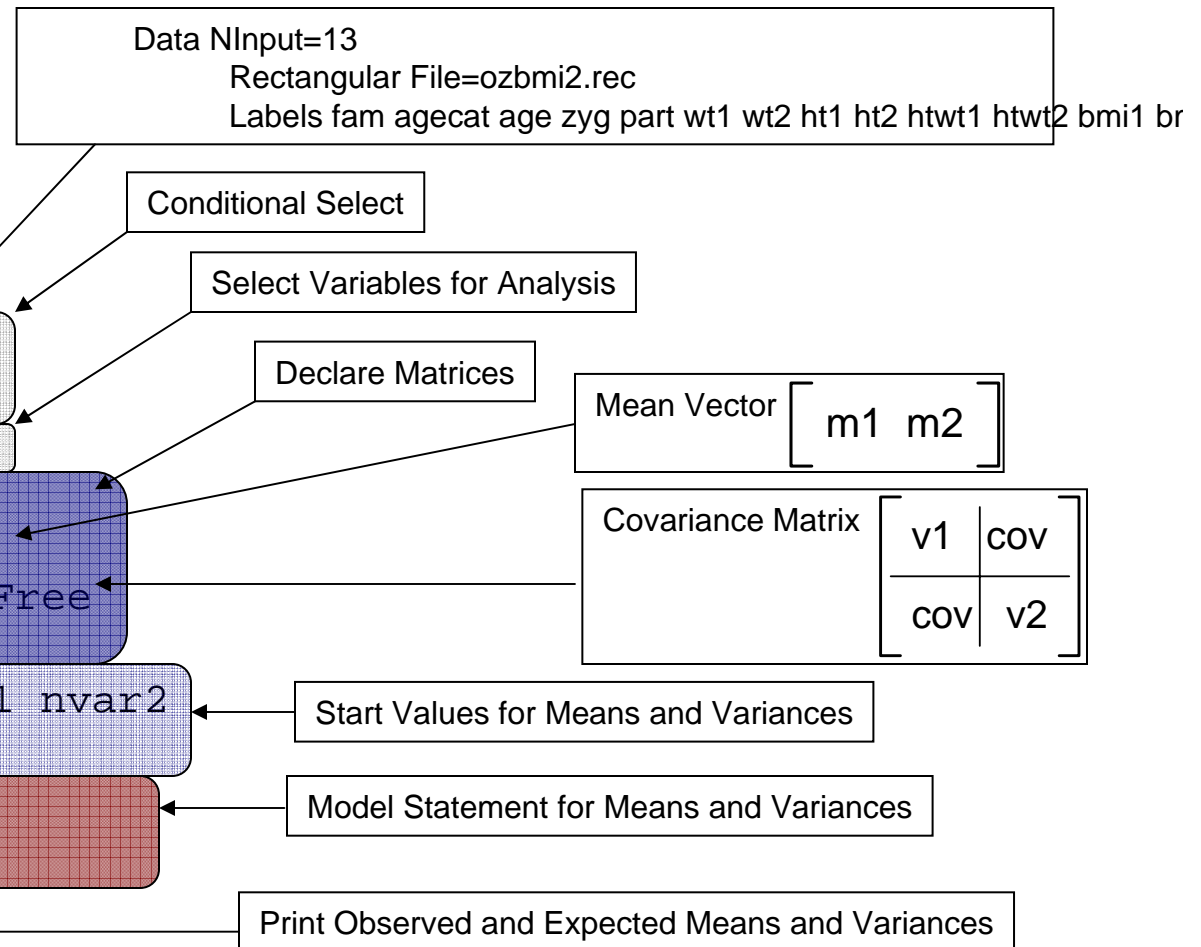
! Estimate means and variances - Saturated model

! OZ BMI data - young females

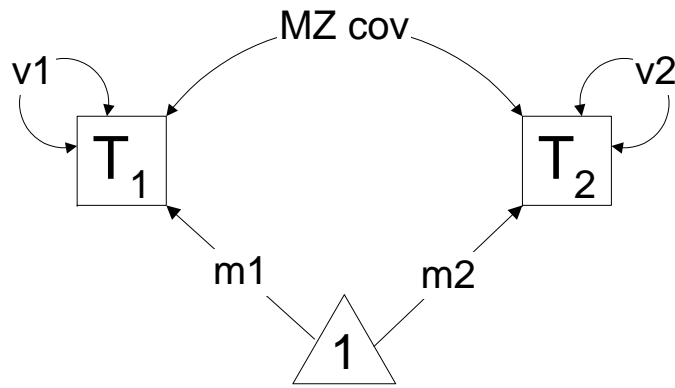
- #NGroups 2
- #define nvar2 2

■ Title 1: MZ data

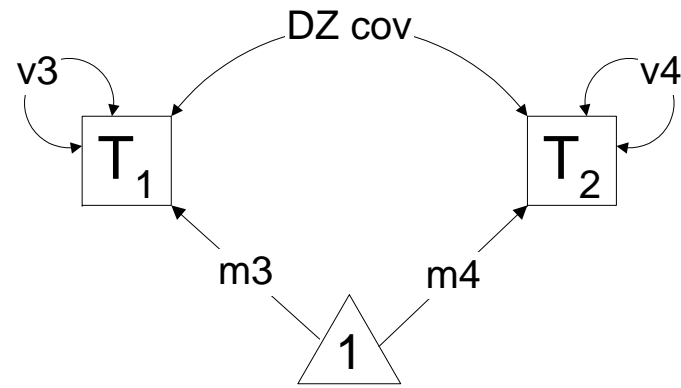
- #include ozbmi2.dat
- Select if zyg =1
- Select if agecat =1
- Select bmi1 bmi2 ;
- Begin Matrices;
- M Full 1 nvar2 Free
- X Symm nvar2 nvar2 Free
- End Matrices;
- Start 20 M 1 1 - M 1 nvar2
- Start 1 X 1 1 X 2 2
- Means M;
- Covariance X;
- Option RSiduals
- End



Saturated Model MZ & DZ



MZ twins



DZ twins

10 parameters: 4 means, 4 variances, 2 covariances

Estimates for BMI yf

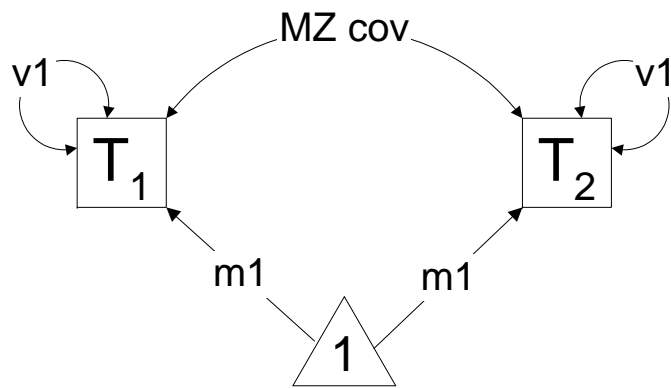
		T1	T2		T1	T2
Saturated model						
mean	MZ	21.34	21.35	DZ	21.45	21.46
cov	T1	.73		T1	.77	
	T2	.59	.79	T2	.24	.82



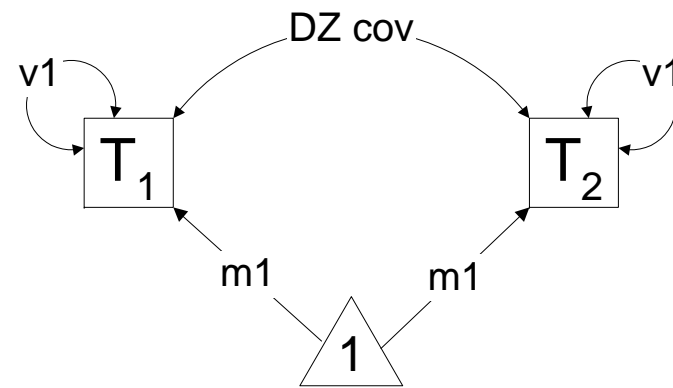
Tests

- Saturated model
- Equality of means
 - Is $m_1 = m_2$?
 - Is $m_{1MZ} = m_{1DZ} = m_{2MZ} = m_{2DZ}$?
- Equality of variances
 - Is $v_1 = v_2$?
 - Is $v_{1MZ} = v_{1DZ} = v_{2MZ} = v_{2DZ}$?

Equality of means, variances



MZ twins



DZ twins

4 parameters: 1 mean, 1 variance, 2 covariances

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

-

```
Option Multiple Issat
```

Indicates that this Model is the Saturated Model

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

Goodness-of-fit for BMI yf

	-2LL	df	np	R^2	df	p
saturated						
eq m/v						
eq m1=m2						
eq means						
eq v1=v2						
eq 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	χ^2	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