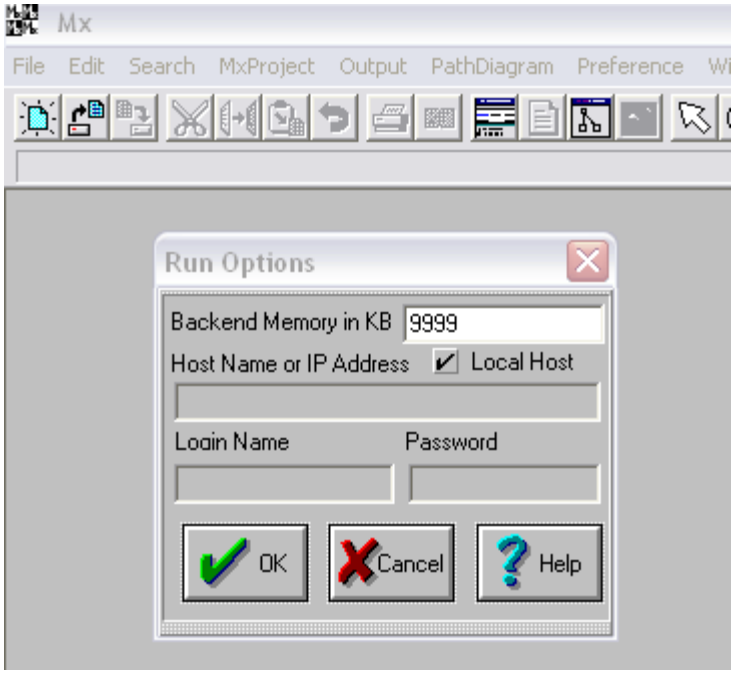
Welcome

- Log on using the username and password you received at registration
- Copy the folder:
  F:/sarah/mon-morning
  To your H drive

Open Mx and increase memory

- Go to preference
  - Host options
  - Back the backend memory larger



Intro to Mx

Sarah Medland – Leuven 2008

This morning

- Fitting a mean and regression with continuous data
- Modelling Ordinal data
- Fitting the regression model with ordinal data

Lets start with the data…

- File: Wednesday.dat
  - Contains 6 variables from the NLTR
  - ntrid zygMZDZ age1 sekse1 AQ1 age2 sekse2 AQ2

ntrid	zygMZDZ	age1	sekse1	AQ1	age2	sekse2	AQ2
2	1	18.12	1	91	18.12	1	95
3	2	18.37	0	90	18.37	1	89
5	2	18.34	0	-1	-4	-4	-1
10	1	18.12	1	-1	18.12	1	-1
15	2	18.52	1	-1	18.5	1	-1
16	2	18.04	1	100	18.04	1	110

Lets start with the data…

ntrid	zygMZDZ	age1	sekse1	AQ1	age2	sekse2	AQ2
2	1	18.12	1	91	18.12	1	95
3	2	18.37	0	90	18.37	1	89
5	2	18.34	0	-1	-4	-4	-1
10	1	18.12	1	-1	18.12	1	-1
15	2	18.52	1	-1	18.5	1	-1
16	2	18.04	1	100	18.04	1	110

If this was a pedigree data file…

ntrid	zygMZDZ	age1	sekse1	AQ1	age2	sekse2	AQ2
2	1	18.12	1	91	18.12	1	95
3	2	18.37	0	90	18.37	1	89
5	2	18.34	0	-1	-4	-4	-1
10	1	18.12	1	-1	18.12	1	-1
15	2	18.52	1	-1	18.5	1	-1
16	2	18.04	1	100	18.04	1	110

Famid	Ind	Father	Mother	Zyg	Sex	Age	Trait
2	1	0	0	0	1	x	x
2	2	0	0	0	2	x	x
2	3	1	2	MZ	1	18.12	91
2	4	1	2	MZ	1	18.12	95

How can we make this data file?

- Assume we have data with 3 variables:

	Family	Individual	Trait
1	1.00	1.00	24.00
2	1.00	2.00	42.00
3	2.00	1.00	21.00
4	2.00	2.00	4.00
5	3.00	1.00	2.00
6	3.00	2.00	1.00
7	4.00	1.00	3.00
8	4.00	2.00	12.00

	Family	Trait.1.00	Trait.2.00
1	1.00	24.00	42.00
2	2.00	21.00	4.00
3	3.00	2.00	1.00
4	4.00	3.00	12.00

How do we make this data?

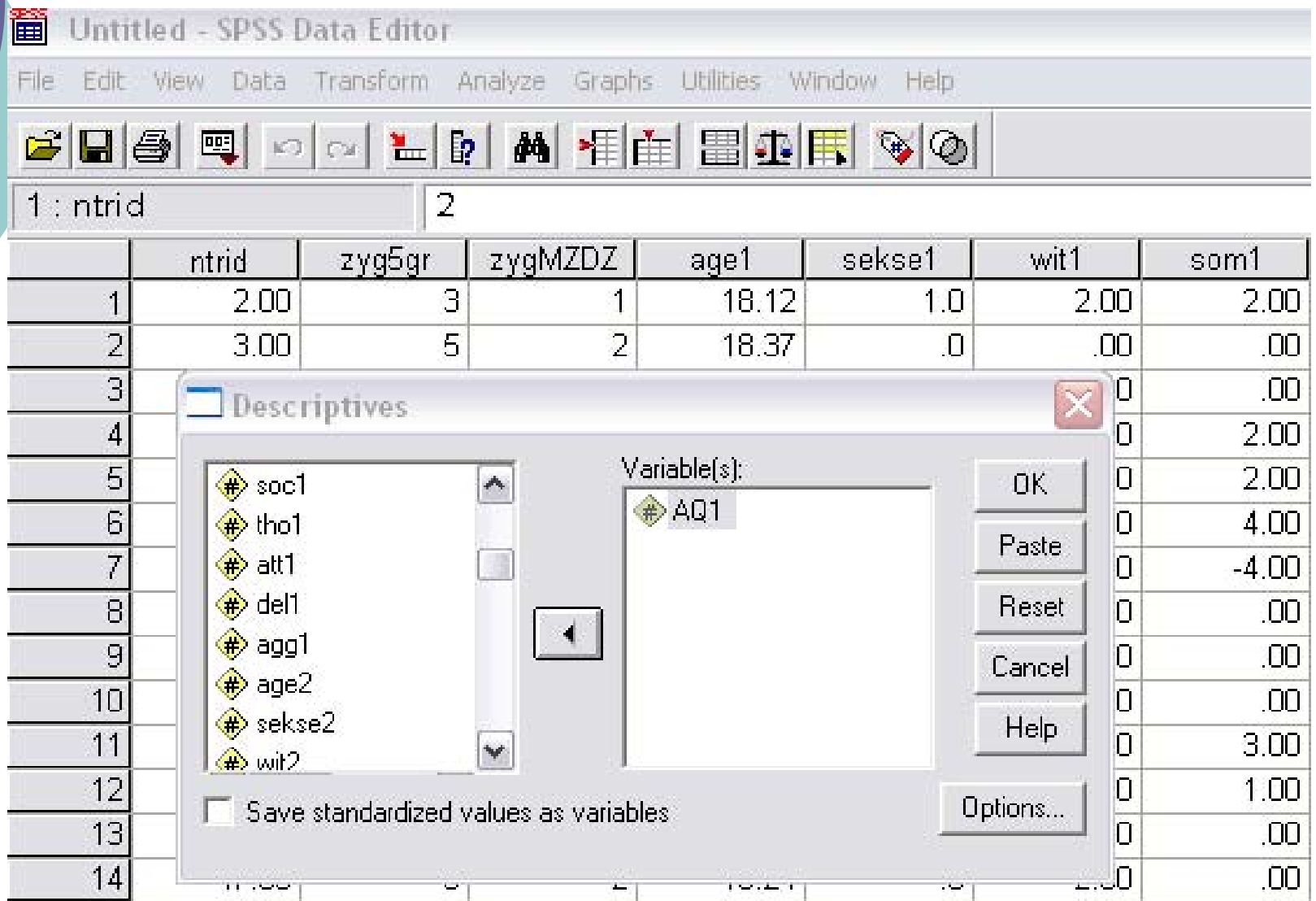
- SPSS

```
SORT CASES BY Family Individual .
CASESTOVARS
 /ID = Family
 /INDEX = Individual
 /GROUPBY = VARIABLE .
```

Means…

- In spss sas etc we calculate the mean
- In Mx and other ML programs we estimate the mean

Spss…


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : ntrid
2
ntrid zyg5gr zygMZDZ age1 sekse1 wit1 som1
1 2.00 3 1 18.12 1.0 2.00 2.00
2 3.00 5 2 18.37 .0 .00 .00
3 .00
4 2.00
5 2.00
6 4.00
7 -4.00
8 .00
9 .00
10 .00
11 3.00
12 1.00
13 .00
14 .00
Descriptives
soc1
tho1
att1
del1
agg1
age2
sekse2
wit2
Variable(s):
AQ1
OK
Paste
Reset
Cancel
Help
Save standardized values as variables
Options...

Spss…

- **Descriptives**

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
AQ1	179	75.00	130.00	102.3408	10.41288
Valid N (listwise)	179				

```
#define nvar 1                    ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1              ! Number of variables per family
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 ;

Begin matrices;
    M Full 1 nvar free           !mean
    V Full nvar nvar free        ! variance
End matrices;

Begin Algebra;
    S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;        ! means model
Covariances V ;          ! variance/covariance model
end
```

Means.mx

```
#define nvar 1                    ! n dependent variables per individual

G1: Singleton (non-pair) data
Dat                                                          mily
mis

Rec
Lab
ntr

select AQ1 ;

Begin matrices;
    M Full 1 nvar free            !mean
    V Full nvar nvar free         ! variance
End matrices;

Begin Algebra;
    S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;        ! means model
Covariances V ;          ! variance/covariance model
end
```

Defining some frequently changed parameters

```
#define nvar 1                          ! n dependent variables per individual

G1: Singleton (non-pair) data
Dat                                                          mily
mis

Rectangular file=wednesday.dat                  ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 ;

Begin matrices;
    M Full 1 nvar free              !mean
    V Full nvar nvar free           ! variance
End matrices;

Begin Algebra;
    S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;          ! means model
Covariances V ;            ! variance/covariance model
end
```

Adding comments !

```
#define nvar 1                    ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1              ! Number of variables per family
missing=-1                                ! Missing values

Rect
Labe
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 ;

Begin matrices;
    M Full 1 nvar free           !mean
    V Full nvar nvar free        ! variance
End matrices;

Begin Algebra;
    S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;        ! means model
Covariances V ;          ! variance/covariance model
end
```

Providing a title

```
#define nvar 1                          ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1                    ! Number of variables per family
missing=-1                                      ! Missing values

Rect
Labe
ntri

sele

Begi


End

Begi

End
```

Directly after the title tell Mx what kind of group it is

- Data
- Calculation
- Constraint

```
! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;          ! means model
Covariances V ;            ! variance/covariance model
end
```

```
#define nvar 1                          ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1                    ! Number of variables per family
missing=-1                                      ! Missing values

Rect
Labe
ntri

sele

Begin matrices;
    M Full 1 nvar free              !mean
    V Full nvar nvar free           ! variance
End matrices;

Begin Algebra;
    S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;       ! means model
Covariances V ;         ! variance/covariance model
end
```

How many variables are in the data file

```
#define nvar 1                          ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1                    ! Number of variables per family
missing=-1                                      ! Missing values

Rect
Labe
ntr

sele

Begin matrices;
    M Full 1 nvar free              !mean
    V Full nvar nvar free           ! variance
End matrices;

Begin Algebra;
    S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;       ! means model
Covariances V ;         ! variance/covariance model
end
```

How many groups in the script

```
#define nvar 1                          ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1                    ! Number of variables per family
missing=-1                                      ! Missing values

Rec
Lab
ntr

select AQ1 ;

Begin matrices;
    M Full 1 nvar free              !mean
    V Full nvar nvar free           ! variance
End matrices;

Begin Algebra;
    S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;       ! means model
Covariances V ;         ! variance/covariance model
end
```

Missing code – default is a .

```
#define nvar 1                          ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1                    ! Number of variables per family
missing=-1                                      ! Missing values

Rectangular file=wednesday.dat                  ! read raw data
Labels
n
s
B
E
B
E
!
Start 100 M 1 1
Start 106 V 1 1

Means M ;           ! means model
Covariances V ;             ! variance/covariance model
end
```

Provide the name of the data file

Rectangular file = continuous

Ordinal file =ordinal/binary

```
#define nvar 1                    ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1              ! Number of variables per family
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 ;

Beg
    V Full nvar nvar free         ! variance
End matrices;

Begin Algebra;
    S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;        ! means model
Covariances V ;          ! variance/covariance model
end
```

List of the variables

```
#define nvar 1                    ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1              ! Number of variables per family
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 ;

Begin matrices;
    M Fu
    V Fu
End matr

Begin Algebra;
    S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;        ! means model
Covariances V ;          ! variance/covariance model
end
```

Tell Mx what to analyse

```
#define nvar 1                    ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1              ! Number of variables per family
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 ;

Begin matrices;
    M Full 1 nvar free           !mean
    V Full nvar nvar free        ! variance
End matrices;

Begin Algebra;
    S = \
End Algeb

! start v
Start 100
Start 106 V 1 1

Means M ;        ! means model
Covariances V ;          ! variance/covariance model
end
```

Tell Mx what matrices you want to use

Matrices: the building blocks

- Many types

```
A Zero 2 3 Free          0 0 0          0 0 0
                         0 0 0          0 0 0

B Unit 2 3 Free          0 0 0          1 1 1
                         0 0 0          1 1 1

C Iden 3 3 Free          0 0 0          1 0 0
                         0 0 0          0 1 0
                         0 0 0          0 0 1

F Diag 3 3 Free          1 0 0          ? 0 0
                         0 2 0          0 ? 0
                         0 0 3          0 0 ?

G Sdiag 3 3 Free         0 0 0          0 0 0
                         1 0 0          ? 0 0
                         2 3 0          ? ? 0

H Stand 3 3 Free         0 1 2          1 ? ?
                         1 0 3          ? 1 ?
                         2 3 0          ? ? 1

I Symm 3 3 Free          1 2 4          ? ? ?
                         2 3 5          ? ? ?
                         4 5 6          ? ? ?

J Lower 3 3 Free         1 0 0          ? 0 0
                         2 3 0          ? ? 0
                         4 5 6          ? ? ?

K Full 2 4 Free          1 2 3 4        ? ? ? ?
                         5 6 7 8        ? ? ? ?
```

Matrices: the building blocks

- Many types
- Denoted by a single letter
  - Elements defined by letter and 3 numbers
    - A 1 2 1 = A matrix group 1 row 2 column 1
- All constants and estimated parameters must be placed in a matrix & Mx must be told what type of matrix it is
- Letters can be reused in subsequent groups

```
#define nvar 1                    ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1              ! Number of variables per family
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 ;

Begin matrices;
    M Full 1 nvar free           !mean
    V Full nvar nvar free        ! variance
End matrices;

Begin Algebra;

End Algebra;

! 
St
St

Means M ;        ! means model
Covariances V ;          ! variance/covariance model
end
```

M & V matrices are both full 1 1

M=[?] V=[?]

```
#define nvar 1                    ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1              ! Number of variables per family
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
```

Tell Mx what algebra you want to do
S will contain the square root of the estimate of V

```
End Matrices;

Begin Algebra;
     S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;         ! means model
Covariances V ;           ! variance/covariance model
end
```

```
#define nvar 1                          ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1                    ! Number of variables per family
missing=-1                                      ! Missing values

Rectangular file=wednesday.dat                  ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 ;
```

Provide some start values to aid estimation

Default is .01

```
! start values
Start 100 M 1 1
Start 106 V 1 1

Means M ;          ! means model
Covariances V ;           ! variance/covariance model
end
```

```
#define nvar 1                    ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1              ! Number of variables per family
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 ;

Begin matrices;
    M Full 1 nvar free           !mean
    V Full nvar nvar free        ! variance
End matrices;
```

Tell Mx how which matrix contains the means and the variance/ covariance matrix

```
Means M ;          ! means model
Covariances V ;          ! variance/covariance model
end
```

```
#define nvar 1                    ! n dependent variables per individual

G1: Singleton (non-pair) data
Data Ninput_vars=8 NGroups=1              ! Number of variables per family
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 ;

Begin matrices;
    M Full 1 nvar free           !mean
    V Full nvar nvar free        ! variance
End matrices;

Begin Algebra;
    S = \sqrt(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
```

Each group needs an end statement

```
Covariances V ;           ! variance/covariance model
end
```

Output

```
  ** Mx startup successful **

!@human;  **MX-PC 1.63h** Job started on 08/10/08 at 17:09:35
 #DEFINE NVAR 1   ! N DEPENDENT VARIABLES PER INDIVIDUAL


 The following MX script lines were read for group    1

 G1: SINGLETON (NON-PAIR) DATA
 DATA NINPUT_VARS=8 NGROUPS=1   ! NUMBER OF VARIABLES PER FAMILY
 MISSING=-1    ! MISSING VALUES

  NOTE: Missing value *string* set to '-1'

 RECTANGULAR FILE=WEDNESDAY.DAT  ! READ RAW DATA
  Rectangular continuous data read initiated
  Note: Maximum ordinal/rectangular record length is:  1000

  NOTE: Rectangular file contained   424 records with data
        that contained a total of   2891 observations

 LABELS
 NTRID ZYGMZDZ AGE1 SEKSE1  AQ1 AGE2 SEKSE2 AQ2
 SELECT AQ1 ;
 BEGIN MATRICES;

  NOTE: Selection yields  179 data vectors for analysis
  NOTE: Vectors contain a total of   179 observations
```

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
AQ1	179	75.00	130.00	102.3408	10.41288
Valid N (listwise)	179				

```
 Summary of VL file data for group  1

                AQ1
    Code      1.0000
  Number    179.0000
    Mean    102.3408
Variance    107.8224
 Minimum     75.0000
 Maximum    130.0000
```

```
G1: Singleton (non-pair) data

 MATRIX M
This is a FULL matrix of orde
             1
1    102.3408

 MATRIX S
This is a computed FULL matri
 [=\SQRT(V)]
             1
1     10.3838

 MATRIX V
This is a FULL matrix of orde
             1
1    107.8224
```

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
AQ1	179	75.00	130.00	102.3408	10.41288
Valid N (listwise)	179				

- Spss assumes this is a sample
- Mx assumes this is a population
- Slightly different algebra

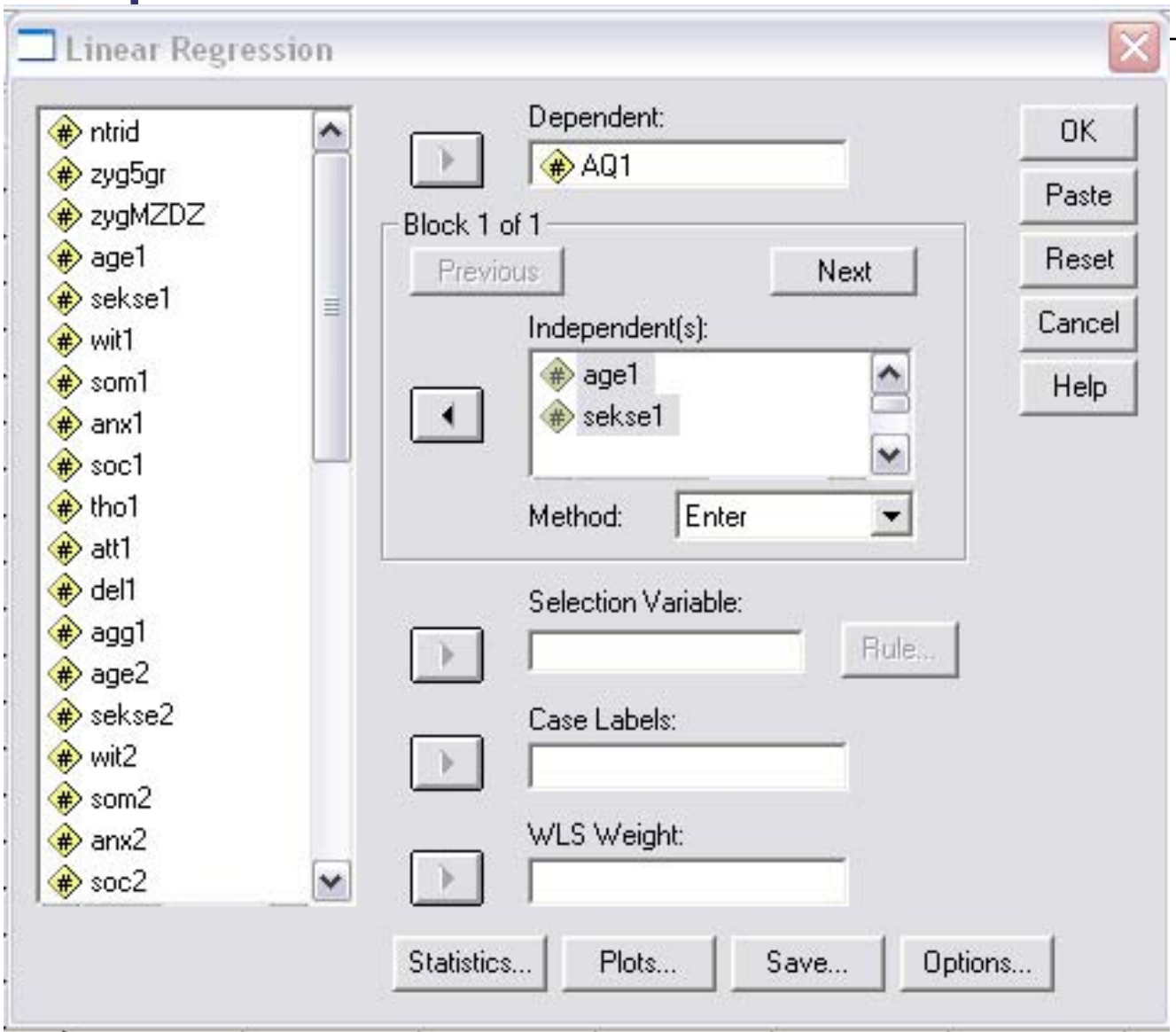
```
G1: Singleton (non-pair) data

 MATRIX M
This is a FULL matrix of orde
              1
1     102.3408

 MATRIX S
This is a computed FULL matri
 [=\SQRT(V)]
              1
1      10.3838

 MATRIX V
This is a FULL matrix of orde
              1
1     107.8224
```

How about regression?

- Y=X*B +C

- Regression speak
  - AutismQuotient = Sex*Beta1 + Age*Beta2 + Intercept
- BG speak
  - AutismQuotient = Sex Effect + Age Effect + Grand Mean

Spss…


Linear Regression
ntrid
zyg5gr
zygMZDZ
age1
sekse1
wit1
som1
anx1
soc1
tho1
att1
del1
agg1
age2
sekse2
wit2
som2
anx2
soc2
Dependent:
AQ1
Block 1 of 1
Previous
Next
Independent(s):
age1
sekse1
Method:
Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
Statistics...
Plots...
Save...
Options...
OK
Paste
Reset
Cancel
Help

regression.mx

```
#define nvar 1                    ! n dependent variables per individual
#define ndef 2                    ! n covariates per individual

G1: Singleton data - Regression
Data Ninput_vars=8 NGroups=1
missing=-1                        ! Missing values

Rectangular file=wednesday.dat
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select age1 sekse1 AQ1 ;
Definition age1 sekse1 ;
```

regression.mx

```
Begin matrices;

! intercept
! aka grand mean, aka intercept, aka observed variable, akak mean
    C Full 1 nvar free

! independent variable
! aka observed covariates for individual
    X Full 1 ndef fix

! beta
! aka covariate effect
    B Full ndef 1 free

! variance
    V Full nvar nvar free

End matrices;
```

regression.mx

```
! tell Mx to put the independent variables in the X matrix
Specify X age1 sekse1 ;

! start values
Start 95 C 1 1
Start 106 V 1 1
Start .5 B 1 1
Start 2  B 2 1


Begin Algebra;

! Write out the regression equation for the observed (dependent) variable - Y
      Y = X*B + C ;

End Algebra;


     Means Y ;  ! means model
     Covariances V ;            ! variance/covariance model
Option multiple
Option issat
end
```

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	sekse1, age1[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: AQ1

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.126[a]	.016	.005	10.38881

a. Predictors: (Constant), sekse1, age1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	305.003	2	152.501	1.413	.246[a]
	Residual	18995.210	176	107.927		
	Total	19300.212	178			

a. Predictors: (Constant), sekse1, age1

b. Dependent Variable: AQ1

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	93.564	65.862		1.421	.157
	age1	.549	3.623	.011	.151	.880
	sekse1	-2.608	1.558	-.125	-1.673	.096

a. Dependent Variable: AQ1

Run regression.mx

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	93.564	65.862		1.421	.157
	age1	.549	3.623	.011	.151	.880
	sekse1	-2.608	1.558	-.125	-1.673	.096

a. Dependent Variable: AQ1

```
G1: Singleton data - Regression

 MATRIX B
This is a FULL matrix of order    2 by    1
              1
1       0.5487
2      -2.6080

 MATRIX C
This is a FULL matrix of order    1 by    1
              1
1      93.5636
```

What does this mean?

	SEKSE1	AGE1
Code	-2.0000	-1.0000
Number	179.0000	179.0000
Mean	0.4581	18.1749
Variance	0.2482	0.0459
Minimum	0.0000	17.6100
Maximum	1.0000	18.9900

- Age Beta = .549
  - For every 1 unit increase in Age the mean shifts .549
  - Grand mean =93.564
  - Mean Age =18.2
- So the mean for 20 year olds is predicted to be:
  - 104.544 = 93.564 + 20*.549

Sex effects?

- Sex Beta = -2.608
- Sex coded Male = 1 Female = 0

- Female Mean:
  - 93.564 = 93.564 + 0*-2.608
- Male Mean:
  - 90.656 = 93.564 + 1*-2.608

How do we get the p-values?

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	93.564	65.862		1.421	.157
	age1	.549	3.623	.011	.151	.880
	sekse1	-2.608	1.558	-.125	-1.673	.096

a. Dependent Variable: AQ1

Set the elements to equal 0 regression.mx

- Do this one at a time!

```
save temp.mxs
drop B 1 1 1
end

get temp.mxs
drop B 1 2 1
end
```

So…

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	93.564	65.862		1.421	.157
	age1	.549	3.623	.011	.151	.880
	sekse1	-2.608	1.558	-.125	-1.673	.096

a. Dependent Variable: AQ1

```
Difference Chi-squared  >>>>>>>>      0.023
Difference d.f.  >>>>>>>>>>>>>>>          1
Probability >>>>>>>>>>>>>>>>>>>>      0.879

Difference Chi-squared  >>>>>>>>      2.826
Difference d.f.  >>>>>>>>>>>>>>>          1
Probability >>>>>>>>>>>>>>>>>>>>      0.093
```

Why bother with Mx?

- Because most stat packages can't handle non-independent data…
  - Non-independence reduces the variance
  - Biases t and F tests

Why bother with Mx?

- Because we want complete flexibility in the model specification…
  - As you see later today

Why bother with Mx?

- Because very few packages can handle ordinal data adequately…

Working with non-independent data

Correlation.mx

Correlation.mx

```
#define nvar 1                    ! n dependent variables per individual
#define nsib 2                    ! number if siblings per family

G1: Corelation between twin pairs
Data Ninput_vars=8 NGroups=1              ! Number of variables per family ___
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 AQ2 ;

Begin matrices;
    M Full 1 nvar free           !mean
    V Symm nsib nsib free        ! variance
End matrices;

Begin Algebra;
    S = \stnd(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1 V 2 2
Start 50 V 2 1
EQ V 1 1 V 2 2

Means M|M ;      ! means model
Covariances V ;          ! variance/covariance model
end
```

Correlation.mx

```
#define nvar 1                          ! n dependent variables per individual
#define nsib 2                          ! number if siblings per family
```

Defining the nsib as the number of siblings per family

```
missing=-1                                      ! Missing values

Rectangular file=wednesday.dat                  ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 AQ2 ;

Begin matrices;
    M Full 1 nvar free                  !mean
    V Symm nsib nsib free               ! variance
End matrices;

Begin Algebra;
    S = \stnd(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1 V 2 2
Start 50 V 2 1
EQ V 1 1 V 2 2

Means M|M ;     ! means model
Covariances V ;         ! variance/covariance model
end
```

Correlation.mx

```
#define nvar 1                    ! n dependent variables per individual
#define nsib 2                    ! number if siblings per family

G1: Corelation between twin pairs
Data Ninput_vars=8 NGroups=1              ! Number of variables per family ___
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 AQ2 ;
```

Selecting the AQ variable for sib 1 and sib 2

```
    V Symm nsib nsib free         ! variance
End matrices;

Begin Algebra;
    S = \stnd(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1 V 2 2
Start 50 V 2 1
EQ V 1 1 V 2 2

Means M|M ;      ! means model
Covariances V ;          ! variance/covariance model
end
```

Correlation.mx

```
#define nvar 1                    ! n dependent variables per individual
#define nsib 2                    ! number if siblings per family

G1: Corelation between twin pairs
Data Ninput_vars=8 NGroups=1              ! Number of variables per family ___
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 AQ2 ;

Begin matrices;
    M Full 1 nvar free            !mean
    V Symm nsib nsib free         ! variance
End matrices;
```

The V matrix is now a symmetric 2 2 matrix

Var1 Cov

Cov Var2

```
Start 100 M 1 1
Start 106 V 1 1 V 2 2
Start 50 V 2 1
EQ V 1 1 V 2 2

Means M|M ;      ! means model
Covariances V ;          ! variance/covariance model
end
```

Correlation.mx

```
#define nvar 1                    ! n dependent variables per individual
#define nsib 2                    ! number if siblings per family

G1: Corelation between twin pairs
Data Ninput_vars=8 NGroups=1              ! Number of variables per family ___
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
n

s

B

End Matrices;

Begin Algebra;
    S = \stnd(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 106 V 1 1 V 2 2
Start 50 V 2 1
EQ V 1 1 V 2 2

Means M|M ;     ! means model
Covariances V ;         ! variance/covariance model
end
```

The S matrix will contain the standardised var/cov matrix (ie the correlation)

1	r
r	1

Correlation.mx

```
#define nvar 1                    ! n dependent variables per individual
#define nsib 2                    ! number if siblings per family

G1: Corelation between twin pairs
Data Ninput_vars=8 NGroups=1              ! Number of variables per family ___
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 AQ2 ;

Begin matrices;
    M Full 1 nvar free           !mean
    V Symm nsib nsib free        ! variance
End matrices;

Begin Algebra;
```

Some extra start values

```
! start values
Start 100 M 1 1
Start 106 V 1 1 V 2 2
Start 50 V 2 1
EQ V 1 1 V 2 2

Means M|M ;      ! means model
Covariances V ;          ! variance/covariance model
end
```

Correlation.mx

```
#define nvar 1                    ! n dependent variables per individual
#define nsib 2                    ! number if siblings per family

G1: Corelation between twin pairs
Data Ninput_vars=8 NGroups=1              ! Number of variables per family ___
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 AQ2 ;

Begin matrices;
    M Full 1 nvar free           !mean
    V Symm nsib nsib free        ! variance
End matrices;

Begin Algebra;
    S = \stnd(V) ; ! compute Standard deviation
End Algebra;
```

EQ V 1 1 V 2 2 sets the variances to be the same for the two siblings.

```
Start 50 V 2 1
EQ V 1 1 V 2 2

Means M|M ;      ! means model
Covariances V ;          ! variance/covariance model
end
```

Correlation.mx

```
#define nvar 1                    ! n dependent variables per individual
#define nsib 2                    ! number if siblings per family

G1: Corelation between twin pairs
Data Ninput_vars=8 NGroups=1              ! Number of variables per family ___
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 AQ2 ;

Begin matrices;
    M Full 1 nvar free           !mean
    V Symm nsib nsib free        ! variance
End matrices;

Begin Algebra;
    S = \stnd(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
```

We are also using the same mean for both siblings

M|M = stack 2 M matrices side by side

```
Means M|M ;      ! means model
Covariances V ;          ! variance/covariance model
end
```

Output – correlation.mxo

```
 Summary of VL file data for group  1

                AQ1         AQ2
   Code       1.0000      2.0000
 Number     179.0000    168.0000
   Mean     102.3408    102.0893
Variance    107.8224    106.8789
Minimum      75.0000     73.0000
Maximum     130.0000    136.0000
```

Notice the small increase in variance when accounting for the non-independence

```
G1: Corelation between twin pairs

 MATRIX M
This is a FULL matrix of order    1 by    1
             1
1    102.1865

 MATRIX S
This is a computed FULL matrix of order    2 b
 [=\STND(V)]
          1       2
1    1.0000  0.3791
2    0.3791  1.0000

 MATRIX V
This is a SYMMETRIC matrix of order    2 by
             1           2
1    107.8418
2     40.8812   107.8418

Your model has    3 estimated parameters and

-2 times log-likelihood of data >>>  2584.422
Degrees of freedom >>>>>>>>>>>>>>>>       344
Akaike's Information Criterion >>>>  1896.422
Bayesian Information Criterion >>>>   390.630
Sample size Adjusted BIC       >>>>   935.445
Deviance Information Criterion >>>>   706.745
```

Re-conceptualising the correlation

```
 MATRIX S
This is a computed FULL m
 [=\STND(V)]
           1         2
1    1.0000    0.3791
2    0.3791    1.0000

 MATRIX V
This is a SYMMETRIC matri
              1           2
1     107.8418
2      40.8812    107.8418
```

- The correlation is the standardised covariance – an index of the amount of variance shared between the 2 variables or 2 individuals
- The variance of each variable is composed of that which is shared and that which is not
  - V=S+N

Re-conceptualising the correlation

```
 MATRIX S
This is a computed FULL m
 [=\STND(V)]
          1        2
1    1.0000   0.3791
2    0.3791   1.0000

 MATRIX V
This is a SYMMETRIC matri
             1           2
1     107.8418
2      40.8812    107.8418
```

- V=S+N
- So another way to model the covariance would be

  S+N | S_

  S | S+N ;

```
#define nvar 1                    ! n dependent variables per individual
#define nsib 2                    ! number if siblings per family

G1: Corelation between twin pairs
Data Ninput_vars=8 NGroups=1              ! Number of variables per family
missing=-1                                ! Missing values

Rectangular file=wednesday.dat            ! read raw data
Labels
ntrid zygMZDZ age1 sekse1  AQ1 age2 sekse2 AQ2

select AQ1 AQ2 ;

Begin matrices;
    M Full 1 nvar free           !mean
    S Full nvar nvar free        ! shared variance
    N Full nvar nvar free        ! non-shared variance
End matrices;

Begin Algebra;
    V = S+N| S _
        S | S+N ;
    C = \stnd(V) ; ! compute Standard deviation
End Algebra;

! start values
Start 100 M 1 1
Start 50 S 1 1 N 1 1

Means M|M ;      ! means model
Covariances V ;          ! variance/covariance model
end
```

Alt_correlation.mx

Results

- Correlation.mx

```
 MATRIX V
This is a SYMMETRIC matrix of order    2 by
             1           2
1    107.8418
2     40.8812   107.8418

Your model has    3 estimated parameters and

-2 times log-likelihood of data >>>  2584.422
Degrees of freedom >>>>>>>>>>>>>>>>       344
```

- Alt_correlation.mx

```
 MATRIX N
This is a FULL matrix of order    1 by    1
             1
1     66.9606

 MATRIX S
This is a FULL matrix of order    1 by    1
             1
1     40.8812

 MATRIX V
This is a computed FULL matrix of order    2 b
 [=S+N|S_S|S+N]
             1           2
1    107.8418    40.8812
2     40.8812   107.8418

Your model has    3 estimated parameters and

-2 times log-likelihood of data >>>  2584.422
Degrees of freedom >>>>>>>>>>>>>>>>       344
```

Twin modelling takes this 1 step further

- Covariance

  Var1 | Cov_

  Cov | Var 2 ;

- MZ

  A+C+E | A+C_

  A+C | A+C+E ;

- DZ

  A+C+E | .5*A+C_

  .5*A+C | A+C+E ;

- A = Additive genetic variance
- C = Common environment variance
- E = unique Environmental variance (including measurement Error)

Working with ordinal data

Binary data

```
1 0 0   25.80 1
1 0 0   21.10 1
1 0 0   21.79 1
1 0 0   21.12 1
1 0 0   32.05 1
1 0 0   37.41 1
1 0 0   33.56 0
1 0 0   32.78 1
1 0 0   39.19 0
```

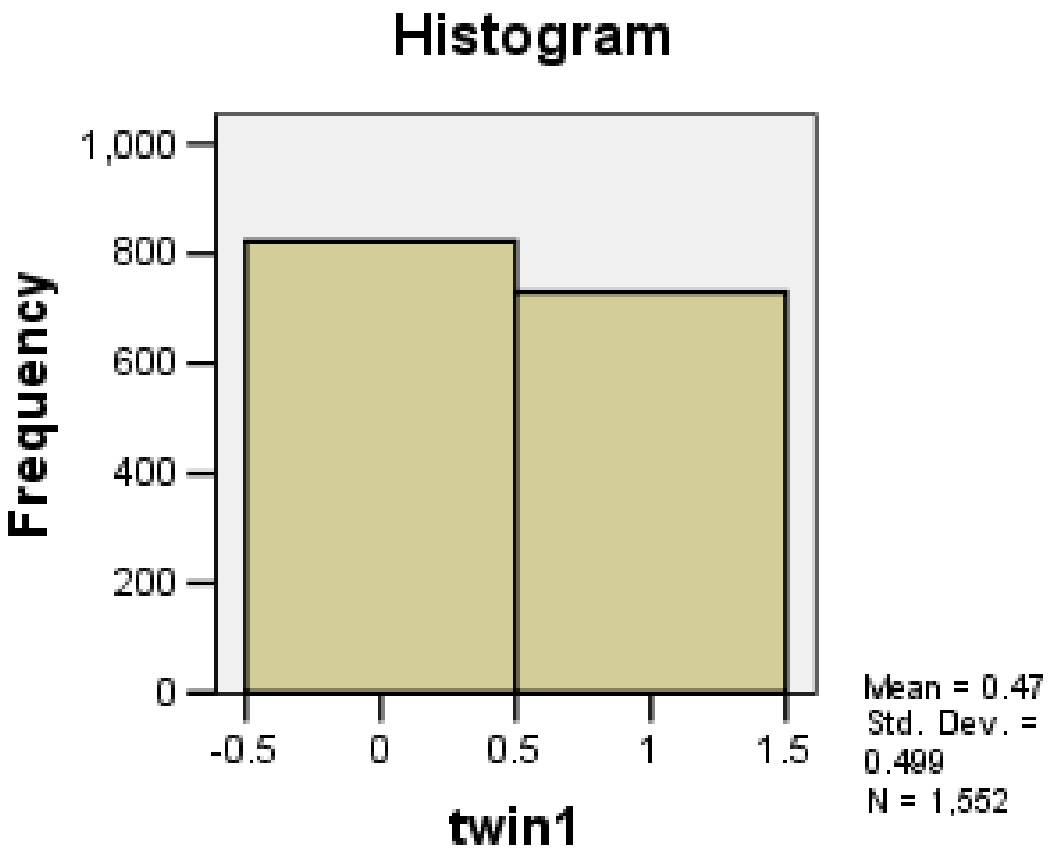
- File: two_cat.dat
- NI=5
- Labels Zyg twin1 twin2 Age Sex
- Trait – smoking initiation
  - Never Smoked/Ever Smoked
  - (Recoded from yesterday)
  - Data is sorted to speed up the analysis

Twin 1 smoking initiation

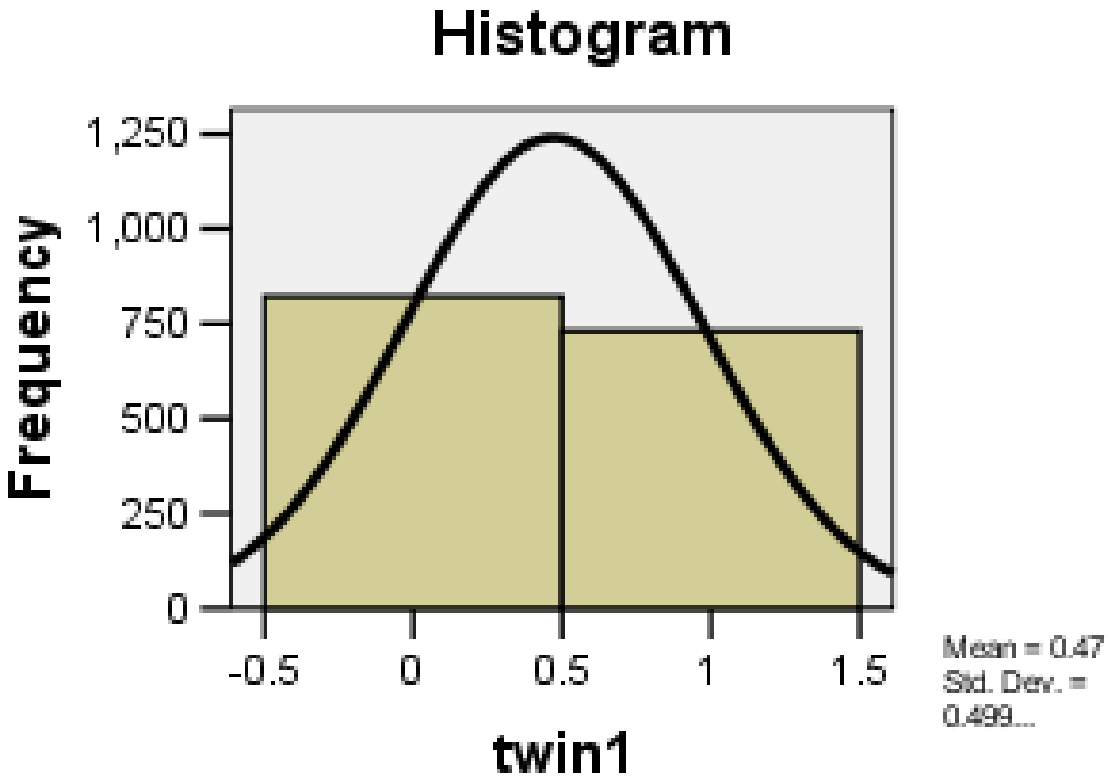
twin1

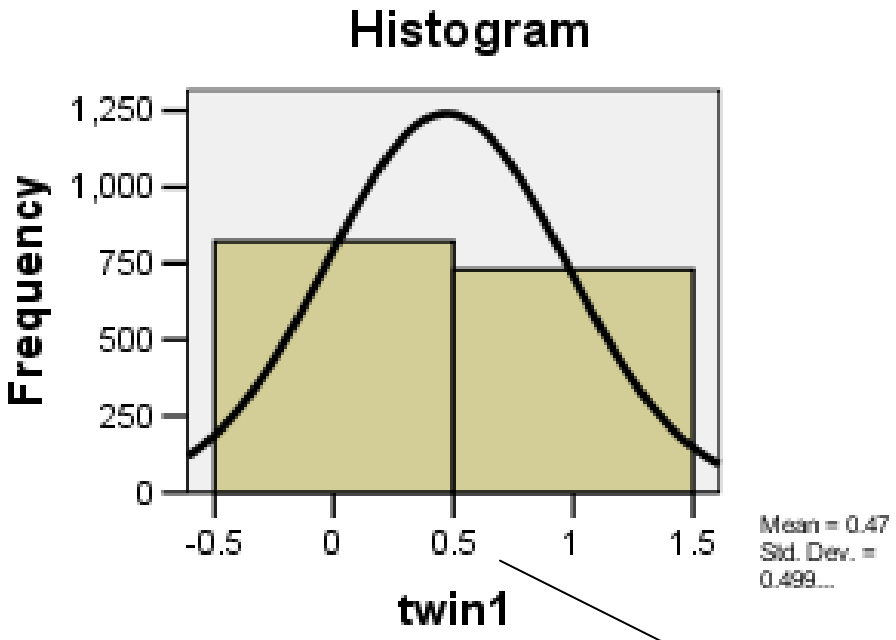
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	822	47.5	53.0	53.0
	1	730	42.2	47.0	100.0
	Total	1552	89.7	100.0	
Missing	System	179	10.3		
Total		1731	100.0		

Twin 1 smoking initiation



Twin 1 smoking initiation

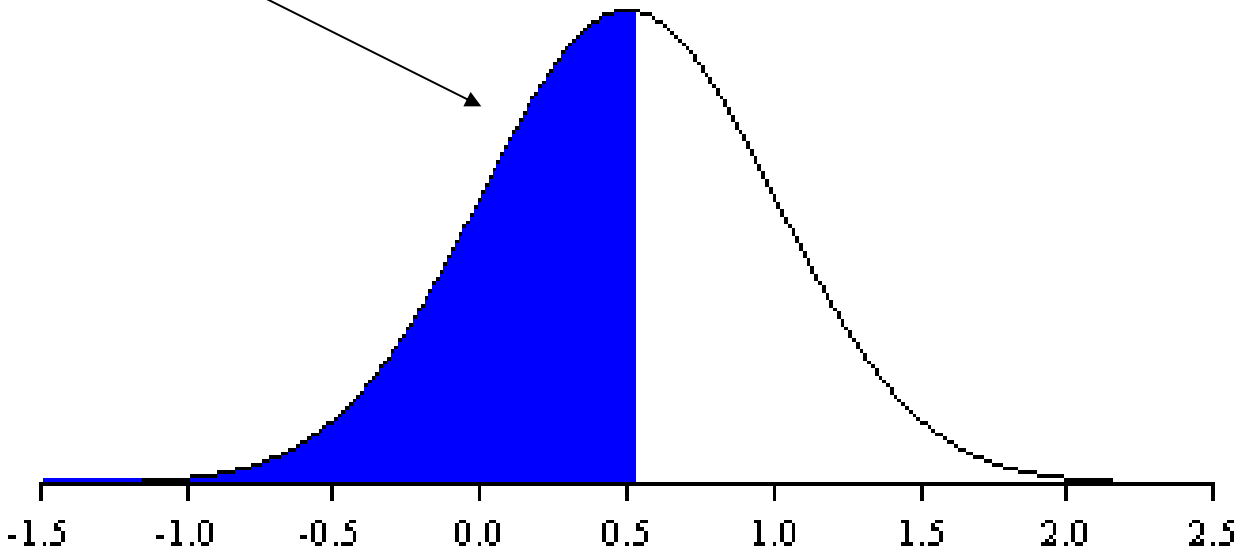


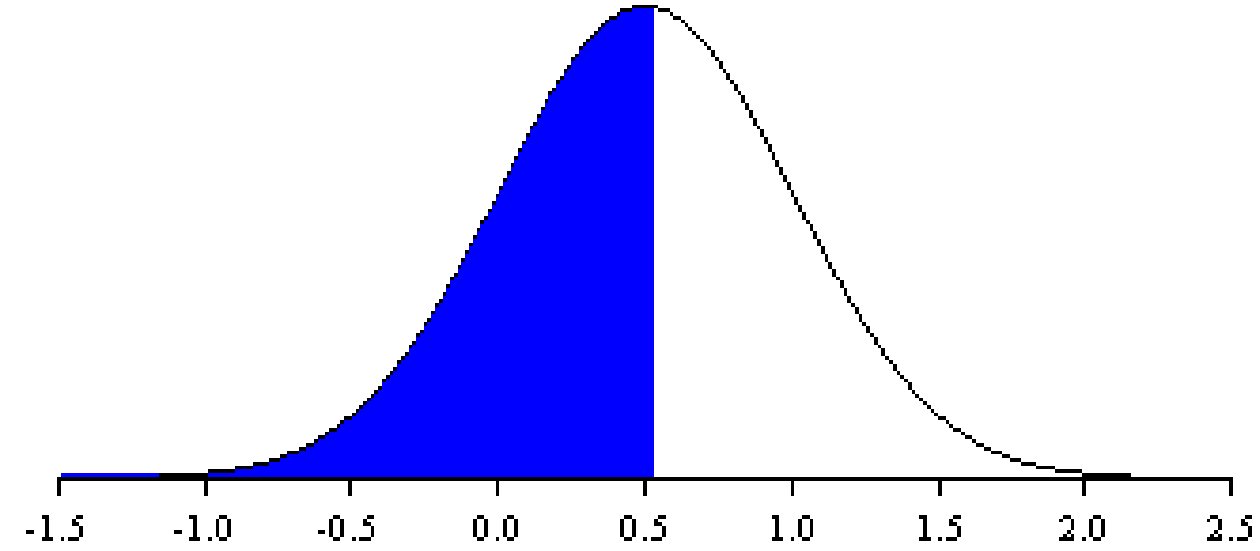


Mean = .47

SD =.499

Non Smokers =53%





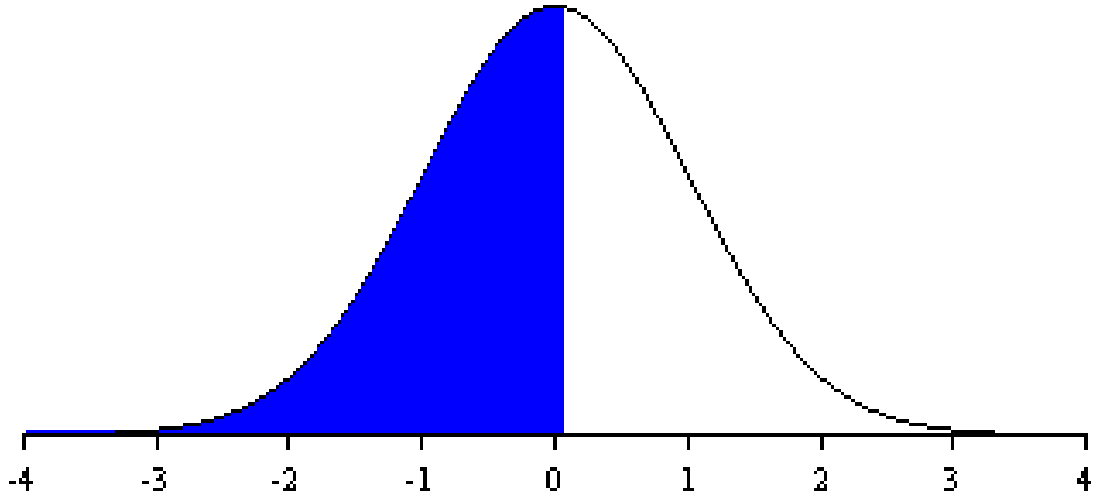
Raw data distribution

Mean = .47

SD =.499

Non Smokers =53%

Threshold =.53



Standard normal distribution

Mean = 0

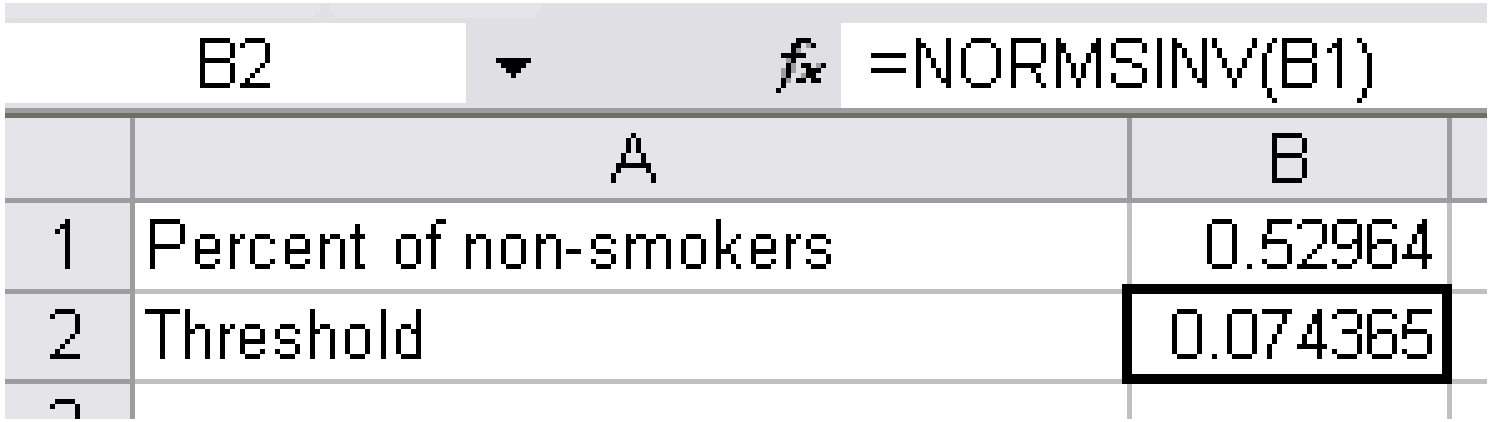
SD =1

Non Smokers =53%

Threshold =.074

Threshold = .074 – Huh what?

- How can I work this out
  - Excell
    - =NORMSINV()



Why do we rescale the data this way?

- Convenience
  - Variance always 1
  - Mean is always 0
  - We can interpret the area under a curve between two z-values as a probability or percentage

Why do we rescale the data this way?

You could use other distributions but **you** would have to specify the fit function

Threshold.mx

```
#define nvar 1

G1: Singleton (non-pair) ordinal data
Data Ninput_vars=5 NGroups=1

Ordinal file=two_cat.dat
Labels
Zyg Twin1 Twin2 Age Sex

select twin1 ;

Begin matrices;
    T Full 1 nvar free          ! Threshold
    V Stand nvar nvar free      ! variance
End matrices;

! start values
Start 0 T 1 1

Thresholds T ;
Covariances V ;          ! variance/covariance model
end
```

B2 =NORMSINV(B1)

	A	B
1	Percent of non-smokers	0.52964
2	Threshold	0.074365

```
G1: Singleton (non-pair) ordinal data

  MATRIX T
 This is a FULL matrix of order    1 by    1
            1
 1    0.0744

  MATRIX V
 This is a STANDARDISED matrix of order    1 by    1
            1
 1    1.0000
```

How about age/sex correction? binary_regression.mx

```
#define nvar 1
#define ndef 2

G1: Singleton (non-pair) ordinal data
Data Ninput_vars=5 NGroups=1

Ordinal file=two_cat.dat
Labels  Zyg Twin1 Twin2 Age Sex

select twin1 Age Sex;
definition Age Sex;

Begin matrices;
    T Full 1 nvar free          ! Threshold
    B Full ndef 1 free          ! Beta
    D Full 1 ndef               ! Age
    V Stand nvar nvar free      ! variance
End matrices;

! start values
Start 0 T 1 1

! place covariate in d matrix
Specify D Age Sex

Thresholds T + D*B ;
Covariances V ;         ! variance/covariance model
end
```

How about age/sex correction?

```
G1: Singleton (non-pair) ordinal data

 MATRIX B
 This is a FULL matrix of order    2 by    1
           1
 1    0.0070
 2   -0.0500

 MATRIX D
 This is a FULL matrix of order    1 by    2
              1           2
 1     37.1300     0.0000

 MATRIX T
 This is a FULL matrix of order    1 by    1
           1
 1   -0.1118
```

What does this mean?

- Age Beta = .007
  - For every 1 unit increase in Age the threshold shifts .007

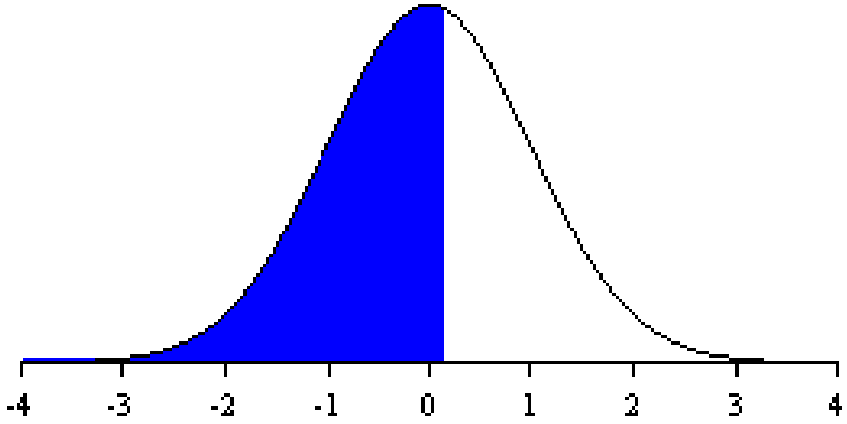
Summary of VL file data for group 1

	SEX	AGE	TWIN1
Code	-2.0000	-1.0000	1.0000
Number	1552.0000	1552.0000	1552.0000
Mean	0.4826	30.0439	0.4704
Variance	0.2497	33.1815	0.2491
Minimum	0.0000	20.0000	0.0000
Maximum	1.0000	39.9800	1.0000

What does this mean?

- Beta = .007
- Threshold is -.1118

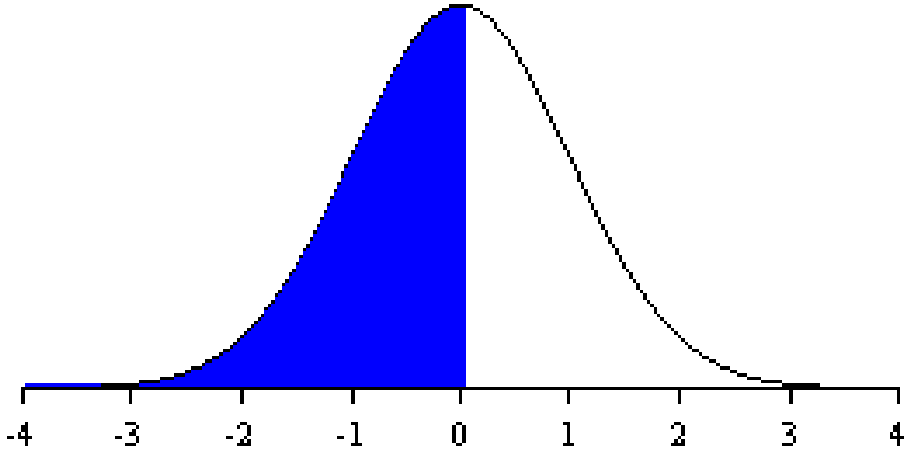
- 38 is +1.38 SD from the mean age
  - The threshold for 38 year olds is: .1544= -.1118 + .007*38
- 22 is -1.38 SD from the mean age
  - The threshold for 38 year olds is: .0422= -.1118 + .007*22



38 year olds

Threshold =.1544

Is the age effect significant?



22 year olds

Threshold = .0422

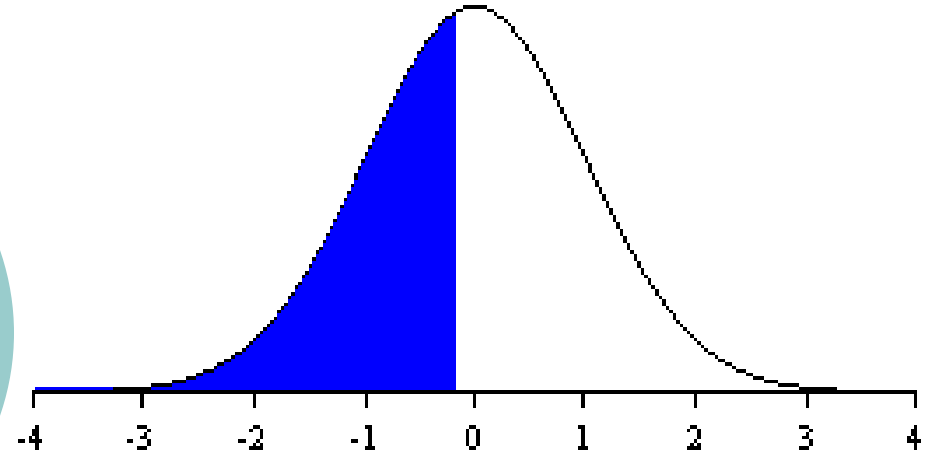
How to interpret this

- The threshold moved slightly to the right as age increases
- This means younger people were more likely to have tried smoking than older people
  - But this was not significant

How about the sex effect

- Beta = -.05
- Threshold = -.1118
- Sex coded Male = 1, Female = 0

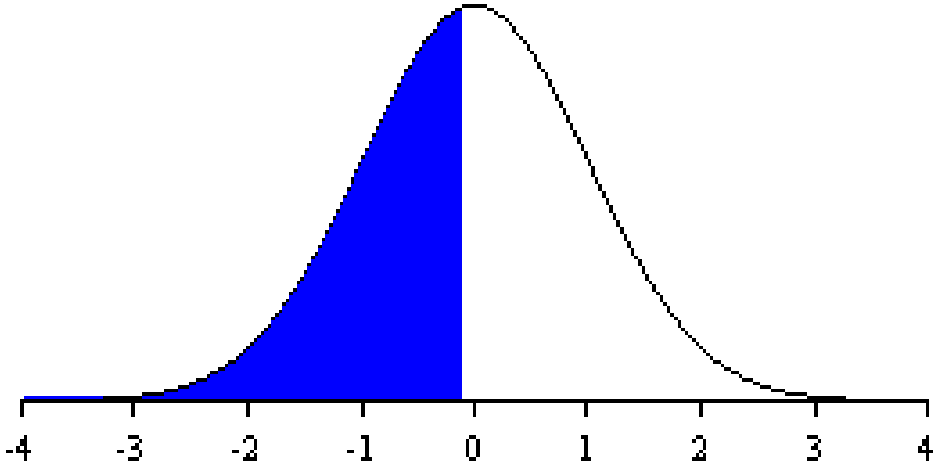
- So the Male threshold is:
  - -.1618= -.1118 + 1*-.05
- The Female threshold is:
  - -.1118 =-.1118 + 0*-.05



Male

Threshold =-.1618

Are males or females more likely to smoke?



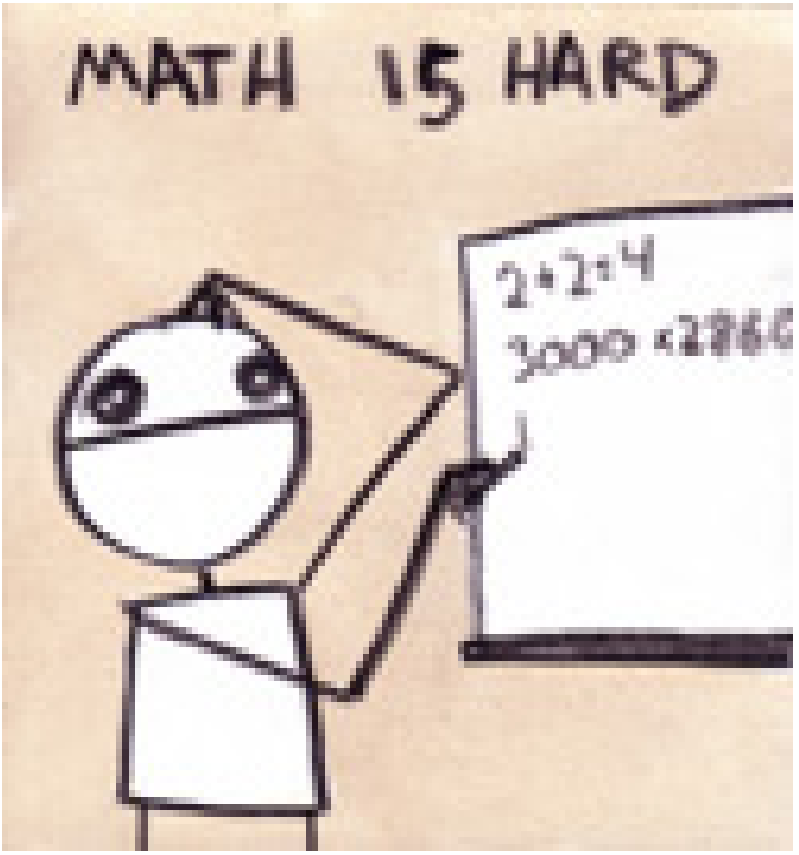
Female

Threshold =-.1118

General Advice/Problem solving

- Scripting styles differ
- Check the parameter numbers
- Check the sample description
- Learn to love the manual
- Comments are your friends

Time for coffee



explodingdog.com