Ordinal (yet again)

Sarah Medland – Boulder 2010

This morning

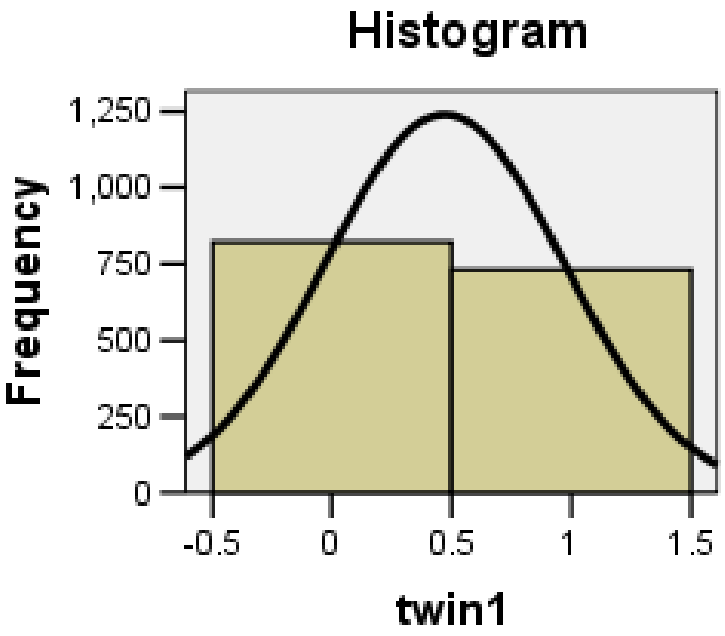
- Fitting the regression model with ordinal data
- Factor Modelling
  - Continuous
  - Ordinal

Including covariates

- At the moment
  - No missingness allowed in the covariates
  - If the person who is missing the cov is also missing the phenotype(s) being modelled - ie incomplete twin pair
    - Replace with the mean

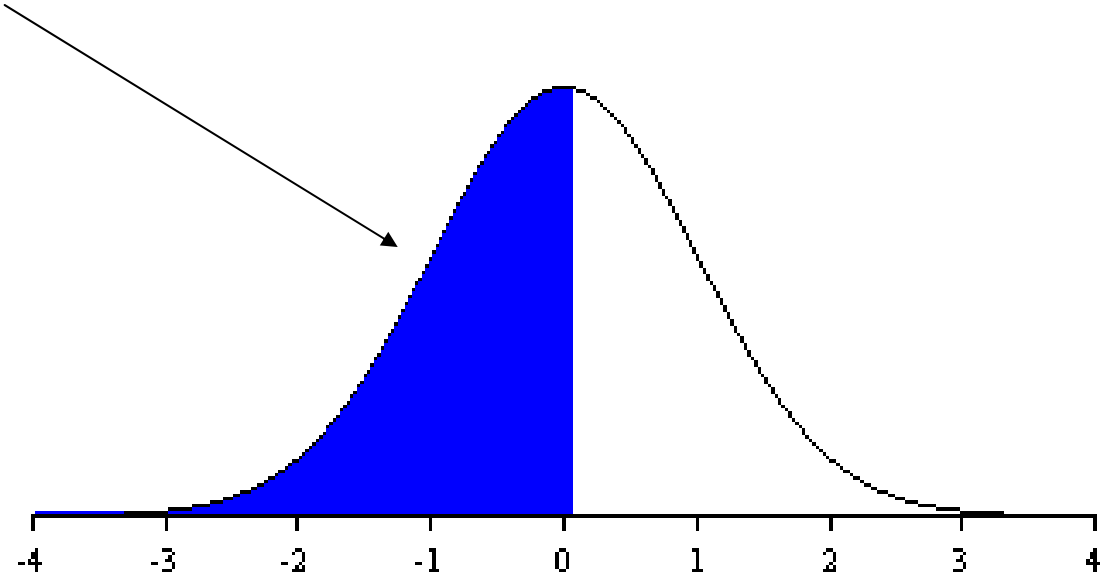
From Tuesday...

- Thresholds T ; $[t_{11}]$



Standard normal distribution

Mean	= 0
SD	=1
Non Smokers	=53%
Threshold	=.074



How about age/sex correction?

- threshold_age_sex.R

```
require(OpenMx)
Canabis <- read.table ('two_cat.dat', header=T )

# Print Descriptive Statistics
# ----------------------------------------------------------------------
summary(Canabis$twin1)
table(Canabis$twin1)

# Select data
# ----------------------------------------------------------------------
Canabis1 <-data.frame(Canabis$twin1, Canabis$Age, Canabis$Sex)
names(Canabis1) <- c("twin1", "Age", "Sex")

# Specify and Run Saturated Model (Tetrachoric correlations) with RawData
# ----------------------------------------------------------------------
nvar <- 1
nthresh <- 1
Vars <-('twin1')
selVars <- ('twin1')
defVars <- c('age','sex')
```

How about age/sex correction?

```
AgeSexRegressionModel<- mxModel("AgeSexRegression",
#Bris<-
mxModel("BrisbaneData",
  # Matrix & Algebra for expected means vector (SND), Thresholds and correlation
        mxMatrix(type="Zero", nrow=nvar, ncol=nvar, name="expMean"),
        mxMatrix(type="Full", nrow=nvar, ncol=nthresh, free=TRUE, values=0,
                 name="ucThresh", label="Uncorrected_Threshold"),
        mxMatrix(type="Stand", nrow=nvar, ncol=nvar, name="expCor"),

#Matrices for age and sex corrections
        mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels="data.Age", name="Age"),
        mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels="data.Sex", name="Sex"),
        mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE, values=0.5, labels="AgeBeta", name="AgeB"),
        mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE, values=0.5, labels="SexBeta", name="SexB"),
        mxAlgebra( expression= AgeB%*%Age + SexB%*%Sex, name="AgeSexR"),
        mxAlgebra( expression= (ucThresh + AgeSexR), name="expThresh", dimnames=list(NA,selVars) ),

        mxData(Canabis1, type="raw"),
        mxFIMLObjective( covariance="expCor", means="expMean", thresholds="expThresh",  dimnames=selVars )),
        mxAlgebra(BrisbaneData.objective, name="-2LL" ),
        mxAlgebraObjective("-2LL")

 )
AgeSexRegressionFit <- mxRun(AgeSexRegressionModel)
AgeSexRegressionSumm <- summary(AgeSexRegressionFit)
AgeSexRegressionSumm
```

```
#Matrices for age and sex corrections
mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE,
        labels="data.Age", name="Age"),
mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE,
        labels="data.Sex", name="Sex"),
mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE, values=0.01,
        labels="AgeBeta", name="AgeB"),
mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE, values=0.01,
        labels="SexBeta", name="SexB"),
mxAlgebra( expression= AgeB%*%Age + SexB%*%Sex, name="AgeSexR"),
mxAlgebra( expression= (ucThresh + AgeSexR), name="expThresh",
        dimnames=list(NA,selVars) ),
```

$$\text{AgeSexR} = [\text{Age}\beta] * [\text{Age}] + [\text{Sex}\beta] * [\text{Sex}]$$

$$\text{expThresh} = [\text{ucThresh}] + [\text{AgeSexR}]$$

How about age/sex correction?

- threshold_age_sex.R

```
> AgeSexRegressionFit <- mxRun(AgeSexRegressionModel)
Running AgeSexRegression
> AgeSexRegressionSumm <- summary(AgeSexRegressionFit)
> AgeSexRegressionSumm
                   name                matrix row col     Estimate   Std.Error
1 Uncorrected_Threshold BrisbaneData.ucThresh   1   1 -0.111843829 0.171943487
2               AgeBeta     BrisbaneData.AgeB   1   1  0.007004344 0.005534061
3               SexBeta     BrisbaneData.SexB   1   1 -0.049993953 0.063750547
```

What does this mean?

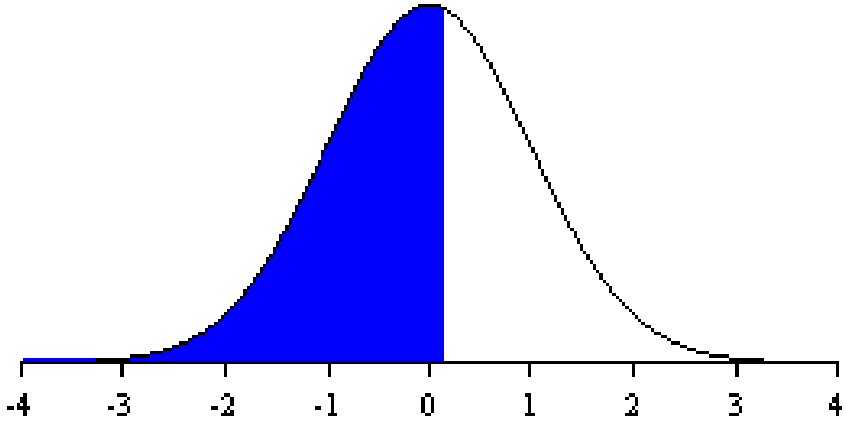
- Age Beta = .007
  - For every 1 unit increase in Age the threshold shifts .007

```
> AgeSexRegressionFit <- mxRun(AgeSexRegressionModel)
Running AgeSexRegression
> AgeSexRegressionSumm <- summary(AgeSexRegressionFit)
> AgeSexRegressionSumm
                   name                  matrix row col     Estimate   Std.Error
1 Uncorrected_Threshold BrisbaneData.ucThresh   1   1 -0.111843829 0.171943487
2               AgeBeta     BrisbaneData.AgeB   1   1  0.007004344 0.005534061
3               SexBeta     BrisbaneData.SexB   1   1 -0.049993953 0.063750547
```

What does this mean?

- Beta = .007
- Threshold is -.1118

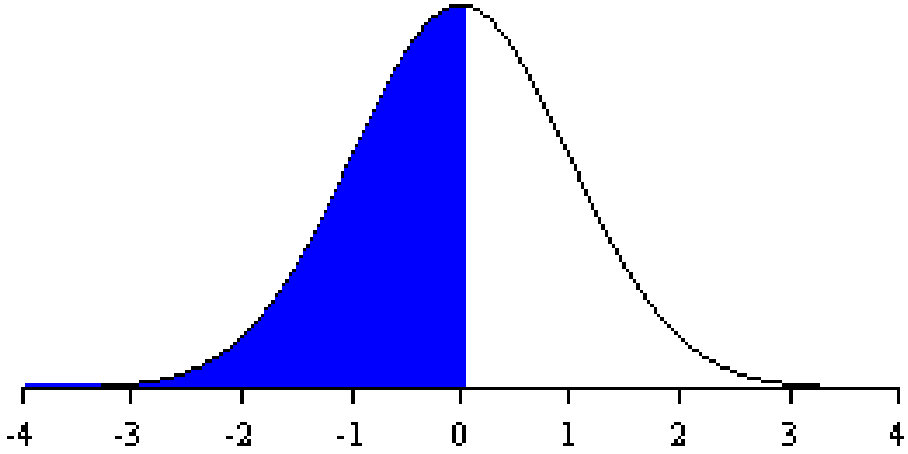
- 38 is +1.38 SD from the mean age
  - The threshold for 38 year olds is: .1544= -.1118 + .007*38
- 22 is -1.38 SD from the mean age
  - The threshold for 38 year olds is: .0422= -.1118 + .007*22



38 year olds

Threshold =.1544

Is the age effect significant?



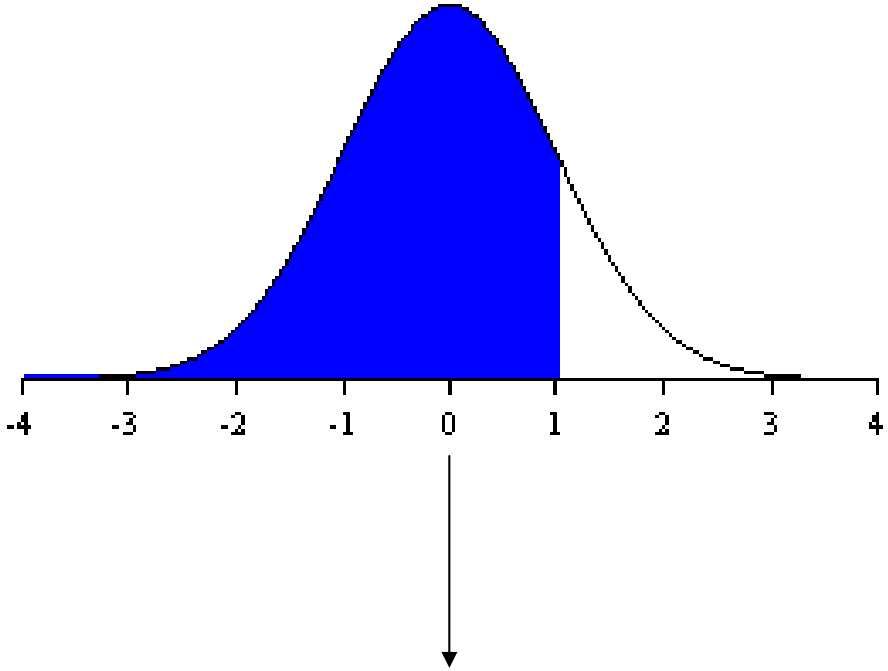
22 year olds

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

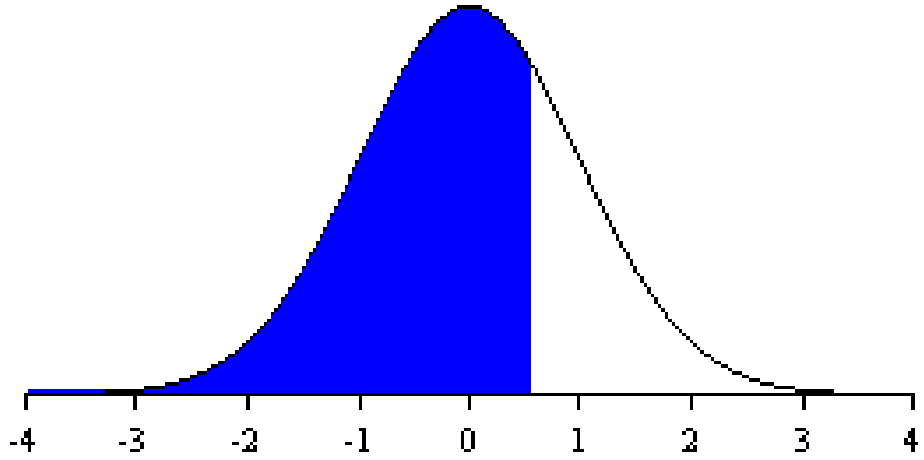
If Beta = .03



38 year olds

Threshold = 1.028

Is the age effect significant?



22 year olds

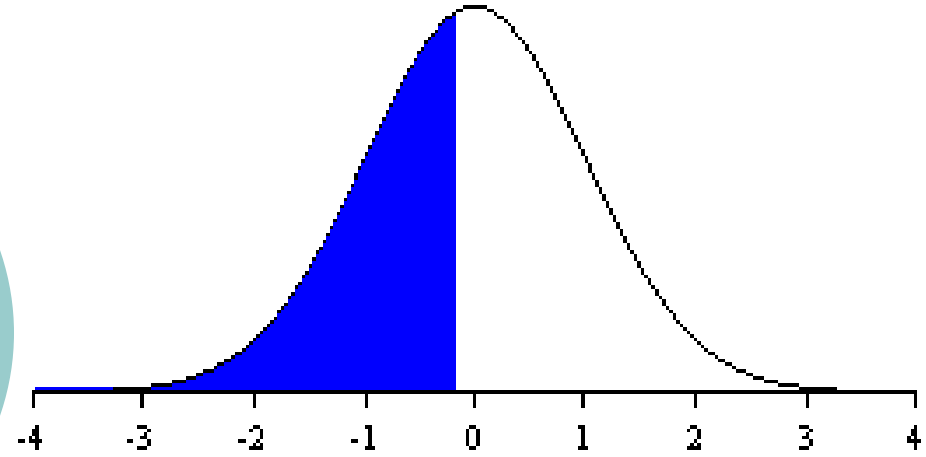
Threshold = .548

How about the sex effect

```
> AgeSexRegressionFit <- mxRun(AgeSexRegressionModel)
Running AgeSexRegression
> AgeSexRegressionSumm <- summary(AgeSexRegressionFit)
> AgeSexRegressionSumm
                   name                matrix row col     Estimate   Std.Error
1 Uncorrected_Threshold BrisbaneData.ucThresh   1   1 -0.111843829 0.171943487
2               AgeBeta     BrisbaneData.AgeB   1   1  0.007004344 0.005534061
3               SexBeta     BrisbaneData.SexB   1   1 -0.049993953 0.063750547
```

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

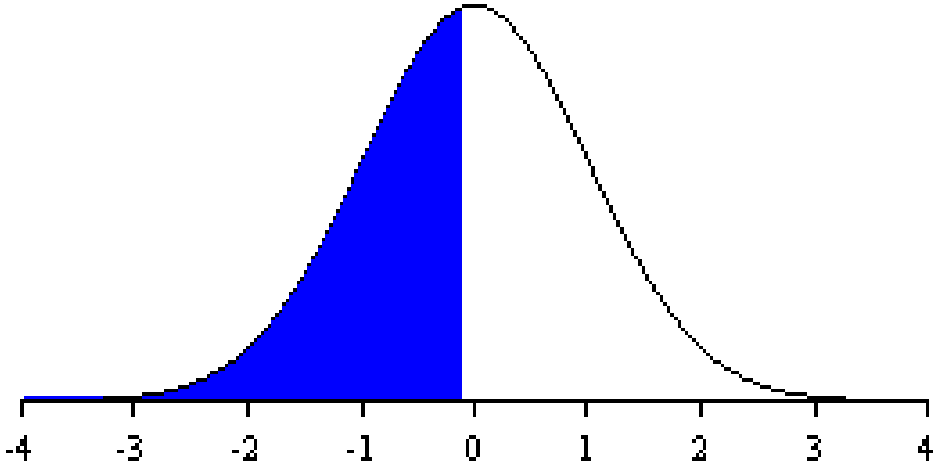
- So the Male threshold is:
  - -.1618= -.1118 + 1*-.05
- The Female threshold is:
  - -.1118 =-.1118 + 0*-.05



Male

Threshold =-.1618

Are males or females more likely to smoke?



Female

Threshold =-.1118

Both effects together

- 38 year old Males:
  - .1042=-.1118 + 1*-.05 + .007*38
- 38 year old Females:
  - .1542=-.1118 + 0*-.05 + .007*38
- 22 year old Males:
  - -.0078=-.1118 + 1*-.05 + .007*22
- 22 year old Females:
  - .0422=-.1118 + 0*-.05 + .007*22

Polychotomous Data?

Binary Data… adding a regression

$$\text{AgeSexR} = [\text{Age}\beta] * [\text{Age}] + [\text{Sex}\beta] * [\text{Sex}]$$

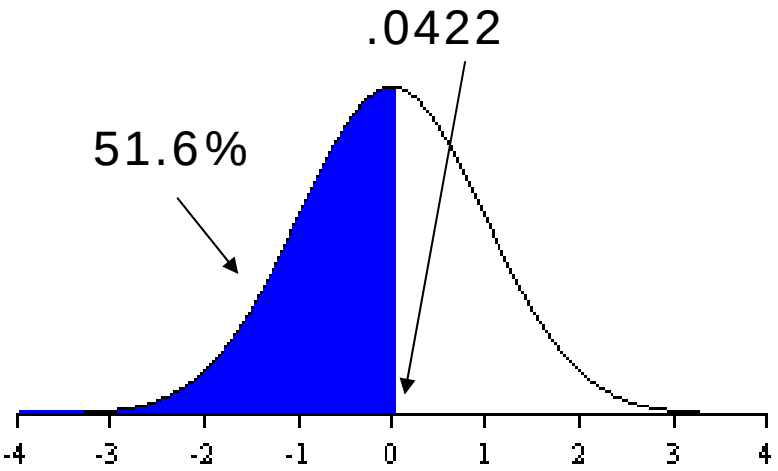
$$\text{expThresh} = [\text{ucThresh}] + [\text{AgeSexR}]$$

$$= [-.1118 + Age * .007 + Sex * -.05]$$

$$\text{if Age} = 22 \text{ and Sex} = 1 \text{ (Male)}$$

$$= [-.1118 + 22 * .007 + 1 * -.05]$$

$$= [.0422]$$

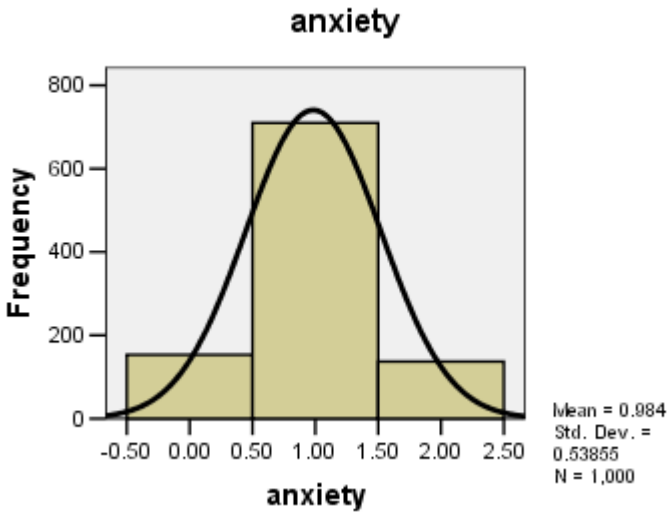


What about more than 2 categories?

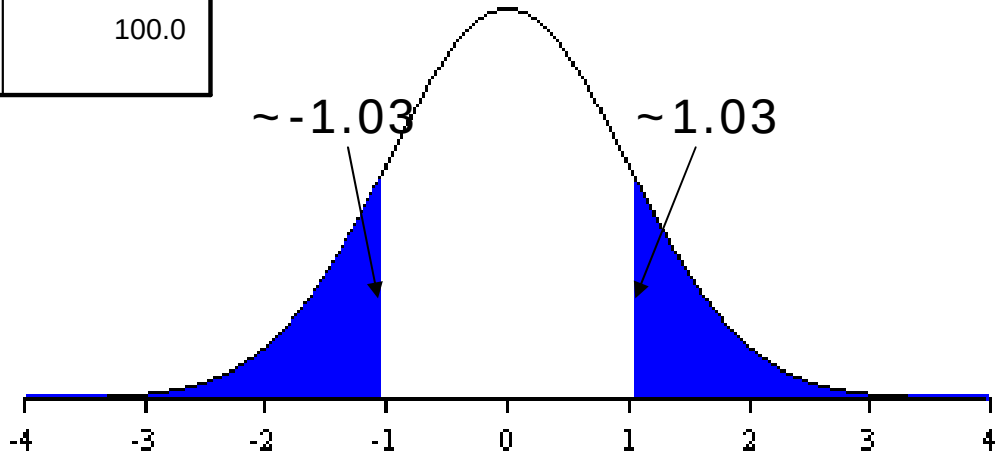
- Thresholds = L*T;

anxiety

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	153	15.3	15.3	15.3
	1.00	710	71.0	71.0	86.3
	2.00	137	13.7	13.7	100.0
	Total	1000	100.0	100.0	



~15% in each tail
Thresholds:



Displacement = ~2.06

What about more than 2 categories?

- Thresholds = L*T;

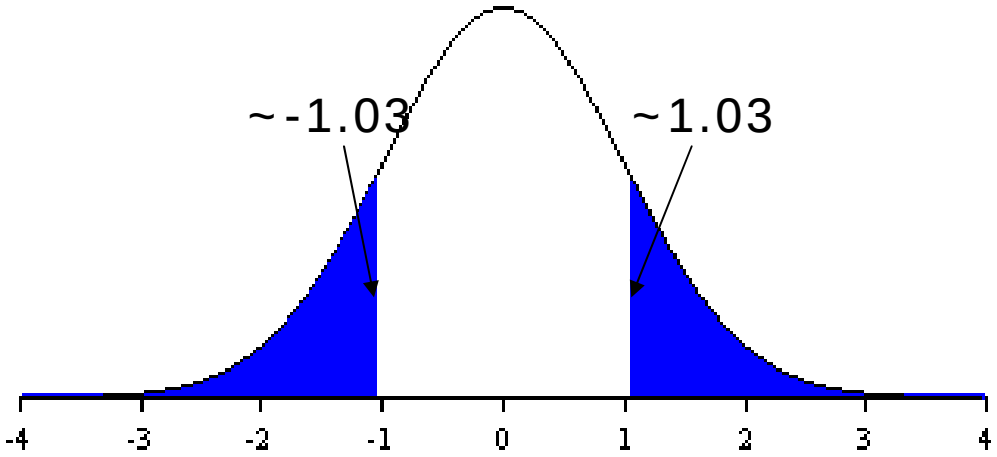
$$= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} t_{11} \\ t_{21} \end{bmatrix}$$

$$= \begin{bmatrix} 1*t_{11} + 0*t_{21} \\ 1*t_{11} + 1*t_{21} \end{bmatrix}$$

$$= \begin{bmatrix} -1.03 \\ -1.03 + 2.06 \end{bmatrix}$$

$$= \begin{bmatrix} -1.03 \\ 1.03 \end{bmatrix}$$

~15% in each tail
Thresholds:



Displacement = ~2.06

Adding a regression

$$\text{AgeSexR} = [\text{Age}\beta] * [\text{Age}] + [\text{Sex}\beta] * [\text{Sex}]$$

$$\text{Duplicate} = \begin{bmatrix} 1 \\ \\ 1 \end{bmatrix} \qquad (\text{unit maxthresh},1)$$

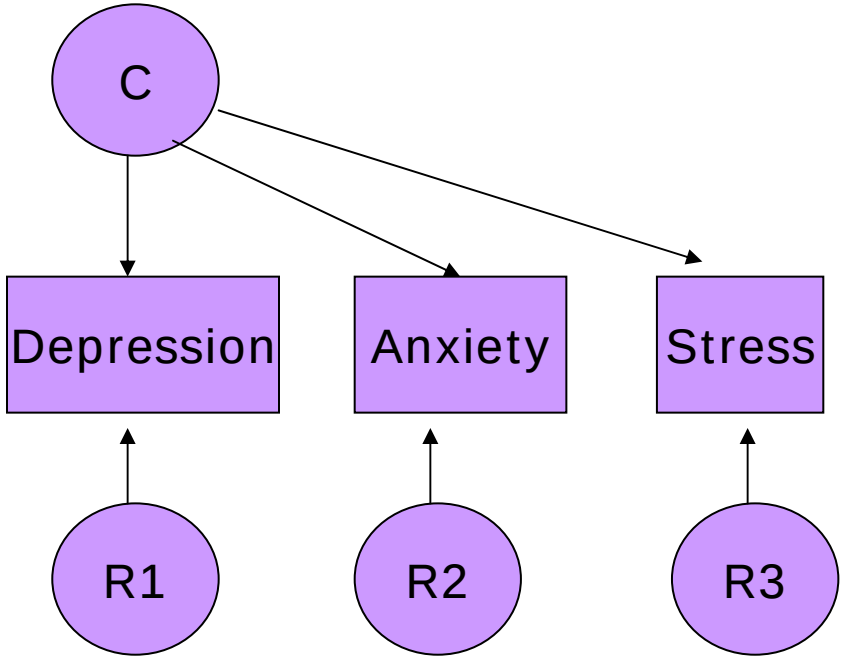
$$\text{expThresh} = ([\text{L}] * [\text{ucThresh}]) + ([\text{Duplicate}] \otimes [\text{AgeSexR}])$$

$$\begin{bmatrix} t11 + \beta sex * sex1 + \beta age * age1 \\ (t11 + t21) + \beta sex * sex1 + \beta age * age1 \end{bmatrix}$$

Multivariate models?

Factor Analysis

- Suppose we have a theory that the covariation between self reports of depression, anxiety and stress levels is due to one underlying factor

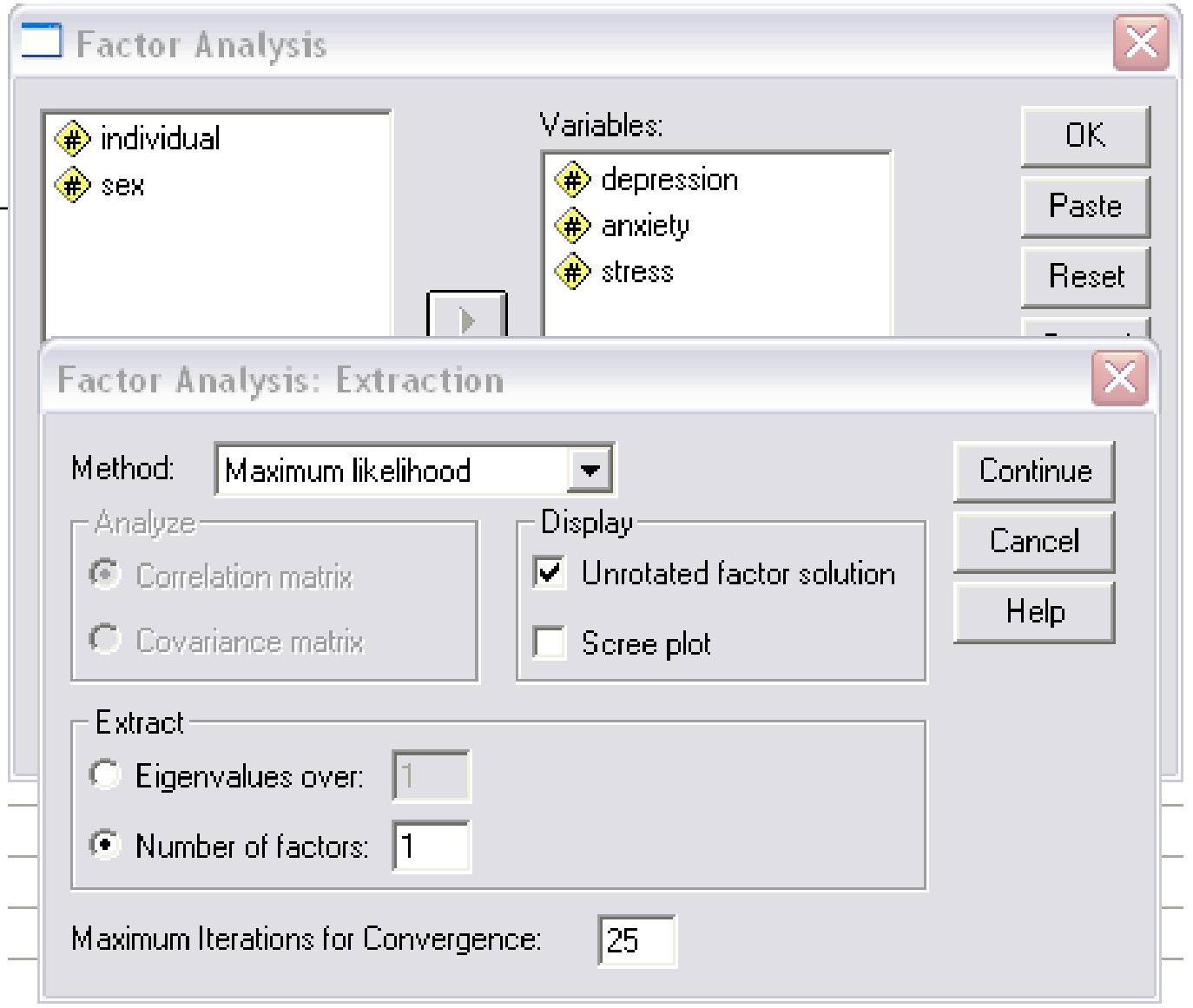


Factor Analysis….

- Our data (simulated)
  - Five variables – Three traits
  - Depression, Anxiety & Stress
  - Transformed to Z-scores

	individual	depression	anxiety	stress	sex
1	1.0	.87	-.49	.52	0
2	2.0	-1.08	.38	-.05	0
3	3.0	-.83	-.21	-1.14	0
4	4.0	-.15	-1.16	-.61	0
5	5.0	1.06	.57	.42	0
6	6.0	-.53	-1.45	-1.71	0
7	7.0	-.53	.33	.68	0
8	8.0	.31	.64	-.52	0
9	9.0	-1.38	-.47	-1.80	0

In Spss..


Factor Analysis
individual
sex
Variables:
depression
anxiety
stress
OK
Paste
Reset
Factor Analysis: Extraction
Method:
Maximum likelihood
Continue
Cancel
Help
Analyze
Correlation matrix
Covariance matrix
Display
Unrotated factor solution
Scree plot
Extract
Eigenvalues over:
1
Number of factors:
1
Maximum Iterations for Convergence:
25

Factor Analysis

Communalities

	Initial	Extraction
depression	.415	.774
anxiety	.325	.408
stress	.257	.319

Extraction Method: Maximum Likelihood.

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.951	65.045	65.045	1.501	50.037	50.037
2	.644	21.466	86.511			
3	.405	13.489	100.000			

Extraction Method: Maximum Likelihood.

Factor Matrix[a]

	Factor
	1
depression	.880
anxiety	.639
stress	.565

Extraction Method: Maximum Likelihood.

a. 1 factors extracted. 5 iterations required.

What do we get?

Factor Matrix[a]

	Factor
	1
depression	.880
anxiety	.639
stress	.565

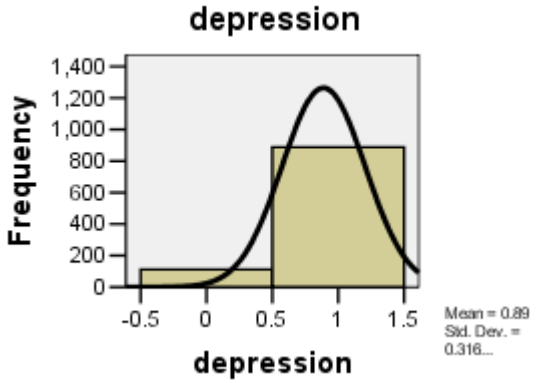
Extraction Method: Maximum Likelihood.

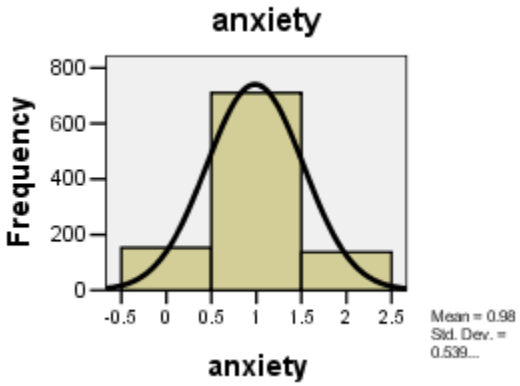
a. 1 factors extracted. 5 iterations required.

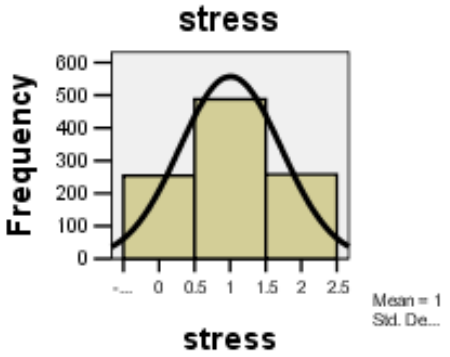
```
 MATRIX S
This is a computed FULL matrix
 [=L%V|(\D2V(R)')%V]
            1         2
1      0.8795    0.4758
2      0.6390    0.7692
3      0.5649    0.8251
```

What if our data was ordinal?

- Depression
  - Yes/No 0/1
- Anxiety and Stress
  - Low / Average / High 0/1/2







Spss says no

Hide details

Data. The variables should be quantitative at the **interval** or **ratio** level. Categorical data (such as religion or country of origin) are not suitable for factor analysis. Data for which Pearson correlation coefficients can sensibly be calculated should be suitable for factor analysis.

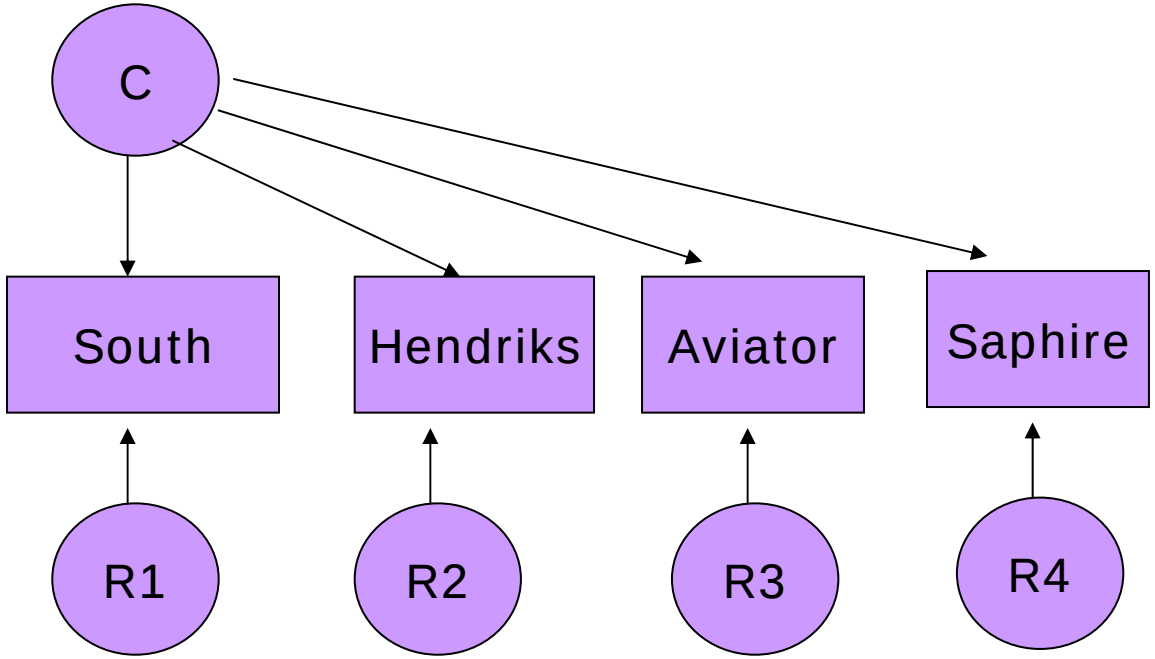
Assumptions. The data should have a bivariate normal distribution for each pair of variables, and observations should be independent. The factor analysis model specifies that variables are determined by common factors (the factors estimated by the model) and unique factors (which do not overlap between observed variables); the computed estimates are based on the assumption that all unique factors are uncorrelated with each

Table 2.2: Classification of correlations according to their observed distribution.

Measurement	Two Categories	Three or more Categories	Continuous
Two	Tetrachoric	Polychoric	Biserial
Three or more	Polychoric	Polychoric	Polyserial
Continuous	Biserial	Polyserial	Product Moment

Factor Analysis

- Suppose we asked people in a nightclub what their favourite drinks were
  - Series of binary variables



OrdinalFactors.r

```
# Fit Multivariate Saturated Model
# ------------------------------------------------------------------------
 OrdinalFactorModel <- mxModel("OrdinalFactor",
    mxModel("Fact",
        mxMatrix( type="Full", nrow=nvar, ncol=nf, free=TRUE, values=.3, name="factor" ),
        mxMatrix( type="Diag", nrow=nvar, ncol=nvar, free=TRUE, values=.3, name="residual" ),
        mxMatrix( type="Stand", nrow=nvar, ncol=nvar, free=TRUE, name="Vconstraint" ),
        mxAlgebra( expression=((factor %*% t(factor))+(residual%*% t(residual))), name="expCov" ),
        mxConstraint( alg1="expCov", "=", alg2="Vconstraint", name="Constraint1"),

        mxMatrix( type="Full", nrow=1, ncol=nvar, free=FALSE, values=0, name="expMean" ),
        mxMatrix( type="Lower", nrow=maxthresh, ncol=maxthresh, free=FALSE, values=1, name="Increment" ),
        mxMatrix( type="Full", nrow=maxthresh, ncol=nvar, free=TRUE, values=c(0.1,-.3,0.1,-.3),
                lbound=-4,name="cutpoints"),
        mxAlgebra( expression= (Increment %*% cutpoints), dimnames=list(thresh,selVars) , name="expThresh"),

        mxData( observed=fullData, type="raw" ),
        mxFIMLObjective( covariance="expCov", means="expMean", dimnames=selVars, thresholds="expThresh")
    ),
    mxAlgebra( Fact.objective, name="-2LL" ),
    mxAlgebraObjective("-2LL")
)
```

Adding Sex and Age effects

Setting up the factor model and the ‘residual’ thresholds

```
 OrdinalFactorModelCOV <- mxModel("OrdinalFactor2",
    mxModel("Fact2",
# specifing the data
        mxData( observed=fullData, type="raw" ),
        mxMatrix( type="Full", nrow=nvar, ncol=nf, free=TRUE, values=c(.5, .7,.8,.8), name="factor" ),
        mxMatrix( type="Diag", nrow=nvar, ncol=nvar, free=TRUE, values=c(.8, .7,.5,.6), name="residual" ),
        mxMatrix( type="Stand", nrow=nvar, ncol=nvar, free=TRUE, name="Vconstraint" ),
        mxAlgebra( expression=((factor %*% t(factor))+(residual%*% t(residual))), name="expCov" ),
        mxConstraint( alg1="expCov", "=", alg2="Vconstraint", name="Constraint1"),

# setting up the basic thresholds model
        mxMatrix( type="Full", nrow=1, ncol=nvar, free=FALSE, values=0, name="expMean" ),
        mxMatrix( type="Lower", nrow=maxthresh, ncol=maxthresh, free=FALSE, values=1, name="Increment" ),
        mxMatrix( type="Full", nrow=maxthresh, ncol=nvar, free=TRUE, values=c(0,-.7,.6,-.5),
                lbound=-4,name="cutpoints"),
        mxAlgebra( expression= (Increment %*% cutpoints), name="Thresh"),
```

```
# setting up the Age and Sex regressions
        mxMatrix( type="Full", nrow=nvar, ncol=1, free=TRUE, values=0.001, name="bSex" ),
        mxMatrix( type="Full", nrow=nvar, ncol=1, free=TRUE, values=0.001, name="bAge" ),
        mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels="data.Age1", name="Age"),
        mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels="data.Sex1", name="Sex"),
        mxAlgebra( expression= t(bAge%*%Age + bSex%*%Sex), name="AgeSexR"),
        mxMatrix( type="Unit", nrow=maxthresh, ncol=1, free=FALSE, name="Duplicate" ),

        mxAlgebra( expression= (Thresh + (Duplicate%x%(AgeSexR))), name="expThresh",
                dimnames=list(NA,selVars) ),
        mxFIMLObjective( covariance="expCov", means="expMean", dimnames=selVars, thresholds="expThresh")
```

$$\mathrm{AgeSexR} = [\mathrm{Age}\beta] * [\mathrm{Age}] + [\mathrm{Sex}\beta] * [\mathrm{Sex}]$$

$$\mathrm{Duplicate} = \begin{bmatrix} 1 \\ \\ 1 \end{bmatrix} \qquad (\text{unit maxthresh},1)$$

$$\mathrm{expThresh} = ([\mathrm{L}] * [\mathrm{ucThresh}]) + ([\mathrm{Duplicate}] \otimes [\mathrm{AgeSexR}])$$

Deliberate Error...

- CODE RED
  - Do NOT pass go do NOT collect $200
- Change start values in script
  - Use values in the code red solution as a guide
    - mxMatrix(type="Full", nrow=maxthresh, ncol=nvar, free=TRUE, values=c(0,-.7,.6,-.5), lbound=-4,name="cutpoints"),

Other useful things...

Debugging

- check.r
  - Can check algebra etc without including data

```
algebraExercises <- mxModel("meh",
        mxMatrix( type="Full", nrow=4, ncol=2, free=FALSE, values=3, name="Betas" ),
        mxMatrix( type="Unit", nrow=1, ncol=1, free=FALSE, name="Duplicate" ),
        mxMatrix( type="Full", nrow=2, ncol=1, free=FALSE, values=c(1,2),
                labels=c('Age','sex'), name="Covariates" ),

        mxMatrix( type="Lower", nrow=1, ncol=1, free=FALSE, values=1, name="Increment"
        mxMatrix( type="Full",  nrow=1, ncol=4, free=FALSE, values=c(1,2,3,4),
                labels=c('t1','t2','t3','t4'), lbound=-4,name="cutpoints"),

        mxAlgebra( expression= (Increment %*% cutpoints), name="check"),


)
answers <- mxRun(algebraExercises)
answers@algebras
result <- mxEval(list(check, try1, try2),answers)
```

Ploting etc using mxEval

```
mxEval(plot(Fact2.Age, Fact2.Sex),OrdinalFactorModelCOV)


#Beter example
matrixA <- mxMatrix("Full", nrow = 1, ncol = 1,
        values = 1, name = "A")
algebraB <- mxAlgebra(A + A, name = "B")

model <- mxModel(matrixA, algebraB)
model <- mxRun(model)
start <- mxEval(-pi * A, model)

mxEval(plot(sin, start, B * pi), model)
##
```

Goodbye & Good Luck