Structural Equation Modeling II

Lecture 11

Multivariate statistics

Psychology 613 – Spring 2022

Featuring content by Allison Tackman

What kinds of models can you test with this method?

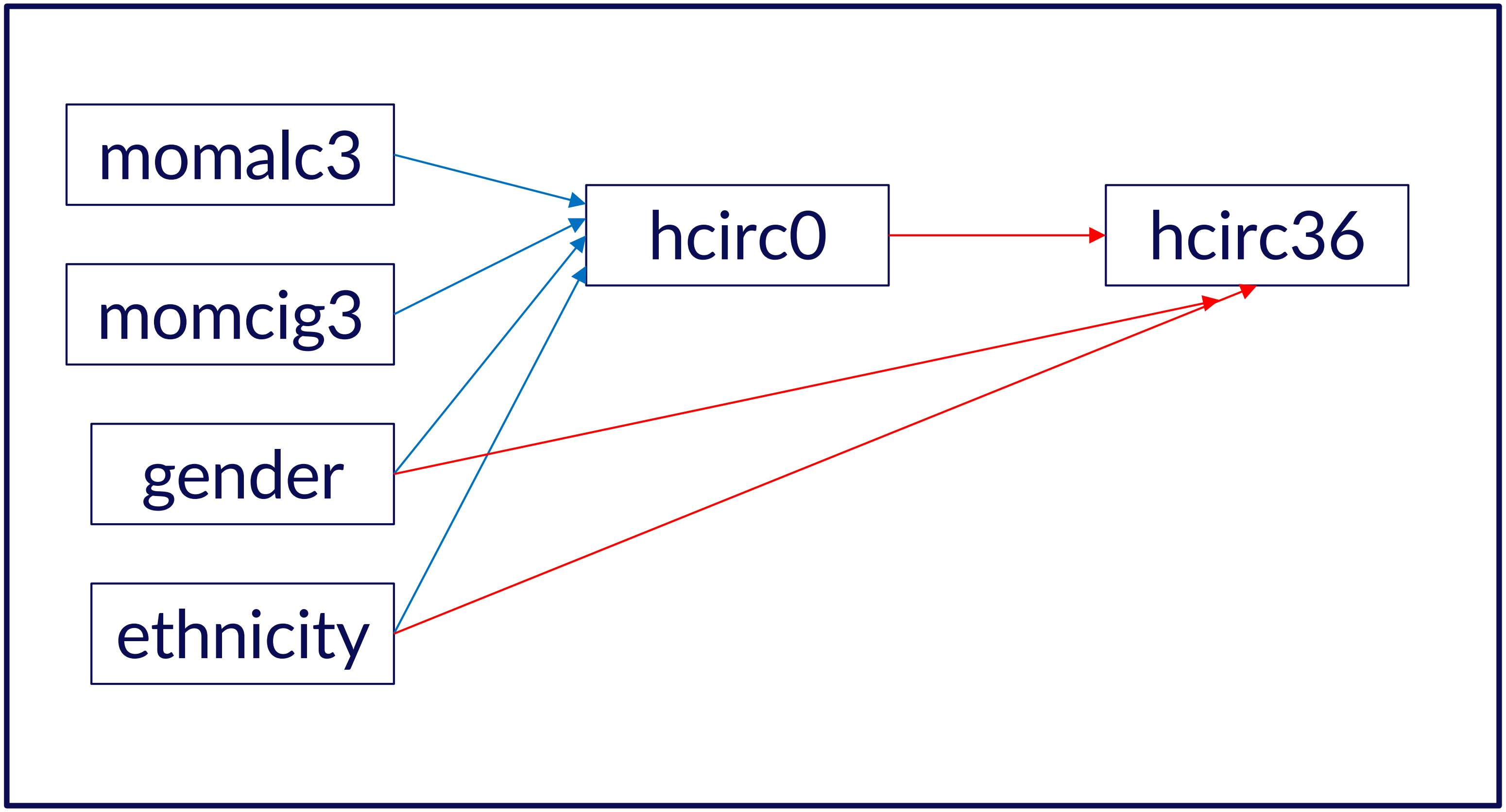
Featuring content by Allison Tackman

What kinds of things can I do with SEM?

- EFA & CFA + measurement invariance
- Path analysis (moderation & mediation)
- Longitudinal data analyses using latent factors
- And so much more…

What kinds of things can I do with SEM?

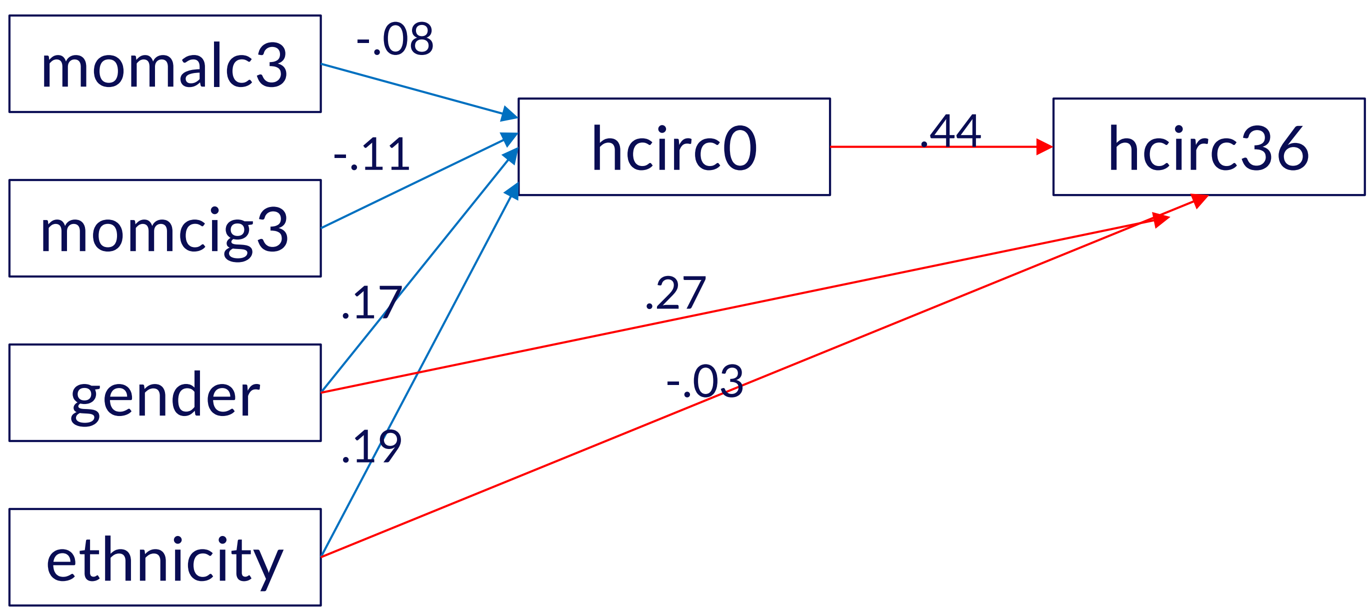
- EFA & CFA + measurement invariance
- Path analysis (moderation & mediation)
- Longitudinal data analyses using latent factors
- And so much more…



Model:
hcirc36 ON hcirc0 gender ethnicity;
hcirc0 ON momalc3 momcig3 gender ethnicity;

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
HCIRC36 ON					
HCIRC0	.415	.036	11.382	.415	.439
GENDER	.762	.107	7.146	.762	.270
ETH	-.094	.107	-.879	-.094	-.033
HCIRC0 ON					
MOMALC3	-.500	.239	-2.090	-.500	-.084
MOMCIG3	-.013	.005	-2.604	-.013	-.108
GENDER	.495	.118	4.185	.495	.166
ETH	.578	.125	4.625	.578	.194



What kinds of things can I do with SEM?

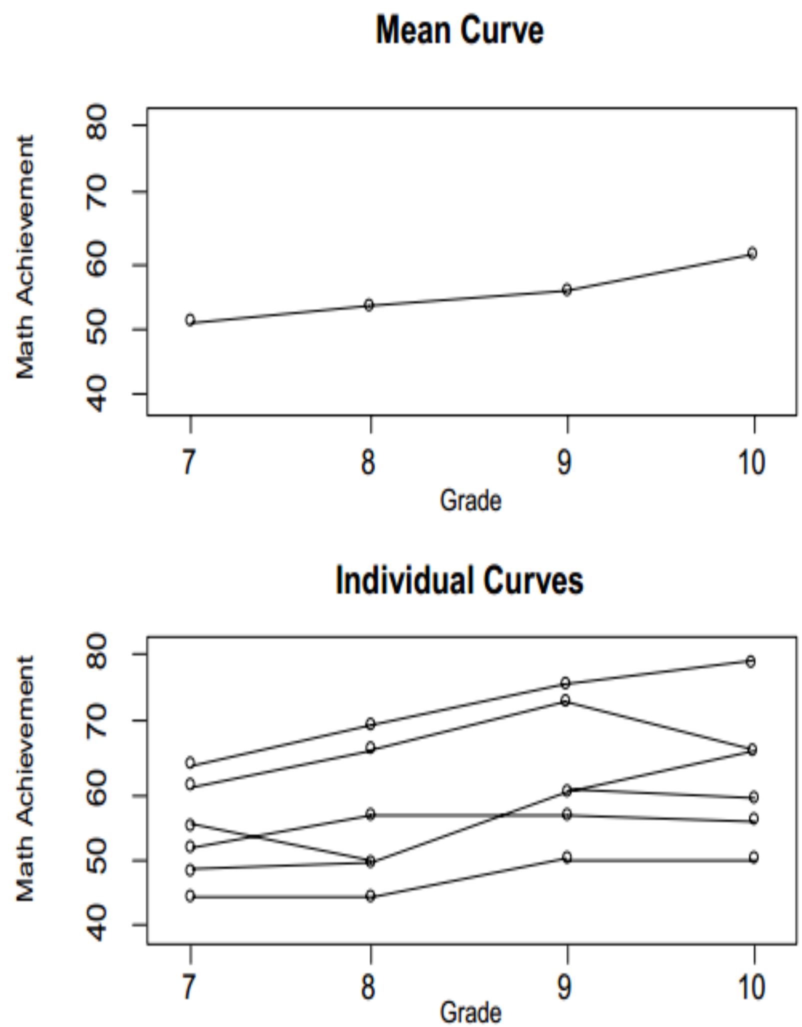
- EFA & CFA + measurement invariance

- Path analysis (moderation & mediation)

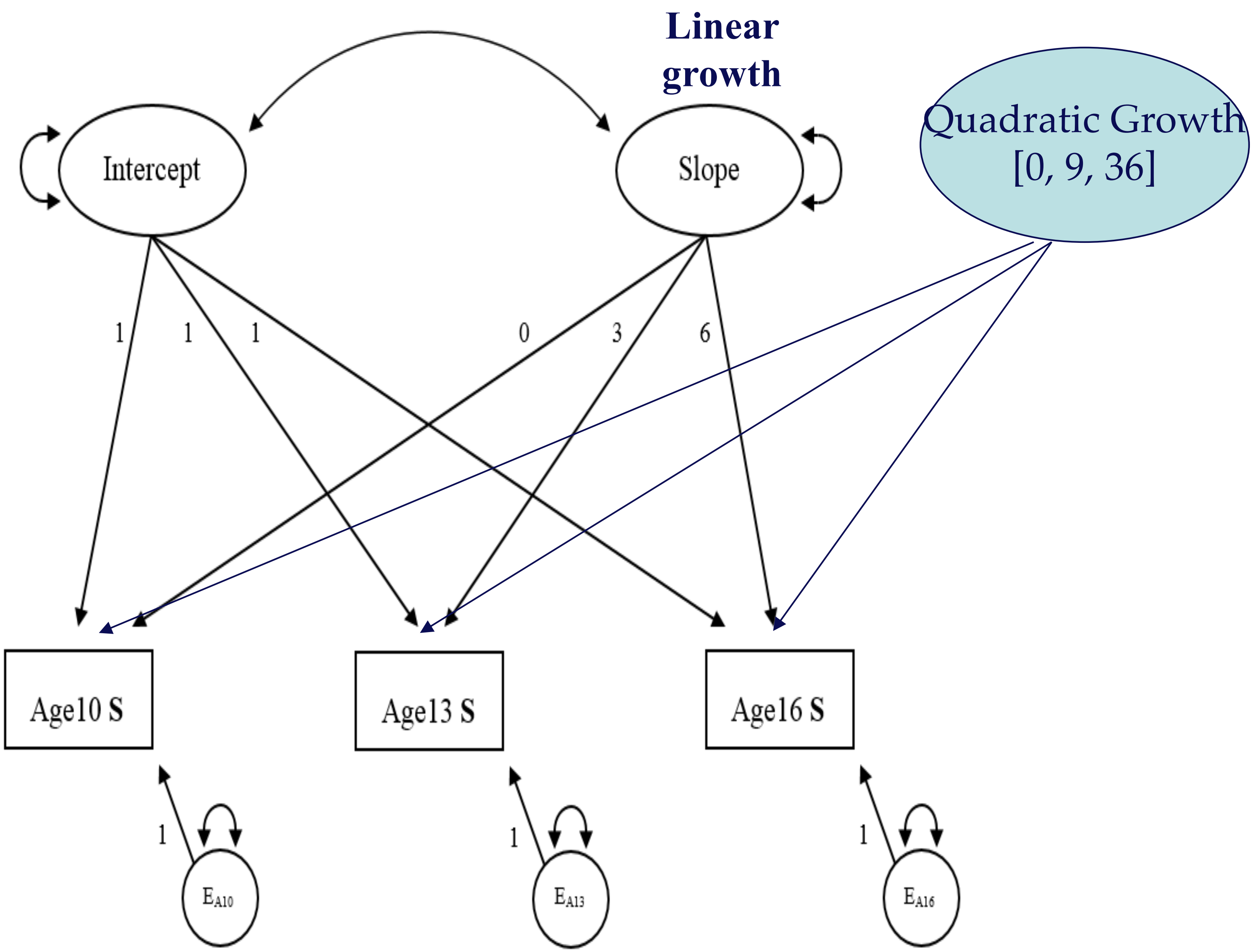
- Longitudinal data analyses using latent factors

- And so much more…

Longitudinal data analyses:
Univariate latent growth curve modeling



Mean Curve
Math Achievement
40
50
60
70
80
7
8
9
10
Grade
Individual Curves
Math Achievement
40
50
60
70
80
7
8
9
10
Grade


Linear growth
Quadratic Growth [0, 9, 36]
Intercept
Slope
1
1
1
0
3
6
Age10 S
Age13 S
Age16 S
1
1
1
E_{A10}
E_{A13}
E_{A16}

Level 1: $y_{ti} = \eta_{0i} + \eta_{1i}\mathrm{x}_t + \varepsilon_{ti}$

Level 2: $\eta_{0i} = \alpha_0 + \zeta_{0i}$

$$\eta_{1i} = \alpha_1 + \zeta_{1i}$$

File Edit View Mplus Plot Diagram Window Help

```
Title: Linear Latent Growth Curve Model Sleep;

Data: File is SA_ALL_REL_DATA.csv;

Variable: Names are gender, t1epsiC, t2epsiC, t3epsiC, T1gradef, T2gradef, T3gradef,
T1hwf, T2hwf, T3hwf, T1SchEng, T2SchEng, T3SchEng, t1carsch, t2carsch, t3carsch,
t1parsch, t2parsch, t3parsch, t1sleep, t2sleep, t3sleep, T1PhysA, T2PhysA, T3PhysA,
t1cesd, t2cesd, t3cesd, t1prsup, t2prsup, t3prsup, t1prdel, t2prdel, t3prdel,
t1pm, t2pm, t3pm, t1dacc, t2dacc, t3dacc, t1macc, t2macc, t3macc;

Missing are all (-99999);

Usevariables are t1sleep t2sleep t3sleep;

Analysis: Type=general;

Model: i s | t1sleep@0 t2sleep@3 t3sleep@6;

Output: sampstat; stdyx;
```

MODEL RESULTS

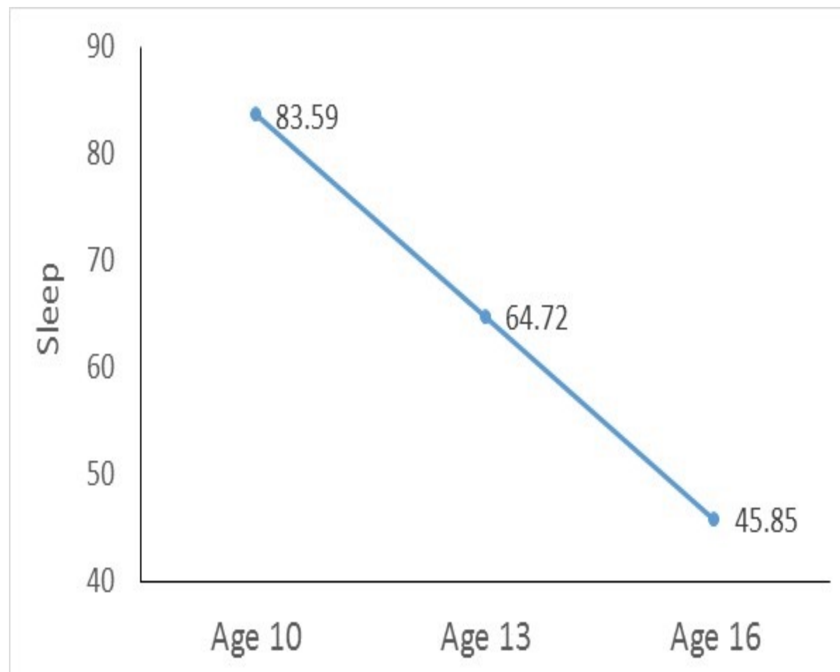
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
I |				
T1SLEEP	1.000	0.000	999.000	999.000
T2SLEEP	1.000	0.000	999.000	999.000
T3SLEEP	1.000	0.000	999.000	999.000
S |				
T1SLEEP	0.000	0.000	999.000	999.000
T2SLEEP	3.000	0.000	999.000	999.000
T3SLEEP	6.000	0.000	999.000	999.000
S WITH				
I	-22.567	12.864	-1.754	0.079
Means				
I	83.591	1.370	61.010	0.000
S	-6.293	0.493	-12.758	0.000
Intercepts				
T1SLEEP	0.000	0.000	999.000	999.000
T2SLEEP	0.000	0.000	999.000	999.000
T3SLEEP	0.000	0.000	999.000	999.000
Variances				
I	142.292	57.624	2.469	0.014
S	10.035	3.926	2.556	0.011
Residual Variances				
T1SLEEP	19.593	53.717	0.365	0.715
T2SLEEP	158.406	38.025	4.166	0.000
T3SLEEP	82.021	74.761	1.097	0.273

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 I        |
    T1SLEEP            1.000      0.000    999.000    999.000
    T2SLEEP            1.000      0.000    999.000    999.000
    T3SLEEP            1.000      0.000    999.000    999.000

 S        |
    T1SLEEP            0.000      0.000    999.000    999.000
    T2SLEEP            3.000      0.000    999.000    999.000
    T3SLEEP            6.000      0.000    999.000    999.000

 S        WITH
    I                -22.567     12.864     -1.754      0.079

 Means
    I                 83.591      1.370     61.010      0.000
    S                 -6.293      0.493    -12.758      0.000

 Intercepts
    T1SLEEP            0.000      0.000    999.000    999.000
    T2SLEEP            0.000      0.000    999.000    999.000
    T3SLEEP            0.000      0.000    999.000    999.000

 Variances
    I                142.292     57.624      2.469      0.014
    S                 10.035      3.926      2.556      0.011

 Residual Variances
    T1SLEEP           19.593     53.717      0.365      0.715
    T2SLEEP          158.406     38.025      4.166      0.000
    T3SLEEP           82.021     74.761      1.097      0.273
```

Pred Value for Age 10 (where Int = 1, and Slo = 0):
83.59(1) + (-6.29)(0) = 83.59

Pred Value for Age 10 (where Int = 1, and Slo = 3):
83.59(1) + (-6.29)(3) = 64.72

Pred Value for Age 10 (where Int = 1, and Slo = 6):
83.59(1) + (-6.29)(6) = 45.85



Longitudinal data analyses:
Univariate latent growth curve modeling with time-invariant covariates

File Edit View Mplus Plot Diagram Window Help

```
Title: Linear Latent Growth Curve Model Sleep;

Data: File is SA_ALL_REL_DATA.csv;

Variable: Names are gender, t1epsiC, t2epsiC, t3epsiC, T1gradef, T2gradef, T3gradef,
T1hwf, T2hwf, T3hwf, T1SchEng, T2SchEng, T3SchEng, t1carsch, t2carsch, t3carsch,
t1parsch, t2parsch, t3parsch, t1sleep, t2sleep, t3sleep, T1PhysA, T2PhysA, T3PhysA,
t1cesd, t2cesd, t3cesd, t1prsup, t2prsup, t3prsup, t1prdel, t2prdel, t3prdel,
t1pm, t2pm, t3pm, t1dacc, t2dacc, t3dacc, t1macc, t2macc, t3macc;

Missing are all (-99999);

Usevariables are t1sleep t2sleep t3sleep gencc;

Define:
if (gender==1) THEN gencc = 0;
if (gender==2) THEN gencc = 1;
! Boys are 0
! Girls are 1

Analysis: Type=general;

Model: i s | t1sleep@0 t2sleep@3 t3sleep@6;
       i s ON gencc;

Output: sampstat; stdyx;
```

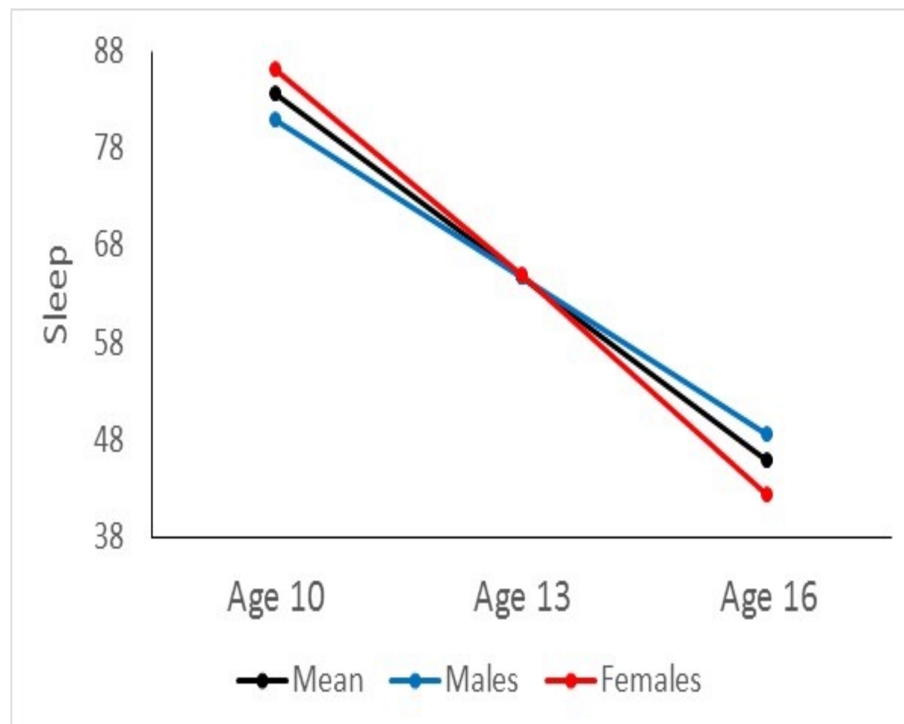
```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 I        |
    T1SLEEP            1.000      0.000    999.000    999.000
    T2SLEEP            1.000      0.000    999.000    999.000
    T3SLEEP            1.000      0.000    999.000    999.000

 S        |
    T1SLEEP            0.000      0.000    999.000    999.000
    T2SLEEP            3.000      0.000    999.000    999.000
    T3SLEEP            6.000      0.000    999.000    999.000

 I        ON
    GENCC              5.166      2.642      1.955      0.051

 S        ON
    GENCC             -1.655      0.963     -1.719      0.086

 S        WITH
    I                -23.870     12.746     -1.873      0.061

 Intercepts
    T1SLEEP            0.000      0.000    999.000    999.000
    T2SLEEP            0.000      0.000    999.000    999.000
    T3SLEEP            0.000      0.000    999.000    999.000
    I                 80.919      1.906     42.448      0.000
    S                 -5.377      0.737     -7.297      0.000

 Residual Variances
    T1SLEEP            3.183     53.594      0.059      0.953
    T2SLEEP          159.766     38.415      4.159      0.000
    T3SLEEP           94.112     76.438      1.231      0.218
    I                151.985     58.174      2.613      0.009
    S                  9.829      3.757      2.616      0.009
```

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 I        |
    T1SLEEP            1.000      0.000    999.000    999.000
    T2SLEEP            1.000      0.000    999.000    999.000
    T3SLEEP            1.000      0.000    999.000    999.000

 S        |
    T1SLEEP            0.000      0.000    999.000    999.000
    T2SLEEP            3.000      0.000    999.000    999.000
    T3SLEEP            6.000      0.000    999.000    999.000

 I        ON
    GENCC              5.166      2.642      1.955      0.051

 S        ON
    GENCC             -1.655      0.963     -1.719      0.086

 S        WITH
    I                -23.870     12.746     -1.873      0.061

 Intercepts
    T1SLEEP            0.000      0.000    999.000    999.000
    T2SLEEP            0.000      0.000    999.000    999.000
    T3SLEEP            0.000      0.000    999.000    999.000
    I                 80.919      1.906     42.448      0.000
    S                 -5.377      0.737     -7.297      0.000

 Residual Variances
    T1SLEEP            3.183     53.594      0.059      0.953
    T2SLEEP          159.766     38.415      4.159      0.000
    T3SLEEP           94.112     76.438      1.231      0.218
    I                151.985     58.174      2.613      0.009
    S                  9.829      3.757      2.616      0.009
```

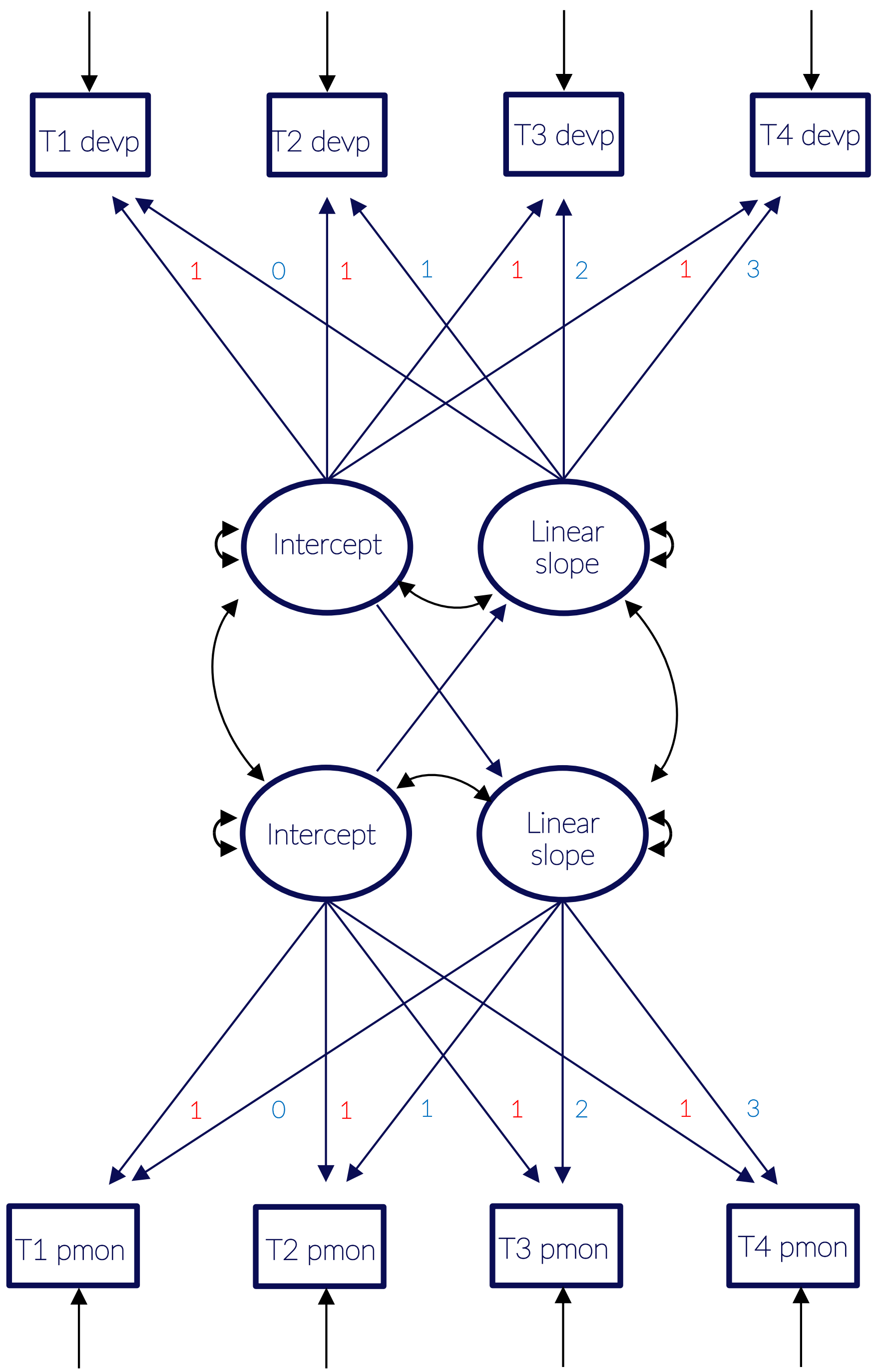
Pred Value for Age 10 (where Int = 1, and Slo = 0):
Males: 80.92(1) + (-5.38)(0) = 80.92
Females: (80.92 + 5.17)(1) + (-5.38 – 1.66)(0) = 86.09

Pred Value for Age 10 (where Int = 1, and Slo = 3):
Males: 80.92(1) + (-5.38)(3) = 64.78
Females: (80.92 + 5.17)(1) + (-5.38 – 1.66)(3) = 64.97

Pred Value for Age 10 (where Int = 1, and Slo = 6):
Males: 80.92(1) + (-5.38)(6) = 48.64
Females: (80.92 + 5.17)(1) + (-5.38 – 1.66)(6) = 42.24



Longitudinal data analyses:
Bivariate latent growth curve modeling


T1 devp
T2 devp
T3 devp
T4 devp
1
0
1
1
1
2
1
3
Intercept
Linear
slope
Intercept
Linear
slope
1
0
1
1
1
2
1
3
T1 pmon
T2 pmon
T3 pmon
T4 pmon

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SMON ON				
IDEVP	0.259	0.080	3.219	0.001
SDEVP ON				
IMON	0.124	0.057	2.172	0.030
IMON WITH				
SMON	-0.010	0.047	-0.220	0.826
IDEVP	-0.426	0.078	-5.432	0.000
IDEVP WITH				
SDEVP	-0.063	0.062	-1.020	0.308
SMON WITH				
SDEVP	-0.020	0.016	-1.270	0.204

Questions of interest:

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SMON ON				
IDEVP	0.259	0.080	3.219	0.001
SDEVP ON				
IMON	0.124	0.057	2.172	0.030
IMON WITH				
SMON	-0.010	0.047	-0.220	0.826
IDEVP	-0.426	0.078	-5.432	0.000
IDEVP WITH				
SDEVP	-0.063	0.062	-1.020	0.308
SMON WITH				
SDEVP	-0.020	0.016	-1.270	0.204

Questions of interest:

1. Do the slopes of the two constructs correlate?

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SMON ON				
IDEVP	0.259	0.080	3.219	0.001
SDEVP ON				
IMON	0.124	0.057	2.172	0.030
IMON WITH				
SMON	-0.010	0.047	-0.220	0.826
IDEVP	-0.426	0.078	-5.432	0.000
IDEVP WITH				
SDEVP	-0.063	0.062	-1.020	0.308
SMON WITH				
SDEVP	-0.020	0.016	-1.270	0.204

Questions of interest:

1. Do the slopes of the two constructs correlate?
2. Do the intercepts of the two constructs correlate?

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SMON ON				
IDEVP	0.259	0.080	3.219	0.001
SDEVP ON				
IMON	0.124	0.057	2.172	0.030
IMON WITH				
SMON	-0.010	0.047	-0.220	0.826
IDEVP	-0.426	0.078	-5.432	0.000
IDEVP WITH				
SDEVP	-0.063	0.062	-1.020	0.308
SMON WITH				
SDEVP	-0.020	0.016	-1.270	0.204

Questions of interest:

1. Do the slopes of the two constructs correlate?
2. Do the intercepts of the two constructs correlate?
3. Does parental monitoring at T1 predict change in deviant peer affiliation over time?

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SMON	ON				
	IDEVP	0.259	0.080	3.219	0.001
SDEVP	ON				
	IMON	0.124	0.057	2.172	0.030
IMON	WITH				
	SMON	-0.010	0.047	-0.220	0.826
	IDEVP	-0.426	0.078	-5.432	0.000
IDEVP	WITH				
	SDEVP	-0.063	0.062	-1.020	0.308
SMON	WITH				
	SDEVP	-0.020	0.016	-1.270	0.204

Questions of interest:

1. Do the slopes of the two constructs correlate?
2. Do the intercepts of the two constructs correlate?
3. Does parental monitoring at T1 predict change in deviant peer affiliation over time?
4. Does deviant peer affiliation at T1 predict change in parental monitoring over time?

Longitudinal data analyses: *Latent class growth analysis / growth mixture modeling*

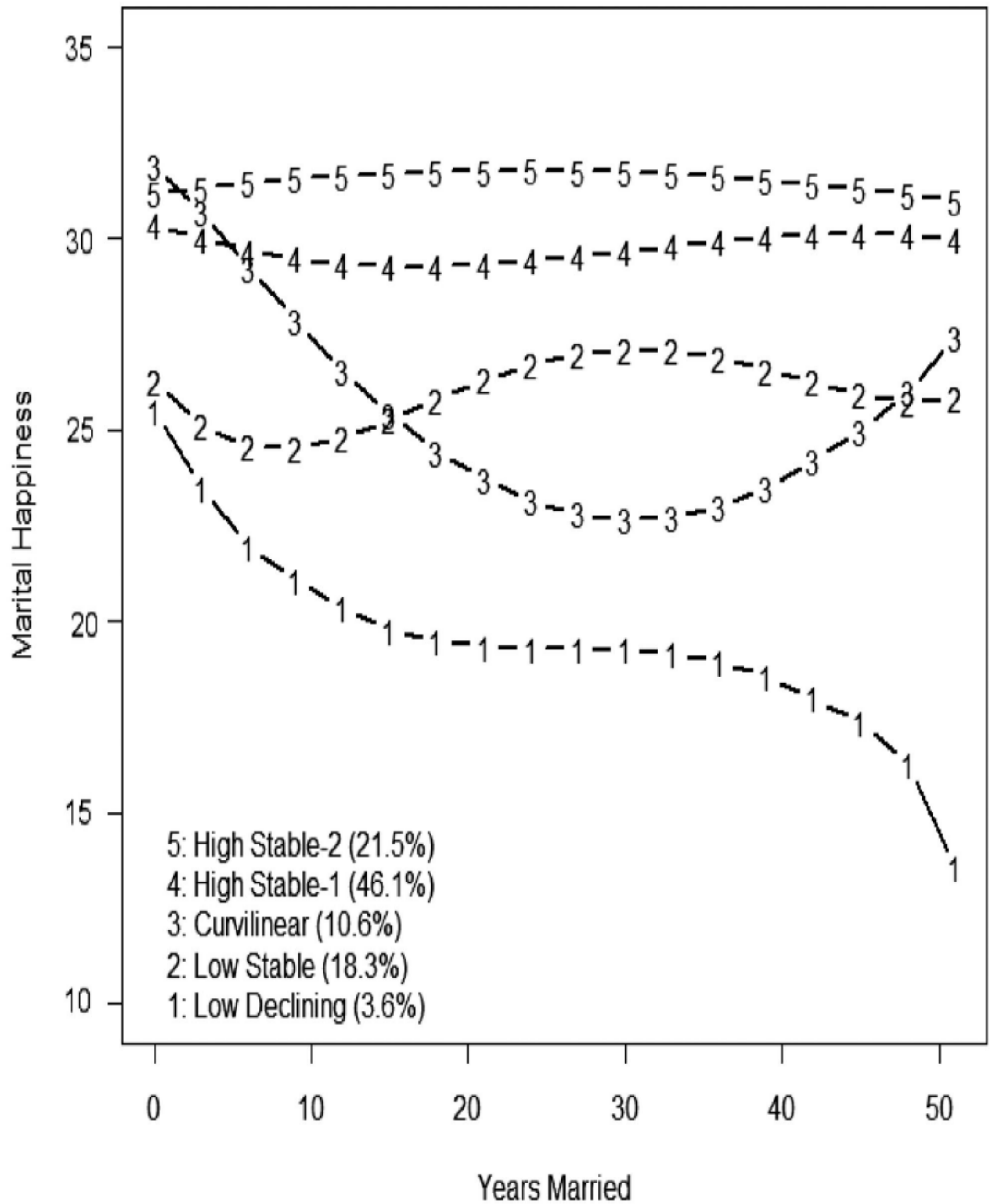
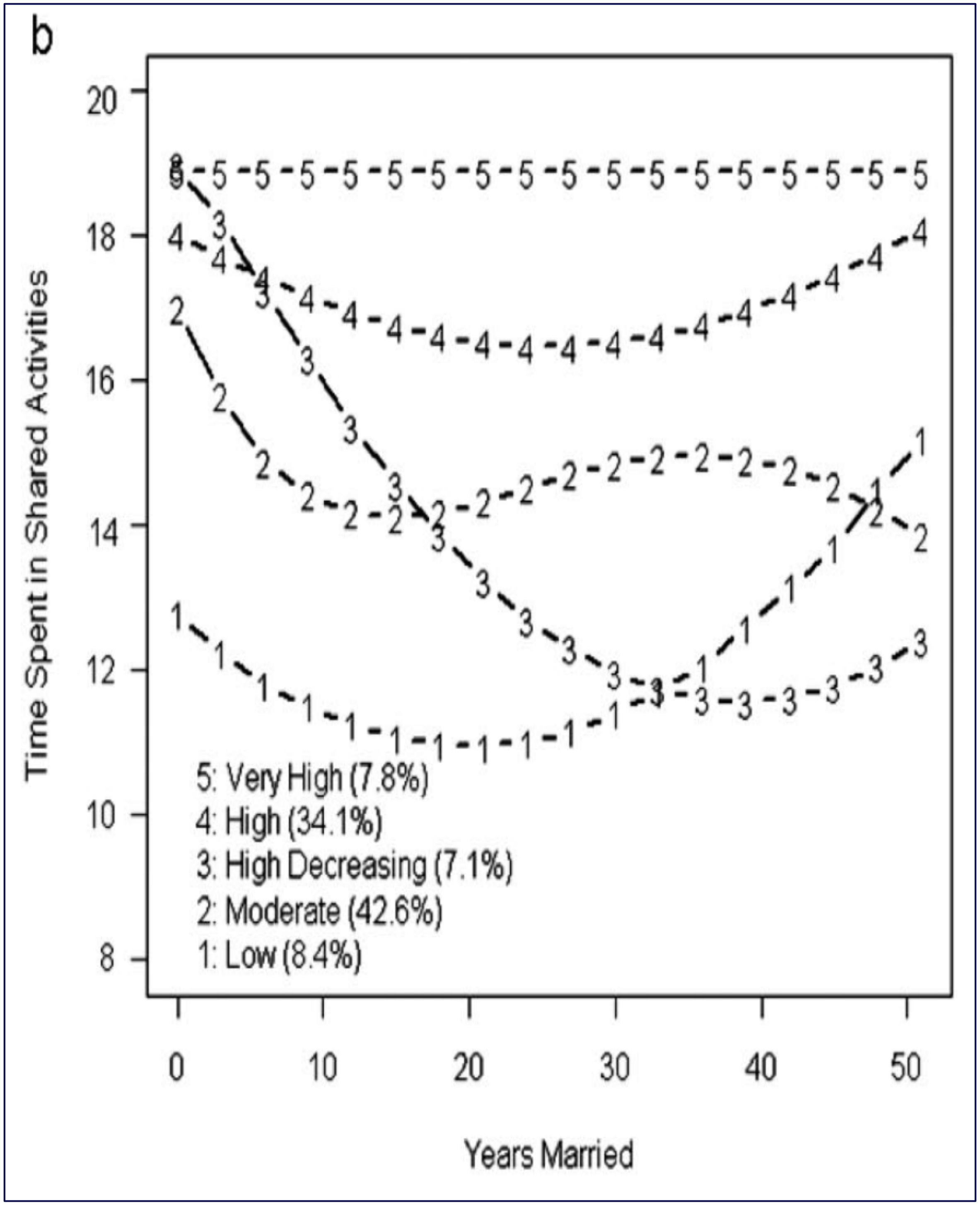
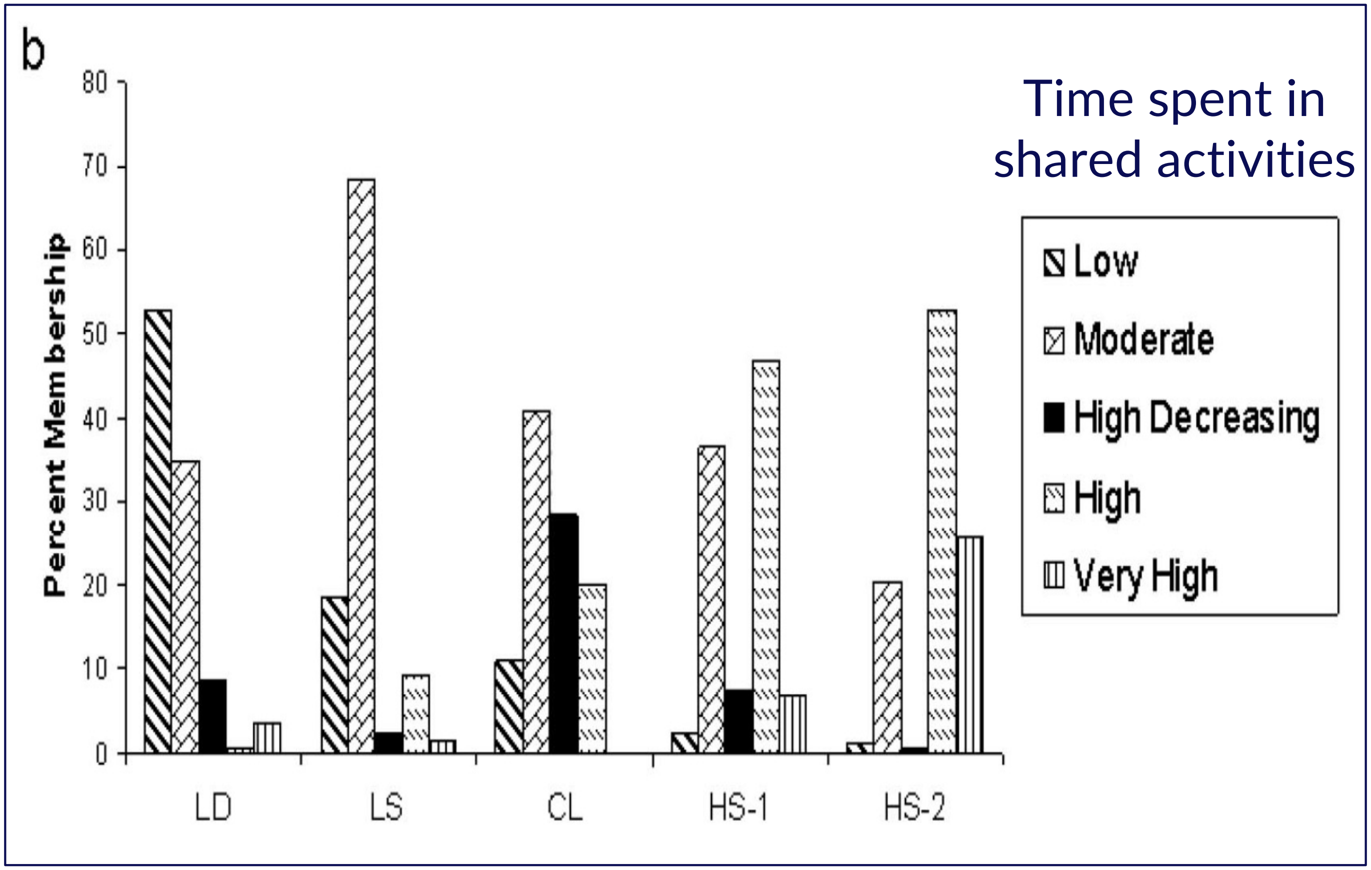
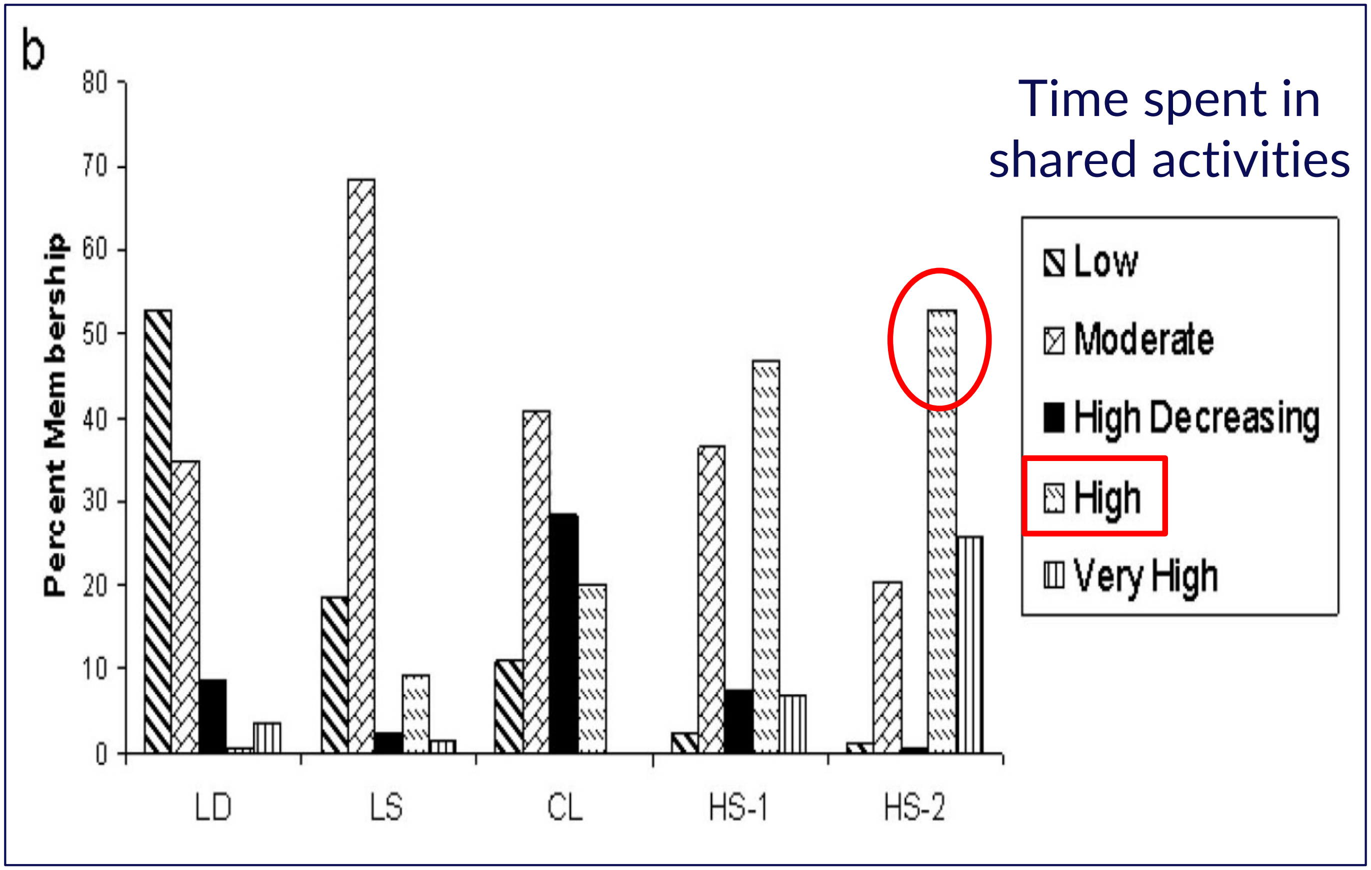


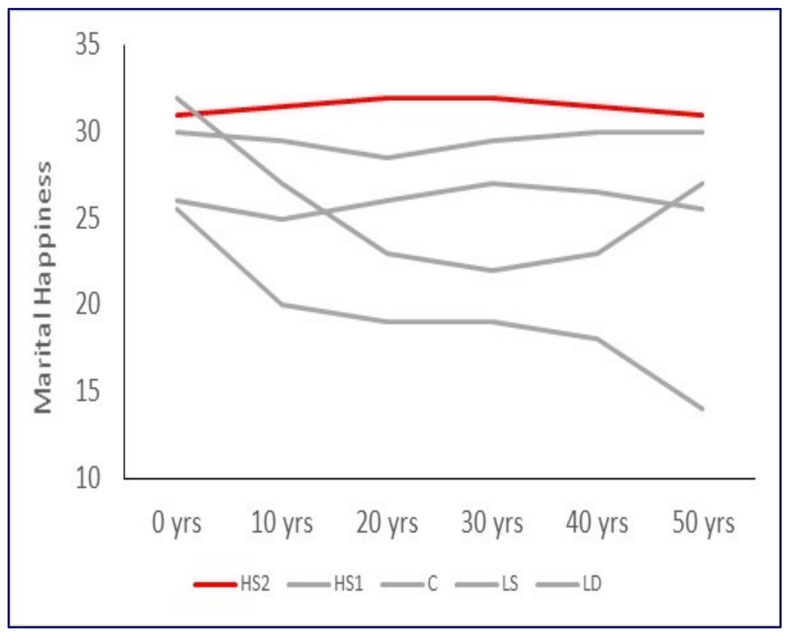
Figure 1. Trajectories of marital happiness.

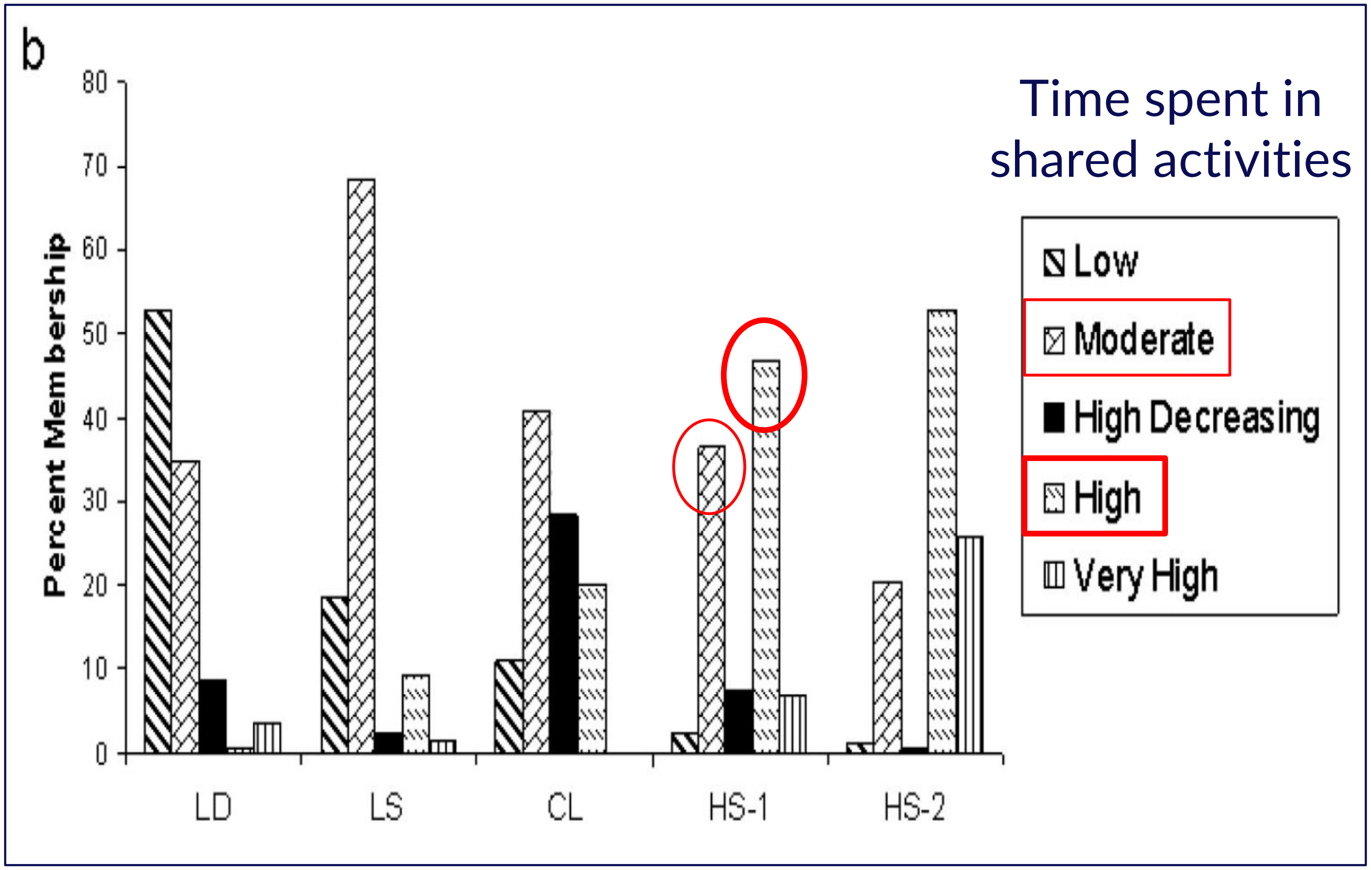

b
Time Spent in Shared Activities
20
18
16
14
12
10
8
0
10
20
30
40
50
Years Married
5: Very High (7.8%)
4: High (34.1%)
3: High Decreasing (7.1%)
2: Moderate (42.6%)
1: Low (8.4%)

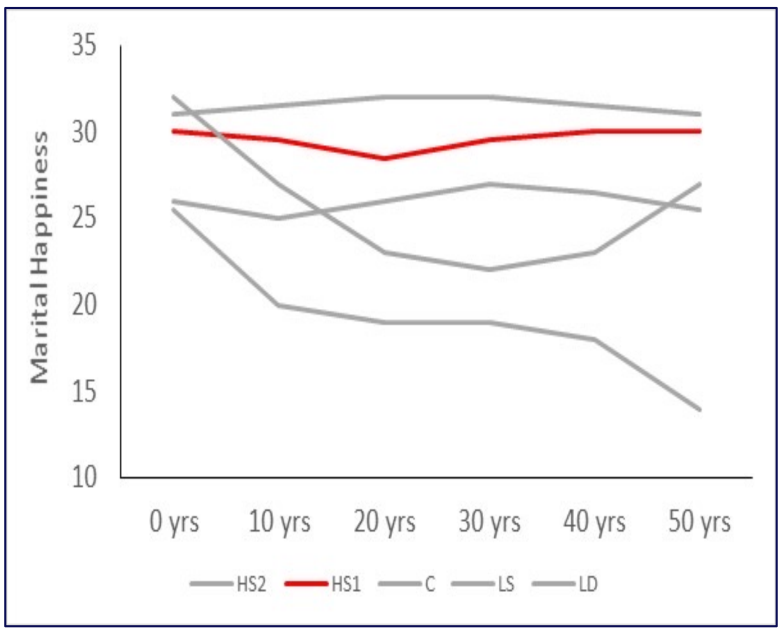


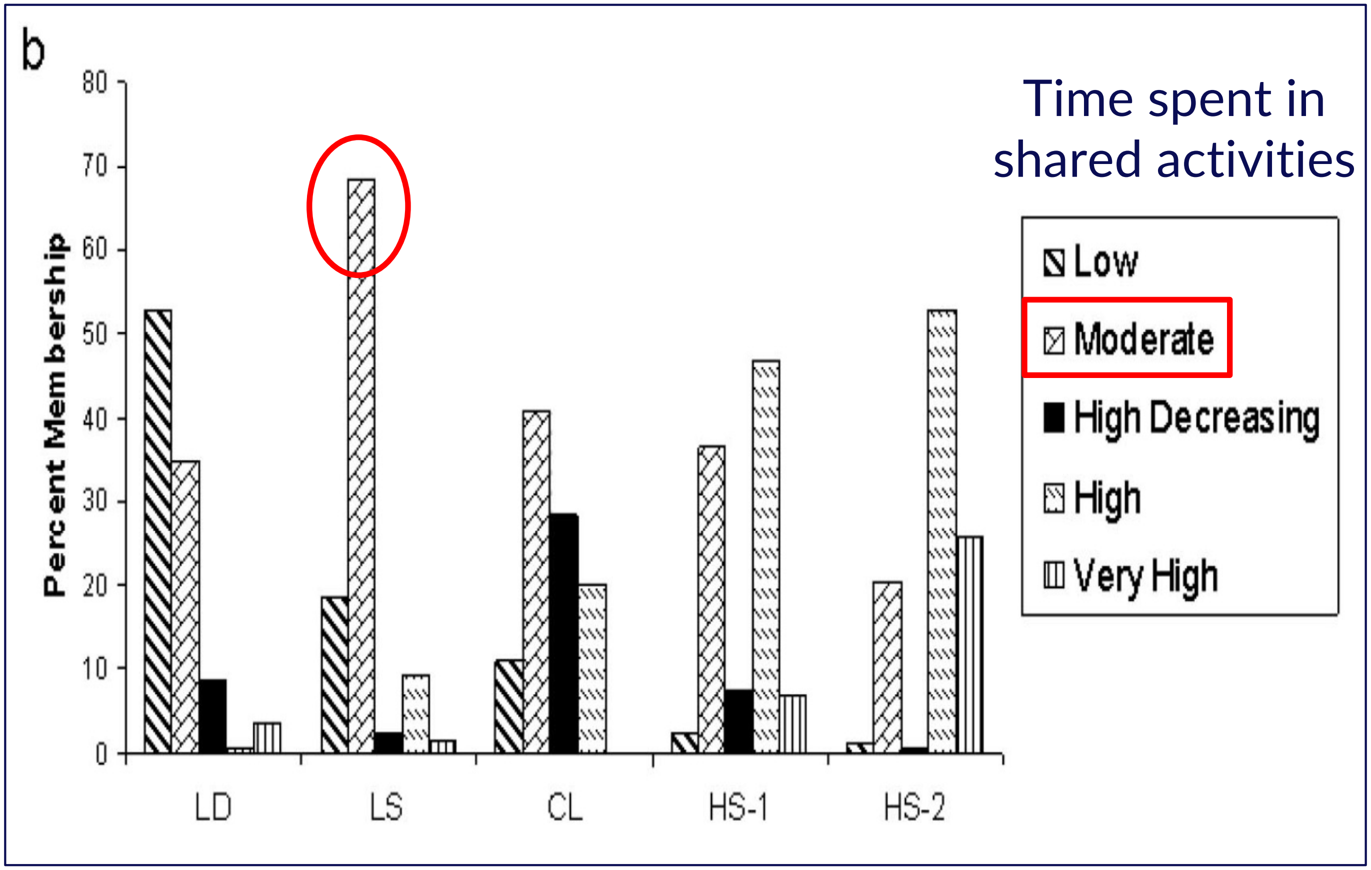
b
Time spent in shared activities
Percent Membership
80
70
60
50
40
30
20
10
0
LD
LS
CL
HS-1
HS-2
Low
Moderate
High Decreasing
High
Very High

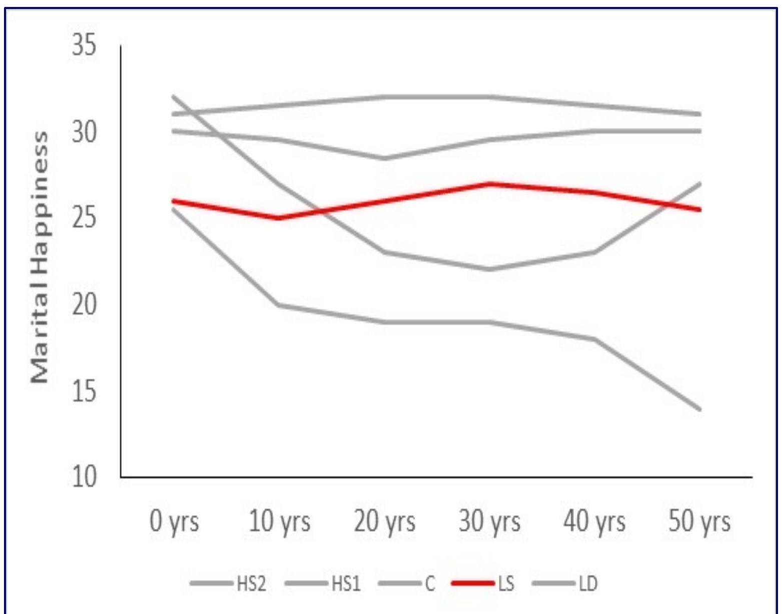

b
Time spent in
shared activities
Percent Membership
0
10
20
30
40
50
60
70
80
LD
LS
CL
HS-1
HS-2
Low
Moderate
High Decreasing
High
Very High

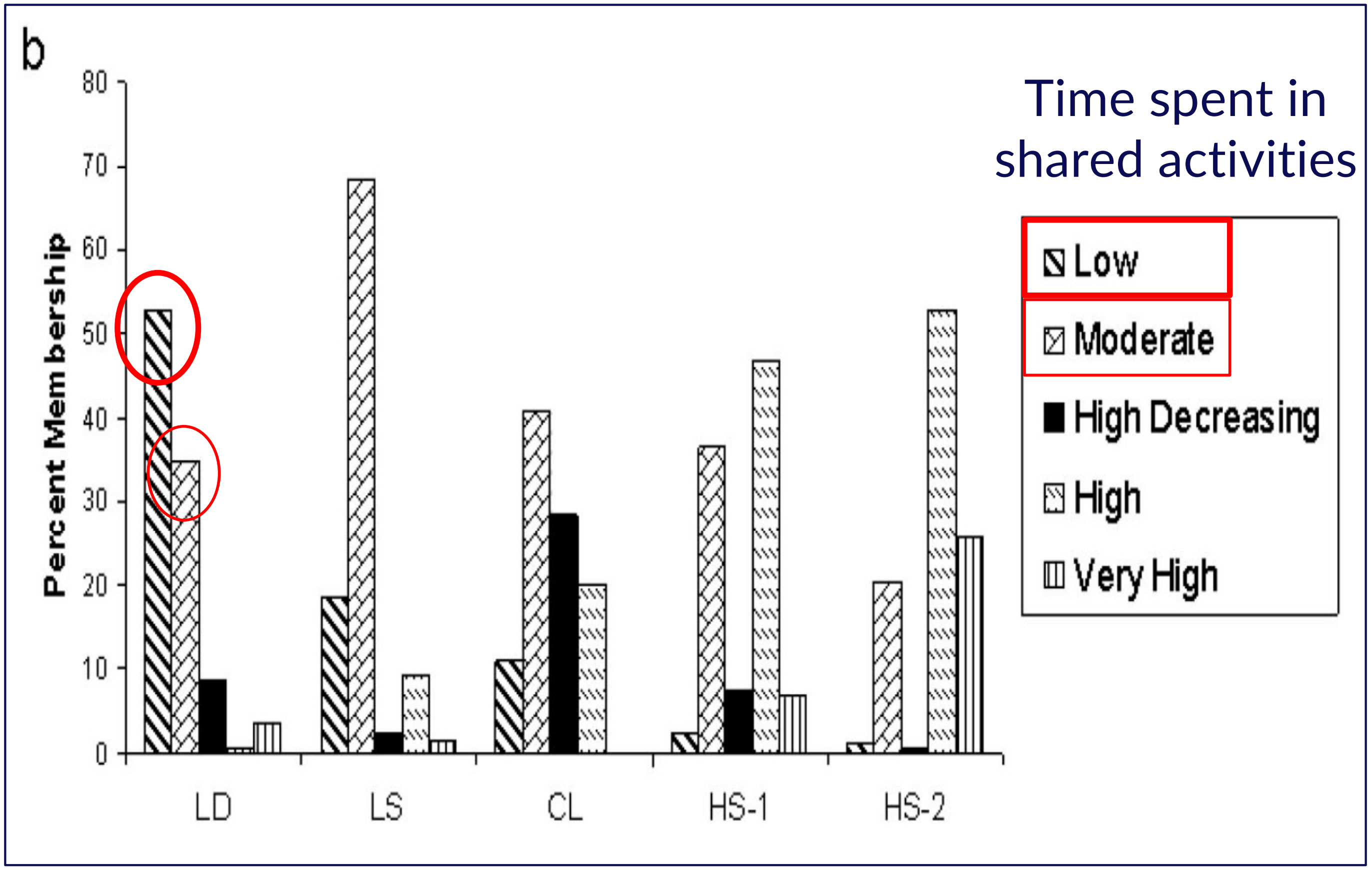

Marital Happiness
35
30
25
20
15
10
0 yrs
10 yrs
20 yrs
30 yrs
40 yrs
50 yrs
HS2
HS1
C
LS
LD

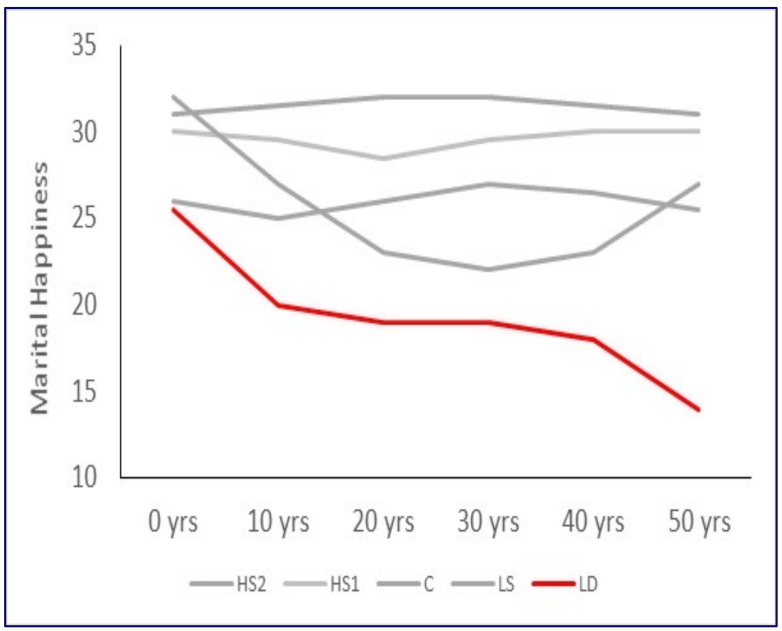

b
Percent Membership
80
70
60
50
40
30
20
10
0
LD
LS
CL
HS-1
HS-2
Time spent in shared activities
Low
Moderate
High Decreasing
High
Very High

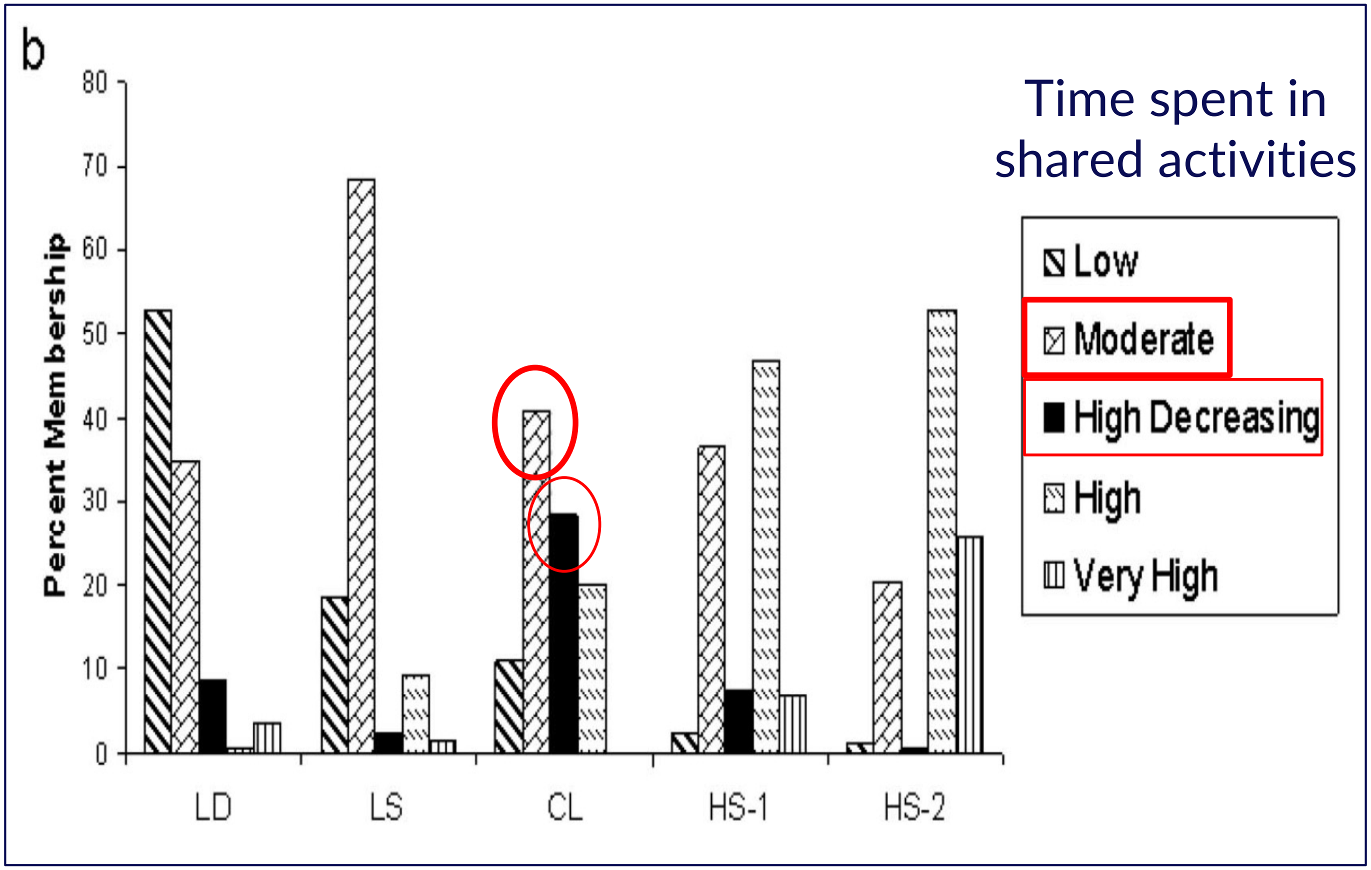

Marital Happiness
35
30
25
20
15
10
0 yrs
10 yrs
20 yrs
30 yrs
40 yrs
50 yrs
HS2
HS1
C
LS
LD

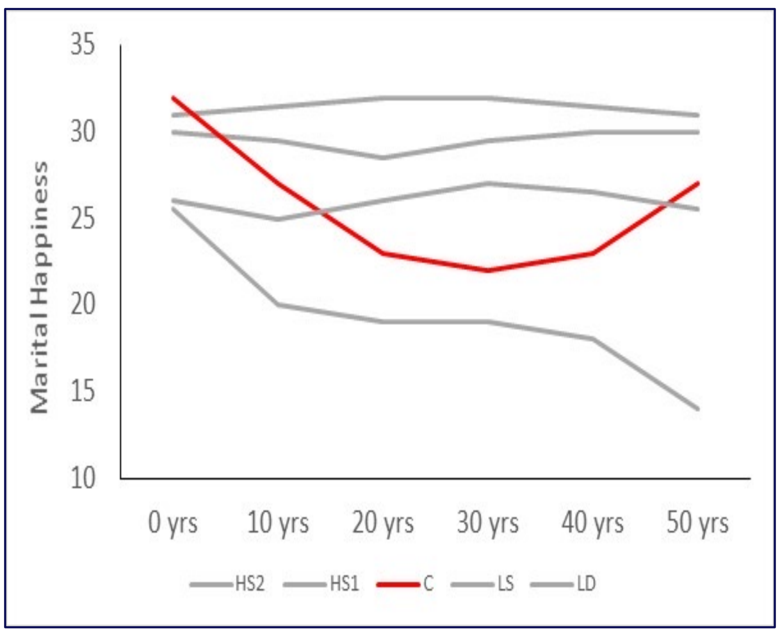

b
Time spent in shared activities
Low
Moderate
High Decreasing
High
Very High
Percent Membership
80
70
60
50
40
30
20
10
0
LD
LS
CL
HS-1
HS-2


Marital Happiness
35
30
25
20
15
10
0 yrs
10 yrs
20 yrs
30 yrs
40 yrs
50 yrs
HS2
HS1
C
LS
LD


b
Time spent in shared activities
Low
Moderate
High Decreasing
High
Very High
Percent Membership
0
10
20
30
40
50
60
70
80
LD
LS
CL
HS-1
HS-2


Marital Happiness
10
15
20
25
30
35
0 yrs
10 yrs
20 yrs
30 yrs
40 yrs
50 yrs
HS2
HS1
C
LS
LD


b
Time spent in shared activities
Percent Membership
80
70
60
50
40
30
20
10
0
LD
LS
CL
HS-1
HS-2
Low
Moderate
High Decreasing
High
Very High


Marital Happiness
35
30
25
20
15
10
0 yrs
10 yrs
20 yrs
30 yrs
40 yrs
50 yrs
HS2
HS1
C
LS
LD

Violations of normality

Lecture 11-b

Violations of normality

Four assumptions of the GLM:

- Independent observations
- Randomly drawn from population
- Error variances are equal across range of IV
  - (i.e., continuous predictor or groups)
- Errors are normally distributed

The Box-Cox transformation

Identifies the exponent that will make a set of residuals closest to normal

Compute using R…

1) Put your data into tab-delimited or SPSS form
2) Import using: read.table or read.spss
3) Parse the data:

Y = data$*C1*   X = data$*C2*

(where *C1* and *C2* are the first and second cols)

The Box-Cox transformation

Compute using R…

4) Then compute the actual transformation:

boxcox(*DV~IV*, lambda=seq(*from*,*to*,*by*), plotit=T)

here,

boxcox(Y~X,lambda=seq(-3,3,.01),plotit=T)

5) Zoom in to figure out the exact value

6) Transform your data (*ynew = y^lambda*)

7) Re-check: boxcox(*ynew*~x)

Box-Cox example

```
> summary(data)
        Length   Class    Mode
subj    23524    -none-   numeric
rt      23524    -none-   numeric
type    23524    factor   numeric

> rt <- data$rt
> type <- data$type
> boxcox(rt~type, lambda = seq(-3,3,.01), plotit=1)
```

Box-Cox example

Box-Cox example

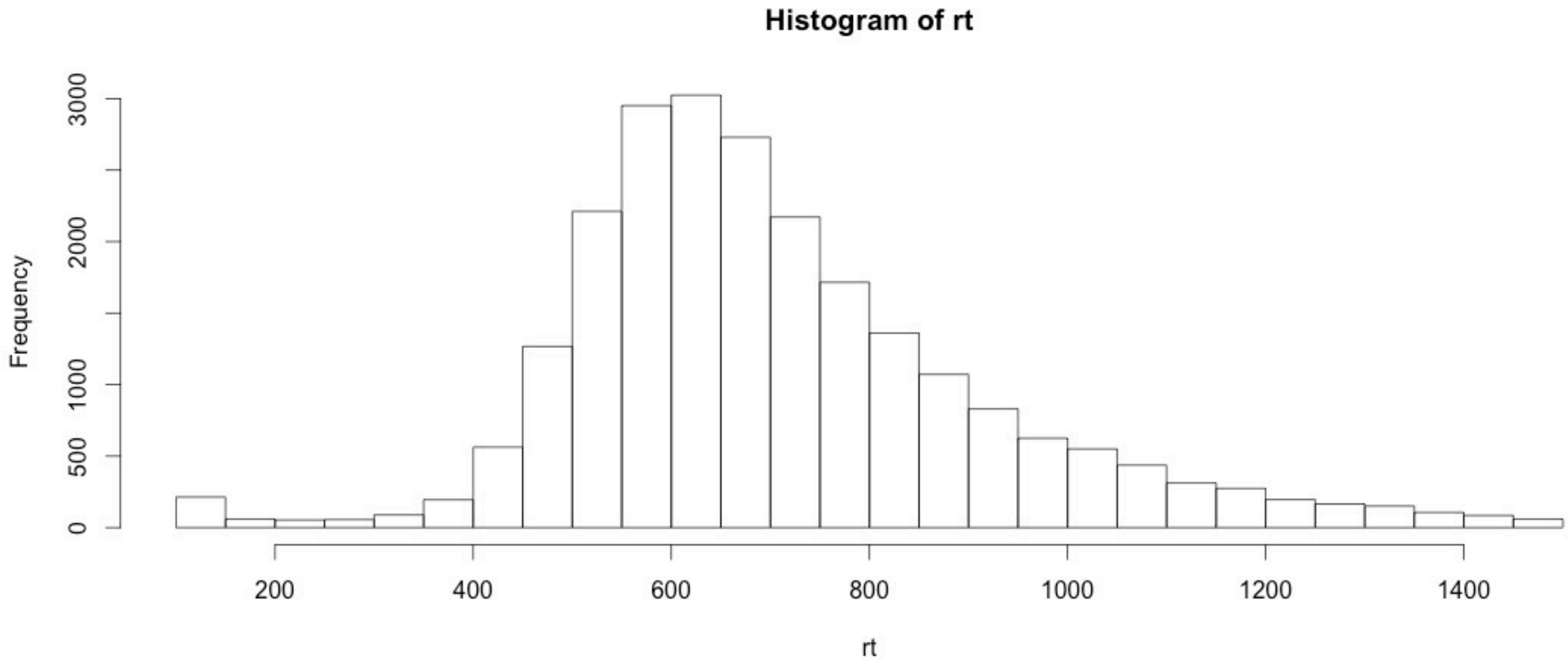
```
> boxcox( rt~type, lambda = seq(0.5,0.75,.01),
                    plotit=1 )
```

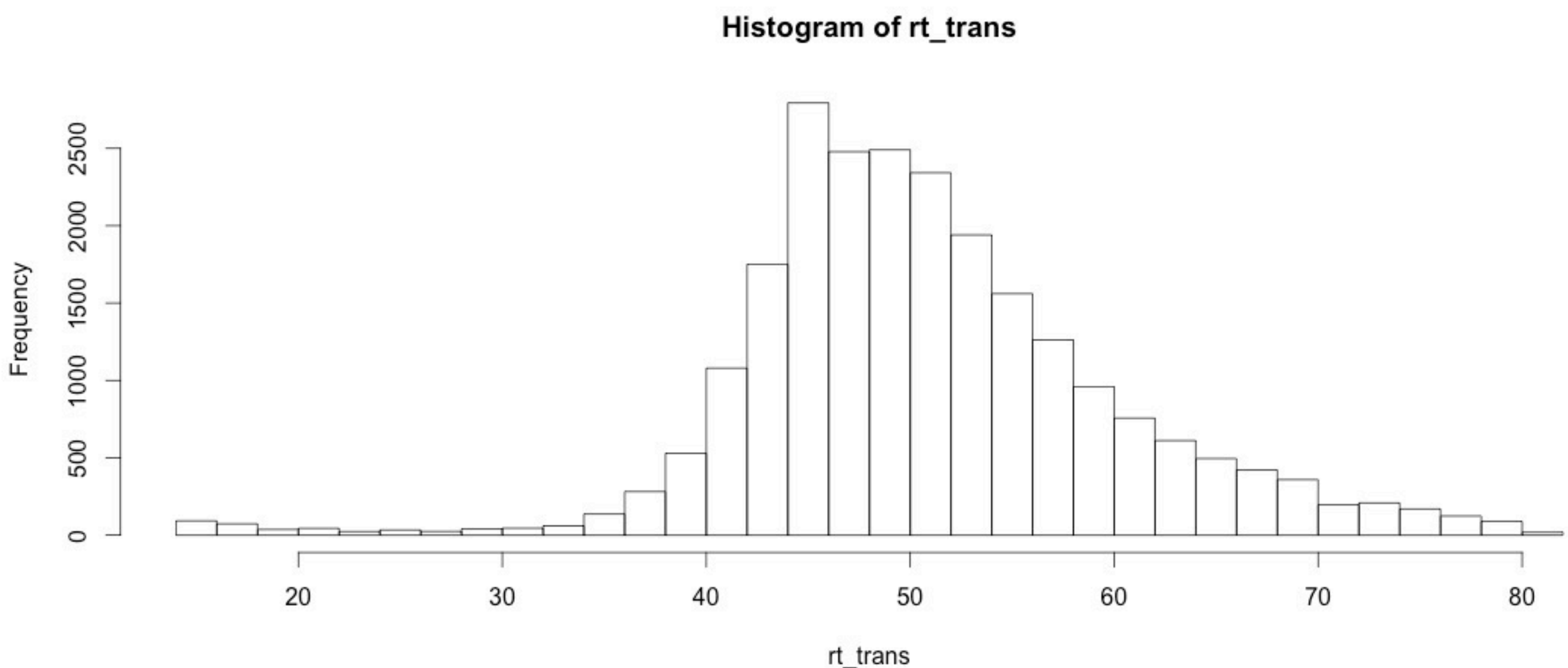
Box-Cox example

Box-Cox example

```
> boxcox( rt~type, lambda =
              seq(0.5,0.75,.01), plotit = 1)
> rt_trans <- rt^0.6
> boxcox( rt_trans~type,
              lambda = seq(-3, 3, .01),
              plotit = 1)
```

Box-Cox example


Histogram of rt
Frequency
0
500
1000
2000
3000
200
400
600
800
1000
1200
1400
rt


Histogram of rt_trans
Frequency
0
500
1000
1500
2000
2500
20
30
40
50
60
70
80
rt_trans