PSY 5939: Longitudinal Data Analysis

Contents

1 Time loadings

1.1 Example data

1.1.1 Substance use example

Substance use: alcohol, cigarettes, marijuana

Ages 15, 16, 17, and 18

Predictors: gender (female), two parent family, others

101 subjects

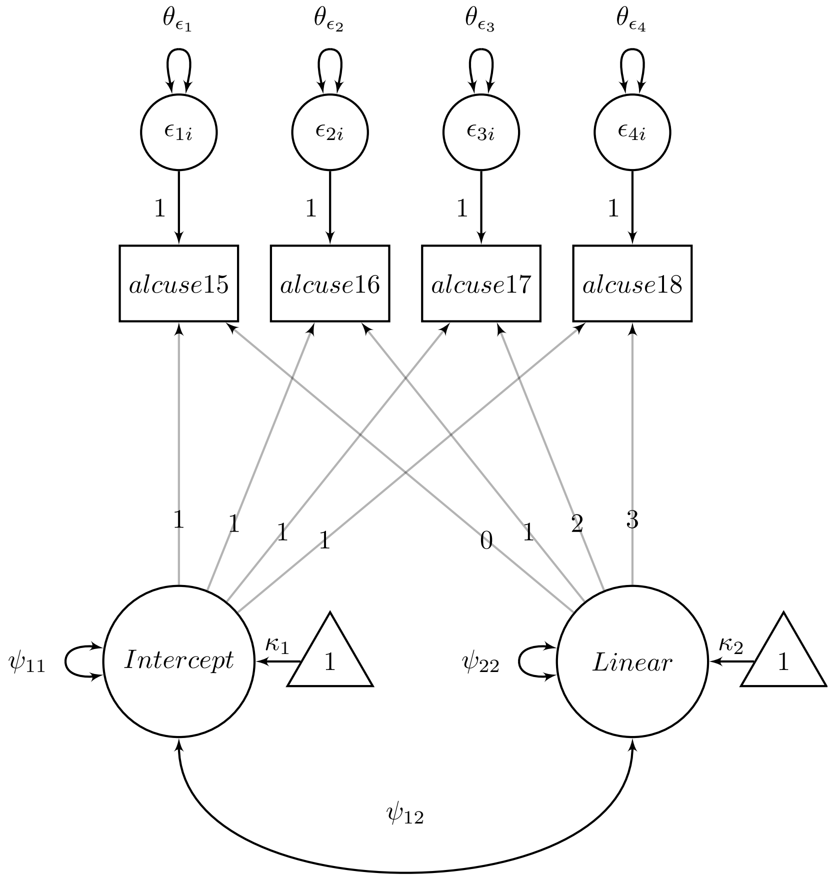
No missing data

Data comes from Singer and Willett book

1.1.2 Substance use data

id	alcuse15	alcuse16	alcuse17	alcuse18	ciguse15	ciguse16	ciguse17	ciguse18	potuse15	potuse16	potuse
1	5	5	5	6	6	6	5	7	5	5	
6	8	8	7	7	9	8	7	6	6	5	
7	5	6	6	6	5	5	5	4	4	4	
19	6	7	8	7	6	6	7	7	6	6	
20	6	7	7	6	7	8	8	9	6	6	
25	7	6	6	6	6	7	7	10	5	5	

1.1.3 Latent growth model



1.2 Time loadings

1.2.1 Time in SEM growth models

Time is defined in SEM growth models using **time loadings** in the Λ matrix
These are the weights you use in Mplus

$$\mathbf{Y_i} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta_i} + \boldsymbol{\varepsilon_i}$$

$$\begin{bmatrix} Y_{1i} \\ Y_{2i} \\ Y_{3i} \\ Y_{4i} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \end{bmatrix}$$

1.2.2 Where do time loadings come from?

Intercept column is always all 1s

Linear slope column depends on:

- *the number of observation points / waves*
- *where you center time*

1.2.3 Center at age 15

Age	Intercept	Linear
15	1	0
16	1	1
17	1	2
18	1	3

1.2.4 Center at age 16

Age	Intercept	Linear
15	1	-1
16	1	0
17	1	1
18	1	2

1.2.5 Center at age 17

Age	Intercept	Linear
15	1	-2
16	1	-1
17	1	0
18	1	1

1.2.6 Center at age 18

Age	Intercept	Linear
15	1	-3
16	1	-2
17	1	-1
18	1	0

1.2.7 Different time loadings

Mplus syntax:

Time centered at age 15
```
int lin | alcuse15@0 alcuse16@1 alcuse17@2 alcuse18@3;
```

Time centered at age 16
```
int lin | alcuse15@-1 alcuse16@0 alcuse17@1 alcuse18@2;
```

Time centered at age 17
```
int lin | alcuse15@-2 alcuse16@-1 alcuse17@0 alcuse18@1;
```

Time centered at age 18
```
int lin | alcuse15@-3 alcuse16@-2 alcuse17@-1 alcuse18@0;
```

1.2.8 Unequally spaced assessments

All participants assessed on the same schedule
But the assessments do not need to be equally spaced

Month	1st session		3rd session		6th session	
	Intercept	Linear	Intercept	Linear	Intercept	Linear
0	1	0	1	-6	1	-36
2	1	2	1	-4	1	-34
6	1	6	1	0	1	-30
12	1	12	1	6	1	-24
24	1	24	1	18	1	-12
36	1	36	1	30	1	0

1.3 Quadratic effects

1.3.1 Quadratic change

Adding a quadratic effect of time involves changing the Λ matrix
A third column that **squares** the "Linear" values

	Age 15			Age 16			Age 17			Age 18		
Age	I	L	Q	I	L	Q	I	L	Q	I	L	Q
15	1	0	0	1	-1	1	1	-2	4	1	-3	9
16	1	1	1	1	0	0	1	-1	1	1	-2	4
17	1	2	4	1	1	1	1	0	0	1	-1	1
18	1	3	9	1	2	4	1	1	1	1	-0	0

You will only need the values from the "Linear" column for Mplus

1.3.2 Quadratic change in Mplus

Syntax changes:

```
int lin quad | alcuse15@0 alcuse16@1 alcuse17@2 alcuse18@3;
```

Additional parameters in the quadratic model:

- Variance of quadratic component
- Covariance between intercept and quadratic
- Covariance between linear slope and quadratic

As with mixed models:

- The quadratic variance component can be problematic
- Remember that centering changes the linear slope but does not affect the quadratic trend

1.4 Estimated time scores

1.4.1 Growth models with estimated time scores

Another alternative: don't specify the shape of change
Let the model figure out the best shape for the data

- Don't impose a specific curve (e.g., quadratic)
- Don't add additional latent parameters

- DO estimate some loadings that you didn't before
- Here, just estimate the first two loadings as 0 and 1

```
MODEL:
int lin | alcuse15@0 alcuse16@1 alcuse17 alcuse18;
```

Alternative version: Estimate the first loading as 0 and the last as 1

```
MODEL:
int lin | alcuse15@0 alcuse16 alcuse17 alcuse18@1;
```

1.4.2 Output for estimated time loadings

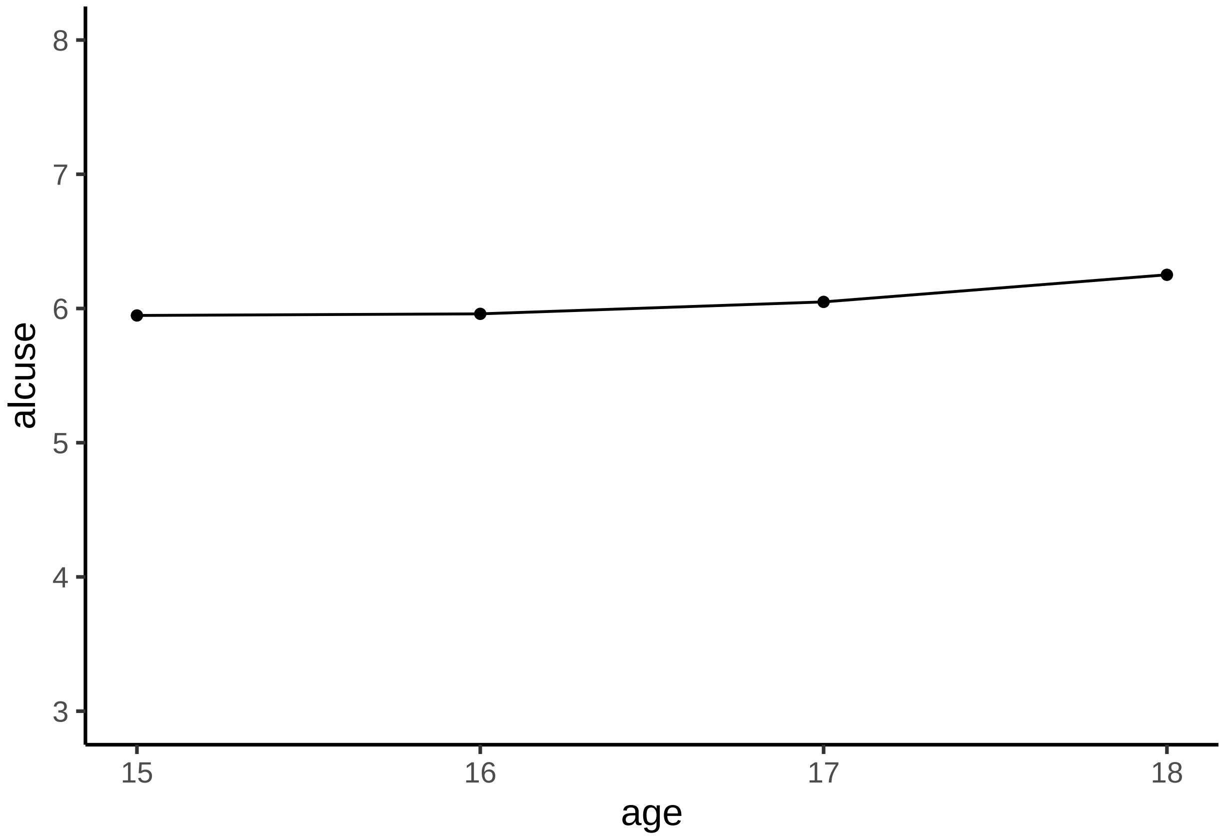
```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 INT      |
    ALCUSE15           1.000      0.000    999.000    999.000
    ALCUSE16           1.000      0.000    999.000    999.000
    ALCUSE17           1.000      0.000    999.000    999.000
    ALCUSE18           1.000      0.000    999.000    999.000
 LIN      |
    ALCUSE15           0.000      0.000    999.000    999.000
    ALCUSE16           1.000      0.000    999.000    999.000
    ALCUSE17           8.067     30.464      0.265      0.791
    ALCUSE18          24.187     96.530      0.251      0.802

...

 Means
    INT                5.948      0.087     68.371      0.000
    LIN                0.013      0.051      0.244      0.807
```

1.4.3 Figuring out means with estimated time loadings

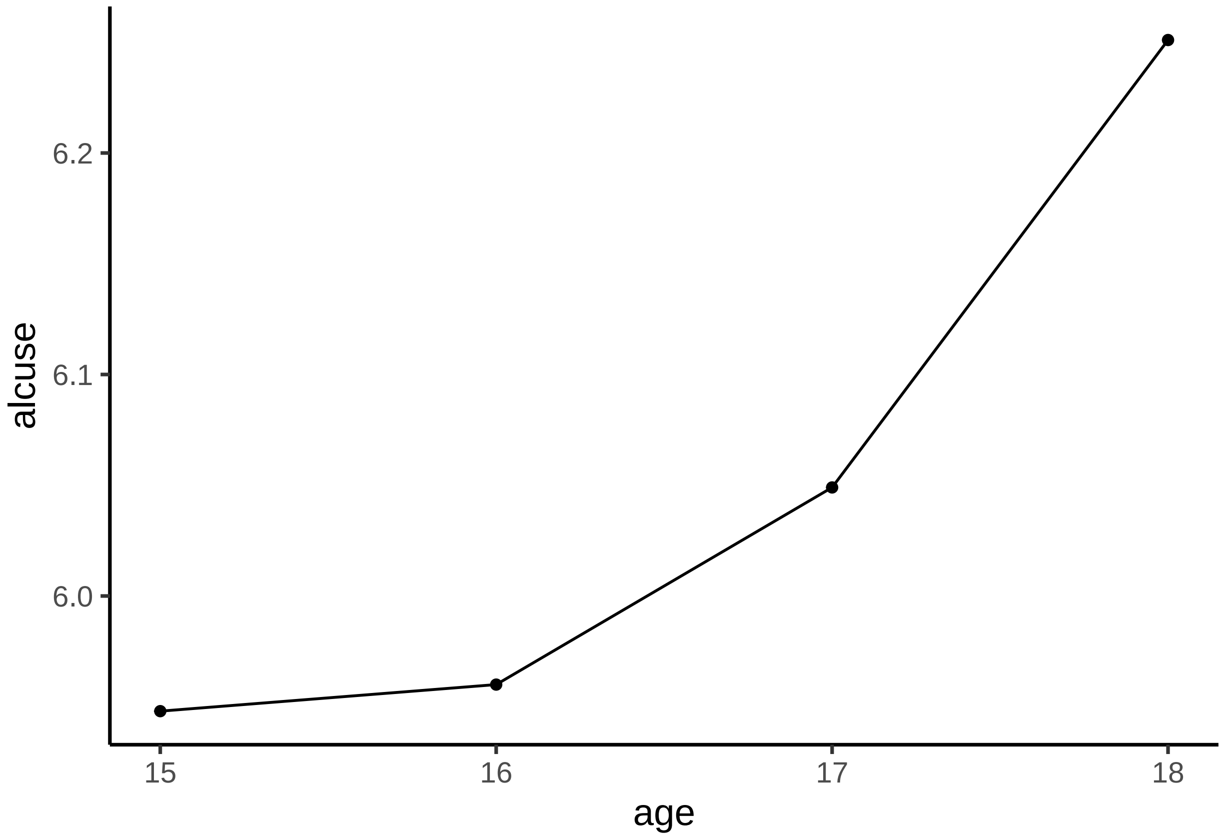
Predicted mean for a specific time =
Intercept + loading × slope

Age 15 : $5.948 + 0.000 \times (0.013) = 5.948$
Age 16 : $5.948 + 1.000 \times (0.013) = 5.960$
Age 17 : $5.948 + 8.067 \times (0.013) = 6.049$
Age 18 : $5.948 + 24.187 \times (0.013) = 6.251$

1.4.4 Predicted mean values



1.4.5 Predicted mean values - rescaled



1.4.6 Output: residual;

```
RESIDUAL OUTPUT

     ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED)

           Model Estimated Means
              ALCUSE15      ALCUSE16      ALCUSE17      ALCUSE18
              ________      ________      ________      ________
                5.948         5.960         6.049         6.251
```

1.5 Person-specific assessment

1.5.1 Person-specific assessment schedules

Many studies do not assess all participants on the same schedule
The assessment time frame can be very important

Use "time scores" instead of constant "time loadings"
Very similar to mixed model "time as a continuous predictor"

$$\Lambda = \begin{bmatrix} 1 & TS_{1i} \\ 1 & TS_{2i} \\ 1 & TS_{3i} \\ 1 & TS_{4i} \end{bmatrix}$$

The i subscript means different values for each subject

1.5.2 What the data look like

id	alcuse1	alcuse2	alcuse3	alcuse4	ciguse1	ciguse2	ciguse3	ciguse4	potuse1	potuse2	potuse3	potuse4
1	5	5	5	6	6	6	5	7	5	5	4	5
2	6	5	6	7	8	7	7	6	5	5	4	5
3	4	6	6	5	5	7	6	6	3	5	4	3
4	6	5	6	7	6	7	6	7	4	5	6	6
5	5	4	4	5	6	7	7	7	4	4	4	5
6	8	8	7	7	9	8	7	6	6	5	5	5

1.5.3 Mplus syntax for "time scores"

To run a growth model with person-specific assessments / time scores, make the following changes:

VARIABLE

1. List the time score variables (`age1c - age4c`) in `USEVARIABLES`
2. `tscores are age1c - age4c;`

MODEL

3. `icept linear | alcuse1 - alcuse4 at age1c - age4c;`

2 Model fit

2.1 Model fit

2.1.1 Relative vs absolute fit

Any SEM model can be assessed for

1. *Relative* fit: How good is a model *compared to another model*?
2. *Absolute* fit: Is a model *good*?

- You might also think about it as: Which is the *best* model? Is that model a *good* model?

We have already talked about **relative fit** in mixed models: LR tests, AIC

- This is virtually identical in SEM growth models

Absolute fit is somewhat specific to SEM models and is a bit more complicated

- Growth models have some additional complications when it comes to fit

2.1.2 Substance use data

Alcohol use at ages 15, 16, 17, and 18

- 101 subjects
- No missing data

We will focus **2 models**:

- "linear": linear model
- "linear eqresvar": linear model w equal residual variances

But we will also consult 2 other models to determine fit:

- "intercept": intercept only model
- "intercept eqresvar": int only w = residual variances

Several other models are also posted

2.1.3 Parameter estimates: "linear"

Linear latent growth model with 4 time points

9 free (estimated) parameters = "number of free parameters" in Mplus output

Two latent factor means: κ_1 is the average intercept, κ_2 is the average linear slope

Two factor variances: ψ_{11} is the intercept variance, ψ_{22} is the slope variance

One factor covariance: ψ_{21} is the covariance between intercept and slope

Four residual variances, one for each time point: $\theta_{\epsilon_{1i}}$, $\theta_{\epsilon_{2i}}$, $\theta_{\epsilon_{3i}}$, $\theta_{\epsilon_{4i}}$

2.1.4 Parameter estimates: "linear eqresvar"

Consider the alternative model where the residual variances are **fixed to be equal** at all time points

This model has fewer free parameters - 6 free parameters:

Two latent factor means (same as last slide)

Two factor variances (same as last slide)

One factor covariance (same as last slide)

One residual variance (**not** the same as last slide)

This model estimates a **single residual variance** across all time points, so instead of 4 different residual variances, there is only one value estimated

2.1.5 Parameter estimates

For a linear latent growth model, the number of factor means, factor variances, and factor covariances does not increase with more time points

- (Adding quadratic terms or predictors will increase parameters)

Since the residual variances are specific to a time point, more time points **increases the number of residual variances estimated**

- 4 time points: 2 + 2 + 1 + 4 = 9 parameters estimated
- 5 time points: 2 + 2 + 1 + 5 = 10 parameters estimated
- 6 time points: 2 + 2 + 1 + 6 = 11 parameters estimated
- 8 time points: 2 + 2 + 1 + 8 = 13 parameters estimated
- 10 time points: 2 + 2 + 1 + 10 = 15 parameters estimated

2.1.6 Why do we care about the # of parameters?

It is a good idea to know how many parameters you should have

- Compare what you think you did to what Mplus produces
- This becomes more important with more complex models

The number of parameters will be important in **comparing models**

If you want to compare the two models we just talked about:

- linear growth model with constrained or equal residual variances ("linear eqresvar")
- linear growth model with no such constraints ("linear")

You will **need to know** the number of free parameters in each model to perform the comparison test (LR test)

2.2 Relative fit

2.2.1 Comparing two models

Frequently, you will want to compare two similar models

- e.g., as part of your model building process

For example:

- Intercept only model vs. linear growth model
- Linear growth model vs. quadratic growth model
- Linear growth model with "free" residual variances vs. linear growth model with constrained residual variances

This is done the same way as for mixed models

- Likelihood ratio test

2.2.2 Likelihood ratio test

Mplus gives the *log-likelihood*, so multiply by -2 to get the -2LL

- Use log-likelihood from "H0 Value" table in Mplus

"linear" model:

```
MODEL FIT INFORMATION

Number of Free Parameters                        9

Loglikelihood

          H0 Value                        -484.077
          H1 Value                        -479.615
```

"linear eqresvar" model:

```
MODEL FIT INFORMATION

Number of Free Parameters                        6

```

```
Loglikelihood

          H0 Value                        -486.846
          H1 Value                        -479.615
```

2.2.3 Likelihood ratio test

The model with **more parameters** has a **smaller -2LL**: closer to perfect fit

Model	# of parameters	-2LL
Estimated / unconstrained residual variance	9	968.154
Equal / constrained residual variance	6	973.692
(Absolute) difference	3	5.538

The critical value for chi-square with 3 df and $\alpha = .05$ is 7.815

- The observed value we got was 5.538 $\rightarrow$ Not significant

Simpler model ("linear eqresvar") is better

- More complex model is not worth extra parameters

- (If the value were larger than 7.815, the model with unconstrained residual variances ("linear") would be preferred)

2.3 Absolute fit

2.3.1 Assessing fit in CFAs

Confirmatory factor analysis relies on "goodness of fit"

- In a very general way, this refers to how well the model **re-creates the observed data** (specifically, the observed covariance matrix)

- A model with "good fit" does a good job re-creating the observed data

- A model with "poor fit" doesn't do a very good job re-creating the observed data

There are a number of "fit statistics" that are used

- Each fit statistic has strengths and weaknesses

- Hopefully, they come to a consensus

- (Frequently do not)

2.3.2 Fit statistics for CFA (1)

Chi-squared test

- observed covariance matrix vs. expected covariance matrix

- significant test means "significant misfit"

- almost always significant with large samples

RMSEA (root mean squared error of approximation)

- model population vs. population covariance matrix
- <.05 is good fit, <.08 is ok fit

2.3.3 Fit statistics for CFA (2)

SRMR (standardized root mean square residual)

- residuals between sample and model covariance matrices
- <.05 is good fit, <.08 is ok fit

CFI (comparative fit index) (**don't use CFI from Mplus output**)

- data vs. hypothesized model, adjusts for sample size
- >.95 is good fit (used to be >.9)

2.3.4 Fit for SEM growth models

The SEM latent growth model is a special case of CFA

But fit indices for growth models have several additional complications

There are three major issues:

- Assessment schedule
- Misfit in covariance versus mean portions of model
- The correct null model

2.3.5 Assessment schedule

SEM fit statistics that are based on the **model-implied covariance matrix** are only defined if the assessment schedule is the same for everyone (i.e., $\mathbf{\Lambda}$ has no i subscripts)

- $\mathbf{\Lambda}$ is the matrix of time loadings

Model-implied covariance matrix: $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T + \boldsymbol{\Theta}_{\boldsymbol{\varepsilon}}$

Model-implied mean vector: $\boldsymbol{\mu} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\kappa}$

- With person-specific loadings, each person would end up with a **different** model-implied covariance matrix and a **different** model-implied mean vector

No way to compute global fit measures with person-specific assessments

- including χ^2, CFI, RMSEA

2.3.6 Misfit in covariance (measurement) vs mean (structural) part of the model

Traditionally, CFAs are based on the covariance structure only

- Not really concerned with the mean structure

Growth models are a special type of CFA

- But growth models incorporate a **mean structure**

If a global fit measure (χ^2, CFI, RMSEA) shows misfit, you don't know if the misfit is in the covariance structure, the mean structure, or both

It is possible to construct models that assess the mean and covariance portions of the model separately

- This method is complicated and a bit beyond the scope of this course

– You need a good knowledge of SEM to understand it

2.3.7 The correct null model

Every time you do a statistical test, you are comparing your obtained test statistic against a **null value**

- Think "null hypothesis"
- Regression coefficient: null hypothesis is that the coefficient = 0

Comparative fit indices (like CFI) have a null model

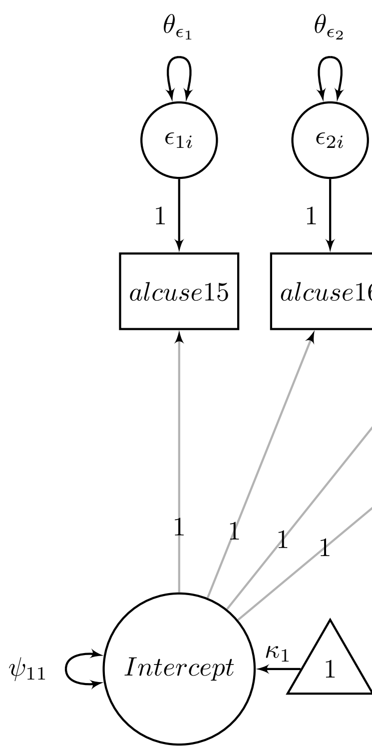
- The null model that SEM programs use for growth models is the wrong null model to compare to

- The null model should be **nested within the model of interest**
  - "Nested" means that you can get to the simpler model by doing nothing but fixing paths in the more complex model to 0

Mplus uses "all variables uncorrelated" as the null
The correct null model is an **intercept only model**

- Widaman & Thompson (2003) in Psychological Methods

2.3.8 Intercept only growth model



2.3.9 CFI for growth model

Calculate CFI using the χ^2 values and their degrees of freedom from you model and its correct null model (intercept only)

$$CFI = \frac{({\chi^2}_{null} - df_{null}) - ({\chi^2}_{model} - df_{model})}{({\chi^2}_{null} - df_{null})}$$

Report this CFI value, not the CFI value that Mplus gives you

(Remember: can't get χ^2 tests with person-specific assessment schedule – this can only be done if all participants are on the same assessment schedule)

2.3.10 CFI for substance use model "linear"

From "linear eqresvar" model:

```
MODEL FIT INFORMATION

Number of Free Parameters                        6
...
Chi-Square Test of Model Fit

          Value                             14.462
          Degrees of Freedom                     8
          P-Value                           0.0705
```

From "intercept eqresvar" model (null):

```
MODEL FIT INFORMATION

Number of Free Parameters                        3
...
Chi-Square Test of Model Fit

          Value                             33.348
          Degrees of Freedom                    11
          P-Value                           0.0005
```

2.3.11 CFI for substance use model "linear"

$$CFI = \frac{({\chi^2}_{null} - df_{null}) - ({\chi^2}_{model} - df_{model})}{({\chi^2}_{null} - df_{null})}$$

$$= \frac{(33.348 - 11) - (14.462 - 8)}{(33.348 - 11)} = 0.711$$

- "null" is the null model: intercept model with equal residual variances
- "model" is the linear model with equal residual variances

Compare 0.711 to (wrong) Mplus value of 0.938

2.3.12 What should I do about model fit?

Same assessment schedule for everyone:

- Make sure to calculate the correct CFI
- Be aware that misfit (if there is misfit) can exist in the covariance structure, the mean structure, or both

Different assessment schedules:

- Both points above, plus global fit statistics can't be calculated

Assessing fit of a growth model is complicated, depends on your specific model

2.4 Model fit summary

2.4.1 Model fit for SEM latent growth models

- Chi-square test: report but usually mostly ignore

- RMSEA: $< .05$ is good, $< .08$ is ok

- SRMR: $<.05$ is good, $<.08$ is ok

- CFI: need to calculate by hand, $> .95$ is good

Recall shortcomings due to:

- Person-specific assessment schedule: no χ^2, CFI, RMSEA
- Difficult to tell if misfit is in structural or measurement portion of model

2.4.2 Model fit for alcohol use linear model

- Chi-square test: $\chi^2(8) = 14.462$, NS
- RMSEA: 0.089 with CI [0.000, 0.162]
- SRMR: 0.208
- CFI (calculated by hand) = 0.711

Chi-square is good, RMSEA is bad but close, SRMR and CFI are bad

- This model is not very good – how can we improve it?
- A bad model can be an **opportunity** to improve your model

2.4.3 Big picture, take home message

There are many things to consider about fit for growth models

In addition to fit statistics (χ^2, CFI, etc.) and likelihood ratio tests, you should look at parameter estimates themselves

Consider a basic quadratic model vs. a quadratic model with quadratic variance fixed to 0:

- LR test: quadratic variance should be fixed to 0

- But quadratic variance is significant

Whether you keep the quadratic variance is more of a judgment call

- Theory may inform your decision

3 Predictors

3.1 Predictors of growth

3.1.1 Adding predictors of growth

Remember including predictors of growth in mixed models

- Very complex
- Easy to get confused
- Many terms in the model

Including predictors of growth in SEM growth model is fairly simple

Recall that we can have **time varying** predictors, **time invariant** predictors, or both

3.2 Time-invariant predictors

3.2.1 Time-invariant predictors

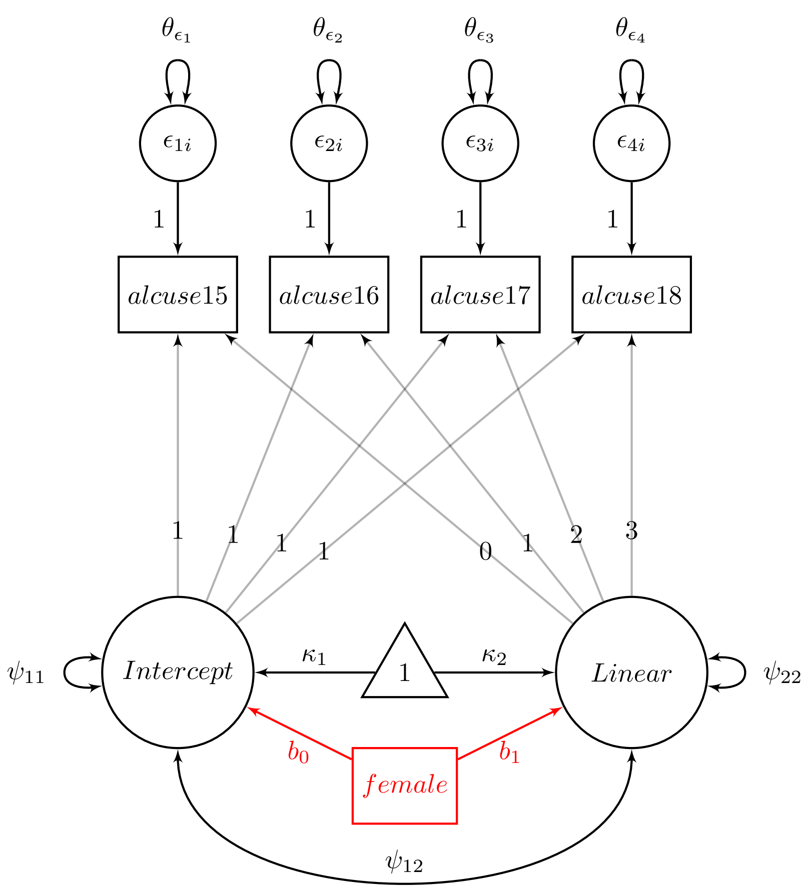
Predictors of growth that **do not** vary over time

We called these "level 2 predictors" in mixed models

Any variable that has the same value at all assessments

Time-invariant predictors are added as **predictors of the latent growth parameters**, η_1 and η_2 (and η_3, if there is a quadratic trend)

3.2.2 Time-invariant predictors



3.2.3 Syntax for time-invariant predictors

Include this line in your MODEL statement:

```
icept linear on gender;
```

`icept linear` are whatever you named your intercept and slope (and quadratic trend, if you have one)

The time-invariant predictor does not **have to** predict all growth parameters if your theory doesn't dictate it

- but it's a good place to start

3.2.4 Time-invariant predictors

"linear eqresvar - invarpred"

- Alcohol use at age 15 through 18
- Linear growth model with equal residual variances
- Gender (female: male = 0, female = 1) is a time invariant predictor of intercept and linear slope
- Time centered at age 15

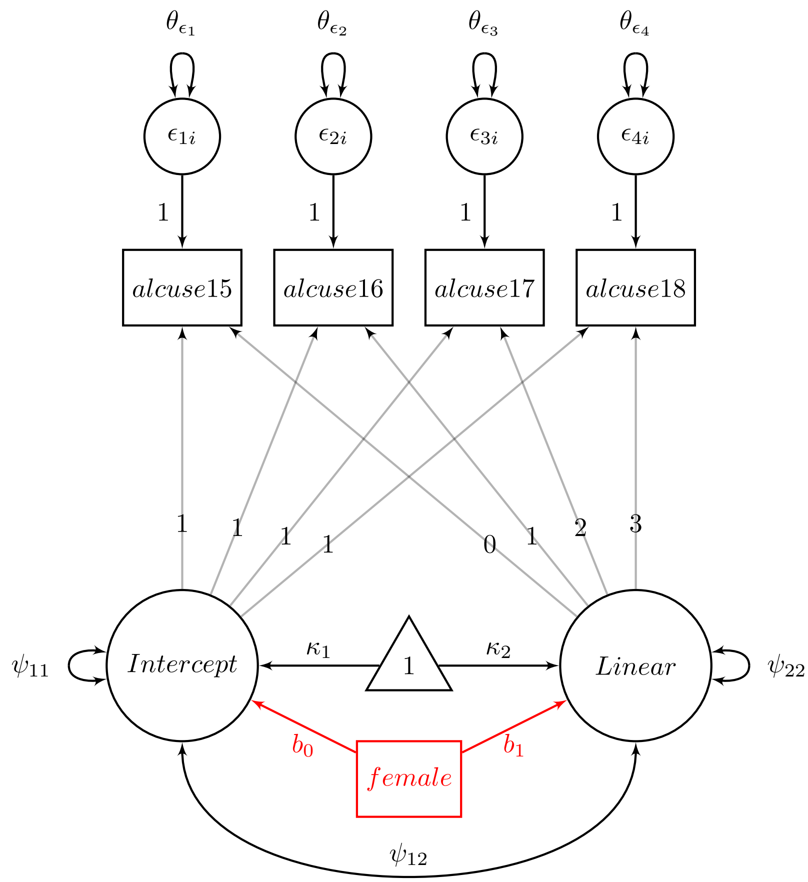
2 latent variable means (intercept, linear slope)
2 latent variable variances (intercept, linear slope)
1 latent variable covariance (Int w Lin)
1 residual variance (same for all time points)
2 effects of gender on growth parameters
= 2 + 2 + 1 + 1 + 2 = **8 estimated parameters**

3.2.5 "linear equresvar - invarpred" results

```
MODEL RESULTS

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

 INT      |
    ALCUSE15           1.000      0.000    999.000    999.000
    ALCUSE16           1.000      0.000    999.000    999.000
    ALCUSE17           1.000      0.000    999.000    999.000
    ALCUSE18           1.000      0.000    999.000    999.000

 LIN      |
    ALCUSE15           0.000      0.000    999.000    999.000
    ALCUSE16           1.000      0.000    999.000    999.000
    ALCUSE17           2.000      0.000    999.000    999.000
    ALCUSE18           3.000      0.000    999.000    999.000

 INT      ON
    FEMALE             0.097      0.182      0.533      0.594

 LIN      ON
    FEMALE             0.126      0.067      1.879      0.060

 LIN      WITH
    INT               -0.087      0.039     -2.240      0.025

 Intercepts
    ALCUSE15           0.000      0.000    999.000    999.000
    ALCUSE16           0.000      0.000    999.000    999.000
```

```
    ALCUSE17            0.000      0.000    999.000    999.000
    ALCUSE18            0.000      0.000    999.000    999.000
    INT                 5.838      0.133     43.922      0.000
    LIN                 0.040      0.049      0.823      0.411

 Residual Variances
    ALCUSE15            0.401      0.040     10.050      0.000
    ALCUSE16            0.401      0.040     10.050      0.000
    ALCUSE17            0.401      0.040     10.050      0.000
    ALCUSE18            0.401      0.040     10.050      0.000
    INT                 0.549      0.120      4.572      0.000
    LIN                 0.033      0.018      1.855      0.064
```

3.2.6 Time-invariant predictors



3.2.7 Intercept interpretation

Average intercept is the average value of the outcome at time = 0, for the group coded 0 on the predictor (gender)

- Average intercept = 5.838 (at age 15, for boys)
- At age 15, **boys** had an average alcohol use score of 5.838, p<.001

Intercept variance tells you how the intercept varies across people
About 95% of people have intercepts in [mean intercept $\pm 2\sqrt{variance}$]

- Intercept variance = 0.549, $p<.001$
- SD = $\sqrt{0.549}$ = 0.741
- At the beginning of the study, about 95% of boys had alcohol use scores between 4.39 and 7.29

3.2.8 Linear slope interpretation (time invariant)

Average linear slope is the average linear increase in the outcome for a 1-unit increase in time for the group coded 0 on the predictor (gender)

- Average slope = 0.040
- The linear change in alcohol use for **boys** is 0.040 points per year, NS

Slope variance tells you how the slope varies across people
About 95% of people have slope in [mean slope $\pm 2\sqrt{variance}$]

- Slope variance = 0.033, $p<.10$
- SD = $\sqrt{0.033}$ =0.182
- About 95% of boys had linear slopes between -0.32 and 0.40

3.2.9 Intercept-slope-quadratic covariance interpretation

Covariance between intercept and linear slope = -0.087, $p<.05$

Intercepts and slopes vary together

- Someone's intercept is related to their slope
- Higher intercept with lower slope
- Lower intercept with higher slope

3.2.10 Gender effects

Gender effect on the intercept = 0.097, NS

- At age 15, **girls** had an average alcohol use score of 5.838 + 0.097 = 5.935
- Girls' alcohol use at age 15 was no different from that of boys

Gender effect on the linear slope = 0.126, $p<.10$

- **Girls** had an average linear growth of 0.040 + 0.126 = 0.166 points per year
- Girls' average linear change was positive and marginally different from that of boys

3.2.11 Equations for boys and girls

Boys:

$$Alco\hat{h}oluse = 5.838 + 0.040(age15)$$

Girls:

$$Alco\hat{h}oluse = 5.935 + 0.166(age15)$$

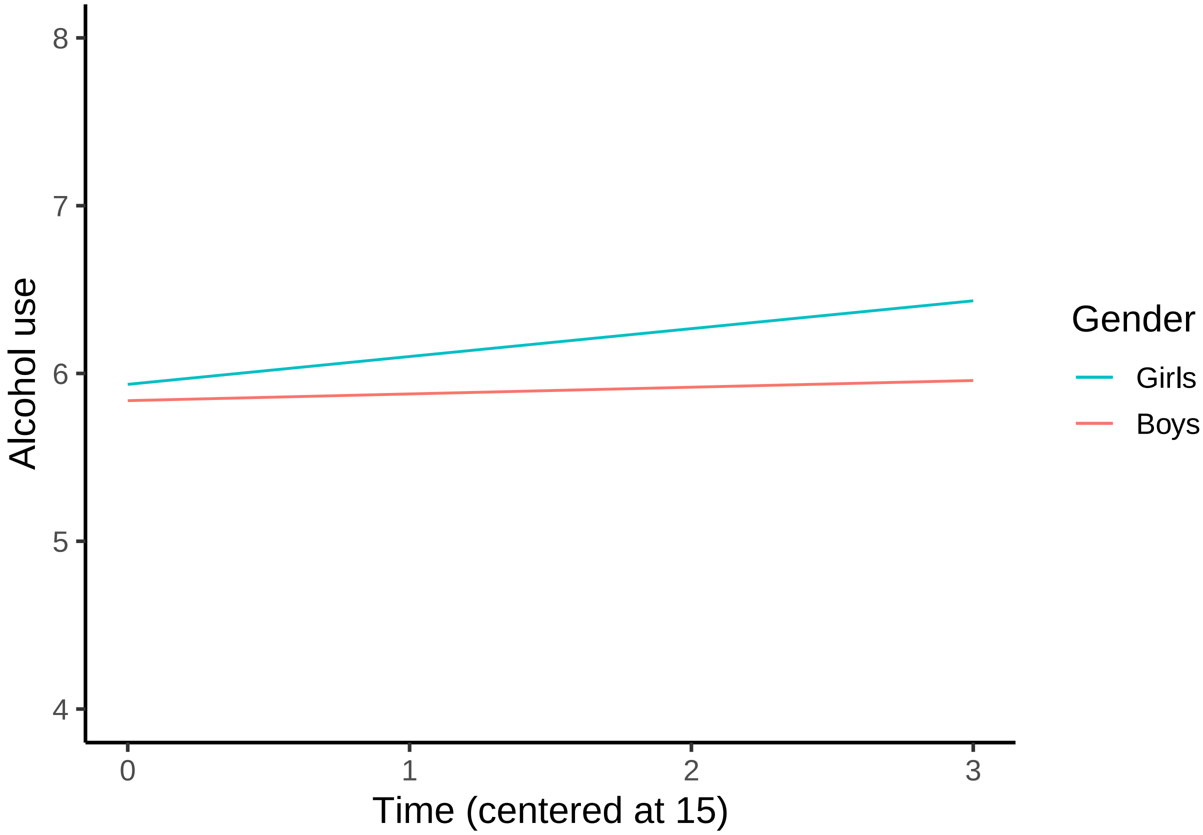
No differences between boys and girls:

- Average alcohol use at age 15

Differences between boys and girls:

- Average linear slope (marginal)

3.2.12 Average lines



3.2.13 Why?

Why would I show you this model where the predictor doesn't seem to predict?

Model comparison: Is the model **with** gender better than the model **without**?

Model	# of parameters	-2LL
With gender	8	964.612
Without gender	6	973.692
(Absolute) difference	2	9.08

The critical value for chi-square with 2 df and $\alpha = .05$ is 5.99

- The observed value we got was 9.08 $\rightarrow$ Significant

More complex model (with gender) is better

- Even if the predictors are not significant predictors, the overall model is better

- Considers all additional parameters at once

3.3 Time-varying predictors

3.3.1 Time-varying predictors

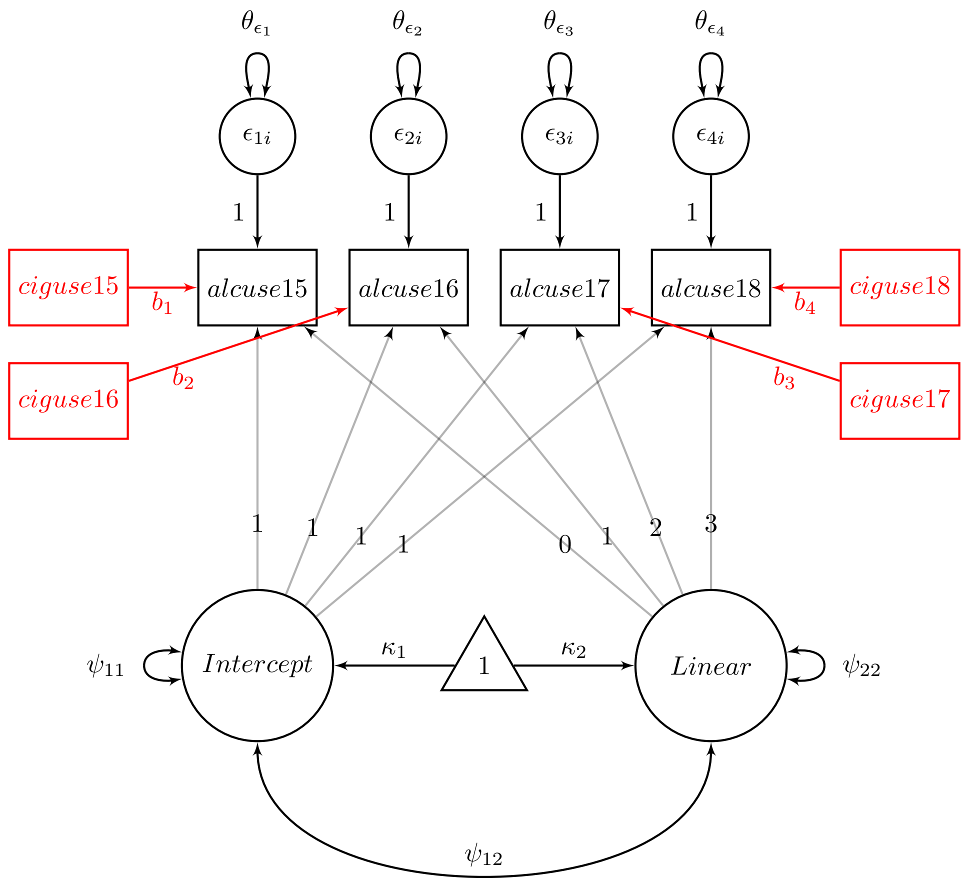
Predictors of growth that **can** change over time

(We called these "level 1 predictors" in mixed models)

Stress, substance use, blood test results, etc.

Time-varying predictors are added as **predictors of their respective outcome variable**

3.3.2 Time-varying predictors



3.3.3 Syntax for time-varying predictors

Include these lines in your MODEL statement:

```
Y1 on X1;
Y2 on X2;
Y3 on X3;
Y4 on X4;
```

`Y1 - Y4` are the outcome variable

`X1 - X4` are the time-varying predictor variables

Can also **lag** time varying predictors:

- `X1 - X4` predict `Y2 - Y5`, etc.

- Requires a little more thought and additional waves

3.3.4 Time-varying predictor example

"linear eqresvar - varypred"

- Alcohol use at age 15 through 18
- Linear growth model with equal residual variances
- Cigarette use is a time varying predictor of alcohol use
- Time centered at age 15
- Gender is not in this model

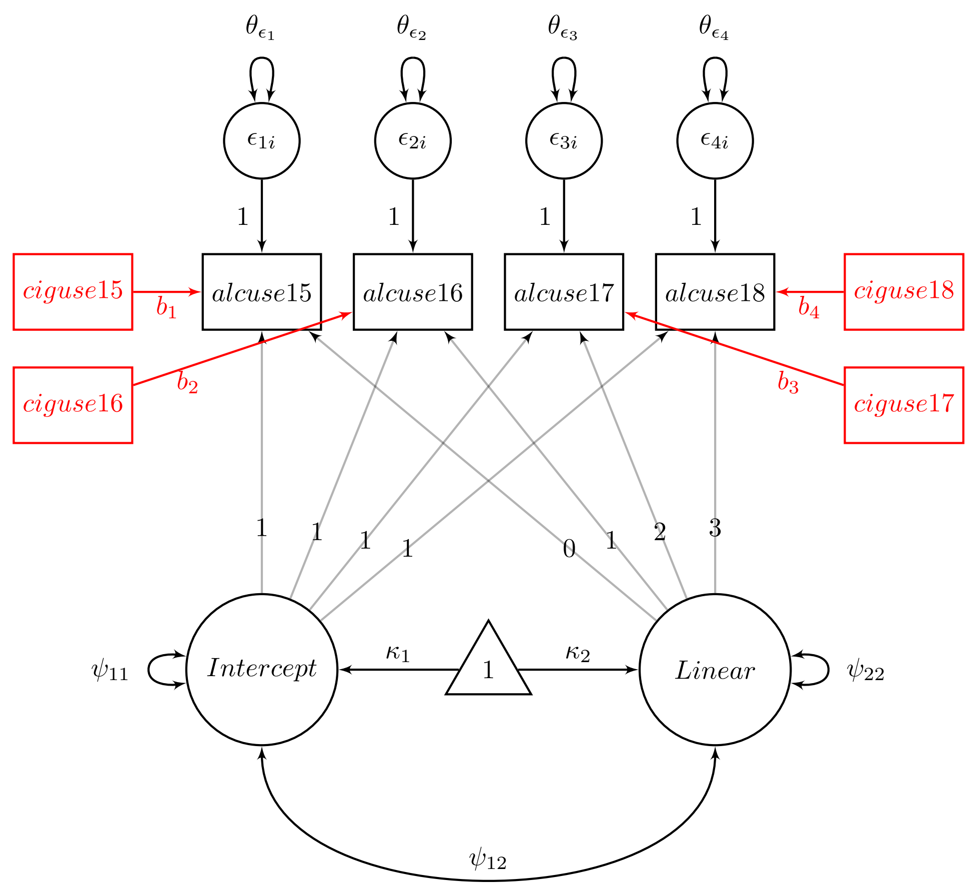
2 latent variable means (intercept, linear slope)
2 latent variable variances (intercept, linear slope)
1 latent variable covariance (Int w Lin)
1 residual variance (same for all time points)
4 effects of cigarette use on alcohol use (at 4 time points)
= 2 + 2 + 1 + 1 + 4 = 10 estimated parameters

3.3.5 "linear eqresvar - varpred" results

```
MODEL RESULTS

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

 INT      |
    ALCUSE15           1.000      0.000    999.000    999.000
    ALCUSE16           1.000      0.000    999.000    999.000
    ALCUSE17           1.000      0.000    999.000    999.000
    ALCUSE18           1.000      0.000    999.000    999.000

 LIN      |
    ALCUSE15           0.000      0.000    999.000    999.000
    ALCUSE16           1.000      0.000    999.000    999.000
    ALCUSE17           2.000      0.000    999.000    999.000
    ALCUSE18           3.000      0.000    999.000    999.000

 ALCUSE15 ON
    CIGUSE15           0.344      0.065      5.295      0.000

 ALCUSE16 ON
    CIGUSE16           0.275      0.046      5.984      0.000

 ALCUSE17 ON
    CIGUSE17           0.225      0.040      5.616      0.000

 ALCUSE18 ON
    CIGUSE18           0.179      0.049      3.662      0.000

 LIN      WITH
```

```
   INT               -0.050      0.033     -1.494      0.135

Means
   INT                3.726      0.419      8.884      0.000
   LIN                0.446      0.176      2.542      0.011

Intercepts
   ALCUSE15           0.000      0.000    999.000    999.000
   ALCUSE16           0.000      0.000    999.000    999.000
   ALCUSE17           0.000      0.000    999.000    999.000
   ALCUSE18           0.000      0.000    999.000    999.000

Variances
   INT                0.352      0.094      3.739      0.000
   LIN                0.029      0.017      1.683      0.092

Residual Variances
   ALCUSE15           0.398      0.040     10.016      0.000
   ALCUSE16           0.398      0.040     10.016      0.000
   ALCUSE17           0.398      0.040     10.016      0.000
   ALCUSE18           0.398      0.040     10.016      0.000
```

3.3.6 Time varying predictors



3.3.7 Time-varying predictor interpretation

Effect of cigarette use at age 15 on alcohol use at age 15 = 0.334, $p<.001$

Effect of cigarette use at age 16 on alcohol use at age 16 = 0.275, $p<.001$

Effect of cigarette use at age 17 on alcohol use at age 17 = 0.225, $p<.001$

Effect of cigarette use at age 18 on alcohol use at age 18 = 0.179, $p<.001$

Cigarette use is positively related to alcohol use at all time points

3.3.8 Intercept interpretation (time varying)

Average intercept is the average value of outcome at time = 0, **controlling for the predictor** (cigarette use)

- Average intercept = 3.726 (at 15), controlling for cigarette use

- At age 15, participants had an average alcohol use score of 3.726, $p<.001$, after controlling for cigarette use

Intercept variance tells you how the intercept varies across people About 95% of people have intercepts in [mean intercept $\pm 2\sqrt{variance}$]

- Intercept variance = 0.352, $p<.001$

- SD = $\sqrt{0.352} = 0.593$

- At the beginning of the study, about 95% of participants had alcohol use scores between 2.56 and 4.89, after controlling for cigarette use

3.3.9 Linear slope interpretation

Average linear slope is the average increase in the outcome for a 1-unit increase in time, **controlling for the predictor**

- Average slope = 0.446, $p<.05$

- At age 15, the linear change in alcohol use is 0.446 points per year, $p<.05$, after controlling for cigarette use

Slope variance tells you how the slope varies across people About 95% of people have slopes in [mean slope $\pm 2\sqrt{variance}$]

- Slope variance = 0.029, $p<.10$

- SD = $\sqrt{0.029} = 0.170$

- About 95% of participants had slopes between 0.11 and 0.78, after controlling for cigarette use

3.3.10 What does "controlling for" mean again?

Controlling for a variable means holding it constant – usually at 0

In this model, that means `ciguse` = 0

- But that value doesn't exist in the data / measure
- `ciguse` ranges from 3 to 8

We have a line that we can plot: $\hat{alcuse} = 3.726 + 0.446 age_{15}$

- But it doesn't mean a lot right now

It is a good idea to **mean center** your predictors

- Then you're holding the predictor constant at **its mean**

3.3.11 Average line