

# Statistical Analysis Using Structural Equation Models

EPsy 8266

Christopher David Desjardins

Research Methodology Consulting Center

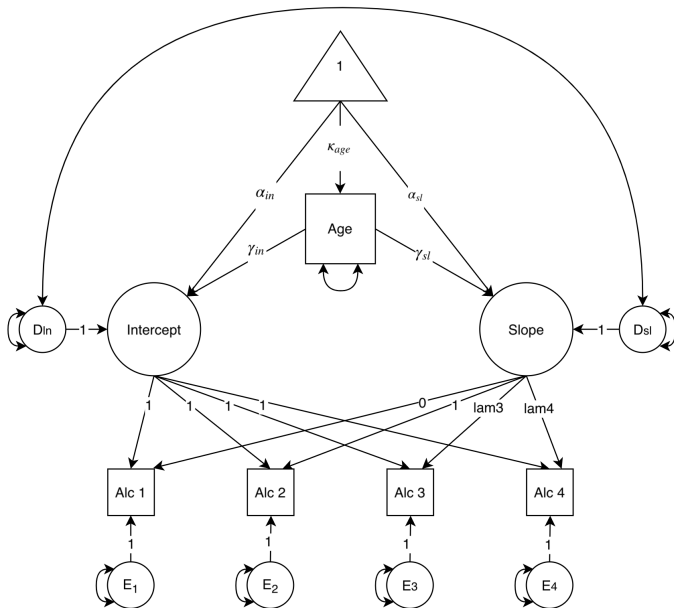
4/11/19

```

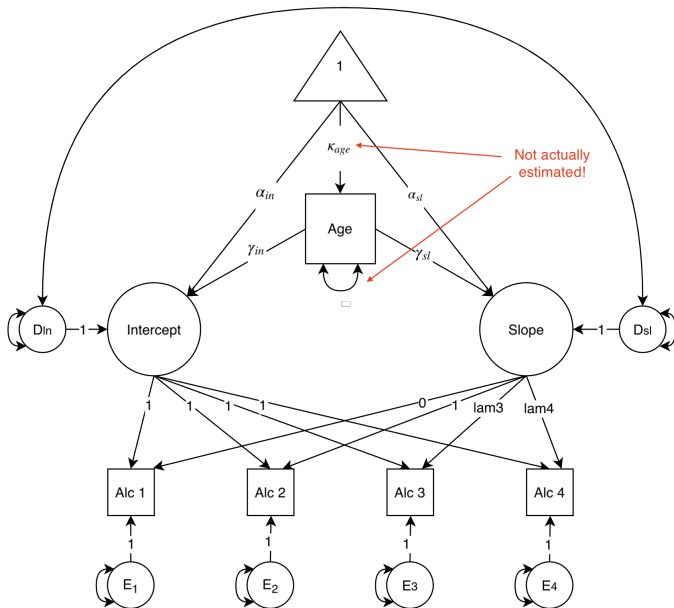
library(lavaan)
alcuse.xbar <- c(0.92, 0.76, 0.73, 0.62, 0.08, 0.08, 0.08, 0.07, 22.69, 0.54, 0.30,
               0.14, 12.72)
alcuse.sigma <- c(1.68, 1.35, 1.26, 1.08, 0.27, 0.27, 0.27, 0.26, 1.38, 0.49, 0.46,
               0.35, 2.02)
lower.cor <- '
1.00
.494 1.00
.440 .519 1.00
.382 .471 .510 1.00
-.074 -.068 -.062 -.035 1.00
-.023 -.048 -.057 -.055 -.086 1.00
.009 -.003 -.036 -.039 -.085 -.083 1.00
.043 .025 .011 -.022 -.081 -.080 -.079 1.00
.032 .020 -.005 .010 .050 .028 .007 .028 1.00
.231, .238 .252 .264 -.048 -.005 -.028 -.003 .011 1.00
-.142 -.146 -.120 -.118 -.085 -.080 -.067 -.071 -.014 -.013 1.00
-.025 -.014 -.028 -.001 .012 .019 -.021 -.013 -.023 .040 -.265 1.00
.012 -.005 -.021 -.018 .006 .006 .009 .049 .222 -.110 -.149 -.125 1.00'
alcuse.cov <- getCov(lower.cor,
                    names = c(paste0("alc", 1:4), paste0("mar", 1:4),
                              "age", "male", "black", "hispanic", "education"),
                    sds = alcuse.sigma)

```

# Age as a predictor



# Age as a predictor



# Prediction model

```
age.mod <- "  
# define factors  
int =~ 1*alc1 + 1*alc2 + 1*alc3 + 1*alc4  
slope =~ 0*alc1 + 1*alc2 + lam3*alc3 + lam4*alc4  
  
# regressions  
int ~ 1 + age  
slope ~ 1 + age  
  
# estimate factor variance/covariance  
int ~~ int + slope  
slope ~~ slope  
  
# assume variances are NOT equivalent across time  
alc1 ~~ alc1  
alc2 ~~ alc2  
alc3 ~~ alc3  
alc4 ~~ alc4  
"  
age.fit <- lavaan(model = age.mod, sample.cov = alcuse.cov,  
                  sample.mean = alcuse.xbar, sample.nobs = 4052)
```

# Fit Measures for Age Model

```
fitmeasures(age.fit, c("chisq", "df", "pvalue", "rmsea", "cfi", "tli", "srmr"))
```

```
##  chisq      df pvalue  rmsea    cfi    tli    srmr  
## 14.606   5.000   0.012   0.022   0.998   0.995   0.010
```

# Residual correlation matrix

```
resid(age.fit, type = "cor")$cov
```

```
##      alc1  alc2  alc3  alc4  age
## alc1  0.000
## alc2  0.001  0.000
## alc3 -0.013  0.014  0.000
## alc4 -0.001 -0.006  0.002  0.000
## age   0.008  0.001 -0.018  0.008  0.000
```

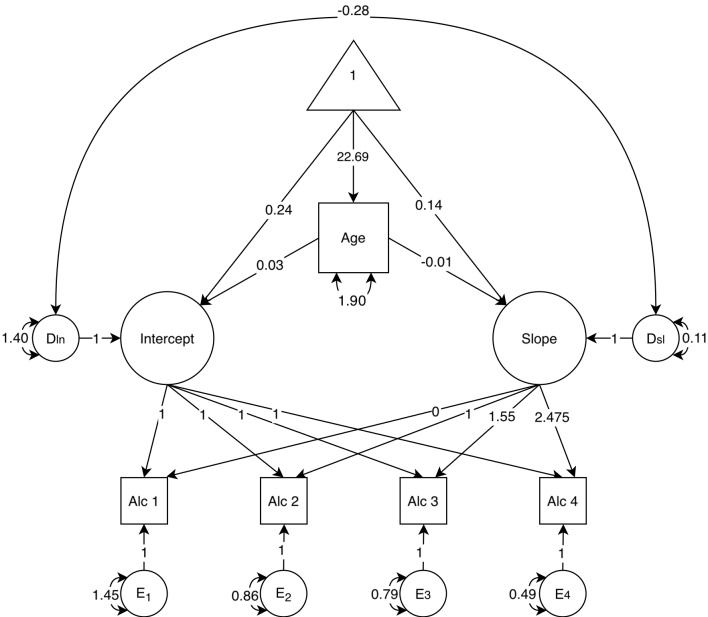
# Parameter Estimates

```
parameterEstimates(age.fit, ci = TRUE, standardized = TRUE, pvalue = FALSE)
```

##	lhs	op	rhs	label	est	se	z	ci.lower	ci.upper	std.lv	std.all	std.nox
## 1	int	=~	alc1		1.000	0.000	NA	1.000	1.000	1.184	0.701	0.701
## 2	int	=~	alc2		1.000	0.000	NA	1.000	1.000	1.184	0.881	0.881
## 3	int	=~	alc3		1.000	0.000	NA	1.000	1.000	1.184	0.942	0.942
## 4	int	=~	alc4		1.000	0.000	NA	1.000	1.000	1.184	1.093	1.093
## 5	slope	=~	alc1		0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000
## 6	slope	=~	alc2		1.000	0.000	NA	1.000	1.000	0.335	0.249	0.249
## 7	slope	=~	alc3	lam3	1.554	0.163	9.510	1.234	1.875	0.520	0.414	0.414
## 8	slope	=~	alc4	lam4	2.475	0.310	7.974	1.866	3.083	0.828	0.765	0.765
## 9	int	~1			0.238	0.389	0.612	-0.525	1.002	0.201	0.201	0.201
## 10	int	~	age		0.029	0.017	1.699	-0.004	0.063	0.025	0.034	0.025
## 11	slope	~1			0.137	0.160	0.854	-0.177	0.451	0.410	0.410	0.410
## 12	slope	~	age		-0.011	0.007	-1.539	-0.025	0.003	-0.033	-0.046	-0.033
## 13	int	~~	int		1.399	0.079	17.608	1.243	1.555	0.999	0.999	0.999
## 14	int	~~	slope		-0.283	0.061	-4.630	-0.402	-0.163	-0.714	-0.714	-0.714
## 15	slope	~~	slope		0.112	0.038	2.937	0.037	0.186	0.998	0.998	0.998
## 16	alc1	~~	alc1		1.451	0.070	20.626	1.313	1.589	1.451	0.509	0.509
## 17	alc2	~~	alc2		0.858	0.030	29.004	0.800	0.916	0.858	0.475	0.475
## 18	alc3	~~	alc3		0.786	0.022	35.313	0.743	0.830	0.786	0.498	0.498
## 19	alc4	~~	alc4		0.488	0.028	17.388	0.433	0.542	0.488	0.416	0.416
## 20	age	~~	age		1.904	0.000	NA	1.904	1.904	1.904	1.000	1.904
## 21	alc1	~1			0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000
## 22	alc2	~1			0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000
## 23	alc3	~1			0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000
## 24	alc4	~1			0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000
## 25	age	~1			22.690	0.000	NA	22.690	22.690	22.690	16.444	22.690



# Fitted model



# The constants in this prediction model

- ▶ The direct effect of the constant onto the intercept factor is 0.239 and onto the slope factor is 0.137.
- ▶ These are not means but intercepts
- ▶ The total effect of the constant, however, are means.
  - ▶ Direct effect of the constant plus the indirect effect through age.
  - ▶ For intercept:  $0.24 + 22.69 * 0.03 = 0.92$
  - ▶ For slope:  $0.14 + 22.69 * -.01 = -.09$

## Activity

Rather than using the predictor age, let's see if biological sex is related to self-reported alcohol use in Year 1 and change in self-reported alcohol use.

The variable is called `male` and takes on values of 1 when a participant is male and 0 when they are a female.

Run the nonlinear curve model with `male` as a predictor and interpret the output.

## Baldwin, Imel, Braithwaite, & Atkins, 2014. "Analyzing Multiple Outcomes in Clinical Research Using Multivariate Multilevel Models"

- ▶ Described multivariate extensions to multilevel models (longitudinal growth curve model)
- ▶ Bivariate example: simulated data to mimic a clinical trial comparing cognitive behavioral therapy to a no-treatment control for the treatment of depression
  - ▶ 3 time points (baseline, midtreatment, posttreatment)
  - ▶ 100 participants (50 per condition), and two outcomes - depression and quality of life.
  - ▶ Time coded as 0, 1, and 2.
  - ▶ Coded treatment as 1 for CBT and 0 for control.
- ▶ R script

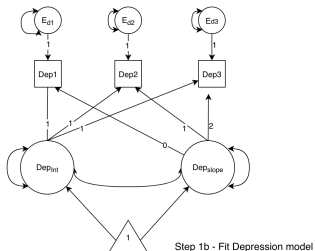
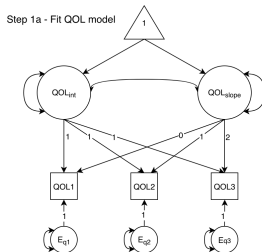
# Modeling strategy

1. Fit linear growth models of depression and quality of life separately
2. Add treatment variable
3. Combine univariate models

```
# Read in data in wide format
baldwin_data <- read.csv("https://bit.ly/2G7ibPf")
# y1: depression
# y2: quality of life

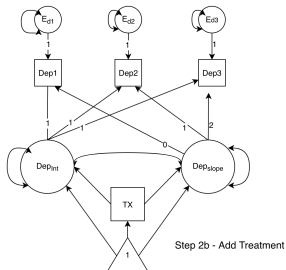
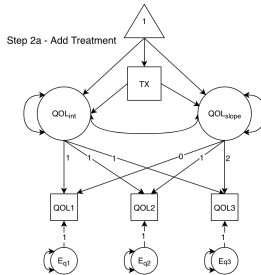
# need to make even wider for lavaan`
baldwin_wide <- reshape(baldwin_data,
                        direction = "wide",
                        v.names = c("y1", "y2"),
                        timevar = "time",
                        idvar = "pid")
colnames(baldwin_wide) <- c("pid", "tx",
                          paste0(rep(c("dep", "qol"), 3), rep(0:2, each = 2)))
```

# Step 1 - Fit models



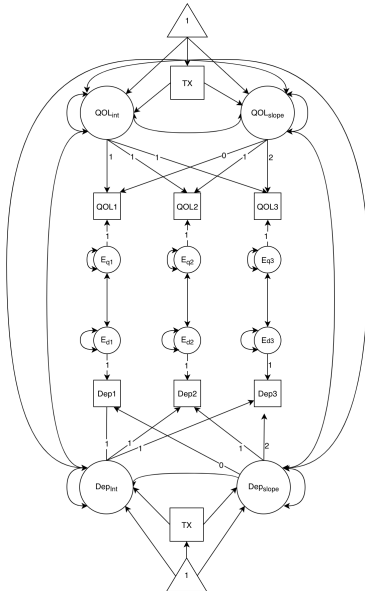
Can constrain residuals and consider autocorrelation, if necessary

## Step 2 - Add treatment





## Step 3 - Combine the two models



# What more?

- ▶ Can test if treatment effect is the same for QOL as depression
  - ▶ Name parameters
- ▶ Residual covariances within time can be constrained
- ▶ Could examine intercept as a predictor of slope
  - ▶ Change bidirectional arrow to a unidirectional arrow
  - ▶  $\text{int} \sim \sim \text{slope} \rightarrow \text{slope} \sim \text{int}$