We thank you for your time spent taking this survey.
Your response has been recorded.

Below is a summary of your responses

Download PDF

Welcome to the Day 6 Practical!

During today's practical we will run through a number of ways to conduct Latent Growth Models in Twin Data.

Before getting started, please spend a couple of minutes getting to know your team members for today. After you introduce yourself, announce the name of the next person to introduce themselves.

During this time, please choose:

1. Someone who will record the responses in Qualtrics
2. Someone who will share their screen for the Rstudio work (can be the same person)
3. Someone who will keep track of the time (probably good to choose a separate person)

Necessary information for this session:

Copy files via SSH
Open the SSH client from the workshop hub or:

```
https://workshop.colorado.edu/ssh
```

Make sure you are in your home folder by typing:

```
pwd
```

Create a directory to hold today's work by typing:

```
mkdir day6
```

Change your directory to the day6 folder:

```
cd day6
```

Copy over the files/exercises from Elizabeth's directory into yours by typing the following (please note that there IS a period that must be included at the end of the second line):

```
cp /faculty/brad/2022/day6/* .
```

Check to make sure you have the following files (with `ls`)

- `lgcData_1.txt`
- `lgcData_2.txt`
- `lgcData_3.txt`
- `lgcFun.R`
- `lgcPrac.R`
- `lgcPracAnswers.R`

Open RStudio in a separate browser tab.

`https://workshop.colorado.edu/rstudio/`

Once you have opened `lgcPrac.R`**, set YOUR working directory**

```
setwd("~/day6")
```

To help monitor the progress of all of the groups we will need to know your room number.

What is your room number?

Answers

You are now ready to begin the tutorial!

Goal 1: Match the datasets with the appropriate latent growth models.

In many situations, you are going to want to determine which model has the best fit to you data.

Accordingly, your first task is to determine which model best fits each dataset.

You are going to fit 3 models:

- Linear Latent Growth Curve Model
- Quadratic Latent Growth Curve Model
- Semiparametric Latent Growth Curve Model

We have simulated 3 datasets:

- lgcData_1.txt
- lgcData_2.txt
- lgcData_3.txt

To focus on model fitting and interpretation (rather than on coding), all of the OpenMx scripts have been placed in a file called `lgcFun.R`. All you will need to do for this practical is to run the code:

```
# Load the necessary packages
require(MASS)
require(OpenMx)
require(psych)

# load the functions to do the analyses
source("lgcFun.R")
```

and you will load all the necessary packages and have a set of functions to run the three models from today's practical. Naturally, if you wish to edit the functions or explore them in more detail feel free to look inside this file, but know that you do not need to do so in order to complete the practical.

Now, read in the 3 datasets.

```
# Read in the data
lgcData_1 <- as.data.frame(read.table("lgcData_1.txt", header = T))
lgcData_2 <- as.data.frame(read.table("lgcData_2.txt", header = T))
lgcData_3 <- as.data.frame(read.table("lgcData_3.txt", header = T))

# Look at the data
head(lgcData_1)
head(lgcData_2)
head(lgcData_3)
```

You will notice that the data have both MZ and DZ twins in the same dataframe. To make separate dataframes for each dataset:

```
# Create two data objects (one for MZ and another for DZ twin
families)
mzData_1 <- lgcData_1[lgcData_1$zyg == "MZ" ,]
dzData_1 <- lgcData_1[lgcData_1$zyg == "DZ" ,]

mzData_2 <- lgcData_2[lgcData_2$zyg == "MZ" ,]
dzData_2 <- lgcData_2[lgcData_2$zyg == "DZ" ,]

mzData_3 <- lgcData_3[lgcData_3$zyg == "MZ" ,]
dzData_3 <- lgcData_3[lgcData_3$zyg == "DZ" ,]
```

Next we need to tell R what variables we plan on analyzing.
There are 6 waves and two twins.

- Each wave is labels [trial1, trial2, trial3 ...].
- Twins are identified by a suffix _1 or _2 [e.g. "trial1_1" indicates for wave 1 for twin 1 or "trial4_2" indicataes wave 4 for twin 2]

The following code will construct an object that tells R which variables we will be analyzing:

```
selVars <- colnames(lgcData_1)[1:12]
selVars
```

To make it easier to switch between the datasets, we can put the data for dataset 1 into generic MZ and DZ data files (as shown below):

```
dataMZ <- mzData_1
dataDZ <- dzData_1
```

This will allow us to change two lines of code (i.e. change mzData_1 to mzData_2 or mzData_3) without having to change anything in the script below.

Now, run the linear LGC model for dataset 1.

Looking at the model parameters (without looking at any of the other models), please list any parameters that may look suspicious (only spend 30-45 seconds on this answer).

There seem to be a few parameters that are very large and potentially not reasonable. While this may not be a sufficient reason to disregard the model entirely, it is something that we should think carefully about. For example, the estimates of c11 and c22 are quite large and negative, which seems suspicious (but not impossible). Similarly the estimates of cs1 and cs6 are extremely large, which may indicate these time points are not particularly "well predicted" by the model.

Do you think that the linear LGC is going to be the best model for this data?

○ Definitely not

● **Probably not**

○ Might or might not

○ Probably yes

○ Definitely yes

Now, run the Quadratic LGC model for dataset 1.

Looking at the model parameters (without looking at any of the other models), please list any parameters that may look suspicious (only spend 30-45 seconds on this answer).

These parameters look much more reasonable. The means are pretty big, so let's look at how they play out as we go forward.

Do you think that the Quadratic LGC is going to be the best model for this data?

○ Definitely not

○ Probably not

● **Might or might not**

○ Probably yes

○ Definitely yes

Now, run the Semiparametric LGC model for dataset 1.

Looking at the model parameters (without looking at any of the other models), please list any parameters that may look suspicious (only spend 30-45 seconds on this answer).

These parameters look pretty reasonable as well. The a22, c22 and e22 estimates look pretty small, however.

Do you think that the Quadratic LGC is going to be the best model for this data?

○ Definitely not

○ Probably not

● **Might or might not**

○ Probably yes

○ Definitely yes

Now compare the fit of the different models for dataset 1.

Use to following code to extract the various AIC and BIC models that are provided by OpenMx.

```
# Compare the models and determine which model fits best for each
datset.
InfoCriteria <- rbind(
c(t(linACElgcSumm$informationCriteria))      ,
c(t(quadACElgcSumm$informationCriteria))     ,
c(t(semiACElgcSumm$informationCriteria))     )

colnames(InfoCriteria) <- c("df Penalty  AIC", "Parameters Penalty
AIC", "Sample-Size Adjusted AIC", "df Penalty  BIC", "Parameters
Penalty BIC", "Sample-Size Adjusted BIC")
rownames(InfoCriteria) <- c("Linear LGC", "Quadratic LGC",
"Semiparametric LGC")
InfoCriteria
```

Which model provides the best fit to dataset 1?

○ The Linear LGC Model

● **The Quadratic LGC Model**

○ The Semi-parametric LGC Model

While the fit statistics make a compelling case for selecting the best fitting model, it is instructive to plot the expected means for each model to understand why one model fits better than another.

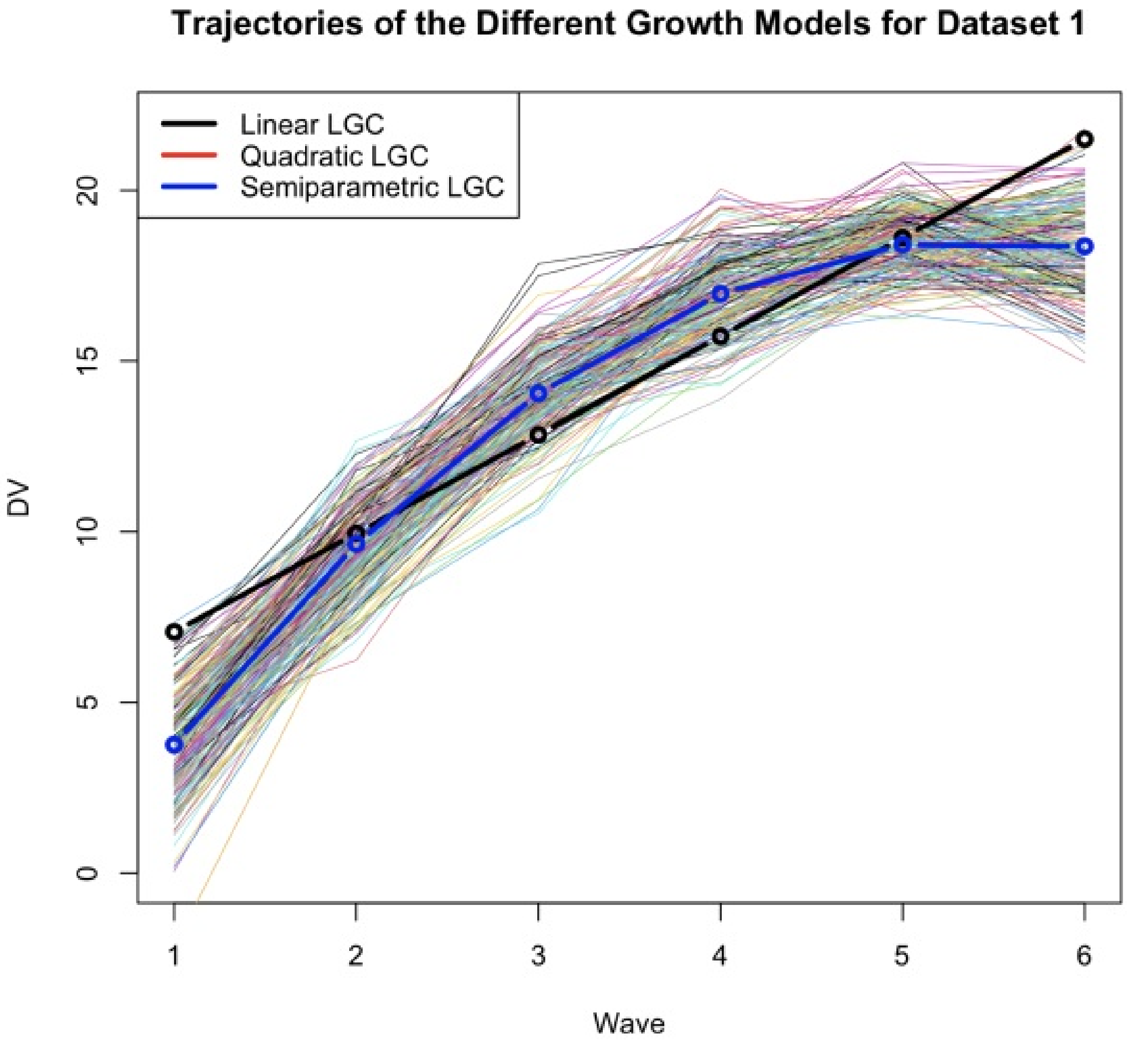
Please conduct the following steps (code to do this is in the practical script):

1. Combine the data from each twin order (e.g. Twin 1/2) and type (MZ/DZ).
2. Make a blank figure (you will probably want to change the ylim argument to better capture the range of the outcomes).
3. Plot a sample of trajectories (200 is ample, but feel free to increase or decrease it if you want to)
4. Overlay the mean trajectories.

What do you notice about the mean trajectories for each model?
(Please write a short explanation)

As we could begin to see by looking at the parameter estimates, the linear results don't match the observe data very well. Also, the expected values for the quadratic and the semi parametric models are almost right on top of each other (making it impossible to distinguish them in the figure). Because the quadratic model fits the data better than the semiparametric model, knowing that the observed means are effectively the same for both models implies that the difference in fit must come from the variance of the latent quadratic factor and its covariance with the linear and intercept factors.

Your figure should look something like this:



Now repeat these steps using dataset 2.

To do this (on lines 35-39), you will need to change lgcData_1 to lgcData_2 like so:

```
selVars <- colnames(lgcData_2)[1:12]
selVars
```

```
dataMZ <- mzData_2
dataDZ <- dzData_2
```

Then you will be able to run the three models for dataset 2.

Looking at the information criteria table, which model fits dataset 2 the best?

● **The Linear LGC Model**

○ The Quadratic LGC Model

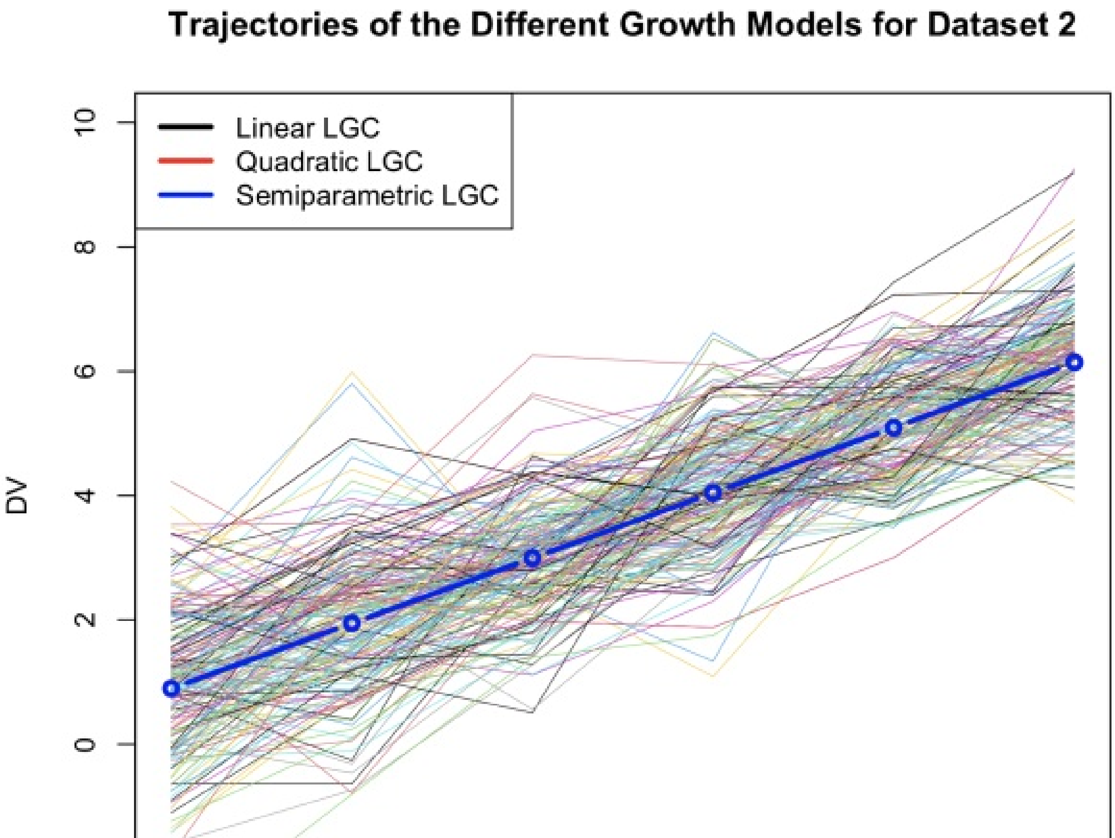
○ The Semi-parametric LGC Model

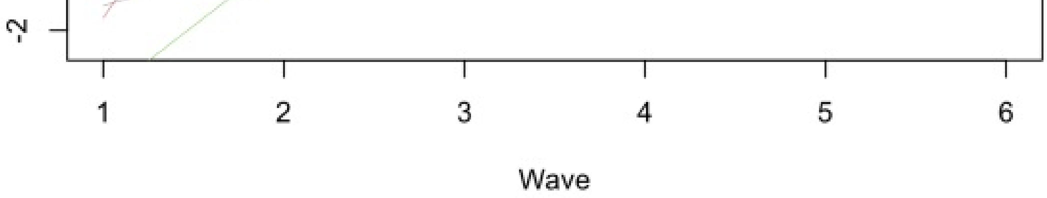
Next, construct a figure of the expected means for data frame 2.

Do you notice anything interesting about the mean trajectories?

All of the mean trajectories are identical. When you look at the quadratic lgc model, you will notice that all of the parameters that are associated with the quadratic term (i.e. means, variances, covariances) are effectively zero. When looking at the factor loadings from the semi parametric model, you can observe a distinct linear trend.

Your figure should look something like this:





Finally, fit the three models to dataframe 3.

Which model fits dataframe 3 the best?

○ The Linear LGC Model

○ The Quadratic LGC Model

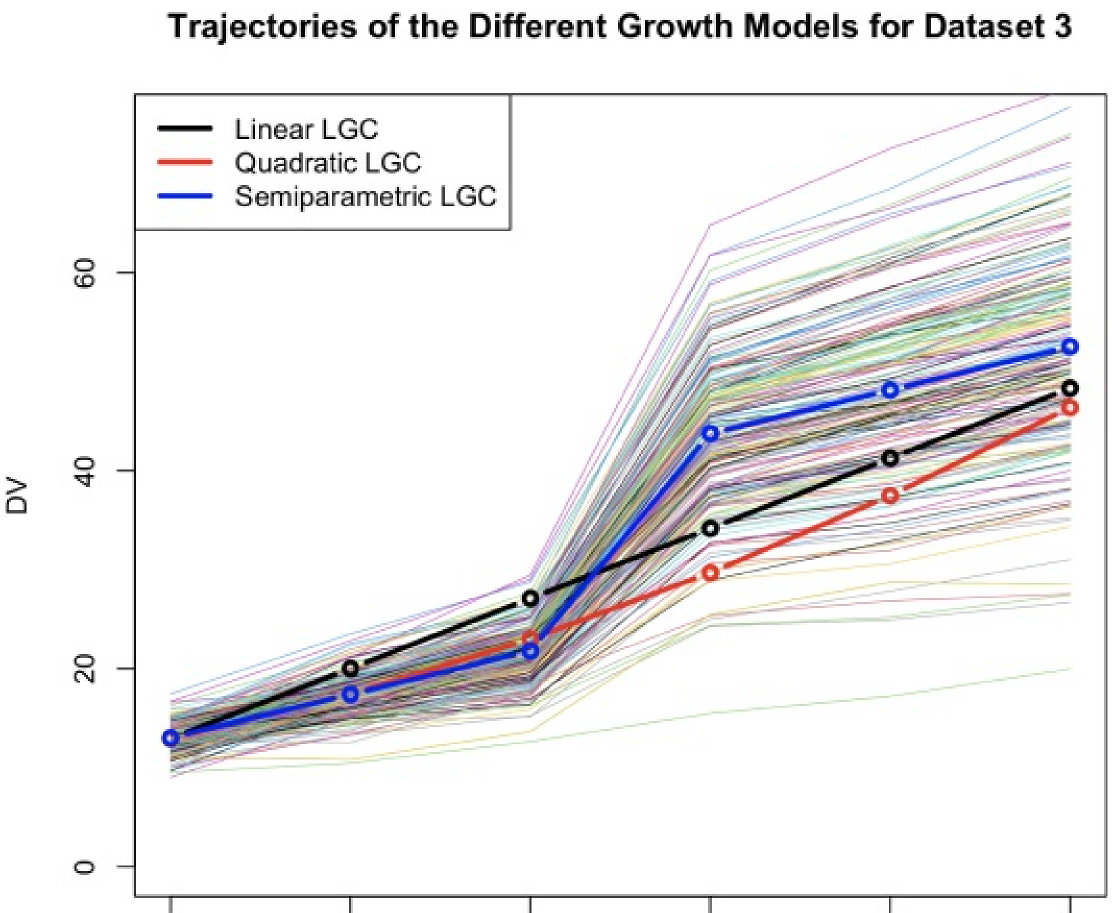
● **The Semi-parametric LGC model**

Next, construct a figure of the expected means for data frame 3.

Do you notice anything interesting about the mean trajectories?

The observed linear and quadratic functions do not come very close to the observed data. Clearly something is going on between wave 3 and 4 that is not captured by the standard models.

Your figure should look something like this:



Excellent Work!

Once you have reached this stage of the practical, please rejoin the main group so that we can have a full discussion of the results.

Goal 2: Interpret the parameters from one of the models

In this section of the practical we are going to take a deep dive into the interpretation of one of the models.

Please indicate which model that you would like to interpret.

- ○ Dataframe 1 - Quadratic LGC Model
- ● **Dataframe 2 - Linear LGC Model**
- ○ Dataframe 3 - Semi-parametric LGC Model

Dataframe 2 - Linear LGC Model

You have chosen to interpret the Linear LGC Model using data from dataset 2.

To ensure that you are running the right model, it is advisable that you re-run the following code:

```
selVars <- colnames(lgcData_2)[1:12]
dataMZ <- mzData_2
dataDZ <- dzData_2

# Run the Linear ACE LGM model
linACElgcModel <- linACElgc(mzData = dataMZ, dzData = dataDZ, selVars
 = selVars, nWaves = 6)
linACElgcFit <- mxRun(linACElgcModel)
linACElgcSumm <- summary(linACElgcFit )
linACElgcSumm
```

First, look at the summary of the fitted model. The summary of the parameters estimates is very useful for providing an overview of the model and other information to assess the quality of the model.

Many people feel that to really understand what is happening in the model, we need to look at the estimates within the context of the specific matrices in detail.

In the boxes below please type in the means of the latent growth factors.

Latent Intercept Mean	8.619999
Latent Linear Mean	4.39

If you would a hint click the hint button below

○ Please give me a hint

What do the means of the latent growth factors tell us? Please write a short interpretation of each latent growth mean.

The mean of the latent intercept (8.61) tells about the average level of the population. This is scaled by the factor loadings so it may not be directly interpretable. If it is multiplied by 0.4082483 (the value of the constant) then you get 3.5191 (which is the mean of the population across all of the waves). The mean of the latent linear factor (4.39) suggests that as the time progresses, we see an increase in the observed behavior by multiplying the linear mean by each wave, we can assess how much change is occurring between waves (On average we would observe a 1.05 difference between waves).

Using the means of the latent growth factors, how can we calculate the expected means for each wave? Please paste the R code for calculating the expected means for each of the six waves in the box below. (Hint it is a single matrix multiplication statement)

```
t(linACElgcFit$output$matrices$lgcLinACE.Mean) %*%
t(linACElgcFit$output$matrices$lgcLinACE.lambda)
```

If you would like to see the algebra, click the hint button below

○ Please give me a hint

What is the COVARIANCE between the latent intercept and the latent linear growth factors?

The hint gives you the following covariance matrix: mxEval(alat + clat + elat, linACElgcFit) [,1] [,2] [1,]

1.548451 -0.579424 [2,] -0.579424 1.008990. Since the question asks for the covariance between the latent intercept and the latent linear growth factor the answer is: -0.58.

If you would like the code to obtain the latent growth covariance matrix, click the hint button below.

● **Please give me a hint**

```
cov2cor(mxEval(alat + clat + elat, linACElgcFit))
```

or

```
linACElgcFit$output$matrices$lgcLinACE.alat +
linACElgcFit$output$matrices$lgcLinACE.clat +
linACElgcFit$output$matrices$lgcLinACE.elat
```

We often interpret the correlations between the latent growth factors rather than the covariances. What is the CORRELATION between the latent intercept and the latent linear growth factors? Please briefly interpret the correlation bewteen the latent intercept and linear growth factors.

cov2cor(mxEval(alat + clat + elat, linACElgcFit)) [,1] [,2] [1,] 1.0000000 -0.4635582 [2,] -0.4635582 1.0000000 The correlation between the intercept and the latent linear growth factor (-0.46) suggest that there is a strong negative relationship between an individuals average level and how rapidly their behavior increases. Thus, the people who start at the highest level appear to increase slower than those who start at a lower level (on average).

If you would like the code to obtain the latent growth correlation matrix, click the hint button below.

● **Please give me a hint**

```
cov2cor(mxEval(alat + clat + elat, linACElgcFit))
```

What proportion of the variance in the latent intercept factor is accounted for by the latent A, C, and E factors?

Proportion accounted for by the Latent A	0.4645128
Proportion accounted for by the Latent C	0.1806472
Proportion accounted for by the Latent E	0.3548400

If you would like the R code to calculate the proportion of variance in the latent growth factors that are accounted for by the latent A, C, and E factors, click the hint button below.

● **Please give me a hint**

```
mxEval( alat /(alat + clat + elat), linACElgcFit)
mxEval( clat /(alat + clat + elat), linACElgcFit)
mxEval( elat /(alat + clat + elat), linACElgcFit)
```

What proportion of the correlation between the latent intercept and latent slope is accounted for by the latent A, C, and E factors?

Proportion accounted for by A	0.4827425
Proportion accounted for by C	0.1551846
Proportion accounted for by E	0.3620729

How would you interpret these proportions of variance?

The results would imply that about 1/2 of the correlation between the latent intercept and the latent slope can be accounted for by additive genetic factors, with approximately 15% and 36% being accounted for by C and E. In many cases the sign of the A, C and E covariances may be different. In such cases, the interpretation of the "coheritabilities" is dubious. As Richard Feynman said: "The first principle is that you must not fool yourself and you are the easiest person to fool."

Approximately what proportion of the phenotypic variance is accounted for by the A, C or E residuals? (a general response collapsing across all 6 specific A's, C's and E's).

A residual	.1 - .25
C residual	.05-.2
E residual	.25-.45

If you would like the R code to calculate the proportion of phenotypic variance that is accounted for by the A, C, and E residuals, click the hint button below.

● **Please give me a hint**

```
mxEval( as /Rt, linACElgcFit)
mxEval( cs /Rt, linACElgcFit)
mxEval( es /Rt, linACElgcFit)
```