This demonstration is for structural equation modelling.

I will not be assessing you on this particular item except at a very conceptual level.

In fact, there is no code challenge for it

because there is a lot more background needed to apply this substantively.

But this is meant so that you would at least have some idea

of what these models are in case you want to follow up.

The package we'll be using is Lavaan.

This is probably the most powerful structural equation modelling package in

R. We'll also use a package called tidySEM,

which stands for structural equation modelling

and it will provide a visualisation function.

This example is adapted from Lavaan tutorial

that you might want to look at in your spare time.

This was data collected at two time points in 1967 and 1971.

It measured the sense of anomie,

which is a condition in which society provides little moral guidance to an individual

and a sense of powerlessness in respondents.

It also measured other demographic information, in particular,

there was a measurement of anomie at two time points and measurement of powerlessness.

Again, at the central time points,

measure of education,

just at one time point.

Education attainment and socioeconomic index,

which is a measure of economic well-being.

Now, the way we're going to load this data is in the form of

a covariance matrix right here.

Then we'll use a function called getCov,

which can be used to load this particular function in a convenient form.

In this form, it takes the string over

the information like this and then it takes the names of each of the variables.

This is what the resulting matrix looks like,

which is exactly what it should look like.

Now, the way we specify the structural equation model is as

in R string and the syntax is similar to our formulas,

but there's certain special operations.

Now remember, structural equation models have latent variables and so we

specify latent variables using this equals tilde notation.

If we want to define more regression like relationship,

then that regression can in fact be with respect to latent variables.

Then we just use tilde like we would in linear model.

Then if you want to say,

but also these two variables should be

correlated in some way over and above other things,

we would put down double tilde like this.

Let's look at the actual model specification.

First of all, we define our output latent variables.

Theta 1, that's alienation 1967.

It's going to be measured by two output observed variables.

Anomie, its latent variable,

measured by anomie and its coefficient is fixed at one and powerlessness,

its coefficient is fixed at 0.833.

Now, these coefficients are fixed for the purposes of this analysis,

but that is not

an exact relationship because there is also an implicit noise term in here.

Same thing with alienation 1967,

a function of anomie

and powerlessness in '71.

Now, we will also have socioeconomic status,

which is going to be a latent variable,

which is a function of education times 1

plus something times economic index.

Now because we don't have a coefficient in front of it,

that means we're not fixing this value,

it's allowed to vary.

Now plus, of course,

noise and this SES is going to serve as an input latent variable

and how so we'll see in a moment.

Then we'll say, we have our latent variables.

We'll say that the alienation '67 is going to be a function

of one's socioeconomic status, also latent variable.

Our alienation in 1971 is going to be function of alienation in '67 times some

coefficient plus the socioeconomic status times some coefficient.

The idea is alienation tomorrow,

which is the function of alienation today plus external factors,

and of course, there's going to be some noise here, again, implicit.

Then we'll also say that anomie in

1967 has the same variance as anomie in 1971.

This is what this notation says.

It's a bit of,

I guess, an abuse of notation,

but we're basically saying that the residual variance of anomie here

is going to be some value Theta 1 and the anomie in '71 is also going to be Theta 1.

They must be equal. The same thing with powerlessness and Theta 2.

Lastly, we will say that anomie

in '67 is going to be correlated with anomie in '71.

That correlation will be Theta 3. We'll call it Theta 3.

We'll also see that the correlation between powerlessness in

'67 and powerlessness in '71,

there's going to be the same correlation as for anomie.

Now, here's the model,

it's pretty straightforward, which is use the function SEM.

We say that we specify a model here, defined here,

we specify the sample covariance as noted earlier

and we specify the number of observations.

We don't need to specify the raw data.

Then we can plot it. Now, I've had to customise this plot nob to make it look good.

If you want to see what each of these does,

I recommend looking in the tutorials for the package,

tidySEM and also just looking at the help.

But here's the basic idea,

here's the reasonable looking plot.

We have socioeconomic status,

which affects alienation in '67 and in

'71 and also education and socioeconomic index.

That's our latent variable.

Now education and SEI,

they are observed variables.

That's why they're in blocks rather than circles.

Then formally based on alienation, in '67,

we observe some anomie '67 and powerlessness in '67 and same thing in '71.

When we think about it,

is that given this model,

so socioeconomic status generates alienation,

alienation generates anomie,

powerlessness and also socioeconomic status generates education and socioeconomic index.

Based on that, given these observed variables at the bottom of this tree,

we can back them out into the original variables,

alienation and socioeconomic status.

Now, there are also coefficients drawn between them

and we can view them in one of two ways.

One, we can, of course, look at this picture and the

other is we can look at the summary.

One thing we can look at immediately is the test for the model,

is there evidence of lack of fit in the model?

This is a chi-square test.

There are nine degrees of freedom and the p-value is not too small, so we're okay.

Now we can look at the coefficients.

Now, the way they're presented here,

we have the various coefficients for the latent variables.

Notice that these are fixed,

so they don't have standard errors or anything like that.

Similarly, it's fixed here,

but for this one, we have a fitted coefficient standard error and a p-value.

We're saying that, yes, in fact,

socioeconomic status does affect or determine the socioeconomic index.

In terms of progressions,

we have our regression coefficients,

standard errors, p-values and so on.

We find a significant relationship on all of these.

Now we can also go back here and look at these in a picture.

The stars indicate a good statistical significance.

Note that for the values that are fixed like this one,

there is no significance.

Now one thing we can interpret here immediately is that

the amount of the socioeconomic status has a negative effect on alienation.

In other words, the higher somebody's socioeconomic status,

the less alienated they are.

The alienation in '67 has a lot of predictive power for alienation in '71.

Socioeconomic status, again, even at that level,

mitigates alienation even after controlling for earlier alienation.

Some covariances we can look at.

There is in fact positive correlation between anomie

in '67 and '71 and same thing with powerlessness.

We can see that here.

Finally, we have the variances.

It's not clear whether it's meaningful to do

hypothesis tests for these, but here they are.

Now on this diagram,

the variance are indicated with these loop to loops.

That's the variance with itself and here are the correlations between these variables.

Now, I think that's all there is I would say about this.

Again, the model seems to be valid.

Here are the summaries again.

That was structural equation models.

Again, I won't be asking you to do anything complicated with them,

so there is no code challenge.

But they are useful to know about and you might find some use in your work,

since they are a very powerful class of models.