# Cheat Sheet: SEM

## Structural Equation Modeling

Measurement and Evaluation of HCC Systems

### Scenario

Use SEM if you want to test the causal structure between experimental manipulations, subjective traits, and observed variables. For this cheat sheet we assume three factors (Fa, Fb, Fc) and two experimental manipulations (varX1 with three levels: u, v and w; and varX2 with two levels s and t).

### Power analysis for SEM

- Conduct the power analyses for the *t* tests, ANOVAs and regressions that you are planning to conduct. Also calculate the minimal number of participants based on the factor model. Take the maximum of these numbers at your minimal *N*.

### Creating dummies or contrasts

- If your experimental manipulations are saved as single categorical variables, you have to recode them as dummies.
- For example, let's say varX1 has values "u", "v" and "w"; and varX2 has values "s" and "t". You can create two dummies for varX1, one for varX2, and two for the interaction:
```
data$v <- data$varX1 == "v"
data$w <- data$varX1 == "w"
data$t <- data$varX2 == "t"
data$vt <- data$v * data$t
data$wt <- data$w *data$t
```
- Alternatively, you can create orthogonal contrasts.
- The first one tests whether "v" and "w" differ from "u":
```
data$uVSvw <- -2/3*(data$varX1 == "u") + 1/3*(data$varX1 == "v") +
1/3*(data$varX1 == "w")
```
- The second one tests whether "w" differs from "v":
```
data$vVSw <- -1/2*(data$varX1 == "v") + 1/2*(data$varX1 == "w")
```
- The third one tests whether "t" differs from "s":
```
data$sVSt <- -1/2*(data$varX2 == "s") + 1/2*(data$varX2 == "t")
```

- The following lines create the interactions of these contrasts:
```
data$uVSvw.sVSt <- data$uVSvw * data$sVSt
data$vVSw.sVSt <- data$vVSw * data$sVSt
```

## Running a MIMIC model

- First run a model with the dummies as predictors for each separate factor:
```
model1 <- ‘
    Fa =~ a1+a3+a4+a5
    Fb =~ b1+b2+b4+b5
    Fc =~ c1+c2+c3+c4+c5
    Fa ~ p1*v+p2*w+t+p3*vt+p4*wt’
```
- Run the model:
```
fit <- sem(model1, data=data, ordered=c(“a1”,”a2”,”a3”,”a4”,”a5”,“b1”,”b2”,
”b3”,”b4”,”b5”,“c1”,”c2”,”c3”,”c4”,”c5”), std.lv=T)
```
- Get the model summary:
```
summary(fit)
```
- To interpret the coefficients, refer to the regression cheat sheet.
- You can conduct an ANOVA test of the interaction effect as follows:
```
lavTestWald(fit, ‘p3==0’; ‘p4==0’)
```
- Remove the interaction effects if they are not significant. You can then conduct an ANOVA test of the main effect as follows:
```
lavTestWald(fit, ‘p1==0’; ‘p2==0’)
```
- Repeat this procedure for Fb and Fc.
- For graphs with error bars, create new dummy variables:
```
data$vs <- (data$varX1 == “v”) * (data$varX2 == “s”)
data$ws <- (data$varX1 == “w”) * (data$varX2 == “s”)
data$ut <- (data$varX1 == “u”) * (data$varX2 == “t”)
data$vt <- (data$varX1 == “v”) * (data$varX2 == “t”)
data$wt <- (data$varX1 == “w”) * (data$varX2 == “t”)
```
- Run the model again, but with the following regression:
```
Fa ~ vs+ws+ut+vt+wt
```
- You can put the coefficients into Excel (the value for condition `us` is zero) to create a graph. The error bars of this graph should be equal to the *SE* of these coefficients.
- Repeat this procedure for Fb and Fc.

## Running a full model

- First construct a theoretical model, based on the effects found in the literature. For example, the literature may show that varX1 -> Fa, varX2 -> Fb, Fa+Fb -> Fc.

- Create a causal order based on these effects, fill in the gaps where needed. The example above does not give us information about the causal order of Fa and Fb, but based on common sense we may argue the following: varX1 and varX2 -> Fa -> Fb -> Fc.
- Construct a saturated model based on this causal order:
```
model1 <- ‘
    Fa =~ a1+a3+a4+a5
    Fb =~ b1+b2+b4+b5
    Fc =~ c1+c2+c3+c4+c5
    Fc ~ Fb+Fa+v+w+t+vt+wt
    Fb ~ Fa+v+w+t+vt+wt
    Fa ~ v+w+t+vt+wt’
```
- Run the model:
```
fit <- sem(model1, data=data, ordered=c(“a1”,”a2”,”a3”,”a4”,”a5”,“b1”,”b2”,
”b3”,”b4”,”b5”,“c1”,”c2”,”c3”,”c4”,”c5”), std.lv=T)
```
- Get the model summary:
```
summary(fit)
```
- Iteratively trim non-significant effects from the model.
  - Start with the least significant and least interesting effects (those that were added for saturation)
  - Work iteratively
  - Manipulations with >2 conditions: remove all dummies at once (if one is significant, keep the others as well)
  - Interaction+main effects: never remove main effect before the interaction effect
  - Check the modification indices as a final step
- Get the model summary of the final model, including the communalities (`rsquare=T`) and the alternative fit meausres (`fit.measures=T`):
```
summary(fit, rsquare=T, fit.measures=T)
```
- Assess the model fit of the final model (see CFA cheat sheet).
- Report the regression $R^2$s for all regressions (they can be found at the end of the `rsquare` results).
- (optional) Conduct ANOVA tests for remaining effects involving manipulations with more than two levels using `lavTestWald`.

## (optional) Getting the total effects

- Label all the regression coefficients in the final model, and define parameters for all paths originating in the experimental conditions, e.g.:
```
model1 <- ‘
    Fa =~ a1+a3+a4+a5
    Fb =~ b1+b2+b4+b5
    Fc =~ c1+c2+c3+c4+c5
```

```
Fc ~ pbc*Fb+pac*Fa
Fb ~ pab*Fa+pvb*v+pwb*w+ptb*t
Fa ~ pva*v+pwa*w+pta*t

v2Fb := pva*pab+pvb
w2Fb := pwa*pab+pwb
t2Fb := pta*pab+ptb

v2Fc := pva*pac + (pva*pab+pvb)*pbc
w2Fc := pwa*pac + (pwa*pab+pwb)*pbc
t2Fc := pta*pac + (pta*pab+ptb)*pbc'
```

- Run this model to get the *b* coefficients, *SE*s, and *p*-values of the total effects. Interpret them as regular regression parameters.

## Reporting

- Report the MIMIC graphs and the final model as a diagram. Furthermore, report on fit and significance as outlined below for the example.
- “We subjected the x factors and the experimental conditions to structural equation modeling, which simultaneously fits the factor measurement model and the structural relations between factors and other variables. The model has a good model fit: chi-square(xxx) = xxx.xxx, p = .xxxx; RMSEA = 0.xxx, 90% CI: [0.xxx, 0.xxx]; CFI = 0.xxx; TLI = 0.xxx.”
- “The model shows that [varX1] and [varX2] each have an independent effect on [Fa] and [Fb], specifically…. [Fa] also has an effect on [Fb], and both [Fa] and [Fb] have an effect on [Fc]. The total effects of [varX1] and [varX2] on [Fb] and [Fc] are significant. Specifically….
- Note: If there is an interaction effect between [varX1] and [varX2], they no longer have an “independent effect”. In that case, try to explain the interaction effect. In many cases, it is best to just refer to a graph of the interaction effect (which you can create in Excel).