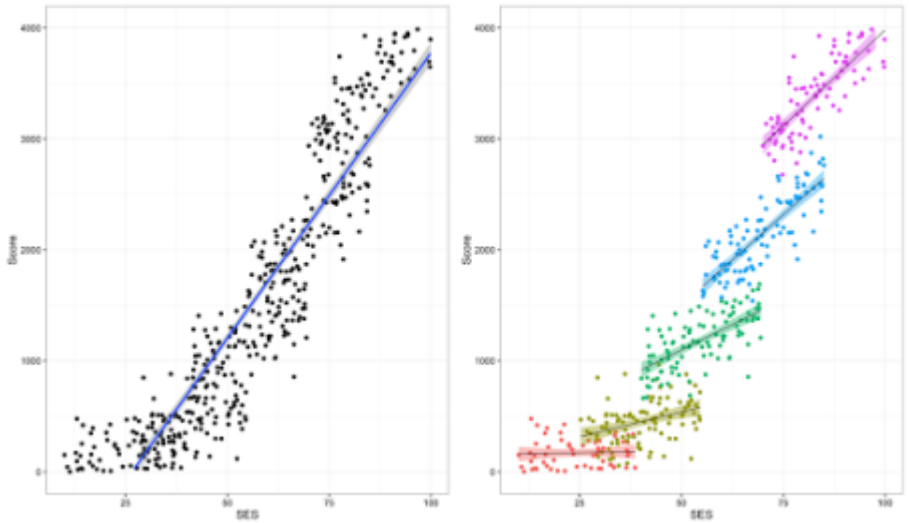
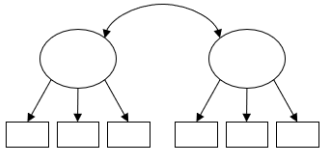
Introduction and Recap

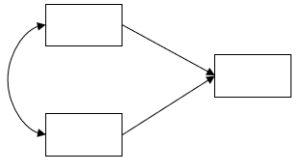


E0430

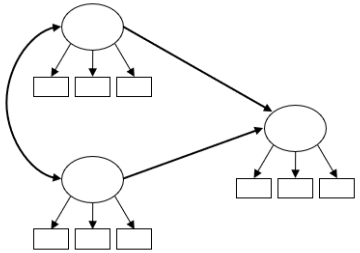
Week 1



Confirmatory Factor Analysis (CFA)



Path Analysis

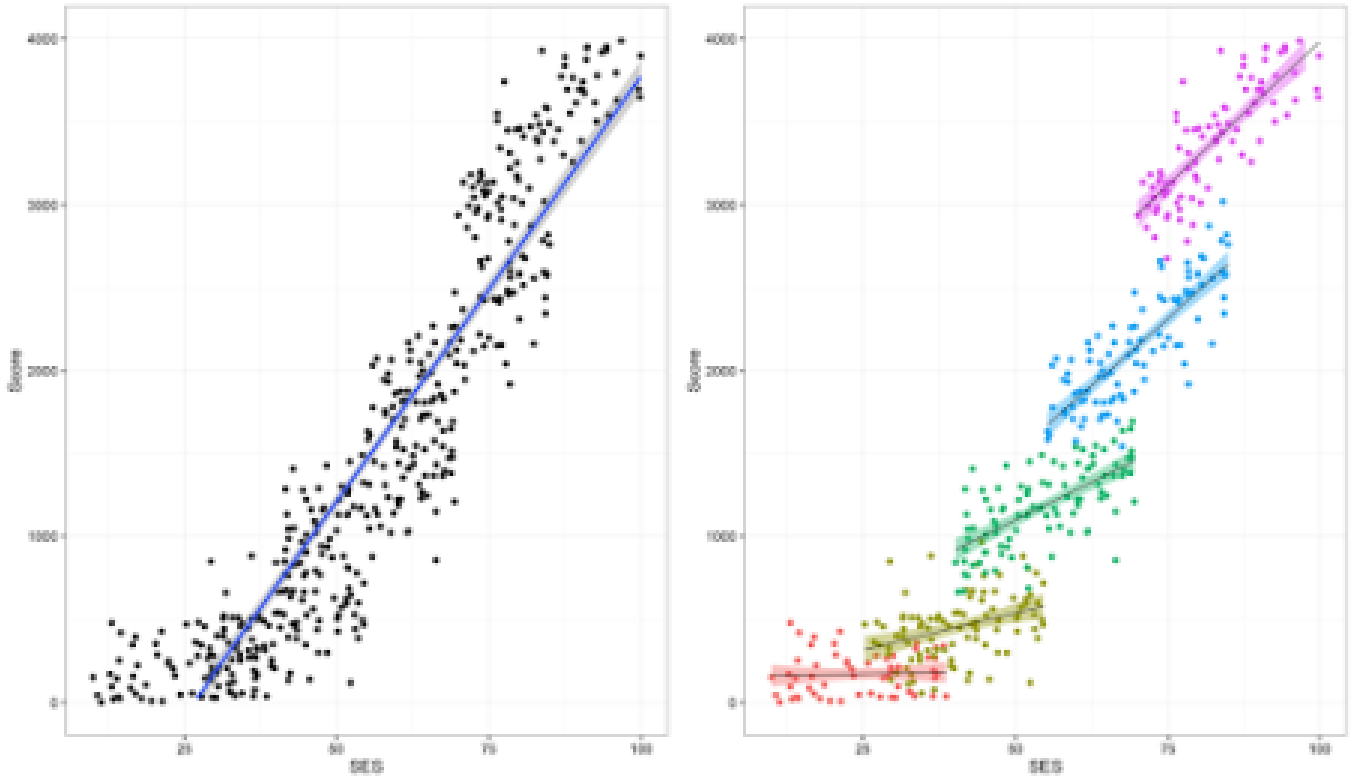


Structural Equation Model (SEM)

HLM approach

- **Hierarchical Linear Modeling (HLM)**
- Synonyms: MLM, mixed-effect or random effects model
- Linear regression that takes into account clustering/nesting of data, i.e., their hierarchical structure
- Clusters = countries, institutions, households, etc.

HLM Example



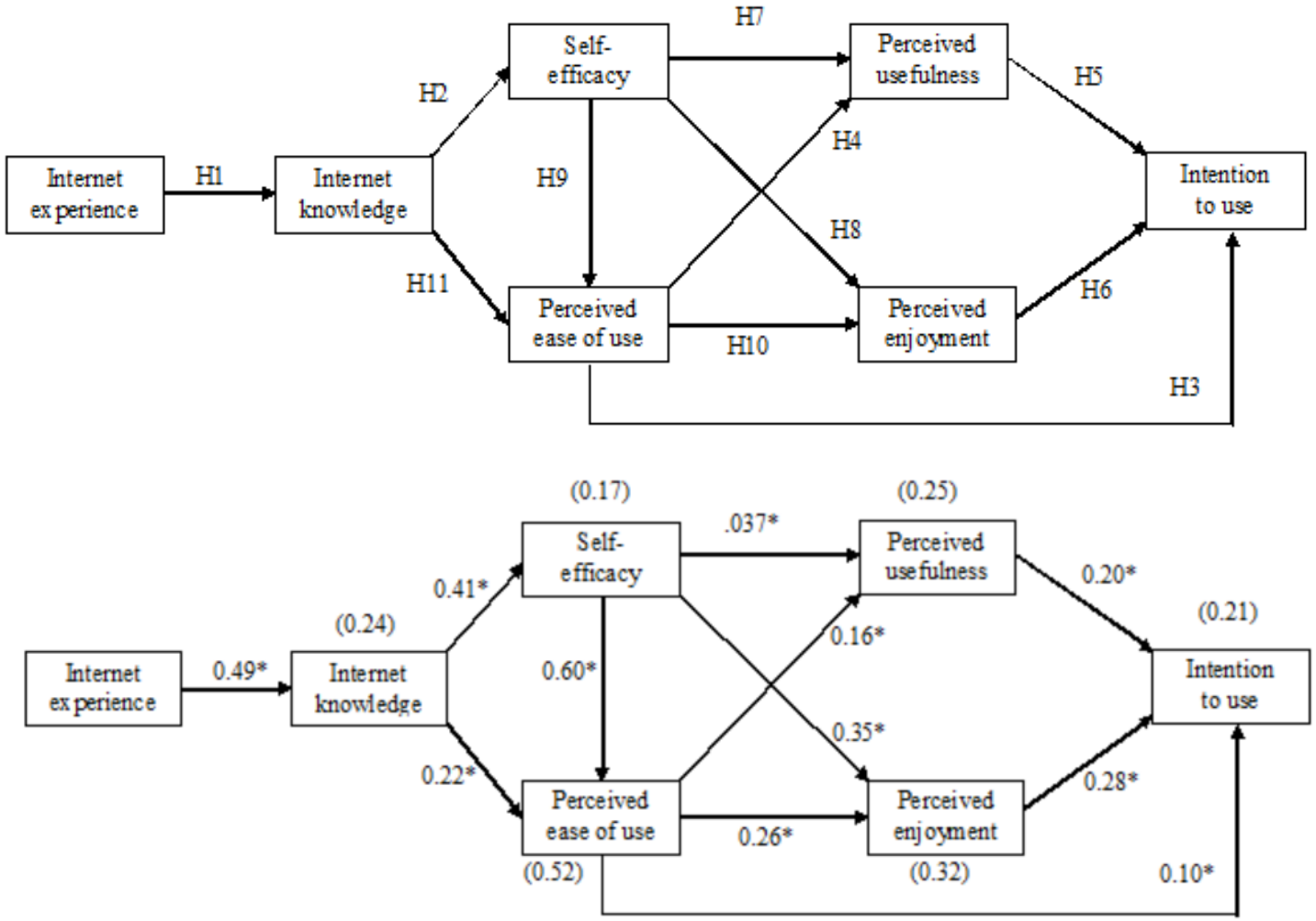
- Positive association between SES and academic achievement overall
- However, the strength and (potentially) significance of the association vary across schools

https://www.r-bloggers.com/2016/10/multilevel-modeling-of-educational-data-using-r-part-1/

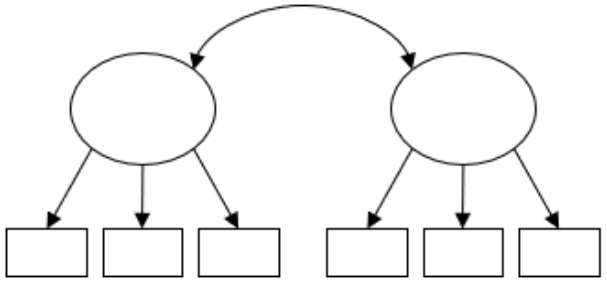
SEM approach

- **Structural Equation Modeling (SEM)**
- Diverse set of methods (path analysis, CFA, SEM, LGM)
- Estimation of complex associations between variables
- Fit between hypothesized model and data
- Modeling latent variables

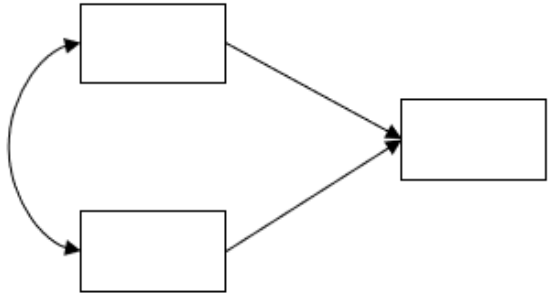
SEM Examples



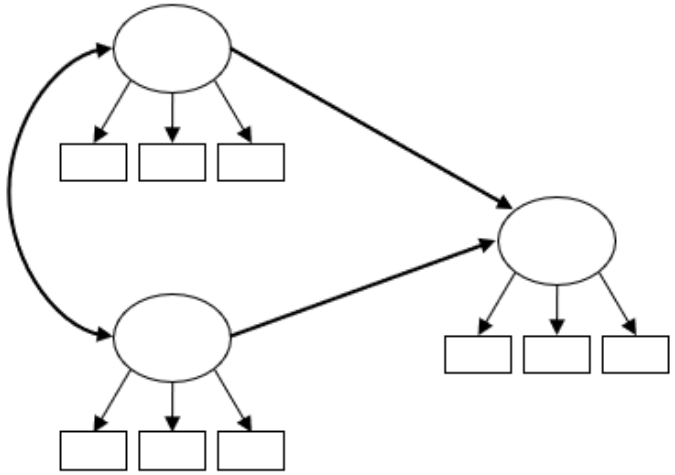
SEM Examples



Confirmatory Factor Analysis (CFA)



Path Analysis



Structural Equation Model (SEM)

Modeling approach

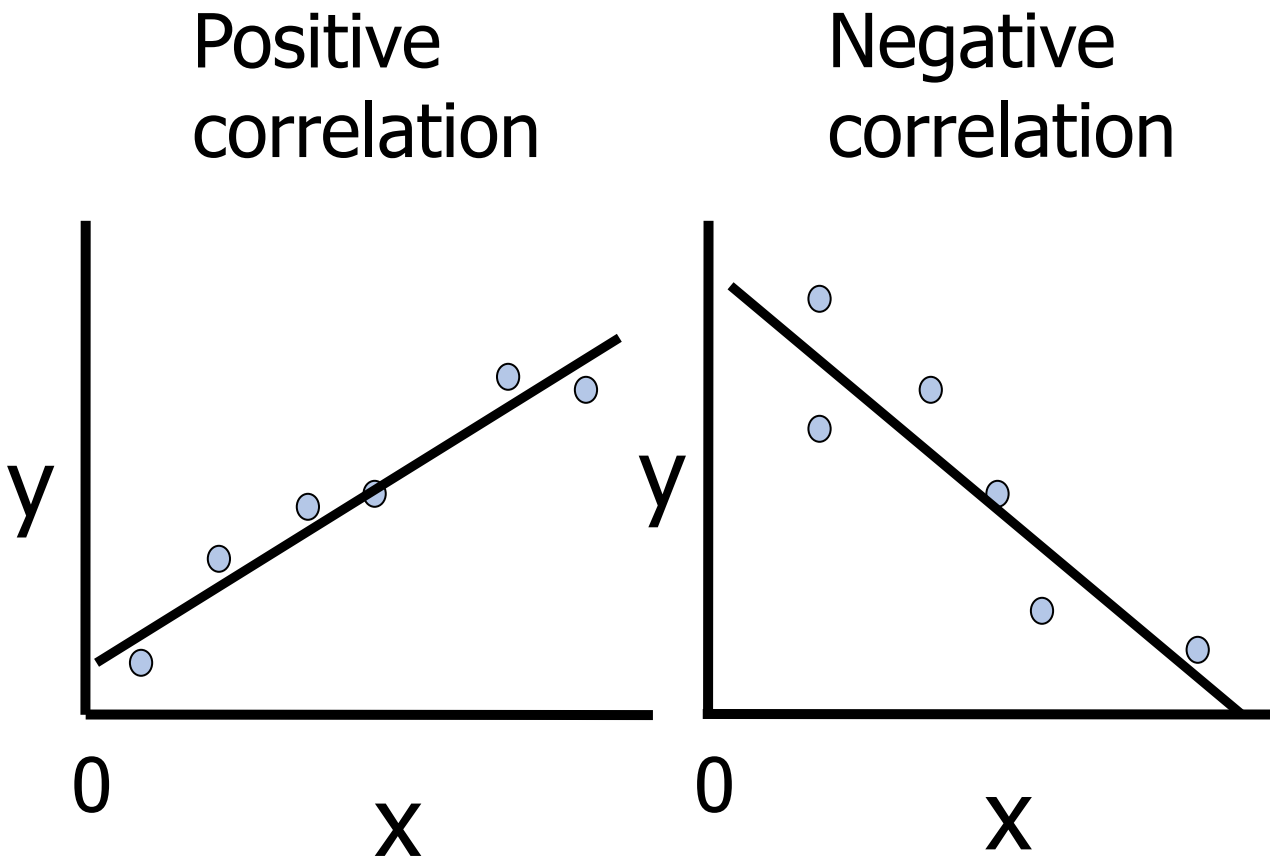
- Theoretical model: hypothesis (or set of hypotheses) about associations between variables
- Theoretical model → Statistical model
- Models are necessarily simplifications and only approximate reality

Common Features of HLM and SEM

- Useful techniques for analysis of large, complex, and/or **longitudinal** data
- Require large sample sizes
- Can be combined (hierarchical SEM)
- Based on covariance/correlation and regression

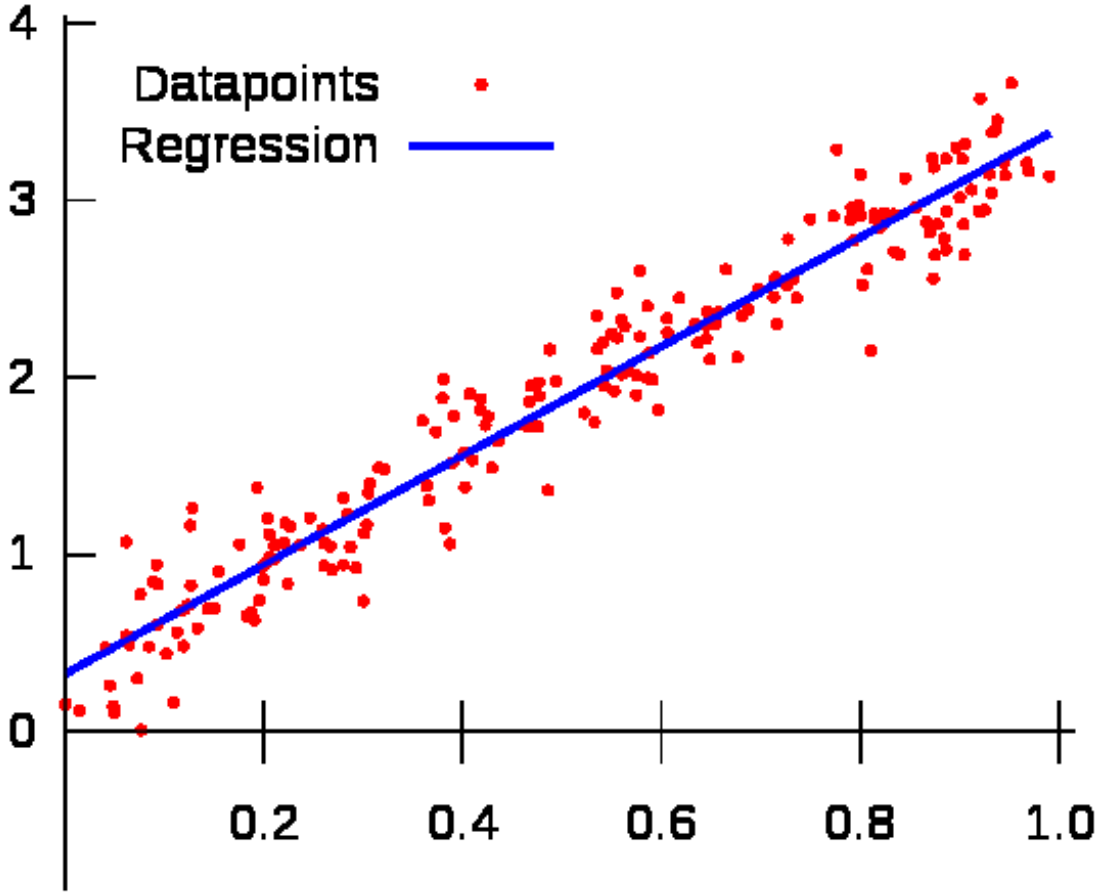
Covariance and Correlation

- Covariance/correlation is a measure of the **joint variability** of two variables
- Linear association between variables
- Covariance = unstandardized
- Correlation = standardized



Regression

- Linear association between variables
- Simple regression (one IV and one DV) is statistically equivalent to correlation between the variables ($r = \beta$)
- Prediction – exposure (IV) and outcome (DV)
- **OLS regression** - minimizing the sum of the squared vertical distances to the data points



Constant/Intercept

Independent Variable

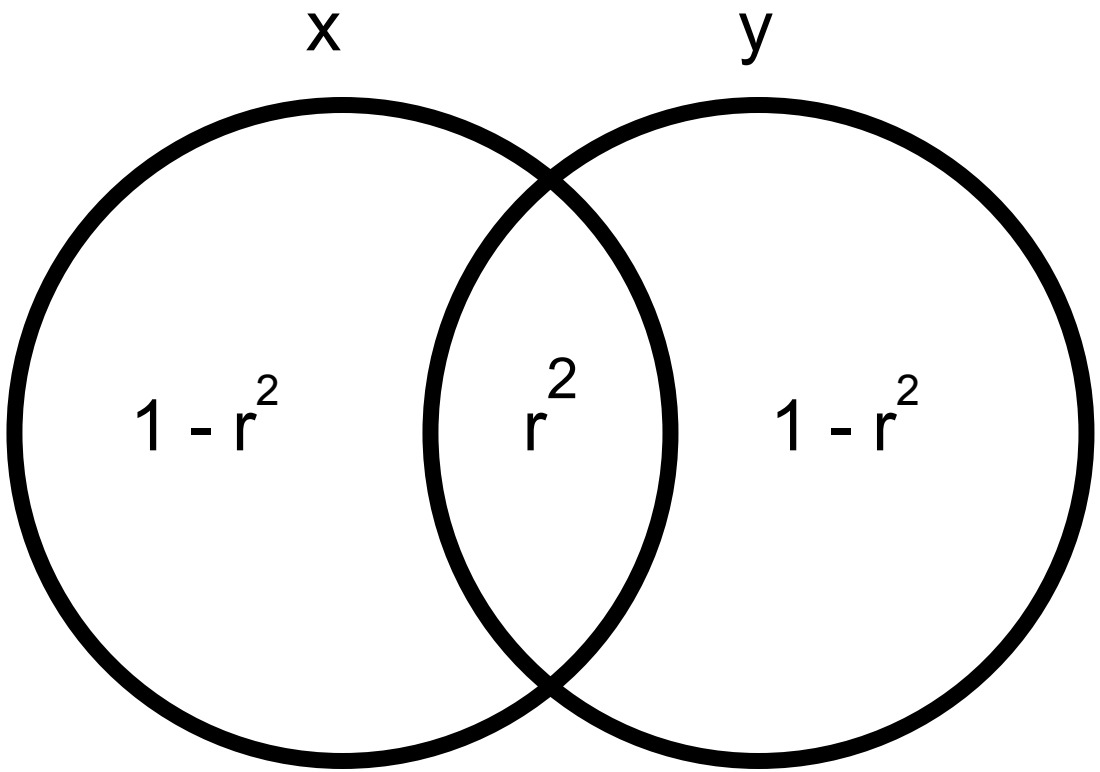
$$Y_i = \beta_0 + \beta_1 X_i$$

Dependent Variable

Slope/Coefficient

Variance explained – R^2

- The total proportion (or %) of the DV variability that is explained by knowing X is called R^2



Practical

1. Open dataset students.sav
2. Test whether there is a statistically significant correlation between **hours spent online** and **hours slept per night**
3. What is the magnitude and significance of the *r* coefficient?
4. Interpret the *r* coefficient
5. Run simple linear regression to test whether **hours spent online predict hours slept per night**
6. What is the magnitude and significance of the regression coefficients (B, β)?
7. What proportion of variance in hours slept per night do hours spent online explain?
8. Summarize the results of the regression analysis in your own words