

Aula 4: Identificar Relações de Covariação

Modelos de Regressão Linear & Multinível

Docente: Amílcar Moreira

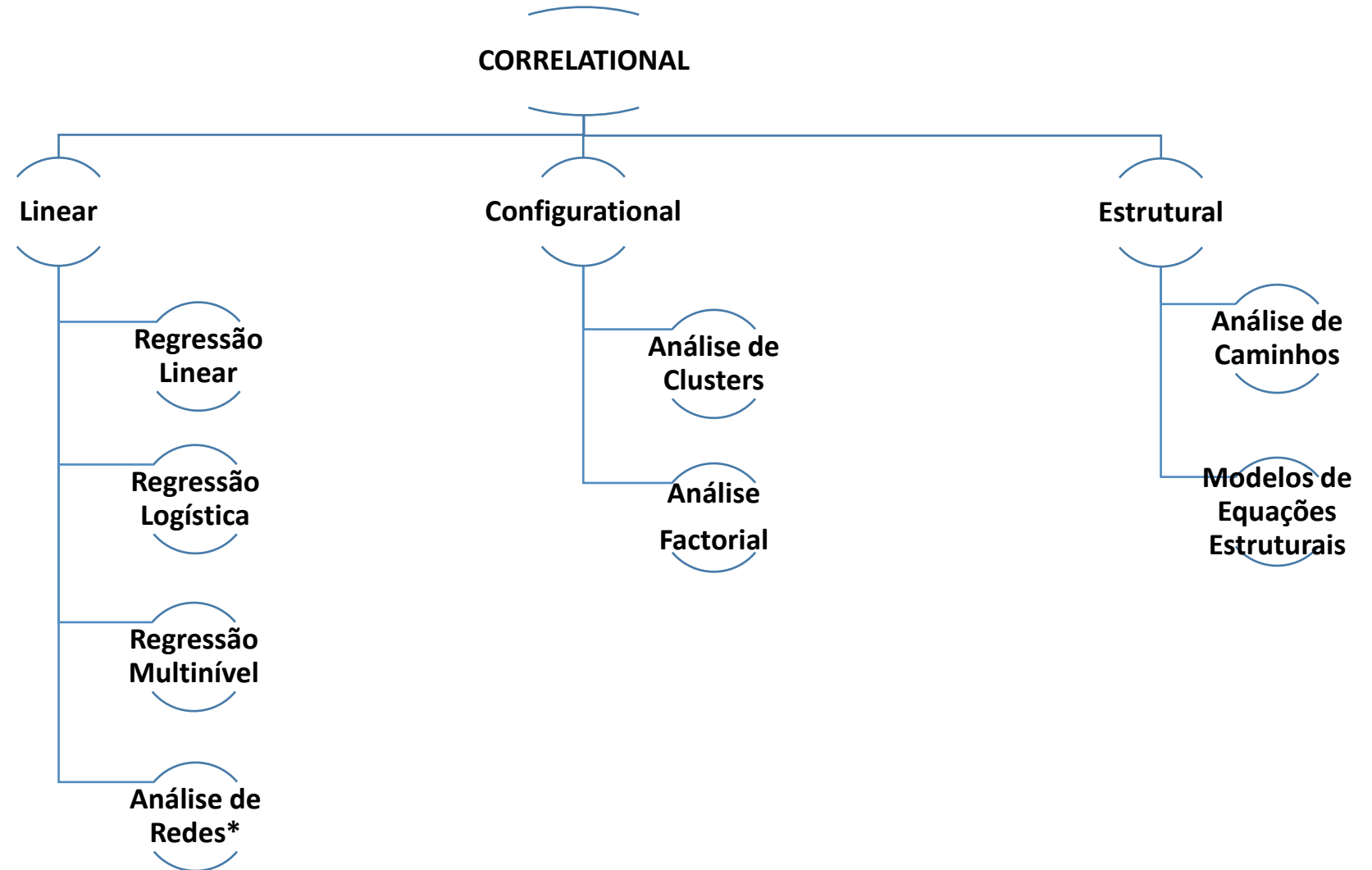
Dia & Hora: 09/10/2023, 20:30-22:30

Sala: Q6, IAPMEI Room

Métodos e Técnicas de Investigação

Aula 5: Identificar Relações de Covariação

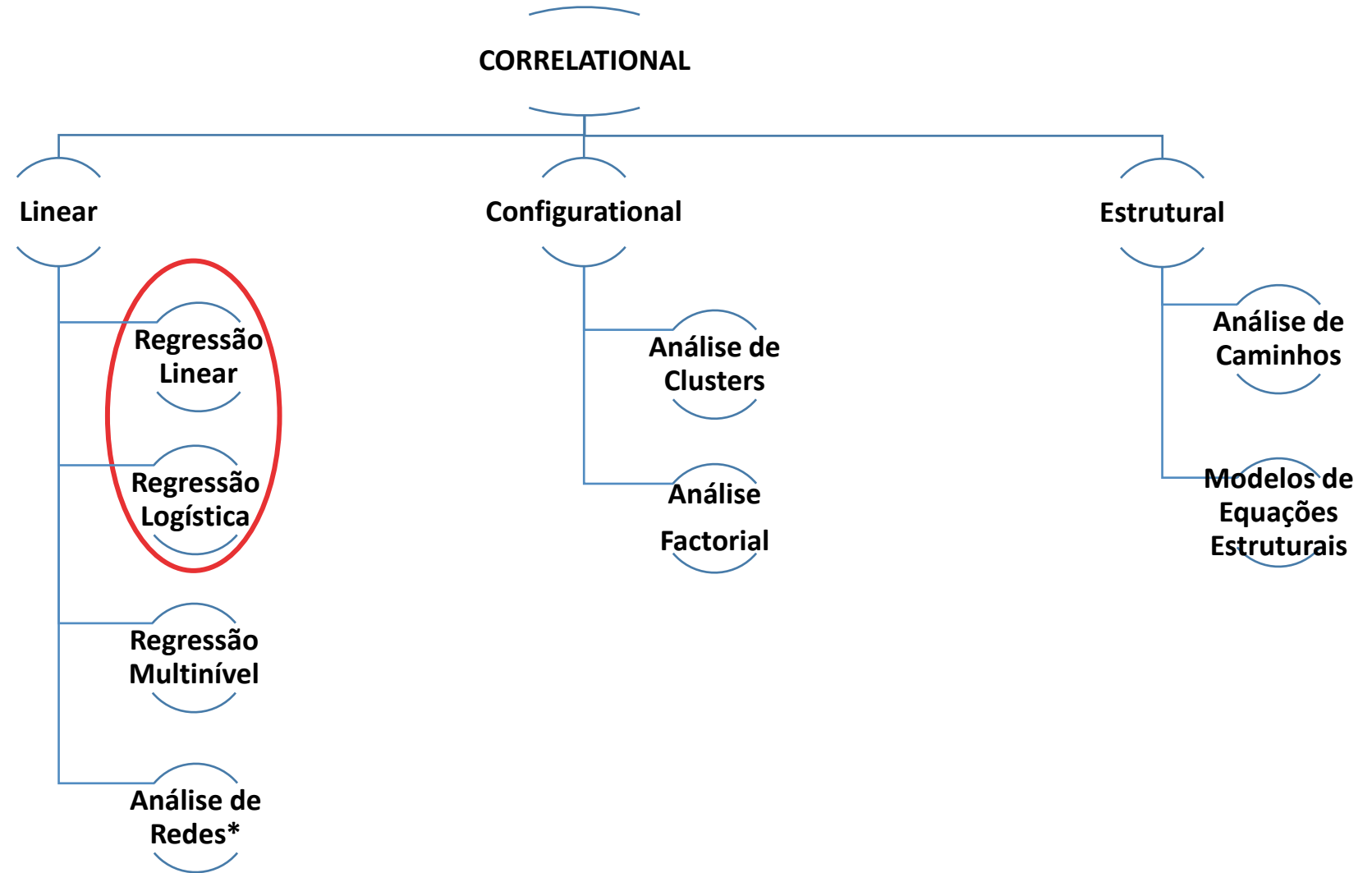
- Nas últimas semanas explorámos diferenças fundamentais no “desenho da investigação” utilizando dados quantitativos.



Métodos e Técnicas de Investigação

Aula 5: Identificar Relações de Covariação

- Nas últimas semanas explorámos diferenças fundamentais no “desenho da investigação” utilizando dados quantitativos.
- Hoje vamos concentrar-nos nas abordagens mais básicas e populares...



Métodos e Técnicas de Investigação

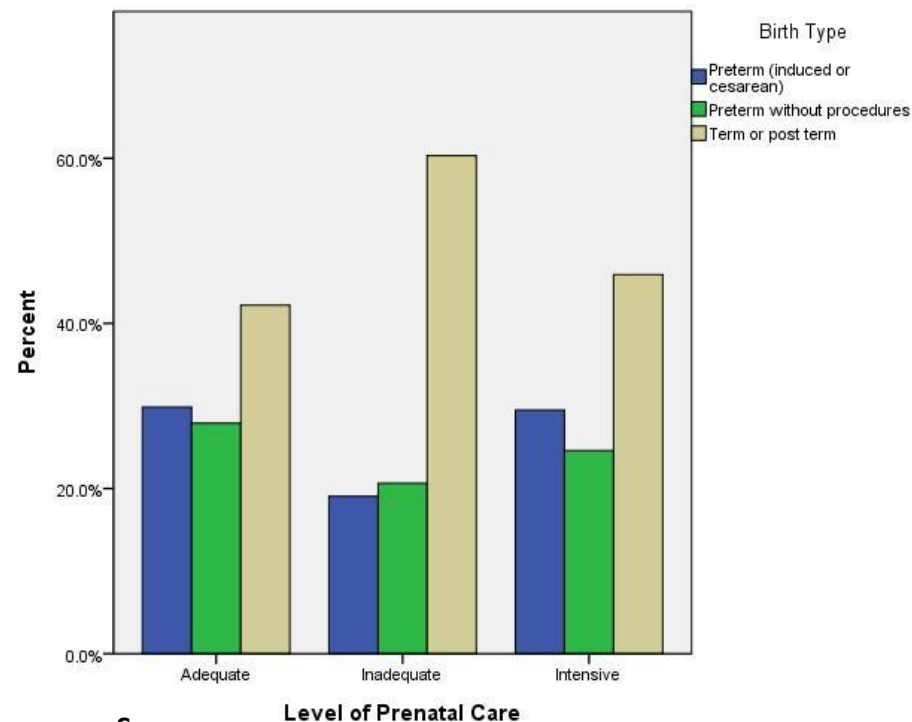
Aula 5: Identificar Relações de Covariação

- **Na investigação social, esforçamo-nos por encontrar relações/padrões de covariação entre variáveis.**

Métodos e Técnicas de Investigação

Aula 5: Identificar Relações de Covariação

- De uma perspetiva mais quantitativa, podemos fazer isso usando uma abordagem mais simplista e bivariada:
 - Podemos utilizar gráficos para avaliar se uma determinada variável influencia a distribuição da nossa variável de interesse;



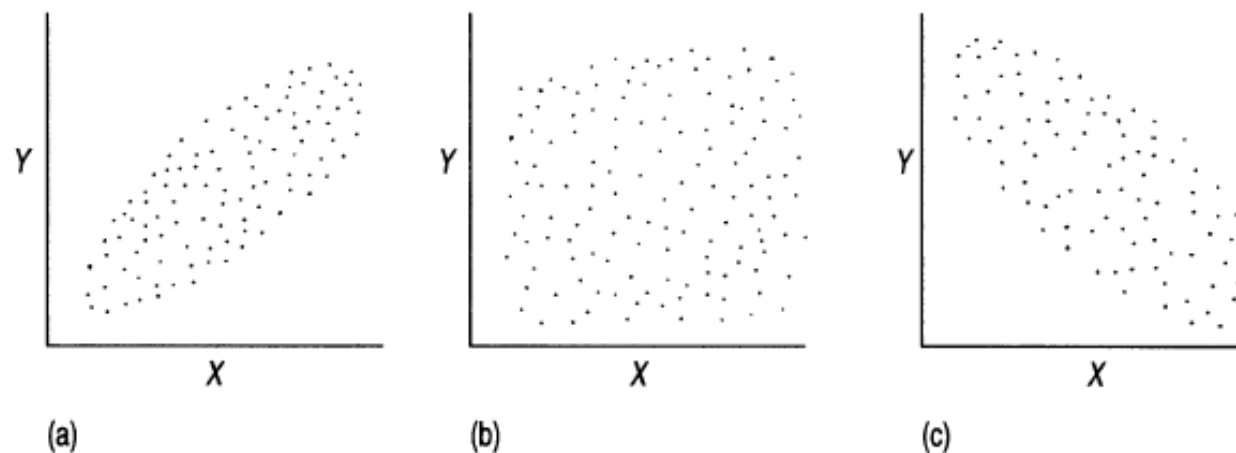
Source:

<http://math150.pbworks.com/w/page/70446470/Clustered%20bar%20graph>

Métodos e Técnicas de Investigação

Aula 5: Identificar Relações de Covariação

- De uma perspetiva mais quantitativa, podemos fazer isso usando uma abordagem mais simplista e bivariada:
 - Podemos utilizar gráficos para avaliar se uma determinada variável varia em função da distribuição da nossa variável de interesse;
 - Podemos também usar gráficos para avaliar se as duas variáveis estão correlacionadas, como e quão forte é essa correlação;



Source: Sarantakos (1998: 392)

Interpretação:

Quando os pontos estão concentrados (a ou c) a correlação é Forte

Quando os pontos estão muito dispersos (b) a correlação é Fraca

Quando os pontos estão inclinados para cima (a) a relação é positiva

Quando os pontos estão inclinados para baixo (c) a relação é negativa

Métodos e Técnicas de Investigação

Aula 5: Identificar Relações de Covariação

- **Uma abordagem mais adequada seria utilizar ferramentas mais sofisticadas...**

- Uma abordagem mais adequada seria utilizar ferramentas mais sofisticadas...

Medidas de Associação / Correlação

- Permitem aferir se existe uma relação sistemática entre duas variáveis.
- Permitem aferir a força da relação sistemática entre duas variáveis.
- Permitem identificar a direção da relação sistemática entre duas variáveis (correlação apenas)

- Uma abordagem mais adequada seria utilizar ferramentas mais sofisticadas...

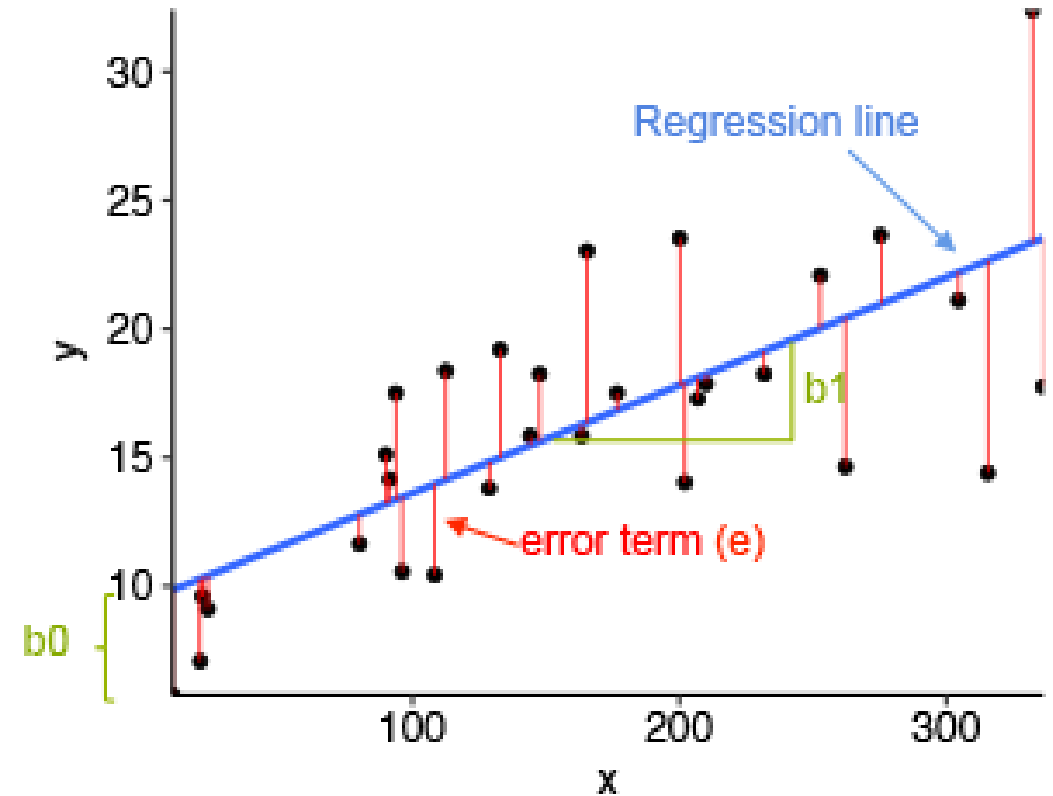
Medidas de Associação / Correlação

- Permitem aferir se existe uma relação sistemática entre duas variáveis.
- Permitem aferir a força da relação sistemática entre duas variáveis.
- Permitem identificar a direção da relação sistemática entre duas variáveis (correlação apenas)

Modelo de Regressão Linear (simples)

- Diferencia entre:
 - Variável Dependente (DV)
 - Variável Independente (IV), que influencia a variável dependente
 - Presume a existência uma relação linear entre as duas variáveis
- O modelo de regressão pode depois ser alargado para incluir de mais do que uma variável independente.

- **Modelo de regressão linear (simples)**
 - A OLS é uma técnica de otimização matemática que procura encontrar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (i.e., dos resíduos)
 - A utilização do OLS como método de estimação vai produzir uma série de medidas que nos vão permitir avaliar o valor explicativo (ou qualidade do ajustamento) do nosso modelo



Resíduo (e) : Diferença entre o valor previsto (i.e., o ponto na linha) e o valor observado

- **Modelo de regressão linear (simples)**

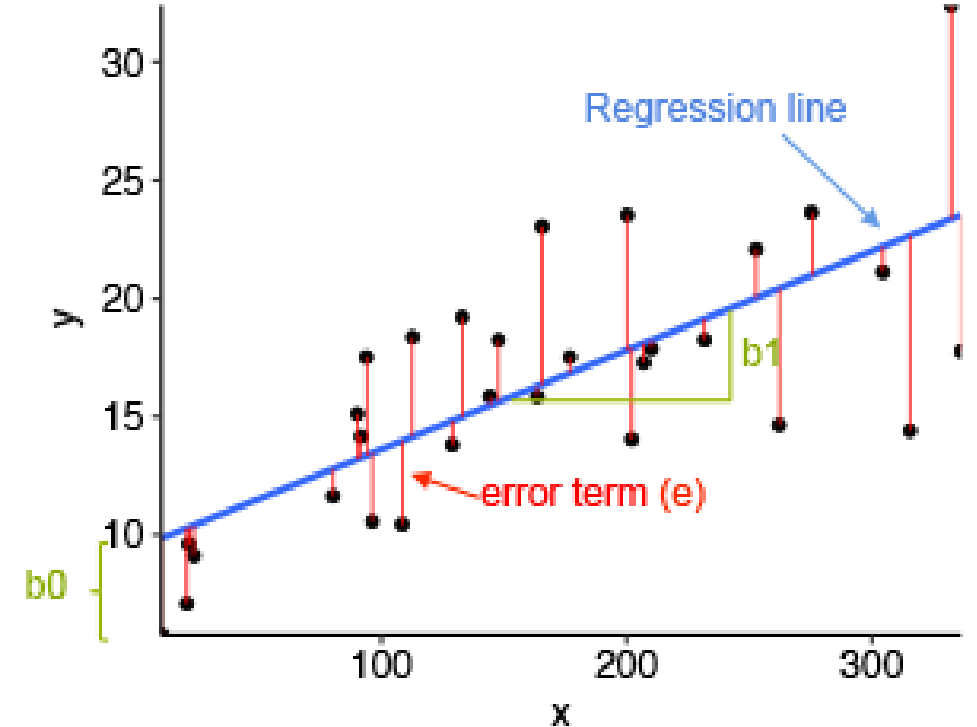
$$Y = \beta_0 + \beta_1 X + E$$

Y Variável Dependente

β_0 Constante (i.e., valor de Y quando $X = 0$)

$\beta_1 X$ Coeficiente Beta da variável X (mede o efeito de uma alteração unitária de X sobre o valor médio da variável Y , quando todas as outras variáveis estão fixas)

E Erro aleatório ou estocástico (reflete a influência de outros factores no no comportamento da variável Y que não podem ser explicadas linearmente pelo comportamento da variável X)



Resíduo (e) : Diferença entre o valor previsto (i.e., o ponto na linha) e o valor observado

A presunção de que o modelo de regressão linear (multivariada) pode estabelecer relações estatisticamente significativas entre variáveis baseia-se na premissa de que um conjunto de suposições é válido.

What are these assumptions?

Linearidade

Independência

Normalidade

Média do **E**rro igual a Zero

Homocedasticidade

E qual é o problema se estes pressupostos não se verificarem?

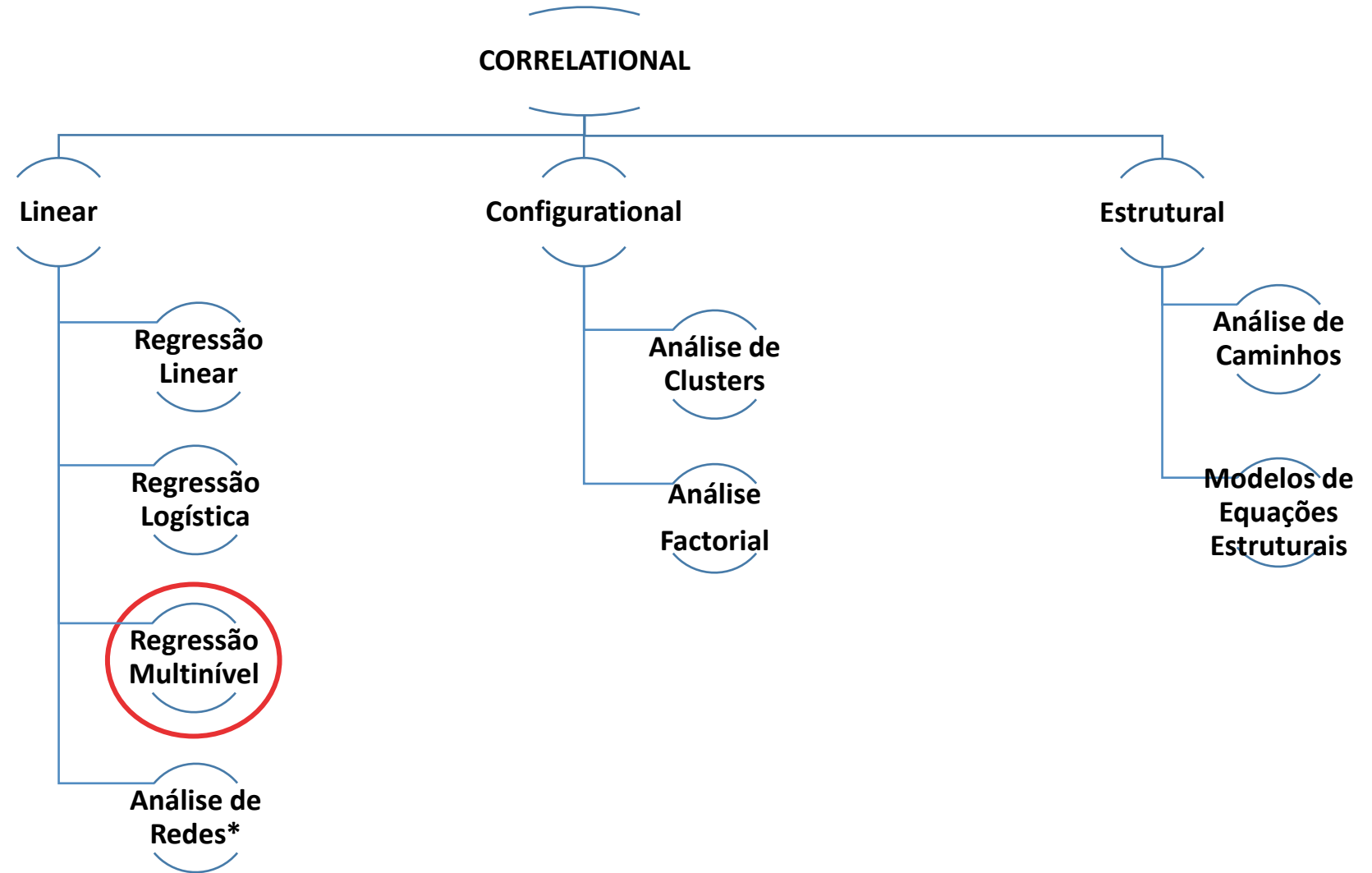
- **Os intervalos de confiança ou os valores de significância podem estar a ser subestimados (i.e. mais pequenos do que na realidade são) ...**

ou seja: estamos a atribuir significância estatística a uma estimativa que na realidade não a terá!

Métodos e Técnicas de Investigação

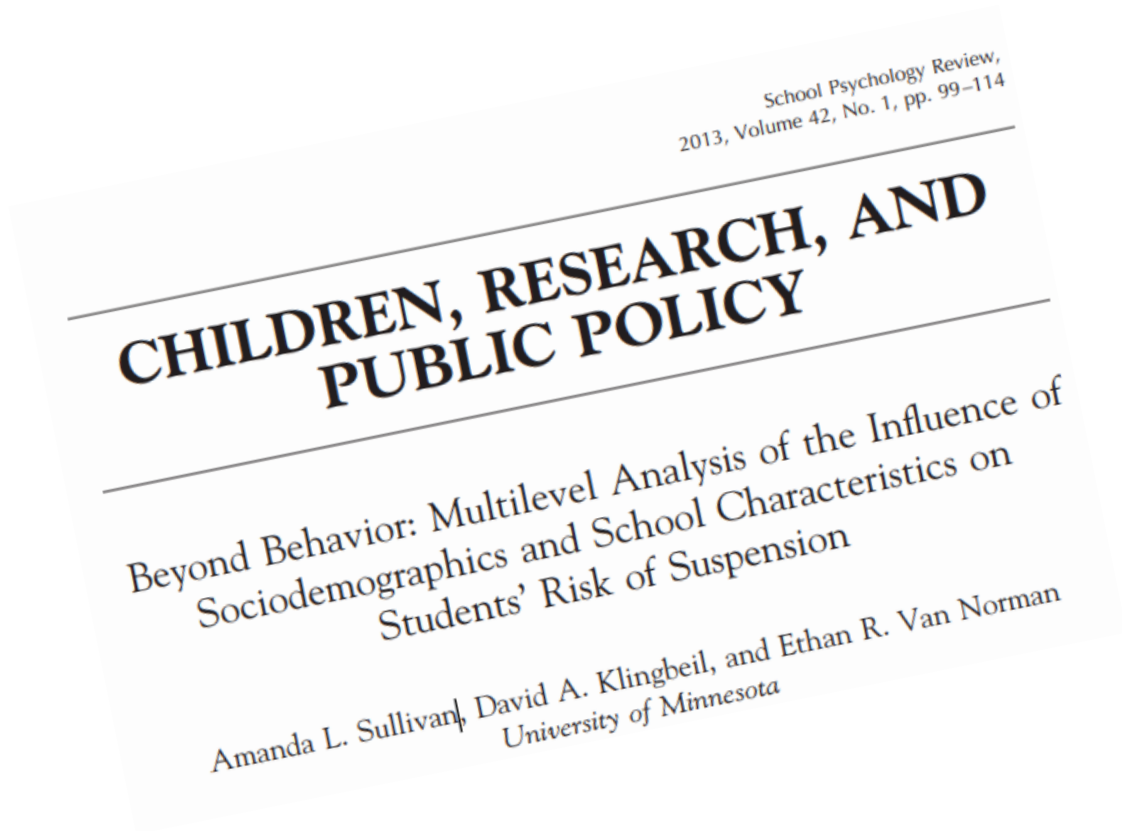
Aula 5: Identificar Relações de Covariação

- Nas últimas semanas explorámos diferenças fundamentais no “desenho da investigação” utilizando dados quantitativos.
- Hoje vamos concentrar-nos nas abordagens mais básicas e populares...



MODELO LINEAR HIERÁRQUICO (HLM)

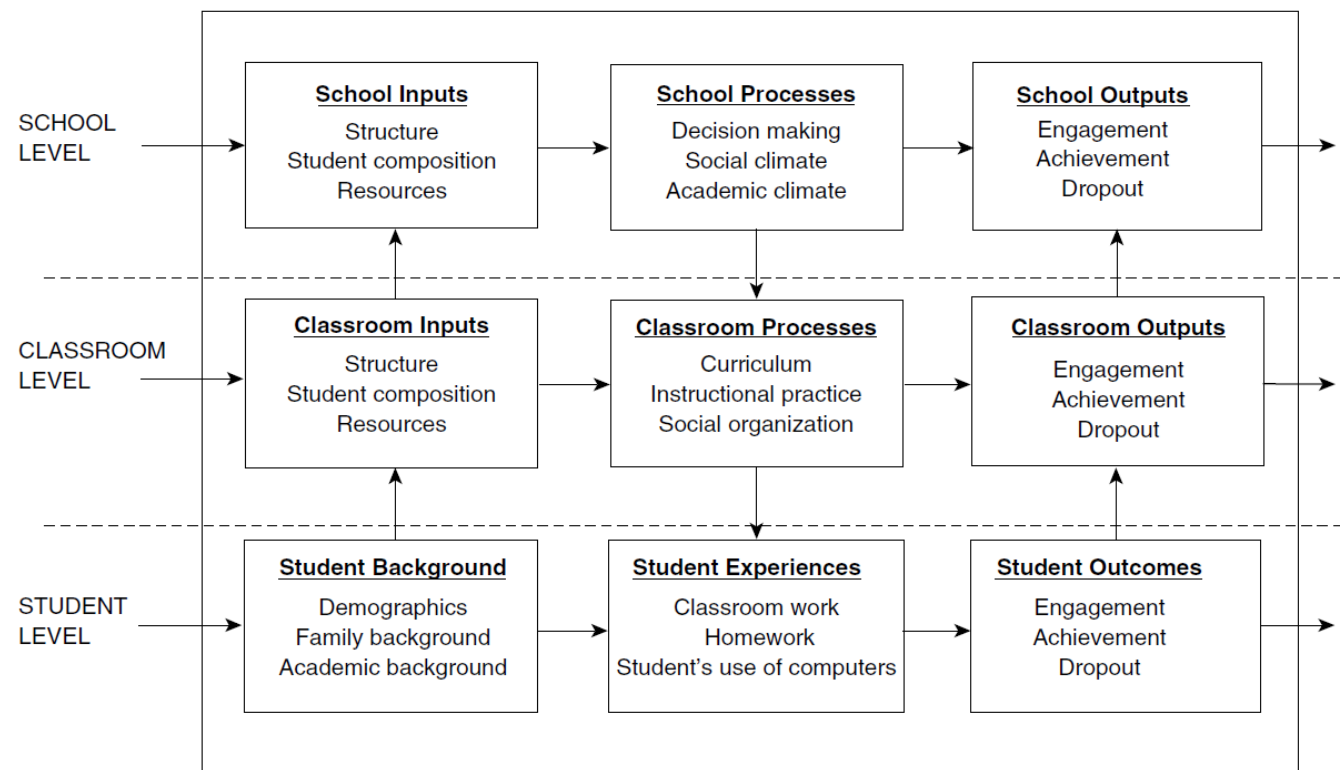
- Extensão do Modelo de Regressão Linear
- Originário das Ciências da Educação e da Psicologia Social



MODELO LINEAR HIERÁRQUICO (HLM)

- Variável dependente definida no nível mais baixo (geralmente o individual)
- Variáveis explicativas em todos os níveis existentes (superior ou inferior)

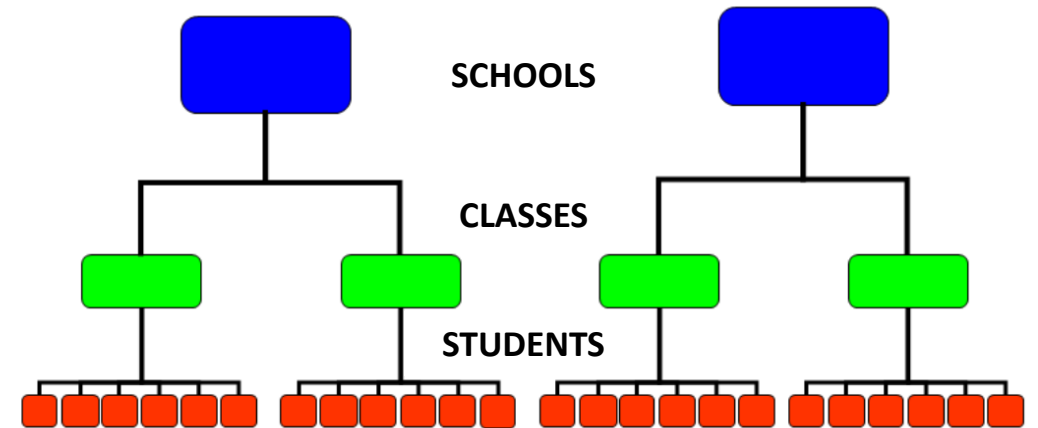
Figure 13.1 A Multilevel Conceptual Framework for Analyzing School Effectiveness



Source: Rumberger & Palardi (2004)

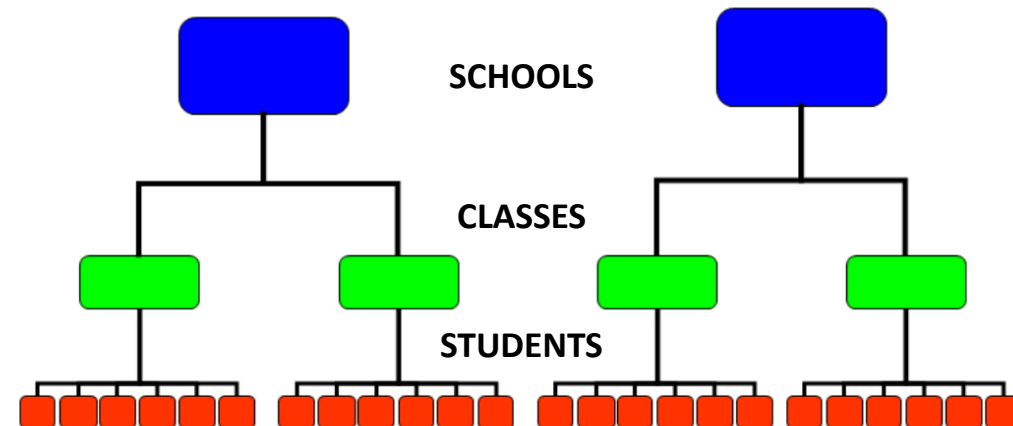
Quando usar este tipo de modelos?

- Quando seus dados têm uma estrutura hierárquica
- Quando os comportamentos/resultados dos indivíduos são influenciados pelo contexto ou grupos aos quais pertencem (pressuposto de dependência)
- Quando você espera que um relacionamento entre variáveis de acordo com o contexto.
- Exemplo:
 - As escolas têm impactos mensuráveis no desempenho dos alunos?
 - Os hospitais têm impactos mensuráveis na prevalência do diabetes?



Por que o modelo de Regressão Linear não consegue lidar com esse tipo de questões?

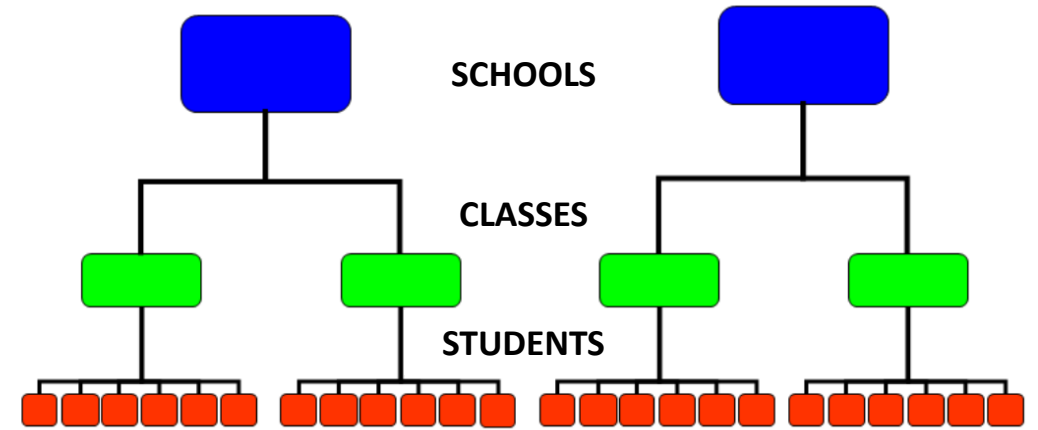
- Este tipo de dados viola duas das principais suposições do Modelo de Regressão Linear
 - Independência de erros: É provável que os indivíduos em grupos partilhem experiências que possam moldar o seu comportamento/resultados. Os erros, portanto, não são independentes;
 - Ausência de multicolinearidade: Admitir a existência de interações entre níveis implica que a autocorrelação está necessariamente presente.



Daí porque o HLM usa o método de Máxima Verossimilhança (Maximum Likelihood method).

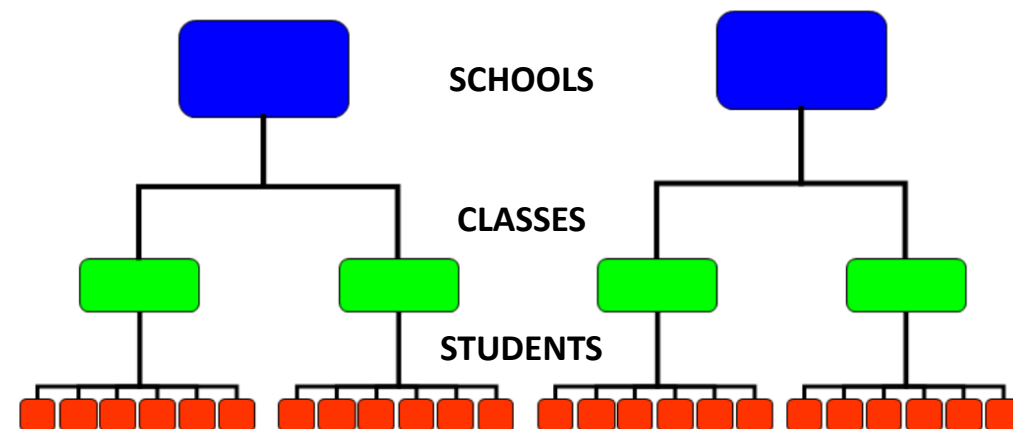
Por que o modelo de Regressão Linear não consegue lidar com esse tipo de questões?

- Se você não tiver cuidado na interpretação dos resultados, poderá cometer a falácia do nível errado, que consiste em analisar os dados em um nível e formular conclusões a outro nível:
 - *Falácia Ecológica* – Fazer inferências para um nível inferior com base em análises realizadas a um nível superior;
 - *Falácia Atomística* – Fazer inferências para um nível superior com base em análises realizadas a um nível inferior.



Questões práticas na implementação do HLM

- **Tamanho da Amostra:** É necessário um número razoável de observações em todos os níveis.
- **Scherbaum and Ferreter (2009)**
 - Aumentar o tamanho da amostra a partir do nível 2 fará mais para aumentar o poder estatístico da amostra do que aumentar a amostra no nível inferior.
- **Regra de 30/30:**
 - Pelo menos 30 unidades de nível superior
 - Pelo menos 30 indivíduos por unidade de nível inferior



Por hoje é tudo...

Até à próxima semana!