

Modelos para Meta-Análise

Revisão Sistemática e Meta-Análise

Marcelo M. Weber & Nicholas A. C. Marino

github.com/nacmarino/maR

Recapitulando

- **Meta-Análise:** "é a análise estatística de uma ampla coleção de resultados de estudos com o propósito de integrar a evidência disponível". (*Glass, 1976*)
- É essencial extrairmos uma métrica de tamanho de efeito e também a sua variância, para que o peso de cada estudo seja proporcional à sua precisão.
- Com estas informações, queremos chegar a um valor que defina a direção e/ou magnitude de um efeito/padrão/processo, a incerteza associada a este valor e fontes que expliquem esta heterogeneidade.

Modelos para Meta-Análise

- Existem diversos tipos de modelo para serem usados, cuja escolha:
 - **Reflete o objetivo da meta-análise:** estimar um efeito vs explorar heterogeneidade;
 - **Define como as estimativas de cada estudo serão combinadas:** todos os estudos são iguais vs estudos diferem em sua 'qualidade';
 - **Descreve as fontes de heterogeneidade entre estudos:** erro amostral vs variação entre estudos;
 - **Determina o tipo de inferência que pode ser feita:** população de estudos analisada vs todos os estudos, até os desconhecidos.
- Desta forma, a escolha do modelo reflete o(s) objetivo(s) da sua meta-análise;
- No geral, os modelos de meta-análise se dividem em dois grandes grupos: *fixed-effects models* e *random-effects models*.

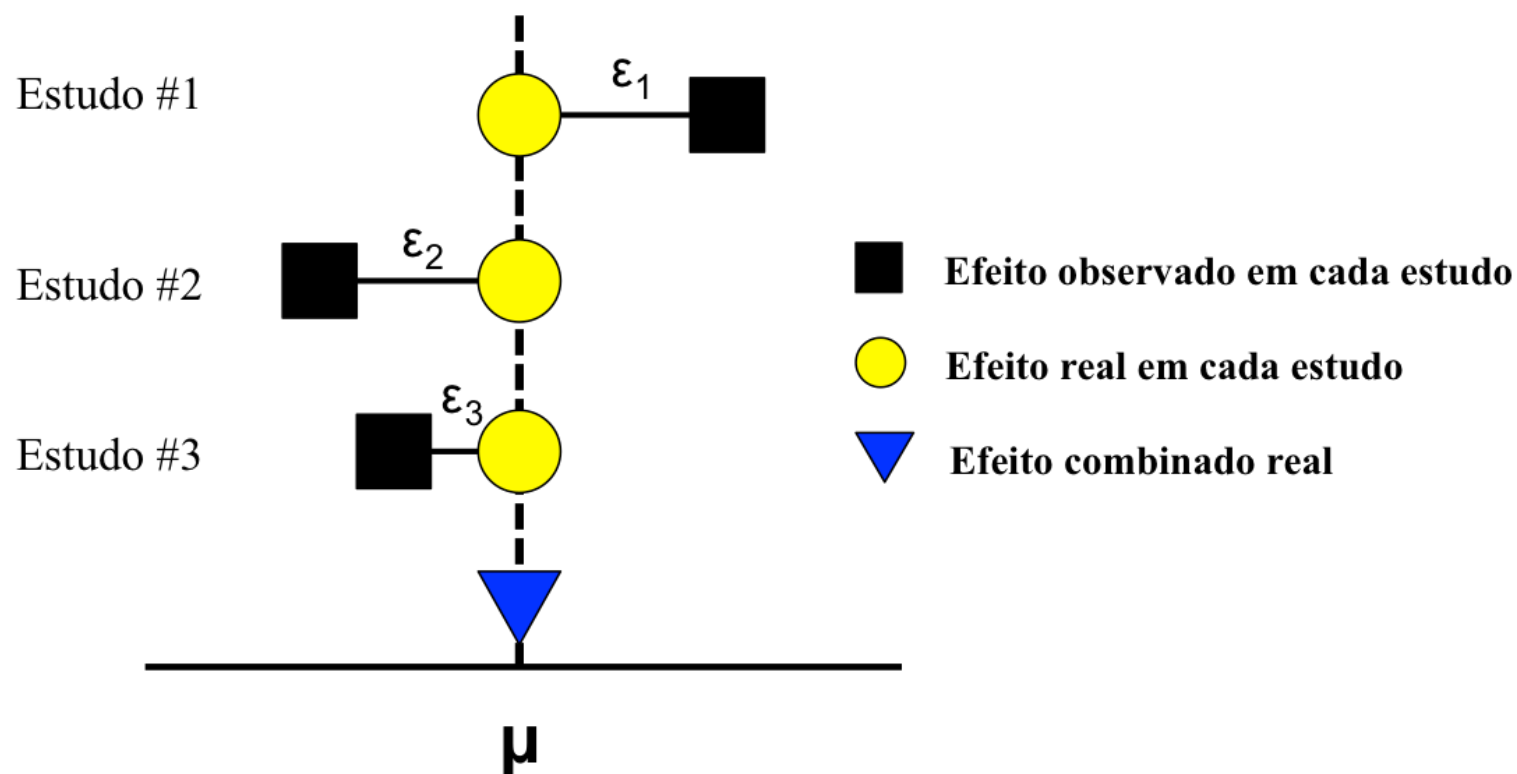
Fixed-effects model

- Usado para estimar a natureza de um efeito;
- Baseado no pressuposto de normalidade;
- Em um fixed-effects model, nós assumimos que:
 - Todos os estudos pertencem a uma mesma população e estão medindo a mesma coisa.
 - A variabilidade existente é explicada unicamente por erro na amostragem em cada estudo;
 - Varibilidade causada por diferenças entre estudos é ínfima ou inexistente.

Fixed-effects model

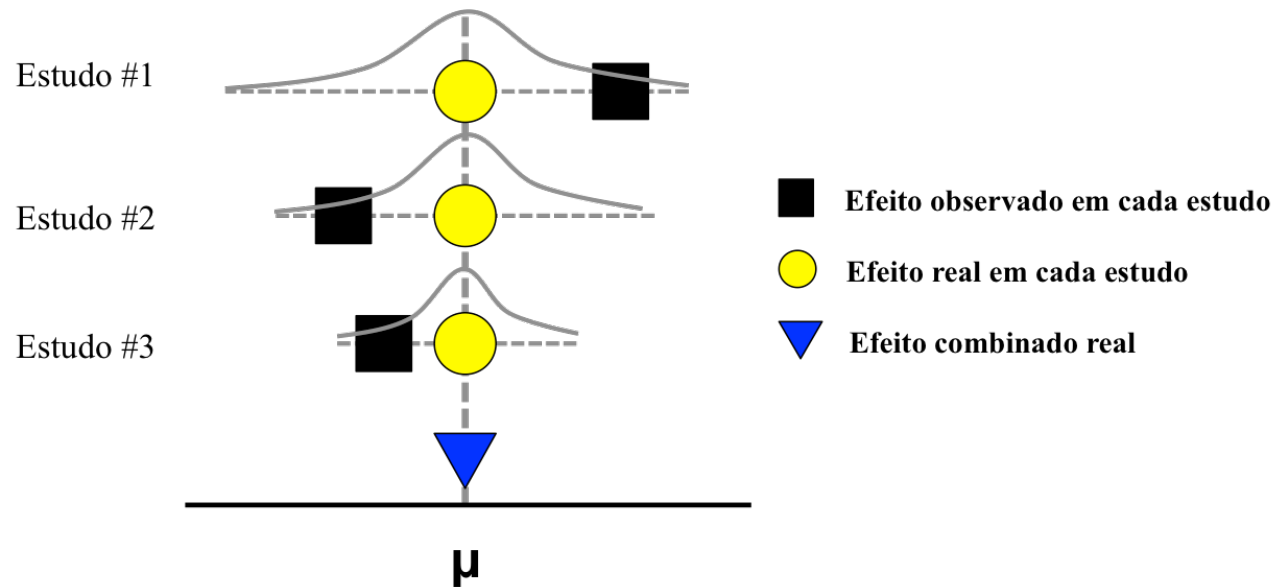
- Exemplo de fixed-effects model:

$$\mu = T_i + \epsilon_i$$



Fixed-effects model

- A incerteza na estimativa de cada estudo está contida dentro da variabilidade do efeito comum que estes estão medindo.
- Generalizações são restritas aos estudos incluídos na meta-análise.



Fixed-effects models

- Cada estudo então contribui com uma estimativa do efeito real, que deve ser ponderado pela sua imprecisão ao medir este efeito.
- No geral, o peso de cada estudo (w_i) em uma meta-análise é proporcional ao inverso de sua variância (s_i): quanto maior variância de um estudo, menor o seu peso ao estimar o efeito real.

$$w_i = \frac{1}{s_i}$$

- Neste sentido, o modelo de meta-análise é semelhante à uma *weighted regression*: o efeito real é calculado com base em cada medida de effect size ($ES_i: d_i, LRR_i, r_i, \dots$) e o peso associado à ele (w_i).

$$\mu = \frac{\sum ES_i \cdot w_i}{\sum w_i} \text{ (onde } \mu, \text{ é o efeito comum entre todos os estudos)}$$

Fixed-effects models

- A variância em torno de T (V_T) é dada por:

$$V_T = \frac{1}{\sum w_i}$$

- Com isto, é possível calcular o intervalo de confiança ao redor de μ :

$$95\% \text{ CI} = \mu \pm 1.96 \sqrt{V_T}$$

- Também podemos testes de significância com o valor estimado de μ (que segue uma distribuição normal):

$$Z = \frac{\mu}{\sqrt{V_T}}$$

No R

- Um fixed-effects model (modelo de efeito fixo) pode ser implementado utilizando o argumento `method = "FE"` na função `rma`, disponível no pacote `metafor`.

```
## carregando dados: efeito da riqueza de especies na biomassa total da comunidade
dados <- read.csv(file = "../98 - dados para exemplos/medias.csv")
```

```
## carregando o metafor
library(metafor)
```

```
## calculando a medida do tamanho do efeito Hedge's d
### Qual o efeito da riqueza de especies na biomassa total da comunidade?
dados <- escalc(measure = "SMD", # opcao de metrica do tamanho do efeito
               mli = media_tratamento, sdli = sd_tratamento, nli = n_tratamento, # infos do tratamento
               m2i = media_controle, sd2i = sd_controle, n2i = n_controle, # infos do controle
               data = dados) # conjunto de dados de onde extrair as informacoes acima
```

```
##      estudo      autor      yi      vi
## 1         11      Maria -1.3334 0.2716
## 2          3  Antonio -0.3242 0.2251
## 3         10      Marcos -0.7577 0.2382
## 4          4  Fernanda -0.8046 0.3088
## 5          6        Joao -0.2034 0.2872
## 6          9      Marcia -0.6654 0.3015
```

No R

- Por hora, vamos focar apenas na interpretação do *Model Results*.

```
rma(yi = yi, vi = vi, method = "FE", data = dados)

##
## Fixed-Effects Model (k = 16)
##
## Test for Heterogeneity:
## Q(df = 15) = 64.3755, p-val < .0001
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
##    0.4294    0.1500    2.8625    0.0042    0.1354    0.7235    **
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

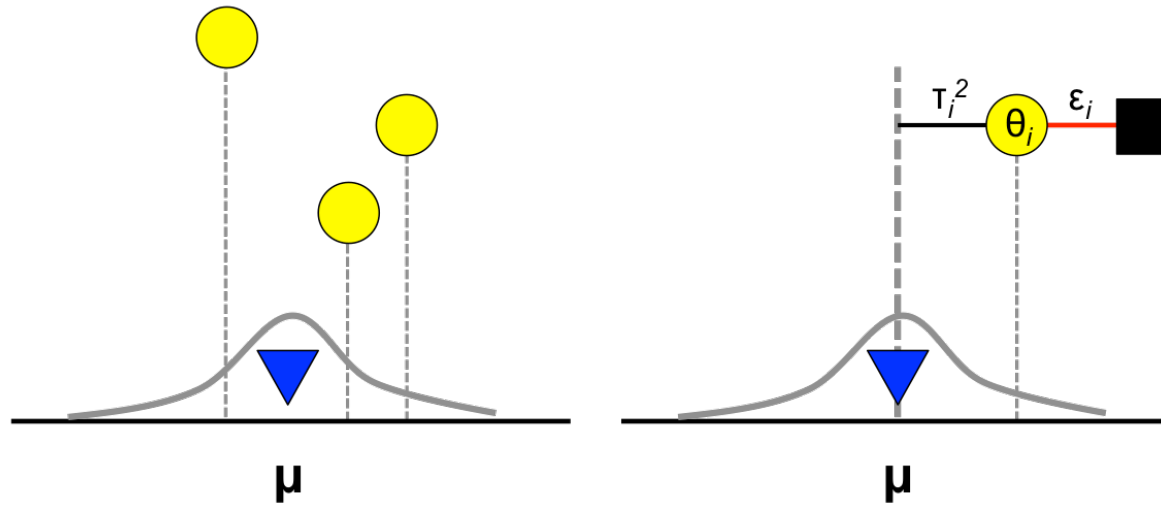
Random-effects model

- Usado para estimar a natureza de um efeito;
- Baseado no pressuposto de normalidade;
- Em um random-effects model:
 - Estudos pertencem a diferentes subpopulações que compõem uma população maior: cada estudo está estimando o verdadeiro efeito em sua subpopulação.
 - A variabilidade existente é explicada por erro na amostragem em cada estudo e por diferenças entre estudos.

Random-effects model

- Em um random-effects model, temos dois componentes de variância:
 - Variância dentro dos estudos: ϵ_i (**within-study variance**)
 - Variância entre os estudos: τ_i^2 (**between-study variance**)
- Você é capaz de explorar onde existe maior variância nos resultados: dentro ou entre estudos - em outras palavras: *consistência* ou *contingência*.

■ Efeito observado em cada estudo ● Efeito real em cada estudo ▼ Efeito combinado real



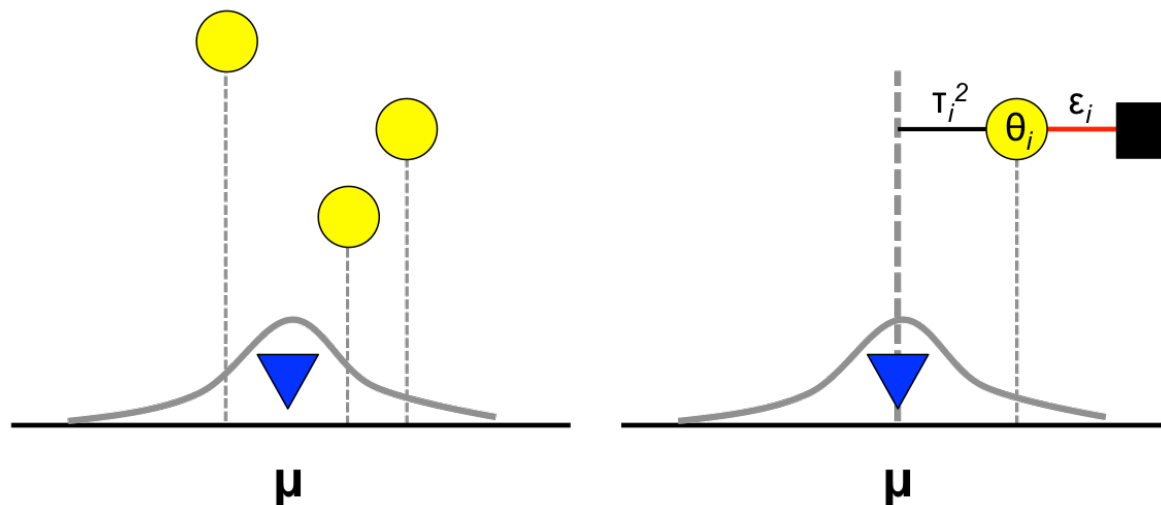
Random-effects model

- Em um random-effects model:

$$\tau_i = \theta_i + \epsilon_i \text{ (efeito real em cada estudo)}$$

$$\mu = \tau_i + \tau_i^2 \text{ (efeito combinado real)}$$

■ Efeito observado em cada estudo ● Efeito real em cada estudo ▼ Efeito combinado real



Random-effects model

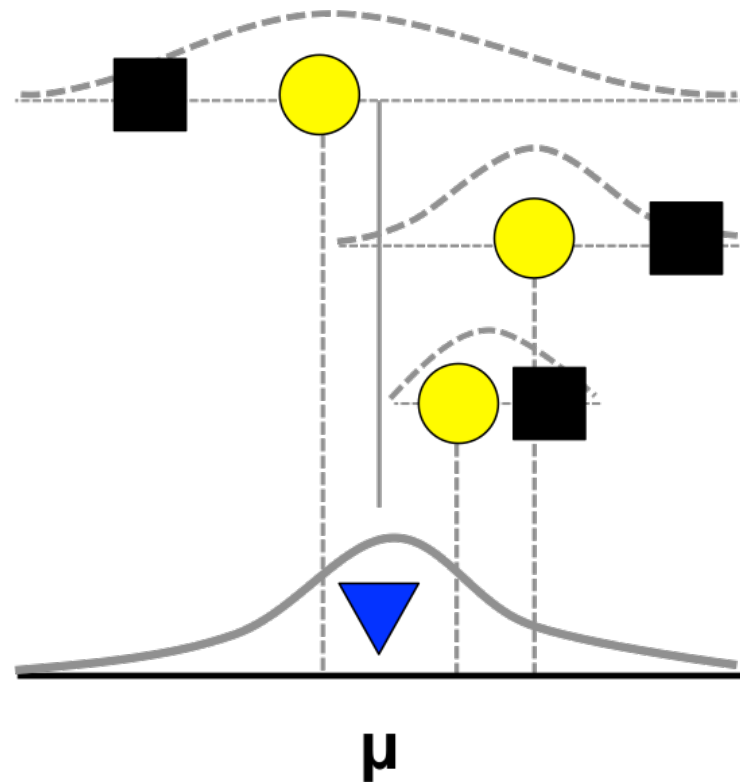
- O efeito comum que os estudos estão medindo contém a variabilidade entre estudos e também dentro de cada estudo.
- Isto afeta diretamente a estimativa do peso de cada estudo no modelo da meta-análise:

$$w_i = \frac{1}{s_i + \tau^2}$$

- Aqui, τ^2 representa o componente da heterogeneidade total que é originária entre estudos (resolução dos cálculos para chegar a este valor está no Box 9.2, Capítulo 9 do livro de Meta-Análise).

Random-effects model

- Generalizações podem ser feitas aos estudos incluídos na meta-análise e também aqueles que são desconhecidos.



No R

- Também usaremos a função `rma` para implementar um random-effects model (modelo de efeitos aleatórios), mas para isso **não** precisamos especificar o argumento `method` - o padrão adotado por ele é o do random-effects model.

```
rma(yi = yi, vi = vi, data = dados)
```

```
##
## Random-Effects Model (k = 16; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 1.2288 (SE = 0.5923)
## tau (square root of estimated tau^2 value):      1.1085
## I^2 (total heterogeneity / total variability):    77.20%
## H^2 (total variability / sampling variability):   4.39
##
## Test for Heterogeneity:
## Q(df = 15) = 64.3755, p-val < .0001
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
## 0.6955 0.3190 2.1800 0.0293 0.0702 1.3208 *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Fixed- vs Random-effects model

- Qual modelo utilizar?
 - RE é mais realístico à meta-análises em ecologia, uma vez que é impossível replicar perfeitamente os experimentos e sistemas de estudos são naturalmente diferentes;
 - No entanto, RE requer a estimativa de um parâmetro adicional relacionado à e entre estudos, que pode ser enviesado por pequeno tamanho amostral ou número de estudos - uso de REML atenua este problema;
 - Um implicação do uso de um termo a mais para a variância: intervalos de confiança estimados com RE models tendem a ser mais amplos do que aqueles estimados com FE models.

Modelo	Estimate	SE	CI.Inferior	CI.Superior
Fixed-effects model	0.429	0.150	0.135	0.723
Random-effects model	0.695	0.319	0.070	1.321

Mixed-effects models

- É utilizado quando queremos não só estimar um efeito, mas também explorar a heterogeneidade ao redor dele;
- Acomoda uma fração aleatória da heterogeneidade (random-effects) - causada por variância entre- e dentro- dos estudos -, assim como uma fração fixa (característica dos estudos que queremos explorar);
- Neste sentido, ao invés de 'jogar' toda a heterogeneidade entre estudos para um termo de variância, parte dela é atribuída a *moderadores* (veremos mais sobre isso na próxima aula);
- No R, basta usarmos a função `rma` no `metafor` e especificar a variável que queremos analisar com o argumento `mods`;
- Você também pode buscar explorar a heterogeneidade ao redor da estimativa do efeito utilizando um fixed-effects model: basta especificar o argumento `method = "FE"` na mesma estrutura do modelo abaixo.

No R

- Podemos implementar um mixed-effects model (modelo de efeitos mistos) através da função `rma`, sem especificar o argumento `method` e definindo moderadores no argumento `mods`.

```
rma(yi = yi, vi = vi, mods = ~ ecossistema, data = dados)
```

```
##
## Mixed-Effects Model (k = 16; tau^2 estimator: REML)
##
## tau^2 (estimated amount of residual heterogeneity):      1.3040 (SE = 0.6424)
## tau (square root of estimated tau^2 value):            1.1419
## I^2 (residual heterogeneity / unaccounted variability): 78.08%
## H^2 (unaccounted variability / sampling variability):    4.56
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 14) = 63.3586, p-val < .0001
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 0.3819, p-val = 0.5366
##
## Model Results:
##
##               estimate      se      zval      pval      ci.lb      ci.ub
## intrcpt           0.9015  0.4618   1.9521  0.0509   -0.0036   1.8066
## ecossistemaTerrestre -0.4034  0.6528  -0.6180  0.5366   -1.6828   0.8760
##
## intrcpt           .
```

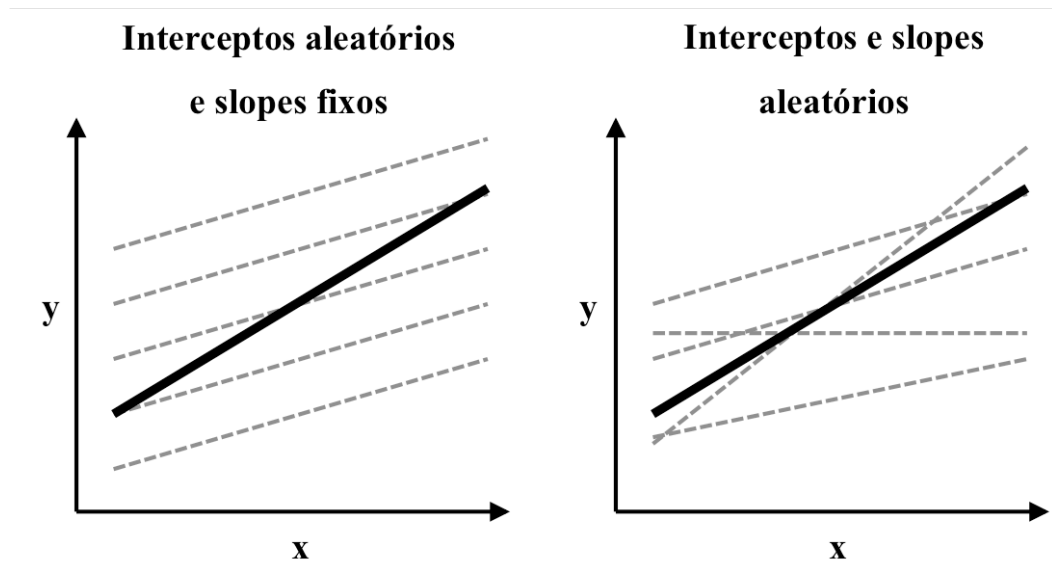
Modelo Hierárquico ou Multinível

- Dentro do contexto meta-analítico, este tipo de modelo permite que se identifique a falta de independência entre observações (e.g., várias observações vindas do mesmo estudo).
- Portanto, dentro deste arcabouço existe parte da heterogeneidade que é oriunda de diferenças entre observações do mesmo estudo (dentro-estudo) e nos efeitos detectados entre estudos (entre-estudos) - similar a um random-effects model.
- No nosso exemplo, algumas observações foram feitas pelo mesmo grupo de pesquisa ou foram publicados no mesmo trabalho.

##	estudo	autor	riqueza_tratamento	ecossistema	yi	vi
## 13	1	Aline	9	Terrestre	1.3707	0.4939
## 7	2	Ana	5	Aquatico	1.8936	0.4138
## 14	2	Ana	9	Terrestre	1.3479	0.4908
## 2	3	Antonio	4	Aquatico	-0.3242	0.2251
## 4	4	Fernanda	5	Aquatico	-0.8046	0.3088
## 9	5	Francisca	6	Aquatico	1.7914	0.4003
## 5	6	Joao	5	Terrestre	-0.2034	0.2872
## 10	7	Jose	7	Terrestre	0.5054	0.2948
## 15	7	Jose	10	Aquatico	2.4570	0.8773
## 11	8	Lucas	8	Terrestre	1.3836	0.3541
## 6	9	Marcia	5	Terrestre	-0.6654	0.3015
## 3	10	Marcos	4	Aquatico	-0.7577	0.2382
## 1	11	Maria	3	Terrestre	-1.3334	0.2716
## 8	11	Maria	6	Aquatico	1.6277	0.3803
## 12	11	Maria	8	Aquatico	2.1690	0.5294
## 16	11	Maria	10	Terrestre	2.2624	0.8199

No R

- Podemos implementar modelos hierárquicos ou multinível através da função `rma.mv`, com pequenas alterações na sintaxe da função que já conhecíamos.
- Estrutura de termos aleatórios: **slope aleatório** | **intercepto_aleatório**:
 - **Intercepto aleatório**: existe uma magnitude comum do efeito de **x** em **y**, mas ela difere de estudo para estudo (uma 'reta' para cada estudo);
 - **Slope aleatório**: existe uma relação entre **x** e **y** comum a todo o conjunto de dados, mas ela difere de estudo para estudo (a 'reta' de cada estudo tem inclinações diferentes).



No R

```
rma.mv(yi = yi, V = vi, random = ~ 1|estudo, data = dados)

##
## Multivariate Meta-Analysis Model (k = 16; method: REML)
##
## Variance Components:
##
##           estim      sqrt  nlvls  fixed  factor
## sigma^2      0.7310  0.8550      11      no  estudo
##
## Test for Heterogeneity:
## Q(df = 15) = 64.3755, p-val < .0001
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
##  0.4398  0.3028  1.4523  0.1464  -0.1538  1.0334
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

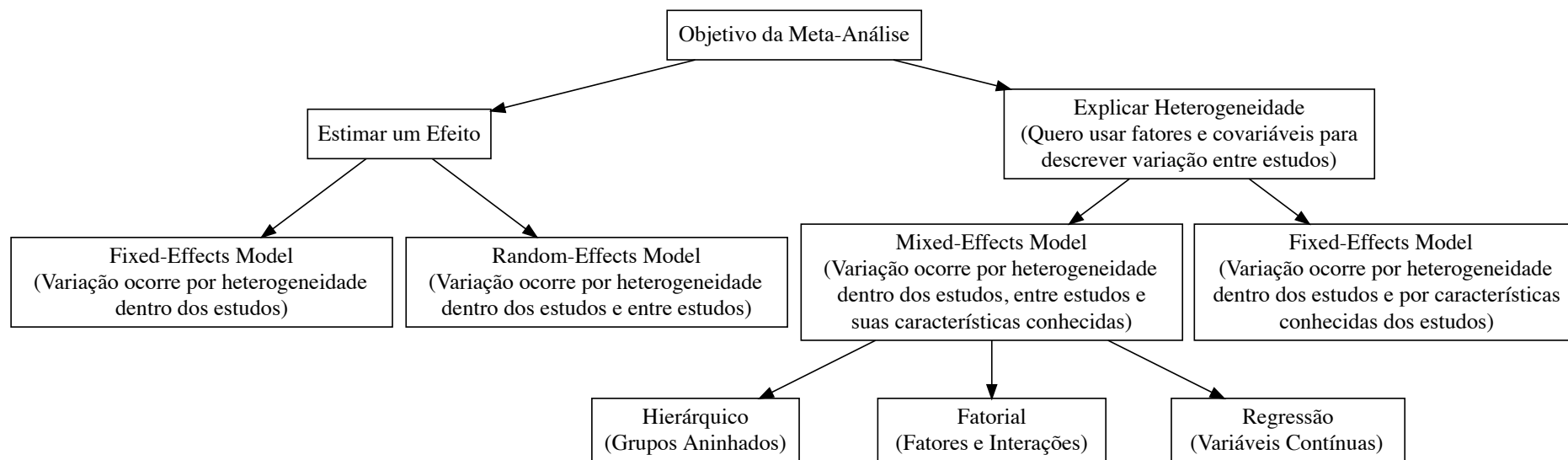
Outros modelos mais complexos

- **Meta-Regressão:** é mais uma terminologia para representar os modelos em que a heterogeneidade é explicada por variáveis contínuas (embora algumas pessoas também usem este termo quando usam variáveis categóricas);
- **Modelo Fatorial:** usado quando a meta-análise segue um desenho fatorial;
- **Modelo via Generalized Linear (Mixed-Effects) Models:** similar aos fixed-, random- e mixed-effects model, mas permite usar outras famílias de distribuição de variáveis aleatórias (no R: `rma.glmm`).

Observações

- Outros pacotes para rodar os modelos de meta-análise: `nlme` e `lme4`.
- Uma alternativa ao uso do inverso da variância como o peso de cada observação é usar o *tamanho amostral* de cada estudo: estudos maiores têm mais peso do que os estudos menores, pois tendem a ser mais precisos.
- Recomenda-se que os intervalos de confiança para a estimativa do efeito também sejam calculados através do método do *bootstrap* paramétrico ou não-paramétrico (ver pacote `boot`).
 - Paramétrico: 'cria' populações de effect size baseado na estimativa do efeito e da variância pelo modelo;
 - Não-paramétrico: 'cria' populações de effect size baseado nas observações existentes no conjunto de dados utilizado na meta-análise.
- Você também pode usar o *bootstrap* para calcular intervalos de confiança para cada nível de moderadores categóricos, enquanto que para moderadores contínuos utilizar o *bootstrap* pode ser um pouco mais complicado.

Só para ficar didático



Resumindo

- Existem diversos tipos de modelos que podem ser usados em uma meta-análise, cada um com um tipo de característica e uso próprio;
- Não existe o melhor modelo para a sua meta-análise, mas sim o modelo que descreve melhor seus objetivos e perguntas;
- Independente do modelo que você use, é boa prática ponderar cada observação/estudo pelo inverso de sua variância.

Literatura Recomendada

1. Nakagawa & Santos, 2012, *Evol Ecol*, Methodological issues and advances in biological meta-analysis
2. Harrison, 2011, *Methods Ecol Evol*, Getting started with meta-analysis
3. Mengersen et al, 2013, Statistical models and approaches to inference, In: *Handbook of meta-analysis in ecology and evolution* (Capítulo 8)
4. Rosenberg, 2013, Moment and least-squares based approaches to meta-analytic inference, In: *Handbook of meta-analysis in ecology and evolution* (Capítulo 9)
5. Mengersen & Schmid, 2013, Maximum likelihood approaches to meta-analysis, In: *Handbook of meta-analysis in ecology and evolution* (Capítulo 10)