**Modelos Mistos**

**LMMs** - Linear **Mixed** Models - Modelos Lineares **Mistos**

**GLMMs** - Generalized Linear **Mixed** Models - Modelos Lineares Generalizados **Mistos**

Os GLMMs e LMMs estendem os modelos para permitir a inclusão de efeitos aleatórios. Neste artigo, exploraremos como e quando aplicarmos, abrangendo as etapas necessárias, sintaxe, interpretação e técnicas avançadas.

Um dos principais pontos fortes da Regressão de Efeitos Mistos é sua capacidade de modelar correlações temporais e espaciais. Por exemplo, quando as medições são feitas dos mesmos sujeitos ao longo do tempo, as observações provavelmente serão mais semelhantes dentro dos sujeitos do que entre os sujeitos. A Regressão de Efeitos Mistos pode modelar efetivamente essa correlação intra-sujeito.

Da mesma forma, na análise de dados espaciais, as observações feitas em locais próximos costumam ser mais semelhantes do que aquelas feitas em locais distantes. Ao incorporar efeitos aleatórios para cada local, podemos modelar essa correlação espacial.

A beleza da Regressão de Efeitos Mistos está em sua flexibilidade. Ele nos permite modelar estruturas hierárquicas complexas em nossos dados, fornecendo um reflexo mais preciso dos cenários do mundo real.

**Oque são Efeitos Fixos e Aleatórios?**

**Efeitos Fixos**

* **Definição**: Efeitos fixos são coeficientes que se assumem serem constantes para toda a população ou amostra. Eles capturam a influência média de uma variável explicativa na variável resposta.
* **Aplicação**: São utilizados quando se quer estimar o efeito de uma variável sobre a resposta de maneira uniforme em toda a população.
* **Exemplo**: Em um estudo sobre o impacto de um medicamento, a dose do medicamento seria um efeito fixo se se assume que o efeito da dose é o mesmo para todos os pacientes.

De maneira geral seriam as variáveis independentes, previsoras, que selecionamos nos modelos mais comuns.

**Efeitos Aleatórios**

* **Definição**: Efeitos aleatórios são coeficientes que permitem variar entre grupos ou clusters. Eles capturam a variabilidade entre esses grupos ou clusters, além do efeito médio estimado pelos efeitos fixos.
* **Aplicação**: Utilizados quando se espera que a relação entre a variável explicativa e a resposta varie de grupo para grupo.
* **Exemplo**: Em um estudo educacional, o efeito de uma metodologia de ensino poderia variar entre diferentes escolas, sendo "escola" um efeito aleatório.

São as fontes de ruido que enquadram os dados em grupos com padrões ou níveis base diferentes, tudo aquilo que causa uma variabilidade não planejada no estudo, exemplo:

Objetivo do estudo: Avaliação da eficácia de diferentes tratamentos em pacientes de diferentes hospitais.

Fonte de erro: “diferentes hospitais”, a variabilidade nos dados causada por esse fato é impactante para o estudo, dado que pacientes dentro do mesmo hospital podem ter características semelhantes e que também o corpo clínico é diferente, o padrão de qualidade pode ser diferente, portanto, deve-se considerar a variabilidade por hospital, segue abaixo o código para modelagem.

lmer(qualidade\_cuidado ~ idade\_paciente + gravidade\_doenca + (1|hospital), data = dados\_hierarquicos)

“Tá, mas, por que não colocar tudo como ‘Hospital’ como covariável?”.   
Desta forma é como se vc disse-se que Hospital tem um efeito, e não que os sujeitos estão condicionados ao efeito que o hospital tem sobre os sujeitos

**Dependência**

* **Definição**: Dependência nos dados ocorre quando as observações não são independentes entre si, o que é comum em dados agrupados, hierárquicos ou longitudinais.  
  Modelos que não levam em conta a dependência podem levar a conclusões incorretas, subestimando a variabilidade real dos dados.  
  Os modelos mistos são versáteis e poderosos para analisar dados com estruturas complexas, permitindo uma melhor compreensão dos fatores que influenciam as variáveis de interesse enquanto capturam a variabilidade inerente a diferentes níveis de agrupamento.  
  + **Estudos agrupados:** Comparar resultados entre diferentes locais onde as respostas individuais dos indivíduos podem variar, mas são influenciadas por características de nível local. (observações dentro de clusters)

lmer(pressao ~ idade + (1|hospital), data = dados\_agrupados)

lmer(nota ~ idade + (1|turma), data = dados\_agrupados)

* + **Estudos hierárquicos:** Mesma coisa dos agrupados, porém possuem uma estrutura de múltiplos níveis de agrupamento, onde grupos menores estão aninhados dentro de grupos maiores.

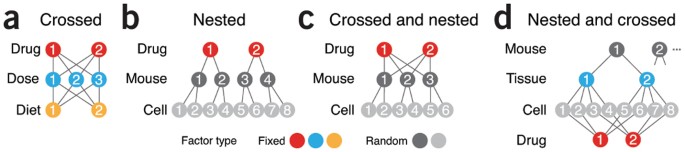
lmer(nota ~ idade + (1|escola/turma), data = dados\_hierarquicos)

* + **Estudos longitudinais:** Modelar o mesmo grupos ou indivíduos ao longo do tempo. (múltiplas medições de indivíduos)

lmer(pressao ~ idade + tempo + (1|id\_paciente), data = dados\_longitudinais)

Os tipos de estruturas e aplicações será detalhado melhor abaixo no próximo tópico.

**Tipos de Estruturas Aleatórias**



[**Designs aninhados | Métodos da Natureza**](https://www.nature.com/articles/nmeth.3137)

1. **Interceptos Aleatórios:**
   * Captura a variação no intercepto entre os grupos.
   * (1 | grouping\_factor)

response ~ fixed\_effects + (1|group)

desempenho ~ idade + tempo + (1|id)

pressao ~ idade + (1|hospital)

É usado quando queremos modelar a variabilidade entre grupos ou perfis individuais em termos de interceptos, permitindo que cada grupo tenha seu próprio intercepto, mas assumindo que a inclinação das variáveis fixas é a mesma para todos os grupos.

Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

1. **Interceptos e Inclinações Aleatórias:**
   * Captura a variação tanto no intercepto quanto na inclinação (slopes) entre os grupos. Modela a variável aleatória com a interação com a variável fixa. Modelando também a inclinação do modelo.
   * (slope | grouping\_factor)

response ~ fixed\_effects + (1 + fixed\_effects|group)

Esse modelo permite que tanto o intercepto quanto a inclinação da variável fixa variem entre os grupos.

É útil quando você acredita que a relação entre a variável dependente e uma ou mais variáveis independentes pode diferir entre os grupos.

**Motivação:**

Aplicável quando há razão para acreditar que a resposta à variável independente (inclinação) não é constante entre os grupos.

Permite modelar a heterogeneidade na resposta dos grupos.

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

response ~ age + (1 + age|clinic), data = df

Neste exemplo, estamos modelando a resposta em função da idade, permitindo que tanto o intercepto quanto a inclinação da idade variem entre as clínicas. Isso pode ser útil se diferentes clínicas têm diferentes tendências de progressão da doença com a idade.

response ~ tratamento + (1 + tratamento|hospital)

Aqui, estamos modelando a resposta ao tratamento, permitindo que o efeito do tratamento (inclinação) e o intercepto variem entre os hospitais.

1. **Inclinações Aleatórias:**
   * Captura a variação na inclinação entre os grupos, mas não no intercepto.
   * (0 + slope | grouping\_factor)

response ~ fixed\_effects + (0 + fixed\_effects | grouping\_factor)

Esse modelo permite que apenas a inclinação (e não o intercepto) varie entre os grupos.

É útil quando você deseja modelar a variabilidade nas inclinações sem alterar os interceptos.

É utilizável quando se considera que os grupos têm, em média, o mesmo nível inicial e que o efeito da variável independente sobre a variável dependente varia entre os grupos.

Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

Exemplos de aplicação:

Imagine que você está investigando como a idade afeta o desempenho acadêmico dos estudantes em diferentes escolas. Algumas escolas podem ter programas que ajudam estudantes mais jovens a se destacarem, enquanto outras podem ter programas que beneficiam estudantes mais velhos. Neste caso, você usaria inclinações aleatórias para modelar essas diferenças:

idade ~ age + (0 + age | school)

Em estudos longitudinais, você pode estar interessado em como o tempo afeta a recuperação dos pacientes em diferentes clínicas. Diferentes clínicas podem ter diferentes protocolos de tratamento que influenciam a taxa de recuperação ao longo do tempo:

recovery ~ time + (0 + time | clinic)

1. **Efeitos Cruzados:**
   * Permite modelar a variação em dois ou mais fatores de agrupamento que não estão aninhados (não hierarquicos).
   * (1 | factor1) + (1 | factor2)

response ~ fixed\_effects + (1 | factor1) + (1 | factor2)

Isso significa que as unidades podem ser agrupadas de maneira independente por duas ou mais variáveis.

As unidades de um grupo podem aparecer em diferentes combinações com as unidades de outro grupo. Em outras palavras, não há uma hierarquia fixa entre os grupos.

Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

Exemplos de aplicação:

Satisfação no trabalho de funcionários em diferentes departamentos e projetos, neste exemplo, os funcionários trabalham em diferentes departamentos e participam de diferentes projetos. Os departamentos e projetos são independentes, e um funcionário pode ser alocado em qualquer combinação de departamento e projeto.

satisfaction ~ experience + (1|department) + (1|project)

obs: dependendo do pacote a sintaxe (1|group1) + (1|group2) não funciona, uma alternativa para contornar esse erro seria criar uma variável nova das combinações das duas variáveis:

df$interation <- factor(interaction(df$group1, df$group2))

modelo <- lmer(response ~ fixed\_effects + (1|interation), data = df)

1. **Efeitos Aleatórios Aninhados:**
   * Captura a variação em fatores de agrupamento que estão aninhados (um fator dentro de outro - Hierárquicos).
   * (1 | higher\_level/lower\_level)

response ~ fixed\_effects + (1| higher\_level/lower\_level)

Há uma dependência hierárquica entre os grupos. Por exemplo, alunos em turmas dentro de escolas. Cada turma está completamente contida dentro de uma escola específica. Captura a variabilidade no desempenho devido a diferenças entre escolas e entre turmas dentro de escolas.

desempenho ~ idade + intervencao + (1|escola/turma)

Outros exemplos:

saude ~ idade + programa\_saude + (1|estado/municipio)

health ~ treatment + (1|hospital/ward)

job\_satisfaction ~ experience + (1|company/department)

A representação gráfica é a mesma do modelo cruzado, porém sem todas as combinações entre os níveis das duas variáveis

Gráfico, Gráfico de linhas, Gráfico de dispersão

Descrição gerada automaticamente

1. **Efeitos Aleatórios Correlacionados:**
   * Inclui correlação entre os efeitos aleatórios de intercepto e inclinação.
   * (intercept + slope || grouping\_factor)

response ~ fixxed\_effects + (1 + fixxed\_effects || group), data = df)

Específica explicitamente que os efeitos aleatórios (interceptos e inclinações) são correlacionados. Na prática, essa é a mesma estrutura do "Modelo com Interceptos e Inclinações Aleatórios" única diferença é o ajuste matemático que permite a correlação entre os efeitos aleatórios por padrão, tanto é, que se investigar a diferença dos modelos através do teste de Anova, não haverá diferença significativa entre os modelos.

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

**Famílias e links de ligação de modelos de Regressão**

As famílias definem a distribuição da variável resposta e a função de variância. Os links conectam o preditor linear à média da variável resposta, por exemplo, binomial() para resultados binários).

**1. Variável Dependente Contínua com Distribuição Normal**

**Família:** Gaussiana (Normal)  
**Função de Ligação:** Identidade  
**Uso:** Modelar dados contínuos.

lm(pressao ~ var1 + var2, data = dados)

ou

glm(pressao ~ var1 + var2, family = gaussian(link = "identity"), data = dados)

**2. Variável Dependente Binária**

**Família:** Binomial  
**Função de Ligação:** Logit  
**Uso:** Modelar a probabilidade de um evento binário.

glm(obito ~ doença\_cardiaca + avc\_previo, family = binomial(link = "logit"), data = dados)

**3. Variável Dependente de Contagem**

**Família:** Poisson  
**Função de Ligação:** Log  
**Uso:** Modelar contagens de eventos. (tbm pode ser usado para eventos binário)

glm(qtd\_avc ~ tratamento + var2, family = poisson(link = "log"), data = dados)

**4. Variável Dependente Contínua com Distribuição Gamma**

**Família:** Gamma  
**Função de Ligação:** Inversa  
**Uso:** Modelar dados contínuos positivos com variância proporcional ao quadrado da média.

glm(despesa\_medica ~ idade + tratamento, family = Gamma(link = "inverse"), data = dados)

**5. Variável Dependente Contínua com Distribuição Inverse.Gaussian**

**Família:** Inverse.Gaussian  
**Função de Ligação:** 1/mu²  
**Uso:** Modelar dados contínuos positivos com variância proporcional ao cubo da média.

glm(tempo\_recuperacao ~ idade + tratamento, family = inverse.gaussian(link = "1/mu^2"), data = dados)

**6. Variável Dependente com Distribuição Quasi**

**Família:** Quasi  
**Função de Ligação:** Identidade  
**Variância:** Constante  
**Uso:** Modelar dados quando a variância não segue uma distribuição específica e deve ser modelada de forma flexível.

glm(variavel\_dependente ~ var1 + var2, family = quasi(link = "identity", variance = "constant"), data = dados)

**7. Variável Dependente Binária com Distribuição Quasibinomial**

**Família:** Quasibinomial  
**Função de Ligação:** Logit  
**Uso:** Modelar a probabilidade de um evento binário quando há sobredispersão nos dados.

glm(sucesso ~ var1 + var2, family = quasibinomial(link = "logit"), data = dados)

**8. Variável Dependente de Contagem com Distribuição Quasipoisson**

**Família:** Quasipoisson  
**Função de Ligação:** Log  
**Uso:** Modelar contagens de eventos quando há sobredispersão nos dados.

glm(numero\_eventos ~ var1 + var2, family = quasipoisson(link = "log"), data = dados)

**Exemplos de Aplicações de Modelos Mistos**

1. **Estudos Médicos**:
   * **Objetivo**: Avaliar o efeito de um tratamento ao longo do tempo em diferentes pacientes.
   * **Modelo**:

glmer(resultado ~ tratamento \* tempo + (1 | paciente), family = binomial(link = "logit"), data = dados\_medicos)

* + **Interpretação**: O modelo avalia como o tratamento e o tempo influenciam o resultado, considerando a variabilidade entre pacientes.

1. **Educação**:
   * **Objetivo**: Estudar o impacto de uma nova metodologia de ensino em diferentes escolas.
   * **Modelo**:

lmer(desempenho ~ metodologia + (1 | escola), data = dados\_educacionais)

* + **Interpretação**: O modelo avalia o impacto da metodologia de ensino no desempenho dos alunos, levando em conta a variabilidade entre escolas.

1. **Ecologia**:
   * **Objetivo**: Modelar a abundância de uma espécie em diferentes locais e anos.
   * **Modelo**:

glmer(abundancia ~ local \* ano + (1 | local/ano), family = poisson(link = "log"), data = dados\_ecologicos)

* + **Interpretação**: O modelo avalia como a abundância da espécie varia entre locais e anos, considerando a dependência hierárquica dos dados.

1. **Economia**:
   * **Objetivo**: Analisar o impacto de políticas econômicas em diferentes regiões.
   * **Modelo**:

lmer(crescimento ~ politica + (1 | regiao), data = dados\_economicos)

* + **Interpretação**: O modelo avalia o efeito das políticas econômicas no crescimento econômico, levando em conta a variabilidade entre regiões.

**Mais Exemplos de Estruturas de Dados Hierárquicos, Longitudinais e Agrupados**

1. **Hierárquicos**:
   * **Exemplo**: Desempenho de estudantes em diferentes escolas e turmas.
   * **Descrição**: Os dados podem estar estruturados em múltiplos níveis, como estudantes dentro de turmas e turmas dentro de escolas.
   * **Aplicação**: Avaliar o impacto de características da escola e da turma no desempenho dos estudantes.
   * **Modelo**:

lmer(desempenho ~ var1 + (1 | escola/turma), data = dados)

1. **Longitudinais**:
   * **Exemplo**: Medidas repetidas de pacientes ao longo do tempo.
   * **Descrição**: Observações repetidas da mesma variável em diferentes pontos no tempo para os mesmos indivíduos.
   * **Aplicação**: Estudar a progressão de uma doença ou o efeito de um tratamento ao longo do tempo.
   * **Modelo**:

lmer(resposta ~ tempo + tratamento + (1 | paciente), data = dados)

1. **Agrupados**:
   * **Exemplo**: Produção agrícola em diferentes fazendas e parcelas.
   * **Descrição**: Dados coletados de diferentes grupos, como fazendas, e subdivisões dentro desses grupos, como parcelas.
   * **Aplicação**: Avaliar o efeito de práticas agrícolas em diferentes regiões e parcelas.
   * **Modelo**:

lmer(producao ~ pratica + (1 | fazenda/parcela), data = dados)

1. **Espaciais**:
   * **Exemplo**: Distribuição de espécies em diferentes regiões geográficas.
   * **Descrição**: Dados coletados de diferentes localizações espaciais, onde as observações dentro de uma localização podem ser mais semelhantes entre si do que entre diferentes localizações.
   * **Aplicação**: Modelar a distribuição espacial de espécies e entender fatores ambientais que afetam essa distribuição.
   * **Modelo**:

lmer(abundancia ~ fator\_ambiental + (1 | regiao/localizacao), data = dados)

1. **Indústria**:
   * **Exemplo**: Qualidade do produto em diferentes lotes de produção.
   * **Descrição**: Observações sobre a qualidade do produto coletadas de diferentes lotes e sublotes.
   * **Aplicação**: Avaliar a consistência da qualidade do produto entre diferentes lotes de produção.
   * **Modelo**:

lmer(qualidade ~ var1 + (1 | lote/sublote), data = dados)

**Fontes**

<https://www.youtube.com/watch?v=pLYR-GCn-ko>

<https://www.youtube.com/watch?v=DDP62EUMRFs>

<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/family>

<https://www.statmethods.net/advstats/glm.html>

<https://stats.stackexchange.com/questions/190763/how-to-decide-which-glm-family-to-use>

<https://stats.stackexchange.com/questions/356046/binomial-vs-quasi-binomial-model>

<https://stats.stackexchange.com/questions/190763/how-to-decide-which-glm-family-to-use>

[Modelos Lineares Mistos (LMM) [Laboratório de Ecologia de Florestas Tropicais] (usp.br)](http://labtrop.ib.usp.br/doku.php?id=cursos:planeco:roteiro:11-lmm_rcmdr)

[Modelos Lineares Mistos [Laboratório de Ecologia de Florestas Tropicais] (usp.br)](http://labtrop.ib.usp.br/doku.php?id=cursos:planeco:roteiro:11-lmm)

[Uma comparação entre ANOVA e modelos lineares mistos para análise de dados de tempo de resposta | Revista da ABRALIN](https://revista.abralin.org/index.php/abralin/article/view/1388)

[Confiabilidade e Viabilidade de Modelos Lineares Mistos em Experimentos Totalmente Cruzados - Michele Scandola, Emmanuele Tidoni, 2024 (sagepub.com)](https://journals.sagepub.com/doi/10.1177/25152459231214454)

Como reportar os resultados ¿