

# Café com estatística e R

## Cronograma de treinamentos

Marcelo Teixeira Paiva

2025-09-22

## **Abstract**

Cronograma proposto dos treinamentos.

# Índice

<b>Estatística Clássica e fundamentos de epidemiologia</b>	<b>4</b>
Treinamento 1: Introdução ao R e variáveis em estudos . . . . .	4
Treinamento 2: estatísticas descritivas e construção de gráficos no R . . . . .	4
Treinamento 3: Probabilidade, variáveis aleatórias e distribuições de probabilidade . . . . .	5
Treinamento 4: Causalidade, amostragem e desenho de questionários . . . . .	6
Treinamento 5: Testes de hipóteses e testes paramétricos . . . . .	8
Treinamento 6: Testes não paramétricos . . . . .	8
Treinamento 7: Medidas de frequência de doenças e de associação . . . . .	8
Treinamento 8: Tipos de estudo . . . . .	9
<b>Técnicas Multivariadas Exploratórias</b>	<b>10</b>
Treinamento 9: Análise de agrupamentos, análise fatorial por componentes principais e análise de correspondência . . . . .	10
<b>Técnicas Multivariadas Confirmatórias ou de Modelagem</b>	<b>11</b>
Treinamento 10: Modelos Lineares Generalizados - Regressão linear simples e múltipla . . . . .	11
Treinamento 11: Modelos Lineares Generalizados - Regressão Logística Binária e Multinomial . .	12
Treinamento 12: Modelos Lineares Generalizados para Dados de Contagem - Regressão de Poisson e Binomial Negativo . . . . .	12
Treinamento 13: Regressão para Dados em Painel em modelos longitudinais . . . . .	12
Treinamento 14: Regressão para Dados em Painel em modelos multinível . . . . .	13

# Lista de Figuras

# Lista de Tabelas

Abaixo está a proposta dos tópicos que serão abordados nos treinamentos.

# Estatística Clássica e fundamentos de epidemiologia

## Treinamento 1: Introdução ao R e variáveis em estudos

- Introdução ao R
- Tipos de variáveis e escalas de mensuração e precisão
- Tipos de variáveis
  - Variáveis não métricas ou qualitativas
  - Variáveis métricas ou quantitativas
- Tipos de variáveis x escalas de mensuração
  - Variáveis não métricas - escala nominal
  - Variáveis não métricas - escala ordinal
  - Variável quantitativa - escala intervalar
  - Variável quantitativa - escala de razão
- Tipos de variáveis x número de categorias e escalas de precisão
  - Variável dicotômica ou binária (dummy)
  - Variável policotômica
  - Variável quantitativa discreta
  - Variável quantitativa contínua

## Treinamento 2: estatísticas descritivas e construção de gráficos no R

- Estatística descritiva univariada
  - Tabela de distribuição de frequências
    - \* Tabela de distribuição de frequências para variáveis qualitativas
    - \* Tabela de distribuição de frequências para dados discretos
    - \* Tabela de distribuição de frequências para dados contínuos agrupados em classes
  - Representação gráfica dos dados
  - Representação gráfica para variáveis qualitativas
    - \* Gráfico de barras
    - \* Gráfico de setores ou pizza
    - \* Diagrama de Pareto
  - Representação gráfica para variáveis quantitativas
    - \* Gráfico de linhas
    - \* Gráfico de pontos ou dispersão
    - \* Histograma

- \* Gráfico de ramo-e-folhas
- \* Boxplot ou diagrama de caixa
- Medidas-resumo
  - \* Medidas de posição ou localização
    - Medidas de tendência central
    - Medidas separatrizes
    - Identificação de existência de outliers univariados
  - \* Medidas de dispersão ou variabilidade
    - Amplitude
    - Desvio-médio
    - Variância
    - Desvio-padrão
    - Erro-padrão
    - Coeficiente de variação
- Medidas de forma
  - \* Medidas de assimetria
  - \* Medidas de curtose
- Estatística descritiva bivariada
  - Associação entre duas variáveis qualitativas
    - \* Tabelas de distribuição conjunta de frequências
  - Medidas de associação
    - \* Estatística qui-quadrado
    - \* Outras medidas de associação baseadas no qui-quadrado
    - \* coeficiente de Spearman
  - Correlação entre duas variáveis quantitativas
    - \* Tabelas de distribuição conjunta de frequências
    - \* Representação gráfica por meio de um diagrama de dispersão
  - Medidas de correlação
    - \* Covariância
    - \* Coeficiente de correlação de Pearson

## Treinamento 3: Probabilidade, variáveis aleatórias e distribuições de probabilidade

- Probabilidade
  - Terminologia e conceitos
    - \* Experimento aleatório
    - \* Espaço amostral
    - \* Eventos
    - \* Uniões, intersecções e complementos
    - \* Eventos independentes
    - \* Eventos mutuamente excludentes
    - \* Classificação de experimentos
  - Definição de probabilidade
  - Fundamentos da probabilidade
  - Axiomas de Kolmogorov e propriedades derivadas dos axiomas
  - Regras básicas da probabilidade
    - \* Campo de variação da probabilidade
    - \* Probabilidade do espaço amostral

- \* Probabilidade de um conjunto vazio
  - \* Regra de adição de probabilidades
  - \* Probabilidade de um evento complementar
  - \* Regra da multiplicação de probabilidades para eventos independentes
  - \* Probabilidade condicional
- Regra da multiplicação de probabilidades
  - \* O teorema de Bayes
- Análise combinatória
  - \* Arranjos
  - \* Combinações
  - \* Permutações
- Variáveis aleatórias e distribuições de probabilidade
  - Variáveis aleatórias
    - \* Variável aleatória discreta
      - Conceito
      - Esperança de uma variável aleatória discreta
      - Variância de uma variável aleatória discreta
      - Momentos de uma variável aleatória discreta
      - Função de distribuição acumulada de uma variável aleatória discreta
    - \* Variável aleatória contínua
      - Conceito
      - Esperança de uma variável aleatória contínua
      - Variância de uma variável aleatória contínua
      - Momentos de uma variável aleatória contínua
      - Função de distribuição acumulada de uma variável aleatória contínua
  - Distribuições de probabilidades para variáveis aleatórias discretas
    - \* Distribuição uniforme discreta
    - \* Distribuição de Bernoulli
    - \* Distribuição binomial
    - \* Distribuição geométrica
    - \* Distribuição binomial negativa
    - \* Distribuição hipergeométrica
    - \* Distribuição Poisson
  - Distribuições de probabilidades para variáveis aleatórias contínuas
    - \* Distribuição uniforme
    - \* Distribuição normal
    - \* Distribuição exponencial
    - \* Distribuição Gama
    - \* Distribuição qui-quadrado
    - \* Distribuição t de Student
    - \* Distribuição F de Snedecor

## Treinamento 4: Causalidade, amostragem e desenho de questionários

- Causalidade
  - Conceitos de múltiplas causas e inferência científica
  - Principais componentes da pesquisa epidemiológica
  - Modelos de causalidade
  - Modelo contrafactual ou de resultados potenciais
  - Evidências de causa observacionais vs. experimentais



- Diagrama de causalidade
- Critérios de causalidade
- Amostragem
  - Amostragem probabilística ou aleatória
    - \* Amostragem aleatória simples sem reposição
    - \* Amostragem aleatória simples com reposição
    - \* Amostragem sistemática
    - \* Amostragem estratificada
    - \* Amostragem por conglomerados
    - \* Amostragem baseada em risco
  - Amostragem não probabilística ou não aleatória
    - \* Amostragem por conveniência
    - \* Amostragem por julgamento ou intencional
    - \* Amostragem por quotas
    - \* Amostragem de propagação geométrica ou bola de neve (snowball)
  - Tamanho da amostra
    - \* Tamanho da amostra aleatória simples
      - Tamanho da amostra para estimar a média de uma população infinita
      - Tamanho da amostra para estimar a média de uma população finita
      - Tamanho da amostra para estimar a proporção de uma população infinita
      - Tamanho da amostra para estimar a proporção de uma população finita
    - \* Tamanho da amostra sistemática
    - \* Tamanho da amostra estratificada
      - Tamanho da amostra estratificada para estimar a média de uma população infinita
      - Tamanho da amostra estratificada para estimar a média de uma população finita
      - Tamanho da amostra estratificada para estimar a proporção de uma população infinita
      - Tamanho da amostra estratificada para estimar a proporção de uma população finita
    - \* Tamanho da amostra por conglomerados
      - Tamanho da amostra por conglomerados em um estágio
      - Tamanho da amostra por conglomerados em dois estágios
- Desenho de questionários
  - Conceituação
  - Desenho das questões
    - \* Questões abertas
    - \* Questões fechadas
      - Checklist
      - Múltiplas opções
      - Escalas e ordenação de importância
  - Descrevendo a questão
  - Estrutura do questionário
  - Piloto e validação
    - \* Pré-teste e validação do questionário
    - \* Confiabilidade em julgamentos de categorias subjetivas
      - Método de Holsti
      - Concordância:  $\pi$  de Scott e  $\kappa$  de Cohen
  - Otimização da taxa de resposta
  - Codificação dos resultados

## Treinamento 5: Testes de hipóteses e testes paramétricos

- Testes de Hipóteses
- Testes paramétricos
- Testes para normalidade univariada
  - Teste de Kolmogorov–Smirnov
  - Teste de Shapiro–Wilk
  - Teste de Shapiro–Francia
- Testes para homogeneidade de variâncias
  - Teste  $\chi^2$  de Bartlett
  - Teste C de Cochran
  - Teste Fmax de Hartley
  - Teste F de Levene
- Testes de hipóteses sobre uma média populacional ( ) a partir de uma amostra aleatória
  - Teste z quando o desvio-padrão populacional ( ) for conhecido e a distribuição for normal
  - Teste t de Student quando o desvio-padrão populacional ( ) não for conhecido
- Teste t de Student para comparação de duas médias populacionais a partir de duas amostras aleatórias independentes
- Teste t de Student para comparação de duas médias populacionais a partir de duas amostras aleatórias emparelhadas
- Análise de variância (ANOVA) para comparação de médias de mais de duas populações
  - ANOVA de um fator (One-Way ANOVA)
  - ANOVA fatorial
    - \* ANOVA de dois fatores (Two-Way ANOVA)
    - \* ANOVA com mais de dois fatores
    - \* ANOVA de dois fatores e medidas repetidas (Two-Way ANOVA repeated measures)

## Treinamento 6: Testes não paramétricos

- Testes para uma amostra
  - Teste binomial
  - Teste qui-quadrado ( $\chi^2$ ) para uma amostra
  - Teste dos sinais para uma amostra
- Testes para duas amostras emparelhadas
  - Teste de McNemar
  - Teste dos sinais para duas amostras emparelhadas
  - Teste de Wilcoxon
- Testes para duas amostras independentes
  - Teste qui-quadrado ( $\chi^2$ ) para duas amostras independentes
  - Teste U de Mann–Whitney
- Testes para k amostras emparelhadas
  - Teste Q de Cochran
  - Teste de Friedman
- Testes para k amostras independentes
  - Teste  $\chi^2$  para k amostras independentes
  - Teste de Kruskal–Wallis

## Treinamento 7: Medidas de frequência de doenças e de associação

- Medidas de frequência da doença

- Contagem, proporção, odds e taxas
- Incidência
- Risco
- Taxa de incidência
  - \* Relação entre risco e incidência
- Taxa de prevalência
- Mortalidade
- Outras medidas de frequência
- Estimando o erro padrão e intervalo de confiança para medidas estimadas da população
- Medidas de associação e efeito
  - Razão de risco
  - Razão de taxas de incidência
  - Odds ratio
  - Medidas de efeito
    - \* Diferença de risco e de incidência
    - \* Fração atribuível
    - \* Risco atribuível populacional
    - \* Fração atribuível populacional
  - Relação entre medidas de associação e os tipos de estudos
  - Testes de hipótese e intervalo de confiança para medidas de associação

## Treinamento 8: Tipos de estudo

- Estudo observacional
- Coorte
- Caso-controle
- Estudos híbridos
- Validade em estudos observacionais
  - Tipos de vieses
  - Viés de seleção
  - Viés de informação
  - Viés de má classificação
  - Erros em medidas de exposição substitutas ou derivadas
  - Impacto do viés de informação no tamanho amostral

# Técnicas Multivariadas Exploratórias

## Treinamento 9: Análise de agrupamentos, análise fatorial por componentes principais e análise de correspondência

- Análise de agrupamentos
  - Definição das medidas de distância ou de semelhança em análise de agrupamentos
    - \* Medidas de distância (dissimilaridade) entre observações para variáveis métricas
    - \* Medidas de semelhança (similaridade) entre observações para variáveis binárias
  - Esquemas de aglomeração em análise de agrupamentos
    - \* Esquemas de aglomeração hierárquicos
    - \* Esquema de aglomeração não hierárquico k-means
  - Análise de agrupamentos com esquemas de aglomeração hierárquicos e não hierárquicos
- Análise Fatorial por Componentes Principais
  - Análise fatorial por componentes principais
  - Correlação linear de Pearson e conceito de fator
  - Adequação global da análise fatorial: estatística Kaiser–Meyer–Olkin (KMO) e teste de esfericidade de Bartlett
  - Definição dos fatores por componentes principais: determinação dos autovalores e autovetores da matriz de correlações e de cálculo dos scores fatoriais
  - Cargas fatoriais e comunalidades
  - Rotação de fatores
- Análise de Correspondência Simples e Múltipla
  - Análise de correspondência simples
    - \* Notação
    - \* Associação entre duas variáveis categóricas e entre suas categorias: teste  $\chi^2$  e análise dos resíduos
    - \* Decomposição inercial: determinação de autovalores
    - \* Definição das coordenadas (scores) das categorias no mapa perceptual
  - Análise de correspondência múltipla
    - \* Notação
    - \* Análise de correspondência múltipla (ACM)
  - Configurações do mapa perceptual de uma análise de correspondência simples

# Técnicas Multivariadas

## Confirmatórias ou de Modelagem

### Treinamento 10: Modelos Lineares Generalizados - Regressão linear simples e múltipla

- Modelos de Regressão Linear Simples e Múltipla
  - Modelos lineares de regressão
    - \* Estimação do modelo de regressão linear por mínimos quadrados ordinários
    - \* Poder explicativo do modelo de regressão:  $R^2$
    - \* Significância geral do modelo e dos parâmetros
    - \* Intervalos de confiança dos parâmetros do modelo e previsões
    - \* Estimação de modelos lineares de regressão múltipla
    - \* Variáveis dummy em modelos de regressão
  - Pressupostos dos modelos de regressão por mínimos quadrados ordinários (MQO ou OLS)
    - \* Normalidade dos resíduos
    - \* Multicolinearidade
      - Causas da multicolinearidade
      - Consequências da multicolinearidade
      - Diagnósticos de multicolinearidade
      - Possíveis soluções para o problema da multicolinearidade
    - \* Heterocedasticidade
      - Causas da heterocedasticidade
      - Consequências da heterocedasticidade
      - Diagnóstico de heterocedasticidade: teste de Breusch–Pagan/Cook–Weisberg
      - Método de mínimos quadrados ponderados
      - Método de Huber–White para erros-padrão robustos
    - \* Autocorrelação dos resíduos
      - Causas da autocorrelação dos resíduos
      - Consequências da autocorrelação dos resíduos
      - Diagnóstico de autocorrelação dos resíduos: teste de Durbin–Watson
      - Diagnóstico de autocorrelação dos resíduos: teste de Breusch–Godfrey
      - Possíveis soluções para o problema da autocorrelação dos resíduos
    - \* Problemas de especificação: linktest e teste RESET

## Treinamento 11: Modelos Lineares Generalizados - Regressão Logística Binária e Multinomial

- Modelos de Regressão Logística Binária e Multinomial
  - Regressão logística binária
    - \* Estimação do modelo de regressão logística binária por máxima verossimilhança
    - \* Significância estatística geral do modelo e dos parâmetros da regressão logística binária
    - \* Construção dos intervalos de confiança dos parâmetros do modelo de regressão logística binária
    - \* Cutoff, análise de sensibilidade, eficiência global do modelo, sensibilidade e especificidade
    - \* Odds e odds ratio
    - \* Pressupostos na regressão logística
    - \* Estatísticas de razão de verossimilhança
    - \* Teste de Wald
  - Regressão logística multinomial
    - \* Estimação do modelo de regressão logística multinomial por máxima verossimilhança
    - \* Significância estatística geral do modelo e dos parâmetros da regressão logística multinomial
    - \* Construção dos intervalos de confiança dos parâmetros do modelo de regressão logística multinomial
  - Modelos de regressão probit

## Treinamento 12: Modelos Lineares Generalizados para Dados de Contagem - Regressão de Poisson e Binomial Negativo

- Regressão Poisson
  - Estimação por máxima verossimilhança
  - Significância estatística geral e dos parâmetros
  - Intervalos de confiança dos parâmetros
  - Teste para verificação de superdispersão em modelos de regressão Poisson
- Regressão binomial negativo
  - Estimação por máxima verossimilhança
  - Significância estatística geral e dos parâmetros
  - Intervalos de confiança dos parâmetros
- Modelos de regressão inflacionados de zeros

## Treinamento 13: Regressão para Dados em Painel em modelos longitudinais

- Modelos Longitudinais de Regressão para Dados em Painel
- Dados longitudinais e decomposição de variância
- Modelos longitudinais lineares
  - Estimação de modelos longitudinais lineares de regressão para dados em painel curto
  - Estimação de modelos longitudinais lineares de regressão para dados em painel longo
- Modelos longitudinais não lineares
  - Estimação de modelos longitudinais logísticos
  - Estimação de modelos longitudinais Poisson e binomial negativo

## Treinamento 14: Regressão para Dados em Painel em modelos multinível

- Modelos Multinível de Regressão para Dados em Painel
- Estruturas aninhadas de dados
- Modelos hierárquicos lineares
  - Modelos hierárquicos lineares de dois níveis com dados agrupados (HLM2)
  - Modelos hierárquicos lineares de três níveis com medidas repetidas (HLM3)
- Modelos hierárquicos não lineares