ACABA PELO AMOR DE DEUS

Testes Estatísticos no R (AULA 10)

Os testes estatísticos são ferramentas fundamentais para avaliar hipóteses e tirar conclusões sobre populações com base em amostras. Eles podem ser realizados no R de maneira eficiente. Abaixo estão os principais testes apresentados:

---

1. Teste de Hipóteses

- Avalia se uma hipótese sobre um parâmetro populacional é verdadeira.
- **Hipótese nula (H_0):** Afirmativa inicial (ex.: média é igual a 30).
- **Hipótese alternativa (H_1):** Contrária à hipótese nula (ex.: média é diferente de 30).
- **Nível de significância (α):** Probabilidade de rejeitar H_0 quando ela é verdadeira. Exemplo: 5% ou 0,05.
- **p-value:** Probabilidade associada ao teste. Rejeita-se H_0 se p-value ≤ α.

---

2. Teste t

- Compara médias em diferentes contextos. Usado quando a amostra é pequena e o desvio padrão é desconhecido.

2.1. Teste t para uma média

Avalia se a média de uma amostra difere de um valor específico.

Código:
```
t.test(amostra, mu = valor_esperado)
```

2.2. Teste t para duas médias (amostras independentes)

Compara médias de dois grupos diferentes.

Código:
```
t.test(grupo1, grupo2, var.equal = TRUE)
```

2.3. Teste t pareado

Compara antes e depois de uma intervenção no mesmo grupo.

Código:
```
t.test(antes, depois, paired = TRUE)
```

---

3. Teste F

- Compara variâncias entre dois grupos para verificar homogeneidade.

- Hipóteses:
    - H_0: Variâncias são iguais.
    - H_1: Variâncias são diferentes.

Código:
```
var.test(grupo1, grupo2)
```

---

4. Teste para Normalidade

- Avalia se os dados seguem uma distribuição normal.
- **Teste Shapiro-Wilk:**
    - H_0: Dados seguem distribuição normal.
    - H_1: Dados não seguem distribuição normal.

Código:
```
shapiro.test(dados)
```

- **Gráfico QQ (qqnorm):** Verifica visualmente a normalidade.

Código:
```
qqnorm(dados)

qqline(dados)
```

Interpretação de Resultados

- **Rejeitar H_0:** Dados indicam que a hipótese nula provavelmente não é verdadeira.
- **Não rejeitar H_0:** Dados não fornecem evidências suficientes para rejeitar H_0.

Análise de variância (parte 1- AULA 11)

A ANOVA é um método estatístico usado para comparar médias de mais de dois grupos. Seu objetivo é determinar se fatores (variáveis independentes) influenciam uma variável de interesse (dependente).

- **Hipóteses:**
    - **H_0:** As médias de todos os grupos são iguais (não há efeito do fator).
    - **H_1:** Pelo menos uma média é diferente (o fator influencia a variável dependente).
- **Condições:**

1. As populações devem seguir distribuição normal.

2. As variâncias dos grupos devem ser homogêneas.
3. As amostras são independentes.

1. Delineamento Inteiramente Casualizado (DIC)

- Usado em experimentos com um único fator.
- Cada tratamento (ou nível do fator) é aplicado aleatoriamente às unidades experimentais.

Teste F:

- Compara a variância entre os grupos (devida ao tratamento) com a variância dentro dos grupos (erro).
- **F calculado** é comparado com **F tabelado**:
  - **Fcal ≤ Ftab:** Não rejeitar H_0 (as médias são iguais).
  - **Fcal > Ftab:** Rejeitar H_0 (há diferença entre as médias).

Exemplo:

- Comparar as vendas de três vendedores.
- Resultado: Se o p-valor > nível de significância (α), não há diferença significativa entre os vendedores.

---

2. Delineamento em Blocos Casualizados (DBC)

- Usado quando há heterogeneidade entre os grupos (blocos), como no caso de experimentos agrícolas.
- Divide-se o material em blocos homogêneos, mas os tratamentos são aleatórios dentro de cada bloco.
- Também chamado de **Two-Way ANOVA** (dois fatores).

---

3. Delineamento em Quadrado Latino

- Controle de duas fontes de variação sistemáticas (linhas e colunas), como no teste de desgaste de pneus com diferentes marcas e posições.

---

4. Experimentos Fatoriais

- Avaliam simultaneamente dois ou mais fatores, verificando não apenas seus efeitos individuais, mas também possíveis **interações** entre eles.
- Exemplo: Estudo sobre distância e ângulo de visada em medições.
  - **Fator principal:** Distância ou ângulo.

    - **Interação:** Verifica se os dois fatores atuam juntos de forma significativa.

Procedimentos no R

- **Definição de fatores:**

```r
fator <- factor(c("A", "B", "C"))
```

- **Análise de Variância:**

```r
resultado <- aov(variavel ~ fator, data = dados)
summary(resultado)
```

- **Interpretação:**
    - Verifique o p-valor:
        - p-valor $\leq \alpha$: Rejeitar H_0 (diferença significativa).
        - p-valor $> \alpha$: Não rejeitar H_0 (sem diferença significativa).

Análise de variância - ANOVA (parte 2 – AULA 12)

- Após identificar que há diferenças significativas entre as médias através do teste F na ANOVA, precisamos saber **quais médias** são diferentes.
- Para isso, usamos **testes de comparação de médias**, como:
    - **Teste de Tukey (HSD) –** mais utilizado.
    - Teste de Duncan.
    - Teste de Scheffé.
    - Teste de Dunnet.
    - Teste de Bonferroni.

1. Teste de Tukey

Características:

- Desenvolvido por **John Wilder Tukey** (1949).
- Comparação rigorosa de médias entre todos os pares.
- Considera a **taxa de erro do conjunto dos testes** como exatamente igual ao nível de significância (α).

Aplicações:

- Usado após uma ANOVA significativa (quando o teste F detecta diferenças).
- Verifica se há **diferenças significativas entre pares de médias de tratamentos**.

Pressupostos:

1. **Independência:** Observações independentes dentro e entre grupos.
2. **Distribuição Normal:** Os grupos devem seguir distribuição normal.
3. **Homogeneidade de variâncias:** A variância dentro dos grupos deve ser constante.

Como Realizar o Teste de Tukey no R

1. **Realize uma ANOVA:**

```r
resultado <- aov(variavel ~ fator, data = dados)
```

2. **Execute o Teste de Tukey:**

```r
TukeyHSD(resultado)
```

3. **Visualize os Resultados Graficamente:**

```r
plot(resultado, cex = 0.6)
```

Interpretação dos Resultados:

- **Intervalo contém o zero:** As médias do par são consideradas **iguais** (não há diferença significativa).
- **Intervalo não contém o zero:** As médias do par são **diferentes** (diferença significativa ao nível de confiança).

Regressão (AULA 13)

A regressão linear simples tem como objetivo estimar uma equação que relacione matematicamente duas variáveis, sendo que uma delas é explicada pela outra.

- A variável explicada geralmente é denominada variável resposta ou variável dependente (Y).
- A variável explicativa é denominada variável explanatória independente (X).

- **Diagrama de Dispersão:**
  - Quando os pontos se agrupam em torno de uma reta, indica relação linear.
- **Exemplo:** Estudar a dilatação de um pilar de concreto em função da temperatura.

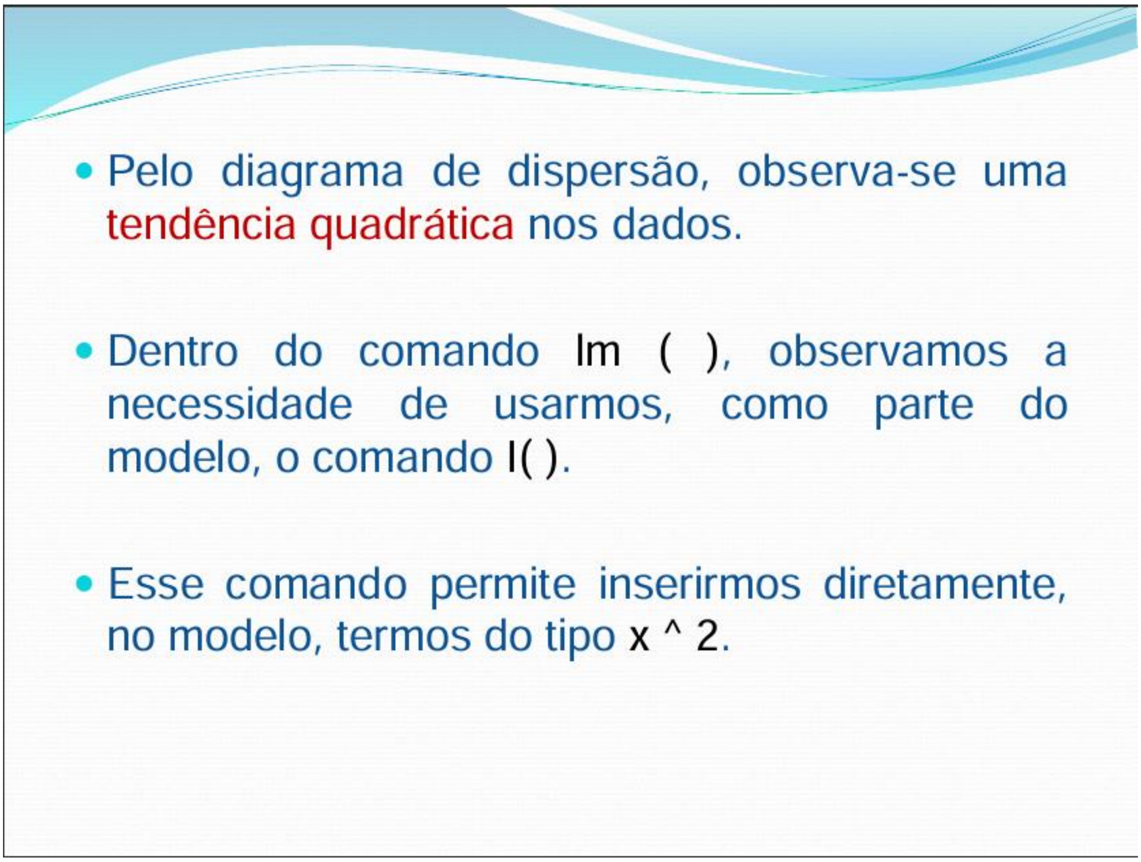
r Copiar código

```r
modelo <- lm(dilat ~ temp, data = dados)
summary(modelo)
plot(temp, dilat)
abline(modelo)
```

- **Resultados:**
  - Coeficiente de determinação (R^2): Mede o ajuste do modelo aos dados.
  - Resíduos: Diferença entre valores observados e previstos.

De Grau Maior que 1

Da mesma forma que o modelo linear ajustado, qualquer modelo de regressão polinomial pode ser obtido com o comando lm( ),que vem do inglês linear models.



- Pelo diagrama de dispersão, observa-se uma tendência quadrática nos dados.

- Dentro do comando lm ( ), observamos a necessidade de usarmos, como parte do modelo, o comando I( ).

- Esse comando permite inserirmos diretamente, no modelo, termos do tipo x ^ 2.

```
> reg<-lm(                    #ajusta uma regressão
+   prd~fert+I(fert^2)) #modelo quadrático
> reg                         #exibe o modelo ajustado

Call:
lm(formula = prd ~ fert + I(fert^2))

Coefficients:
(Intercept)         fert    I(fert^2)
   15.51667      2.95720     -0.02716
```

c b a

$$y = ax^2+bx+c$$

2. Regressão Polinomial Múltipla

- **Definição:** Uma variável resposta (Y) é relacionada a duas ou mais variáveis explicativas ($X_1, X_2, \ldots$).
- **Modelo Matemático:**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon$$

- **Exemplo:** Prever o preço de revenda de carros com variáveis como:
  - Idade.
  - Quilometragem.
  - Condição de pneus, entre outros.

r Copiar código

```r
modelo <- lm(preco ~ idade + km + consumo, data = dados)
summary(modelo)
```

3 - Modelos Não Polinomiais

Apesar de os modelos polinomiais serem úteis em muitas situações, há casos em que a disposição dos pontos no diagrama de dispersão, ou mesmo o problema do qual os dados foram obtidos, indique a necessidade ou exigência de modelos mais específicos.

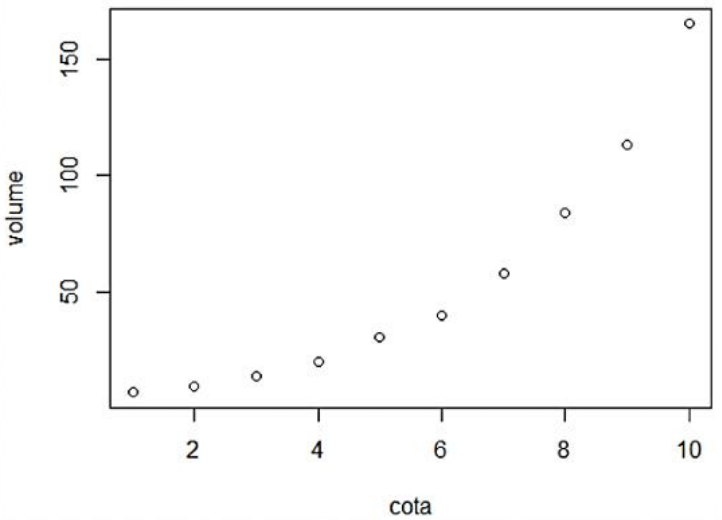
Modelo Exponencial Exemplo:

Num projeto de construção de urna barragem é de grande interesse equacionar a relação entre a cota do nível de água e o volume armazenado quando esta cota é atingida.

Essa relação é obtida a partir de um diagrama cota volume, estimado através do levantamento topográfico, com suas respectivas curvas de nível, da região onde será construída a barragem.

Suponha os dados a seguir, com a, cota dada em metros e o volume em quilômetros cúbicos:

```
cota<-c(1,2,3,4,5,6,7,8,9,10)
volume<-c(7,10,14,20,31,40,58,84,113,165)
dados<-data.frame(cota,volume)
plot(dados)
```



- Apesar de ser possível ajustar um modelo polinomial para os dados em questão, um modelo mais apropriado seria baseado na função $y = a.e^{b.x}$

```
funcao<-volume~a*exp(b*cota) #modelo
exponencial<-nls(            #ajust. modelos não lineares
  funcao,                    #modelo a ajustar
  dados,                     #conjunto de dados
  start = c(a=1,b=1))        #valores iniciais dos estimadores
exponencial
```

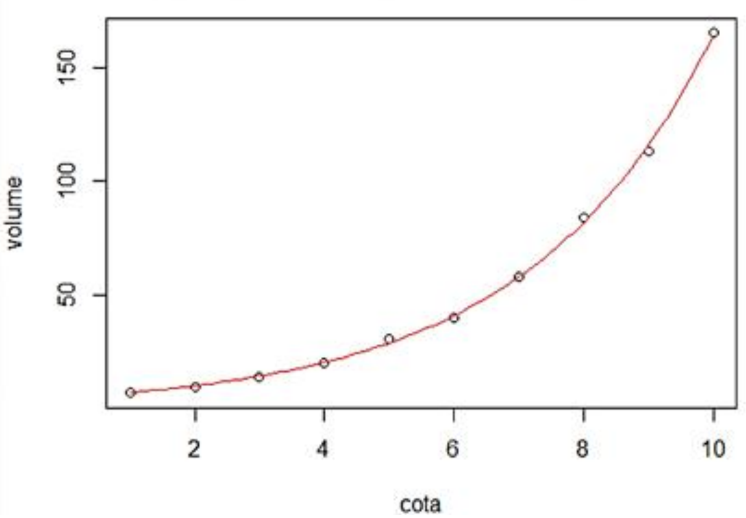
```
Nonlinear regression model
  model: volume ~ a * exp(b * cota)
   data: dados
     a      b
5.1164 0.3467
 residual sum-of-squares: 19.44

Number of iterations to convergence: 14
Achieved convergence tolerance: 2.694e-07
```

- A obtenção de estimativa dos mínimos quadrados dos coeficientes se dá através do comando nls ( ), do inglês *Nonlinear Least Squares*.

- Comando ??regression no console abre mais opções de modelo de regressão.

- Desenhando a curva ajustada:

```
curve(                          #desenha a curva
  5.1164*exp(0.3467*x),         #equação ajustada
  1,                            #lim. inf. do eixo das abcissas
  10,                           #limite superior
  add = T,                      #acrescentar no gráfico anterior
  col=2)                        #cor vermelha
```



Dicas Gerais para Regressão no R

- **Gráficos:** Sempre visualize os dados com plot() para identificar padrões.
- **Resumo do Modelo:** Use summary() para avaliar os coeficientes e R2R^2R2.
- **Resíduos:** Verifique resíduos para garantir qualidade do ajuste.