Estatística Básica e Introdução ao R

Profª. Dra. Natalia Giordani

3. Testes de hipóteses

- Ferramenta utilizada para responder afirmações que se desejam verificar
  - Formulação da hipótese
  - Levantamento de dados (amostra)
  - Análise para confirmar, ou não, a hipótese testada
  - Resultado carrega uma probabilidade de erro

3. Testes de hipóteses

- Exemplo
  - Hipótese de pesquisa: alteração no fluxo de compra do site resulta em uma maior conversão de compra do serviço oferecido
  - Transformação da hipótese de “pesquisa” em hipótese estatística
    - Hipótese estatística é uma suposição que é feita sobre o parâmetro de interesse da população e que pode ser rejeitada ou não rejeitada, de acordo com o que observamos na amostra
    - A decisão de rejeitar ou não uma hipótese é feita levando em consideração o nível de significância adotado e o p-valor

3. Testes de hipóteses

- Tipos de hipóteses estatísticas
  - Hipótese nula (H_0): estabelece a ausência de diferença entre os parâmetros
  - Hipótese alternativa (H_1): é a hipótese contrária a hipótese nula.
- Teste de hipóteses é um procedimento estatístico através do qual se rejeita ou não uma hipótese associando à conclusão um risco máximo de erro
  - Erro de afirmar que existe uma diferença quando ela efetivamente não existe (α)

3. Testes de hipóteses

- Tipos de testes
  - Bilateral: hipótese nula considera a igualdade e a hipótese alternativa, a diferença
    - H_0: conversão compra fluxo novo = conversão compra fluxo atual
    - H_1: conversão compra fluxo novo ≠ conversão compra fluxo atual
  - Unilateral: hipótese alternativa assume uma direção (< , > , ≤ , ≥)
    - H_0: conversão compra fluxo novo = conversão compra fluxo atual
    - H_1: conversão compra fluxo novo > conversão compra fluxo atual

3. Testes de hipóteses

- Tipos de erro

Verdade	Conclusão do teste	
	Não se rejeita H_0	Rejeita-se H_0
H_0 é verdadeira	Decisão correta Probabilidade: $1 - \alpha$	Decisão errada **Erro tipo I** Probabilidade: α
H_0 é falsa	Decisão errada **Erro tipo II** Probabilidade: β	Decisão correta Probabilidade: $1 - \beta$ = Poder do teste

3. Testes de hipóteses

- Valor p
  - Sinônimo: nível descritivo amostral
  - **Refere-se a probabilidade de se obter uma estatística de teste igual ou mais extrema que aquela observada em uma amostra supondo a veracidade da hipótese nula**
  - $p < \alpha$: rejeitamos H_0
    - Resultado estatisticamente significativo
    - Valores usuais para α: 0,01; **0,05**
  - Definições baseadas apenas em valor p = **PERIGO!**

3. Testes de hipóteses

- Tipos de testes
  - Paramétricos
    - Exigem que algumas suposições sejam cumpridas para garantir resultados confiáveis
    - Em geral: suposição de normalidade da variável estudada
  - Não paramétricos
    - Não fazem suposições quanto à distribuição da variável
    - Comumente utilizados em amostras menores

3.1 Teste t – amostras independentes

- Objetivo
  - Comparar **médias** de duas **amostras independentes**
- Hipóteses
  - H_0: $\mu_A = \mu_B$ **ou** $(\mu_A - \mu_B) = 0$
  - H_1: $\mu_A \neq \mu_B$ **ou** $(\mu_A - \mu_B) \neq 0$
- Cuidados com uso desse teste
  - Distribuição da variável em cada um dos grupos deve ser Normal
  - Homogeneidade de variâncias

3.1 Teste t – amostras independentes

- Exemplo: duas técnicas de vendas (A e B) são aplicadas por dois grupos distintos de vendedores. A hipótese da empresa é de que a técnica B produza resultados melhores.

	Técnica A	Técnica B
N. médio de vendas	67	76
N. vendedores	50	50

H_0: μ_A - μ_B = 0

```
> t.test(vendas  ~ tecnica, data =  exemplo_vendas)

        Welch Two Sample t-test

data:  vendas by tecnica
t = -12.841, df = 13.722, p-value = 4.941e-09
alternative hypothesis: true difference in means between group A and group B is not equal to 0
95 percent confidence interval:
 -10.62283  -7.57717
sample estimates:
mean in group A mean in group B
           66.5            75.6
```

REJEITAMOS H_0

Conclusão: há evidências significativas ($p < 0,05$) de que a técnica B produz melhores resultados do que a A!

3.2 Teste de Wilcoxon ou Mann-Whitney

- Objetivo
  - Comparar duas **amostras independentes** em termos de localização (mediana ou média)
- Hipóteses
  - H_0: $M_A = M_B$ **ou** $(M_A - M_B) = 0$
  - H_1: $M_A \neq M_B$ **ou** $(M_A - M_B) \neq 0$
- Sobre esse teste
  - Baseado nos postos dos valores obtidos combinando-se as duas amostras
  - Não faz suposições sobre a distribuição

3.2 Teste de Wilcoxon ou Mann-Whitney

- Exemplo: um fazendeiro está interessado em verificar se o uso de um novo fertilizante (FN) aumenta a produção de milho quando comparado ao fertilizante atual (FA). Para isso, plantou 20 canteiros com cada fertilizante e registrou suas produções (sacas).

	FA	FN
Produção mediana	7	6,8

H_0: M_A - M_B = 0

```
> wilcox.test(producao ~ fertilizante, data = exemplo_wilcoxon, exact = FALSE)

        Wilcoxon rank sum test with continuity correction

data:  producao by fertilizante
W = 68, p-value = 0.1785
alternative hypothesis: true location shift is not equal to 0
```

Conclusão: há evidências de que a produção de milho dois fertilizantes é similar (p = 0,18).

NÃO REJEITAMOS H_0

3.3 Teste t – amostras dependentes

- Objetivo
  - Comparar **médias** de duas **amostras dependentes**
- Hipóteses
  - H_0: $\mu_D = 0$
  - H_1: $\mu_D \neq 0$
- Cuidados com uso desse teste
  - Distribuição das diferenças deve ser Normal

3.3 Teste t – amostras dependentes

- Exemplo: operadores de certo tipo de máquina são treinados em máquinas de marcas diferentes. O tempo gasto na realização da mesma tarefa em cada máquina foi medido. Qual das duas máquinas a empresa deve comprar objetivando ser mais eficiente?

Operador	Máquina A	Máquina B	Diferença
1	80	75	5
2	72	70	2
...	...	...	...

H_0: μ_D = 0

- Informação complementar: μ_A = 76; μ_B = 71

```
> t.test(Pair(maq_a, maq_b) ~ 1, data = exemplo_tpareado)

        Paired t-test

data:  Pair(maq_a, maq_b)
t = 5.9761, df = 4, p-value = 0.00394
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.677059 7.322941
sample estimates:
mean of the differences
                      5
```

REJEITAMOS H_0

Conclusão: há evidências significativas (p < 0,05) de que o tempo gasto na realização da mesma tarefa não é igual entre as máquinas A e B (demora-se mais na máquina A).

3.4 Teste dos postos sinalizados de Wilcoxon

- Objetivo
  - Comparar duas **amostras dependentes** em termos de localização (mediana ou média)
- Hipóteses
  - H_0: $M_D = 0$
  - H_1: $M_D \neq 0$
- Sobre esse teste
  - Não faz suposições sobre a distribuição

3.4 Teste dos postos sinalizados de Wilcoxon

- Exemplo: um psicólogo do esporte, a fim de verificar se ouvir música afeta a duração das sessões de exercícios de atletas, registrou o tempo de 10 sessões de exercícios enquanto ouviam e não ouviam música.

Atleta	Com música	Sem música	Diferença
1	45	38	7
2	38	40	-2
...	...	...	...

H_0: M_D = 0

- Informação complementar: M_{com} = 43; M_{sem} = 39

```
> wilcox.test(x = wilcox_pareado$musica_sim, y = wilcox_pareado$musica_nao, paired = TRUE, exact = FALSE)

        Wilcoxon signed rank test with continuity correction

data:  wilcox_pareado$musica_sim and wilcox_pareado$musica_nao
V = 47.5, p-value = 0.04657
alternative hypothesis: true location shift is not equal to 0
```

REJEITAMOS H_0

Conclusão: há evidências significativas (p = 0,047) de que ouvir música aumente a duração das sessões.

3.5 Teste para comparação de proporções

- Objetivo
  - Comparar **proporções** de duas populações com mesmo atributo
- Hipóteses
  - H_0: $p_1 = p_2$
  - H_1: $p_1 \neq p_2$

3.5 Teste para comparação de proporções

- Exemplo: para lançamento da nova embalagem do shampoo X a área de criação estuda duas propostas, P1 e P2. Para isso, selecionou aleatoriamente clientes em duas lojas distintas e perguntou se havia notado o shampoo. Será que ambas as propostas são atraentes?

Proposta	Notaram?		Total
	Sim	Não	
1	168	232	400
2	180	420	600
	34,8%	68,2%	1000

H_0: $p_1 = p_2$

- Informação complementar: $p_1 = 42\%$; $p_2 = 30\%$

```
> prop.test(exemplo_prop)

	2-sample test for equality of proportions with continuity correction

data:  exemplo_prop
X-squared = 14.707, df = 1, p-value = 0.0001256
alternative hypothesis: two.sided
95 percent confidence interval:
 0.05722104 0.18277896
sample estimates:
prop 1 prop 2
  0.42   0.30
```

REJEITAMOS H_0

Conclusão: há evidências significativas (p = 0,047) de que as propostas não são igualmente atraentes. P1 é mais notada que P2.

3.6 Teste χ^2 de independência

- Objetivo
  - Verificar **associação** entre variável linha e variável coluna (variáveis qualitativas – tabelas de contingência)
- Hipóteses
  - H_0: não há associação entre as variáveis = são independentes
  - H_1: há associação entre as variáveis = não são independentes
- Cuidados com uso desse teste
  - Garantir que todas as caselas tenham valores > 5, caso contrário, optar pelo teste exato de Fisher
  - Para tamanhos de amostra pequenos, utilizar correção de continuidade de Yates

3.6 Teste χ^2 de independência

- Exemplo: companhia de seguro analisou frequência com que 2.000 segurados usaram hospitais e deseja verificar se o uso do hospital independe do sexo do segurado.

Usou hospital	Sexo	
	Feminino	Masculino
Sim	150	100
Não	850	900

H_0: não há associação entre as variáveis (são independentes)

```
> chisq.test(exemplo_chisq)

        Pearson's Chi-squared test with Yates' continuity correction

data:  exemplo_chisq
X-squared = 10.976, df = 1, p-value = 0.000923
```

REJEITAMOS H_0

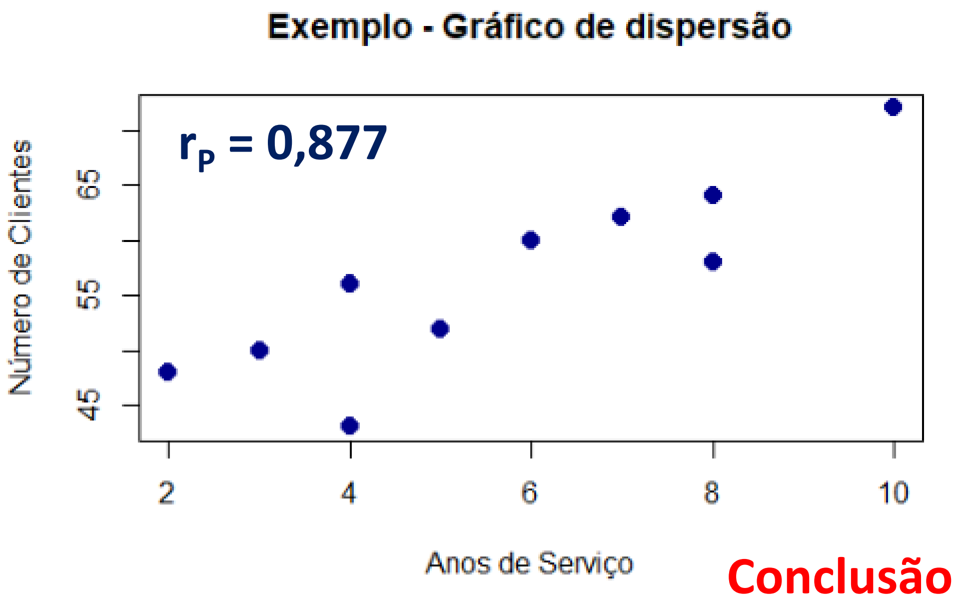
Conclusão: há evidências significativas ($p < 0,05$) de que sexo e uso de hospital não são independentes!

3.7 Teste para o coeficiente de correlação

- Objetivo
  - Determinar a existência de correlação entre as variáveis
- Hipóteses
  - H_0: $\rho = 0$ (variáveis são independentes)
  - H_1: $\rho \neq 0$
- Cuidados com uso desse teste
  - Suposição de distribuição Normal bivariada (variáveis avaliadas)
  - Correlação (e associação) não implica causalidade!

3.7 Teste para o coeficiente de correlação

- Exemplo: número de anos de serviço por número de clientes de agentes de uma companhia de seguro.




```
> cor.test(x = dados.ex4_4$anos_servico,
+          y = dados.ex4_4$n_clientes,
+          method = "pearson")

        Pearson's product-moment correlation

data:  dados.ex4_4$anos_servico and dados.ex4_4$n_clientes
t = 5.1573, df = 8, p-value = 0.0008665
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5517820 0.9705991
sample estimates:
      cor
0.8767952
```

REJEITAMOS H_0

Conclusão: há evidências significativas ($p < 0,05$) de que existe dependência entre anos de experiência e número de clientes.

3.8 ANOVA de 1 via (Anova one-way)

- Objetivo
  - Comparar **médias** de três ou mais **amostras independentes**
- Hipóteses
  - H_0: $\mu_1 = \mu_2 = \ldots = \mu_k = \mu$
  - H_1: $\mu_i \neq \mu_j$, para algum par (i, j)
- Cuidados com uso desse teste
  - Distribuição da variável quantitativa deve ser Normal
  - Variâncias amostrais semelhantes
    - Esse teste é robusto para desvios da suposição de homocedasticidade

3.8 ANOVA de 1 via (Anova one-way)

- Exemplo: num experimento sobre eficácia de um regime para emagrecer, homens, todos pesando cerca de 100 kg e de biotipos semelhantes, são submetidos a três regimes. Após um mês, verifica-se a perda de peso(kg) de cada um.

Dados:

ID	Redução peso	Regime
1	7,13	A
2	10,91	B
...	...	...

- Perdas médias na amostra: A = 6 kg; B = 12 kg; C = 13 kg

- H_0: $\mu_A = \mu_B = \mu_C$

```
> anova <- aov(reducao ~ regime, data = dados)
> summary(anova)
            Df Sum Sq Mean Sq F value   Pr(>F)
regime       2 172.93   86.47   35.42 2.08e-06 ***
Residuals   15  36.62    2.44
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

REJEITAMOS H_0

Qual regime se mostrou melhor?

Conclusão: há evidências significativas (p < 0,05) de que existe diferenças na redução média de peso entre os diferentes regimes.

3.8 ANOVA de 1 via (Anova one-way)

- **Qual regime se mostrou melhor?**
  - A comparação das diferenças entre as médias é realizada através dos testes de comparações múltiplas – Teste de Tukey

```
> TukeyHSD(anova, conf.level=.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = reducao ~ regime, data = dados)

$regime
        diff        lwr      upr     p adj
B-A 5.976667  3.633559 8.319775 0.0000228
C-A 7.043333  4.700225 9.386441 0.0000033
C-B 1.066667 -1.276441 3.409775 0.4810906
```

Comparação	Diferença entre médias	Conclusão
Regime A vs Regime C	13,2 – 6,16 = 7,04	Médias diferem
Regime A vs Regime B	12,1 – 6,16 = 5,94	Médias diferem
Regime B vs Regime C	13,2 – 12,1 = 1,1	Médias não diferem

Regimes B e C apresentaram os melhores resultados!

Perdas médias na amostra:
A = 6 kg
B = 12 kg
C = 13 kg

3.9 Teste de Kruskal-Wallis (ANOVA não paramétrica)

- Objetivo
  - Comparar três ou mais **amostras independentes** em termos de localização
- Hipóteses
  - H_0: $M_1 = M_2 = \ldots = M_K = M$
  - H_1: $M_i \neq M_j$, para algum par (i, j)
- Quando usar este teste?
  - Quando forem violadas de forma importante as suposições de normalidade e homocedasticidade
  - Complementação: teste de Dunn

3.9 Teste de Kruskal-Wallis (ANOVA não paramétrica)

- Exemplo: uma empresa deseja verificar se o perfil de consumo de um novo produto lançado em três grupos distintos de clientes (jovens, adultos e idosos) é igual. Para isso, registrou o número de vezes que cada cliente comprou o produto em um mês.

ID	Grupo	Qtde
1	Adolescente	20
2	Adulto	12
3	Idoso	8
...	...	...

- Quantidade mediana de compras/mês: adolescente = 5,5; adulto = 10,5; idoso = 9

- H_0: $M_{Adolescente} = M_{Adulto} = M_{Idoso}$

```
> kruskal.test(consumo ~ grupo, data = dados)

        Kruskal-Wallis rank sum test

data:  consumo by grupo
Kruskal-Wallis chi-squared = 0.88103, df = 2, p-value = 0.6437
```

NÃO REJEITAMOS H_0

Conclusão: não há evidências (p = 0,64) de que existe diferenças no consumo do produto entre os grupos.

Vamos a prática!

- Dados reais: a Eagle Boys é uma rede de pizzarias australiana que, há muitos anos, realiza uma campanha publicitária afirmando que suas pizzas grandes são maiores do que as da maior rede de pizzas australiana, a Domino's Pizza.

Campanha publicitária de fev/2012

Vamos a prática!

- Dentre as afirmações da campanha:
  1. Nossas pizzas grandes são maiores do que as deles – 29,17 cm x 27,44 cm de diâmetro
  2. Nossas pizzas grandes tem 12 polegadas (30,48 cm) de tamanho real

- Como essas afirmações podem ser verificadas?
- Que dados são necessários e como poderiam ser obtidos?

Vamos a prática!

- Dados disponibilizados:

Nome da variável	Descrição	Observação
ID	Identificador da pizza	
Company	Nome da rede de pizzarias	Dominos ou EagleBoys
CrustDescription	Tipo da massa	Tipos: classic , deep pan, mid crust, thin crust, thin Ncrispy
Topping	Sabor da pizza	Tipos: Hawaiian, Supreme, SuperSupremo, BBQ Meat lovers
Diameter	Diâmetro da pizza	Medido em centímetros

Vamos a prática!

- Afirmação 1:

#Independent laboratory testing conducted by A.C.M. Laboratory Pty Ltd from 26/5/11 to 1/6/11 [illegible]251 deep pan, mid/classic crust and thin crust large pizzas purchased from 21 stores across NSW, QLD, WA and SA, from 24/5/11 to 31[illegible], found that Eagle Boys' deep pan, mid crust and thin crust BBQ Meatlovers, Super Supremo and Hawaiian Large Pizzas were, on average, 29.17cm in diameter and Dominos' equivalent large pizzas were, on average, 27.44cm in diameter. A.C.M. Laboratory Pty Ltd is an independent provider of dimensional calibration services and is accredited by the National Association of Testing Authorities.