Pesquisa Quantitativa

Lupércio França Bessegato
Mestrado em Administração/UFJF

Comparações de Médias

Roteiro Geral

1. Introdução
2. Coleta de dados
3. Modelos probabilísticos
4. Distribuições amostrais e estimação
5. Testes de significância
6. **Comparações de médias**
7. Tabelas de contagem
8. Análise de regressão
9. Referências

Roteiro do Módulo

5. Comparações de médias:
   a) Amostras emparelhadas
   b) Testes não paramétricos
   c) Duas amostras independentes
   d) Mais de duas amostras

Amostras Emparelhadas

Amostras Emparelhadas

- As amostras são dependentes
  - √ Observações combinam-se em pares
    - – Observação na amostra 1 de equipara com observação na amostra 2
    - – Membros dos pares estão relacionados de alguma maneira
  - √ Dados de pares emparelhados
  - √ Dependência ocorre geralmente quando cada amostra tem os mesmos sujeitos

Exemplos

- Estudos observacionais longitudinais
  - √ Mesmos sujeitos fazem parte dos dois grupos (antes e após um período)
  - √ Exemplo:
    - – Condicionamento físico antes e 4 semanas após início de treinamento de ciclismo

- Estudos experimentais com medidas repetidas
  - √ Sujeitos fazem parte dos dois grupos (antes e após um "tratamento")
  - √ Exemplo:
    - – Teste de desempenho de veículos com dois tipos diferentes de combustível

Importante

- Sempre deve-se examinar cuidadosamente como os dados foram produzidos para ver se há duas amostras independentes ou duas amostras emparelhadas (pareadas)

Análise de Amostras Emparelhadas

- Analisam-se as diferenças entre as medidas observadas por sujeito
  - √ Transforma-se em problema de uma amostra
  - √ Representam-se graficamente as diferenças
  - √ Compara-se a diferença da média dos dois grupos, aplicando teste t no conjunto de diferenças

Exemplo

- Uso de telefone celular e tempo de reação
  - √ Amostras com 32 estudantes universitários
  - √ Simulação de situações de direção (com e sem celular), em que alvo brilhava na cor verde ou vermelha, em períodos irregulares
    - Pisar no freio o mais rápido possível com luz verde
  - √ Resposta:
    - Tempo médio de reação em várias tentativas
- Investigar se o uso do telefone celular prejudica a reação dos motoristas

- Tempos de reação (em ms)

Usou Celular?				Usou Celular?		
Sim	Não	Diferença	Estudante	Sim	Não	Diferença
636	604	32	17	626	525	101
623	556	67	18	501	508	-7
615	540	75	19	574	529	45
672	522	150	20	468	470	-2
601	459	142	21	578	512	66
600	544	56	22	560	487	73
542	513	29	23	525	515	10
554	470	84	24	647	499	148
543	556	-13	25	456	448	8
520	531	-11	26	688	558	130
609	599	10	27	679	589	90
559	537	22	28	960	814	146
595	619	-24	29	558	519	39
565	536	29	30	482	462	20
573	554	19	31	527	521	6
554	467	87	32	536	543	-7

- Estatísticas Descritivas - Saída R

```
> library(pastecs)# fornece resumo de estatísticas descritivas
> aggregate(tempo ~ uso, data = celular, stat.desc, basic = F)

uso tempo.median   tempo.mean tempo.SE.mean tempo.CI.mean.0.95    tempo.var
1   N  527.0000000  534.5625000    11.7463908         23.9569220 4415.2862903
2   S  569.0000000  585.1875000    15.8473335         32.3208497 8036.4153226
  tempo.std.dev tempo.coef.var
1    66.4476207      0.1243028
2    89.6460558      0.1531920
```

- Estatísticas descritivas – Quadro resumo

Desvio = Tempo de reação com celular – tempo sem celular

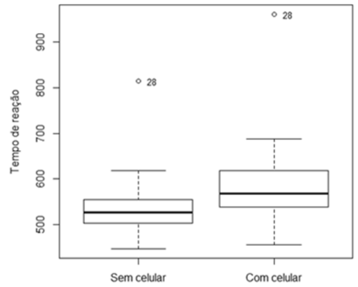
Grupos	$\bar{x}$	s_X
Com celular	585,2	66,45
Sem celular	534,6	89,65
Diferença	50,6	52,49

$$\text{ep}(\bar{x}_D) = \frac{52,48579}{\sqrt{32}} = 9,278 \ .$$

13

Pesquisa Quantitativa - 2016

- Box-plot

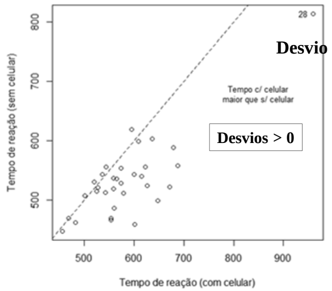


√ O estudante 28 é um valor atípico

14

Pesquisa Quantitativa - 2016

- Plot dos tempos por sujeito



Desvio = Tempo com celular – tempo sem celular

√ Estudante 28 é claramente atípico

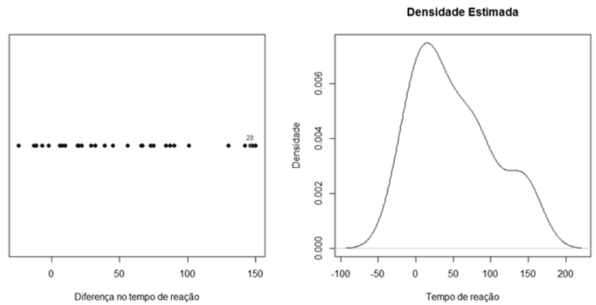
√ Há diferenças negativas!

– Tempo médio de reação com celular é menor que sem celular!

15

Pesquisa Quantitativa - 2016

- Distribuição das diferenças:



√ O estudante 28 não tem uma diferença atípica

√ Pequena anomalia na densidade

– Há um grupo com maior diferença de reação

16

Pesquisa Quantitativa - 2016

- Verificação normalidade das diferenças


```
> shapiro.test(dif)

 Shapiro-Wilk normality test

data:  dif
W = 0.92194, p-value = 0.02348
```

√ Dados apresentam desvios (extremidades)

√ Forte evidência de dados não serem normais

- Parâmetro de interesse:
  - √ $\mu_D = \mu_C - \mu_S$
  - √ μ_C: tempo médio de reação da população de motorista com uso de celular
  - √ μ_S: tempo médio de reação da população de motorista sem uso de celular
- Hipóteses:
  - √ H_0: $\mu_D = 0$
  - √ H_1: $\mu_D \neq 0$

- Considerações para uso do teste t:
  - √ É robusto no presente caso, pois:
    - Hipótese é bilateral
    - Sem presença de valores atípicos
    - Amostra de tamanho moderado sem assimetria pronunciada

- Execução do teste – comando:

```
        # Criação do objeto com os resultados do teste t emparelhado

celular$uso <- relevel(celular$uso, "S")   # Dá precedência ao nível "S"
dif.t <- with(celular, t.test(tempo ~ uso, alternative = "two.sided",
                              mu = 0, paired = TRUE, conf.level = 0.95))
```

- Resultado - saída:

```
> dif.t

 Paired t-test

data:  tempo by uso
t = 5.4563, df = 31, p-value = 5.803e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 31.70186 69.54814
sample estimates:
mean of the differences
                 50.625
```

```
> dif.t

 Paired t-test

data:  tempo by uso
t = 5.4563, df = 31, p-value = 5.803e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 31.70186 69.54814
sample estimates:
mean of the differences
                 50.625
```

- Conclusão:
  - √ Existe forte evidência de que o tempo médio de reação é maior quando o motorista usa o celular
  - √ Tempo médio de reação da população usando celular está entre 32 e 70 ms acima do que aquele sem o uso do celular

21

Pesquisa Quantitativa - 2016

- Teste t sem o estudante 28:

```
> > with(subset(celular, estudante != 28), t.test(tempo[uso == "S"], tempo[uso ==
"N"], mu = 0,
+     alternative = "two.sided", paired = TRUE))

 data:  tempo[uso == "S"] and tempo[uso == "N"]
t = 5.2595, df = 30, p-value = 1.119e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 29.08544 66.01133
sample estimates:
mean of the differences
               47.54839
```

Estatísticas amostrais	**Diferenças (completo)**	**Diferenças (sem est. 28)**
$\bar{x}_D$	50,63	47,55
t_0	5,456	5,260
p-valor	5,8e-6	1,1e-5
IC com 95%	[37,7; 69,5]	[29,1; 66,0]

Conclusão não é afetada pelo estudante 28

22

Pesquisa Quantitativa - 2016

Importante

- Sujeitos no experimento provavelmente eram de uma amostra voluntária
  - √ Conclusões inferenciais são aproximadas

23

Pesquisa Quantitativa - 2016

Exemplo

- Examinando pares de medidas de cabeças
  - √ Medição das cabeças de 18 pilotos de caça para compra de capacetes
  - √ Interesse:
    - Determinar a adequação de uso de calibrador de cartolina (barato) em vez de calibrador metálico (caro e desconfortável)
- Há diferença sistemática nas medidas produzidas pelos dois tipos de calibradores?

24

Pesquisa Quantitativa - 2016

- Estatísticas Descritivas - Saída R

```
> > apply(cabecas[-1], 2, stat.desc, basic = F)

                cartolina        metal
median       153.50000000 152.50000000
mean         154.55555556 152.94444444
SE.mean        1.37251790   1.30477309
CI.mean.0.95   2.89575964   2.75283060
var           33.90849673  30.64379085
std.dev        5.82310027   5.53568341
coef.var       0.03767642   0.03619408
```

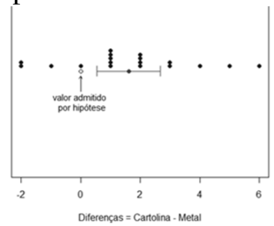
- Estatísticas descritivas – Quadro resumo

Desvio = Cartolina – Metal

Grupos	$\overline{x}$	s_X
Cartolina	154,56	5,82
Metal	152,94	5,54
Diferença	1,611	2,146

$$\text{ep}(\bar{x}_D) = \frac{2.145827}{\sqrt{18}} = 0,506 \ .$$

- Gráfico de pontos



√ Há pontos empilhados no gráfico
  - Dados arredondados até mm mais próximo

√ Centro dos dados está à direita do 0

√ Gráfico não dá motivo para duvidar da aplicabilidade de procedimentos t

- Verificação normalidade das diferenças


```
> shapiro.test(dif)

 Shapiro-Wilk normality test

data:  dif
W = 0.95735, p-value = 0.5515
```

√ Amostra pequena com empates

√ Teste e gráfico não revelam qualquer evidência contra suposição de normalidade

- Parâmetro de interesse:
  - √ $\mu_D = \mu_C - \mu_M$
  - √ μ_C: medidas com calibrador de cartolina (mm)
  - √ μ_M: medidas com calibrador metálico (mm)
- Hipóteses:
  - √ H_0: $\mu_D = 0$
  - √ H_1: $\mu_C \neq 0$

- Execução do teste – comando:

```
        # Criação do objeto com os resultados do teste t emparelhado

dif.t <- with(cabecas, t.test(cartolina, metal, alternative = "two.sided",
                              mu = 0, paired = TRUE, conf.level = 0.95))
```

- Resultado - saída:

```
> dif.t

 Paired t-test

data:  cartolina and metal
t = 3.1854, df = 17, p-value = 0.005415
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.5440163 2.6782060
sample estimates:
mean of the differences
               1.611111
```



```
> dif.t

 Paired t-test

data:  cartolina and metal
t = 3.1854, df = 17, p-value = 0.005415
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.5440163 2.6782060
sample estimates:
mean of the differences
               1.611111
```

- Conclusão:
  - √ Evidência muito forte de que o tipo de calibrador faz diferença
  - √ Com 95% de confiança, os calibradores de cartolina fornecem medidas maiores em média que as dos calibradores metálicos, por entre 0,5 mm e 2,7 mm.

Testes Não Paramétricos

Testes Não Paramétricos

- Não fazem qualquer suposição a respeito da distribuição que produz os dados
  - √ Tendem a não ser sensíveis à presença de *outliers*
- Testes paramétricos:
  - √ Dependem da distribuição subjacente dos dados
  - √ O teste t é um teste paramétrico
    - – Considera que a distribuição subjacente é normal

Testes Não Paramétricos – Classes

- Há várias classes de testes não paramétricos
- Testes não paramétricos baseados na ordem
  - √ Não consideram distâncias das observações entre si
  - √ Mediana usada como medida padrão de locação

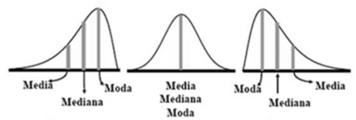
- Alguns testes não paramétricos:
  - √ Teste do sinal
  - √ Teste de Wilcoxon de postos sinalizados
  - √ Teste de Wilcoxon-Mann-Whitney
  - √ Tese de Kruskal-Wallis
  - √ Teste de Levene
- Intervalos de confiança fornecidos pelos pacotes computacionais
  - √ Em geral por meio de inversão do teste

Testes Não Paramétricos – Uso

- Por que não são usados sempre?
  - √ Os testes paramétricos são superiores quando suas suposições são válidas
    - – Superiores no sentido de, mais provavelmente, perceber afastamentos de H_0
      (testes mais poderosos)
  - √ Devem ser usados quando os dados parecem tão não normais que há sérios receios no uso de testes paramétricos

Teste do Sinal

- Mediana populacional ($\tilde{\mu}$):
  - √ Valor do meio da distribuição
    - – Mesma probabilidade de obter valor acima da mediana $\tilde{\mu}$, como abaixo dela.



Teste de Significância para a Mediana

- Testar que a mediana é 0
  - √ H_0: $\tilde{\mu} = 0$
- Se H_0 for verdadeira
  - √ Observação teria mesma probabilidade de ser positiva ou negativa

    P(+) = P(–) = 0,5 = p
    - Ou seja, testar H_0 seria testar se p = ½
  - √ Evidência contra H_0:
    - Grande desequilíbrio de sinais + e –
    - Teste usa teoria binomial na distribuição dos dados

Estrutura do Teste

- Equivalência das hipóteses:

$$H_0: \tilde{\mu} = 0 \text{ vs. } \tilde{\mu} \neq 0 \qquad H_0: p = \frac{1}{2} \text{ vs. } p \neq \frac{1}{2}$$
$$H_0: \tilde{\mu} = 0 \text{ vs. } \tilde{\mu} > 0 \quad \Longrightarrow \quad H_0: p = \frac{1}{2} \text{ vs. } p > \frac{1}{2}$$
$$H_0: \tilde{\mu} = 0 \text{ vs. } \tilde{\mu} < 0 \qquad H_0: p = \frac{1}{2} \text{ vs. } p < \frac{1}{2}$$

- Empates:
  - √ Observações iguais à mediana hipotética
  - √ Diferenças emparelhadas iguais a zero
- Empates são ignorados
  - √ Ignorar empates é considerado uma aproximação

- Intervalos de confiança não paramétrico:
  - √ Em geral, usa-se o procedimento de inversão do teste
  - √ Construído reunindo todos os valores hipotético do parâmetro que não são rejeitados no nível de significância especificado

Exemplo

- Examinando pares de medidas de cabeças
  - √ Interesse:
    - Determinar a adequação de uso de calibrador de cartolina (barato) em vez de calibrador metálico (caro e desconfortável)
  - √ Hipótese nula:
    - Mediana das medidas dos 2 tipos de calibradores é 0

Diferença = Cartolina – Metal

- Empate:
  - √ Há um sujeito para o qual as medidas são iguais (diferença = 0)
  - √ Ignora-se este sujeito
    - São utilizados as 17 observações restantes
  - √ Sinais
    - Calibradores de cartolina (+)
      - Medida superior em 14 casos
    - Calibrador de metal (–)
      - Medida superior em 3 casos
  - √ Seria como lançar uma moeda 17 vezes e obter 14 caras.

41

Pesquisa Quantitativa - 2016

- Se H_0 for verdadeira:
  - √ p = ½
  - √ P{14 ou mais caras em 17 lançamentos}

$$P\{Y \geq 14\} \text{ , com } Y \sim \text{binomial}(n = 17; p = 0,5)$$
$$P\{Y \geq 14\} = 0,00636 \text{ .}$$

- p-valor (H_1 é bilateral)
  - √ p = 2 x 0,00636 = 0,0127
  - √ p-valor (teste t) = 0,005415
- Conclusão:
  - √ Forte evidência para rejeitar H_0
  - √ Mesma conclusão do teste t

42

Pesquisa Quantitativa - 2016

- Execução do teste – comando:

```
      # Criação do objeto com os resultados do teste do sinal

library(BSDA)    # Para comando do teste do sinal
dif.sign <- SIGN.test(dif, md = 0, alternative = "two.sided", conf.level = 0.95)
```

- Resultado - saída:

```
> dif.sign

 One-sample Sign-Test

data:  dif
s = 14, p-value = 0.01273
alternative hypothesis: true median is not equal to 0
95 percent confidence interval:
 1.000000 2.707563
sample estimates:
median of x
        1.5
```

43

Pesquisa Quantitativa - 2016

- Intervalo de 95% confiança para a mediana das diferenças
  - √ [1,0; 2,7]
- IC de 95% para a média
  - √ Resultado do procedimento paramétrico (intervalo t)
  - √ [0,5; 2.7]

44

Pesquisa Quantitativa - 2016

Teste de Wilcoxon de Postos Sinalizados

- Outro teste não paramétrico com funcionamento similar
  - √ Testa as mesmas hipóteses
  - √ É melhor para detectar afastamentos de H_0

- Execução do teste – comando:

```
      # Criação do objeto com os resultados do teste do sinal

dif.wilc <- > with(cabecas, wilcox.test(cartolina, metal, mu = 0,
                    alternative = "two.sided", paired=TRUE, exact = FALSE,
                    conf.int = TRUE, conf.level = 0.95))
```

- Resultado - saída:

```
> dif.wilc

 Wilcoxon signed rank test with continuity correction

data:  cartolina and metal
V = 130.5, p-value = 0.01052
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 0.5000399 2.9999637
sample estimates:
(pseudo)median
      1.500075
```


- P-valor do teste:
  - √ Wilcoxon: 0,011
  - √ Sinais: 0,0127
- Conclusão:
  - √ A mesma em ambos os testes
- Intervalo de 95% confiança para a mediana das diferenças
  - √ Wilcoxon: [0,5; 3,0]
  - √ Sinais: [1,0; 2.7]

Teste do Sinal – Uso

- Condições para uso:
  - √ Pode ser usado sempre, desde que
    - as observações sejam independentes (caso de uma amostra)
    - os diferentes pares de observações sejam independentes (caso de amostras emparelhadas)
  - √ O teste é muito sensível ao afastamento da hipótese de independência

- Vantagens:
  - √ Livre de distribuição (*distribution-free*)
    - Válido para dados de qualquer distribuição (normal ou não)
  - √ Não é sensível a valores atípicos
- Desvantagens:
  - √ O teste t é superior quando suas suposições forem válidas

49

Pesquisa Quantitativa - 2016

- Em geral, testes não paramétricos são úteis quando:
  - √ Têm-se amostras muitos pequenas ($n \leq 10$)
    - Há menor capacidade de verificar a validade das suposições do teste t ou de detectar valores atípicos
  - √ Execução de teste unilateral em situações que se suspeita de que os dados sejam oriundos de distribuição assimétrica
  - √ Existência de observação suspeita (grande ou pequena) que julgamos ser real em vez de apenas um erro.

50

Pesquisa Quantitativa - 2016

Duas Amostras Independentes

51

Pesquisa Quantitativa - 2016

Duas Amostras Independentes

- Comparação de parâmetros de amostras independentes de duas populações diferentes:
  - √ Comparar as duas médias populacionais
  - √ Comparar as duas dispersões populacionais

52

Pesquisa Quantitativa - 2016

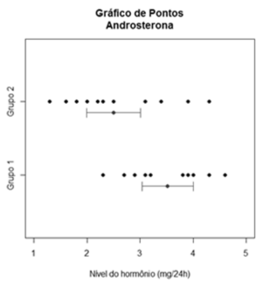
Questões de Desenho do Estudo

- Experimentos x Estudos Observacionais
  - √ Dados são analisados da mesma maneira
  - √ Distinção se torna importante quando vamos interpretar os resultados da análise
  - √ Importante:
    - – Deve-se ter muito cuidado ao fazer inferências causais a partir de dados observacionais

Questões de Análise

- É importante representar graficamente os dados:
  - √ Examinar um gráfico apropriado é a maneira mais rápida de ver o que está acontecendo
  - √ Pode evitar o aparecimento de falsos juízos perigosos

- Exemplo:



  - √ Apenas a comparação entre as médias não permitiria observar a superposição significativa do nível do hormônio entre os dois grupos

Questões de Análise

- Aplicabilidade de ferramentas formais de análise:
  - √ Todos os métodos formais de análise fazem suposições:
    - – As suposições do método que desejamos usar são satisfeitas?
    - – Se não, este fato pode nos causar problemas?

Questões de Análise

- Outras características que não as médias:
  - √ Há outras características dos dados que não as médias que seriam interessante de se comparar?
  - √ Que procedimento adotar com tal finalidade?

Comparação de Duas Médias

- Suposição
  - √ Duas amostras aleatórias das populações:
    - População 1: $N(\mu_1, \sigma_1)$
    - População 2: $N(\mu_2, \sigma_2)$
- Intervalo de confiança para $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \pm t \times \text{ep}(\bar{x}_1 - \bar{x}_2) \ .$$

- Teste t para H_0: $\mu_1 - \mu_2 = 0$

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2)}{\text{ep}(\bar{x}_1 - \bar{x}_2)} \ .$$

- Cálculo manual:
  - √ Erro padrão: $\text{ep}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \ .$
  - √ Graus de liberdade: $\text{gl} = \min\{(n_1 - 1), (n_2 - 1)\} \ .$
  - √ Aproximação conservadora do procedimento de Welch (cálculo complicado)

- Cálculo computacional:
  - √ Variabilidade das populações:
    - Caso 1: variâncias iguais (homocedasticidade)
    - Caso 2: variâncias diferentes (heterocedasticidade)
  - √ Cada pacote apresenta *defaults* diferentes
  - √ Recomenda-se como escolha de rotina considerar as variâncias das populações diferentes (heterocedasticidade)

- Variâncias desiguais – Heterocedasticidade
  - √ Erro padrão: $\text{ep}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.
  - √ Graus de liberdade – Aproximação de Satterthwaite

$$\text{gl} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2} .$$

    - – Em amostragens repetidas, o valor fornecido muda de amostra a amostra
    - – Frequentemente obterá valores não inteiros (saídas)
  - √ Estatística de teste t_0 não tem distribuição exata
  - √ Cálculo manual garante uma aproximação conservadora dos graus de liberdade

61

- Variâncias iguais – Homocedasticidade
  - √ Erro padrão: $\text{ep}(\bar{x}_1 - \bar{x}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$.
  - √ s_p^2: estimativa combinada da variância única:

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} .$$

    - – Combina a informação das duas amostras para formar uma única estimativa
  - √ Graus de liberdade: $\text{gl} = n_1 + n_2 - 2$.
  - √ Hipótese de homocedasticidade deve sempre ser verificada
    - – Na prática, raramente é verdadeira

62

- Teste é razoavelmente robusto às variâncias diferentes (heterocedasticidade) quando os tamanhos das amostras são iguais
  - √ Perde-se bastante quando a suposição de variâncias iguais é falsa e os tamanhos das amostras são diferentes
- Há autores que recomendam o uso rotineiro do procedimento de Welch
  - √ Variâncias diferentes

63

Robustez do Teste t de Duas Amostras

- Testes t e intervalos t para duas amostras são ainda mais robustos diante de não normalidade do que os testes t para 1 amostra
  - √ Especialmente quando as formas das duas distribuições são semelhantes e os tamanhos das amostras são iguais
    - – Mesmo para amostras tão pequenas quanto $n_1=n_2=5$
  - √ Pode- se aplicar para $n_1 + n_2$ as recomendações sobre tamanhos amostrais do caso de 1 amostra

64

Testes Não Paramétricos de Duas Amostras

- Alternativas mais conhecidas:
  - √ Teste da soma de postos de Wilcoxon
  - √ Teste de Mann-Whitney
- Ambos são equivalentes e produzem os mesmos p-valores

- Testes são baseados apenas na ordenação dos dados
  - √ Livre de distribuição
  - √ Testa-se a igualdade das medianas populacionais
  - √ Intervalos de confiança fornecidos são para a diferença das duas medianas
- Intervalos de confiança não paramétricos
  - √ Intervalos de confiança obtidos pelos pacotes por meio da inversão do teste

Exemplo

- Nível de hormônio masculino
  - √ Níveis de androsterona urinária (mg/24h)
  - √ Grupos de homens com boa saúde física:
    - Grupo 1: 11 homens
    - Grupo 2: 15 homens
- Os níveis médios dos grupos são iguais?

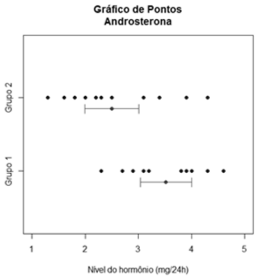
- Estatísticas Descritivas - Saída R

```
> library(pastecs)# fornece resumo de estatísticas descritivas
> aggregate(nivel ~ orientacao, data = androsterona, stat.desc, basic = F)

  orientacao nivel.median nivel.mean nivel.SE.mean nivel.CI.mean.0.95 nivel.var
1        Hom    2.3000000  2.5000000     0.2382476          0.5109902 0.8514286
2        Het    3.8000000  3.5181818     0.2173469          0.4842790 0.5196364
  nivel.std.dev nivel.coef.var
1     0.9227289      0.3690915
2     0.7208581      0.2048951
```

- Estatísticas descritivas – Quadro resumo

Grupos	$\bar{x}$	s_X	n
1	3,518	0,721	11
2	2,500	0,923	15

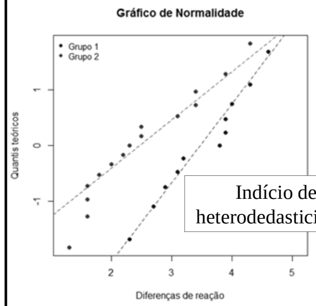
- Dados amostrais:



√ Grupo 2 tem média mais baixa, alguns dos indivíduos deste grupo têm níveis mais altos do que a maioria do grupo 1

√ Leve assimetria e maior dispersão no grupo 2

- Verificação normalidade das amostras



Indício de heterodedasticidade

```
> shapiro.test(nivel.1)

 Shapiro-Wilk normality test

data:  nivel.2
W = 0.95746, p-value = 0.7396

> shapiro.test(nivel.2)

 Shapiro-Wilk normality test

data:  nivel.2
W = 0.92782, p-value = 0.2531
```

√ Grupo 2 apresenta desvios (extremidades)

√ Testes não têm evidência contra normalidade

- Considerações para aplicação do teste t:
  - √ Avaliação empírica
    - Possibilidade de comportamento não normal para os dados
  - √ Gráficos de probabilidade estão razoáveis e teste formais não apresentaram evidência contrária à suposição de normalidade
    - Em amostras pequenas pode ser difícil verificar normalidade dos dados
  - √ Estamos também protegidos pela robustez dos testes t e intervalos t.

- Parâmetro de interesse:
  - √ $\mu_1 - \mu_2$: diferença desconhecida entre os valores verdadeiros dos níveis médios de androsterona nos dois grupos
- Hipóteses
  - √ H_0: $\mu_1 - \mu_2 = 0$ vs. H_1: $\mu_1 - \mu_2 \neq 0$

- Teste t – Cálculo manual:
  - √ Erro padrão:
    - Grupos independentes

$$\begin{aligned} \mathrm{ep}(\bar{x}_1 - \bar{x}_1) &= \sqrt{\frac{s_1^2}{n_N} + \frac{s_1^2}{n_S}} \\ &= \sqrt{\frac{0,721^2}{11} + \frac{0,923^2}{15}} \\ &= 0,32257\ . \end{aligned}$$

- Estatística de teste:

$$\begin{aligned} t_0 &= \frac{\text{estimativa} - \text{valor admitido por hipótese}}{\text{erro padrão}} \\ &= \frac{(3,518 - 2,500) - 0}{0,32257} = 3,15887\ . \end{aligned}$$

- p-valor:
  - √ Distribuição amostral: t_{10} [gl=min(11–1,15–1)=10

$$p - valor = 2 \times \Pr\{T \geq 3,15887\} = 0,010178.$$

73

Pesquisa Quantitativa - 2016

√ IC de 95% para a verdadeira diferença ($\mu_1 - \mu_2$)

$$\begin{aligned} &(\bar{x}_1 - \bar{x}_2) \pm t_{10} \times \mathrm{ep}(\bar{x}_1 - \bar{x}_2) = \\ &(3,518 - 2,500) \pm 2,228139 \times 0,32257 = \\ &[0,30; 1,74]\ . \end{aligned}$$

74

Pesquisa Quantitativa - 2016

- Teste t – variâncias diferentes:

```
       # Criação do objeto com os resultados do teste t para diferença

dif.t <- t.test(nivel ~ orientacao, mu = 0, alternative = "two.sided",
                 paired = FALSE, conf.level = 0.95, data = androsterona)
```

- Resultado - saída:

```
> dif.t

 Welch Two Sample t-test

data:  nivel by codigo
t = 3.1572, df = 23.862, p-value = 0.004278
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3523851 1.6839786
sample estimates:
mean in group 1 mean in group 2
         3.518182         2.500000
```

75

Pesquisa Quantitativa - 2016

```
> dif.t

Welch Two Sample t-test

data:  nivel by codigo
t = 3.1572, df = 23.862, p-value = 0.004278
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3523851 1.6839786
sample estimates:
mean in group 1 mean in group 2
         3.518182         2.500000
```

- Conclusão:
  - √ Evidência muito forte de uma diferença significativa entre as médias verdadeiras
  - √ Com 95% de confiança, nível médio do hormônio é maior no grupo 1 não menos que 0, 4 mg/24h e e não mais que 1,7 mg/24h do que o nível médio do grupo 2

76

Pesquisa Quantitativa - 2016

- Teste t combinado – variâncias iguais:

```
        # Criação do objeto com os resultados do teste t para diferença

        # para a diferença - procedimento combinado
dif.comb <- t.test(nivel ~ orientacao, mu = 0, alternative = "two.sided",
                            var.equal = T, paired = FALSE, conf.level = 0.95,
                            data = androsterona)
```

- Resultado - saída:

```
> dif.comb

 Two Sample t-test

data:  nivel by grupo
t = 3.0372, df = 24, p-value = 0.005679
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.326298 1.710066
sample estimates:
mean in group 1 mean in group 2
         3.518182          2.500000
```

77

Pesquisa Quantitativa - 2016

- Teste não paramétrico – Mann-Whitney:

```
        # Criação do objeto com os resultados do teste não paramétrico

        # Teste de Mann-Whitney-Wilcoxon
dif.wilc <- wilcox.test(nivel ~ orientacao, data = androsterona, mu = 0,
                          paired = FALSE, alternative = "two.sided",
                          exact = FALSE, conf.int = TRUE, conf.level = 0.95)
```

- Resultado - saída:

```
> dif.wilc

Wilcoxon rank sum test with continuity correction

data:  nivel by orientacao
W = 132.5, p-value = 0.01003
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 0.3999628 1.8000446
sample estimates:
difference in location
              1.100007
```

78

Pesquisa Quantitativa - 2016

- Comparação dos resultados:

Teste	p-valor	Intervalo de confiança
t – Welch	0,0044	[0,35; 1,68]
T combinado	0,0057	[0,33; 1,71]
Mann-Whitney	0,0100	[0,40; 1,80]
t – cálculo manual	0,0102	[0,30; 1,74]

√ Valores são bem semelhantes

√ Mesmas conclusões dos testes

√ Pequenas diferenças nos intervalos de confiança

79

Pesquisa Quantitativa - 2016

Mais de Duas Amostras

80

Pesquisa Quantitativa - 2016

Análise de Variância de um Fator

- Problema mais geral:
  - √ Comparar a média de várias (k) populações ou grupos usando amostras independentes
- Análise de variância de um fator:
  - √ Situações em que há apenas um fator (ou variável categórica) definindo pertinência a grupo

81
Pesquisa Quantitativa - 2016

Teste F para Anova de um Fator

- Hipóteses:
  - √ H_0: Todas as k médias verdadeiras são iguais
  - √ H_1: Existem diferenças entre algumas delas
- Formalização:
  - √ H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$
  - √ H_1: pelo menos uma diferença entre as médias

82
Pesquisa Quantitativa - 2016

Exemplo

- Comparação de quatro métodos de leitura:
  - √ Investigação sobre os efeitos de diferentes métodos para compreensão de leitura
  - √ Amostra de 50 alunos entre 13 e 14 anos
  - √ Métodos:
    - Mapeamento: uso de diagramas para observar e relacionar pontos principais
    - Varredura: leitura de introduções, passando os olhos para uma visão geral, antes de ler em detalhes
    - Ambos
    - Nenhum

83
Pesquisa Quantitativa - 2016

- Interesse:
  - √ Saber a efetividade dos métodos
- Medidas:
  - √ Idade de leitura (teste Gapadol)
  - √ Obter idades de leitura antes e depois do período de instrução (6 semanas)
- Classificação dos estudantes:
  - √ 4 grupos, dependendo de qual os métodos de leitura eles utilizaram antes do 2º teste
- Resposta:
  - √ Diferença dos escores de teste de leitura
    (em anos de idade de leitura)

84
Pesquisa Quantitativa - 2016

- Questão:
  - √ Os aumentos, em média, são maiores para os alunos que usaram mapeamento, varredura ou ambos?
- Cuidado da pesquisa:
  - √ Se usada a idade de leitura após período de instrução, quaisquer diferenças poderiam ser apenas devido a diferenças em habilidade de leitura entre os grupos

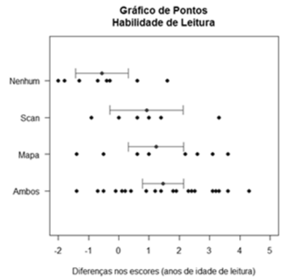
- Estatísticas Descritivas - Saída R

```
> library(pastecs)# fornece resumo de estatísticas descritivas
> aggregate(aumento ~ metodo, data = leitura, stat.desc, basic = F)

 metodo aumento.median aumento.mean aumento.SE.mean aumento.CI.mean.0.95
1     Ambos      1.6000000    1.4590909       0.3290850            0.6843697
2      Mapa      1.0000000    1.2333333       0.4160298            0.9156755
3    Nenhum     -0.7000000   -0.5555556       0.3782677            0.8722868
4 Varredura      1.0000000    0.9142857       0.4920456            1.2039921
  aumento.var aumento.std.dev aumento.coef.var
1   2.3825325       1.5435454        1.0578816
2   2.0769697       1.4411696        1.1685159
3   1.2877778       1.1348030       -2.0426453
4   1.6947619       1.3018302        1.4238768
```

Grupos	$\overline{x}$	s_X
Ambos	1,459	1,544
Mapa	1,233	1,441
Scan	0,914	1,302
Nenhum	–0,556	1,135

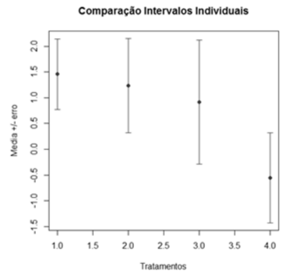
Variabilidade aparenta ser a mesma nos grupos

- Visualização gráfica:



  - √ Escores para cada um dos 3 grupos parecem deslocados para a direita (em relação a nenhum)
  - √ Aparentemente todos os tratamentos parecem ter algum efeitos
    - – Efeitos semelhantes

- Gráfico dos erros:



  - √ Quando os IC's individuais não se sobrepõem, o IC para a diferença não conterá 0
    - – Parece haver pelo menos uma diferença estatisticamente significante no gráfico

Tabela Anova

- Tabela Anova – comando:

```
# Criação do objeto com o ajuste

leitura.mod = lm(aumento ~ metodo, data = leitura)
```

- Resultado - saída:

```
> anova(leitura.mod)

Analysis of Variance Table

Response: aumento
          Df Sum Sq Mean Sq F value   Pr(>F)
metodo     3 27.062  9.0205   4.445 0.007977 **
Residuals 46 93.351  2.0294
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

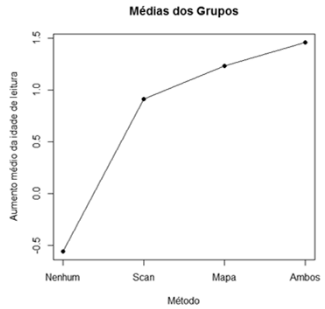
Há pelo menos uma diferença entre as médias verdadeiras

√ Evidência muito forte contrária à hipótese de igualdade das médias

89

Pesquisa Quantitativa - 2016

- Plot das médias:



√ Aparentemente, há diferenças entre os aumentos médios de idade de leitura entre os grupos

90

Pesquisa Quantitativa - 2016

- Importante:
  - √ Teste não indica onde estão as diferenças, nem sua intensidade
  - √ Para estimar o tamanho das diferenças são necessários intervalos de confiança

91

Pesquisa Quantitativa - 2016

Interpretação do p-valor do Teste F

- p-valor grande:
  - √ Diferenças observadas entre as médias amostrais podem ser explicadas simplesmente em termos da variação amostral
- p-valor pequeno:
  - √ Evidência de que existem diferenças reais entre pelo menos duas médias verdadeiras

    (sem indicação de qual diferença, nem de sua intensidade)

92

Pesquisa Quantitativa - 2016

Suposições da Anova

- Todos os grupos têm a mesma variância verdadeira
- Estimativa combinada do desvio padrão é empregada no cálculo dos erros padrão para as médias individuais e para as diferenças entre elas

93

Pesquisa Quantitativa - 2016

- Diferenças (controle = Nenhum)
  - √ erros padrão individuais (baseados em s_p)

```
> sp <- summary(leitura.mod)$sigma
> sp
 [1] 1.424557
```

```
> summary(leitura.mod)

Call:
lm(formula = aumento ~ metodo, data = leitura)

Residuals:
    Min      1Q  Median      3Q     Max
-2.8591 -1.0229 -0.1444  0.9602  2.8409

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.5556     0.4749  -1.170 0.248045
metodoAmbos      2.0146     0.5637   3.574 0.000839 ***
metodoMapa       1.7889     0.6282   2.848 0.006560 **
metodoVarredura  1.4698     0.7179   2.047 0.046355 *
```

Intercepto: média grupo “Nenhum”

Coeficientes: diferença entre a média do grupo e “Nenhum”

Erro padrão: calculado com s_p

- √ Refere-se ao teste da diferença de médias entre o grupo indicado na linha e o grupo “Nenhum”
  - Evidência forte de diferenças significativa: (“Mapa” – “Nenhum”) e (“Ambos” – “Nenhum”)

94

Pesquisa Quantitativa - 2016

- IC’s de 95% para a diferença da média de cada grupos com o grupo de controle (“Nenhum”):
  - √ Erro padrão calculado a partir de s_p

```
> confint(leitura.mod)

                     2.5 %    97.5 %
(Intercept)     -1.51138352 0.4002724
metodoAmbos      0.88002958 3.1492633
metodoMapa       0.52444734 3.0533304
metodoVarredura  0.02476521 2.9149173
```

- √ Sugere-se que as médias de “Nenhum” e “Varredura” sejam iguais
- √ E o nível de confiança global nestas 3 comparações de diferenças?

95

Pesquisa Quantitativa - 2016

- Comparações emparelhadas de Fisher
  - √ Compara todas as diferenças entre médias

```
> MSerro <- anova(leitura.mod)[["Mean Sq"]] [2]
> DFerro <- anova(leitura.mod)[["Df"]][2]
> LSD.test(leitura.mod, "metodo", DFerro, MSerro, console = TRUE)

metodo,  means and individual ( 95 %) CI
             aumento      std  r        LCL       UCL  Min Max
Ambos      1.4590909 1.543545 22  0.8477413 2.0704405 -1.4 4.3
Mapa       1.2333333 1.441170 12  0.4055620 2.0611046 -1.4 3.6
Nenhum    -0.5555556 1.134803  9 -1.5113835 0.4002724 -2.0 1.6
Varredura  0.9142857 1.301830  7 -0.1695213 1.9980928 -0.9 3.3
alpha: 0.05 ; Df Error: 46
Critical Value of t: 2.012896

Means with the same letter are not significantly different

Groups, Treatments and means
a        Ambos    1.459091
a        Mapa     1.233333
a        Varredura       0.9142857
b        Nenhum          -0.5555556
```

Sugere que “Ambos”, “Mapa” e “Scan” têm a mesma média verdadeira

96

Pesquisa Quantitativa - 2016

- Comparações emparelhadas de Fisher:
  - √ IC de cada diferença têm 95% de cobertura
  - √ Chance de cobertura de todos os IC's é menor
    - Family-wise error: 0,198, nas 6 comparações do slide anterior
    - Significa que a chance de todos os 6 IC's cobrirem o verdadeiro valor da diferença é 1 em 5
  - √ Quanto mais comparações simultâneas, pior o problema

- Comparações emparelhadas de Tukey
  - √ Nível global de confiança: 95%

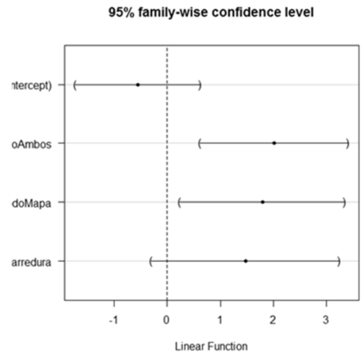
```
> leitura.mod2 <- aov(aumento ~ metodo, data = leitura)
> TukeyHSD(x = leitura.mod2, "metodo", conf.level=0.95)
 Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = aumento ~ metodo, data = leitura)

$metodo
                       diff        lwr      upr     p adj
Ambos-Nenhum      2.0146465  0.5121756 3.517117 0.0045033
Mapa-Nenhum       1.7888889  0.1145029 3.463275 0.0321119
Varredura-Nenhum  1.4698413 -0.4437427 3.383425 0.1859365
Mapa-Ambos       -0.2257576 -1.5884421 1.136927 0.9709011
Varredura-Ambos  -0.5448052 -2.1925747 1.102964 0.8145109
Varredura-Mapa   -0.3190476 -2.1249529 1.486858 0.9650797
```

Todos os intervalos, exceto dois contém 0

  - √ Diminui a precisão dos intervalos
    - Taxa individual é 0,0106

- Intervalos que não contém zero
  - √ "Mapa" e "Nenhum":
    - Verdadeiro aumento no escore do teste é maior, em média, para o grupo "Mapa" do que para o grupo "Nenhum" por entre 0,1 e 3,5
    - (limite inferior de confiança muito próximo de 0)
  - √ "Ambos" e "Nenhum"
    - [0,5; 3,5]: pode ser lido de maneira análoga
  - √ Há evidência que os métodos são efetivos, mas há informação insuficiente nos dados para se poder dizer qual deles é melhor

- IC individuais para nível de confiança global de 95%



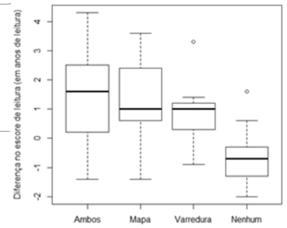
Teste F

- Aplica-se a situações em que têm-se k amostras independentes de cada uma de k populações
- Suposição
  - √ Dados do i-ésimo grupo $\sim N(\mu_i, \sigma)$
  - √ Mesmos desvios padrão das populações subjacente

Tabela Anova

- Interpretação:

```
Analysis of Variance Table

Response: aumento
          Df Sum Sq Mean Sq F value   Pr(>F)
metodo     3 27.062  9.0205   4.445 0.007977
Residuals 46 93.351  2.0294
```



- √ Linha do fator:
  - Variabilidade entre os grupos
- √ Linha dos resíduos (erros)
  - Variabilidade dentre dos grupos

Notação

x_{ij} : j-ésima observação no i-ésimo grupo

$\bar{x}_{i\cdot}$: média amostral do grupo i

s_i: desvio padrão do grupo i

n_i: quantidade de observações do grupo i

$n = \sum n_i$: total global de observações

$\bar{x}_{\cdot\cdot}$: média global

Fonte de Variação	Soma de Quadrados	gl	Média Quadrática[a]	F_0
Entre	$\sum n_i(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2.$	k – 1	s_B^2	$f_0 = \frac{s_B^2}{s_W^2}$
Dentro	$\sum (n_i - 1)s_i^2.$	n – k	s_W^2	
Total	$\sum\sum (x_{ij} - \bar{x}_{\cdot\cdot})^2.$	n – 1		

[a]Média quadrática = (soma de quadrados)/gl.

- P-valor:
  - √ $p = \Pr\{F \geq f_0\}$

Quantidades da Tabela Anova

- s_B^2: média quadrática entre os grupos

$$s_B^2 = \frac{\sum n_i(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2}{k-1}.$$

  - √ Medida de variabilidade das médias amostrais (quão afastadas elas estão)
- s_W^2: média quadrática dentro dos grupos

$$s_W^2 = \frac{\sum (n_i - 1)s_i^2}{n-k}.$$

  - √ Média podenrada dos desvios padrão dos grupos
    - (reflete a variabilidade dentro das amostras)

106

Pesquisa Quantitativa - 2016

- f_0: estatística F

$$f_0 = \frac{s_B^2}{s_W^2}.$$

  - √ Compara a variabilidade das médias amostrais com a variabilidade interna dos grupos
- Evidência contra H_0:
  - √ Valores de f_{0} que sejam grandes se H_0 fosse verdadeira
- Quando k = 2, $f_0 = t_0^2$
  - √ t_0 do teste t combinado
  - √ p-valor é o mesmo do teste t bilateral com estimativa combinada

107

Pesquisa Quantitativa - 2016

- Quando H_0 é verdadeira:

$$f_0 \sim F_{\mathrm{gl}_1, \mathrm{gl}_2}.$$

  - √ gl_1: graus de liberdade no numerador
  - √ gl_2: graus de liberdade no denominador
- p-valor:

$$\text{p-valor} = \Pr\{F \geq f_0\}.$$

108

Pesquisa Quantitativa - 2016

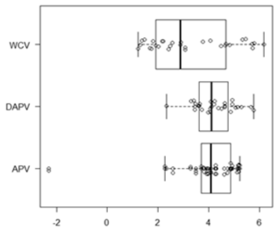
Exemplo

- Vacinas contra coqueluche
  - √ Comparação de vacina padrão (WCV) com uma nova vacina (APV) e DAPV (vacina APV com o dobro de potência).
  - √ Estudo aleatorizado, duplo cego, com 91 crianças de 17 a 19 meses
  - √ Variável medida:
    - Concentração de anticorpos de coqueluche medida no sangue, 1 mês após imunização
    - Medidas na escala log(UI/mL)

109

Pesquisa Quantitativa - 2016

- Avaliação gráfica:



√ Valor atípico (–2,30) no grupo APV foi um erro

√ Indício de diferença entre a vacina experimental e a vacina padrão (WCP)

√ Pouca ou nenhuma diferença entre as duas diferentes potências da vacina experimental (APV e DAPV)

√ Interessante estimar os tamanhos das diferenças

110

Pesquisa Quantitativa - 2016

- Anova – todo o conjunto de dados:

```
# Criação do objeto com o ajuste

coqueluche.mod1 <- lm(concentracao ~ vacina, data = coqueluche)
```

- Resultado - saída:

```
> anova(coqueluche.mod1)

Analysis of Variance Table

Response: concentracao
          Df  Sum Sq Mean Sq F value  Pr(>F)
vacina     2  13.519  6.7597  3.9224 0.02335 *
Residuals 88 151.653  1.7233
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

√ Evidência contrária à hipótese de igualdade das médias

111

Pesquisa Quantitativa - 2016

- Anova – sem valor atípico:

```
# Criação do objeto com o ajuste

# valor atipico: linha 25
coqueluche.mod2 <- lm(concentracao ~ vacina, data = coqueluche[-25,])
```

- Resultado - saída:

```
> anova(coqueluche.mod2)

Analysis of Variance Table

Response: concentracao
          Df  Sum Sq Mean Sq F value   Pr(>F)
vacina     2  15.829  7.9143  6.1752 0.003106 **
Residuals 87 111.502  1.2816
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

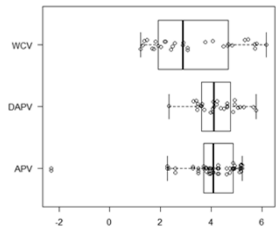
Há pelo menos uma diferença entre as médias verdadeiras

√ Sinaliza evidência clara de que existem diferenças entre as médias verdadeiras

112

Pesquisa Quantitativa - 2016

- Boxplot:



113

Pesquisa Quantitativa - 2016

- Comparações emparelhadas de Tukey
  - √ Nível global de confiança: 95%

```
> coqueluche.mod3 <- aov(concentracao ~ vacina, coqueluche[-25,])
> TukeyHSD(x = coqueluche.mod3, "vacina", conf.level=0.95)
 Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = concentracao ~ vacina, data = coqueluche[-25, ])

$vacina
                diff        lwr        upr     p adj
DAPV-APV  0.07097887 -0.6264043  0.7683621 0.9680650
WCV-APV  -0.85082759 -1.5538061 -0.1478490 0.0135082
WCV-DAPV -0.92180645 -1.6131583 -0.2304546 0.0057456
```

DAPV – APV contém 0

  - √ Não há diferença entre vacinas DAPV e APV
  - √ Grupo controle diferente dos grupos tratamento

114

Pesquisa Quantitativa - 2016

Suposições Subjacentes no Teste F

- Independência das amostras:
  - √ Suposição crucial
  - √ Em geral, só sabemos que ela é verdadeira quando as amostras consistem em indivíduos ou objetos ou processos fisicamente não relacionados
  - √ Dados emparelhados podem ser incorretamente analisados como se fossem provenientes de amostras independentes

115

Pesquisa Quantitativa - 2016

- Dados normalmente distribuídos
  - √ Teste F com robustez semelhante a do teste t para 2 amostras
    - – Especialmente, quando tamanhos das amostras são muito semelhantes

116

Pesquisa Quantitativa - 2016

- Homocedasticidade:
  - √ Mesmo nível de variabilidade subjacente (variância ou desvio padrão) em cada grupo
  - √ Regra prática (conservadora):
    - – Desvios desta hipótese não prejudicarão a análise se a razão do maior desvio padrão para o menor não for maior que 2
  - √ Embora o teste F seja razoavelmente robusto diante da heterocedasticidade, os IC’s não o são
    - – IC’s longos demais para grupos menos variáveis
    - – IC’s estreitos demais para grupos mais variáveis
  - √ Solução:
    - – IC’s para as médias individuais e para as diferenças

117

Pesquisa Quantitativa - 2016

- O que fazer se as suposições forem gravemente violadas?
  - √ Opções discutidas para o teste t
  - √ Teste não paramétrico alternativo para teste F
    - Teste de Kruskal-Wallis
  - √ Alternativas computacionais para comparação de mais de duas médias sem suposição de homocedasticidade

118

Pesquisa Quantitativa - 2016

Testes para Diferenças de Dispersão

- Testes baseados nos desvios padrão são suspeitos
  - √ Muito sensíveis à não normalidade dos dados
- Testes não paramétricos robustos para testar se as populações têm dispersão diferentes
  - √ Utilizam outra medida de dispersão

119

Pesquisa Quantitativa - 2016

Teste de Levene

- Utiliza desvios absolutos
  - √ Variância utiliza desvios quadráticos
- Calcula-se para cada grupo:

$$y_{ij} = |x_{ij} - \bar{x}_{i\cdot}|.$$

  - √ Com dispersão maior num grupo do que em outro
    - Desvio absolutos médio será maior

120

Pesquisa Quantitativa - 2016

- Procedimento:
  - √ Usar teste F para testar se a média verdadeira dos desvios absolutos (y_{ij}) são iguais para todos os grupos
    - É o mesmo que testar a igualdade das dispersões
- Versão robusta do teste de Levene
  - √ Usa mediana amostral no cálculo dos desvios
    (Desvios medianos)
  - √ Aplica-se Anova nos desvios medianos

121

Pesquisa Quantitativa - 2016

Exemplo

- Vacina contra coqueluche
  - √ Teste de Levene

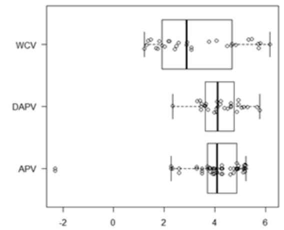
```
> library("car")# para o comando do teste de Levene
> # teste robusto de Levene
> leveneTest(concentracao ~ vacina, data = coqueluche, center = median)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2  5.4234 0.00601 **
      88
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

  - √ Sinaliza uma forte evidência de pelo menos uma diferença nas dispersões dos grupos

122

Pesquisa Quantitativa - 2016

- Box-plot :



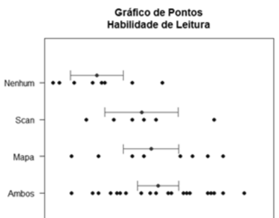
- √ Grupo de controle parece muito mais variável do que os grupos de tratamento
- √ Seu desvio padrão é cerca de 2 vezes maior
  - Poderíamos até confiar no teste F
- √ Contudo, não confiaríamos nos IC's obtidos a partir do programa Anova com um fator

123

Pesquisa Quantitativa - 2016

Exemplo – Verificação das Suposições

- Comparação de quatro métodos de leitura



Grupos	$\bar{x}$	s_X
Ambos	1,459	1,544
Mapa	1,233	1,441
Scan	0,914	1,302
Nenhum	–0,556	1,135

- √ Desvios padrão amostrais são muito semelhantes
  - Razão maior/menor < 2
- √ Gráficos de pontos não evidencia comportamento não normal severo
- √ Preocupação:
  - Alguma aglomeração em "Mapas"
  - Maior observação "Scan" é atípico? (mas variação é a mesma de outros)

Pesquisa Quantitativa - 2016

124

- Conclusão:
  - √ Não se observa nenhuma violação importante das suposições no modelo Anova
  - √ Resultados anteriores poderiam ser confirmados

125

Pesquisa Quantitativa - 2016

- Verificação formal da homocedasticiade
  - √ Teste de Levene:

```
> library("car")# para o comando do teste de Levene
> # teste robusto de Levene
> leveneTest(aumento ~ metodo, data = leitura, center = median)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.8733 0.4618
      46
```

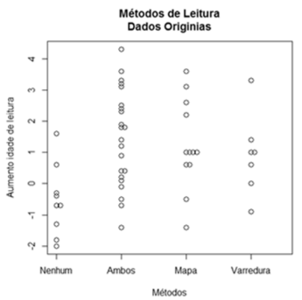
  - √ P-valor do teste confirma a avaliação de homocedasticidade (variâncias iguais em todos os grupos)

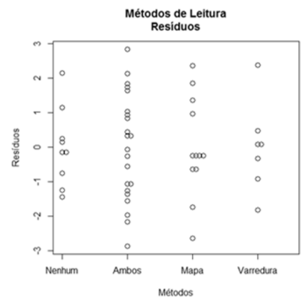
Investigações Adicionais da Adequação do Modelo

- Modelo de população: $X_{ij} \sim \mathrm{N}(\mu_i, \sigma_i),\ j = 1, 2, \ldots, n_i$
  $i = 1, 2, \ldots, k.$
- Modelo de Medição: $X_{ij} = \mu_i + \epsilon_{ij}$
  $\epsilon_{ij} \sim \mathrm{N}(0, \sigma_i)$.
  - √ μ_i: média verdadeira do grupo i
  - √ $\boldsymbol{\epsilon}_{ij}$: perturbações (erros) aleatórias

- Erros e resíduos:
  - √ $U_{ij} = X_{ij} - \mu_i$: erro (perturbação)
    - distância entre as observações e suas médias
  - √ $\hat{u}_{ij} = x_{ij} - \bar{x}_{ij}$: resíduo
    - Estima a distância verdadeira
- Análise dos resíduos:
  - √ $\hat{u}_{ij}$ vs. grupo i
  - √ Gráfico de normalidade de todos os resíduos

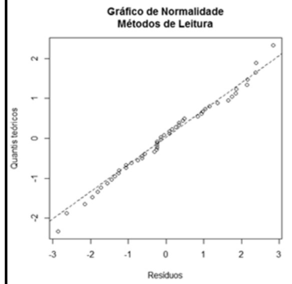
- $\hat{u}_{ij}$ vs. grupo i
  - √ Para pesquisar diferenças nas dispersões entre grupos





  - √ Pontos deslocados (médias no mesmo nível 0)
  - √ Gráfico não apresenta indícios de heterocedasticidade
    - Teste de Levene formalizou essa indicação

- Gráfico de normalidade de todos os $\hat{u}_{ij}$:
  - √ Se adotada a hipótese de homocedasticidade




```
> shapiro.test(leitura.mod$residuals)

Shapiro-Wilk normality test

data:  leitura.mod$residuals
W = 0.98346, p-value = 0.704
```

Teste não apresenta qualquer evidência contra a suposição de normalidade

  - √ Gráfico de probabilidade normal dos resíduos indica que eles parecem bem comportados

130

Pesquisa Quantitativa - 2016

Exemplo

- Vacinas contra coqueluche:
  - √ Pressuposto de homocedasticidade

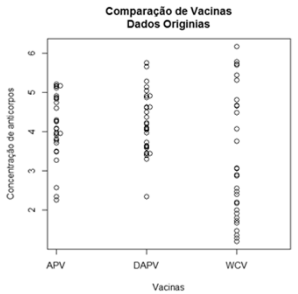
```
> library("car")# para o comando do teste de Levene
> # teste robusto de Levene
> leveneTest(concentracao ~ vacina, data = coqueluche, center = median)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2  5.4234 0.00601 **
      88
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

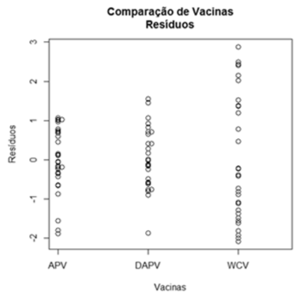
  - √ Teste de Levene apresenta forte evidência de violação da hipótese de homocesdasticidade

132

Pesquisa Quantitativa - 2016

- $\hat{u}_{ij}$ vs. grupo i
  - √ Pesquisa gráfica de diferenças nas dispersões



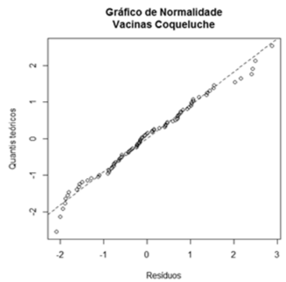


  - √ Grupo controle (WCV) com maior dispersão
  - √ Gráfico apresenta indícios de heterocedasticidade
    - Teste de Levene formalizou essa indicação

133

Pesquisa Quantitativa - 2016

- Gráfico de normalidade de todos os $\hat{u}_{ij}$:
  - √ Se adotada a hipótese de homocedasticidade




```
> shapiro.test(coqueluche.mod2$residuals)

 Shapiro-Wilk normality test

data:  coqueluche.mod2$residuals
W = 0.98133, p-value = 0.2231
```

Teste não apresenta qualquer evidência contra a suposição de normalidade

  - √ Gráfico de probabilidade normal dos resíduos indica que eles parecem bem comportados

134

Pesquisa Quantitativa - 2016

Teste de Kruskal-Wallis

- Equivalente não paramétrico para ANOVA
  - √ Variável medida em escala ordinal ou numérica
- Usado se:
  - √ Houver pressupostos comprometidos
    - normalidade e homocedasticidade.
  - √ Quantidades mínimas
    - Sujeitos: 5
    - Grupos: 3

- Uso:
  - √ Verificar se as amostras têm a mesma distribuição.
- Procedimento:
  - √ Baseado nos postos (ranks) das observações em cada grupo.
- Hipóteses:
  - √ H_0: grupos têm a mesma distribuição
  - √ H_1: grupos não têm a mesma distribuição
- Se todas distribuições tiverem mesma forma
  - √ H_0 será a igualdade das medianas dos grupos

Exemplo

- Vacina contra coqueluche
  - √ Teste de Kruskal-Wallis

```
> kruskal.test(concentracao ~ vacina, data = coqueluche)

 Kruskal-Wallis rank sum test

data:  concentracao by vacina
Kruskal-Wallis chi-squared = 6.3843, df = 2, p-value = 0.04108
```

  - √ Evidência contra a hipótese de mesma distribuição dos grupos de vacinas

Referências

Bibliografia Recomendada

- AGRESTI, A.; FINLAY, B. *Métodos estatísticos para as ciências sociais.* Penso, 2012.
- MOORE, D. S.; MCCABE, G. P.; DUCKWORTH, W. M.; SLOVE, S. L. *A prática da estatística empresarial: como usar dados para tomar decisões*. LTC, 2006.
- WILD, J. W.; SEBER, G. A. F. *Encontros com o acaso: um primeiro curso de análise de dados e inferência*. LTC, 2004.