

GRUPO3_ANALISES

2022-07-16

Sumário

- Objetivo
- Metodos
- Resultados
- Conclusão

Objetivo

O objetivo é avaliar identificar a associação entre as variáveis do exame e as variáveis do questionário, relacionar os valores das latências, LV, LA, LD, LE, LF e L0 com o **Score Semântico**, entre outras questões solicitadas, para esta análise foi utilizado a linguagem R e software Rstudio.

Metodo

Foi Feito a Análise exploratória e os tratamentos dos dados para buscar relações entre as variáveis, a limpeza dos dados foi feita no software Rstudio, utilizando a linguagem R para as demais análises. Iniciamos o tratamento dos dados substituindo nas variáveis quantitativas onde tinha valores ausentes colocamos os valores da média de cada variável respectivamente, em seguida foi feito alguns boxplot com o objetivo de observar principalmente o comportamento da média e mediana, por fim foi criado um modelo de regressão, executamos o teste tukey, criamos alguns gráficos para observar o comportamento dos resíduos.

Carregando os pacotes

Vamos iniciar carregando os pacotes que iremos utilizar

```
#PACOTES UTILIZADOS
require(readODS)
require(ggplot2)
require(cowplot)
require(dplyr)
require(readxl)
require(PMCMRplus)
require(car)
```

Carregando o Banco de Dados

Vamos fazer o carregamento do banco de dados contendo as informações sobre 30 crianças observadas

```
dados <- read_excel("Dados.xlsx", sheet=1)
```

Tratamento dos dados

Os valores faltantes foram substituídos pela média de sua respectiva variável

Foram identificadas NAS em nossa amostra, para prosseguirmos foi decidido substituir os valores ausentes por a média referente a suas respectivas colunas.

```
dados$LV[is.na(dados$LV)] <- round(mean(dados$LV, na.rm = T), 2)
dados$LC[is.na(dados$LC)] <- round(mean(dados$LC, na.rm = T), 2)

dados$LD[is.na(dados$LD)] <- round(mean(dados$LD, na.rm = T), 2)
dados$LE[is.na(dados$LE)] <- round(mean(dados$LE, na.rm = T), 2)
dados$LF[is.na(dados$LF)] <- round(mean(dados$LF, na.rm = T), 2)
dados$LO[is.na(dados$LO)] <- round(mean(dados$LO, na.rm = T), 2)

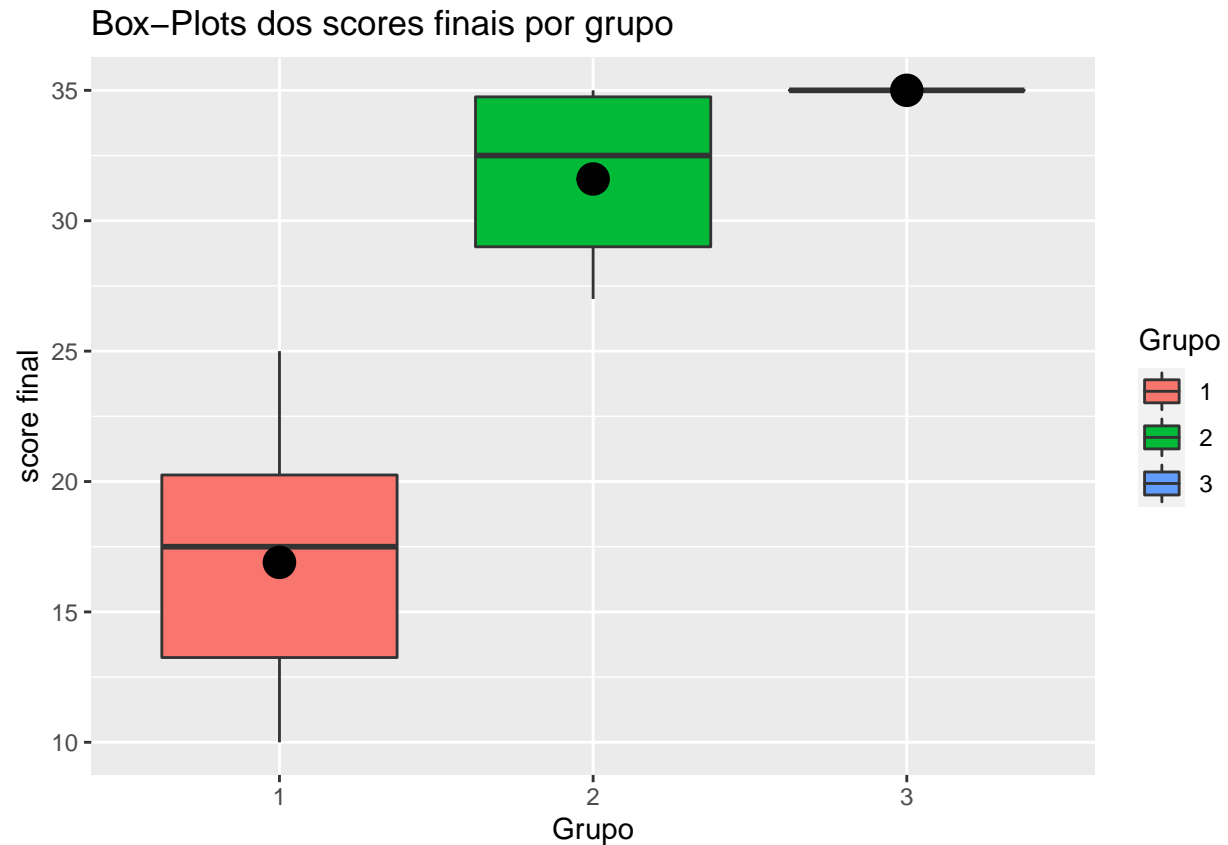
dados$AC[is.na(dados$AC)] <- round(mean(dados$AC, na.rm = T), 2)
dados$AD[is.na(dados$AD)] <- round(mean(dados$AD, na.rm = T), 2)
dados$AF[is.na(dados$AF)] <- round(mean(dados$AF, na.rm = T), 2)
# verificando se ainda temos algum valor ausente
any(is.na(dados))
```

```
## [1] FALSE
```

```
#como podemos observar não temos mais nenhum NA na nossa amostra
```

Podemos observar que agora não temos mais nenhum NA na nossa amostra, logo, podemos prosseguir com as nossas análises

```
## Warning: 'fun.y' is deprecated. Use 'fun' instead.
```



Pode-se observar que o grupo 3 apresenta uma maior média em relação aos demais , o próximo passo, é verificar se essa diferença é de fato significativa , paara isso , vamos utilizar o teste Tukey.

Fazendo o teste de tukey para verificar se há significância entre as médias dos grupos para a variável de score final

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = dados$'score final' ~ dados$Grupo)
##
## $'dados$Grupo'
##      diff      lwr      upr    p adj
## 2-1 14.7 10.7692562 18.630744 0.0000000
## 3-1 18.1 14.1692562 22.030744 0.0000000
## 3-2  3.4 -0.5307438  7.330744 0.0996508
```

Pelo teste Tukey vemos que a diferença entre o grupo 1 em relação aos demais é significativa, ou seja, de fato são diferentes, enquanto as médias dos grupos 2 e 3 são estatisticamente iguais.

Testando pelo Teste Nemenyi

```
## Warning in kwAllPairsNemenyiTest.default(c(14, 23, 13, 10, 18, 21, 17, 10, :
## Ties are present, p-values are not corrected.
```

```
##           1           2
## 2 9.768445e-03      NA
## 3 7.804371e-06 0.1769774
```

O grupo 1 foi diferente dos demais, o 2 e o 3 são iguais pelo teste de nemenyi

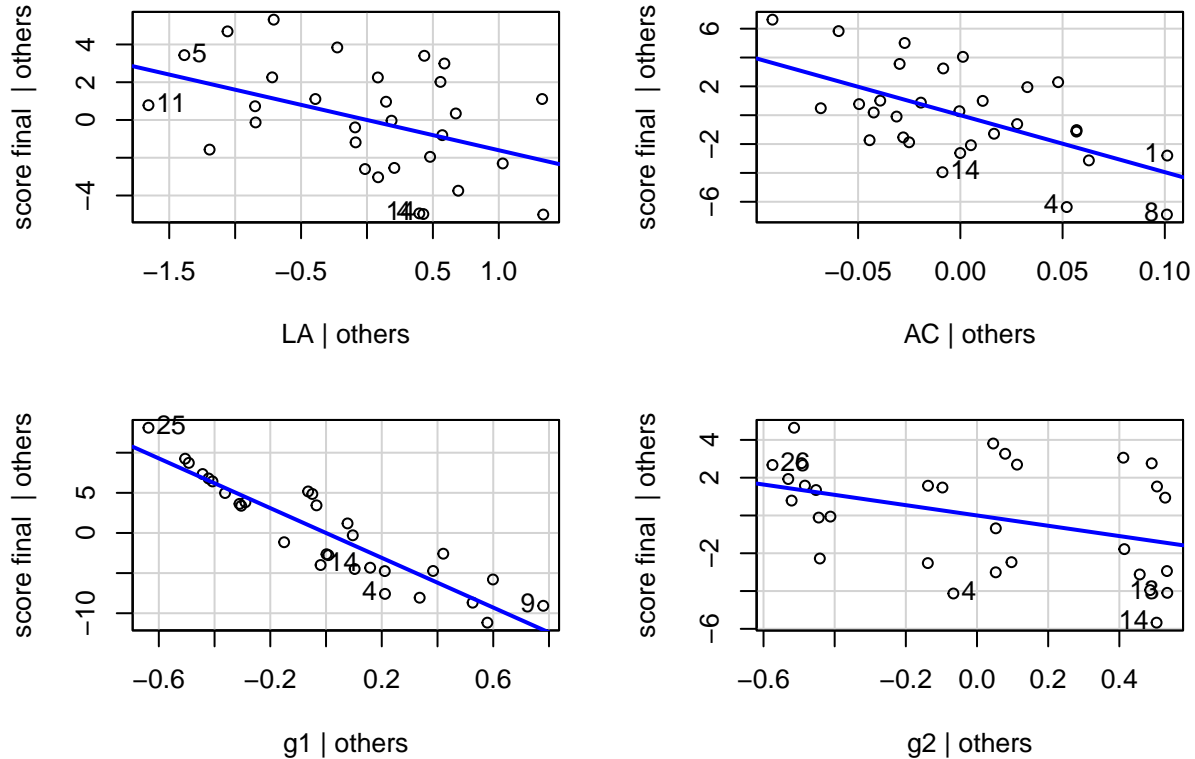
Regressão linear para os dados do exame com o score final e com os grupos

agora vamos construir um modelo de regressão linear para tentar explicar o **score final** (nossa variável predizida) através de nossas variáveis preditoras

```
##
## Call:
## lm(formula = 'score final' ~ LA + AC + g1 + g2, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.3113 -2.1250 -0.2124  2.1444  4.1789
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  52.6503     6.2441   8.432 8.89e-09 ***
## LA           -1.6036     0.6743  -2.378 0.02536 *
## AC          -39.4715    10.8274  -3.646 0.00122 **
## g1          -15.4870     1.3956 -11.097 3.77e-11 ***
## g2           -2.7270     1.2706  -2.146 0.04175 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.803 on 25 degrees of freedom
## Multiple R-squared:  0.9103, Adjusted R-squared:  0.8959
## F-statistic: 63.42 on 4 and 25 DF,  p-value: 1.006e-12
```

As variáveis LA e AC são estatisticamente significantes, de acordo com o teste t. A utilização dos grupos também apresentou significância

Added-Variable Plots



Teste de Normalidade Shapiro Wilk

- Hipóteses do teste

H_0 : os resíduos são normais $p \geq 0,05$ H_1 : os resíduos não são normais $p < 0,05$

```
##
## Shapiro-Wilk normality test
##
## data:  r1$residuals
## W = 0.95272, p-value = 0.1998
```

De acordo com o teste shapiro-wilk a 5% de significância Os resíduos para essa regressão são normais

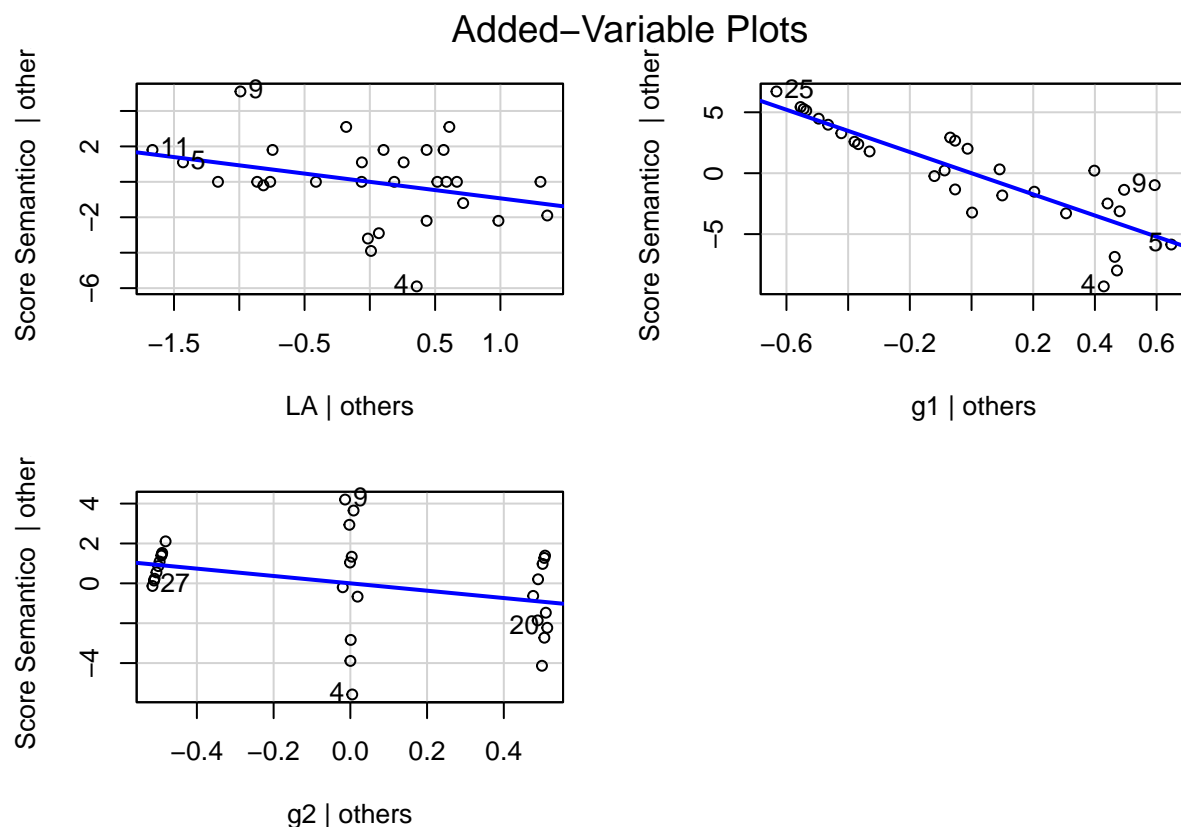
Relacionando os Valores de Latencia Com o Score Semântico, e Analisando em Relação ao Grupo

```
##
## Call:
## lm(formula = 'Score Semantico' ~ LA + g1 + g2, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -5.5655 -0.9206 0.0578 1.1898 4.1766
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.3981     4.9586   5.525 8.45e-06 ***
## LA           -0.9318     0.5444  -1.712  0.0989 .
## g1           -8.6826     1.0422  -8.331 8.20e-09 ***
## g2           -1.8447     1.0136  -1.820  0.0803 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.266 on 26 degrees of freedom
## Multiple R-squared:  0.7823, Adjusted R-squared:  0.7571
## F-statistic: 31.14 on 3 and 26 DF,  p-value: 9.237e-09
```

após a elaboração do segundo modelo percebemos que apenas **g1** foi significativo , LA e g2 podem ou não continuar no modelo ,ficando a critério de debate a remoção ou não .

```
avPlots(r1);shapiro.test(r1$residuals)
```



```
##
## Shapiro-Wilk normality test
##
## data:  r1$residuals
## W = 0.9777, p-value = 0.7618
```

Box-Plots dos scores semanticos por grupo

