

# Resolução Lista 3 - TRI

Allan Vieira - 14/0128492

1) Expressar a proficiência dos alunos da noite na escala de proficiências dos alunos da manhã utilizando diferentes métodos

Leitura e carregamento:

```
# library(dplyr)
library(readr)
library(magrittr)
library(ggplot2)
# library(reshape2)
# library(tibble)

# prova:
saresp.loc <- 'http://www.ufpr.br/~aanjos/TRI/sinape/dados/saresp.dat'
dados <- read_fwf(saresp.loc, fwf_widths(c(1, -1, 1, -1, 12, 3, -1, rep(1,30)))), skip=3, na=' ')

dados <- dados %>%
  dplyr::select(c(-2,-4,-7)) %>% # usar namespace dos pacotes para não dar erro
  magrittr::set_colnames(c('grupo', 'escola', 'id', 'turno', paste('i',1:30,sep="")))
# sol em: https://stackoverflow.com/questions/28100780/use-with-replacement-functions-like-colname
s

# gabarito:
gab <- read_fwf(saresp.loc, fwf_widths(c(-20, rep(1, 30)))), n_max=3, na=' ')

gab2 <- gab %>%
  dplyr::select(-1) %>%
  magrittr::set_colnames(c(paste('i',1:30,sep="")))
```

Verificando os data.frames:

```
head(dados)
```

|                             | grupo<br><int> | escola<br><int> | id<br><chr>  | turno<br><chr> | i1<br><chr> | i2<br><chr> | i3<br><chr> | i4<br><chr> | i5<br><chr> | ▶ |
|-----------------------------|----------------|-----------------|--------------|----------------|-------------|-------------|-------------|-------------|-------------|---|
| 1                           | 1              | 1               | 011001138433 | m07            | A           | D           | D           | C           | A           |   |
| 2                           | 1              | 1               | 011002964093 | m07            | D           | B           | A           | C           | A           |   |
| 3                           | 1              | 1               | 011004154243 | m07            | A           | B           | D           | C           | A           |   |
| 4                           | 1              | 1               | 011005367283 | m07            | D           | C           | D           | C           | A           |   |
| 5                           | 1              | 1               | 011007519633 | m07            | D           | B           | D           | C           | A           |   |
| 6                           | 1              | 1               | 011008054863 | m07            | D           | D           | B           | C           | A           |   |
| 6 rows   1-10 of 35 columns |                |                 |              |                |             |             |             |             |             |   |

```
head(gab2)
```

|   | i1<br><chr> | i2<br><chr> | i3<br><chr> | i4<br><chr> | i5<br><chr> | i6<br><chr> | i7<br><chr> | i8<br><chr> | i9<br><chr> | ▶ |
|---|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|---|
| 1 | A           | B           | D           | C           | A           | A           | B           | C           | D           |   |
| 2 | A           | D           | C           | D           | A           | A           | D           | B           | C           |   |

|                             | i1<br><chr> | i2<br><chr> | i3<br><chr> | i4<br><chr> | i5<br><chr> | i6<br><chr> | i7<br><chr> | i8<br><chr> | i9<br><chr> |  |
|-----------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|--|
| 3                           | B           | D           | C           | A           | C           | A           | B           | A           | C           |  |
| 3 rows   1-10 of 31 columns |             |             |             |             |             |             |             |             |             |  |

Estimação dos parâmetros dos itens para os três turnos:

```

# preparando os dados

# separando os turnos
lista_turnos <- dados %>% dplyr::group_by(turno) %>% dplyr::do(data = (.)) %>% dplyr::select(data)
%>% purrr::map(identity)

turnos <- c("manha", "noite", "tarde")
for(i in 1:3){
  assign(turnos[i], lista_turnos$data[[i]])
}

gab3 <- matrix(9,nrow(gab2),ncol(gab2))

for (i in 1:nrow(gab3)) {
  for (j in 1:ncol(gab3)) {
    if (gab2[i,j]=="A") gab3[i,j] <- 1
    if (gab2[i,j]=="B") gab3[i,j] <- 2
    if (gab2[i,j]=="C") gab3[i,j] <- 3
    if (gab2[i,j]=="D") gab3[i,j] <- 4
  }
}

resp.manha <- manha[,5:34]
resp.manha <- as.matrix(resp.manha)
resp.m <- matrix(9,nrow(resp.manha),ncol(resp.manha))
resp.m[resp.manha=="A"] <- 1
resp.m[resp.manha=="B"] <- 2
resp.m[resp.manha=="C"] <- 3
resp.m[resp.manha=="D"] <- 4

for (i in 1:nrow(resp.m)) {
  for (j in 1:ncol(resp.m)) {
    if ((resp.m[i,j]!=gab3[1,j])&&(resp.m[i,j]!=9)) resp.m[i,j] <- 0
    if (resp.m[i,j]==gab3[1,j]) resp.m[i,j] <- 1
  }
}

resp.m[resp.m==9] <- NA

resp.tarde <- tarde[,5:34]
resp.tarde <- as.matrix(resp.tarde)
resp.t <- matrix(9,nrow(resp.tarde),ncol(resp.tarde))
resp.t[resp.tarde=="A"] <- 1
resp.t[resp.tarde=="B"] <- 2
resp.t[resp.tarde=="C"] <- 3
resp.t[resp.tarde=="D"] <- 4

for (i in 1:nrow(resp.t)) {
  for (j in 1:ncol(resp.t)) {
    if ((resp.t[i,j]!=gab3[2,j])&&(resp.t[i,j]!=9)) resp.t[i,j] <- 0
    if (resp.t[i,j]==gab3[2,j]) resp.t[i,j] <- 1
  }
}

resp.t[resp.t==9] <- NA

resp.noite <- noite[,5:34]
resp.noite <- as.matrix(resp.noite)
resp.n <- matrix(9,nrow(resp.noite),ncol(resp.noite))
resp.n[resp.noite=="A"] <- 1
resp.n[resp.noite=="B"] <- 2
resp.n[resp.noite=="C"] <- 3
resp.n[resp.noite=="D"] <- 4

for (i in 1:nrow(resp.n)) {

```

```

for (j in 1:ncol(resp.n)) {
  if ((resp.n[i,j]!=gab3[3,j])&&(resp.n[i,j]!=9)) resp.n[i,j] <- 0
  if (resp.n[i,j]==gab3[3,j]) resp.n[i,j] <- 1
}
}

resp.n[resp.n==9] <- NA

```

```

# estimativas

library(irtoys)

##### Obtencao das estimativas dos parametros dos itens para turno da manha via pacote "irtoys"

#resp.m.tpm <- tpm(resp.m,constraint=cbind(1:ncol(resp.m),1,0.25))
resp.m.tpm <- tpm(resp.m)
par.m.est <- coef(resp.m.tpm) # cc, bb, aa

### estimacao da proficiencia (funcao "eap" serve apenas para o modelo 3PL)
### matriz com a estimativa da proficiencia e o erro-padrao (1/sqrt(informacao(theta.est)))

theta.m.est <- eap(resp.m, cbind(par.m.est[,3],par.m.est[,2],par.m.est[,1]), qu=normal.qu())
prof.m.est <- theta.m.est[,1]

##### Obtencao das estimativas dos parametros dos itens para turno da tarde via pacote "irtoys"

#resp.t.tpm <- tpm(resp.t,constraint=cbind(1:ncol(resp.t),1,0.25))
resp.t.tpm <- tpm(resp.t)
par.t.est <- coef(resp.t.tpm) # cc, bb, aa

### estimacao da proficiencia (funcao "eap" serve apenas para o modelo 3PL)
### matriz com a estimativa da proficiencia e o erro-padrao (1/sqrt(informacao(theta.est)))

theta.t.est <- eap(resp.t, cbind(par.t.est[,3],par.t.est[,2],par.t.est[,1]), qu=normal.qu())
prof.t.est <- theta.t.est[,1]

##### Obtencao das estimativas dos parametros dos itens para turno da noite via pacote "irtoys"

#resp.n.tpm <- tpm(resp.n,constraint=cbind(1:ncol(resp.n),1,0.25))
resp.n.tpm <- tpm(resp.n)
par.n.est <- coef(resp.n.tpm) # cc, bb, aa

### estimacao da proficiencia (funcao "eap" serve apenas para o modelo 3PL)
### matriz com a estimativa da proficiencia e o erro-padrao (1/sqrt(informacao(theta.est)))

theta.n.est <- eap(resp.n, cbind(par.n.est[,3],par.n.est[,2],par.n.est[,1]), qu=normal.qu())
prof.n.est <- theta.n.est[,1]

```

### item a) Equalização via Regressão Linear

```

##### Equalizacao via regressão linear
# Equalização
#plot(par.m.est[15:19,3],par.n.est[15:19,3])
#reg.mn.a <- lm(par.n.est[15:19,3]~par.m.est[15:19,3]-1)

reg.mn.b <- lm(par.n.est[15:19,2]~par.m.est[15:19,2]) # soh usando os itens comuns (linhas de 15 a 19)

reg.mn.b$coefficients

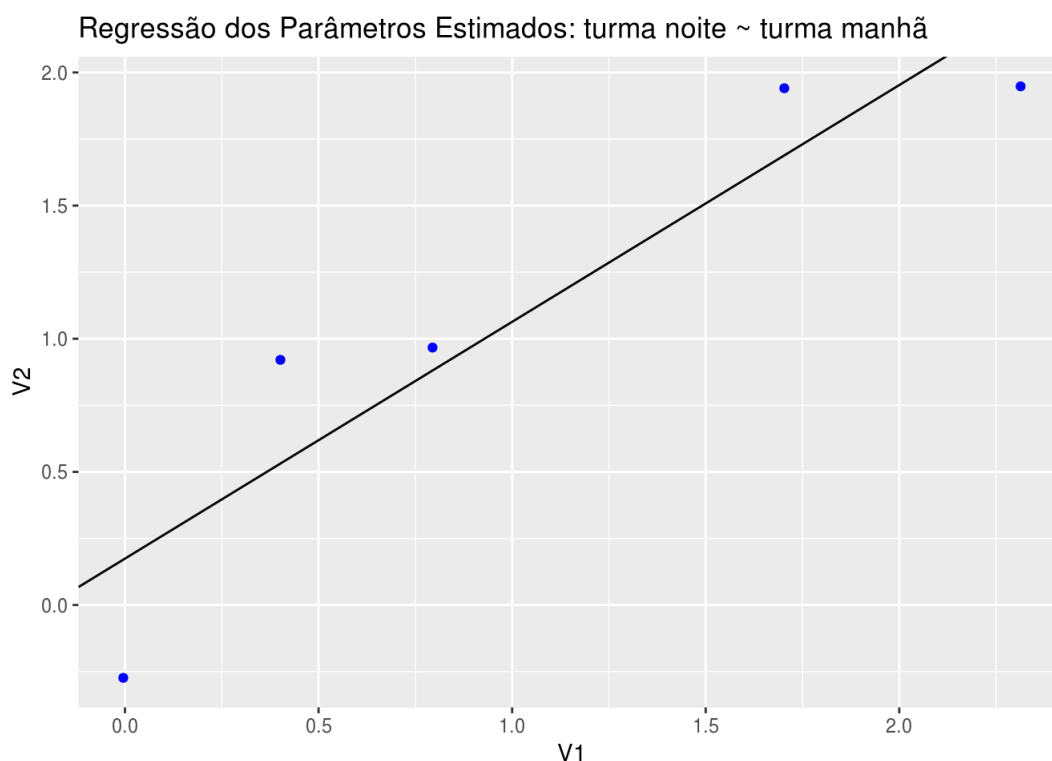
```

```
##      (Intercept) par.m.est[15:19, 2]
##      0.1743214      0.8891807
```

```
df_par <- as.data.frame(cbind(par.m.est[15:19,2], par.n.est[15:19,2]), row.names = FALSE)

df_coef <- as.data.frame(cbind(reg.mn.b$coefficients[1], reg.mn.b$coefficients[2]))

library(ggplot2)
library(magrittr)
df_par %>%
  ggplot()+
  geom_point(aes(x=V1, y=V2), colour="blue")+
  geom_abline(intercept = df_coef$V1, slope=df_coef$V2)+
  labs(title="Regressão dos Parâmetros Estimados: turma noite ~ turma manhã")
```



Temos, portanto, a seguinte relação entre os parâmetros de dificuldade  $b$  da turma da noite e turma da manhã:

$$b_{noite} = 0.1743214 + 0.8891807 \times b_{manhã}$$

Novos parâmetros da turma da noite - equalizados pelos parâmetros da turma da manhã, podem ser obtidos da seguinte forma:

$$b_{noite_{new}} = \frac{b_{noite} - 0.1743214}{0.8891807}$$

$$a_{noite_{new}} = a_{noite} \times 0.8891807$$

Os novos parâmetros da turma da noite equalizados pelos da manhã são apresentadas na tabela abaixo:

```
## colocando parametro b da noite na escala da manha

par.n.est.b.eq <- (par.n.est[,2]-reg.mn.b$coefficients[1])/reg.mn.b$coefficients[2]

## colocando parametro b da noite na escala da manha

par.n.est.a.eq <- par.n.est[,3]*reg.mn.b$coefficients[2]

tibble::as_tibble(list("Item" = paste0("item", 1:30),
  "b" = par.n.est.b.eq, "a" = par.n.est.a.eq))
```

| Item            | b        | a       |
|-----------------|----------|---------|
| 1-10 of 30 rows |          |         |
|                 | Previous | 123Next |

As proficiências equalizadas estimadas podem ser obtidas por:

```
## proficiencia dos alunos da noite na escala da manha
prof.n.est.eq <- (prof.n.est-reg.mn.b$coefficients[1])/reg.mn.b$coefficients[2]
```

$$\theta_{noite_{new}} = \frac{\theta_{noite} - 0.1743214}{0.8891807}$$

item b) Equalização via método Média-Desvio

```
##### Equalizacao pelo metodo media-desvio

alfa <- sd(par.m.est[15:19,2])/sd(par.n.est[15:19,2])

beta <- mean(par.m.est[15:19,2]) - alfa*mean(par.n.est[15:19,2])
```

Utilizando o método Média-Desvio, o qual consiste em utilizar a média e o desvio padrão do parâmetro  $b$  dos itens comuns a ambos os turnos, temos:

$$\alpha = \frac{\sigma_{b_{manh\tilde{a}}}}{\sigma_{b_{noite}}} = \frac{0.9513018}{0.9168048} = 1.037627$$

$$\beta = \bar{b}_{manh\tilde{a}} - \alpha \times \bar{b}_{noite} = 1.041596 - 1.037627 \times 1.100489 = -0.1003009$$

Em seguida, calculamos as proficiências da noite equalizadas pelas da manhã com o referido método:

```
## proficiencia dos alunos da noite na escala da manha

prof.nm <- alfa*prof.n.est+beta
```

As proficiências equalizadas são obtidas por:

$$\theta_{noite_{new}} = \alpha \times \theta_{noite} + \beta = 1.037627 \times \theta_{noite} - 0.1003009$$

item c) Gráficos

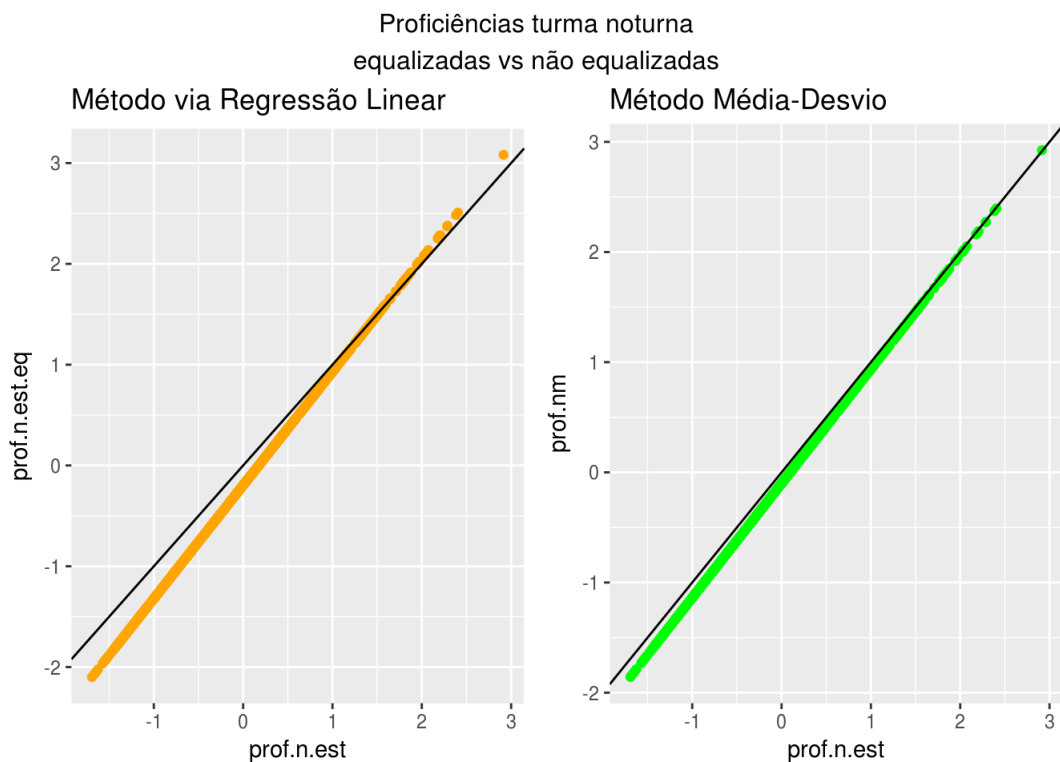
Abaixo é apresentado o gráfico de dispersão com as proficiências  $\theta$  da escala da noite equalizadas e sem equalização:

```
## proficiencia dos alunos da noite na escala da manha

# equalizacao via regressao
library(ggplot2)
p1 <- ggplot()+
  geom_point(aes(x=prof.n.est, y=prof.n.est.eq), colour="orange")+
  geom_abline(intercept = 0, slope=1)+
  labs(title="Método via Regressão Linear")

# equalizacao metodo media-desvio
p2 <- ggplot()+
  geom_point(aes(x=prof.n.est, y=prof.nm), colour="green")+
  geom_abline(intercept = 0, slope=1)+
  labs(title="Método Média-Desvio")

library(gridExtra)
grid.arrange(p1, p2, ncol=2, top="Proficiências turma noturna\nnequalizadas vs não equalizadas")
```



Dos gráficos anteriores, constata-se que o método Média-Desvio resulta em um ajuste mais refinado das proficiências ao longo de toda escala de  $\theta$ . Ainda sim, é possível dizer que o ajuste para as proficiências mais altas é melhor e praticamente se sobrepõe à reta 0-1. O gráfico do método via Regressão Linear, possui um pior ajuste à reta 0-1 para os valores mais baixos e mais altos de  $\theta$ . Seu melhor ajuste ocorre para valores intermediários de proficiência, mas resulta em um ajuste consideravelmente pior do que aquele ao se utilizar o método Média-Desvio.