

# Prob 5 CP 2

15 de junho de 2016

## Análise do efeito da beleza na pontuação média da didática dos professores

De posse desses dados, queremos avaliar se a beleza possui um efeito significativo no score dos professores, levando em conta os demais fatores que foram identificados como tendo possíveis efeitos, que são as variáveis rank, ethnicity, gender, language, age, cls\_\*, pic\_outfit e pic\_color. Para isso faça o seguinte:

1. Comece a análise como se começa toda análise: faça um descritivo

### Análise dos dados

A partir do conjunto de dados fornecido, serão analisadas as seguintes variáveis para responder as perguntas:

- **score**: Pontuação média na avaliação docente
- **bty\_avg**: Média de beleza dos professores
- **gender**: Sexo dos professores
- **ethnicity**: Etnia dos professores

Foram escolhidos duas variáveis além das solicitadas que podem influenciar em conjunto com a beleza na pontuação dos professores.

Nossa hipótese é que o sexo e etnia dos professores influenciam no efeito que a beleza tem sob a pontuação. Mas antes ver algumas informações sobre os dados.

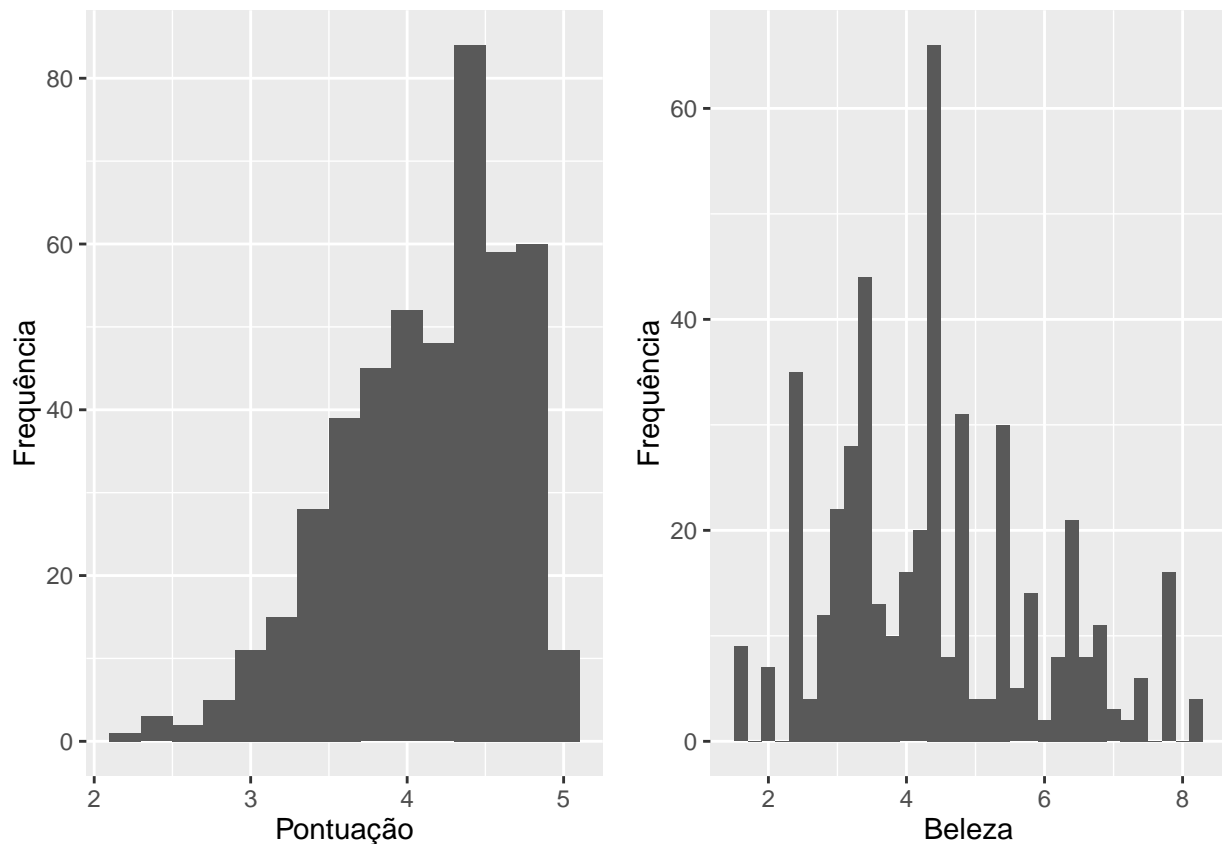
```
eval_df <- eval %>% select(score, bty_avg, gender, ethnicity)
summary(eval_df)
```

score		bty_avg		gender		ethnicity	
Min.	:2.300	Min.	:1.667	female:	195	minority	: 64
1st Qu.	:3.800	1st Qu.	:3.167	male	:268	not minority:	399
Median	:4.300	Median	:4.333				
Mean	:4.175	Mean	:4.418				
3rd Qu.	:4.600	3rd Qu.	:5.500				
Max.	:5.000	Max.	:8.167				

Podemos perceber que existem mais professores do sexo masculino na nossa amostra, além de termos poucos professores que representam a minoria étnica, o que já era esperado. Vamos analisar com mais detalhes as variáveis score e beleza média.

Abaixo temos dois gráficos onde podemos visualizar a distribuição das variáveis pontuação e beleza. É bem evidente a diferença entre como os professores são avaliados pelas suas capacidades e como são avaliados pela sua beleza.

```
gridExtra::grid.arrange(
  ggplot(eval_df, aes(x=score)) +
    geom_histogram(binwidth = 0.2) +
    xlab("Pontuação") +
    ylab("Frequência"),
  ggplot(eval_df, aes(x=bty_avg)) +
    geom_histogram(binwidth = 0.2) +
    xlab("Beleza") +
    ylab("Frequência"),
  ncol = 2
)
```



Logo, isso indica que há pouca chance de se ter um efeito forte da beleza sob a pontuação dos professores. A seguir iremos aplicar regressão linear para entender melhor essa relação.

## Relação entre Beleza e Score

Para explicar o efeito que a beleza exerce sobre o score, definimos o seguinte modelo:

$$score = B_0 + B_1 \cdot bty\_avg$$

Abaixo podemos ver o efeito que a beleza exerce sob a pontuação do professor.

```
eval_simples <- eval %>% select(score, bty_avg)
modelEval <- lm(score ~ bty_avg, data = eval_simples)
```

```
confint(modelEval)
```

```
                2.5 %      97.5 %  
(Intercept) 3.73070764 4.02996827  
bty_avg      0.03462292 0.09865116
```

```
summary(modelEval)
```

Call:

```
lm(formula = score ~ bpy_avg, data = eval_simples)
```

Residuals:

```
      Min       1Q   Median       3Q      Max  
-1.9246 -0.3690  0.1420  0.3977  0.9309
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)  3.88034    0.07614   50.96 < 2e-16 ***  
bpy_avg      0.06664    0.01629    4.09 5.08e-05 ***  
---
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

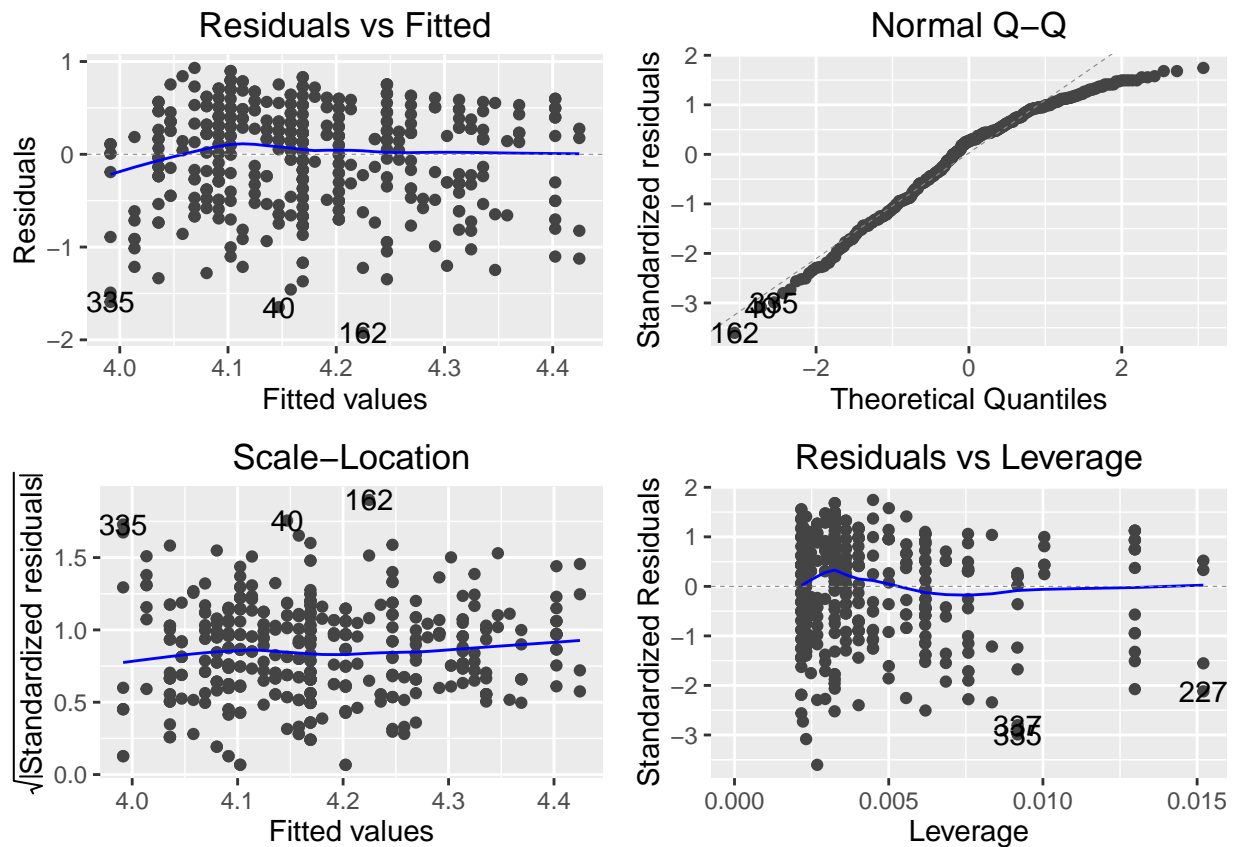
Residual standard error: 0.5348 on 461 degrees of freedom

Multiple R-squared: 0.03502, Adjusted R-squared: 0.03293

F-statistic: 16.73 on 1 and 461 DF, p-value: 5.083e-05

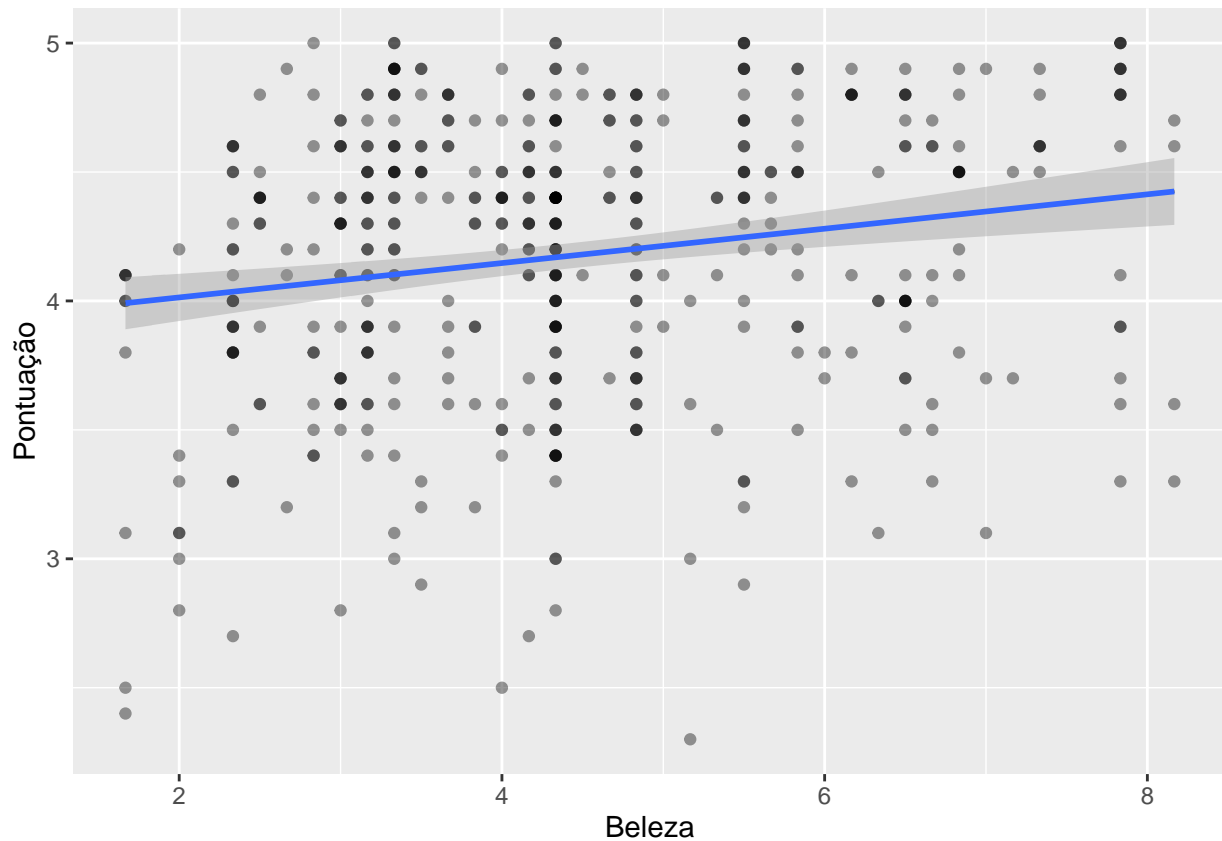
Como podemos ver acima, temos o *R-squared* com cerca de 3%, o que explica muito pouco sobre o efeito da beleza na pontuação dos professores. Abaixo é possível perceber que os pressupostos da regressão linear foram atendidos. Temos poucos *outliers* e os resíduos se comportam como esperado.

```
autoplot(modelEval)
```



Por fim, a visualização abaixo deixa bem mais claro que, para a amostra observada, não há correlação entre a beleza e o score.

```
ggplot(eval_simples, aes(x = bty_avg, y = score)) +
  geom_point(alpha = 0.4) +
  geom_line(aes(y = predict(modelEval, eval_simples))) +
  stat_smooth(method = "lm") +
  xlab("Beleza") +
  ylab("Pontuação")
```



## Influência de outras variáveis

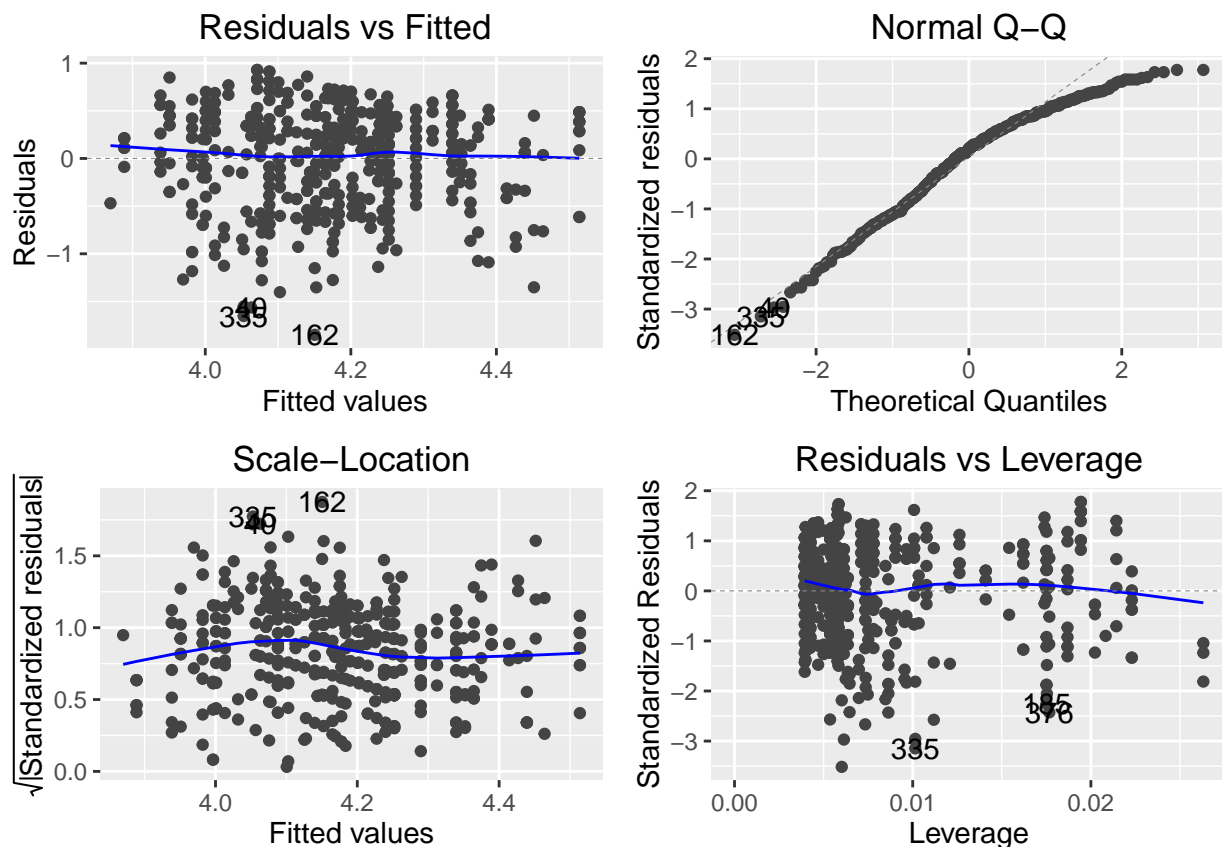
Agora será analisado a influência das variáveis etnia e genero no efeito que a beleza exerce sob o score.

Nosso modelo provavelmente irá aplicar o sexo masculino na variável de genero como visto em sala de aula. Acredito que essa variável terá uma influência positiva sobre os efeitos da beleza, bem como a variável de etnia também terá efeito positivo sobre o modelo. Esse modelo é definido por:

$$score = B_0 + B_1 \cdot bty\_avg + B_2 \cdot gender + B_3 \cdot ethnicity$$

```
bty_df <- eval %>% select(score, gender, ethnicity, bty_avg)

model_bty <- lm(score ~ bty_avg + gender + ethnicity, data = bty_df)
autoplot(model_bty)
```



Os pressupostos necessários para satisfazer a regressão linear parecem estar dentro do esperado. Poucos *outliers* e os resíduos tem uma distribuição semelhante a normal.

Agora vamos verificar os efeitos das variáveis no modelo. A partir do sumário desse modelo, é possível afirmar que houve uma melhora no efeito das variáveis adicionadas em relação a pontuação. Temos o *R-squared* com cerca de 6% e também houve um acréscimo no efeito da beleza sobre o score.

```
summary(model_bty)
```

Call:

```
lm(formula = score ~ bty_avg + gender + ethnicity, data = bty_df)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.8504	-0.3669	0.1104	0.4107	0.9287

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.65789	0.10388	35.212	< 2e-16 ***
bty_avg	0.07479	0.01624	4.606	5.32e-06 ***
gendermale	0.16415	0.05046	3.253	0.00123 **
ethnicitynot minority	0.10609	0.07159	1.482	0.13906

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.528 on 459 degrees of freedom

Multiple R-squared: 0.0636, Adjusted R-squared: 0.05748  
F-statistic: 10.39 on 3 and 459 DF, p-value: 1.255e-06

Porém, percebemos que o intervalo de confiança da variável *ethnicity* passa pelo zero, isso quer dizer que essa variável não exerce nenhum efeito em nosso modelo. Portanto, ela será descartada.

```
confint(model_bty)
```

	2.5 %	97.5 %
(Intercept)	3.45374105	3.8620322
bty_avg	0.04288023	0.1066951
gendermale	0.06498371	0.2633181
ethnicitynot minority	-0.03459934	0.2467832

Dessa forma, temos o seguinte modelo ajustado:

$$\text{score} = B_0 + B_1 \cdot \text{bty\_avg} + B_2 \cdot \text{gender}$$

```
model_bty <- lm(score ~ bty_avg + gender, data = bty_df)
summary(model_bty)
```

Call:

```
lm(formula = score ~ bty_avg + gender, data = bty_df)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.8305	-0.3625	0.1055	0.4213	0.9314

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.74734	0.08466	44.266	< 2e-16 ***
bty_avg	0.07416	0.01625	4.563	6.48e-06 ***
gendermale	0.17239	0.05022	3.433	0.000652 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5287 on 460 degrees of freedom

Multiple R-squared: 0.05912, Adjusted R-squared: 0.05503

F-statistic: 14.45 on 2 and 460 DF, p-value: 8.177e-07

## Conclusões

Por fim, continuamos com pouco efeito das variáveis utilizadas em relação ao score. Esta análise de dados nos dá um forte indicativo de que não há relação entre a beleza e o score que os professores atingem em suas avaliações docentes. Isso parece fazer sentido, já que tem outras características mais interessantes que podem ter relação com a pontuação do professor, algumas delas seriam: empatia, domínio do conteúdo, assiduidade, pontualidade, metodologia de avaliação, entre outras. Essas seriam boas métricas que podem mostrar fortes relacionamentos com o score do professor.