

Trabalho 4

Daniel Krügel

2023-02-20

Estarei testando o pacote `gtsummary` para criar as tabelas deste trabalho, portanto deixarei todos os chunks de códigos com `echo = TRUE`.

Questão 13

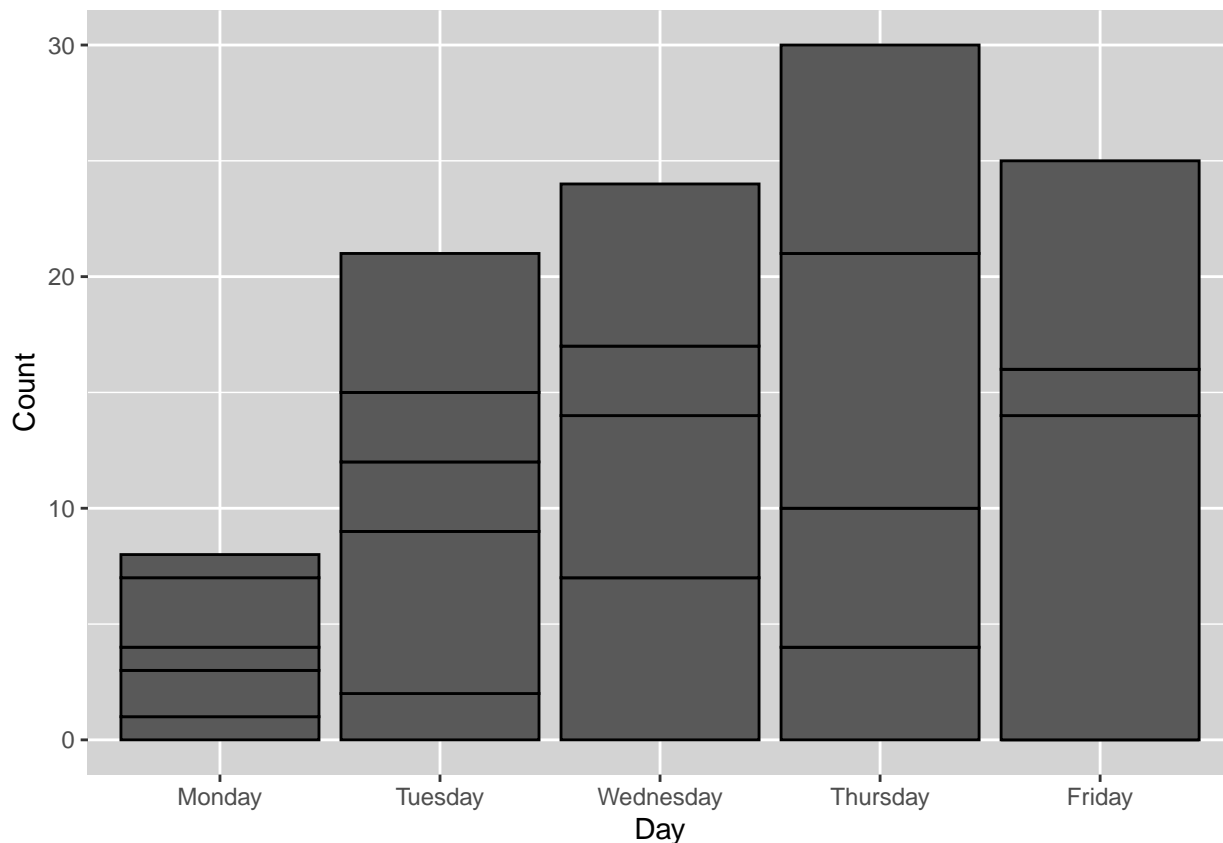
a)

A população de inferência é a população que frequenta o starbucks no período das 8:00 até as 8:30, a amostra foi quem frequenta o starbucks de Lincoln, NE.

b)

```
starbucks$Day <- factor(starbucks$Day,
                        levels = c('Monday', "Tuesday", "Wednesday", "Thursday", "Friday"))

starbucks %>%
  ggplot(aes(x = Day, y = Count)) +
  geom_bar(stat = "identity", color = "black") +
  theme(
    panel.background = element_rect(fill = "lightgrey",
                                     colour = "lightgrey",
                                     linewidth = 0.5, linetype = "solid"),
    panel.grid.major = element_line(linewidth = 0.5, linetype = 'solid',
                                     colour = "white"),
    panel.grid.minor = element_line(linewidth = 0.25, linetype = 'solid',
                                     colour = "white"))
```



c)

```
fit1 <- glm(Count ~ Day,
            family = poisson(link='log'),
            data = starbucks)
```

```
fit1 %>%
  tbl_regression(add_estimate_to_reference_rows = F) %>%
  add_q()
```

```
## add_q: Adjusting p-values with
## `stats::p.adjust(x$table_body$p.value, method = "fdr")`
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

| Characteristic | log(IRR) | 95% CI | p-value | q-value |
|----------------|----------|-----------|---------|---------|
| Day | | | | |
| Monday | — | — | | |
| Tuesday | 0.97 | 0.19, 1.8 | 0.020 | 0.020 |
| Wednesday | 1.1 | 0.34, 2.0 | 0.007 | 0.009 |
| Thursday | 1.3 | 0.59, 2.2 | <0.001 | 0.004 |
| Friday | 1.1 | 0.39, 2.0 | 0.005 | 0.009 |

O pacote omitiu o intercepto, já pedi para ele exponenciar os coeficientes pois foi utilizado a função log, o valor do intercepto, que equivale a segunda-feira, é 0.4700036

I)

Lembrando que usamos a função de ligação log, então para conseguir as estimativas diárias precisamos passar um exponencial nos coeficientes.

```
fit1 %>%
  tbl_regression(exponentiate = T) %>%
  add_global_p()
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

| Characteristic | IRR | 95% CI | p-value |
|----------------|------|------------|---------|
| Day | | | 0.005 |
| Monday | — | — | |
| Tuesday | 2.62 | 1.21, 6.31 | |
| Wednesday | 3.00 | 1.41, 7.13 | |
| Thursday | 3.75 | 1.80, 8.78 | |
| Friday | 3.13 | 1.47, 7.41 | |

II)

```
xtable>Anova(fit1))
```

```
## % latex table generated in R 4.1.2 by xtable 1.8-4 package
## % Mon Feb 20 21:25:57 2023
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrr}
## \hline
## & LR Chisq & Df & Pr(>$Chisq) & \\
## \hline
## Day & 15.00 & 4 & 0.0047 & \\
## \hline
## \end{tabular}
## \end{table}
```

Utilizando o pacote no qual estou aprendendo a usar, o valor do teste LR é omitido, porém o p-valor é exibido a cada covariável.

```
fit1 %>%
  tbl_regression() %>%
  add_global_p()
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

| Characteristic | log(IRR) | 95% CI | p-value |
|----------------|----------|-----------|---------|
| Day | | | 0.005 |
| Monday | — | — | |
| Tuesday | 0.97 | 0.19, 1.8 | |
| Wednesday | 1.1 | 0.34, 2.0 | |

| Characteristic | log(IRR) | 95% CI | p-value |
|----------------|----------|-----------|---------|
| Thursday | 1.3 | 0.59, 2.2 | |
| Friday | 1.1 | 0.39, 2.0 | |

O Teste LR demonstrou que a variável Day é relevante para a contagem de pessoas presentes no Starbucks com um P-valor de 0.0046982

III)

IV)

Aqui temos a média estimada para cada dia da semana

```
quart <- summary(starbucks$Count)
quart[4]*fit1$coefficients[1:5]
```

```
## (Intercept) DayTuesday DayWednesday DayThursday DayFriday
##      2.030416      4.169149      4.746005      5.709985      4.922356
```

E aqui vemos a média observada em cada dia da semana

```
starbucks %>%
  group_by(Day) %>%
  reframe(total = mean(Count))
```

```
## # A tibble: 5 x 2
##   Day      total
##   <fct>    <dbl>
## 1 Monday      1.6
## 2 Tuesday      4.2
## 3 Wednesday    4.8
## 4 Thursday      6
## 5 Friday       5
```

Os intervalos de confiança para a média de cada dia da semana:

```
K <- matrix(data = c(1*quart[4], 0, 0, 0, 0,
                     0, 1*quart[4], 0, 0, 0,
                     0, 0, 1*quart[4], 0, 0,
                     0, 0, 0, 1*quart[4], 0,
                     0, 0, 0, 0, 1*quart[4]), nrow = 5, ncol = 5, byrow = TRUE)
linear.combo <- mcprofile(object = fit1, CM = K)
ci.beta <- confint(object = linear.combo, level = 0.95)
Estimativas <- ci.beta$confint
rownames(Estimativas) <- c("Segunda", "Terça", "Quarta", "Quinta", "Sexta")
Estimativas
```

```
##           lower      upper
## Segunda -2.1739801  5.208082
## Terça    0.1431574  8.835735
## Quarta   0.8217293  9.359534
## Quinta   1.9299624 10.247206
## Sexta    1.0267539  9.520816
```

d)

Ambas as formas de escrita do teste de hipóteses descrevem um modelo linear de variáveis, porém quando levado para modelos lineares generalizados precisamos levar em conta a função de ligação utilizada no modelo. Como utilizamos a função canônica de ligação para o GLM da família Poisson, sua link é a Função logarítmica, o que deve de ser levado a diante no teste de razões LRT seria algo como:

$$H_0 : e_1^\mu = e_2^\mu = e_3^\mu = e_4^\mu = e_5^\mu$$

$$H_1 : \mu \neq \mu \neq \mu \neq \mu \neq \mu$$