

Trabalho 4

Daniel Krügel

2023-02-16

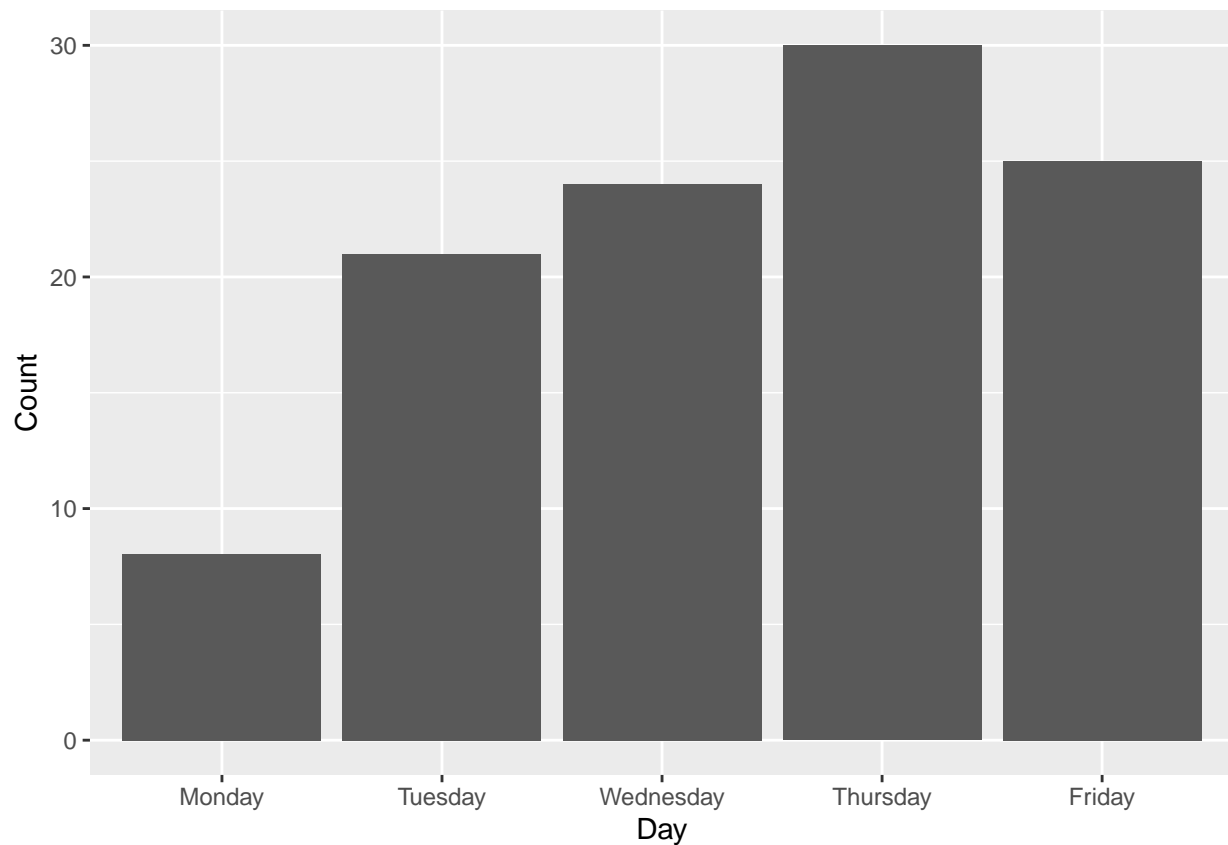
Qestão 13

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")
```



c)

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

summary(fit1)

##
## Call:
## glm(formula = Count ~ Day, family = poisson(link = "log"), data = starbucks)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4641  -0.8832  -0.5099   0.9392   3.2908
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.4700     0.3536   1.329 0.183726
## DayTuesday     0.9651     0.4155   2.323 0.020188 *
## DayWednesday   1.0986     0.4082   2.691 0.007123 **
## DayThursday    1.3218     0.3979   3.322 0.000895 ***
## DayFriday      1.1394     0.4062   2.805 0.005030 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 87.432  on 24  degrees of freedom
## Residual deviance: 72.431  on 20  degrees of freedom
## AIC: 150.9
##
## Number of Fisher Scoring iterations: 5
```

I)

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

```
exp(
  coef(fit1)
)
```

	(Intercept)	DayTuesday	DayWednesday	DayThursday	DayFriday
	1.600	2.625	3.000	3.750	3.125

Aqui o intercepto equivale a segunda feira.

II)

```
Anova(fit1)

## Analysis of Deviance Table (Type II tests)
##
## Response: Count
##      LR Chisq Df Pr(>Chisq)
```

```
## Day    15.002  4    0.004698 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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 -2.1678897  5.204594
## Terça    0.1480009  8.829256
## Quarta   0.8264062  9.353099
## Quinta   1.9344049 10.240834
## Sexta    1.0313840  9.514393
```

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:

$$H_0 : e^\mu = e^\mu = e^\mu = e^\mu = e^\mu$$

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