

Projeto IMDB - Enunciado 2

Natanael Magalhães Cardoso*

31/05/2021

Pergunta 01 (ok)

```
IC_media_Z <- function(dados, confianca, dp) {  
  xbarra = mean(dados, na.rm=TRUE)  
  z = qnorm((1 + confianca) / 2)  
  n = length(dados)  
  e = z * (dp / sqrt(n))  
  LS = xbarra + e  
  LI = xbarra - e  
  return(c(LI, LS))  
}  
IC_media_Z(imdb2$Lucro, 0.95, 55e6)
```

```
## [1] 14853397 20182038
```

O intervalo de confiança da média considerando variância populacional conhecida é $IC = [14.853.397; 20.182.038]$.

Pergunta 02 (ok)

```
IC_media_t <- function(dados, confianca) {  
  xbarra = mean(dados, na.rm=TRUE)  
  n = length(dados)  
  t = qt((1 + confianca) / 2, (n - 1))  
  e = t * (sd(dados, na.rm=TRUE) / sqrt(n))  
  LS = xbarra + e  
  LI = xbarra - e  
  return(c(LI, LS))  
}  
IC_media_t(imdb2$Lucro, 0.95)
```

```
## [1] 14467782 20567653
```

O intervalo de confiança da média considerando variância populacional desconhecida é $IC = [14.467.782; 20.567.653]$.

*nUSP: 8914122. Usando dados para o segundo grupo (datas entre 2000-2010)

Pergunta 03 (ok)

```
imdb2 <- imdb2 %>% mutate(
  Aprovado=ifelse(nota_imdb >= 6.5, TRUE, FALSE)
)
df_g2 <- imdb2 %>% mutate(
  Aprovado=ifelse(nota_imdb >= 6.5, TRUE, FALSE)
)
```

Pergunta 04 (ok)

```
IC_prop <- function(dados, confianca) {
  n = length(dados)
  z = qnorm((1 + confianca) / 2)
  p_chapeu = table(dados) / n
  e = z * sqrt((p_chapeu * (1 - p_chapeu)) / n)
  LS = p_chapeu + e
  LI = p_chapeu - e
  return(c(LI, LS))
}
IC_prop(imdb2$Aprovado, 0.80)
```

```
##      FALSE      TRUE      FALSE      TRUE
## 0.4692034 0.4991362 0.5008638 0.5307966
```

```
IC_prop(imdb2$Aprovado, 0.95)
```

```
##      FALSE      TRUE      FALSE      TRUE
## 0.4608234 0.4907562 0.5092438 0.5391766
```

```
IC_prop(imdb2$Aprovado, 0.99)
```

```
##      FALSE      TRUE      FALSE      TRUE
## 0.4532159 0.4831488 0.5168512 0.5467841
```

Para Aprovado = TRUE, os intervalos de confiança para proporção são $IC_{80\%} = [0,499;0,531]$, $IC_{95\%} = [0,491;0,539]$ e $IC_{99\%} = [0,483;0,547]$ para os níveis de confiança de 80, 95 e 99%, respectivamente. Foi notado que o intervalo **diminuiu** com o aumento do nível de confiança. Consequentemente, para Aprovado = FALSE, o intervalo **aumentou** com o aumento do nível de confiança.

Pergunta 05 (ok)

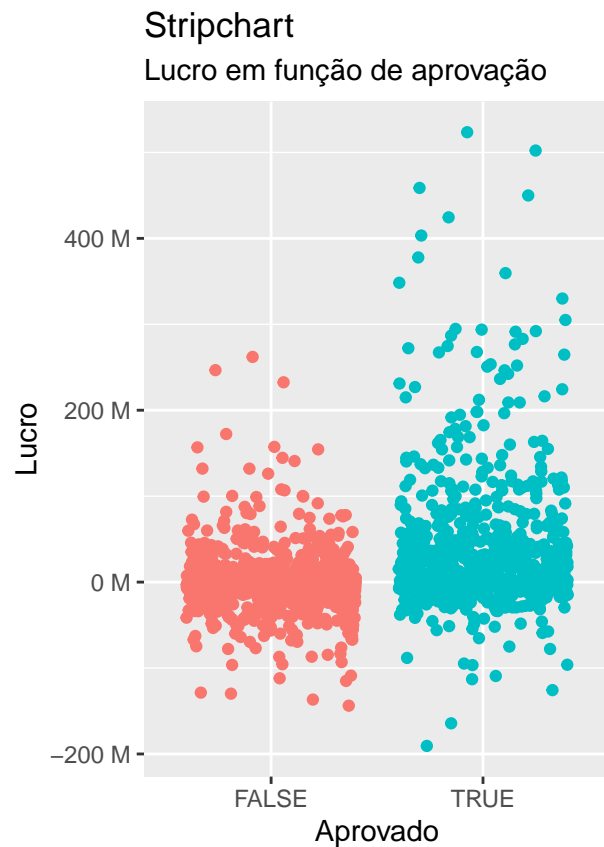
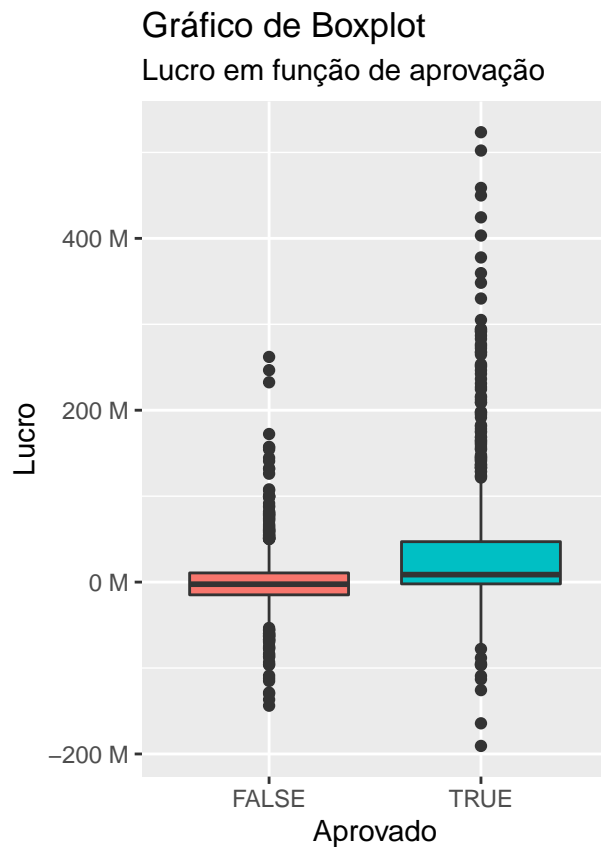
```
IC_var <- function(dados, confianca) {
  gl = length(dados) - 1
  S = var(dados, na.rm=TRUE)
  chi_lower = qchisq((1 + confianca) / 2, gl)
  chi_upper = qchisq((1 - confianca) / 2, gl)
  LI = sqrt(gl * S / chi_lower)
  LS = sqrt(gl * S / chi_upper)
  return(c(LI, LS))
}
IC_var(imdb2$Lucro, 0.95)
```

```
## [1] 60830134 65146228
```

O intervalo de confiança da variância considerando um nível de confiança de 95% é $IC = [60.830.134; 65.146.228]$.

Pergunta 06

```
p1 <- df_g2 %>% ggplot(aes(x=Aprovado, y=Lucro)) +  
  geom_boxplot(aes(fill=Aprovado)) +  
  theme(legend.position="none") +  
  scale_y_continuous(labels=unit_format(unit="M", scale=1e-6)) +  
  labs(  
    x="Aprovado",  
    y="Lucro",  
    title="Gráfico de Boxplot",  
    subtitle="Lucro em função de aprovação"  
  )  
  
p2 <- df_g2 %>% ggplot(aes(x=Aprovado, y=Lucro)) +  
  theme(legend.position="none") +  
  geom_jitter(aes(colour=Aprovado)) +  
  scale_y_continuous(labels=unit_format(unit="M", scale=1e-6)) +  
  labs(  
    x="Aprovado",  
    y="Lucro",  
    title="Stripchart",  
    subtitle="Lucro em função de aprovação"  
  )  
  
ggarrange(p1, p2, ncol=2)
```



Pergunta 07

Pergunta 08

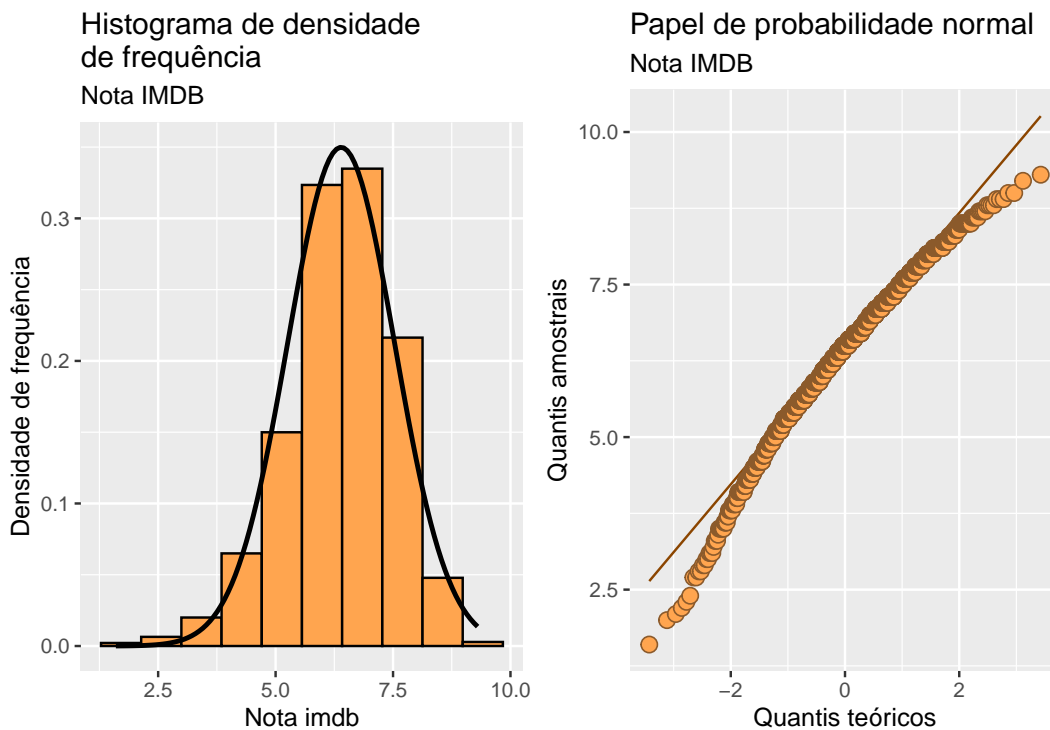
Pergunta 09

```
fit <- fitdistr(df_g2$nota_imdb, "normal")
MI <- fit$estimate[1]
SIGMA <- fit$estimate[2]

plot3 <- df_g2 %>% ggplot(aes(x=nota_imdb)) +
  geom_histogram(aes(y=..density..), bins=10, colour="black", fill="tan1") +
  labs(
    x="Nota imdb",
    y="Densidade de frequência",
    title="Histograma de densidade\nde frequência",
    subtitle="Nota IMDB"
  ) +
  stat_function(
    fun=dnorm,
    args=list(mean=MI, sd=SIGMA),
    size=1,
    col="black",
    linetype=1
  )
```

```
plot4 <- df_g2 %>% ggplot(aes(sample=nota_imdb)) +
  geom_qq_line(colour="darkorange4") +
  geom_qq(shape=21, colour="tan4", fill="tan1", size=3) +
  labs(
    x="Quantis teóricos",
    y="Quantis amostrais",
    title="Papel de probabilidade normal",
    subtitle="Nota IMDB"
  )

ggarrange(plot3, plot4, ncol=2)
```



Pergunta 10

```
x = df_g2$nota_imdb
KS_teste <- ks.test(x, "pnorm", mean=mean(x), sd=sd(x))
KS_teste
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: x
## D = 0.052888, p-value = 0.0002108
## alternative hypothesis: two-sided
```