

ANOVA - Estudo com ativos financeiros

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Programa de Pós-Graduação em Computação Aplicada

Disciplina: Análise Estatística de Dados e Informações - AEDI
Professor: João Gabriel de Moraes Souza
Aluno: Felipe Schiavon de Oliveira (matrícula: 20/0077104)

Introdução

Informações sobre os dados

Para a realização da atividade, optou-se por selecionar ações de grandes farmacêuticas envolvidas no processo de produção de vacinas contra o Covid-19: Pfizer (PFE), Moderna (MRNA), Johnson & Johnson (JNJ) e AstraZeneca (AZN). Utilizou-se, também, o Índice Nasdaq Composite (^IXIC), para fins de comparação.

Considerando que só há dados disponíveis da Moderna a partir de 07/12/2018, definiu-se essa data como o início e o dia 29/08/2021 para a data de fim para a coleta de dados. Utilizou-se os dados da variável Close, que corresponde ao valor da ação no fechamento.

ANOVA

A ANOVA - Análise de Variância é um procedimento estatístico usado para comparar a distribuição de grupos em amostras independentes. Visa, essencialmente, verificar se existe uma diferença significativa entre as médias e se os fatores exercem influência em alguma variável dependente.

Para avaliar as diferenças, define-se uma hipótese nula - H_0 (chamada também de hipótese sem efeitos) que representará, para fins deste trabalho, que não existe diferença entre a taxa de retorno dos ativos. Dessa forma, as médias entre os ativos não seriam estatisticamente diferentes.

O pressuposto para isso é de que a variabilidade das observações em cada grupo em relação à média do grupo é a mesma. Isso significa que supomos que há homocedasticidade entre as distribuições.

A hipótese alternativa - H_1 , representa, neste caso, que pelo menos um ativo não possui média igual à dos outros ativos.

Hipóteses:

- H_0 (hipótese nula): Não existe diferença entre a taxa de variação de preços dos ativos.
- H_1 (hipótese alternativa): Há pelo menos um ativo com taxa de variação de preço diferente dos demais.

Importando bibliotecas

```
suppressMessages(library(car))
suppressMessages(library(agricolae))
suppressMessages(library(dplyr))
suppressMessages(library(ggplot2))
suppressMessages(library(FSA))
```

Criando funções

```
wilcox_teste <- function(f., groups, values, data, verbose = FALSE){
  # f.: formula, groups: character(1), values: character(2)
  # results: htest

  index <- (data[[groups]] %in% values)
  data_sub <- data[index,]

  test <- wilcox.test(f., data = data_sub, exact = FALSE)

  if (verbose) {
    cat("=====\n")
    msg <- paste0(
      values, collapse = ' VS '
    )
    cat("Testando diferenças para: ", msg, '\n')
    print(test)
    cat("-----\n")
    return(invisible(test))
  }

  return(test)
}
```

Carregando dados

```
dados = read.csv("dados/taxa_retorno_grupos.csv", stringsAsFactors = TRUE)
head(dados)
```

```
##   X ativo      taxa
## 1 0   PFE  0.010928919
## 2 1   PFE -0.008783761
## 3 2   PFE  0.003408314
## 4 3   PFE  0.009284464
## 5 4   PFE -0.017276250
## 6 5   PFE -0.015753398
```

Verificação da homogeneidade

Para a verificação da homogeneidade das distribuições, serão aplicados dois testes estatísticos: teste de Levene e teste de Bartlett.

Teste de Levene

O teste de Levene nos permite averiguar a homogeneidade das variâncias entre os ativos.

```
leveneTest(taxa ~ ativo, dados, center=mean)

## Levene's Test for Homogeneity of Variance (center = mean)
##           Df F value    Pr(>F)
## group      4 279.08 < 2.2e-16 ***
##           3420
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para avaliar se a hipótese nula será rejeitada, devemos comparar se o valor p é inferior ao intervalo de confiança definido. Para este trabalho, será considerado o intervalo de confiança de 5% em todos os testes, ou seja, 0.05.

Como o valor p obtido no teste de Levene (2.2e-16) é inferior ao intervalo de confiança (0.05), rejeita-se a hipótese nula (H_0 - não existe diferença entre a taxa de variação de preços dos ativos). Ou seja, não há homocedasticidade entre as distribuições dos ativos. As variâncias são diferentes nos grupos, uma vez que a significância associada ao teste é inferior a 0.05, e assume-se a heterocedasticidade da distribuição.

Teste de Bartlett

```
bartlett.test(taxa ~ ativo, dados)

##
## Bartlett test of homogeneity of variances
##
## data: taxa by ativo
## Bartlett's K-squared = 1945.9, df = 4, p-value < 2.2e-16
```

Além do teste de Levene, realizou-se também o teste de Bartlett para verificar a homogeneidade das variâncias.

Como o valor p obtido no teste de Bartlett (2.2e-16) também é inferior ao intervalo de confiança (0.05), rejeita-se a hipótese nula (H_0 - não existe diferença entre a taxa de variação de preços dos ativos). Ou seja, não há homocedasticidade entre as distribuições dos ativos e assume-se que a distribuição dos dados possui uma estrutura heterocedástica.

ANOVA

Realiza-se, agora, a Análise de Variância - ANOVA.

```
anova <- aov(taxa ~ ativo, data = dados)
resumo <- summary(anova);resumo

##           Df Sum Sq Mean Sq F value    Pr(>F)
## ativo      4 0.0143 0.003580    4.8 0.000734 ***
## Residuals 3420 2.5510 0.000746
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Observando-se o resultado obtido ao calcular o valor p (0.000734) da ANOVA, verifica-se que há aproximadamente 0,07% de chance de errar caso a hipótese nula seja rejeitada. Considerando o valor do intervalo de confiança estipulado ($\alpha = 0,05$ ou 5%), pode-se rejeitar a hipótese nula.

A hipótese nula (H_0) é que os erros são homocedásticos. Ou seja, rejeitar a hipótese nula implica concluir que não há homocedasticidade e, assim, as médias entre os ativos não são iguais. Assume-se, portanto, a heterocedasticidade das distribuições, em que as variâncias não são iguais para todas as observações.

Antes de realizar o teste de hipótese, deve-se verificar se as premissas para a ANOVA de um fator foram atendidas. Os dados da amostra são independentes, pois correspondem à taxa de variação individual de cada ativo.

Porém, é necessário confirmar todas as suposições para a ANOVA:

- Homocedasticidade
- Normalidade dos dados

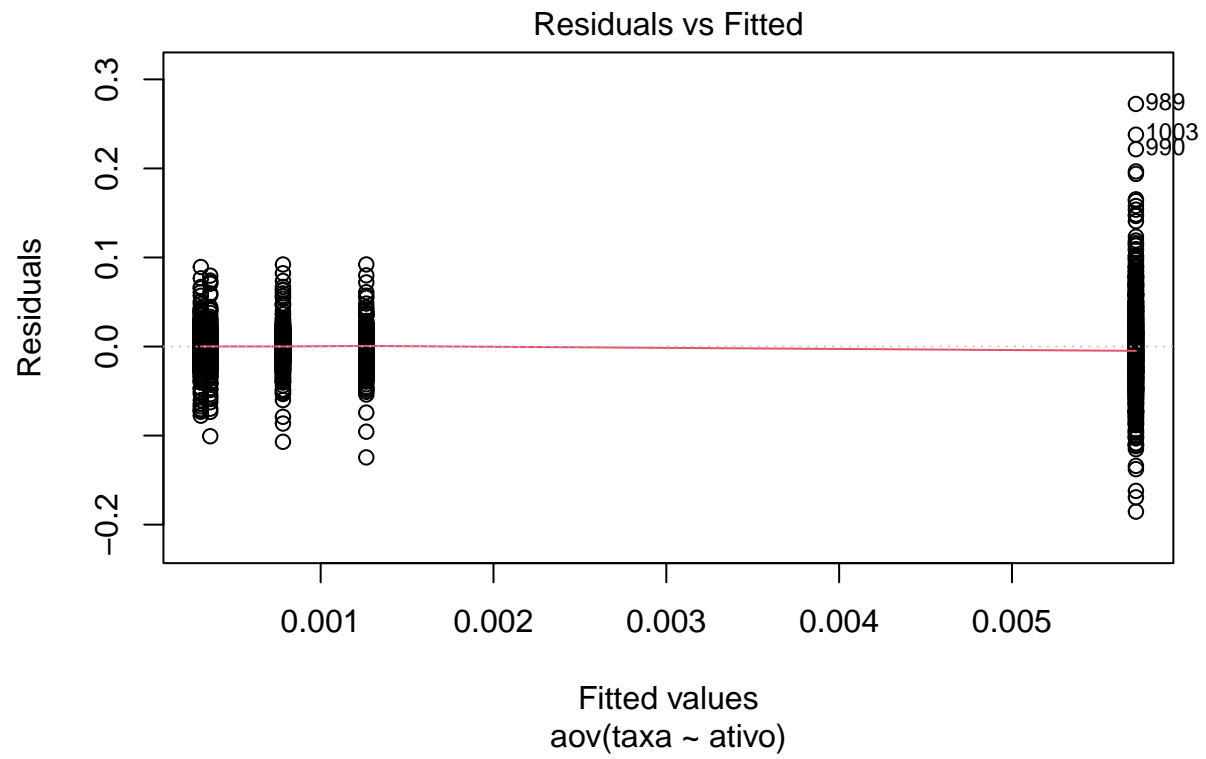
Verificou-se, anteriormente, que o pressuposto de homocedasticidade não foi atendido (por meio da realização do Teste de Levene e de Bartlett). Será realizada, portanto, a ANOVA com correção dos dados utilizando, por exemplo, a matriz de variância-covariância dos coeficientes ajustada para heterocedasticidade (Matriz de White)

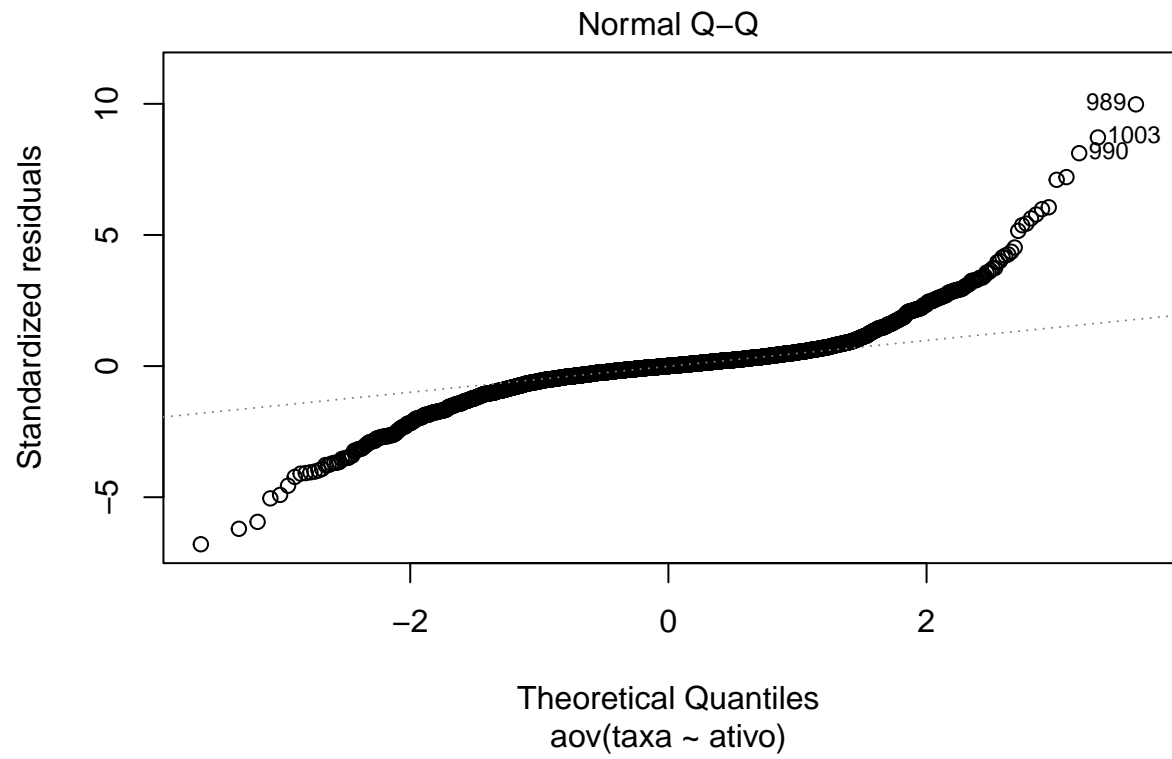
Para fins da verificação de normalidade dos dados, outras avaliações e testes serão realizados. Caso a normalidade não seja confirmada, serão realizados testes não paramétricos.

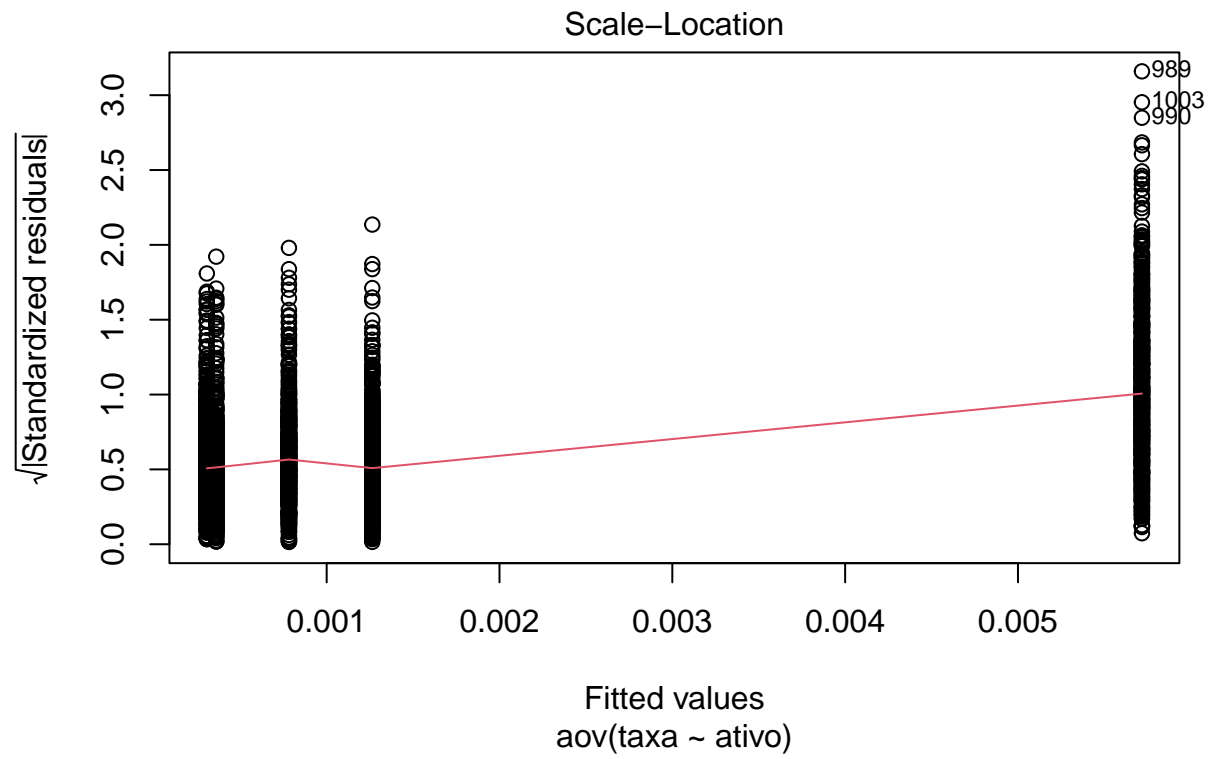
Antes da realização dos testes para verificação de normalidade, serão realizadas análises visuais para verificação tanto da homogeneidade quanto para a normalidade.

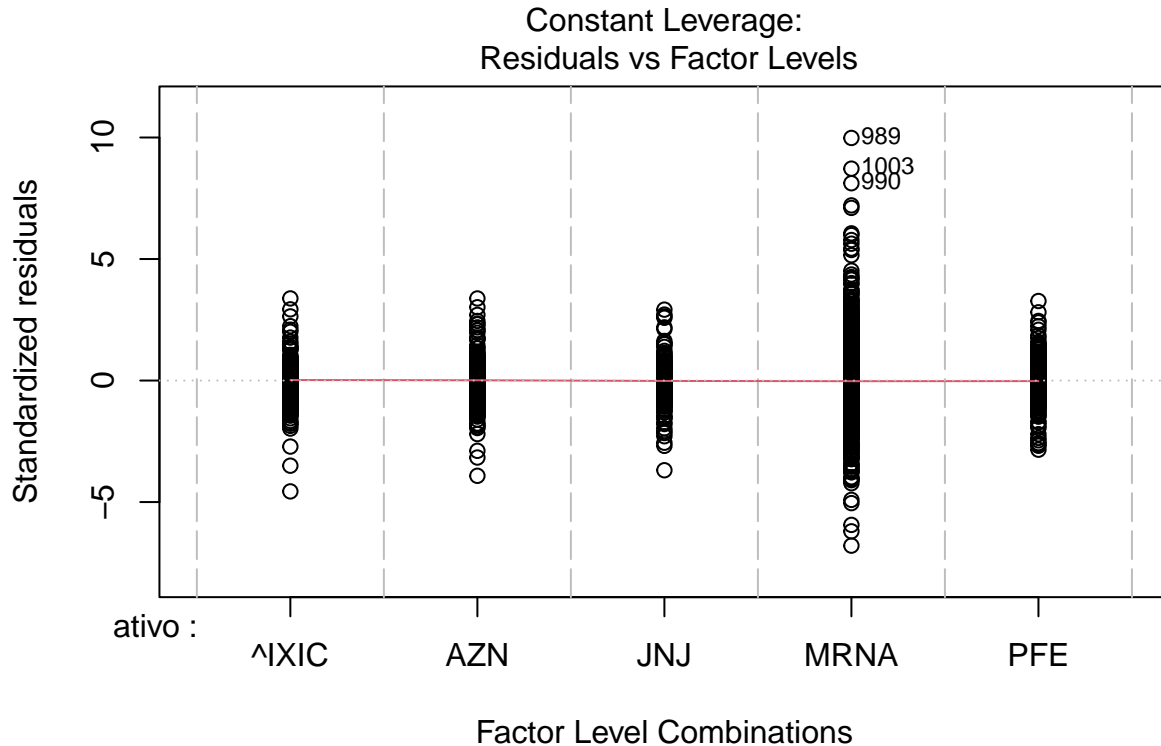
Análise visual de homogeneidade e normalidade

```
plot(anova)
```









Avaliação de homogeneidade dos resíduos

Gráfico Residual x Fitted

No gráfico Residual x Fitted espera-se que a curva de resíduos fique próximo de zero. Utiliza-se esse gráfico para verificar a suposição de que os resíduos são distribuídos aleatoriamente e têm variância constante. Idealmente, os pontos devem cair aleatoriamente em ambos os lados de 0, sem padrões reconhecíveis nos pontos.

Ele mostra a variância dos resíduos em relação ao intervalo de dados e, no caso da distribuição dos ativos, não se verifica forte tendência de alteração.

Gráfico Scale-Location

Dentre os gráficos apresentados nesta seção, este é o mais indicado para análise de homogeneidade, pois o gráfico Scale-Location mostra se os resíduos são distribuídos igualmente ao longo dos dados das variáveis de entrada (preditor). A suposição de variância igual (homocedasticidade) pode ser verificada com este gráfico, caso represente uma linha horizontal com pontos espalhados aleatoriamente.

No gráfico Scale-Location dos ativos, a linha não é horizontal, o que indica que as distribuições não são homocedásticas. Esse indício será confirmado por meio dos testes quantitativos.

Gráfico Constant-Leverage: Residuals x Factor Levels

O gráfico Residuals x Factor Levels verifica se a variância não considerada pelo modelo é diferente para diferentes níveis de um fator, auxiliando na detecção de outliers. Se tudo estiver bem, o gráfico deve exibir uma dispersão aleatória. A curvatura pronunciada pode indicar uma contribuição sistemática do fator independente que não é contabilizado pelo modelo.

No caso da distribuição dos ativos, não se verifica alteração na contribuição dos fatores.

Finaliza-se, portanto, a avaliação de homogeneidade, após a realização do teste de Levene, do teste de Bartlett e das análises visuais observando-se os gráficos Residual x Fitted, Constant-Leverage: Residuals x Factor Levels e, especialmente, o gráfico Scale-Location.

Avaliação de normalidade dos resíduos

Gráfico Normal Q-Q

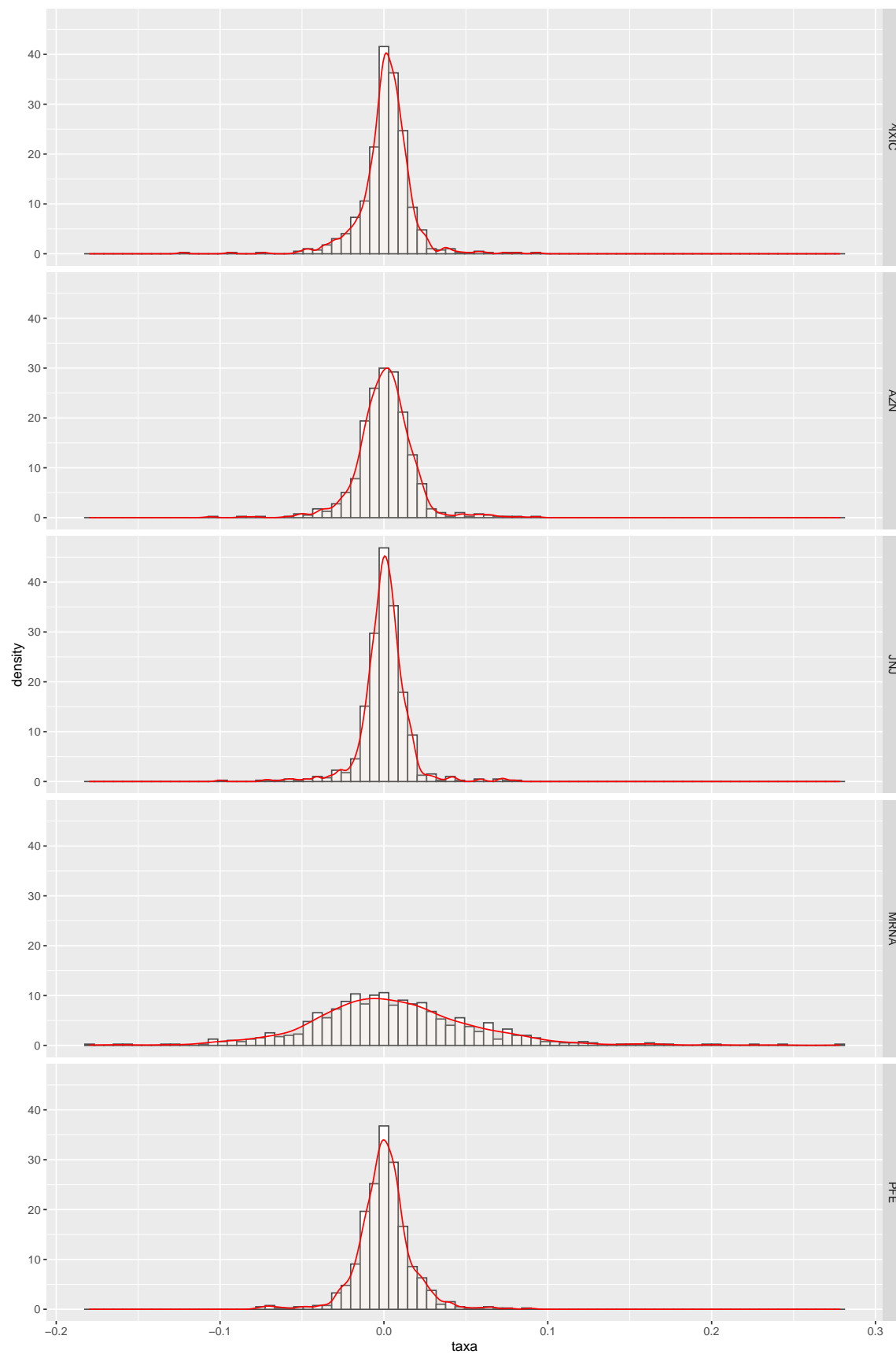
Uma das formas de se verificar a suposição de normalidade é por meio de um gráfico de probabilidade normal (gráfico Q-Q). Esse gráfico consiste em uma técnica exploratória para verificar o pressuposto de distribuição (no caso normal) para o conjunto de dados. Se os dados seguirem a distribuição normal, os pontos do gráfico formarão, aproximadamente, uma linha reta.

Considerando o gráfico Q-Q plotado acima, não se pode supor que as taxas de retorno diária dos ativos correspondem, aproximadamente, à uma linha reta. Dessa forma, a princípio, pode-se depreender de que as distribuições não se configuram como gaussianas ou normais.

Gráficos com função densidade de probabilidade

Os gráficos com a função densidade de probabilidade dos ativos também são técnicas exploratórias para verificar a distribuição dos dados. Por meio deles, é possível analisar, visualmente, se os dados seguem o formato da curva de distribuição normal.

```
ggplot(dados, aes(x = taxa)) +  
  geom_histogram(aes(y = ..density..),  
                 bins = 80, color = "grey30", fill = "white") +  
  geom_density(alpha = .2, fill = "antiquewhite3", color = "red") +  
  facet_grid(ativo ~ .)
```



Verifica-se que todas as distribuições dos ativos possuem uma cauda longa, para ambos os lados, o que não caracteriza uma distribuição normal.

Após a realização de testes visuais de normalidade (Gráfico Normal Q-Q e Gráficos com Função Densidade de Probabilidade), faz-se necessária a aplicação de testes quantitativos para verificar a suposição de normalidade dos dados.

Verificação da normalidade

Teste de Shapiro-Wilk

O teste de Shapiro-Wilk avalia uma amostra de dados e quantifica a probabilidade de eles terem sido extraídos de uma distribuição normal (gaussiana). Para o teste, a hipótese nula é que a distribuição é normal.

Vamos calcular a estatística e o valor p de cada ativo. Com base no valor p, pode-se rejeitar ou não a hipótese nula.

```
dados %>%
  group_by(ativo) %>%
  summarise(statistic = shapiro.test(taxa)$statistic,
            p.value = shapiro.test(taxa)$p.value)
```

```
## # A tibble: 5 x 3
##   ativo statistic p.value
##   <fct>      <dbl>   <dbl>
## 1 ^IXIC      0.870 2.22e-23
## 2 AZN        0.917 7.13e-19
## 3 JNJ        0.848 4.17e-25
## 4 MRNA       0.961 1.87e-12
## 5 PFE        0.927 9.40e-18
```

Após a realização do cálculo da estatística e do valor p de cada ativo usando o teste de Shapiro-Wilk, verifica-se que o valor p de cada ativo é menor que o intervalo de confiança definido (0.05). Dessa forma, rejeita-se a hipótese nula e confirma-se que nenhum dos ativos possui distribuição normal ou gaussiana.

Teste de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov também pode ser utilizado para verificar se uma amostra pode ser considerada como proveniente de uma população com distribuição normal. Como ele é particularmente indicado para uso em distribuições contínuas, adequa-se bem às distribuições dos ativos.

```
dados %>%
  group_by(ativo) %>%
  summarise(statistic = ks.test(taxa, "pnorm", mean=mean(taxa),
                                sd=sd(taxa))$statistic,
            p.value = ks.test(taxa, "pnorm", mean=mean(taxa),
                                sd=sd(taxa))$p.value)
```

```
## # A tibble: 5 x 3
##   ativo statistic      p.value
##   <fct>      <dbl>      <dbl>
```

```
## 1 ^IXIC    0.112  0.0000000599
## 2 AZN      0.0744 0.00101
## 3 JNJ      0.115  0.0000000305
## 4 MRNA     0.0614 0.0114
## 5 PFE      0.0924 0.0000167
```

Após realizar-se o teste de Kolmogorov-Smirnov, verifica-se que o valor p de cada ativo é menor o intervalo de confiança definido (0.05). Dessa forma, rejeita-se a hipótese nula e confirma-se que nenhum dos ativos possui distribuição normal ou gaussiana.

Dessa forma, finaliza-se a verificação de normalidade, após a realização de testes exploratórios (Gráfico Normal Q-Q e Gráficos com Função Densidade de Probabilidade) e de testes quantitativos (Teste de Shapiro-Wilk e Teste de Kolmogorov-Smirnov).

ANOVA com correção (Matriz de White)

Como verificou-se, anteriormente, que o pressuposto de homocedasticidade não foi atendido (por meio da realização do teste de Levene e teste de Bartlett), será realizada a ANOVA com correção dos dados utilizando a matriz de variância-covariância dos coeficientes ajustada para heterocedasticidade (Matriz de White) para verificar se as médias entre os grupos são iguais.

A matriz coloca pesos nos valores extremos na tentativa de equalizar a diferença das distribuições e fazer com que a média seja menos influenciada por esses valores extremos. Dessa forma, a heterocedasticidade dos dados é tratada para que se possa realizar a ANOVA.

```
reg_mean <- lm(taxa ~ ativo, data = dados)
anova_white <- Anova(reg_mean, white.adjust = TRUE)
anova_white
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: taxa
##           Df      F Pr(>F)
## ativo      4 1.9961 0.09242 .
## Residuals 3420
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Enquanto o valor p da ANOVA sem a correção de White foi de 0.000734, após a correção foi de 0.09242. Nesse caso, o valor p é superior ao intervalo de confiança definido (0.05) e, assim, não se rejeita a hipótese nula de que as médias entre as distribuições são iguais.

Síntese:

- Há heterocedasticidade entre os grupos.
- As médias entre os grupos são iguais.

Teste de Tukey

O teste de comparação múltipla de Tukey é um dos vários testes que podem ser usados para determinar quais médias entre um conjunto de médias diferem do restante.

Dessa forma, compara-se os valores da distribuição de cada ativo entre os demais, considerando o intervalo de confiança atribuído de 0.05, para verificar a diferença de média entre eles.

```
TukeyHSD(anova)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = taxa ~ ativo, data = dados)
##
## $ativo
##              diff              lwr              upr              p adj
## AZN-^IXIC -4.822488e-04 -0.0045098810 0.003545383 0.9975402
## JNJ-^IXIC -9.028848e-04 -0.0049305170 0.003124747 0.9732683
## MRNA-^IXIC 4.453295e-03 0.0004256630 0.008480927 0.0215948
## PFE-^IXIC -9.573009e-04 -0.0049849331 0.003070331 0.9669223
## JNJ-AZN -4.206360e-04 -0.0044482681 0.003606996 0.9985592
## MRNA-AZN 4.935544e-03 0.0009079119 0.008963176 0.0074375
## PFE-AZN -4.750521e-04 -0.0045026843 0.003552580 0.9976804
## MRNA-JNJ 5.356180e-03 0.0013285479 0.009383812 0.0026670
## PFE-JNJ -5.441612e-05 -0.0040820483 0.003973216 0.9999996
## PFE-MRNA -5.410596e-03 -0.0094382283 -0.001382964 0.0023208
```

Verifica-se que a variação da média do ativo MRNA comparada aos demais ativos indica que esse ativo difere das médias dos demais. Isso pode ser verificado pelo valor p na comparação dos ativos com o MRNA, com resultados menores do que o intervalo de confiança (0.05):

MRNA-^IXIC: valor p = 0.0215948

MRNA-AZN: valor p = 0.0074375

MRNA-JNJ: valor p = 0.0026670

PFE-MRNA: valor p = 0.0023208

Os valores p, quando a comparação é feita com o ativo MRNA, são menores do que 0.05, indicando que, nesses casos, a média entre os ativos é diferente.

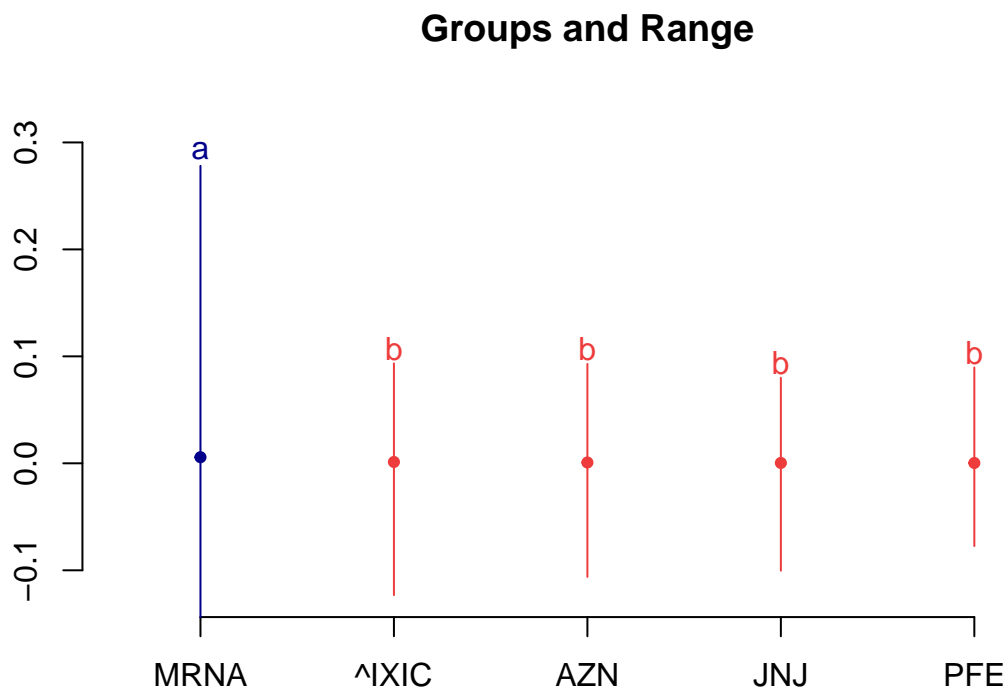
Por meio desse teste, descobre-se, portanto, qual distribuição tem média divergente das demais, sabendo-se que as distribuições são heterocedásticas.

```
tuk_ativo <- HSD.test(
  anova, "ativo", group = T, alpha = 0.05, console = T
)
```

```
##
## Study: anova ~ "ativo"
##
## HSD Test for taxa
##
## Mean Square Error: 0.0007458982
##
## ativo, means
##
##              taxa              std              r              Min              Max
## ^IXIC 0.0012640406 0.01618547 685 -0.12321331 0.09345998
## AZN 0.0007817918 0.01763517 685 -0.10619674 0.09294690
```

```
## JNJ    0.0003611558 0.01469350 685 -0.10037877 0.07997719
## MRNA   0.0057173358 0.05151227 685 -0.17966902 0.27810651
## PFE    0.0003067397 0.01694431 685 -0.07734643 0.08960696
##
## Alpha: 0.05 ; DF Error: 3420
## Critical Value of Studentized Range: 3.859716
##
## Minimun Significant Difference: 0.004027632
##
## Treatments with the same letter are not significantly different.
##
##          taxa groups
## MRNA  0.0057173358      a
## ^IXIC 0.0012640406      b
## AZN    0.0007817918      b
## JNJ    0.0003611558      b
## PFE    0.0003067397      b
```

```
plot(tuk_ativo)
```



Com base no Teste de Tukey, pode-se, também, plotar um gráfico mostrando a que grupos cada ativo pertence. Tem-se, portanto, conforme a análise quantitativa mostrou, que o ativo MRNA faz parte de um grupo diverso dos demais ativos em análise.

Porém, ressalte-se que o Teste de Tukey tem como pressuposto a existência de normalidade e homocedasticidade das distribuições. Por isso, serão realizados outros testes, não paramétricos, para as características da distribuição em análise.

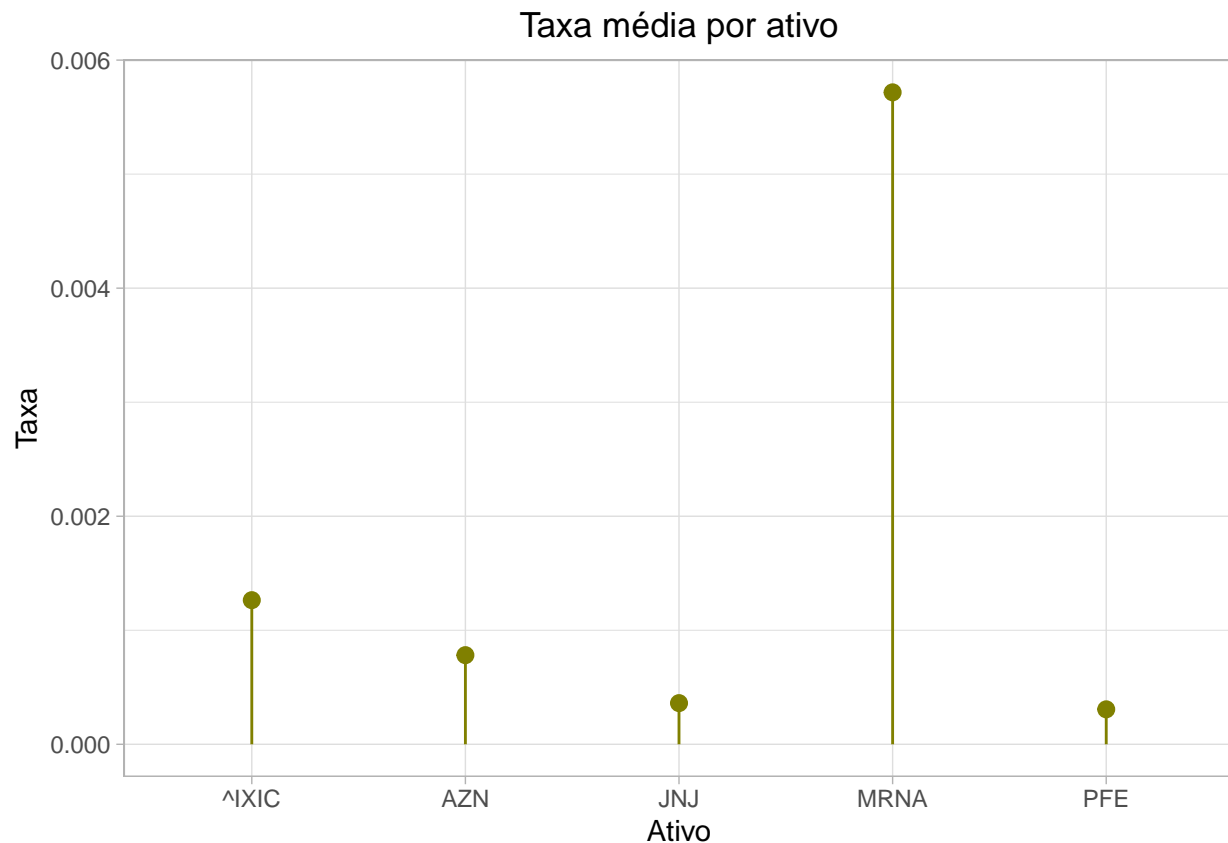
Taxa média por ativo

Antes da realização dos testes não paramétricos, pode-se verificar a diferença entre a taxa média dos ativos.

```
# Média do índice por ativo
medias <- dados %>%
  group_by(ativo) %>%
  summarise(taxa_media = mean(taxa), .groups = 'drop')
medias_nulo <- medias %>%
  mutate(taxa_media = 0)
medias_n <- rbind(medias, medias_nulo)

p1 <- medias %>%
  ggplot(aes(x = ativo, y = taxa_media)) +
  geom_point(color = '#808000', size = 2.5) +
  geom_line(data = medias_n, aes(x = ativo, y = taxa_media), color = '#808000') +
  theme_light() +
  labs(
    x = 'Ativo', y = 'Taxa',
    title = 'Taxa média por ativo'
  ) +
  theme(
    plot.title = element_text(hjust = .5)
  )

print(p1)
```



Todas médias são muito próximas de zero, mas verifica-se que a média do ativo MRNA é superior à média dos demais. Porém, os testes não paramétricos podem utilizar uma referência diferente da média para a comparação entre as distribuições, como a mediana ou o posto (ranque), conforme testes a seguir.

Testes não paramétricos

Os testes não paramétricos são métodos de distribuição livre, que não dependem de suposições extraídas dos dados fornecidos por uma distribuição normal de probabilidade.

Teste não paramétrico de Wilcoxon-Mann-Whitney

Por exemplo, ao invés de aplicar o Teste de Tukey, apropriado para situações em que os dados são provenientes de uma distribuição normal e em que há homocedasticidade, recomenda-se a aplicação de um teste não paramétrico, como o de Wilcoxon-Mann-Whitney, quando não se verifica nem a homocedasticidade nem a normalidade dos dados.

O Teste de Wilcoxon-Mann-Whitney é aplicado em situações em que se tem um ou mais pares de amostras independentes e se quer verificar se as populações que deram origem a essas amostras podem ser consideradas semelhantes ou não. Ele não é baseado nas médias, mas sim nos postos (ranques) dos valores obtidos combinando-se as duas amostras. Isso é feito ordenando-se esses valores, do menor para o maior, independentemente do fato de qual população cada valor provém. Assim, utilizando-se os postos, verifica-se se os grupos são os mesmos ou não.

```
dados <- dados %>%
  mutate(ativo = as.character(ativo))

resumo_mw <- dados %>%
  group_by(ativo) %>%
  summarise(
    N = n(),
    taxa_media = mean(taxa, na.rm = TRUE),
    std = sd(taxa, na.rm = TRUE),
    .groups = 'drop'
  ) %>%
  arrange(desc(taxa_media)) %>%
  as.data.frame()
print(resumo_mw)
```

```
##   ativo   N taxa_media      std
## 1  MRNA 685 0.0057173358 0.05151227
## 2  ^IXIC 685 0.0012640406 0.01618547
## 3   AZN 685 0.0007817918 0.01763517
## 4   JNJ 685 0.0003611558 0.01469350
## 5   PFE 685 0.0003067397 0.01694431
```

```
# Gerando pares de ativos para teste
ativos <- unique(dados$ativo)
ativos <- combn(
  x = ativos, m = 2, simplify = FALSE
)
```



```

names(ativos) <- sapply(
  ativos, function(x){paste(x, collapse = '-')})
)

# Testando pares de ativos
testes_wilcox <- lapply(
  X = ativos, FUN = wilcox_teste, f. = taxa ~ ativo,
  groups = 'ativo', data = dados
)

resumo_wilcox <- data.frame(
  comparacao = names(testes_wilcox),
  statistic_W = sapply(testes_wilcox, `[`, i = 'statistic'),
  p_valor = sapply(testes_wilcox, `[`, i = 'p.value'),
  row.names = NULL
)

print(list(resumo_wilcox))

```

```

## [[1]]
##      comparacao statistic_W    p_valor
## 1      PFE-MRNA    242579.5 0.27657295
## 2      PFE-JNJ    239485.0 0.50578889
## 3      PFE-AZN    240682.0 0.40716490
## 4  PFE-^IXIC    252668.0 0.01366622
## 5      MRNA-JNJ    227625.0 0.33994609
## 6      MRNA-AZN    228355.0 0.39278942
## 7  MRNA-^IXIC    231430.0 0.66385877
## 8      JNJ-AZN    237548.0 0.68852543
## 9      JNJ-^IXIC    251323.0 0.02247627
## 10  AZN-^IXIC    244514.0 0.17629197

```

Com base no resultado obtido, verifica-se, primeiramente, que o valor p é inferior à 0.05 na comparação dos seguintes ativos: PFE-^IXIC e JNJ-^IXIC. Assim, seria possível rejeitar a hipótese nula apenas na comparação entre esses ativos, resultando que a média entre esses ativos é diferente.

Além disso, ao comparar-se os resultados obtidos entre o teste de Tukey e o teste de Wilcoxon-Mann-Whitney, verifica-se que foram diferentes. Enquanto no teste de Tukey obteve-se quatro comparações com valor p inferior à 0.05 (todas as que comparava-se o ativo MRNA), no teste de Wilcoxon-Mann-Whitney obteve-se apenas duas: PFE-^IXIC e JNJ-^IXIC.

Teste não paramétrico de Kruskal-Wallis

O teste de Kruskal-Wallis também é uma alternativa não paramétrica ao teste ANOVA.

Hipótese nula (H_0) para o teste: as populações tendem a apresentar valores similares da variável em questão.

Hipótese alternativa (H_a) para o teste: pelo menos duas das k populações tendem a apresentar valores da variável em questão diferentes entre si.

```

kruskal.test(taxa ~ ativo, data = dados)

```

```
##
```

```
## Kruskal-Wallis rank sum test
##
## data: taxa by ativo
## Kruskal-Wallis chi-squared = 6.2811, df = 4, p-value = 0.1791
```

Como o valor de p (0.1791) é maior que o nível de significância 0.05, não há provas suficientes para rejeitar a hipótese nula de que as medianas da população são todos iguais. Pode-se concluir, portanto, que não existem diferenças significativas entre os ativos.

Teste não paramétrico de Dunn

O teste de comparação múltipla de Dunn pode ser usado para identificar quais medianas são significativas em relação a outras. É, também, um teste não paramétrico (um teste de “distribuição livre”) que não assume que seus dados vêm de uma distribuição específica.

É necessário realizar duas ressalvas na realização desse teste.

- 1) Assume-se que o teste de Dunn pode ser compreendido como um teste para diferença de mediana apenas quando as distribuições possuem dados contínuos e quando são consideradas idênticas, exceto por uma diferença na localização (postos).
- 2) Ele é utilizado após o teste de Kruskal-Wallis, na situações em que o teste de K-W permite rejeitar H_0 . Não foi o caso para este trabalho, mas será realizado o teste para avaliar os resultados e confirmar a similaridade entre as distribuições.

A hipótese nula (H_0) para o teste é que não há diferença entre os grupos (os grupos podem ser iguais ou desiguais em tamanho).

A hipótese alternativa (H_a) para o teste é que há uma diferença entre os grupos.

```
dunnTest(taxa ~ ativo,
         data = dados,
         method = "bonferroni")
```

##	Comparison	Z	P.unadj	P.adj
## 1	^IXIC - AZN	1.3288961	0.18388225	1.0000000
## 2	^IXIC - JNJ	1.7279829	0.08399128	0.8399128
## 3	AZN - JNJ	0.3990868	0.68982921	1.0000000
## 4	^IXIC - MRNA	0.4669401	0.64054275	1.0000000
## 5	AZN - MRNA	-0.8619560	0.38871173	1.0000000
## 6	JNJ - MRNA	-1.2610429	0.20729341	1.0000000
## 7	^IXIC - PFE	2.1433671	0.03208363	0.3208363
## 8	AZN - PFE	0.8144710	0.41537516	1.0000000
## 9	JNJ - PFE	0.4153842	0.67786068	1.0000000
## 10	MRNA - PFE	1.6764271	0.09365457	0.9365457

Assumindo-se um intervalo de confiança de 0.05, verifica-se que nenhuma comparação de ativos é estatisticamente diferente entre si.