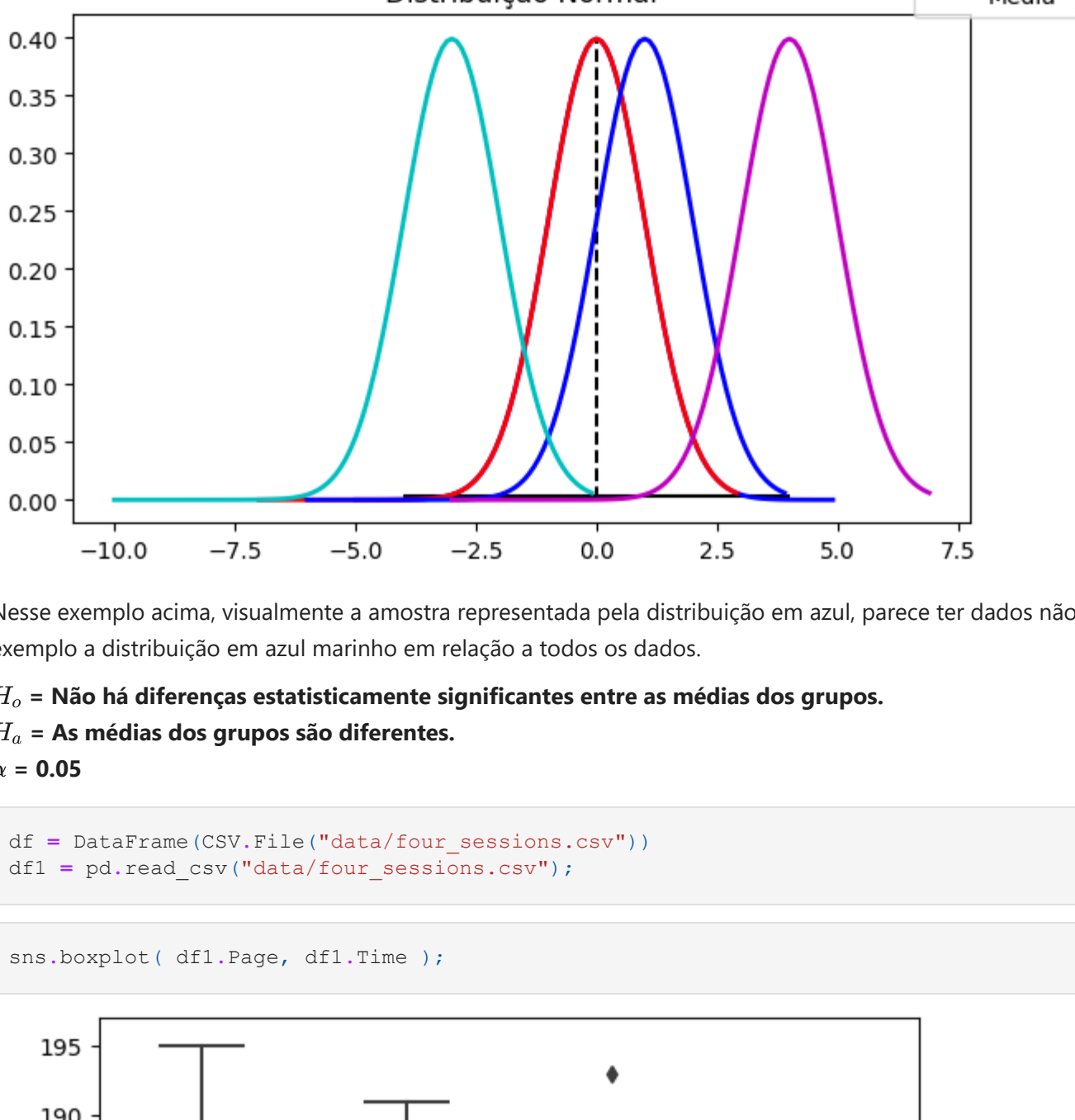


Distribuição Normal



Nesse exemplo acima, visualmente a amostra representada pela distribuição em azul, parece ter dados não tão importantes quanto por exemplo a distribuição em azul marinho em relação a todos os dados.

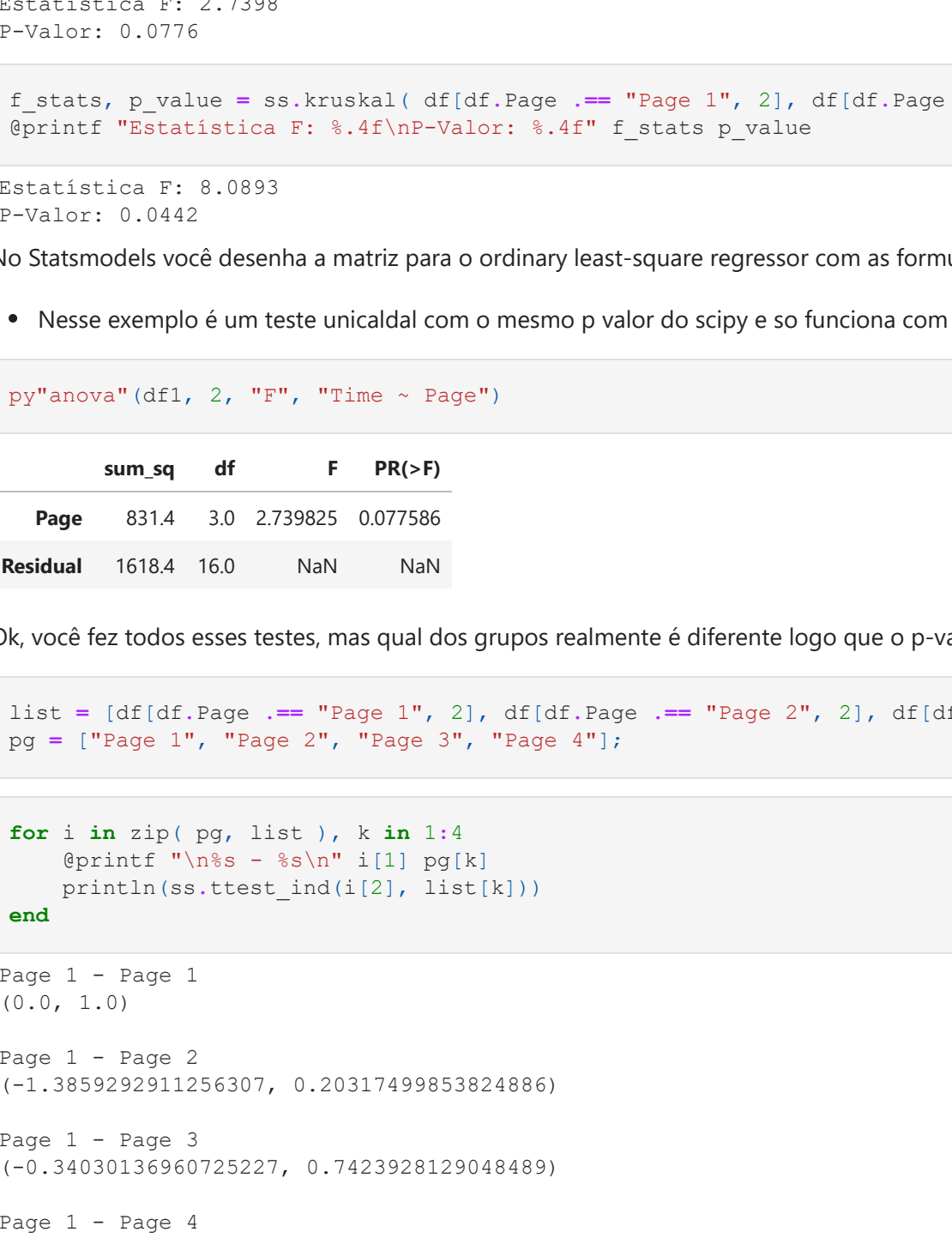
H_0 = Não há diferenças estatisticamente significantes entre as médias dos grupos.

H_a = As médias dos grupos são diferentes.

$\alpha = 0.05$

```
In [289]: df = DataFrame(CSV_File("data/four_sessions.csv"))
df1 = pd.read_csv("data/four_sessions.csv");
```

```
In [210]: sns.boxplot( df1.Page, df1.time );
```



```
In [290]: f_stats, p_value = ss.f_oneway( df[df.Page == "Page 1", 2], df[df.Page == "Page 2", 2], df[df.Page == "Page 3", 2], df[df.Page == "Page 4", 2] )
f_stats p_value
Estadística F: 1.445
p-Valor: 0.0776
```

```
In [296]: f_stats, p_value = ss.kruskal( df[df.Page == "Page 1", 2], df[df.Page == "Page 2", 2], df[df.Page == "Page 3", 2], df[df.Page == "Page 4", 2] )
f_stats p_value
Estadística F: 1.445
p-Valor: 0.0442
```

No Statsmodels você desenha a matriz para o ordinary least-square regressor com as formulas.

- Nesse exemplo é um teste unidacal com o mesmo p valor do scipy e so funciona com um dataframe do pandas.

```
In [291]: py"anova"( df1, 2, "P", "Time ~ Page")
```

```
Out[291]:
      sum_sq      df      F      PR(>F)
Page      831.4      3.0  2.739825  0.077586
Residual 1618.4     16.0      NaN      NaN
```

Ok, você fez todos esses testes, mas qual dos grupos realmente é diferente logo que o p-valor foi pequeno?

```
In [362]: list = [df[df.Page == "Page 1", 2], df[df.Page == "Page 2", 2], df[df.Page == "Page 3", 2], df[df.Page == "Page 4", 2]]
pg = ["Page 1", "Page 2", "Page 3", "Page 4"];
```

```
In [360]: for i in zip( pg, list ), k in 1:4
0:printf "\n%5s - %5s\n" i[1] pg[k]
println(as.ttest_ind(i[2], list[k]))
end
```

Page 1 ~ Page 1
(0.0, 1.0)

Page 1 ~ Page 2
(-1.3859292931256307, 0.20317499853824886)

Page 1 ~ Page 3
(-0.34030136960725227, 0.7423928129048489)

Page 1 ~ Page 4
(1.136191733208663, 0.28096106091142226)

Page 2 ~ Page 1
(1.3859292931256307, 0.20317499853824886)

Page 2 ~ Page 2
(0.0, 1.0)

Page 2 ~ Page 3
(1.2650143347089609, 0.24146881279653873)

Page 2 ~ Page 4
(4.950587917224828, 0.00112005643441126)

Page 3 ~ Page 1
(0.34030136960725227, 0.7423928129048489)

Page 3 ~ Page 2
(-1.2650143347089609, 0.24146881279653873)

Page 3 ~ Page 3
(0.0, 1.0)

Page 3 ~ Page 4
(1.9782125034804172, 0.08327876770509358)

Page 4 ~ Page 1
(-1.136191733208663, 0.28096106091142226)

Page 4 ~ Page 2
(-4.950587917224828, 0.00112005643441126)

Page 4 ~ Page 3
(-1.9782125034804172, 0.08327876770509358)

Page 4 ~ Page 4
(0.0, 1.0)

Somente em relação a página 4 com a página 2 o p-value ficou pequeno.

3.4.1. Bonferroni & Sidak Correction

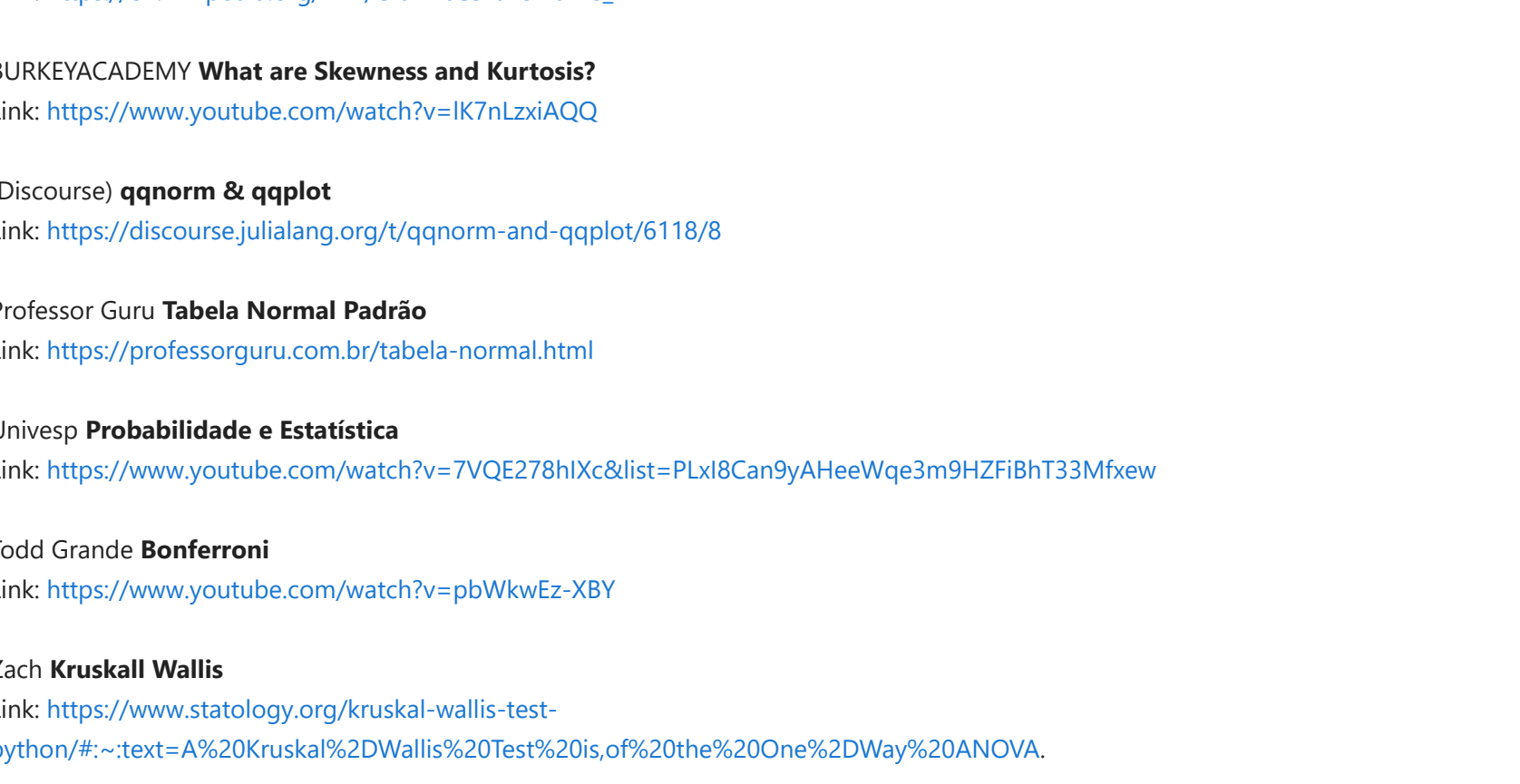
Quando se trabalha com a Anova ou múltiplos testes t, geralmente podem superestimar a significância, ou seja, maior a chance de encontrar p valores pequenos, uma forma de corrigir esse problema é com a correção de bonferroni, que nada mais é que dividi o nível de significância pelo número unicos de comparações feitas, por exemplo

Com a correção de Sidak, a formula é

Outro comumente utilizado também é o Tukey's test

```
In [427]: py"tukey"( df1.Time, df1.Page, 0.05)
```

Multiple Comparisons Between All Pairs (Tukey)



Link: <http://wiki.cmc.usp.br/images/1/13/Chisq-hsher.pdf>

x.0. Referências

PETER BRUCE & ANDREW BRUCE **Estatística prática para cientistas de dados: 50 conceitos essenciais**.

Link: <https://www.amazon.com.br/Estat%C3%AAdica-Pr%C3%A1tica-Para-Cientistas-Dados/dp/855080603X>

DAVID MATOS **8 Conceitos Estatísticos Fundamentais Para Data Science**.

Link: <https://www.cienciaedados.com/8-conceitos-estatisticos-fundamentais-para-data-science/>

IGOR SOARES **Correlação não implica em Causalidade**.

Link: <https://medium.com/@felipemaiapolo/correla%C3%A7%C3%A3o-n%C3%A3o-implica-em-causalidade-8459179ad1bc>.

annahaensch **Número de Casos de Divórcio em Maine**

Link: <https://blogs.ams.org/blogonmathblogs/2017/04/10/divorce-and-margarine/>

Wikipédia **Cramer's V**

Link: https://en.wikipedia.org/wiki/Cram%C3%A9r%27s_V

BURKEYACADEMY **What are Skewness and Kurtosis?**

Link: <https://www.youtube.com/watch?v=IK7hLzxiAQQ>

(Discourse) **qqnorm & qqplot**

Link: <https://discourse.julialang.org/t/qqnorm-and-qqplot/6118/8>

Professor Guru **Tabela Normal Padrão**

Link: <https://professorguru.com.br/tabela-normal.html>

Univesp **Probabilidade e Estatística**

Link: <https://www.youtube.com/watch?v=7VQE278hXc&list=PLxI8Can9yAHeeWqe3m9HZf8hT33Mfxew>

Todd Grande **Bonferroni**

Link: <https://www.youtube.com/watch?v=pbWkwEz-XBY>

Zach **Kruskal Wallis**

Link: <https://www.statology.org/kruskal-wallis-test-python/#~:text=A%20kruskal%2Dwallis%20test%20is,o%20the%20One%2DWay%20ANOVA.>

USP **Teste Exato de Fisher**

Link: <http://wiki.icmc.usp.br/images/7/73/Chisq-fisher.pdf>