

GRUPO i

BOOTCAMP "NEURON"

CAROLINNE
PINHEIRO

EBRAIM
CARVALHO

Dr. MICHAEL
NAZARKOVSKY

GABRIELA
PIOVESAN

IGO
BARROS

ESTRUTURA DA APRESENTAÇÃO

01

INTRODUÇÃO

03

RESULTADOS

02

MÉTODOS

04

CONCLUSÕES
& INSIGHTS

INTRODUÇÃO

- Extrair insights a partir do tratamento de uma base de dados de “Tips” (port. “Gorjeta”), utilizando os dados;
- Analisar os dados para fazer conclusões e ajudar o negócio.

Dados quantitativos

Conta total (Total_bill)

Número dos Clientes por única visita (Size_of_group)

Gorjeta (Tip) - **target**

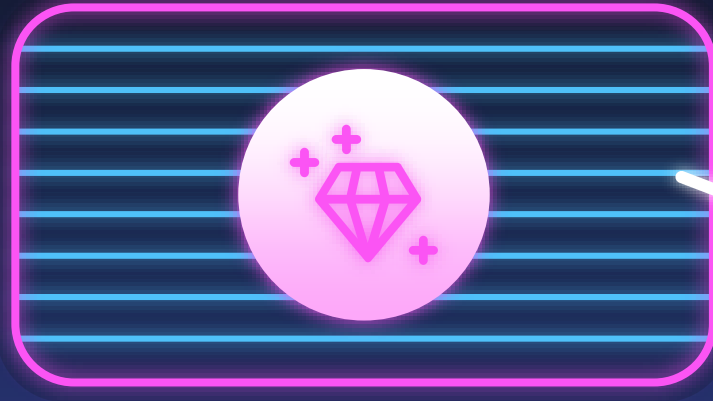
Dados categóricos

Gênero (Sex)

Fumante ou não (Smoker)

Dia da semana (Day)

Almoço/Jantar (Time)



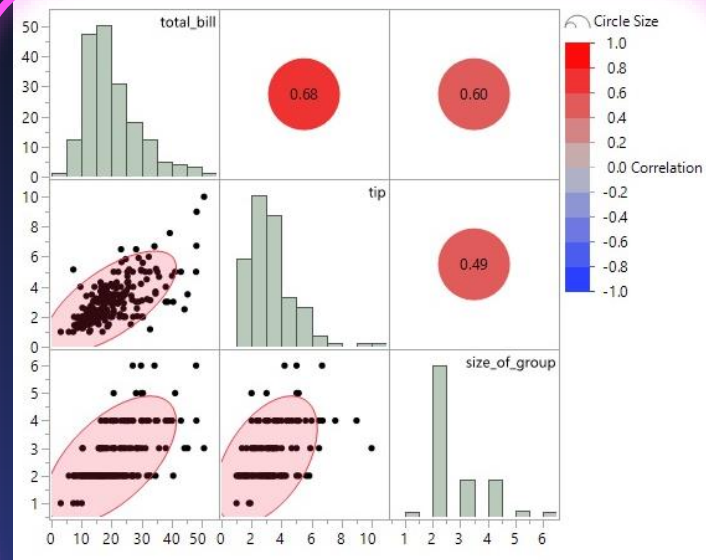
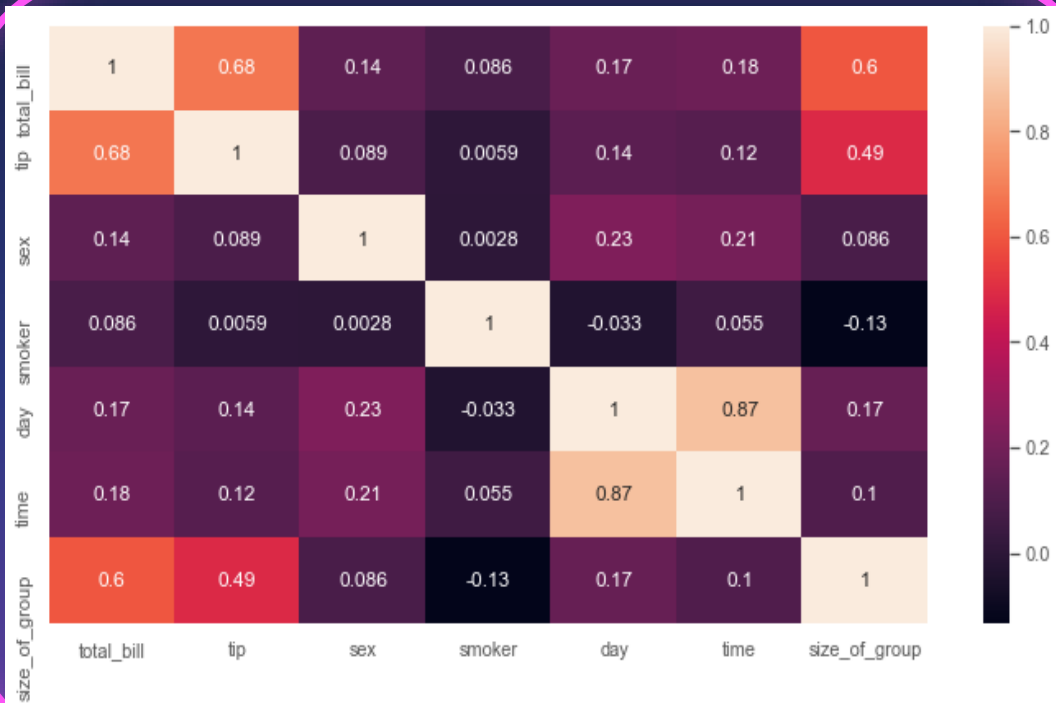
MÉTODOS

Aplicar os métodos de matemática para analisar as variâncias e correlações:

- Means Comparison (Tukey-Kramer);
- ANOVA;
- Welch's Method;
- Equality of Variances tests.

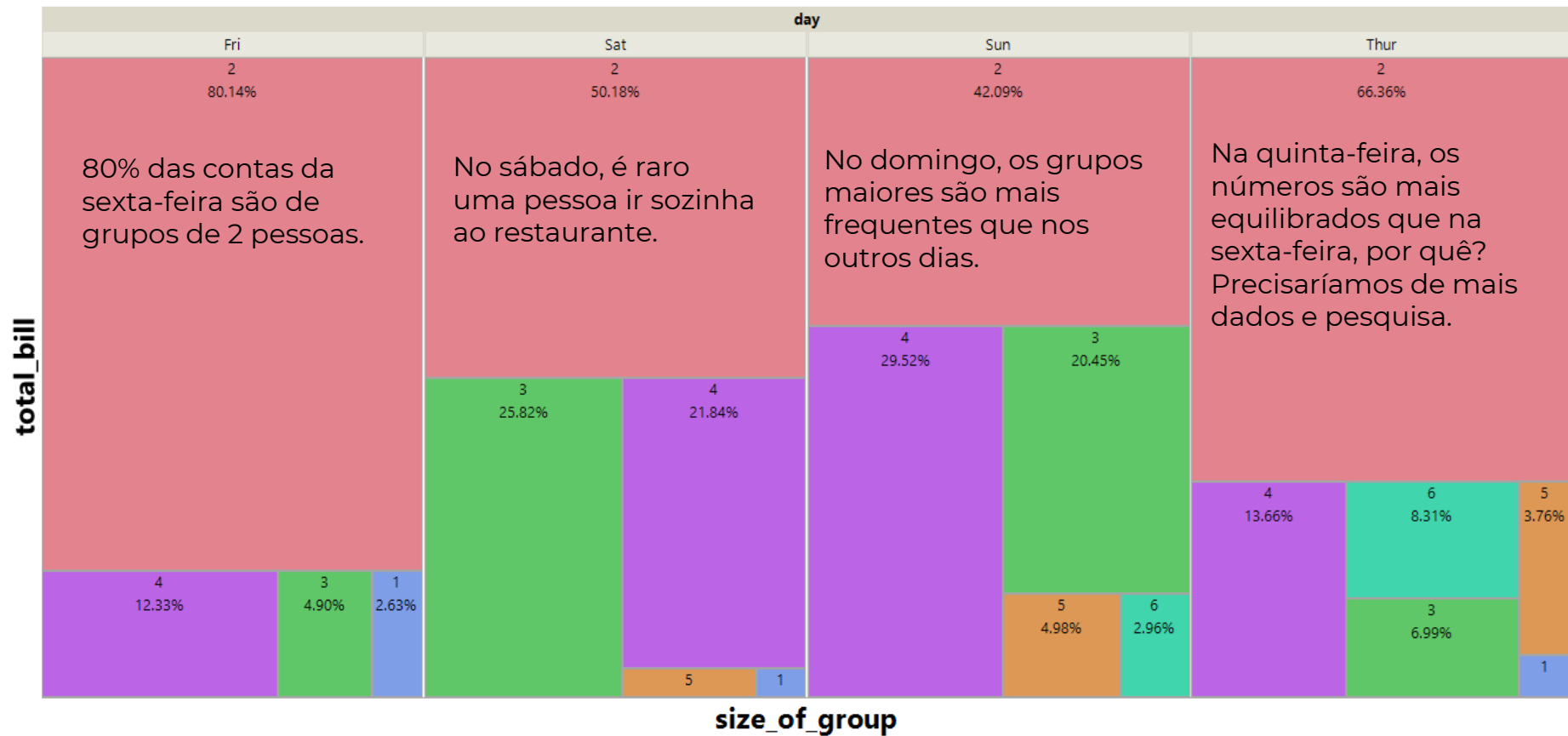
RESULTADOS

MATRIZES DE CORRELAÇÃO

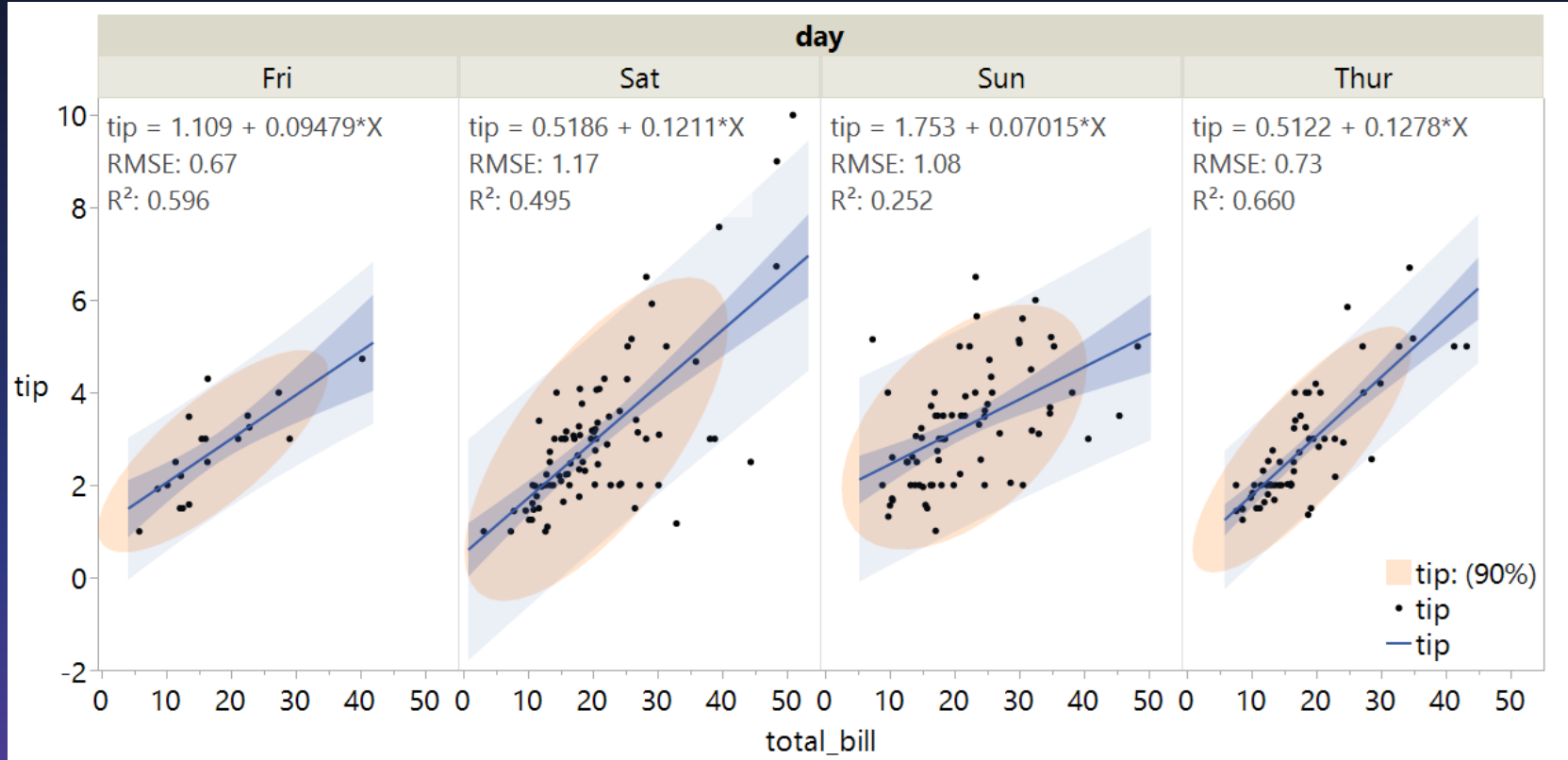


Utilizando um mapa de calor, podemos ver uma menor correlação de fumantes com o tamanho do grupo. Em contrapartida, o total da conta apresenta maior relação com a gorjeta oferecida.

% of Total_bill vs. Size_of_group por Days



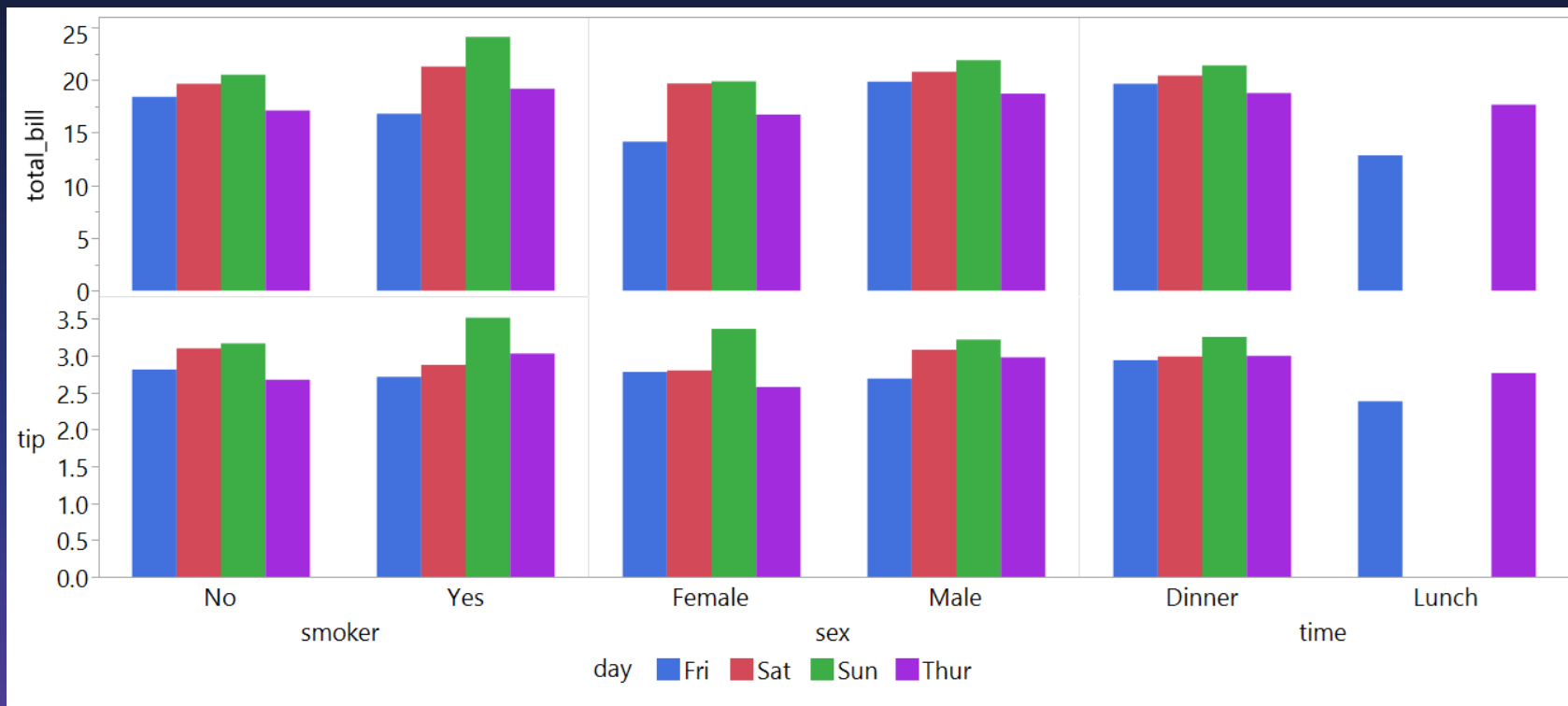
Tips vs. Total_bill



Em Sat e Sun as correlações são piores. Mas em todos dias existe correlação

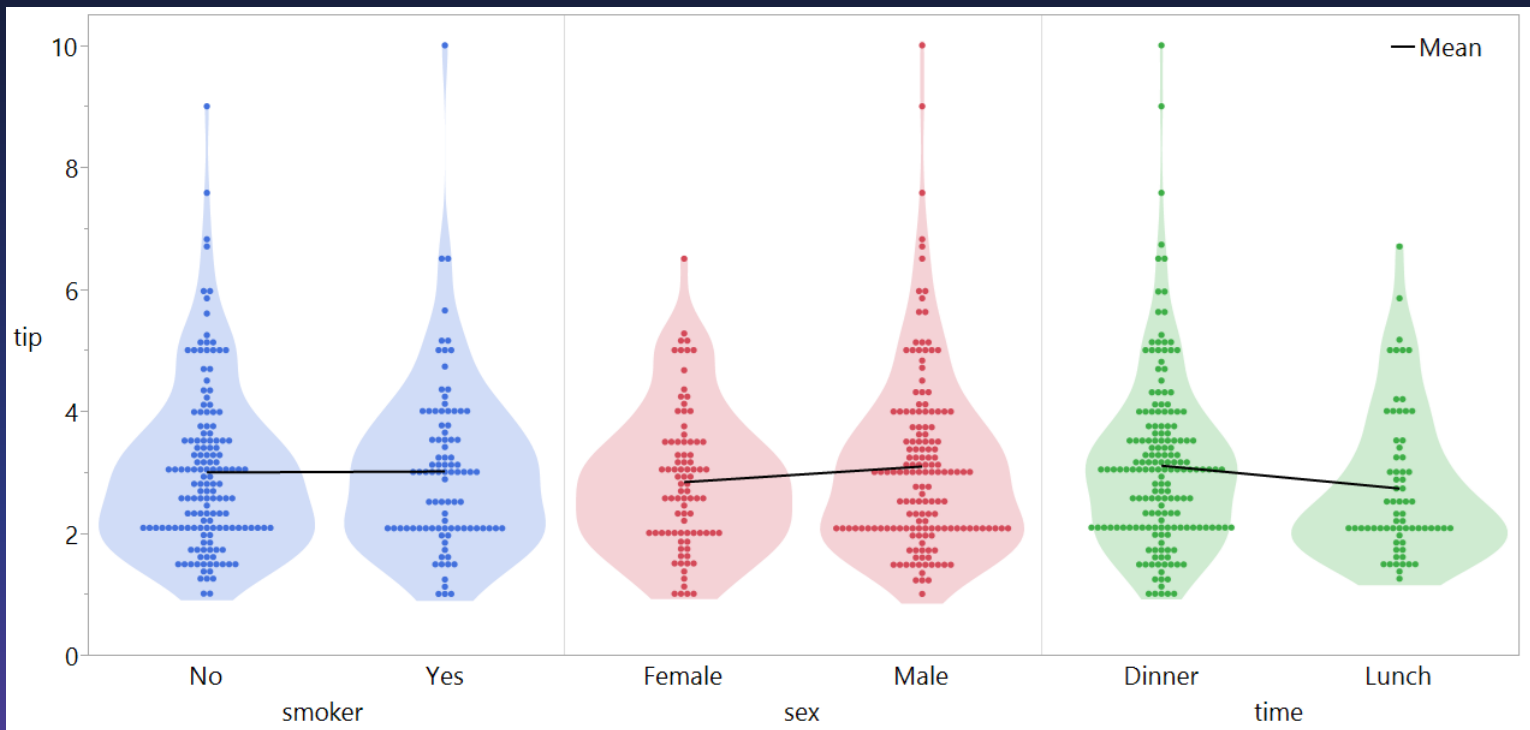
Tips and Total_Bill vs. "smoker", "sex", "time"

stratificados por Days



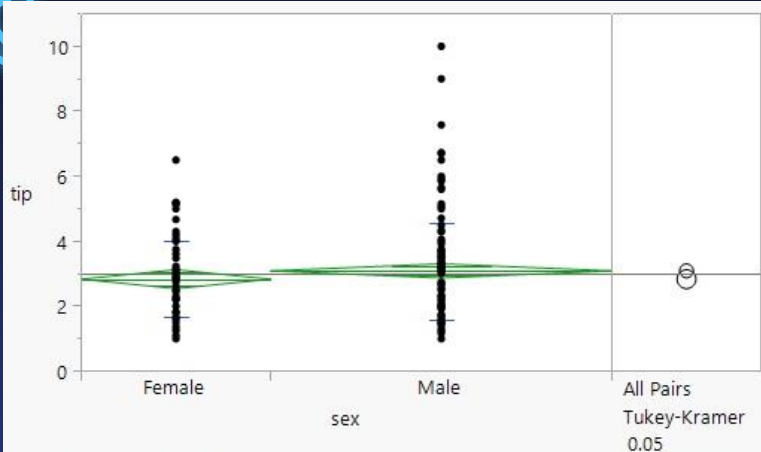
Tem problemas em Fri & Thur!

Tips vs. Smoker, Sex e Time - análise visual



Será que esses fatores podem influenciar ao Tips? Vamos comparar as variâncias!

Tip vs Sex – não significativo!



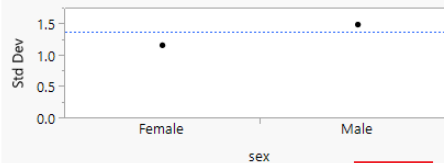
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
sex	1	3.67353	3.67353	1.9262	0.1665
Error	242	461.53894	1.90719		
C. Total	243	465.21248			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Female	87	2.83345	0.14806	2.5418	3.1251
Male	157	3.08962	0.11022	2.8725	3.3067

Tests that the Variances are Equal



Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[5]	2.1893	1	242	0.1403
Brown-Forsythe	1.9910	1	242	0.1595
Levene	2.3572	1	242	0.1260

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
2.2187	1	215.71	0.1378

t Test

1.4895

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD
Confidence Quantile

q*	Alpha
1.96982	0.05

HSD Threshold Matrix

Abs(Dif)-HSD

	Male	Female
Male	-0.30703	-0.10742
Female	-0.10742	-0.41246

Positive values show pairs of means that are significantly different.

Connecting Letters Report

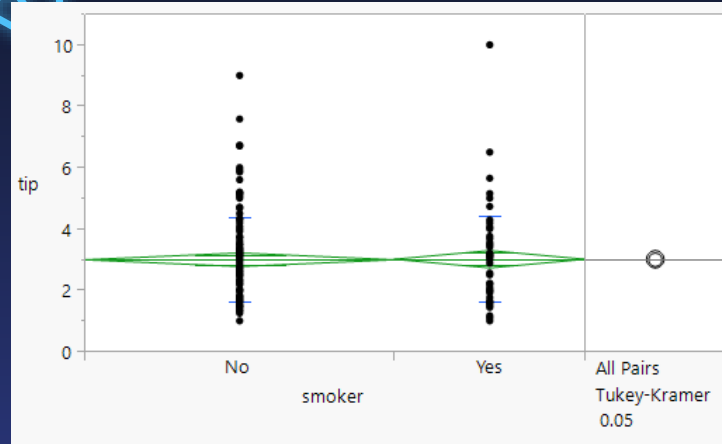
Level	Mean
Male A	3.0896178
Female A	2.8334483

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Male	Female	0.2561696	0.1845789	-0.107417	0.6197559	0.1665

Tips vs Smoker – não significativo!



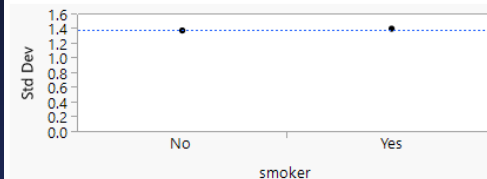
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
smoker	1	0.01635	0.01635	0.0085	0.9266
Error	242	465.19613	1.92230		
C. Total	243	465.21248			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
No	151	2.99185	0.11283	2.7696	3.2141
Yes	93	3.00871	0.14377	2.7255	3.2919

Tests that the Variances are Equal



Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[5]	0.0125	1	242	0.9109
Brown-Forsythe	0.0205	1	242	0.8862
Levene	0.0522	1	242	0.8195

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
0.0084	1	192.26	0.9269

t Test
0.0918

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD Confidence Quantile

q*	Alpha
1.96982	0.05

HSD Threshold Matrix

Abs(Dif) - HSD

	Yes	No
Yes	-0.40051	-0.34314
No	-0.34314	-0.31431

Positive values show pairs of means that are significantly different.

Connecting Letters Report

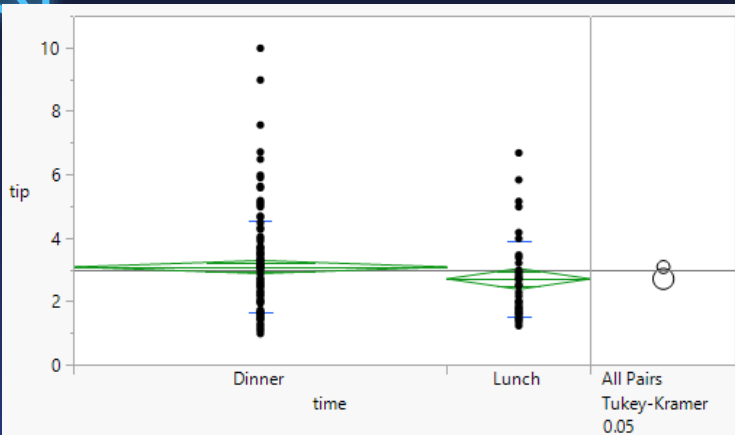
Level	Mean
Yes A	3.0087097
No A	2.9918543

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Yes	No	0.0168554	0.1827575	-0.343143	0.3768541	0.9266

Tips vs. Time - não tão simples!



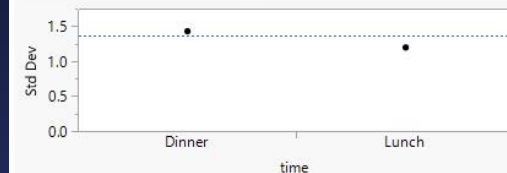
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
time	1	6.88218	6.88218	3.6338	0.0578
Error	242	458.33030	1.89393		
C. Total	243	465.21248			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Dinner	176	3.10267	0.10373	2.8983	3.3070
Lunch	68	2.72809	0.16689	2.3993	3.0568

Tests that the Variances are Equal



Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.S]	0.9218	1	242	0.3380
Brown-Forsythe	0.8810	1	242	0.3489
Levene	0.4469	1	242	0.5044

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
4.2408	1	144.07	0.0413*

t Test
2.0593

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD
Confidence Quantile

q*	Alpha
1.96982	0.05

HSD Threshold Matrix

Abs(Dif)-HSD

	Dinner	Lunch
Dinner	-0.28898	-0.01249
Lunch	-0.01249	-0.46491

Positive values show pairs of means that are significantly different.

Connecting Letters Report

Level	Mean
Dinner	A 3.1026705
Lunch	A 2.7280882

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Dinner	Lunch	0.3745822	0.1965014	-0.012489	0.7616538	0.0578

Welch's Test, baixos p -valores em ANOVA e Tukey-Kramer testes têm que ser anotados

A QUESTÃO DE DAYS (DIAS DA SEMANA)

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
2.58714	0.05

HSD Threshold Matrix

Abs(Dif)-HSD

	Sun	Sat	Thur	Fri
Sun	-0.5783	-0.2977	-0.1264	-0.3940
Sat	-0.2977	-0.5405	-0.3708	-0.6444
Thur	-0.1264	-0.3708	-0.6403	-0.8981
Fri	-0.3940	-0.6444	-0.8981	-1.1566

Positive values show pairs of means that are significantly different.

Connecting Letters Report

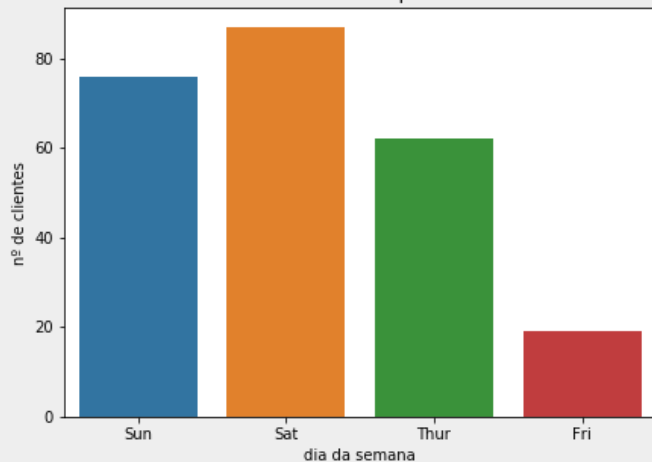
Level	Mean
Sun	A 3.2551316
Sat	A 2.9931034
Thur	A 2.7714516
Fri	A 2.7347368

Levels not connected by same letter are significantly different.

Ordered Differences Report

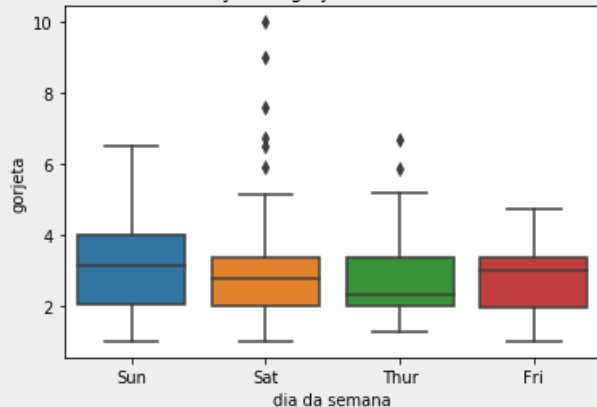
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Sun	Fri	0.5203947	0.3534319	-0.393984	1.434773	0.4558
Sun	Thur	0.4836800	0.2358112	-0.126397	1.093757	0.1724
Sun	Sat	0.2620281	0.2163490	-0.297698	0.821754	0.6204
Sat	Fri	0.2583666	0.3489346	-0.644377	1.161110	0.8806
Sat	Thur	0.2216518	0.2290157	-0.370844	0.814148	0.7679
Thur	Fri	0.0367148	0.3613245	-0.898083	0.971513	0.9996

Dias da semana x qtd clientes



O melhor dia da semana para o restaurante fazer uma promoção, seria na sexta, uma vez que é o dia em que há menos clientes.

Distribuição de gorjeta x dia de semana



Nota-se que os dias da semana não influenciam o valor das gorjetas.



CONCLUSÕES & INSIGHTS

1. Não há diferenças **estatisticamente** significativas em Tip entre Sex, Smoker, Time e Days. Porém há algumas observações a relação de Time e Days;
2. Tem correlação positiva entre Tips, Total_bill e Size_of_the_Group;
3. A maior contribuição ao Tips e Total_bill é feita pelo grupo das 2 pessoas nas horas do jantar;
4. A maior contribuição ao Tips e Total_bill nos dias observados é feita pelo grupo das 2 pessoas com menores percentagens de contribuição dele em Sat e Sun;
5. Já que Tips (gorjeta) e Size_of_group (quantidade dos clientes no grupo) são os menores em Fri (sexta-feira), faz sentido fazer as promoções nesse dia para os grupos a partir de 5 pessoas - colegas, amigos, happy hour, aniversários.
6. Especialmente promoções para o almoço na sexta-feira, onde a contribuição para o total das contas e das gorjetas são as menores, focando na atração de grupos de 5 pessoas (grupos de trabalho, amigos, faculdade...), assim o restaurante poderá melhorar os resultados.



**AGRADECEMOS
A ATENÇÃO!**