Statistical Analysis of Coffee Data

Presented by Shuting Lin

2022.1.4

Outline

1. Data preprocessing

2. Data Visualization

3. What factors affect daily average sales quantity?

4. ANOVA - Time Aspect

Does interaction exist between categories and time intervals?
Does interaction exist between channels and time intervals?

5. ANOVA - Space Aspect

Does interaction exist between categories and Taipei districts?
Does interaction exist between channels and Taipei districts?

6. Conclusion

1. Data preprocessing

Data Preprocessing

Original data size: 4,706,364

- 369,706 duplicates
- 106,965 entries not in {Latte, Americano}
- 100,119 entries whose unit_price > 300

(based on recalculated unit_price)

= Final dataset size: 4,129,574 (87.74 %)

o_idx	id	name	quant	uniprice	totprice	invo_price	county_district
25727	1	大冰拿鐵N	88	55	4840	312	彰化縣伸港鄉
40697	1	代銷一中經典美式	1	3500	3500	10	彰化縣大城鄉
50599	1	隨時取–拿鐵熱咖啡(大)買50送15	1	2750	2750	2576	彰化縣鹿港鎮
87744	1	美式熱咖啡(大)	1	10000	10000	10000	彰化縣秀水鄉
90329	1	濃萃拿鐵冰咖啡(大)	45	60	2700	1980	彰化縣永靖鄉

(a)

(b)

65 ↘

↘ 42.3

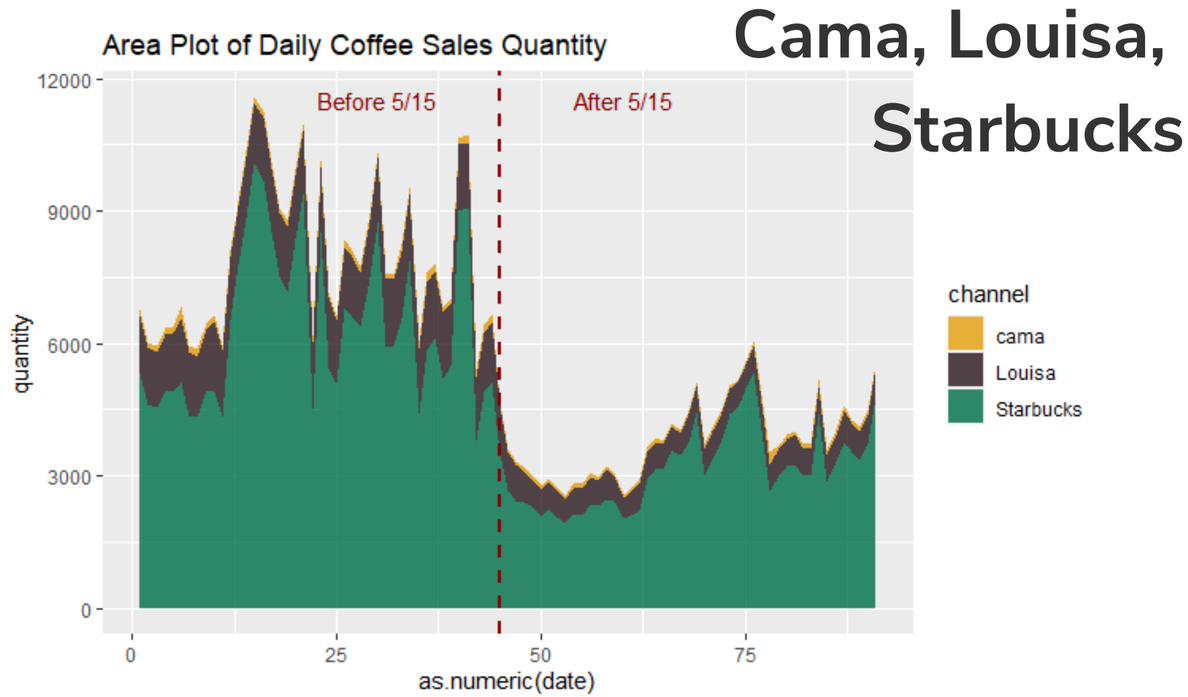
2. Data Visualization

Coffee Sales from April to June

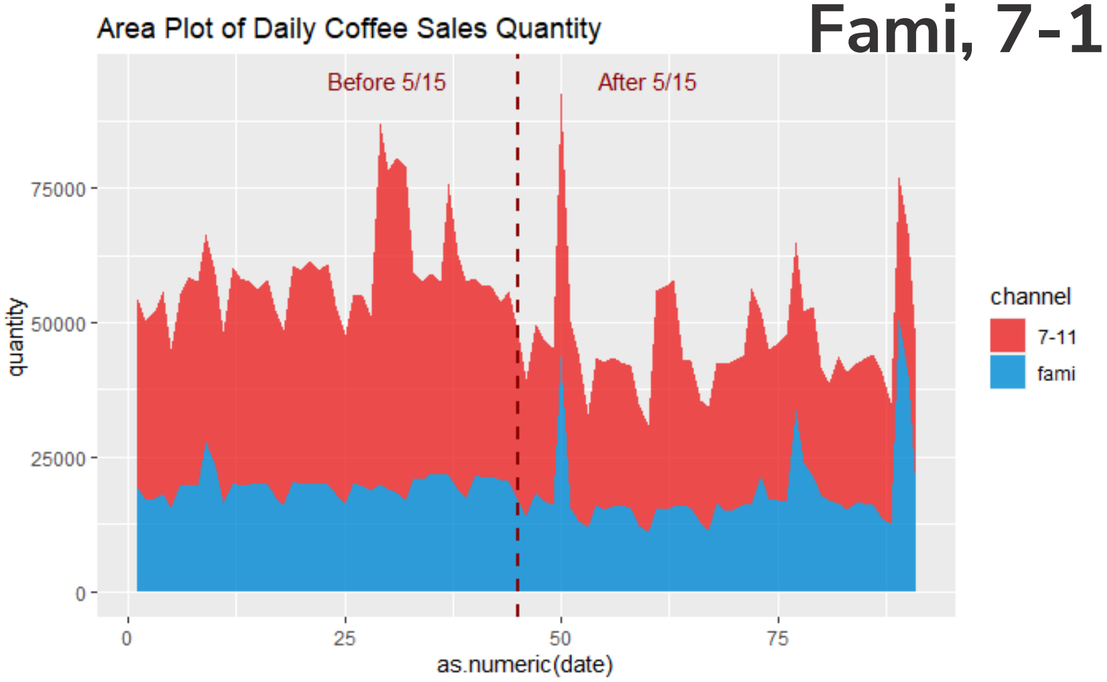
A look into 5 Channels

Cama, Louisa, Starbucks

7-11, FamiliyMart



Cama, Louisa, Starbucks



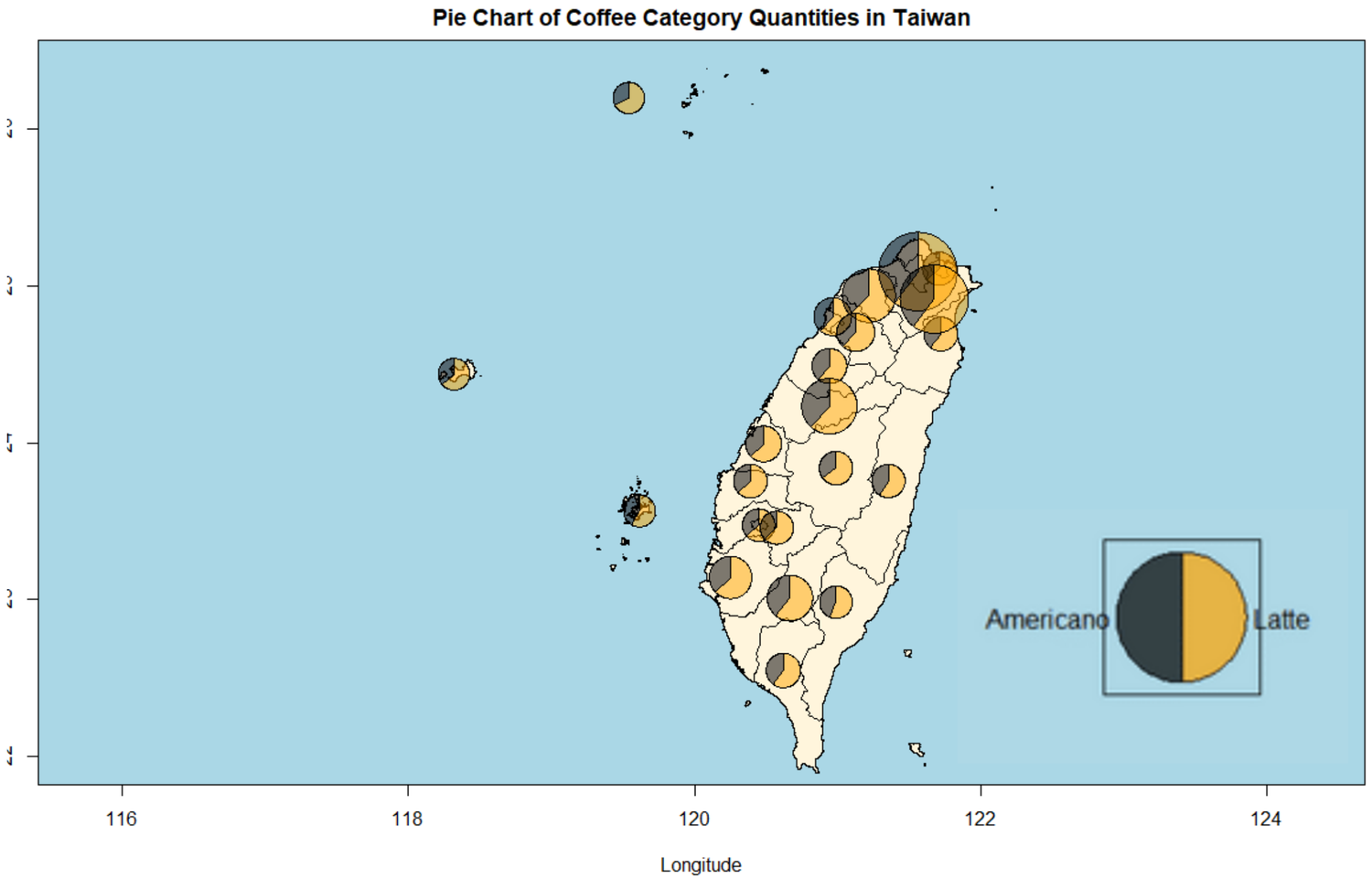
Fami, 7-11

Coffee Sales from April to June

A look into 2 Categories

Americano

Latte



3. What factors affect daily average sales quantity?

What factors affect coffee sales quantity?

category

Americano or Latte

channel

7-11, FamilyMart, Cama, Louisa
and Starbucks

county

22 counties in Taiwan

is_Alert

Indicator variable,
before or after 5/15. Before: 0, After: 1.

weekday

From Monday to Sunday

unit_price

Average Unit Price,
considering drink size and promotion

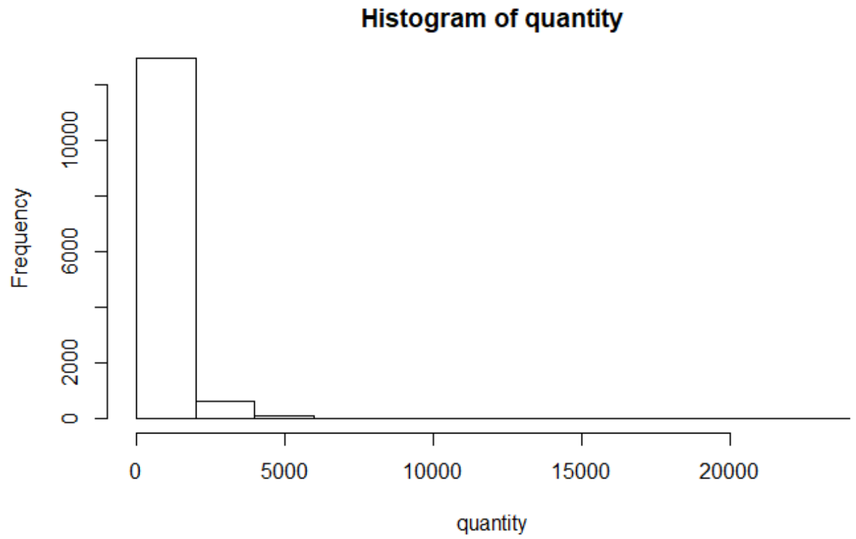
interval

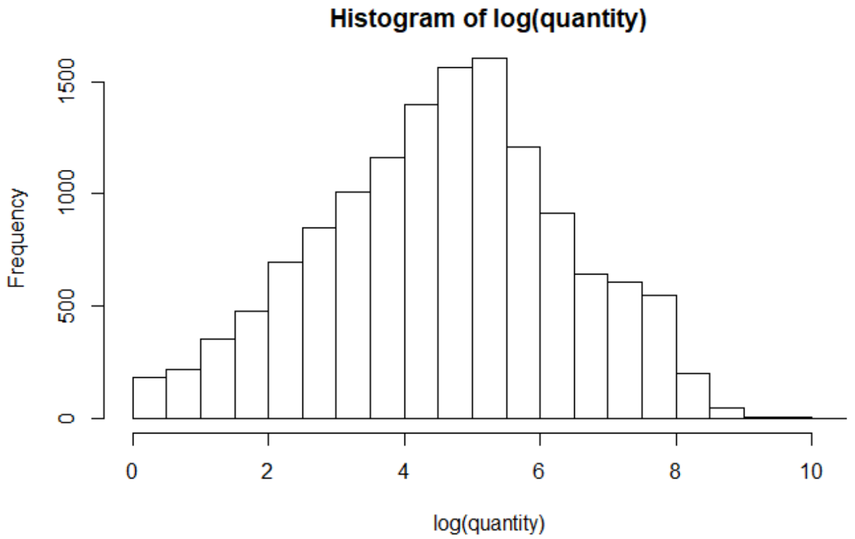
6-11 Morning
11-16 Noon
16-21 Afternoon

What factors affect daily average sales quantity?

Dataset and EDA

date <fctr>	weekday <fctr>	county <fctr>	channel <fctr>	category <fctr>	is_Alert <int>	quantity <int>	total_price <int>	unit_price <dbl>
2021/4/1	Thursday	Changhua	7-11	Americano	0	290	12802	44.14483
2021/4/1	Thursday	Changhua	7-11	latte	0	522	28920	55.40230
2021/4/1	Thursday	Changhua	Starbucks	Americano	0	21	2415	115.00000
2021/4/1	Thursday	Changhua	Starbucks	latte	0	84	12585	149.82143
2021/4/1	Thursday	Changhua	fami	Americano	0	125	5633	45.06400

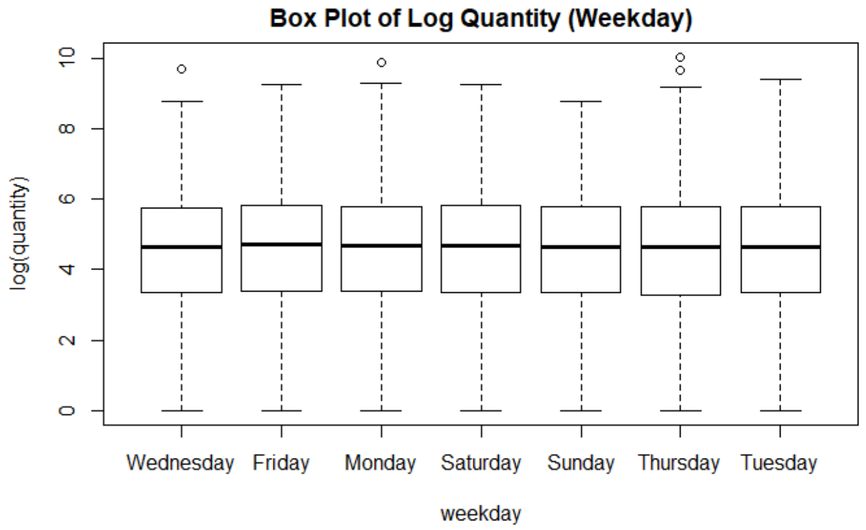




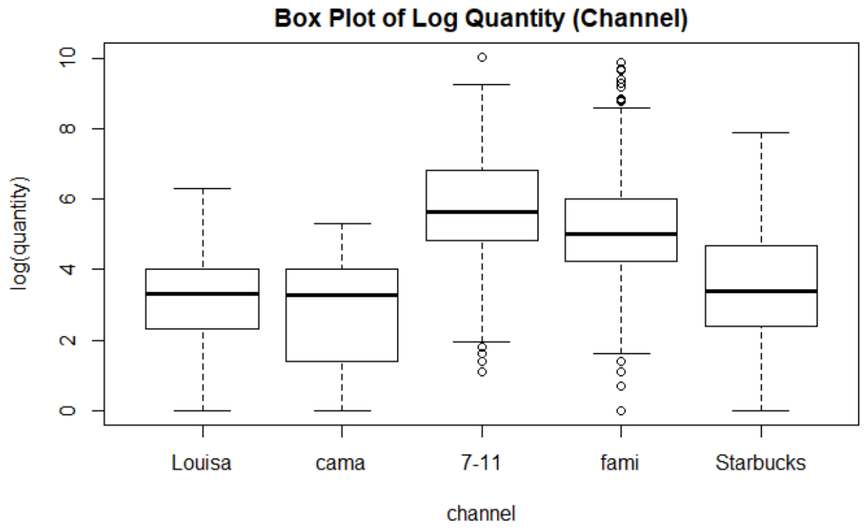
What factors affect daily average sales quantity?

EDA

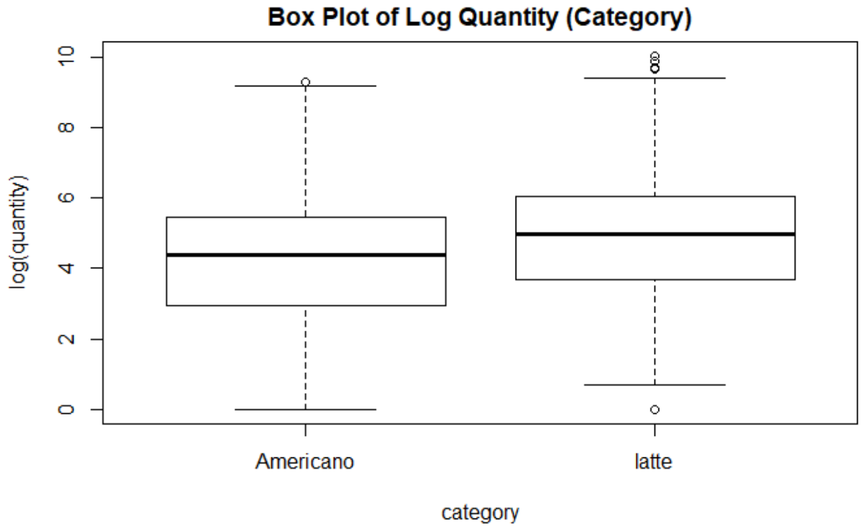
Weekday



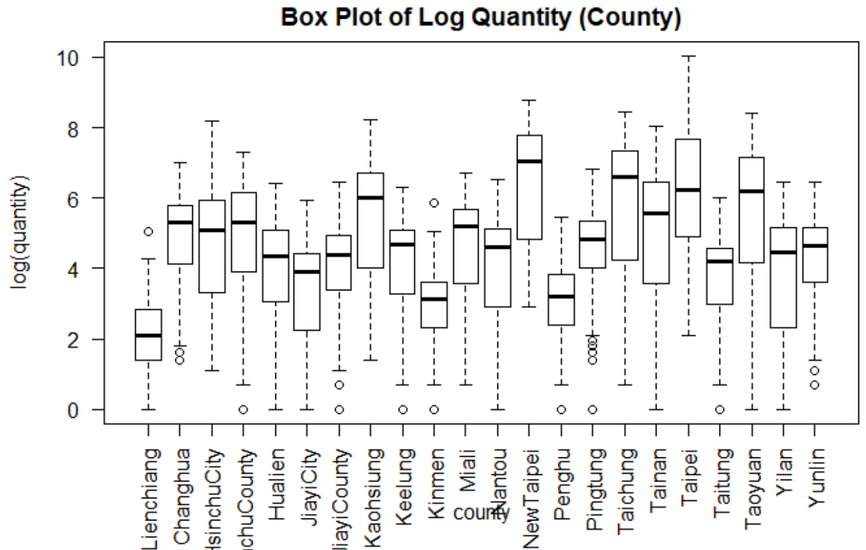
Channel



Category



County



What factors affect daily average sales quantity?

Proposed Linear Regression Model

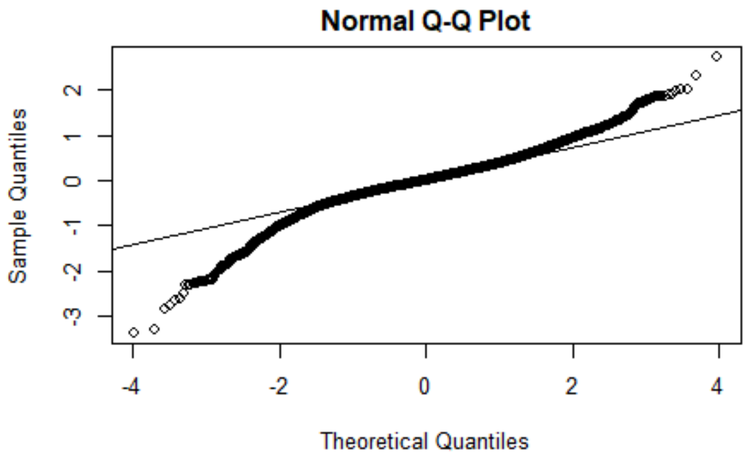
```
log(quantity) ~
                (weekday + county + channel * category * unit_price) * is_Alert
```

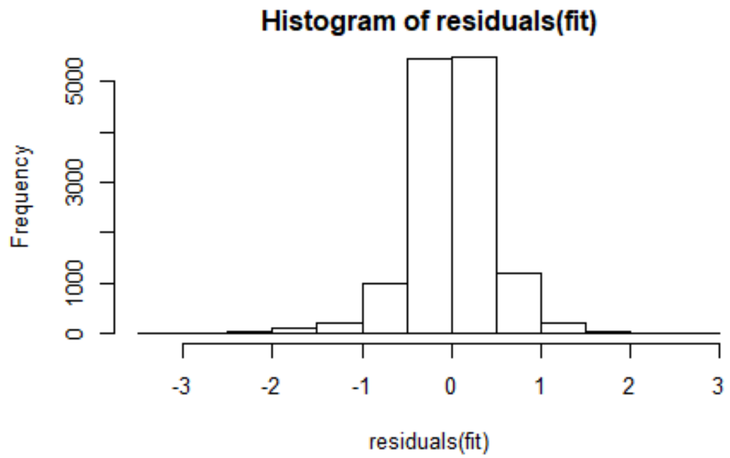
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	11.950362	0.257537	46.403	< 2e-16	***
weekdayMonday	0.083580	0.021383	3.909	9.32e-05	***
weekdayFriday	0.112840	0.020659	5.462	4.79e-08	***
weekdaySaturday	0.142814	0.021545	6.629	3.52e-11	***
weekdaySunday	0.143239	0.021496	6.664	2.77e-11	***
weekdayThursday	0.019165	0.020633	0.929	0.352986	
weekdayTuesday	0.058945	0.021366	2.759	0.005808	**
countyLienchiang	-5.651312	0.037364	-151.249	< 2e-16	***
countyChanghua	-2.734437	0.037224	-73.460	< 2e-16	***
countyHsinchuCity	-2.305356	0.033745	-68.316	< 2e-16	***
countyHsinchuCounty	-2.591141	0.035494	-73.002	< 2e-16	***
countyHualien	-3.518968	0.036749	-95.756	< 2e-16	***
countyJiayiCity	-3.723451	0.033881	-109.898	< 2e-16	***
countyJiayiCounty	-3.444939	0.036766	-93.699	< 2e-16	***
countyKaohsiung	-1.442800	0.033827	-42.652	< 2e-16	***
countyKeelung	-3.447490	0.036742	-93.831	< 2e-16	***
countyKinmen	-5.115235	0.037705	-135.665	< 2e-16	***
countyMiali	-3.034757	0.036883	-82.280	< 2e-16	***
countyNantou	-3.635675	0.036901	-98.525	< 2e-16	***
countyNewTaipei	-0.550216	0.033672	-16.340	< 2e-16	***
countyPenghu	-4.623364	0.036997	-124.965	< 2e-16	***
countyPingtung	-2.995404	0.036907	-81.161	< 2e-16	***
countyTaichung	-1.065206	0.033737	-31.574	< 2e-16	***
countyTainan	-1.873403	0.033781	-55.457	< 2e-16	***
countyTaitung	-3.860441	0.036783	-104.952	< 2e-16	***
countyTaoyuan	-1.224964	0.033763	-36.282	< 2e-16	***
countyYilan	-3.143439	0.033326	-94.324	< 2e-16	***
countyYunlin	-3.272297	0.036929	-88.610	< 2e-16	***
channelLouisa	-7.309857	0.453053	-16.135	< 2e-16	***
channelcama	-7.523387	0.618042	-12.173	< 2e-16	***
channelfami	-4.960488	0.366298	-13.542	< 2e-16	***
channelStarbucks	-6.354890	0.453066	-14.026	< 2e-16	***
categorylatte	-1.800792	0.462771	-3.891	0.000100	***
unit_price	-0.077626	0.005851	-13.268	< 2e-16	***
is_Alert	-2.194136	0.345875	-6.344	2.31e-10	***
channelLouisa:categorylatte	-2.291225	0.725537	-3.158	0.001592	**
channelcama:categorylatte	0.719722	1.090117	0.660	0.509121	
channelfami:categorylatte	1.693518	0.649300	2.608	0.009111	**
channelStarbucks:categorylatte	1.203588	0.777326	1.548	0.121556	
channelLouisa:unit_price	0.080225	0.009009	8.905	< 2e-16	***
channelcama:unit_price	0.065292	0.011861	5.505	3.76e-08	***
channelfami:unit_price	0.093299	0.008292	11.251	< 2e-16	***
channelStarbucks:unit_price	0.081480	0.006720	12.125	< 2e-16	***
categorylatte:unit_price	0.054828	0.009174	5.976	2.34e-09	***
weekdayMonday:is_Alert	-0.098732	0.029311	-3.368	0.000758	***
weekdayFriday:is_Alert	-0.151584	0.029343	-5.166	2.43e-07	***
weekdaySaturday:is_Alert	-0.169012	0.029419	-5.745	9.39e-09	***
weekdaySunday:is_Alert	-0.239006	0.029374	-8.137	4.41e-16	***
weekdayThursday:is_Alert	-0.097417	0.029298	-3.325	0.000886	***
weekdayTuesday:is_Alert	-0.062752	0.029272	-2.144	0.032071	*
countyLienchiang:is_Alert	-0.007824	0.053150	-0.147	0.882971	
countyChanghua:is_Alert	0.186051	0.051746	3.595	0.000325	***
countyHsinchuCity:is_Alert	0.267016	0.047052	5.675	1.42e-08	***
countyHsinchuCounty:is_Alert	0.342751	0.049560	6.916	4.86e-12	***
countyHualien:is_Alert	-0.072723	0.051392	-1.415	0.157077	
countyJiayiCity:is_Alert	0.211184	0.047777	4.420	9.94e-06	***
countyJiayiCounty:is_Alert	0.137565	0.051271	2.683	0.007304	**
countyKaohsiung:is_Alert	0.225628	0.047120	4.788	1.70e-06	***
countyKeelung:is_Alert	0.134176	0.051353	2.613	0.008990	**
countyKinmen:is_Alert	0.403570	0.052748	7.651	2.13e-14	***
countyMiali:is_Alert	0.252670	0.051414	4.914	9.01e-07	***
countyNantou:is_Alert	0.115281	0.051743	2.228	0.025901	*
countyNewTaipei:is_Alert	0.228420	0.046991	4.861	1.18e-06	***
countyPenghu:is_Alert	-0.019755	0.051879	-0.381	0.703369	
countyPingtung:is_Alert	0.113539	0.051392	2.209	0.027172	*
countyTaichung:is_Alert	0.179669	0.047064	3.818	0.000135	***
countyTainan:is_Alert	0.228746	0.047083	4.858	1.20e-06	***
countyTaitung:is_Alert	0.183834	0.051330	3.581	0.000343	***
countyTaoyuan:is_Alert	0.195413	0.047054	4.153	3.30e-05	***
countyYilan:is_Alert	-0.003259	0.047020	-0.069	0.944745	
countyYunlin:is_Alert	0.078988	0.051506	1.534	0.125159	
channelLouisa:is_Alert	0.533820	0.554031	0.964	0.335303	
channelcama:is_Alert	-2.132551	1.165857	-1.829	0.067396	.
channelfami:is_Alert	2.148821	0.518799	4.142	3.47e-05	***
channelStarbucks:is_Alert	1.613894	0.611206	2.641	0.008288	**
categorylatte:is_Alert	1.732155	0.551266	3.142	0.001681	**
unit_price:is_Alert	0.041618	0.007706	5.401	6.75e-08	***
channelLouisa:categorylatte:unit_price	0.004593	0.012668	0.363	0.716924	
channelcama:categorylatte:unit_price	-0.026438	0.017444	-1.516	0.129643	
channelfami:categorylatte:unit_price	-0.045277	0.012845	-3.525	0.000425	***
channelStarbucks:categorylatte:unit_price	-0.043848	0.010349	-4.237	2.28e-05	***
channelLouisa:categorylatte:is_Alert	1.005629	0.866998	1.160	0.246111	
channelcama:categorylatte:is_Alert	4.425522	1.599493	2.767	0.005668	**
channelfami:categorylatte:is_Alert	-3.157069	0.879574	-3.589	0.000333	***
channelStarbucks:categorylatte:is_Alert	-1.861369	0.985405	-1.889	0.058921	.
channelLouisa:unit_price:is_Alert	-0.030485	0.011068	-2.754	0.005888	**
channelcama:unit_price:is_Alert	0.030780	0.021535	1.429	0.152945	
channelfami:unit_price:is_Alert	-0.045843	0.011603	-3.951	7.82e-05	***
channelStarbucks:unit_price:is_Alert	-0.043806	0.008901	-4.921	8.69e-07	***
categorylatte:unit_price:is_Alert	-0.039541	0.011025	-3.586	0.000337	***
channelLouisa:categorylatte:unit_price:is_Alert	0.006714	0.015167	0.443	0.658020	
channelcama:categorylatte:unit_price:is_Alert	-0.061498	0.026385	-2.331	0.019778	*
channelfami:categorylatte:unit_price:is_Alert	0.065761	0.017363	3.787	0.000153	***
channelStarbucks:categorylatte:unit_price:is_Alert	0.041477	0.012685	3.270	0.001079	**

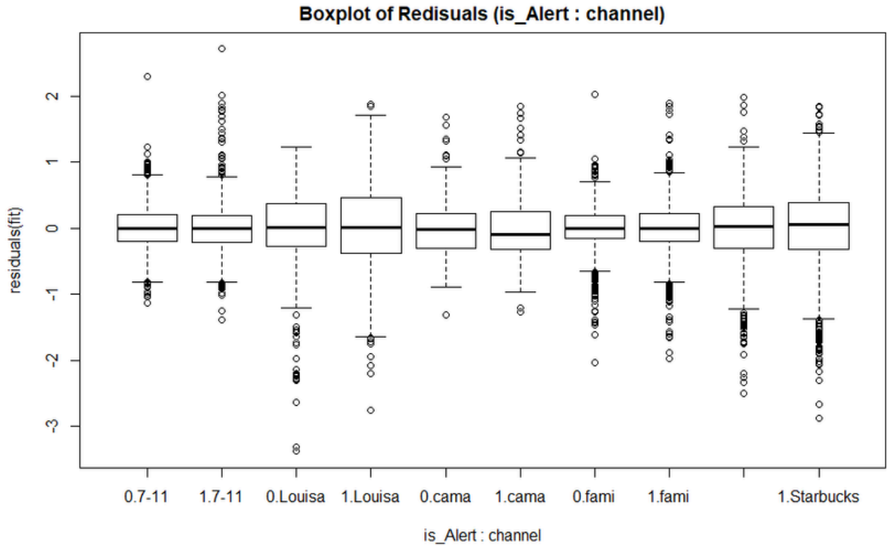
```
Residual standard error: 0.4559
    on 13577 degrees of freedom

Multiple R-squared: 0.9378

Adjusted R-squared: 0.9373

F-statistic: 2200
    on 93 and 13577 DF

p-value: < 2.2e-16
```

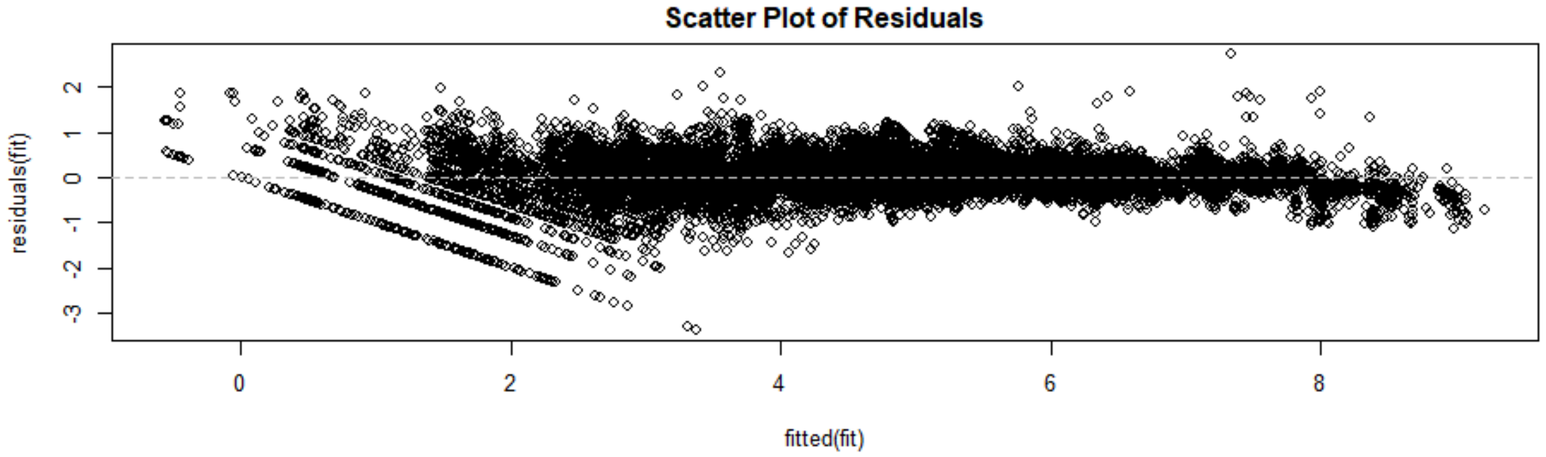
What factors affect daily average sales quantity?

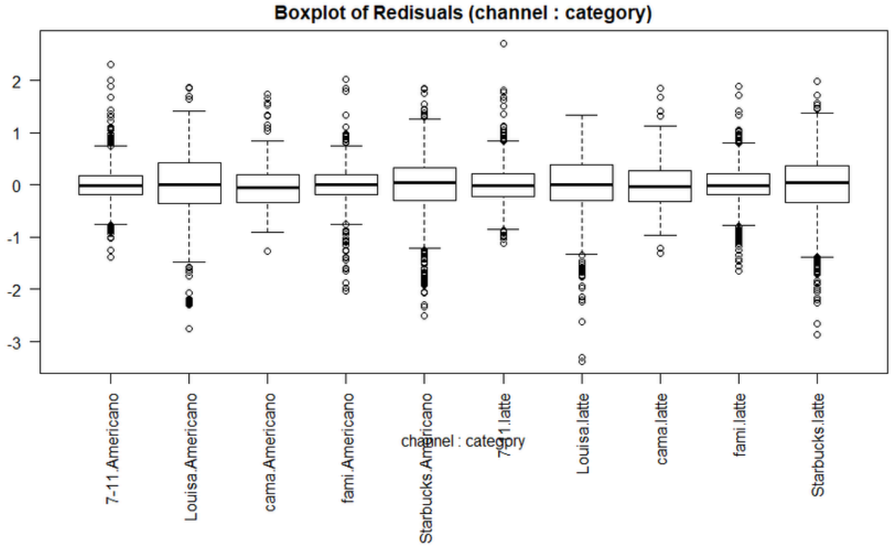
3+1 Regression Assumptions











Interpretation

Weekday	County	Channel	Category	is_Alert	Formula and Interpretation
Sunday	Taipei	7-11	Latte	0	$\log(\widehat{\text{quantity}}) = 10.2928 - 0.0228\text{unit\_price}$ $\widehat{\text{quantity}} = \exp(10.2928 - 0.0228\text{unit\_price})$
					We expect daily coffee sales quantity to increase by 2.3 % for each one **decrease** of unit price.
Sunday	Taipei	7-11	Latte	1	$\log(\widehat{\text{quantity}}) = 9.5918 - 0.0207\text{unit\_price}$ $\widehat{\text{quantity}} = \exp(9.5918 - 0.0207\text{unit\_price})$
					We expect daily coffee sales quantity to increase by 2 % for each one **decrease** of unit price.
Sunday	Taipei	Starbucks	Latte	0	$\log(\widehat{\text{quantity}}) = 5.1415 + 0.0148\text{unit\_price}$ $\widehat{\text{quantity}} = \exp(5.1415 + 0.0148\text{unit\_price})$
					We expect daily coffee sales quantity to increase by 1.4 % for each one **increase** of unit price.
Sunday	Taipei	Starbucks	Latte	1	$\log(\widehat{\text{quantity}}) = 4.1930 + 0.0145\text{unit\_price}$ $\widehat{\text{quantity}} = \exp(4.1930 + 0.0145\text{unit\_price})$
					We expect daily coffee sales quantity to increase by 1.4 % for each one **increase** of unit price.

ANOVA

Category

Channel

Time Aspect

Time Interval

Does interaction exist between categories and time intervals?

Does interaction exist between channel and time intervals?

Space Aspect

Taipei District

Does interaction exist between category and Taipei districts?

Does interaction exist between channels and Taipei districts?

4. ANOVA - Time Aspect

Does interaction exist between categories and time intervals?

Does interaction exist between channels and time intervals?

Does interaction exist between categories and time intervals?

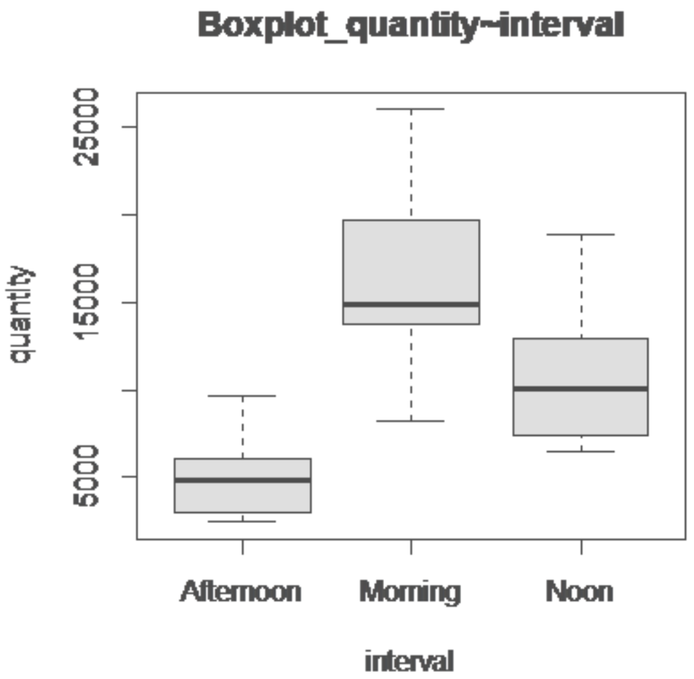
Dataset and EDA

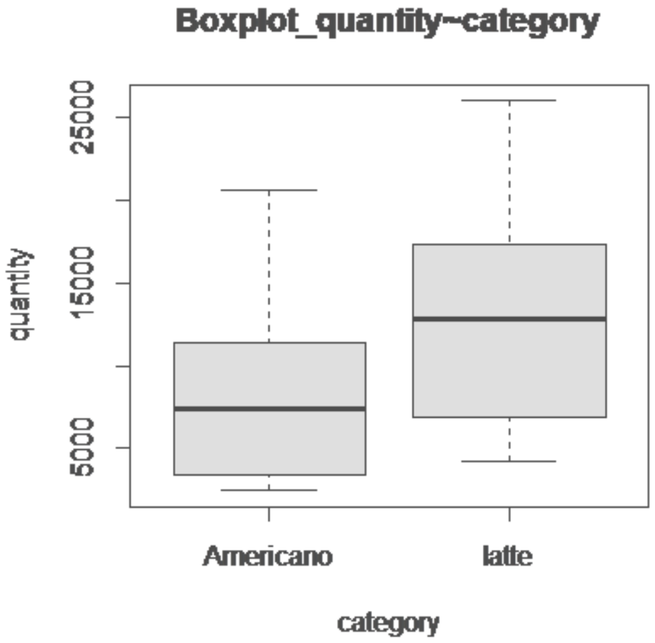
Factor A: category (Americano / Latte)

Factor B: interval (Morning: 6 - 11 Noon: 11 - 16 Afternoon: 16-21)

Response variable: log Quantity

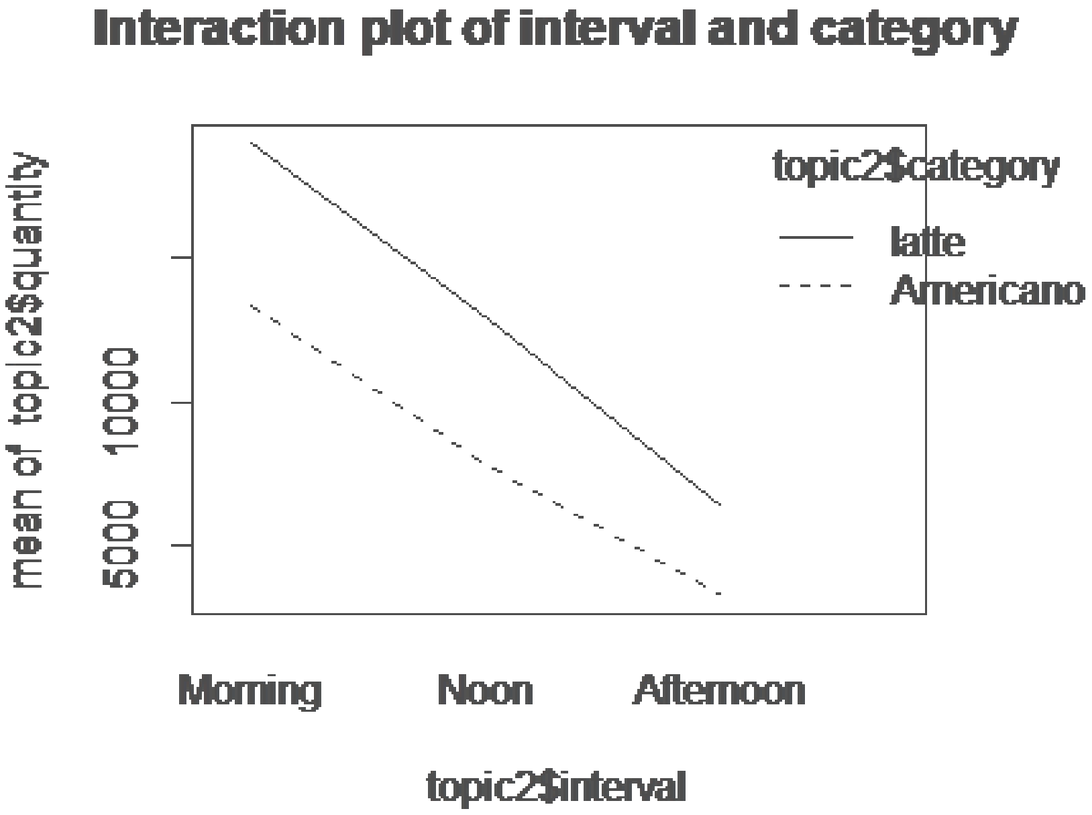
date	category	interval	quantity
2021/4/1	Americano	Afternoon	2975
2021/4/1	Americano	Morning	12895
2021/4/1	Americano	Noon	6519
2021/4/1	latte	Afternoon	5501
2021/4/1	latte	Morning	19185
2021/4/1	latte	Noon	10927

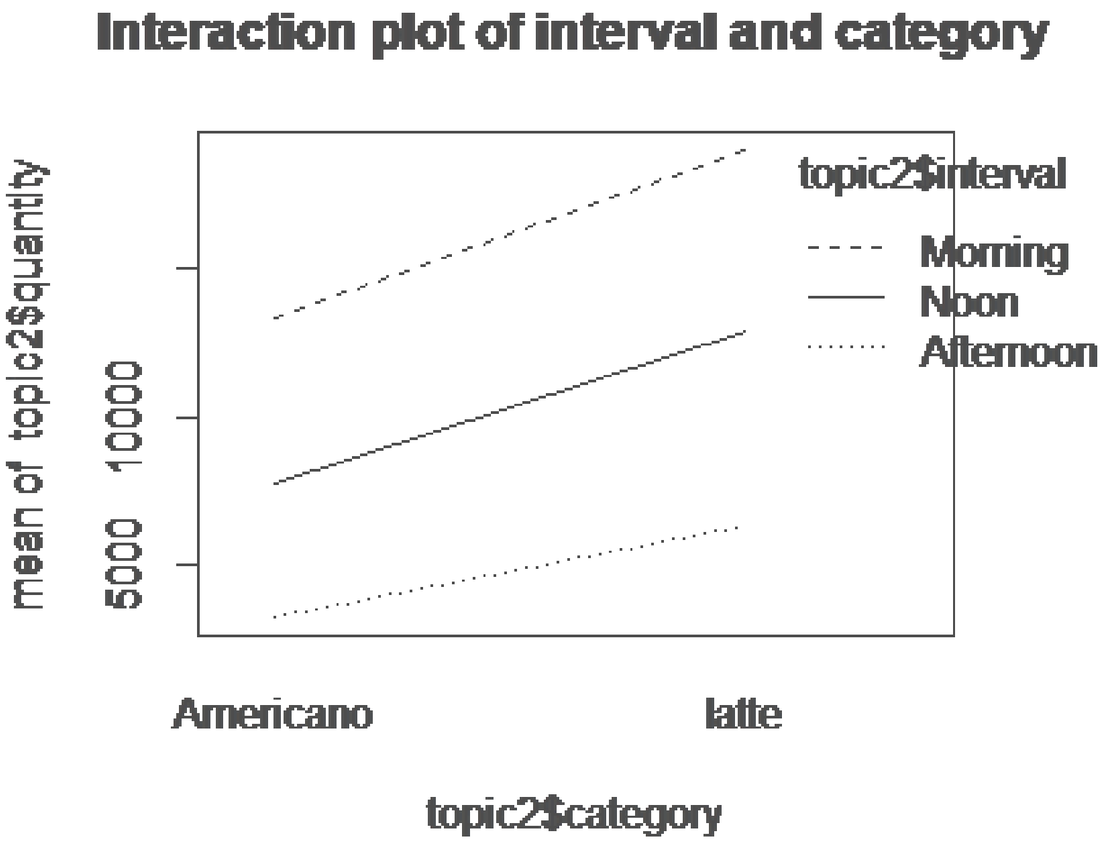




Does interaction exist between categories and time intervals?

Interaction plot





Does interaction exist between categories and time intervals?

ANOVA

```
Call:
lm(formula = log(topic2$quantity) ~ topic2$interval * topic2$category)

Residuals:
     Min       1Q   Median       3Q      Max
-0.46826 -0.08979  0.00511  0.08177  0.44920

Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                                  8.07865    0.02354 343.132  < 2e-16 ***
topic2$intervalMorning                       1.40588    0.03330  42.224  < 2e-16 ***
topic2$intervalNoon                          0.86930    0.03330  26.108  < 2e-16 ***
topic2$categorylatte                         0.66354    0.03330  19.928  < 2e-16 ***
topic2$intervalMorning:topic2$categorylatte -0.30729    0.04709  -6.526 3.56e-10 ***
topic2$intervalNoon:topic2$categorylatte    -0.15155    0.04709  -3.218  0.00145 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1562 on 258 degrees of freedom
Multiple R-squared:  0.9339,    Adjusted R-squared:  0.9326
F-statistic: 728.9 on 5 and 258 DF,  p-value: < 2.2e-16
```

R-squared: 0.9339

Analysis of Variance Table

Model 1: logq ~ topic2$interval + topic2$category
Model 2: logq ~ topic2$interval * topic2$category

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	260	7.3313				
2	258	6.2926	2	1.0388	21.295	2.756e-09 ***

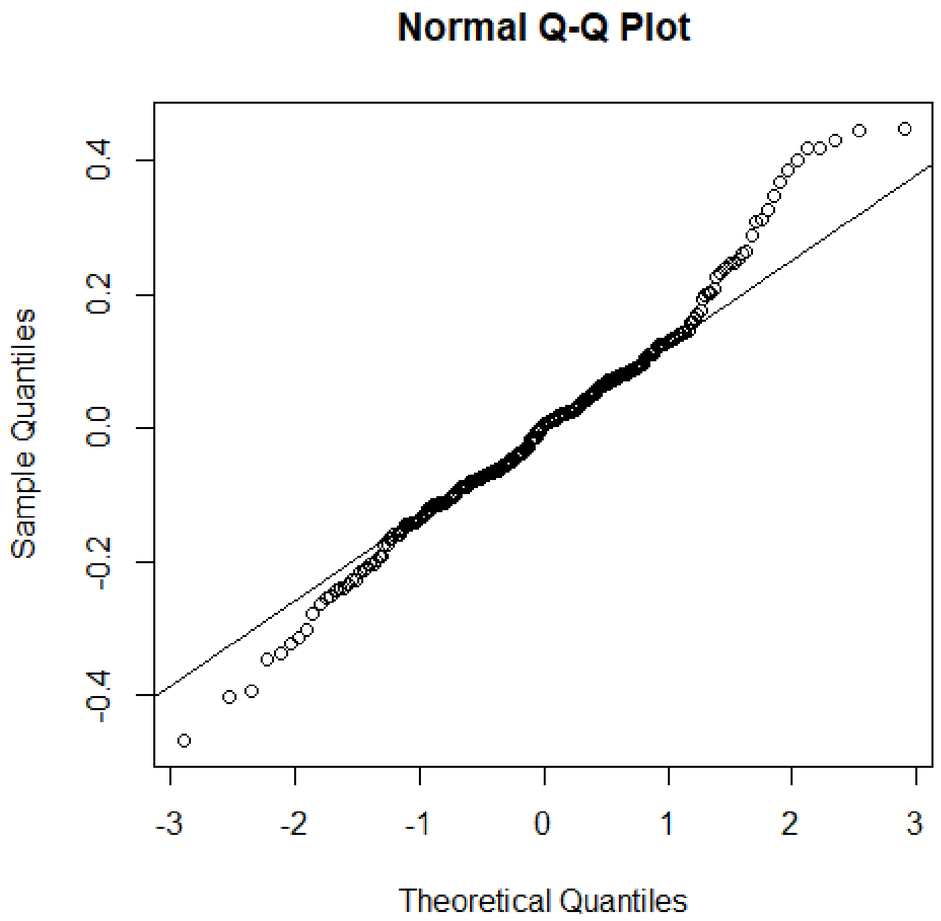
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

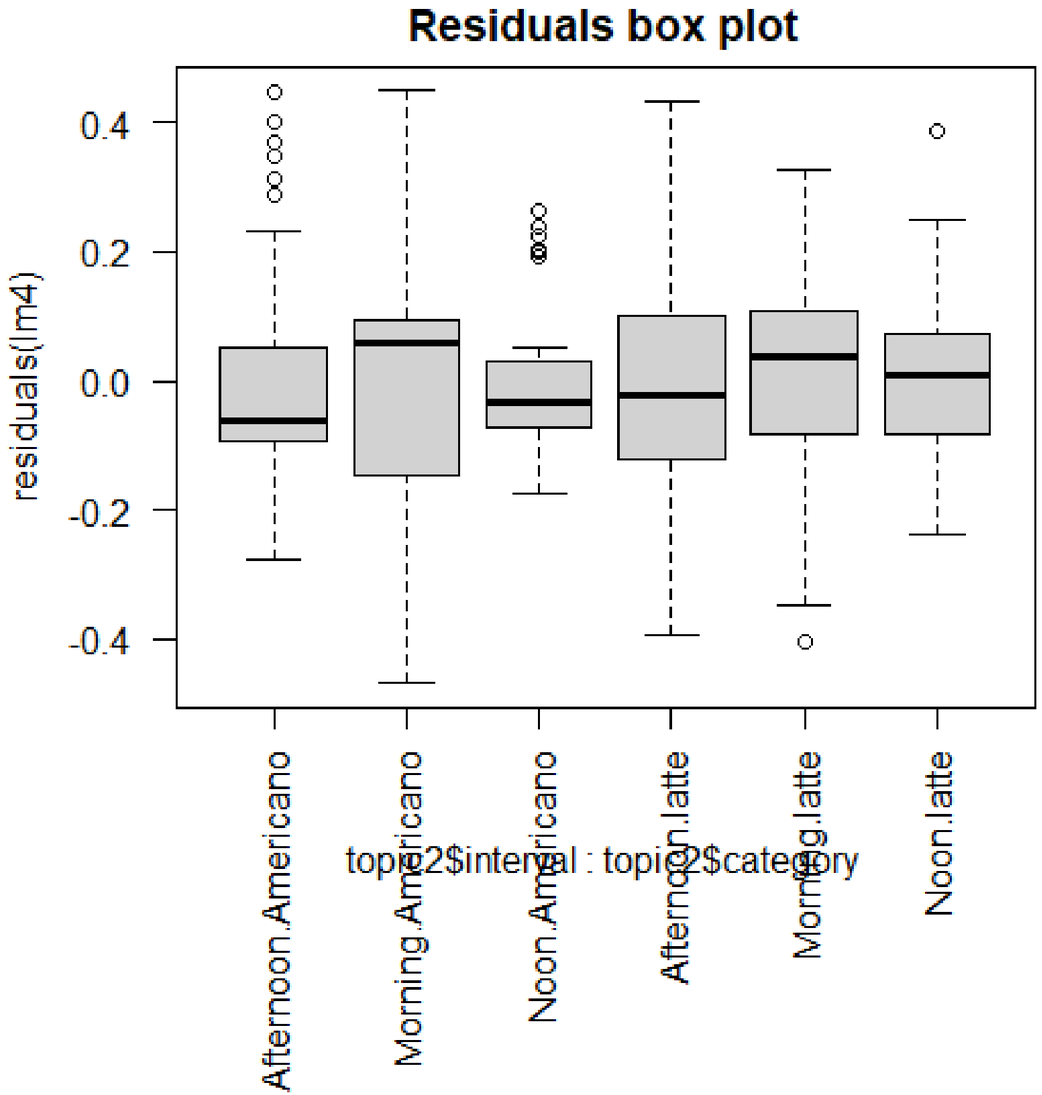
--> lm2 is better

--> interaction exists

Does interaction exist between categories and time intervals?

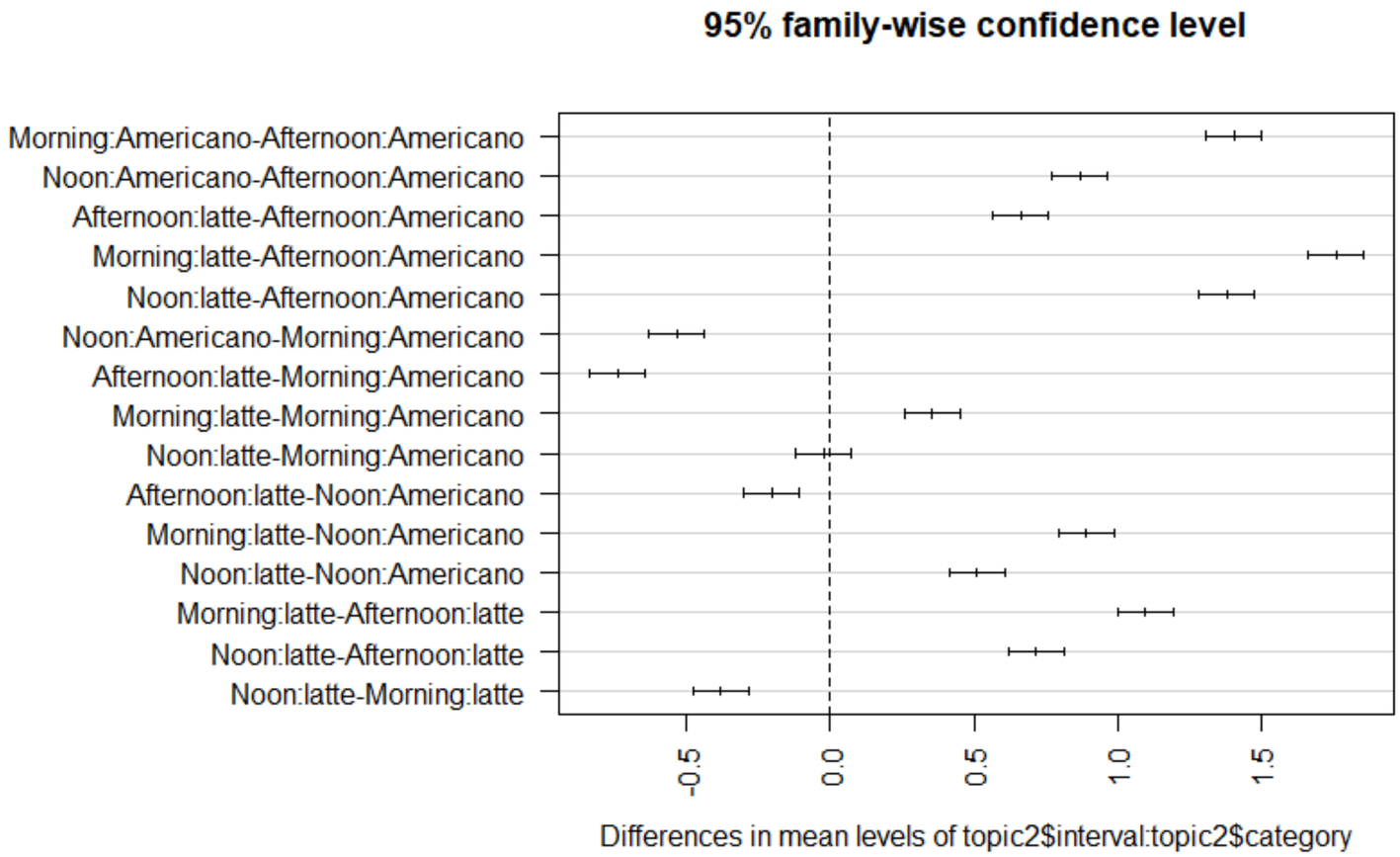
Check Assumptions





Does interaction exist between categories and time intervals?

Multiple Comparisons of Mean



Interpretation :

Non-significant :

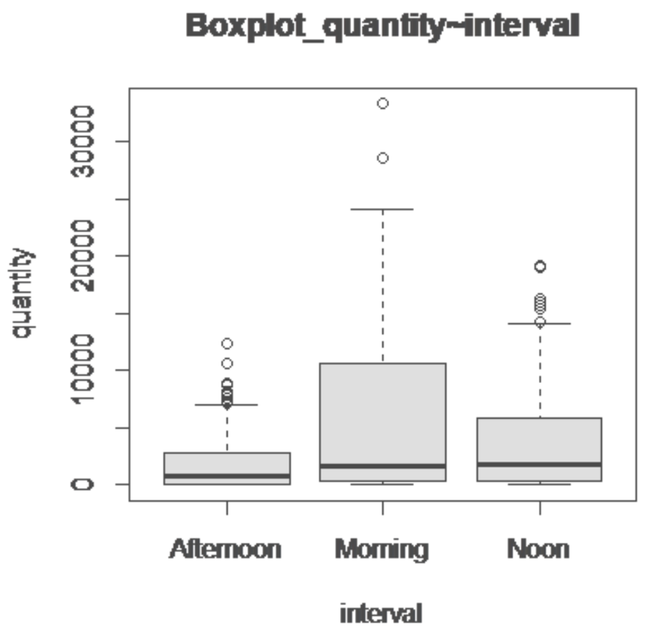
Noon:latte-Morning:Americano

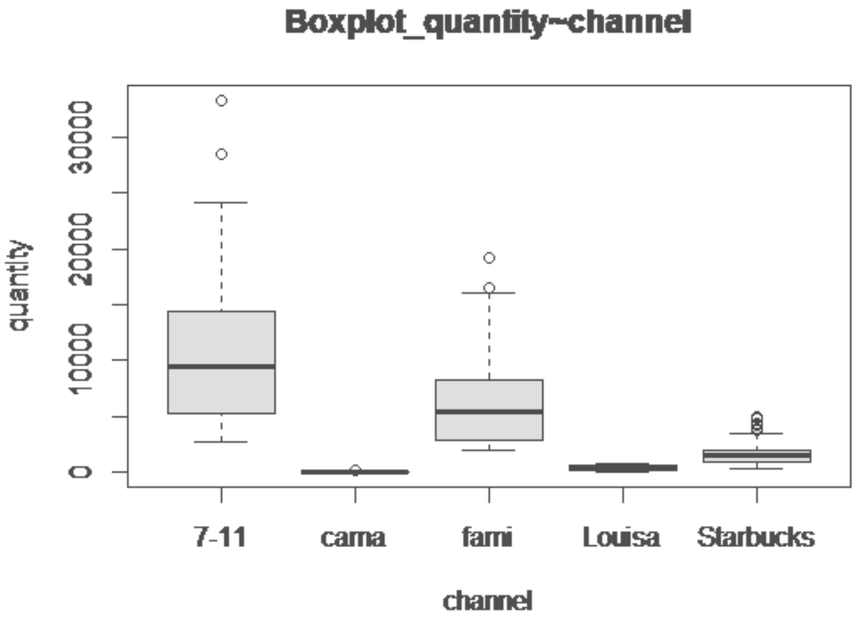
Does interaction exist between channels and time intervals?

Dataset and EDA

Factor A: channel (7-11/ FamilyMart/ Starbucks/ Louisa/ Cama)
Factor B: interval (Morning: 6-11 Noon: 11-16 Afternoon: 16-21)
Response variable: log Quantity

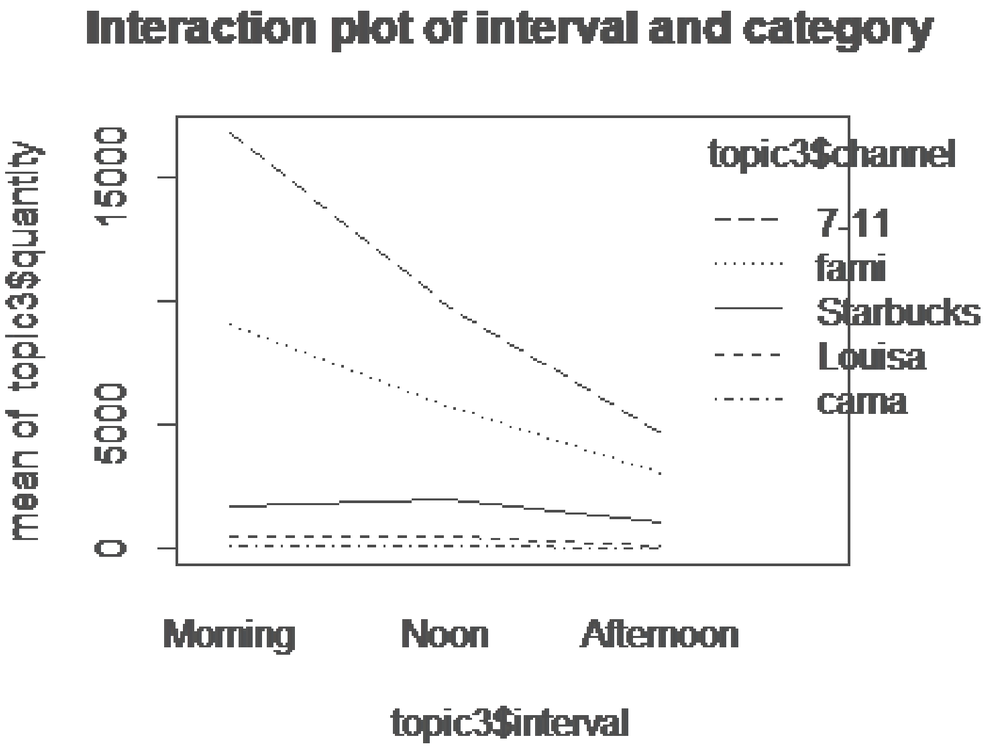
date	channel	interval	quant_
2021/4/1	7-11	Afternoon	4487
2021/4/1	7-11	Morning	18741
2021/4/1	7-11	Noon	9511
2021/4/1	Louisa	Afternoon	176
2021/4/1	Louisa	Morning	599
2021/4/1	Louisa	Noon	527

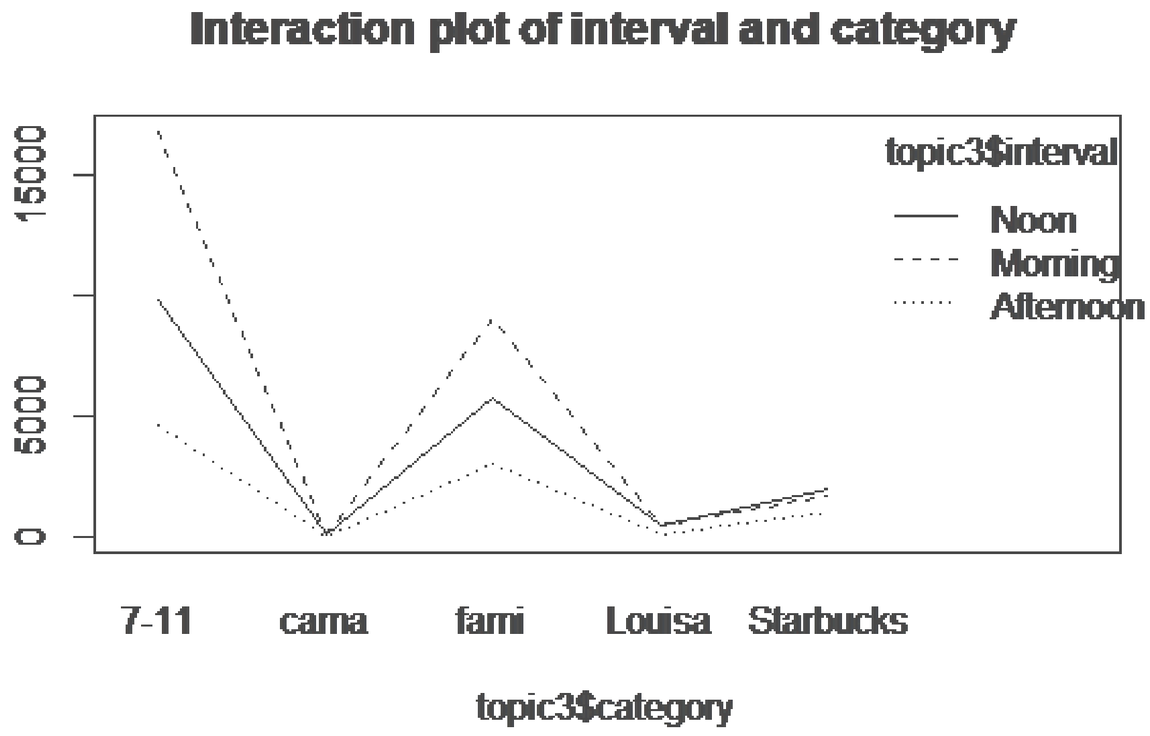




Does interaction exist between channels and time intervals?

Interaction plot





The lines above are not approximately parallel → substantial interaction exists

Does interaction exist between channels and intervals?

ANOVA

```
Call:
lm(formula = logq ~ topic3$interval * topic3$channel)

Residuals:
     Min       1Q   Median       3Q      Max
-2.02767 -0.26540 -0.01536  0.27602  1.49160

Coefficients:
                                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                                      2.02767    0.04727  42.892  < 2e-16 ***
topic3$intervalMorning                           1.90129    0.06667  28.517  < 2e-16 ***
topic3$intervalNoon                              1.91816    0.06667  28.770  < 2e-16 ***
topic3$channel7-11                               6.36971    0.06667  95.538  < 2e-16 ***
topic3$channelfami                               5.90441    0.06667  88.559  < 2e-16 ***
topic3$channelLouisa                             2.49714    0.06667  37.454  < 2e-16 ***
topic3$channelStarbucks                          4.67654    0.06667  70.143  < 2e-16 ***
topic3$intervalMorning:topic3$channel7-11       -0.60344    0.09416  -6.409 2.02e-10 ***
topic3$intervalNoon:topic3$channel7-11          -1.14735    0.09416 -12.185  < 2e-16 ***
topic3$intervalMorning:topic3$channelfami       -0.75024    0.09416  -7.968 3.41e-15 ***
topic3$intervalNoon:topic3$channelfami          -1.23554    0.09416 -13.122  < 2e-16 ***
topic3$intervalMorning:topic3$channelLouisa     -0.40772    0.09416  -4.330 1.60e-05 ***
topic3$intervalNoon:topic3$channelLouisa        -0.41867    0.09416  -4.446 9.45e-06 ***
topic3$intervalMorning:topic3$channelStarbucks -1.23129    0.09416 -13.077  < 2e-16 ***
topic3$intervalNoon:topic3$channelStarbucks     -1.12647    0.09416 -11.964  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4485 on 1349 degrees of freedom
Multiple R-squared:  0.9604,    Adjusted R-squared:   0.96
F-statistic:  2339 on 14 and 1349 DF,  p-value: < 2.2e-16
```

R-squared: 0.9604

Analysis of Variance Table

Model 1: logq ~ topic3$interval + topic3$channel

Model 2: logq ~ topic3$interval * topic3$channel

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1357	343.52				
2	1349	271.33	8	72.188	44.863	< 2.2e-16 ***

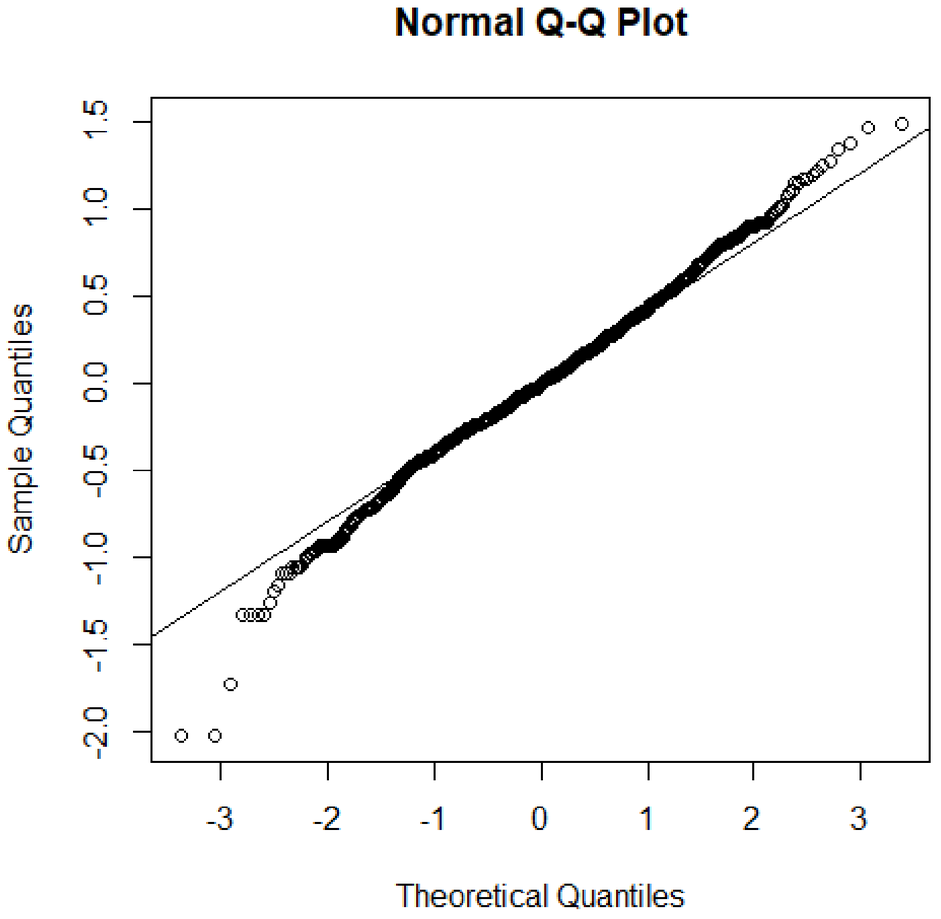
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

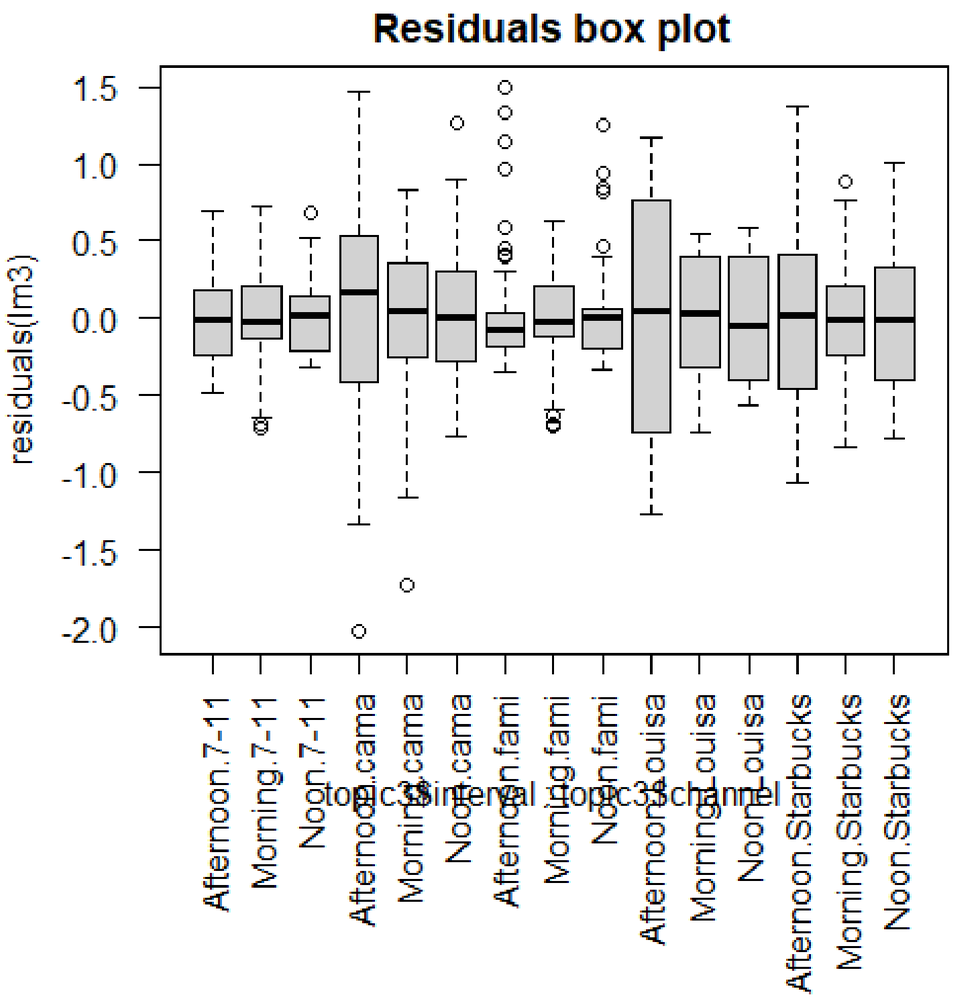
--> lm2 is better

--> interaction exists

Does interaction exist between channels and intervals?

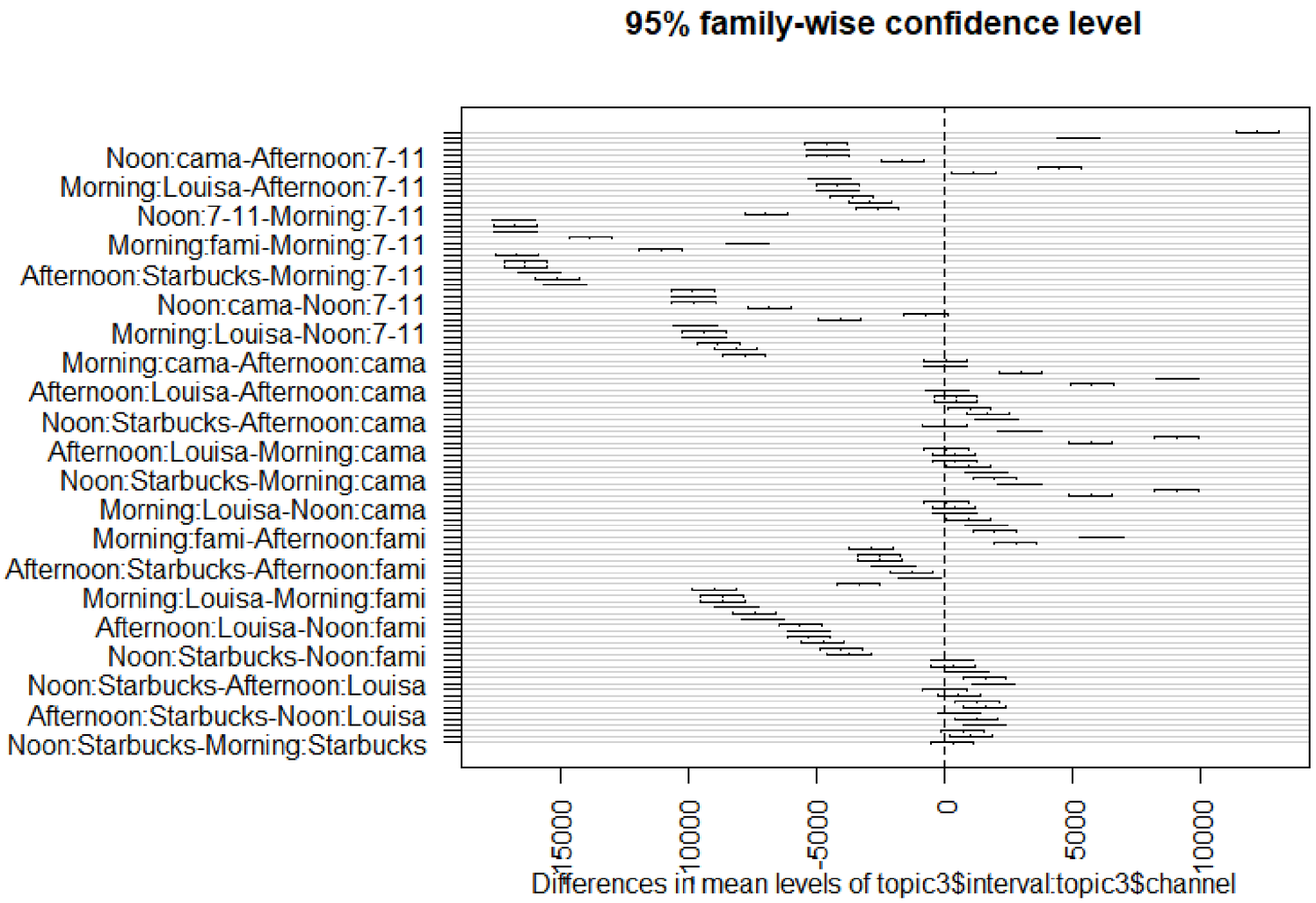
Check Assumptions





Does interaction exist between channels and intervals?

Multiple Comparisons of Mean



Non-sigificant :

Morning:cama-Afternoon:cama
Noon:cama-Afternoon:cama
Noon:cama-Morning:cama
Morning:Louisa-Afternoon:Louisa
Noon:Louisa-Afternoon:Louisa
Noon:Louisa-Morning:Louisa
Morning:Starbucks-Afternoon:Starbucks
Noon:Starbucks-Morning:Starbucks

Afternoon:Starbucks-Morning:Louisa
Afternoon:Starbucks-Noon:Louisa
Afternoon:Louisa-Afternoon:cama
Morning:Louisa-Afternoon:cama
Noon:Louisa-Afternoon:cama
Afternoon:Louisa-Morning:cama
Morning:Louisa-Morning:cama
Noon:Louisa-Morning:cama
Afternoon:Louisa-Noon:cama
Morning:Louisa-Noon:cama
Noon:Louisa-Noon:cama
Morning:fami-Noon:7-11

5. ANOVA -Space Aspect

Does interaction exist between categorys and Taipei districts?

Does interaction exist between channels and Taipei districts?

Does interaction exist between categorys and Taipei districts?

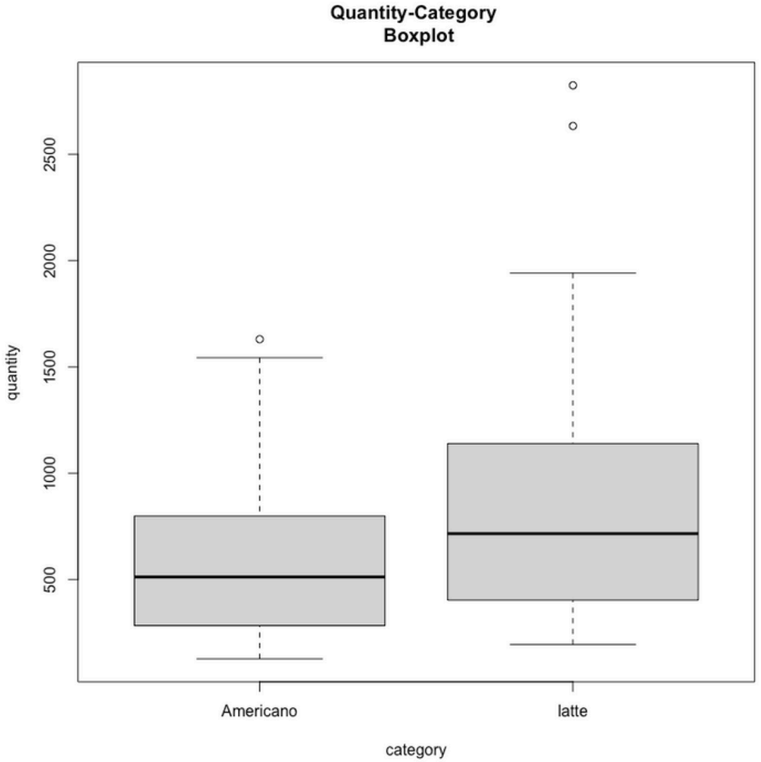
Dataset and EDA

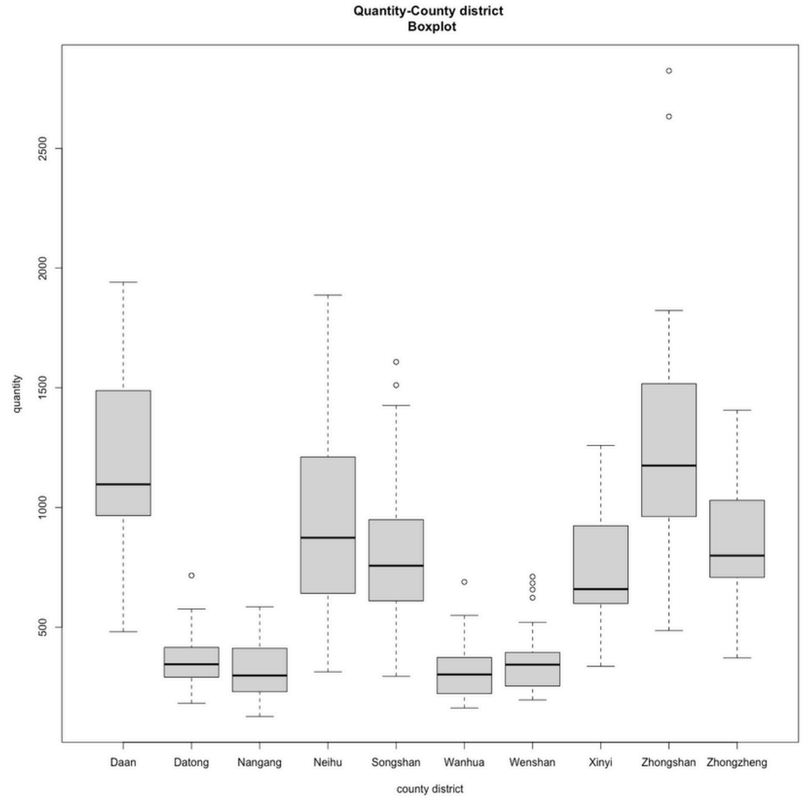
Factor A: category (Americano / Latte)

Factor B: Taipei_district

Response variable: log Quantity

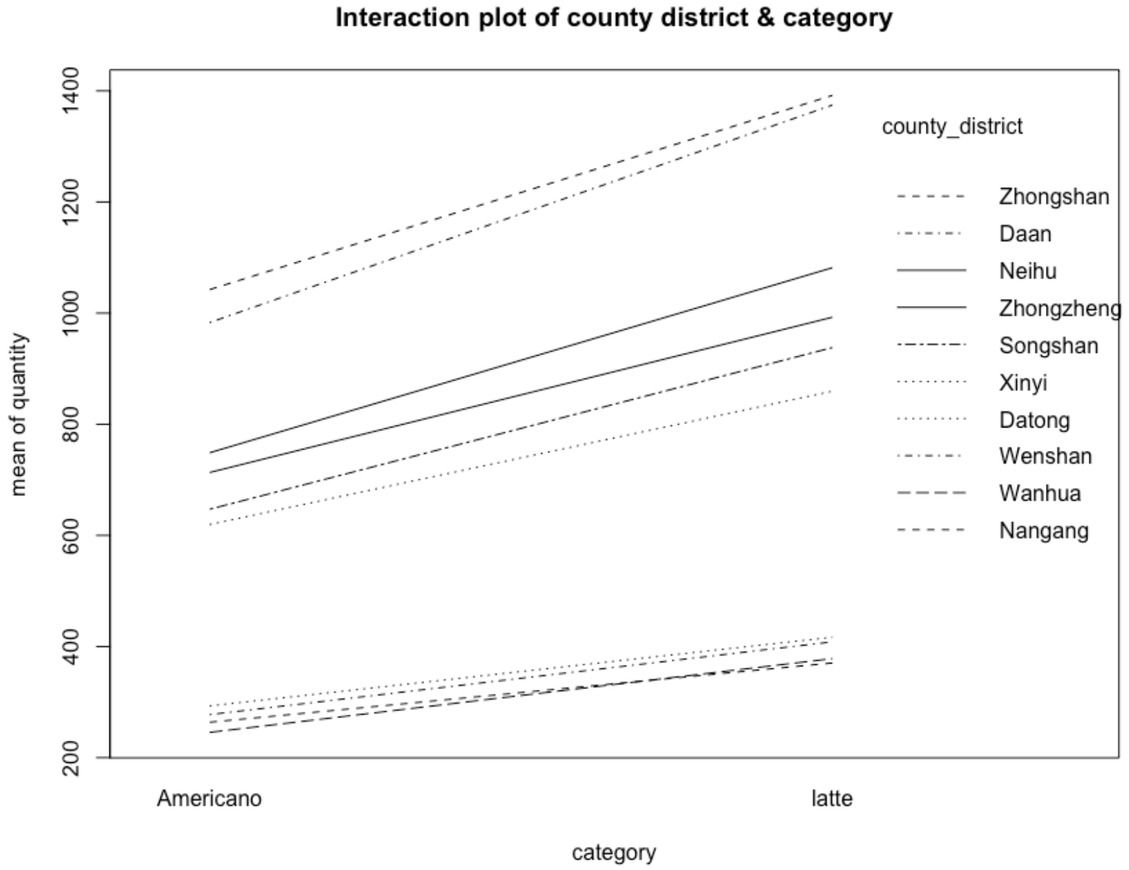
date	county_distric	category	quant_
2021/4/1	Datong	Americano	251
2021/4/1	Datong	latte	387
2021/4/1	Daan	Americano	954
2021/4/1	Daan	latte	1417
2021/4/1	Nangang	Americano	310
2021/4/1	Nangang	latte	402
2021/4/1	Neihu	Americano	755
2021/4/1	Neihu	latte	1110
2021/4/1	Songshan	Americano	616
2021/4/1	Songshan	latte	915
2021/4/1	Wanhua	Americano	196
2021/4/1	Wanhua	latte	333
2021/4/1	Wenshan	Americano	221
2021/4/1	Wenshan	latte	347
2021/4/1	Xinyi	Americano	646
2021/4/1	Xinyi	latte	847

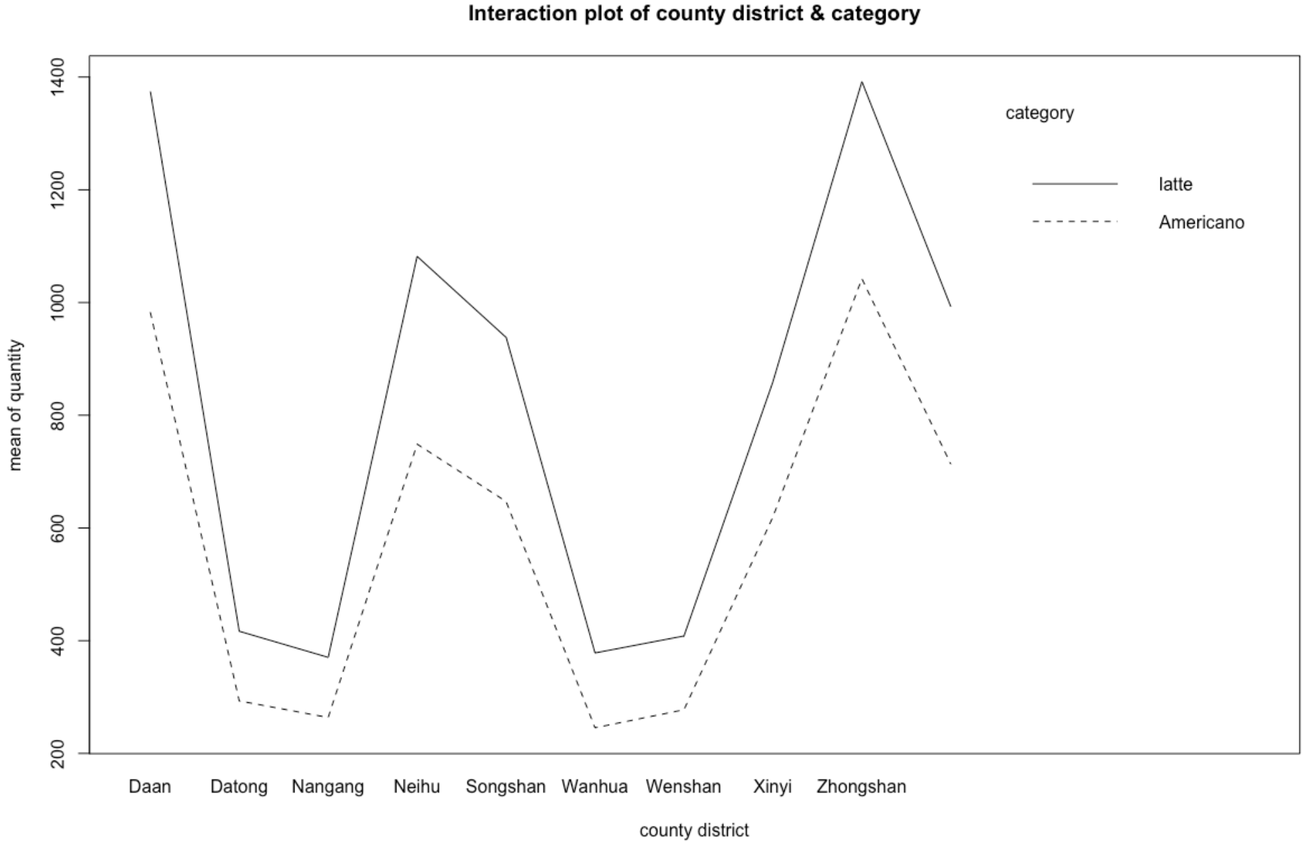




Does interaction exist between categories and Taipei districts?

Interaction plot





Does interaction exist between categories and Taipei districts?

ANOVA

```
Call:
lm(formula = log(quant_) ~ category + county_district)

Residuals:
     Min       1Q   Median       3Q      Max
-0.81840 -0.15128  0.05658  0.17058  0.75486

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                5.51234    0.03068 179.658  < 2e-16 ***
categorylatte              0.35237    0.01852  19.025  < 2e-16 ***
county_districtDaan        1.34012    0.04137  32.395  < 2e-16 ***
county_districtDatong      0.15244    0.04137   3.685 0.000243 ***
county_districtNangang     0.01074    0.04137   0.260 0.795146
county_districtNeihu       1.05226    0.04137  25.437  < 2e-16 ***
county_districtSongshan    0.92238    0.04137  22.297  < 2e-16 ***
county_districtWenshan     0.11027    0.04137   2.665 0.007830 **
county_districtXinyi       0.87254    0.04137  21.092  < 2e-16 ***
county_districtZhongshan   1.35738    0.04161  32.623  < 2e-16 ***
county_districtZhongzheng  1.02449    0.04137  24.765  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2744 on 867 degrees of freedom
Multiple R-squared:  0.8049,    Adjusted R-squared:  0.8026
F-statistic: 357.6 on 10 and 867 DF,  p-value: < 2.2e-16
```

R-squared: 0.8049

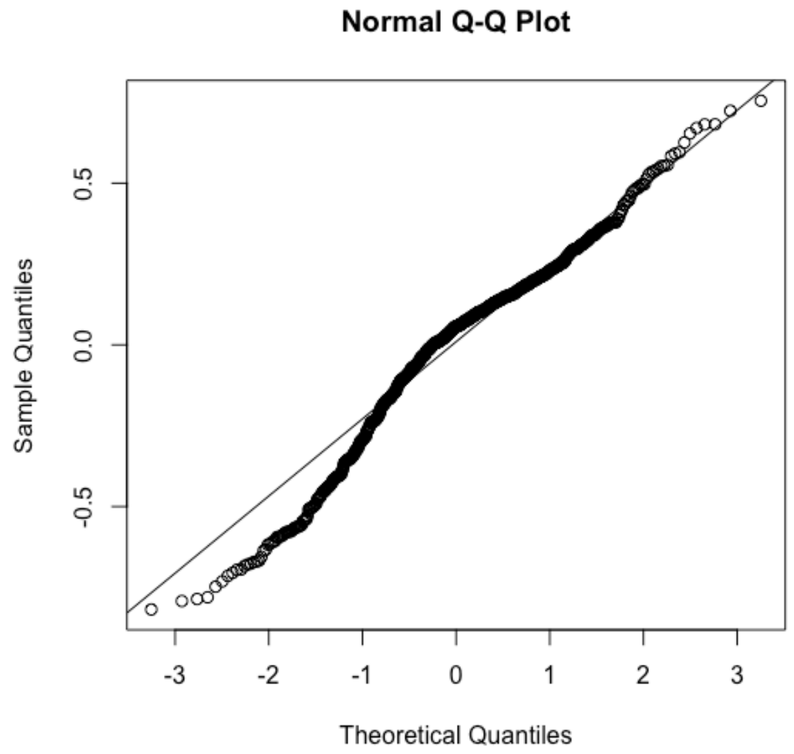
```
> anova(mod2,mod1)
Analysis of Variance Table

Model 1: log(quant_) ~ category + county_district
Model 2: log(quant_) ~ category * county_district
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    869 68.127
2    860 67.745  9   0.38263 0.5397 0.8461
```

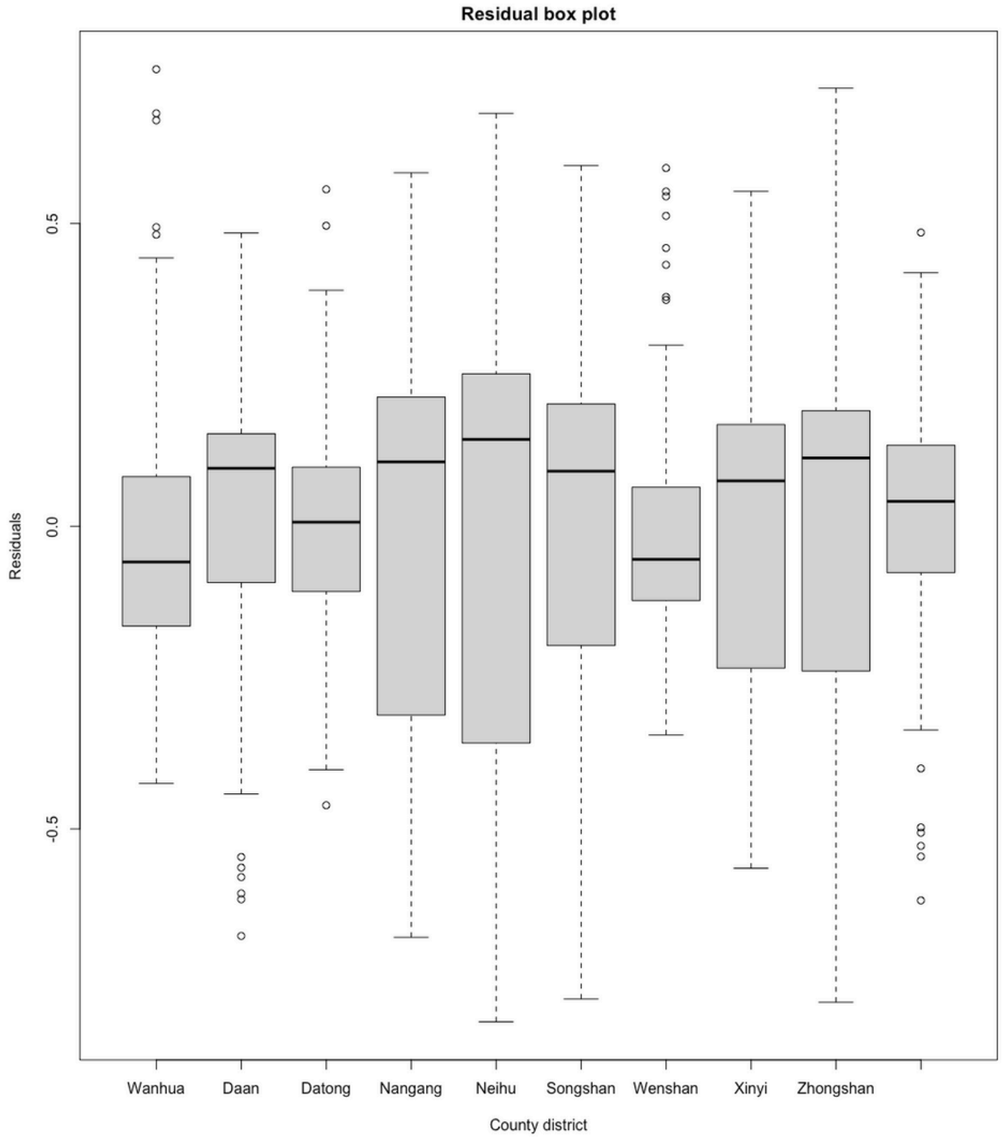
--> lm1 is better

--> without interaction

Does interaction exist between categories and Taipei districts?

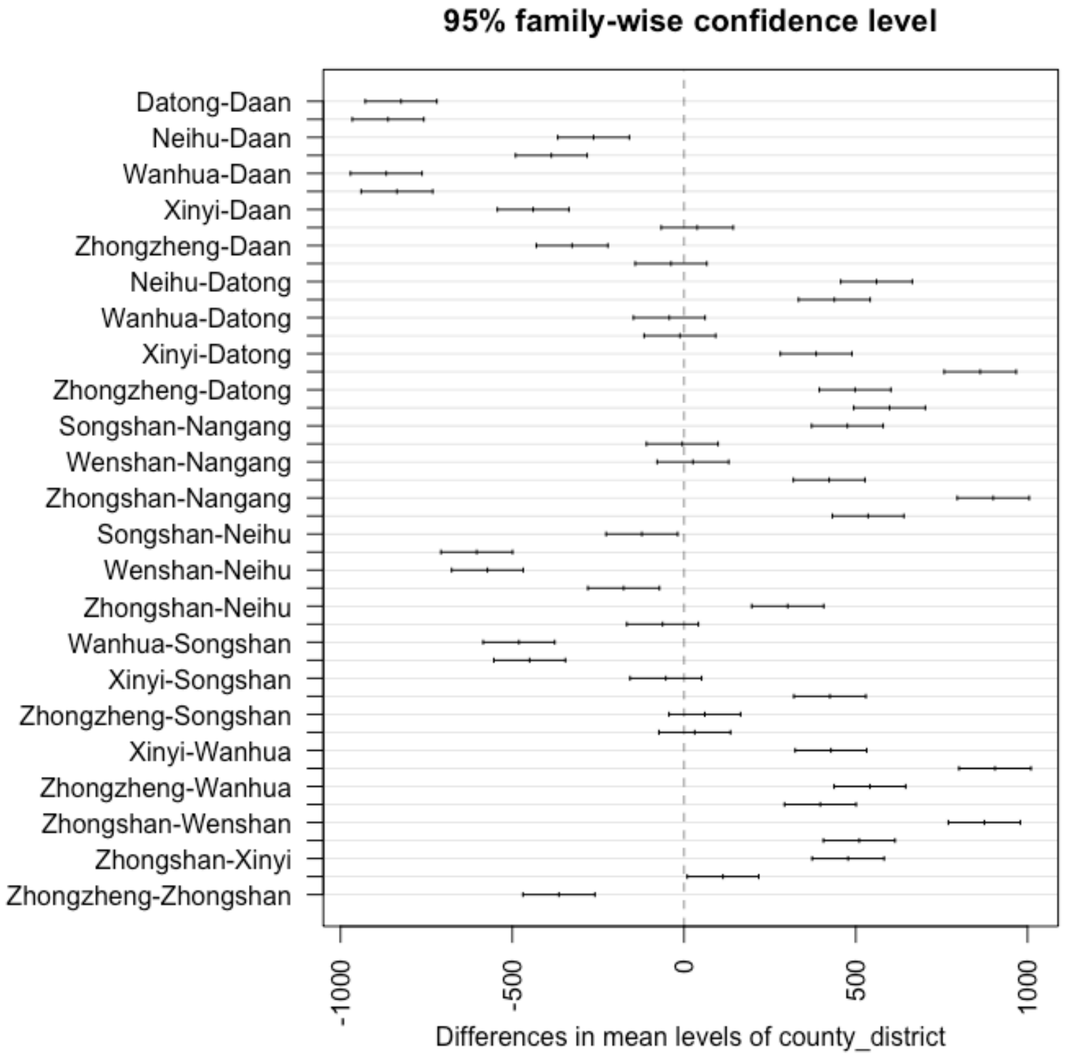
Check Assumptions





Does interaction exist between categories and Taipei districts?

Multiple Comparisons of Mean



Non-sigificant :

Zhongshan-Daan
Wenshan-Datong
Wanhua-Nangang
Wenshan-Nangang
Songshan-Neihu
Zhongzheng-Neihu
Xinyi-Songshan
Zhongzheng-Songshan
Wenshan-Wanhua

Does interaction exist between channels and Taipei districts?

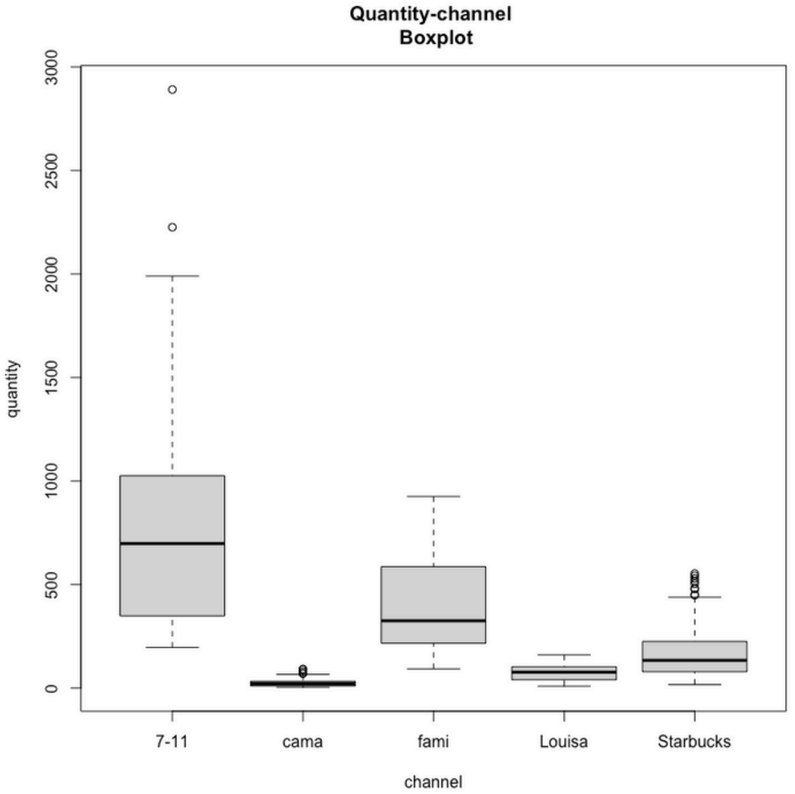
Dataset and EDA

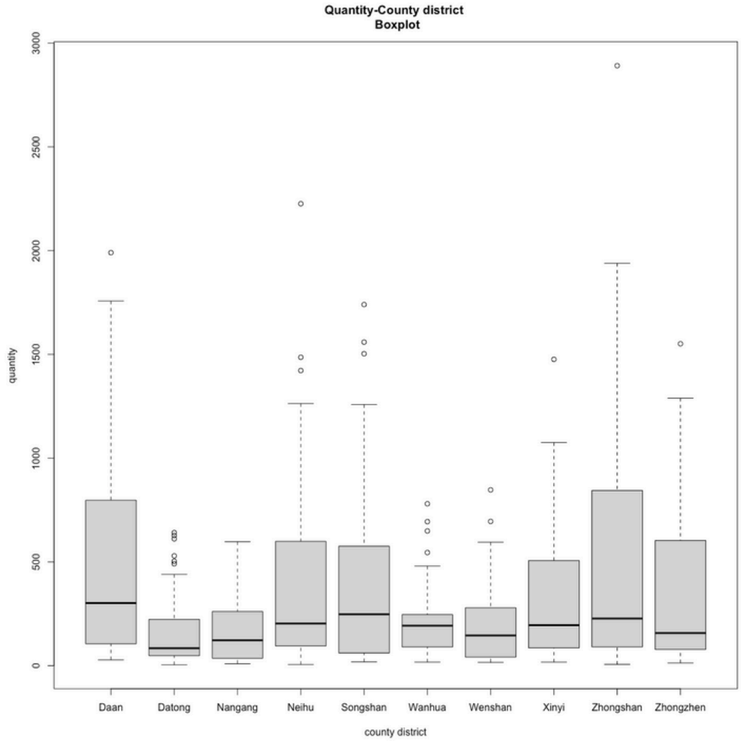
Factor A: channel (7-11/ FamilyMart/ Starbucks/ Louisa/ Cama)

Factor B: Taipei_district

Response variable: log Quantity

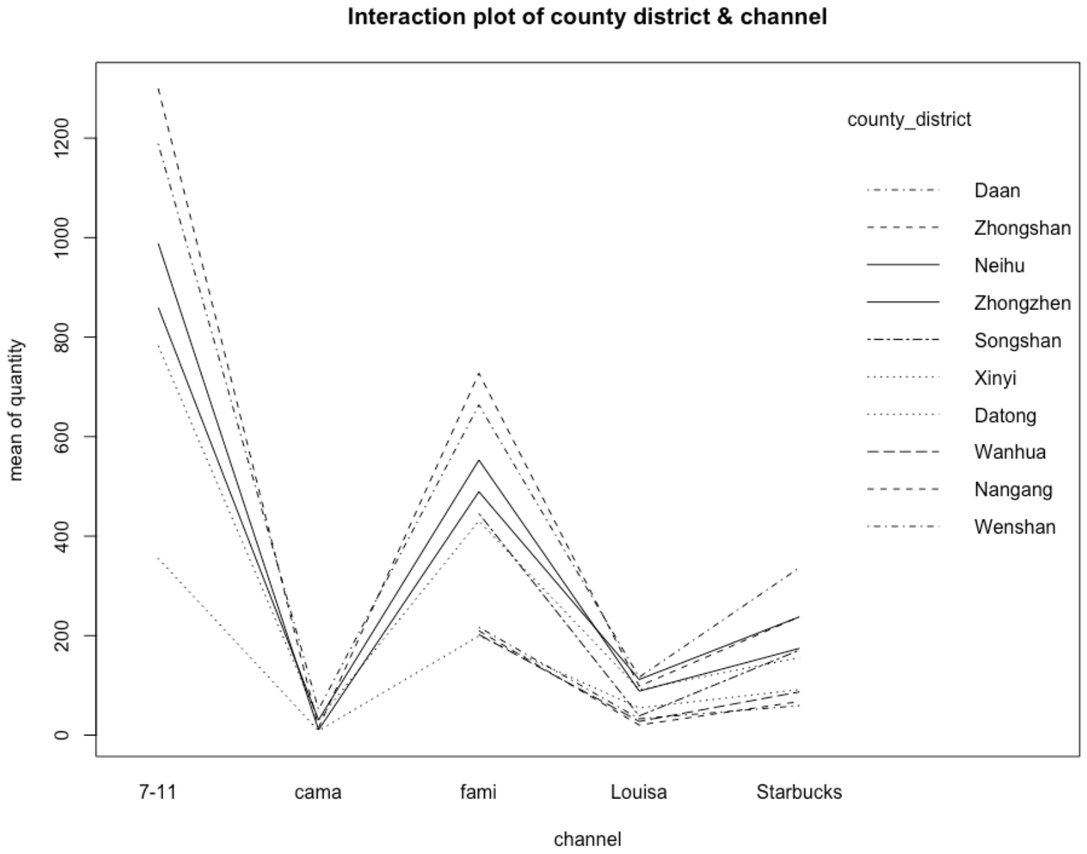
date	county_distric	channel	quant_
2021/4/1	Daan	7-11	1221
2021/4/1	Daan	Louisa	96
2021/4/1	Daan	Starbucks	313
2021/4/1	Daan	cama	47
2021/4/1	Daan	fami	694
2021/4/1	Datong	7-11	329
2021/4/1	Datong	Louisa	46
2021/4/1	Datong	Starbucks	70
2021/4/1	Datong	cama	7
2021/4/1	Datong	fami	186
2021/4/1	Nangang	7-11	406
2021/4/1	Nangang	Louisa	15
2021/4/1	Nangang	Starbucks	56
2021/4/1	Nangang	fami	235

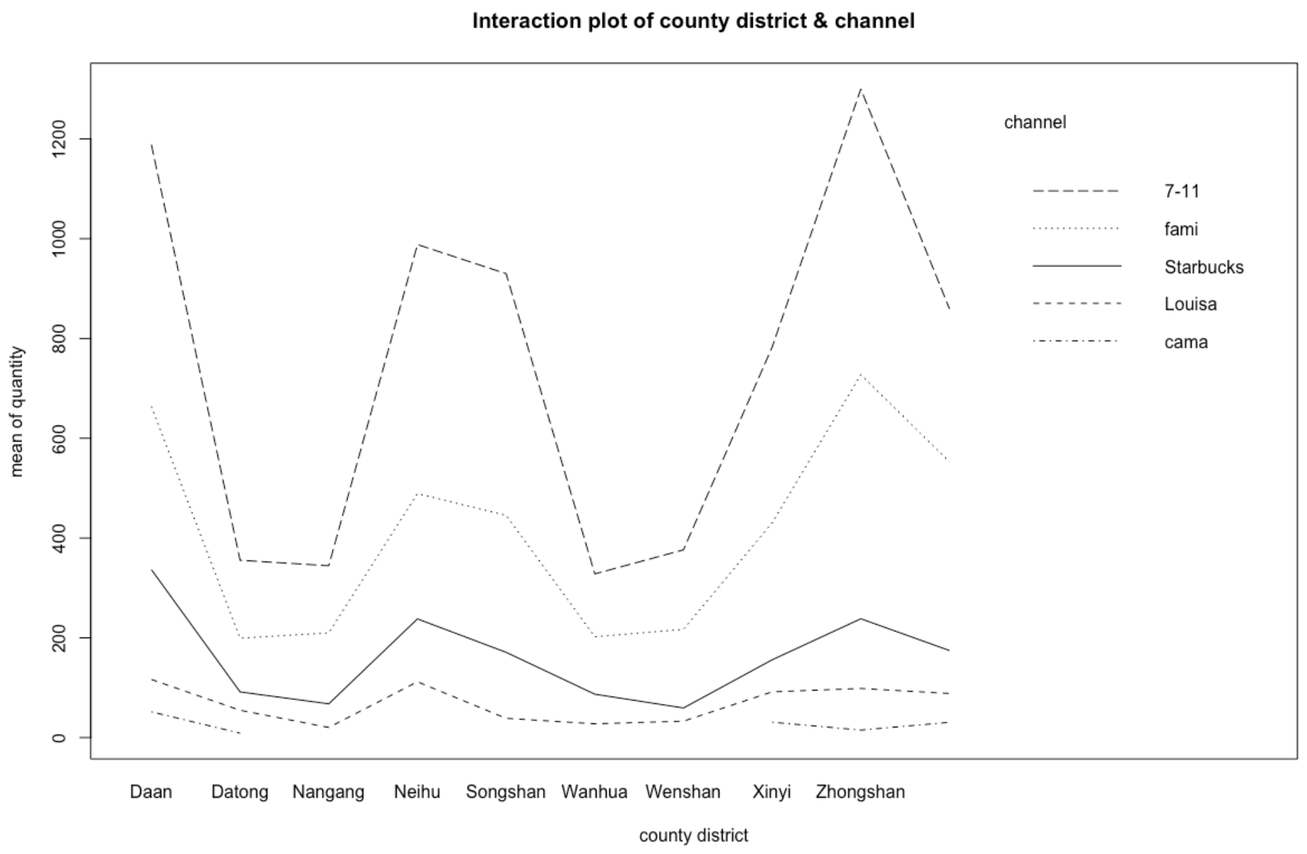




Does interaction exist between channels and Taipei districts?

Interaction plot





Does interaction exist between channels and Taipei districts?

ANOVA

```
Call:
lm(formula = log(quant_) ~ channel * county_district, data = topic5a)

Residuals:
     Min       1Q   Median       3Q      Max
-1.24091 -0.16833  0.03406  0.19775  1.01961

Coefficients: (4 not defined because of singularities)
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                                7.05113    0.04844 145.572  < 2e-16 ***
channelcama                               -3.13357    0.06850 -45.745  < 2e-16 ***
channelfami                               -0.60686    0.06850  -8.859  < 2e-16 ***
channelLouisa                             -2.30166    0.06850 -33.600  < 2e-16 ***
channelStarbucks                          -1.28147    0.06850 -18.707  < 2e-16 ***
county_districtDatong                     -1.20958    0.06850 -17.658  < 2e-16 ***
county_districtNangang                    -1.24361    0.06850 -18.155  < 2e-16 ***
county_districtNeihu                      -0.21184    0.06850  -3.092 0.002014 **
county_districtSongshan                   -0.26376    0.06850  -3.850 0.000122 ***
county_districtWanhua                     -1.31434    0.06850 -19.187  < 2e-16 ***
county_districtWenshan                    -1.15560    0.06850 -16.870  < 2e-16 ***
county_districtXinyi                      -0.41888    0.06850  -6.115 1.17e-09 ***
county_districtZhongshan                   0.06448    0.06850   0.941 0.346649
county_districtZhongzhen                  -0.32239    0.06850  -4.706 2.70e-06 ***
channelcama:county_districtDatong         -0.58714    0.09745  -6.025 2.02e-09 ***
channelfami:county_districtDatong          0.04316    0.09688   0.446 0.655983
channelLouisa:county_districtDatong        0.45176    0.09688   4.663 3.33e-06 ***
channelStarbucks:county_districtDatong    -0.08263    0.09688  -0.853 0.393812
channelcama:county_districtNangang              NA         NA      NA       NA
channelfami:county_districtNangang         0.09083    0.09688   0.938 0.348562
channelLouisa:county_districtNangang      -0.59382    0.09914  -5.990 2.51e-09 ***
channelStarbucks:county_districtNangang   -0.45193    0.09914  -4.559 5.48e-06 ***
channelcama:county_districtNeihu          -1.42291    0.09745 -14.601  < 2e-16 ***
channelfami:county_districtNeihu          -0.14939    0.09716  -1.538 0.124314
channelLouisa:county_districtNeihu         0.14764    0.09688   1.524 0.127680
channelStarbucks:county_districtNeihu     -0.17065    0.09688  -1.762 0.078307 .
channelcama:county_districtSongshan             NA         NA      NA       NA
channelfami:county_districtSongshan       -0.14154    0.09688  -1.461 0.144158
channelLouisa:county_districtSongshan     -0.87263    0.09688  -9.008  < 2e-16 ***
channelStarbucks:county_districtSongshan  -0.44976    0.09688  -4.643 3.67e-06 ***
channelcama:county_districtWanhua               NA         NA      NA       NA
channelfami:county_districtWanhua          0.17237    0.09688   1.779 0.075356 .
channelLouisa:county_districtWanhua       -0.16117    0.13176  -1.223 0.221429
channelStarbucks:county_districtWanhua    -0.04315    0.09688  -0.445 0.656083
channelcama:county_districtWenshan              NA         NA      NA       NA
channelfami:county_districtWenshan         0.08523    0.09688   0.880 0.379056
channelLouisa:county_districtWenshan      -0.13000    0.09688  -1.342 0.179770
channelStarbucks:county_districtWenshan   -0.57824    0.09688  -5.969 2.84e-09 ***
channelcama:county_districtXinyi          -0.08792    0.10295  -0.854 0.393190
channelfami:county_districtXinyi           0.01342    0.09688   0.139 0.889835
channelLouisa:county_districtXinyi         0.17908    0.09688   1.849 0.064672 .
channelStarbucks:county_districtXinyi     -0.42909    0.09716  -4.416 1.06e-05 ***
channelcama:county_districtZhongshan      -1.38938    0.09716 -14.301  < 2e-16 ***
channelfami:county_districtZhongshan       0.03741    0.09745   0.384 0.701130
channelLouisa:county_districtZhongshan    -0.23852    0.09688  -2.462 0.013901 *
channelStarbucks:county_districtZhongshan -0.42015    0.09688  -4.337 1.52e-05 ***
channelcama:county_districtZhongzhen      -0.22033    0.09688  -2.274 0.023054 *
channelfami:county_districtZhongzhen       0.17431    0.09688   1.799 0.072128 .
channelLouisa:county_districtZhongzhen     0.03879    0.09688   0.400 0.688882
channelStarbucks:county_districtZhongzhen -0.33133    0.09688  -3.420 0.000639 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3213 on 1906 degrees of freedom
Multiple R-squared:  0.9423,    Adjusted R-squared:  0.941
F-statistic:   692 on 45 and 1906 DF,  p-value: < 2.2e-16
```

R-squared: 0.9423

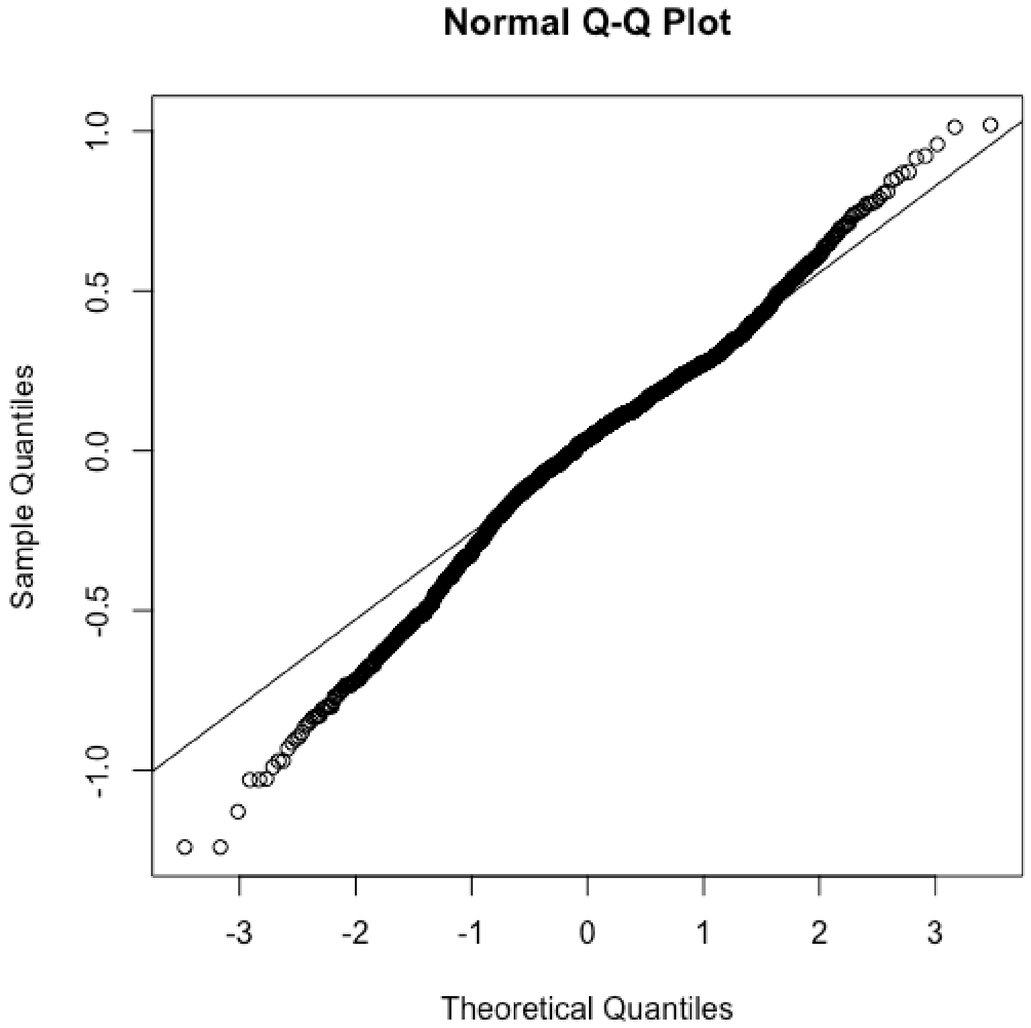
```
> anova(mod2_,mod1_)
Analysis of Variance Table

Model 1: log(quant_) ~ channel + county_district
Model 2: log(quant_) ~ channel * county_district
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1   1938 307.39
2   1906 196.76 32    110.63 33.488 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

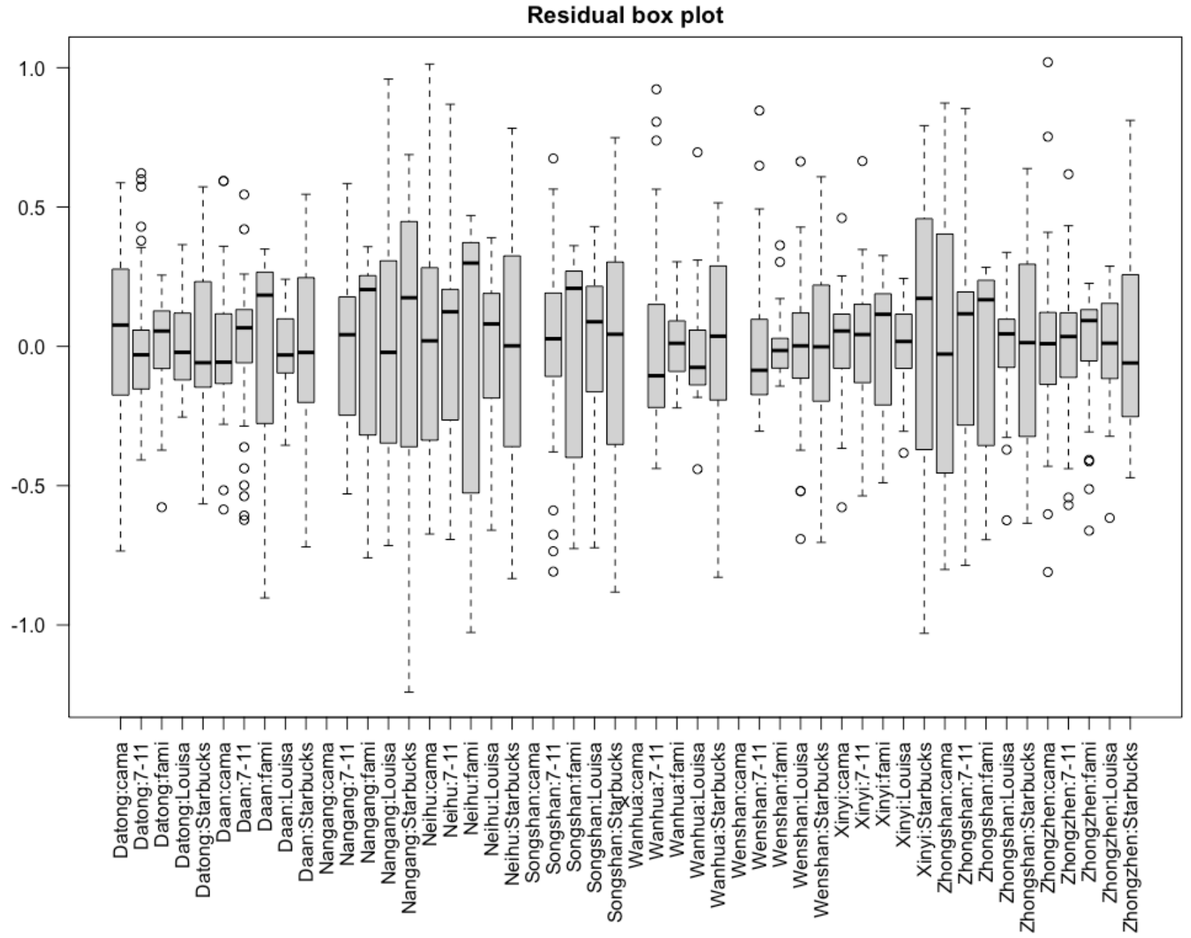
--> lm2 is better

--> interaction exists

Does interaction exist between channels and Taipei districts?

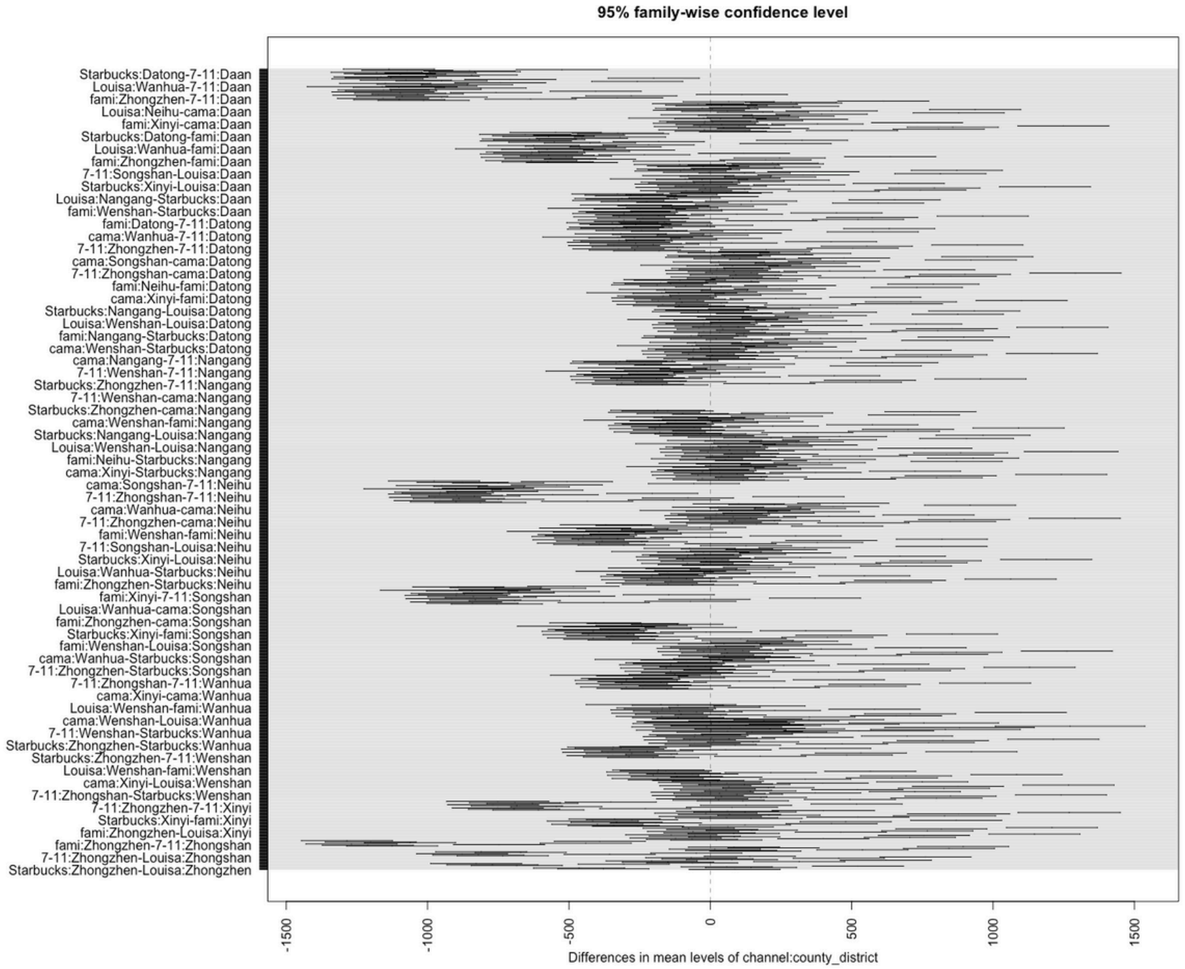
Check Assumptions





Does interaction exist between channels and Taipei districts?

Multiple Comparisons of Mean



Non-sigificant :

7-11:Zhongshan-7-11:Daan
Louisa:Daan-cama:Daan
cama:Datong-cama:Daan
fami:Datong-cama:Daan
Louisa:Datong-cama:Daan
Starbucks:Datong-cama:Daan
fami:Nangang-cama:Daan
Louisa:Nangang-cama:Daan
Starbucks:Nangang-cama:Daan

,etc.

6. Conclusion

Conclusion

1. **What factors affect daily average sales quantity?**
   - We propose a sufficient model, which can explain 93.78% variance of the response variable.
   - log(quantity) ~ (weekday + county + channel * category * unit_price) * is_Alert
2. **Does interaction exist between categories and time intervals?**
   - Factors & interaction: significant.
   - Best-sale combination: latte sold in the morning.
3. **Does interaction exist between channels and time intervals?**
   - Factors & interaction: significant.
   - Best-sale combination: 7-11 coffee in the morning.
4. **Does interaction exist between categories and Taipei districts?**
   - Factors : significant.
   - Best-sale combination: latte sold in Zhongshan.
5. **Does interaction exist between channels and Taipei districts?**
   - Factors & interaction: significant.
   - Best-sale combination: 7-11 coffee sold in Zhongshan.

Thank You