

Which messaging app is the most used and why ?

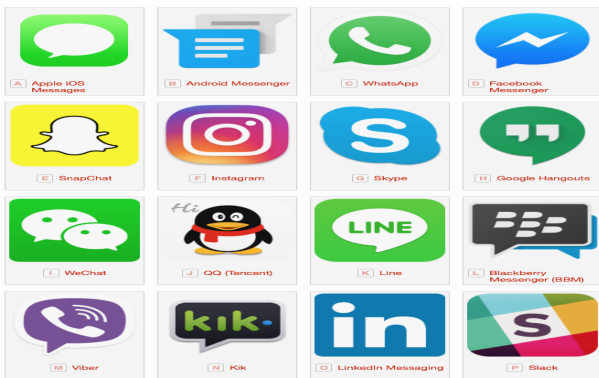
Benjamin Giraudon, Dina Capelle

ENAC,IENAC16

April 23, 2019

- 1 Introduction
 - Purpose
- 2 Variables
 - Binary Variables
 - Category variables
 - Ranking Variables
 - Counting variables
 - Dummy variables
 - Dependent variables
- 3 First model
- 4 Second model
- 5 Third Model
- 6 Fourth Model
- 7 Conclusion

Question of interest



Purpose

- collect data on people's behaviour
- create an Econometric model of this behaviour
- Analyze the answers on Eviews

Data Gathering

Our survey

- Use of a Google form shared on social medias
- n=473 answers collected in 7 days

Étude du marché des applications de messagerie instantanées

Toutes les données récoltées seront supprimées à la fin de notre projet.
Dans le cadre de notre projet en économétrie, on a quelques petites questions à vous poser.
Merci pour votre participation !
Bergamini & Dina

**obligatoire*

Genre *

Homme

Femme

Age *

Votre réponse _____

Nationalité *

Française

Figure: Our survey

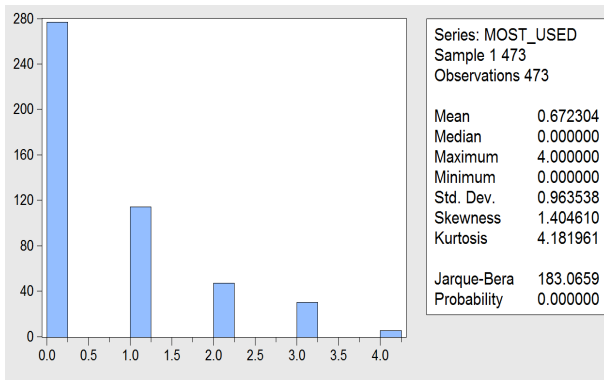
Binary Variables

- GENDER : 64% of women
- NATIONALITY: 96% of French people

Category variables

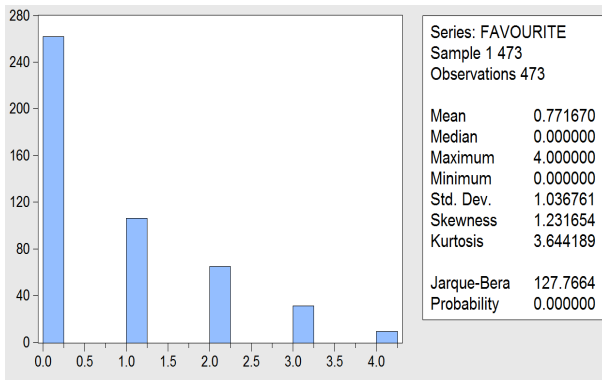
- REGION
- SMARTPHONE
- FAVOURITE
- MOST_USED

Most used app



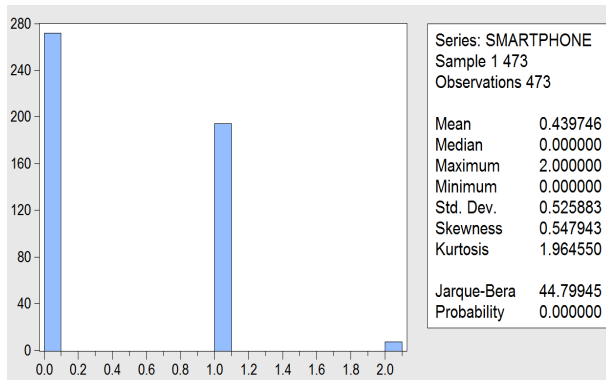
- 0 : Messenger
- 1 : SMS
- 2 : Whatsapp
- 3 : Snapchat
- 4 : Instagram

Favourite



- 0 : Messenger
- 1 : SMS
- 2 : Whatsapp
- 3 : Snapchat
- 4 : Instagram

Smartphone



- 0 :Android
- 1 :Apple
- 2 :Others

Region

Région

287 observations

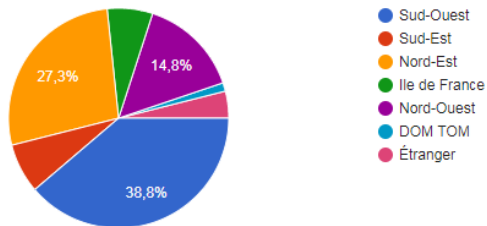


Figure: Geographical distribution

Ranking Variables

- NO_ADS
- DATA_POLICY
- FUNCTIONS
- NETWORK
- EASE_OF_USE

No Ads

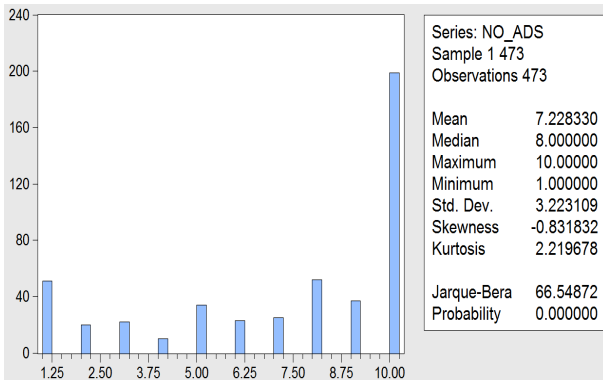


Figure: Majority of 10/10

Data Policy

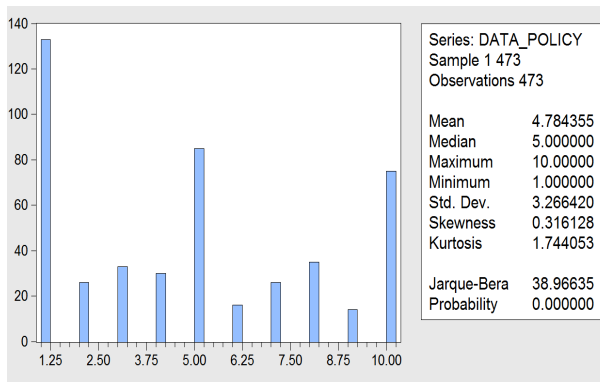


Figure: 1/10, most people disagree or don't care about data policy

Functions

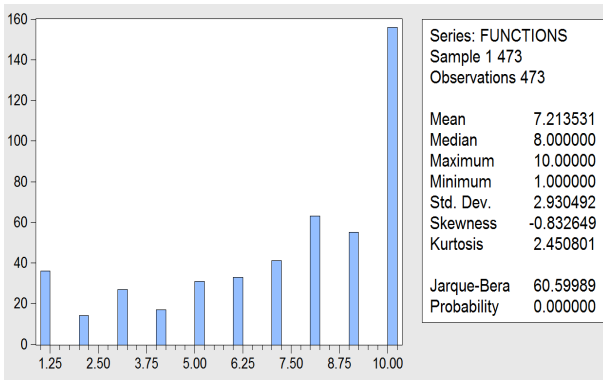


Figure: 10/10

Network

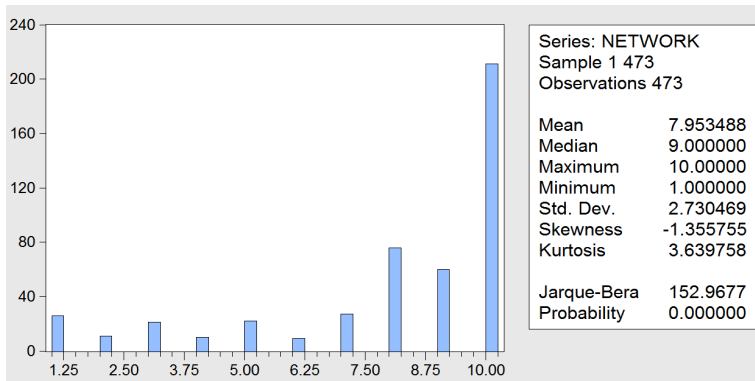


Figure: Most people's favourite app is chosen according to what their network is using

Ease of use

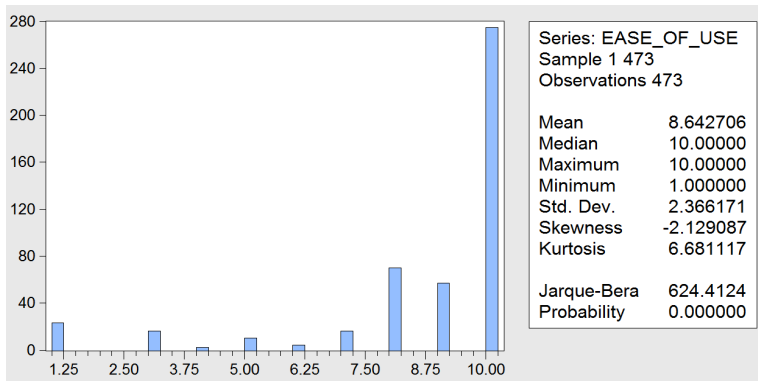


Figure: The ease of use is a determining factor in why an app is our favourite

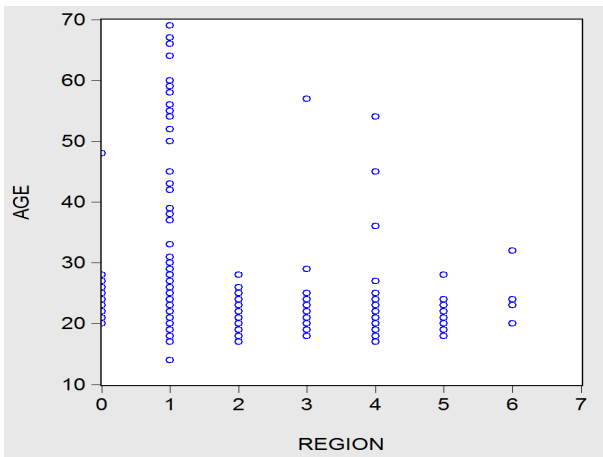
Age

Range: 14-99 years Mean: 23.5 Median:21

Normality for Age ?

$P(\text{JB})=0$ so the distribution isn't normal.

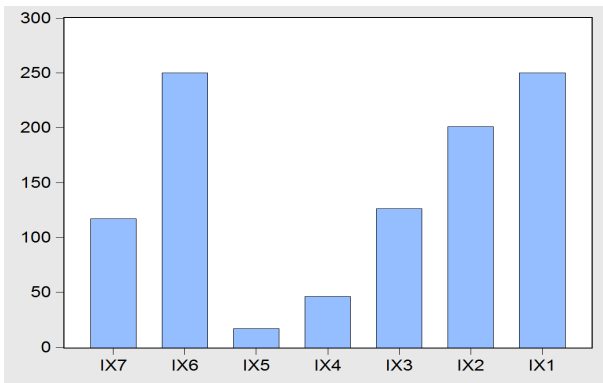
Age vs Region



Dummy variables

- variable $S \times 1$: S stands for Snapchat, 1 stands for "is used for communicating with friends"
- 2 stands for "is used to share media"
- 3 stands for "I know it but I don't use it"
- 4 stands for "communicate within a group"
- 5 stands for "communicate with coworkers"
- 6 stands for "organize events"
- 7 stands for "communicate with people in an other country"

Instagram



Snapchat

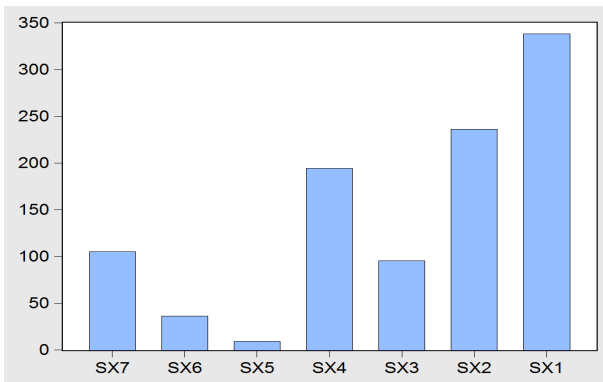
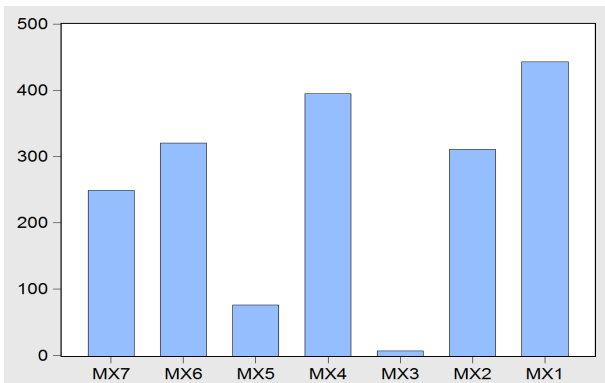
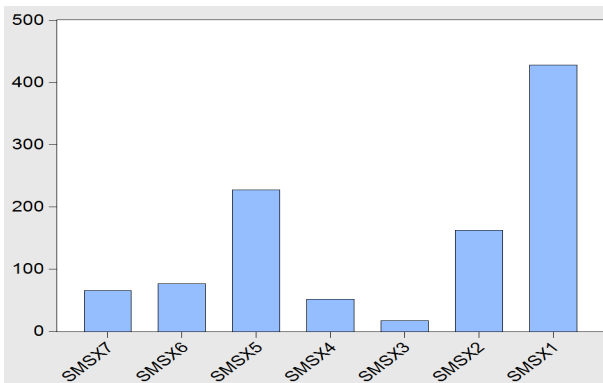


Figure:

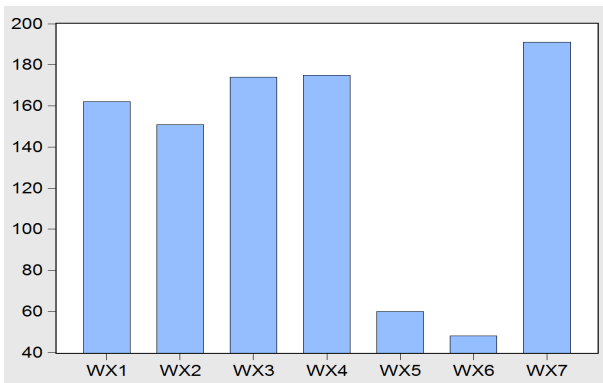
Messenger



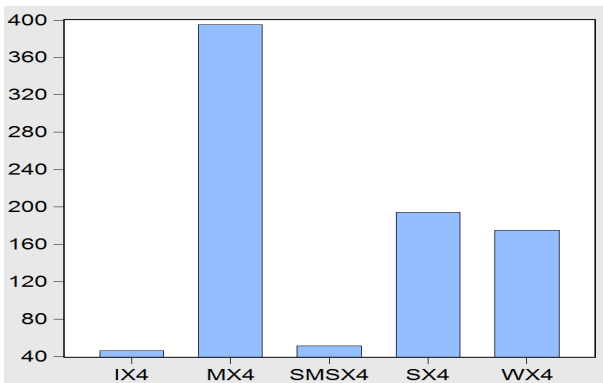
SMS



Whatsapp



GROUPS



PROFESSIONAL

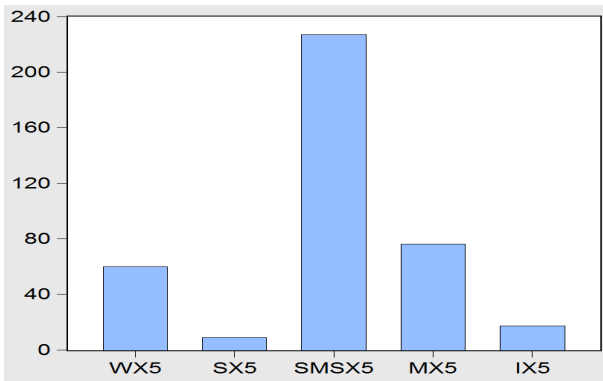


Figure:

FOREIGN

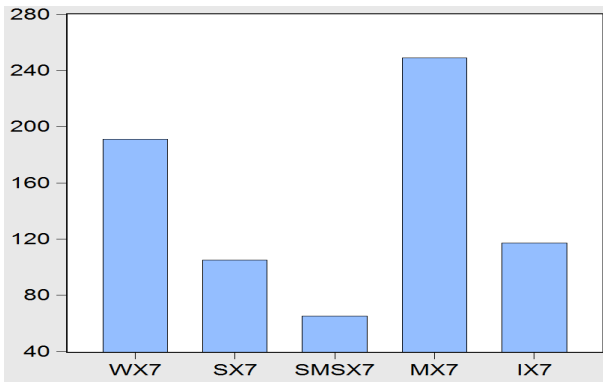


Figure:

What will be our Y ?

Dependent variable

Regarding our topic, we observed that the most relevant variables giving us the best explanatory model are:

- age
- favourite=messenger
- favourite=SMS

First Linear Regression : Age

Dependent Variable: AGE
 Method: Least Squares
 Date: 04/07/19 Time: 17:25
 Sample: 1 473
 Included observations: 473

	Coefficient	Std. Error	t-Statistic	Prob.
C	23.78166	3.047183	7.804472	0.0000
FAV_W	4.178068	2.947490	1.417500	0.1570
FAV_M	0.238993	2.767682	0.086351	0.9312
FAV_S	-1.254032	3.098546	-0.404716	0.6859
FAV_SMS	3.931761	2.924957	1.344212	0.1795
FUNCTIONS	-0.031966	0.168614	-0.189579	0.8497
DATA_POLICY	0.163511	0.141867	1.152564	0.2497
NETWORK	0.092015	0.184392	0.499016	0.6180
EASE_OF_USE	-0.387496	0.245058	-1.581242	0.1145
NO_ADS	0.040117	0.159183	0.252018	0.8011

R-squared	0.075121	Mean dependent var	23.51163
Adjusted R-squared	0.057143	S.D. dependent var	8.347103
S.E. of regression	8.105104	Akaike info criterion	7.043780
Sum squared resid	30415.73	Schwarz criterion	7.131710
Log likelihood	-1655.854	Hannan-Quinn criter.	7.078365
F-statistic	4.178474	Durbin-Watson stat	1.485609
Prob(F-statistic)	0.000033		

Observations

Prob(F-Stat)=0,00003

$\overline{R^2} = 0.057$

$k = 10$

counter – intuitive

Variables	Expected Effect	Observed effect	Significant
FAV_W	?	4.17	No
FAV_M	=	0.24	No
FAV_S	-	-1.25	No
FAV_SMS	+	3.9	No
FUNCTIONS	-	-0.03	No
DATA_POLICY	?	0.16	No
NETWORK	?	0.09	No
EASE_OF_USE	+	-0.39	No
NO_ADS	?	0.04	No

Table: Comparison of real and marginal effects

Second Linear Regression: FAVOURITE=Whatsapp

Dependent Variable: FAV_W
 Method: Least Squares
 Date: 04/07/19 Time: 17:07
 Sample: 1 473
 Included observations: 473

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.137027	0.082004	1.670973	0.0954
GENDER	0.030106	0.028174	1.068569	0.2858
NATIONALITY	-0.153440	0.065696	-2.335589	0.0199
FUNCTIONS	0.007672	0.005437	1.411050	0.1589
DATA_POLICY	-0.000945	0.004769	-0.198120	0.8430
NETWORK	-0.017094	0.006615	-2.584333	0.0101
EASE_OF_USE	-0.012158	0.008733	-1.392177	0.1645
NO_ADS	0.023357	0.005461	4.276810	0.0000
WX1	0.229747	0.033246	6.910495	0.0000
WX2	0.129349	0.035403	3.653591	0.0003
WX5	0.135865	0.044703	3.039282	0.0025
WX7	0.015193	0.030321	0.501067	0.6166
SMARTPHONE_AP	0.014240	0.027886	0.510632	0.6099
R-squared	0.315261	Mean dependent var		0.137421
Adjusted R-squared	0.297399	S.D. dependent var		0.344655
S.E. of regression	0.288895	Akaike info criterion		0.381591
Sum squared resid	38.39169	Schwarz criterion		0.495900
Log likelihood	-77.24625	Hannan-Quinn criter.		0.426551
F-statistic	17.64910	Durbin-Watson stat		0.789084
Prob(F-statistic)	0.000000			

Observations

Prob(F-Stat)=0

$R^2 = 0.30$

$k = 13$

counter – intuitive

Variables	Expected Effect	Observed effect	Significant
GENDER	?	0.03	No
NATIONALITY	-	-0.15	Yes
FUNCTIONS	+	0.007	No
DATA_POLICY	+	-0.001	No
NETWORK	+	-0.02	Yes
EASE_OF_USE	+	-0.01	No
NO_ADS	+	0.02	Yes
WX1	+	0.23	Yes
WX2	+	0.12	Yes
WX5	+	0.13	Yes
WX7	+	0.016	No
SMARTPHONE=APPLE	?	0.014	No

Table: Comparison of real and marginal effects

Third Linear Regression: FAVOURITE=SMS

Dependent Variable: FAV_SMS
 Method: Least Squares
 Date: 04/07/19 Time: 17:07
 Sample: 1 473
 Included observations: 473

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.034158	0.107926	-0.316499	0.7518
GENDER	0.002869	0.033765	0.084979	0.9323
NATIONALITY	0.034868	0.078250	0.445597	0.6561
FUNCTIONS	-0.066906	0.006449	-10.37540	0.0000
DATA_POLICY	0.037056	0.005636	6.574755	0.0000
NETWORK	0.010581	0.007842	1.349284	0.1779
EASE_OF_USE	0.013994	0.010380	1.348230	0.1782
NO_ADS	0.022347	0.006495	3.440792	0.0006
SMSX1	0.117504	0.062599	1.877106	0.0611
SMSX2	0.067693	0.035722	1.894962	0.0587
SMSX5	0.014201	0.034703	0.409220	0.6826
SMSX7	0.108535	0.051798	2.095322	0.0367
SMARTPHONE_AP	0.025783	0.032975	0.781901	0.4347
R-squared	0.340997	Mean dependent var		0.224101
Adjusted R-squared	0.323805	S.D. dependent var		0.417431
S.E. of regression	0.343258	Akaike info criterion		0.726429
Sum squared resid	54.19990	Schwarz criterion		0.840738
Log likelihood	-158.8005	Hannan-Quinn criter.		0.771389
F-statistic	19.83531	Durbin-Watson stat		0.737704
Prob(F-statistic)	0.000000			

Observations

Prob(F-Stat)=0

$R^2 = 0.32$

$k = 13$

*counter – intuitive
 on SMSX.*

Variables	Expected Effect	Observed effect	Significant
GENDER	?	0.003	No
NATIONALITY	?	0.035	No
FUNCTIONS	-	-0.07	Yes
DATA_POLICY	+	0.04	Yes
NETWORK	+	0.01	No
EASE_OF_USE	+	0.01	No
NO_ADS	+	0.02	Yes
SMSX1	+	0.12	No
SMSX2	+	0.07	No
SMSX5	+	0.01	No
SMSX7	-	0.11	Yes
SMARTPHONE=APPLE	?	0.03	No

Table: Comparison of real and marginal effects

Fourth Linear Regression: FAVOURITE=MESSENGER

Dependent Variable: FAV_M
 Method: Least Squares
 Date: 04/07/19 Time: 17:08
 Sample: 1 473
 Included observations: 473

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.268750	0.134563	1.997202	0.0464
GENDER	-0.061123	0.039905	-1.531718	0.1263
NATIONALITY	0.142533	0.092709	1.537433	0.1249
FUNCTIONS	0.051172	0.007657	6.682907	0.0000
DATA_POLICY	-0.049123	0.006680	-7.353868	0.0000
NETWORK	0.004300	0.009362	0.459295	0.6462
EASE_OF_USE	-0.002334	0.012356	-0.188934	0.8502
NO_ADS	-0.029988	0.007706	-3.891245	0.0001
MX1	0.204375	0.078685	2.597379	0.0097
MX2	0.118582	0.044125	2.687444	0.0075
MX5	0.046715	0.051996	0.898438	0.3694
MX7	0.049552	0.040725	1.216771	0.2243
SMARTPHONE_AP	-0.113033	0.040067	-2.821069	0.0050
R-squared	0.346889	Mean dependent var		0.553911
Adjusted R-squared	0.329851	S.D. dependent var		0.497611
S.E. of regression	0.407358	Akaike info criterion		1.068849
Sum squared resid	76.33253	Schwarz criterion		1.183158
Log likelihood	-239.7828	Hannan-Quinn criter.		1.113809
F-statistic	20.36010	Durbin-Watson stat		0.674113
Prob(F-statistic)	0.000000			

Observations

Prob(F-Stat)=0

$\overline{R^2} = 0.33$

$k = 13$

counter – intuitive

Variables	Expected Effect	Observed effect	Significant
GENDER	=	-0.06	No
NATIONALITY	+	0.14	No
FUNCTIONS	+	0.05	Yes
DATA_POLICY	-	-0.05	Yes
NETWORK	+	0.004	No
EASE_OF_USE	+	-0.002	No
NO_ADS	-	-0.03	Yes
MX1	+	0.2	Yes
MX2	+	0.12	Yes
MX5	-	0.05	No
MX7	+	0.05	No
SMARTPHONE=APPLE	?	-0.11	Yes

Table: Comparison of real and marginal effects

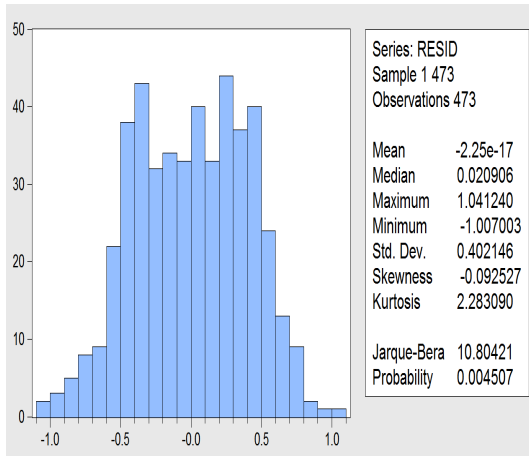
Normality of the residuals ?

Block 2

$$JB = 10.8 > \chi_{0.95}^2(2) = 5.99$$

rH_0

The classical assumption fails with this set of data.



Conclusion

Positive points

- Lots of answers, descriptive analysis of our data is relevant.

Mixed points

- Marginal effects are often counter-intuitive
- Hard to achieve high R^2

Limits

- Fairly homogeneous population in age and nationality
- Majority of category data: linear analysis are limited