Why do people buy shoes?

And what are the differences beetween boys and girls?

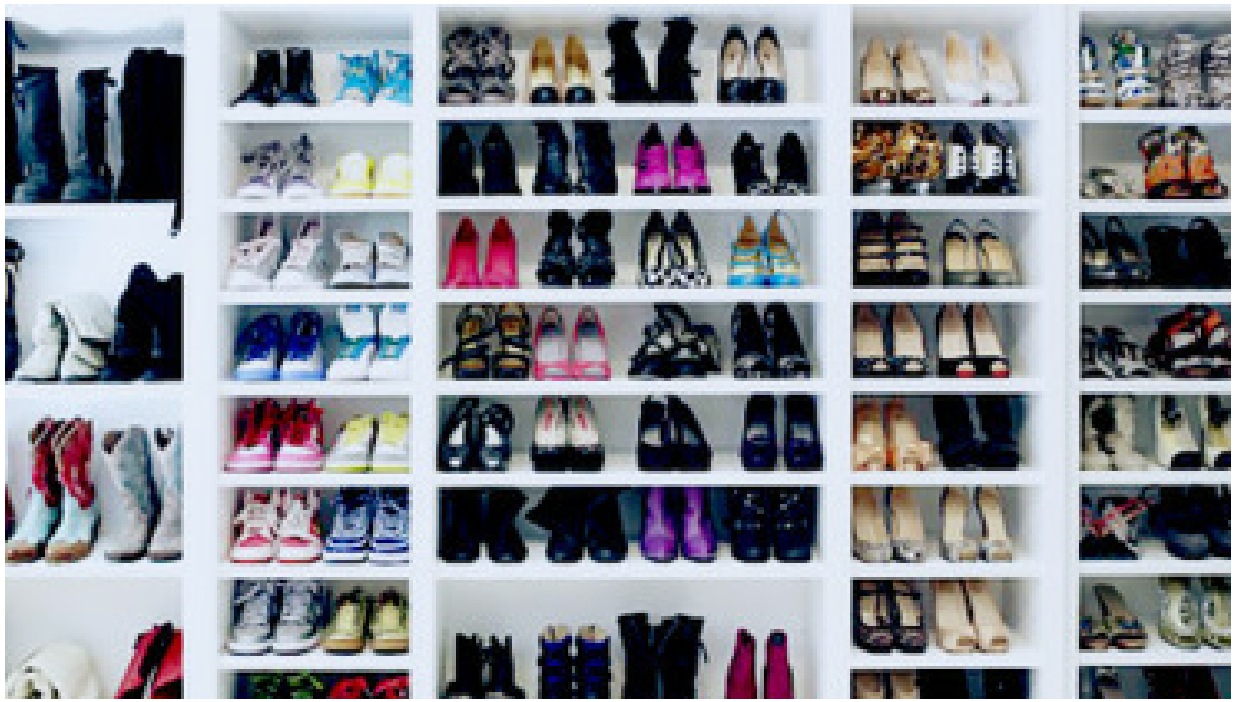
Bérénice Mallet - Floriane Mounié

Econometrics : Application Project

May 21st 2014

Introduction

What individual characteristics can influence our shoes purchases ?



We had to :

1. Choose some individual characteristics that might influence shoe purchase

We had to :

1. Choose some individual characteristics that might influence shoe purchase
2. Realise a survey

We had to :

1. Choose some individual characteristics that might influence shoe purchase
2. Realise a survey
3. Create a data base for Eviews with the survey answers

Explanatory Variables	Expected Effect
Age	=
Friends	+
Store	=
Budget	+
Shoe Budget	+
Couple	-
Job	+
Children	-
Facebook	+
Pairs	+
Sex	-
Sales	+
Events	+
Clothe Budget	+
Zone	+

First model

A linear model : First Equation

At first, we tried to modelise shoe purchase with linear dependance on all other variables.

EQU 01

$$\widehat{purchase}_i = \widehat{\beta_0} + \widehat{\beta_1} age_i + \widehat{\beta_2} friends_i + \widehat{\beta_3} store_i + \widehat{\beta_4} budget_i$$
$$+\widehat{\beta_5} shoebudget_i + \widehat{\beta_6} couple_i + \widehat{\beta_7} job_i + \widehat{\beta_8} children_i$$
$$+\widehat{\beta_9} facebook_i + \widehat{\beta_{10}} pairs_i + \widehat{\beta_{11}} sex_i + \widehat{\beta_{12}} sales_i + \widehat{\beta_{13}} events_i$$
$$+\widehat{\beta_{14}} clothebudget_i + \widehat{\beta_{15}} zone_i$$

First model
Linear Regression Output

Dependent Variable: ACHAT
Method: Least Squares
Date: 05/15/14 Time: 18:26
Sample: 1 135
Included observations: 135

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.975966	1.018889	0.957872	0.3401
AGE	-0.016388	0.018550	-0.883472	0.3788
AMIS	-0.000392	0.000758	-0.516824	0.6062
BOUTIQUE	-0.014009	0.160495	-0.087288	0.9306
BUDGET	-6.17E-06	0.000181	-0.034036	0.9729
CHAU	0.004085	0.002115	1.931943	0.0557
COUPLE	0.295366	0.188242	1.569075	0.1193
EMPLOI	0.261512	0.127625	2.049070	0.0427
ENFANTS	0.075008	0.353613	0.212120	0.8324
FB	-0.468879	0.559881	-0.837461	0.4040
PAIRES	0.068360	0.013974	4.891901	0.0000
SEXE	-0.361318	0.221881	-1.628432	0.1061
SOLDES	-0.001053	0.003291	-0.320047	0.7495
SORTIES	0.061448	0.025100	2.448109	0.0158
VTMT	0.000759	0.000890	0.852034	0.3959
ZONE	0.029212	0.105024	0.278142	0.7814

R-squared	0.561912	Mean dependent var	2.533333
Adjusted R-squared	0.506690	S.D. dependent var	1.381109
S.E. of regression	0.970036	Akaike info criterion	2.887918
Sum squared resid	111.9754	Schwarz criterion	3.232247
Log likelihood	-178.9345	Hannan-Quinn criter.	3.027844
F-statistic	10.17565	Durbin-Watson stat	2.193371
Prob(F-statistic)	0.000000		

First model
Real effects

Explanatory Variables	Expected Effect	Real Effect
Age	=	-
Friends	+	-
Store	+	-
Budget	+	=
Shoe Budget	+	+
Couple	-	+
Job	+	+
Children	-	+
Facebook	+	-
Pairs	+	+
Sex	-	-
Sales	+	-
Events	+	+
Clothe Budget	+	+
Zone	+	+

First model
R-squared

- For this estimation, **R-squared** has a value of 0.562
  We will try to improve this by not taking into account some redundant characteristics.
- The **correlation matrix** reveals that some characteristics are strongly correlated, and so redundant.

First model
Correlation matrix

Correla

	AGE	AMIS	BOUTIQUE	BUDGET	CHAU	COUPLE	EMPLO
AGE	1.000000	-0.502833	0.230535	0.739609	0.149932	0.261884	0.55075
AMIS	-0.502833	1.000000	0.048551	-0.375701	-0.004852	-0.071572	-0.38041
BOUTIQUE	0.230535	0.048551	1.000000	0.168861	-0.200917	0.056331	0.23002
BUDGET	0.739609	-0.375701	0.168861	1.000000	0.240077	0.281092	0.47481
CHAU	0.149932	-0.004852	-0.200917	0.240077	1.000000	0.113071	0.13925
COUPLE	0.261884	-0.071572	0.056331	0.281092	0.113071	1.000000	0.32812
EMPLOI	0.550752	-0.380410	0.230028	0.474810	0.139251	0.328120	1.00000
ENFANTS	0.751718	-0.474881	0.104016	0.633254	0.219402	0.254197	0.48914
FB	-0.560417	0.346031	-0.168269	-0.404142	-0.020282	-0.188982	-0.35484
OBS	0.221726	-0.153223	0.076745	0.242184	-0.029634	0.132386	0.11920
PAIRES	0.079938	-0.000321	-0.172190	0.082315	0.518435	0.170136	0.15871
SEXE	-0.042192	0.187767	0.165432	0.065677	-0.152017	-0.158486	-0.10683
SORTIES	-0.332444	0.462161	-0.170211	-0.232827	-0.026621	0.001633	-0.30281
SOLDES	-0.040593	-0.097750	-0.003311	0.055278	-0.114886	0.141864	0.03466
ZONE	-0.154952	0.187655	-0.209646	-0.049305	0.019309	0.000608	-0.32807
VTMT	0.077556	0.004444	-0.211521	0.155962	0.707960	0.132196	0.08250

First model
Correlated variables

Two variables were removed from the model :

- Age was to much correlated to budget
- Children was also correlated to budget

First model
Wald Test

Some variables might not have a significant influence on shoe purchase. We performed a **Wald test** on the nullity hypothesis of the β_k.

Wald Test:
Equation: EQ04

Test Statistic	Value	df	Probability
F-statistic	0.760444	(8, 121)	0.6381
Chi-square	6.083555	8	0.6379

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	-0.000251	0.000730
C(3)	-0.036098	0.157781
C(4)	-9.35E-05	0.000146
C(7)	0.245513	0.124981
C(8)	-0.288791	0.519596
C(11)	-0.000358	0.003182
C(13)	0.000734	0.000883
C(14)	0.031434	0.104449

Restrictions are linear in coefficients.

First model

Insignificant Variable

A high p-value means that the nullity hypothesis is not rejected. Four variables are removed of the model :

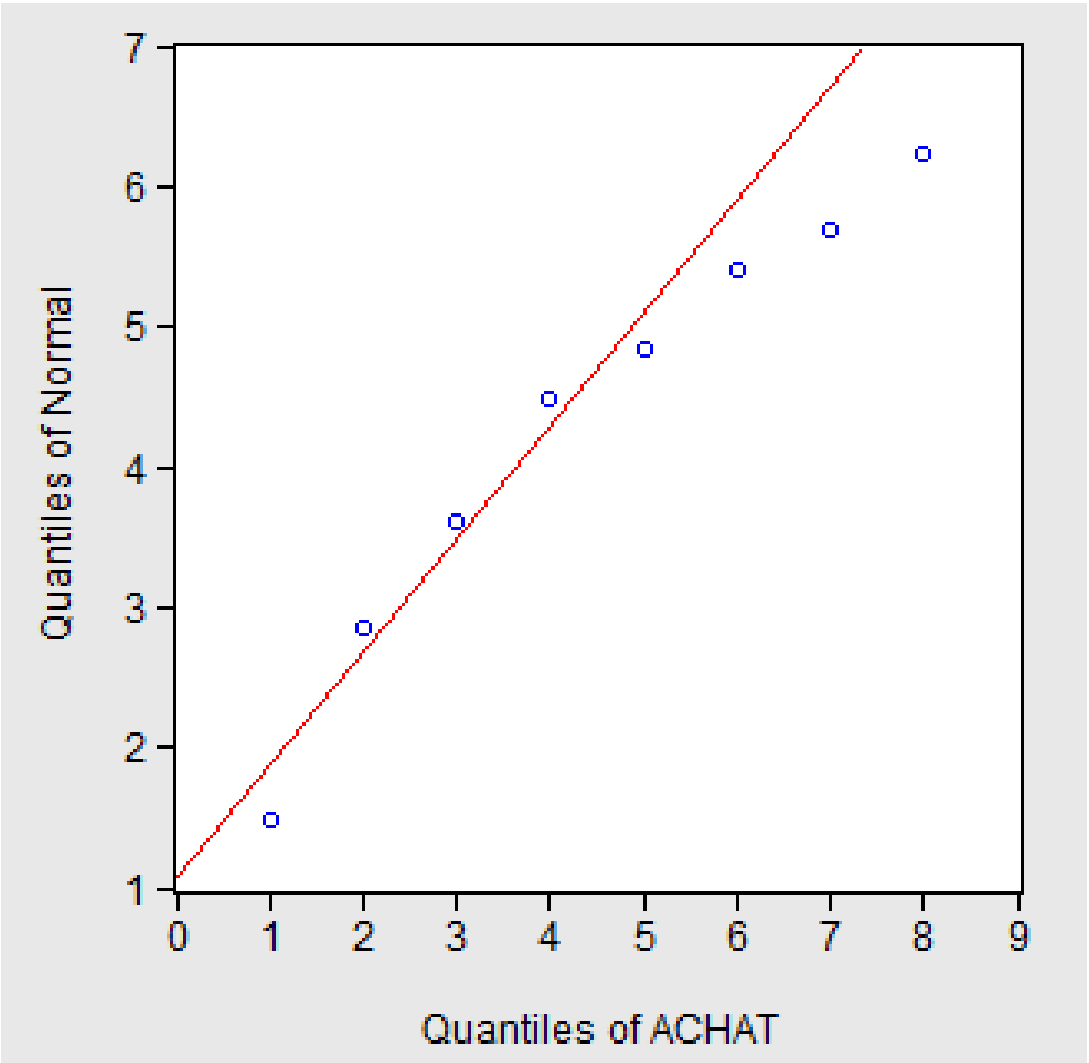
- Store
- Facebook
- Zone

Here is the new linear equation :

EQU 02

$$\widehat{purchase}_i = 0.158 - 0.0004 friends_i - 0.0001 budget_i$$
$$+0.004 shoebudget_i + 0.304 couple_i + 0.236 job + 0.069 pairs_i$$
$$-0.0337 sex_i - 0.0001_i sales + 0.063 events_i$$
$$+0.0006 clothebudget_i$$

First model
Normality, Quantile Quantile Plot



Women's Model
Distinct Population

One of the reasons our model is not exactly accurate migth be that boys and girls have a very different behavior for shoes purchase.

Women's Model
Distinct Population

One of the reasons our model is not exactly accurate migth be that boys and girls have a very different behavior for shoes purchase.

There are two different populations, so we need to build a different model for women !

Women's Model

Linear Regression

Using the same variable, another linear regression was performed focusing on women population. Using the same methods,

- The variables Age, Children and Clothebudget are removed because of correlation.
- Store, Facebook, Couple and Zone are removed because they are insgnificant.

Women's Model
Women's Equation

Here is the final linear equation :

EQU 03

$$\widehat{purchase}_i = 0.0187 - 0.0011 friends_i - 0.00001 budget_i$$
$$+0.0083 shoebudget_i + 0.524 job + 0.0509 pairs_i$$
$$-0.0072 sales + 0.0917 events_i$$

A few remarks :

- Of course, women buy a lot more shoes than men
- Girls are less sensible than boys to budget constraint
- They buy a higher proportion of their shoes during sales.

Women's Model
Heteroscedasticity

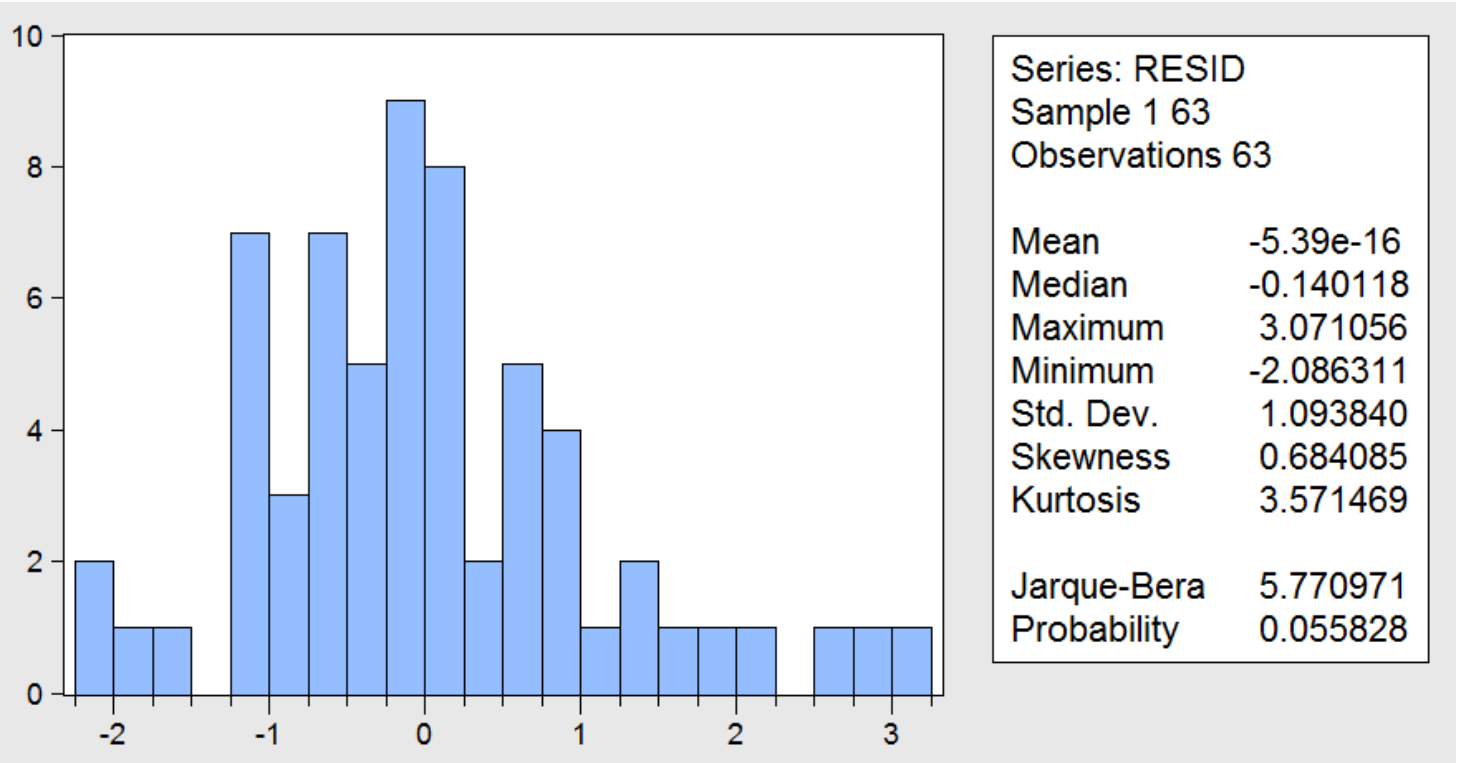
The **White's test** result is that we do reject the heteroscedasticity hypothesis. ($p > 0.05$)
The model can be considered homoscedastic.

Heteroskedasticity Test: White

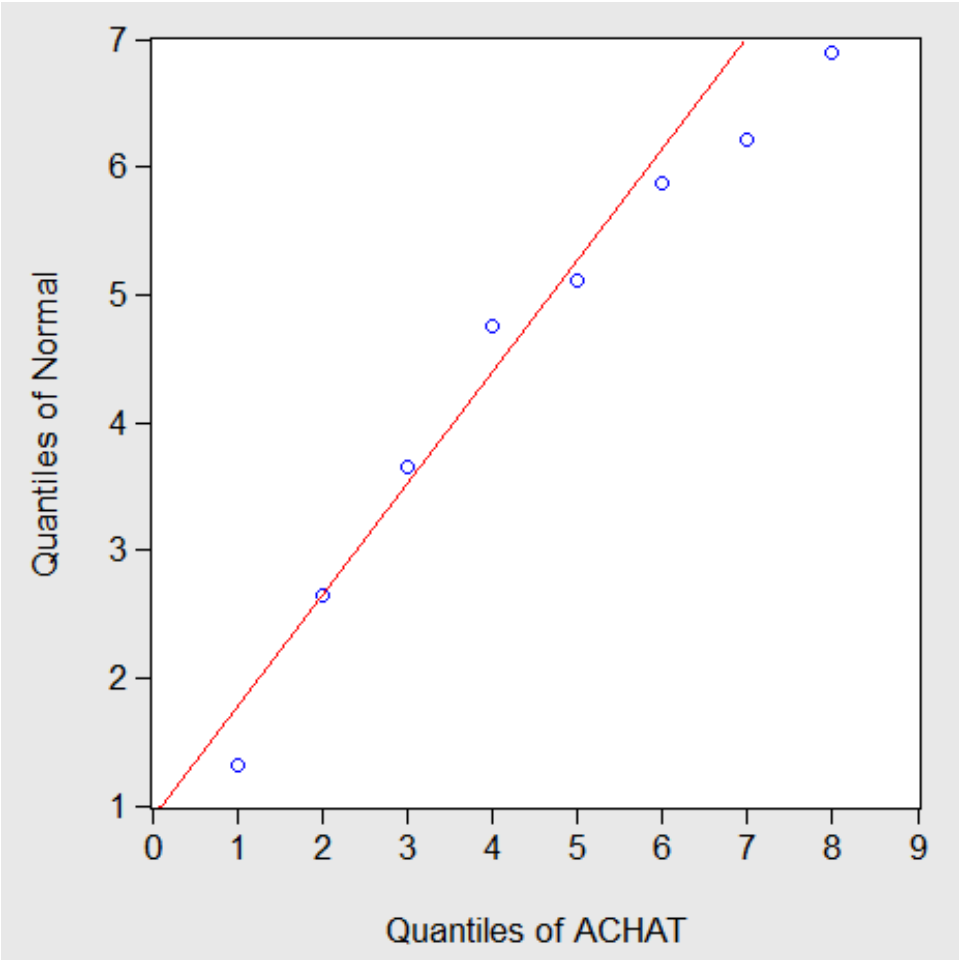
F-statistic	1.354587	Prob. F(44,18)	0.2458
Obs*R-squared	48.38694	Prob. Chi-Square(44)	0.3003
Scaled explained SS	45.70734	Prob. Chi-Square(44)	0.4011

Women's Model
Error

The **Jarque-Bera's test** result ($p > 5$ % and Stat < 6) indicates that the residual error can be identified as following a **normal law**.



Comparison with normal law :



Conclusion

The final model reveals that the more influcial charateristics on women's purchase are :

- her budget for shoes
- wether she has a job or not
- the number of pairs she already owns
- the number of activities an leisures she have

But more surprisingly neither the place of living nor the being in a relationship have an impact on shoe purchase.

Thank you for your attention !