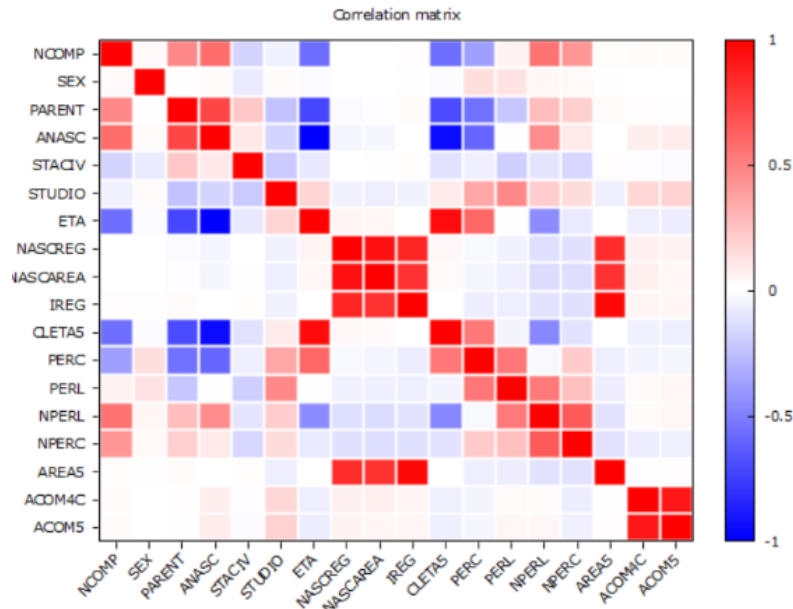


## TAKE HOME (ALBERTO SARTINI)

I built a logistic regression model for LFP as dependent variable, that's because LFP is a dichotomous variable.

I dropped all the variables with too many NAs, then I plotted the matrix of correlations and decided to drop the most correlated variables.



Subsequently I transformed the variable STACIV into dummy variables where:

1=MARRIED

2=SINGLE

3=DIVORCED

Forcing me to drop 4=WIDOW

The current list of explanatory variables would be: NCOMP, MARRIED, SINGLE, DIVORCED, STUDIO, ETA, PERC, PERL, NPERC, NPERL, AREA5, ACOM4C.

The following is the stepwise logistic regression of the full sample

	coefficient	std. error	z	p-value
const	-4.91819	0.213030	-23.09	6.28e-118 ***
NCOMP	0.279703	0.0377143	7.416	1.20e-013 ***
SEX	0.501522	0.0768930	6.522	6.92e-011 ***
SINGLE	0.660727	0.0955306	6.916	4.63e-012 ***
DIVORCED	1.29650	0.191542	6.769	1.30e-011 ***
STUDIO	0.383492	0.0220102	17.42	5.48e-068 ***
PERC	-1.78225	0.122624	-14.53	7.35e-048 ***
PERL	8.03723	0.160316	50.13	0.0000 ***
NPERL	-0.732357	0.0536499	-13.65	2.00e-042 ***
AREA5	0.297719	0.0292303	10.19	2.31e-024 ***
ACOM4C	-0.0983075	0.0399910	-2.458	0.0140 **
Mean dependent var	0.436107	S.D. dependent var	0.495918	
McFadden R-squared	0.720860	Adjusted R-squared	0.719749	
Log-likelihood	-2762.070	Akaike criterion	5546.139	
Schwarz criterion	5629.499	Hannan-Quinn	5573.848	
Number of cases 'correctly predicted' = 13560 (93.9%)				
f(beta'x) at mean of independent vars = 0.248				
Likelihood ratio test: Chi-square(10) = 14265.7 [0.0000]				

As we can see the McFadden R-squared is 0.72, this can lead us to say that the model is decent.

PERL has a value of 8, this means that if a person has work income than the probability to participate to the labour force increases by a factor of  $e^8$  which is a lot and that makes sense.

If we look at SEX we can see that if you are a male it's likely that you participate to the labour force; also STUDIO has a positive coefficient, the higher is your educational level the higher are the chances

Now let's see the two models for the male (left) and female (right) subsets

	coefficient	std. error	z	p-value		coefficient	std. error	z	p-value
const	-4.94273	0.606603	-8.148	3.69e-016 ***	const	-4.70157	0.455802	-10.31	6.03e-025 ***
NCOMP	0.275493	0.0553792	4.975	6.54e-07 ***	NCOMP	0.195437	0.0616642	3.169	0.0015 ***
MARRIED	1.02655	0.462666	2.219	0.0265 **	SINGLE	0.714097	0.205985	3.467	0.0005 ***
SINGLE	1.17101	0.494437	2.368	0.0179 **	DIVORCED	1.08647	0.288663	3.764	0.0002 ***
DIVORCED	2.29068	0.527701	4.341	1.42e-05 ***	STUDIO	0.514425	0.0349157	14.73	3.94e-049 ***
STUDIO	0.264091	0.0318130	8.301	1.03e-016 ***	ETA	-0.0205433	0.00560988	-3.662	0.0003 ***
ETA	0.0219369	0.00514620	4.263	2.02e-05 ***	PERC	-1.08012	0.192724	-5.604	2.09e-08 ***
PERC	-3.27263	0.211230	-15.49	3.85e-054 ***	PERL	8.22385	0.266931	30.81	1.99e-208 ***
PERL	8.32950	0.216198	38.53	0.0000 ***	NPERL	-0.681955	0.0838510	-8.133	4.19e-016 ***
NPERL	-0.756283	0.0713271	-10.60	2.89e-026 ***	AREAS	0.248797	0.0443713	5.607	2.06e-08 ***
AREAS	0.333898	0.0398626	8.376	5.47e-017 ***					
ACOM4C	-0.117134	0.0545795	-2.146	0.0319 **					
Mean dependent var	0.497871	S.D. dependent var	0.500030		Mean dependent var	0.373343	S.D. dependent var	0.483726	
McFadden R-squared	0.706396	Adjusted R-squared	0.704018		McFadden R-squared	0.749030	Adjusted R-squared	0.746918	
Log-likelihood	-1481.742	Akaike criterion	2987.484		Log-likelihood	-1188.090	Akaike criterion	2396.181	
Schwarz criterion	3070.200	Hannan-Quinn	3015.930		Schwarz criterion	2464.951	Hannan-Quinn	2419.850	

Both of them do a great job in prediction, it's interesting that in both models PERL maintains the same "big" magnitude and ETA doesn't seem to play a big role, that's because the LFP ranges from 15 y.o. to 65 y.o. PERC has a negative coefficient in both samples, that's because if you receive a subsidy you don't need to work.

The stepwise regression maintains the MARRIED variable for the male subset but drops it for the female, I think that this happens because marriage plays a more important role in participating in the labour force when you are a man, what I mean is that if a man is married it is highly likely that he works in order to be able to support the family but the same thing is usually not true for the woman ( I hope my comment is not misinterpreted, I do not want to sound sexist at all).

It's interesting to notice that STUDIO coefficient value in the female sample is twice as big as the one in the male sample and the p-value is way smaller; this could mean that education plays a more important role in labour participation for women, in other words it may seem that it would be harder for a woman to have a job if she doesn't reach a certain educational level.

Lastly, let's take a look to the interaction model.

	coefficient	std. error	z	slope
const	-4.71192	0.414843	-11.36	
NCOMP	0.240277	0.0411551	5.838	0.0594273
MARRIED	0.813391	0.323059	2.518	0.198109
SINGLE	0.915032	0.318930	2.869	0.224491
DIVORCED	2.04193	0.408665	4.997	0.425153
STUDIO	0.275588	0.0259673	9.514	0.0681607
ETA	0.0209002	0.00459995	4.544	0.00516921
PERC	-3.26666	0.188982	-17.29	-0.657530
PERL	8.27824	0.168333	49.18	0.961290
NPERL	-0.727175	0.0543178	-13.39	-0.179850
AREAS	0.354807	0.0351885	10.08	0.0877536
ACOM4C	-0.106102	0.0409449	-2.591	-0.0262421
MARRIED_SEX_0	-0.661125	0.243774	-2.712	-0.158556
DIVORCED_SEX_0	-0.767861	0.430028	-1.786	-0.175347
STUDIO_SEX_0	0.228678	0.0382492	5.979	0.0565585
ETA_SEX_0	-0.0406718	0.00582846	-6.978	-0.0100593
PERC_SEX_0	2.20430	0.220225	10.01	0.497219
AREAS_SEX_0	-0.129307	0.0428082	-3.021	-0.0319811
Mean dependent var	0.436107	S.D. dependent var	0.495918	
McFadden R-squared	0.730141	Adjusted R-squared	0.728322	
Log-likelihood	-2670.238	Akaike criterion	5376.475	
Schwarz criterion	5512.883	Hannan-Quinn	5421.817	

Number of cases 'correctly predicted' = 13611 (94.2%)  
f(beta\*x) at mean of independent vars = 0.247  
Likelihood ratio test: Chi-square(17) = 14449.4 [0.0000]

We can confirm what I said above, It could be likely that a married woman doesn't work and it's also likely that a divorced woman doesn't work because she may receive an income from her ex-husband but, again I am generalizing. STUDIO maintains a positive effect while AREA5 has a negative coefficient, this means that a woman living in the south has less chance to work.