Assessing the structure of Health Care System in UK using Statistical Modelling Methods

Author

Carlos Utrilla Guerrero, MSc Manchester University, SRMS.

Supervisor assignment

Prof. Dr Mark Elliot
Cathie Marsh Institute for Social Research
Manchester University
E-mail: mark.elliot@manchester.ac.uk

Course

Introduction to Statistical Modelling

Introduction

In recent year, there have been numerous attempts to define, measure, and analyze illness in various context. The vast majority literature on the concept of health inequalities, indicate a lack of agreement about the conceptualization. Thus, health inequalities can be used to draw: (1) health differences between individuals, (2) health differences between population groups and (3) health differences between those occupying unequal positions in the dominant social hierarchies. Previous literature as a reference suggest that both levels – health inequalities as individual level, what matter the varies of health between individuals- and – health differences among social groups as a systematic differences in the health of groups occupying unequal positions in society – are important concept of measuring population health.

The research presented here, builds on this type of work from geographical dimension to individual approach. To explore health variability across individuals and of different population groups categorized a priori by their relative social status, typically reflected by measure of social class or material/social deprivation. The basis for using social and economic factors as categorizing variables is the overwhelming evidence of their important as health determinants across diverse health outcomes. (Berkman L, Kawachi I, 2000).

Most of the statistical models that attempted to measure illness has been developed using individual-level microdata to make inferences. However, it should be noted that individuals live in households that reside within neighborhood and communities, might be subject to the influences of grouping.

The research is set out as follows. In the section how we measure illness and Determinant? We briefly review the research that has been conducted and highlighter some key points from this studies. In addition, we discuss the theoretical reasons and relevant empirical works in the field of illness measure. This section will release the justification of the variables added on the models. The section Data and Method attempt to present the data, variables, and method that we used in order to build different statistical modeling of healthy variability in Britain. The final section further discusses and summarizes the finding of our results.

How we measure illness and Determinants?

The influence of socio-economic status and area characteristics on healthy variations has been the subject of investigation during the last 30 years. Previous studies have created an ongoing debate whether studying social inequalities in health inherently prejudges causality, whether it produces mask intra-group variation and does not allow for scientific inquiry into other determinants of health. However, Murray et al. proposed new approaches to illustrate such inequalities based on healthy inequalities comparison across geographical regions but not social group. Indeed, geographical comparisons in that involve a priori selection of a categorizing variables base on knowledge indicating its likely relevance. (Kaplan GA,1996) Even so, the prevalence of these geographical variations in relation to socioeconomic and other social conditions can provide important insights into population distributions of health. Moreover, the majority of analysis reports show an axiomatic relation between certain types of social categorization and health inequalities. The author arguments is that social groups and health gradients according to these groupings are more important because groups at the bottom of the social gradient have disadvantages in other spheres of well-being such as income, wealth or education. The interest to study about health differences between social groups stems not from the health differences themselves, but from their covariance with other socioeconomic variables. Nonetheless, the lack of conceptual clarity in the use of terms describing social position has made it difficult to compare groups over time. Furthermore, it is critical to grasp that social status is often a crucial determinant as key dominance of health.

Measuring people' health claim methodological and ethical challenges, specially health conditions such as 'disability' and 'mental illness' (Hilary Graham, 2004) might be lies within individual, and not the status or circumstances in which area they are living.

Methodological issues such as those described above and the critical elements to define theoretical social class dimension is long-standing debates in sociology. It is important to consider the implication of the ethical perspective for setting public health priorities according health inequalities classified by social status. Therefore, some authors argue that this approach is accepted in high income countries, thus, according this argument, deprivation is not a main determinant of health, even for the poor segment of the society. However, the importance to select a relative socioeconomic status refers to an individual's makes mandatory to choose variety of proxies such as income, education, car ownership, wealth or occupation as imperfect measure of socioeconomic status. (C.J.L. Murray,1 E.E. Gakidou, & J. Frenk ,1999).

In addition, and from a regional science perspective, further analyses reported inequalities in illness across socioeconomic characteristic of the population as covariates tenure, car ownership and area deprivation. On the other hand, others investigation has been carried out separating out the effect of individual factors, such as lifestyle, housing tenure, from the contextual effects of the area. Along the vast literature, suggest that those exposed to social disadvantage, raised level of limiting long-term illness (LLTI) than more advantaged groups. In addition, other studies may well yield further evidence about socioeconomic differential in health, whether the household own its home or rents, how many automobiles it has access to (see Whitehead, Haynes for discussion), whether any household members are unemployed, and whether the household resides in a poverty area. However, regarding automobile ownership, also may suffers from a potential endogeneity problem because healthier households may choose to live further away from services and therefore may have a higher demand for automobiles (Jonathan S. Feinstein, 1993: pp.300).

Feinstein reviewed the most widely discussed explanations of both inequalities in health outcomes and the relationship between socioeconomic status and health, focusing primarily on issues and studies relevant in the United States. He suggested two dimensions refers to the underlying characteristic of the persons (or households) that may cause differences in health status, and divides these characteristics onto two distinct groups. We only take in account the first dimension related with access to resource, materialism.

Figure. 1 Feinstein Dimensions

		SOURCE OF INEQUALITY	
		Life span	Access to and utilization of health care system
TYPE OF EXPLANATION	Materialistic	Housing, overcrowding, sanitation, transit mode, occupational	Ability to purchase health care, ability to purchase pharmaceuticals
	Behaviorial	Diet, smoking, exercise regime, leisure activities, risk taking, alcohol and substance	Comprhensive medical information, following intructions, self-diagnostic, and

In terms of socioeconomic position, refers to an individual place in the social hierarchies built around education, occupation and income. These factors are high important to consider because their influence on whether an individual differs in life chances. These three components could be used to produce a hierarchical classification of socioeconomic status: from unskilled manual job to professional job, from no qualifications to degree-level qualifications and from low income to high income. In addition, Housing tenure and household assets, procure additional measure of socioeconomic status. (Marmot M, Kogevinas M, Elston MA, 1987, 8:111–35)

.

In order to address theoretical and practical issues claiming in the definition of social status appears an deprivation index studies. However, there are drawbacks as well to these measures. Deprivation research is related to disadvantage suffered by an individual, concerning the living condition of the community. According with Townsend, this index is a measure of material deprivation. Carstairs deprivation index is very similar to Townsend index and differs from it only because contains variables more representative of Scotland situation. (Carstairs and Morris, 1991). Furthermore, Forrest and Gordon in their study developed two different measures of deprivation, with the distinction between material and social deprivation (Forrest and Gordon 1993). Finally, Index of Multiple Deprivation was performed in based on the idea that deprivation is made up of separate dimensions. However, deprivation index could make misunderstanding in the interpretation. For instance, people living in more rich areas are classified as not disadvantaged and, on the other hand, affluent people living in poor areas are considered disadvantaged. Hence, Index of deprivation also makes it hard to identify whether health effect is caused by individual's socioeconomic conditions or the wider neighborhood. Francisca on her study point out that it's an important limitation: there is evidence that living in a poor area, with run-down housing, poor local services and high levels of crime, takes an additional toll on health, over and above the poor personal circumstances of the people who live there.

Data and Methods

Data

The aim of this report presented here is to specify a vast of statistical modeling that will help us understand the nature and extent of variation in healthy at different level. We would like to have information on all the variables that appear to have relationship with illness, as discussed in the section Determinants of Healthy inequalities for individual and region level, in order to test the extent to which variations in healthy can be explained by compositional factors. Despite that there are ongoing important efforts to build comprehensive databases of illness variation; we have used of two different resources depending on the context. Thus, we made use the data available in Office for National Statistics ONS (2001) census data. The ONS is used for many researchers as a reference base for many statistical series such as population estimates and projections and sample surveys, enables to central and local government, healthy authorities and many other organizations to target their resources more effectively and to plan housing, education, health and transport services for years to come. We use ONS (2001) to analyze data on healthy variations and its possible determinants at the district level. The letter data set is Census Samples of Anonymised Records (SARs). The SARs is a 3 per cent representative sample of the population of England and Wales containing comprehensive of datasets drawn from the 1991 and 2001. It covers information on age, gender, ethnicity, health, employment status, housing, amenities, family type, geography, social class and education. These include the illness question (LLMIT), which is based on the responses a long term health problem or disability that limited their day-to-day activities and that had lasted, or was expected to last, at least 12 months, were asked to assess whether their daily activities were limited a lot, a little or not at all by such a health problem. In this study, we used the response of LLMIT summarized in a percentage of illness population in the case of ONS data set and, scale 0 to 1, in the case of SARs data set.

The datasets described above were emerged in two single micro data file, which was then analyzed in the software package SPSS. We fitted model of healthy measure LLMIT described above treated as continuous and dichotomous variable depending on analyzed context. In addition, some of the variables described in the Appendix were used to derive interaction term in order to capture better understanding and contextual influences. For instance, in district-level unemployment rate from the Census were multiplied by lone parents rate and unsuitable accommodation. Similar interaction terms were considered in our individual model for the following variables with the objective to figure out some assumption about individual variability in healthy: *marital status, economic status and tenure.* An additional derived methodology was to create a material index deprivation in order to identify which districts are considered the most/least deprived.

Methods

The datasets that we worked comprises two levels of observation. As noted in the previous section, we made used two different statistical modeling. Firstly, we created a regression analysis. In particular this approach can be used to discover different characteristics, such environmental or other factors 366 metropolitan districts, which are relevant to consider determinant of health variability. As mentioned in previous subsection, our dependent limiting long term was treated as continuous variable and we employed linear model for simplicity. The sample that we used comprised 366 districts in 47 metropolitan areas. Thus we fitted the model regression (called area level model) with five explanatory variables.

(1)

$$Log(Llmit)=\beta_0 + \beta_1{*}X_1 + \beta_1{*}X_1 + \beta_1{*}X_1 + \beta_1{*}X_1 + \beta_1{*}X_1 + u$$

It was also possible to include two interaction terms to explore links between variables. In the context of district research, we explored the impact of such interaction terms for employment, housing tenure and lone parent. (see Model 2 below). On the basis of our regression model, we were able to estimate the hypothetical pattern of healthy variation in district level as well as identify the goodness predictor of our model (standardized residual). We then systematically introduced five explanatory variables which each selection was based on theoretical considerations that were briefly reviewed in the section Determinant of health inequalities.

Secondly, a binary logistic regression was performed to predict whether level of healthy among individuals in Britain is reflect by some individual characteristic of the sample respondent. In this research context, we used the second data resource of SARs. In this case, the sample that we used comprised 5000 individual, nested in twelve regions. Hence, we conducted logistic regression as it enables us to address questions such as: who is more likely to be considered unhealthy? What are the factors that influence different types of individual 'illness'? Therefore, there are many factors and conditions involved and the subjective approach might be vary the 'type' of individual that we are looking for Thus, all the variables are selected with some theoretical reason as are demonstrated on previous studies (see Determinant of healthy inequality, individual section above).

Specifically, all logistic models were performed with limiting long term illness as the response variable(Y=1 YES illness, Y=0 NO illness):

(2)

$$\left(\frac{p}{1-p}\right) = logit(p) = \ \alpha \ + \ \beta 1 * X1$$

Indeed, in order to address the large number of potential covariates and avoid over-fitting the logistic regression models, we examined the impact of the risk factors by creating a core model and then adding interaction effect mentioned in above subsection. In particular, this interaction effect attempts to address a subjective hypothetical assumption that might be considered a relevant determinant of healthy variability. We refer whether an individual have this two situation: (1) as more payment commitments you have, more likely to be sick like stress or psychology problems, especially whether you have a job or no. In other words, could be explained variability in healthy among individual by interaction term between be unemployment and live in rented house?, 2) what would be happened if we have also 'family duties', for instance, be married? and (3) do you live in London? How do you feel?

Ones we had our multi lineal regression model and logistic model as reflecting district and individual models, we have considered that a comparative perspective among different contexts would make to enhance the interpretation our report.

Therefore, on one hand, we included a Index Deprivation in order to locate those district area that need environmental improve situation and, on the other hand, areas that are considered least deprived. We built an index of deprivation following the British historical model. In general, we decided to develop a material deprivation index, which represents a objective measure, independent from the consequences. There are a large vast of studies that have helped to create with our Index. Moreover, all variables were taken in consideration concerning previous academy studies and reports, especially ONS Report, Index Deprivation in UK, or The English Index Deprivation by Department for Communities and Local Government. In addition, Townsend and Carstairs were taken into consideration. (See specification previous subsection, Determinant of health inequalities, part of Deprivation).

On the other hand, we classified the district according our Index Deprivation through cluster analysis. The aim in this part of the report was to discover which areas has either similitude or different among them.

Modeling Results

This section presents the results of four different statistical modeling regarding the research context. First, we discuss the results of the Multi lineal Models Regression that was described in Equation *(1)*. Within first section, we show Model 1 and then, we add the interaction term as we described before. Secondly, we go through individual level with our Logistic Binary Regression that was noted in Equation *(2)*. Again, we show, firstly our core model and subsequently, the interactions term. After then, as third stage is presented the outcomes regarding the 'Material' Index Deprivation and the Cluster Classification.

We should mention that for each method, we have carried out some exploratory data analysis (EDA) to test whether the data sets needed some variable transformation in order to deal with some assumption depending on statistic modeling.

Linear Regression

Exploratory Data Analysis

We check whether our model is valid and unbiased. Thus, it requires the following assumptions to be met: normality of residuals, independence of residuals, no multicollinearity among predictors, and homogeneity of residual variance.

Exploratory Analysis shows that all of the variables selected have a positive skewed with many outliers in the right tail of the distribution. Specially, Tower Hamlets in persons living in household over 1,5 could be an extreme value to drop in our analysis but we decided to follow up with this since we have not robust information about the neighborhood (50% more people living over 1,5 than the second). To deal with normal distribution of our dependent variable, we made the log transformation. Regarding the variables that we observe, it is rather normal to have this kind of distribution. Thus, we have been considered to not drop these outliers.

We notice that there is a strong positive correlation between percentage of lone parent families and percentage of unemployed persons ($r=.907$, $p< .001$), over of .80 threshold, so we have taken into account for our model. The rest of the explanatory variables, have a moderate positive correlation with our respondent variable (see appendix A).

Models

Table 2 gives the result of two models of estimated fixed effect of number of demographic, economic position and social status in district-level covariates of illness. Second model correspond to interaction term.

Table 2. Models specification:

Model Summary[c]									
					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.851[a]	.724	.720	.10297	.724	188.588	5	360	.000
2	.875[b]	.765	.760	.09527	.041	31.269	2	358	.000

As we note, Model 1 with all linear terms get 72.5% of the variance of the respondent variable ($p<.001$). After adding the interaction terms, the explanatory power is greater, which has $R^2_{adj}=.760$, improving in 5,6% of the previous model. Thus, the Model 2 has been preceded for the interpretation of variability in ward level.

Table 3. Multi Linear Regression Model

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	2.437	.005		452.767	0.000	2.426	2.448
	age60p_centered	.033	.002	.615	20.685	.000	.030	.036
	unemp_centered	.107	.009	.806	12.057	.000	.090	.124
	density_centered	-.032	.007	-.161	-4.248	.000	-.046	-.017
	unsuit_centered	-.122	.017	-.259	-7.349	.000	-.155	-.089
	lpfam_centered	.005	.004	.107	1.467	.143	-.002	.012
2*	(Constant)	2.461	.006		421.300	0.000	2.450	2.473
	age60p_centered	.034	.002	.626	21.717	.000	.031	.037
	unemp_centered	.106	.008	.797	12.885	.000	.090	.122
	density_centered	-.020	.007	-.102	-2.854	.005	-.034	-.006
	unsuit_centered	-.144	.020	-.305	-7.044	.000	-.184	-.103
	lpfam_centered	.016	.004	.324	4.433	.000	.009	.023
	unemp*lpfam_centered (Int1)	-.005	.001	-.361	-7.363	.000	-.007	-.004
	unemp*unsuit_centered (Int2)	.015	.008	.096	2.038	.042	.001	.030

a. Dependent Variable: log_llti

** Interaction effect were computed substracting the main mean of each variable in order to enhance the aid and interpretation of the model*

As we can be seen, all coefficients are highly significant at 95% significance level, excepting the second interaction term between unsuitable accommodation and unemployment, but it is still significance. These results suggest that: (1) one percent of increase in unemployment rate of the area, increase 0.11 the limiting long term. It interesting to note that people who live in unsuitable accommodation has a negative effect to our respondent variable and, with the interaction term unemployment, it is converted to .015. The explanation would be that with this interaction, we are only take in account people in age of work thus, those people who live these conditions and have not job, make increase in .015 the limiting long term of the region. On contrary, the first interaction obtained the opposite effect, lone structure family increase the level of unhealthy areas but when we mix with unemployment, decrease this level. However, this interaction has less relative importance (β=0.96) than first interaction (β=-3.61).Therefore, to illustrate such interaction effect you can understand the relationship, having a look to the regression equation (Appendix A, Int

Figures). The overall interpretation with the effect of centered our covariates is 2.461 as average of limiting long term (in log term) as ward level when all the covariates and factors are holding in their respective variables. We can see that low % persons in unsuitable accommodation (blue line) have strong regression effect (R^2=0.528). The correlation drops as same time as % persons in unsuitable in one area (correlation high %unsuit=0.374).

Finally, the standardized residual saved on the model, allow us to identify the endpoints of limiting long term illness expectations. Then, we identify on Table 3. that Lancashire has considered the lower expectation of limiting long term illness with 2.0 of standardized residual and, on the opposite side, Hertfordsh have been predicted as worst healthy area than the actual values.

Table 4: Standardized residual (Zresid)

LLTI EXPECTED CLASSIFICATION

EXPECTED	N	ward	Actual Values	Predicted Values	Standardized Residual
LOWER	1	Lancashire	2.63	2.44	1.99999
	2	Durham	2.76	2.59	1.80327
	3	South York	2.78	2.67	1.11528
HIGHER	45	Berkshire	2.11	2.19	-0.81890
	46	West Midla	2.55	2.63	-0.82511
	47	Hertfordsh	2.21	2.33	-1.25075

Logistic Regression Analysis

<u>Exploratory Data Analysis</u>

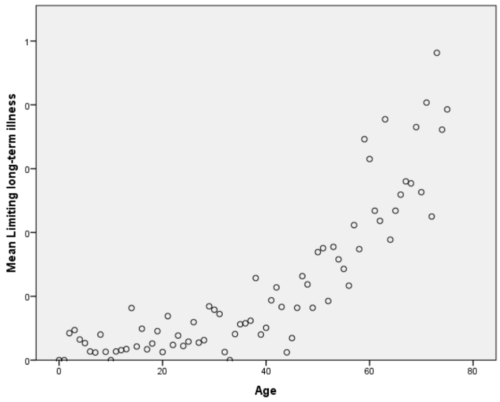
- ***Sex***

Females are more likely to be unhealthy than males (relative odds of 1.21).

- ***Age***

Exploring age showed a mean age of 53.09 for illness cases and 32.76 for no illness cases. Indeed, we tested for the linearity assumption for this continuous variable (age against the log of age, as described by Field, 2009). The variable has been met the assumption of linearity as expected given the log transformation. Therefore, we decided to include this variable as it potential had a strong explanatory role. Furthermore, a scatter plot was performed to understand the relationship between the variable “age” and the dependent variable “LLMIT”. Finally, the chart below shows a exponential growth of illness people when increase the age.

Figure 2: Exponential



- ***Ethnic group (MSG):***

Although included social group as determinant to explain this health variability would be to accept some ethic issues, we have thought that, very often, minorities social group are located in deprived living condition due lack of resources in term of incomes. Thus, we have seen appropriate to add this covariate to explain the variability of healthy. We have done cross tabulation of the new variable recode in MSG(minority social group) where 1 correspond to be in YES MSG and 0 correspond to Not be MSG(white) showing odds ratio of a person from a BME group being illness were 0.0479 compared to 0.1157 for someone not from this group (relative odd of 2.41).

- ***Marital status***

The variable “marital status” has five categories, two of them “married” and “remarried” refer to the same situation, and as a consequence these categories were collapsed.

- ***Economic position:***

This variable was filtered out permanent sick. The objective is to attempt to discover which economic position is more likely to be illness.

- ***Tenure***

Tenure was recoded from ten to two categories to fit the theoretical assumption that owner occupiers would be less likely to be illness than those renting. The variable was recoded as 0=owner occupied, 1= rented. Cross tabulation showed a strong relationship to illness, where social renters were more likely to be illness that owner

occupiers (relative odds of 2.14).

- ***London+No London:***

This variable was recoded from region.

<u>Multicollinearity</u>

In order to assess the assumptions in logistic regression analysis, the selected variables were subjected to a Linear Regression analysis with the sole purpose of evaluating multicollinearity in the model (Menard, 2002). The VIF and Tolerance values indicate that there is no evidence of multicollinearity in the model. (See appendix B)

In the context of the research presented here, we provided information on individual level. Table 4 showed categorical variables selected for the logit model. Nonetheless we had variables information concerning socioeconomic status like number of car, country of birthday; we have though enough to include these covariates as proxies for socioeconomic status.

Table 5: Dummy variables

Categorical Variables Codings

		Frequency	Parameter coding (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Economic position (primary)	Employee FT	1589	0	0	0	0	0	0	0	0
	Employee PT	387	1	0	0	0	0	0	0	0
	Self-emp with	90	0	1	0	0	0	0	0	0
	Self-emp without	199	0	0	1	0	0	0	0	0
	Govt scheme	38	0	0	0	1	0	0	0	0
	Unemployed	248	0	0	0	0	1	0	0	0
	Student	171	0	0	0	0	0	1	0	0
	Retired	532	0	0	0	0	0	0	1	0
	Other inactive	517	0	0	0	0	0	0	0	1
Marital status	Single	1029	0	0	0					
	Married	2297	1	0	0					
	Divorced	234	0	1	0					
	Widowed	211	0	0	1					
Lon=noLon	no london	3332	0							
	London	439	1							
MEGroup	No	3623	0							
	Yes	148	1							
type of tenure recode	owner	2743	0							
	rented	1028	1							
Sex	Male	1797	0							
	Female	1974	1							

Model 1.

Table 6: Variables Model 1

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Model 1.	age	.057	.006	84.886	1	.000	1.059	1.046	1.072
	Female	.018	.149	.015	1	.901	1.019	.760	1.364
	Single			28.053	3	.000			
	Married	-.930	.195	22.697	1	.000	.394	.269	.578
	Divorced	-.204	.275	.553	1	.457	.815	.476	1.397
	Widowed	-.638	.262	5.946	1	.015	.528	.316	.882
	MEG(yes)	-.417	.478	.759	1	.384	.659	.258	1.683
	Employee FT			62.290	8	.000			
	Employee PT	-.154	.333	.212	1	.645	.858	.446	1.648
	Self-emp with	.324	.540	.360	1	.549	1.383	.480	3.985
	Self-emp without	-.094	.446	.045	1	.832	.910	.380	2.179
	Govt scheme	.943	.759	1.543	1	.214	2.567	.580	11.362
	Unemployed	1.081	.275	15.468	1	.000	2.948	1.720	5.054
	Student	.411	.546	.568	1	.451	1.509	.518	4.396
	Retired	1.190	.227	27.390	1	.000	3.286	2.104	5.130
	Other inactive	1.362	.225	36.744	1	.000	3.905	2.514	6.066
	Constant	-5.213	.303	296.741	1	.000	.005		

Model 1 was created with limiting long term illness as the response variable (coded 1 YES illness against 0 NO illness) and the explanatory variables of: age, sex, marital status, social group and economic position. This model is significant different from 0 with 14 df,($p<0.0005$) and explained between 11% (Cox & Snell R Square) and 24.7% (Nagelkerke R Square) of the variation in health (see Appendix B, summary model table).

The equation model 1 is:

$$\begin{aligned}\boldsymbol{Logit}(\boldsymbol{lltil}) = {} & -5.213 - .057age + .018female - .930married - .204divorced \\ & - .638widowed - .417MEG(yes) - .154employeePT \\ & + .324selfempwithGB - .094selfempwithoutGB - .943Govts\,cheme \\ & + 1.081unemployed - .411student - 1.190retired - 1.362otherinactive\end{aligned}$$

Looking at table 5, the following variables have significant negative main effect on illness: be married, be in minorities ethnic groups (not significance), and either to be not unemployed, or student or retired or other inactive. In other words, a person who are older, female (there is not significant effect, but tend to be positive more than men), and not working, are considered the most likely to be illness.

It is also interesting to note that according to this model, the respondents who were unemployed, higher than average, are likely to be unhealthy compared with individuals, who are working, especially employee FT. This result suggest that a person that has achieved to get into the social rules as a workers, it help to feel more healthy and comfortable because, to get a job, might be considerate as social and economic reward and, subsequently, an enhance of mental health. In addition, be engaged has a negative main effect, increasing the probability to be sick of those who are married against single in .384 times.

Interaction Term Models

The next step was to further investigate the importance of the context in determining individual illness adding some interaction term to the previous model. The aim of theses interactions is justify our assumption of: (1) as more payment commitments you have, more likely to be sick like stress or psychology problems, especially whether you have a job or no. In other words, could be explained variability in healthy among individual by interaction term between be unemployment and live in rented house? We have gone beyond and we have added a third condition that might be affecting to health variability: (2) what would be happened if we have also family duties, for instance, be married?

To enhance the relationship between levels, we added a third (3) model. We wanted to test whether an individual live in the heavy stressful life of London or not. In principle, it seems that live in the capital could cause some stress related illness underlying to the fact to be on continue pressure of the social and work life.

The outcomes presented on the table 3 on Appendix B (Omnibus Tests of Model Coefficients Table) shows that the addition of the interaction term (1) and (2), were significant for the model and hence, it does improve the fit of the model.

Table 6: Variables Interaction term Model (1): could be explained variability in health by unemployment and live in rented house?

Model Int(1)	B	S.E.	Wald	df	Sig.	Exp(B)	EXP(B) Lower	EXP(B) Upper
age	.058	.006	83.740	1	.000	1.059	1.046	1.073
sex(1)	.007	.152	.002	1	.965	1.007	.748	1.355
mstatus			17.339	3	.001			
mstatus(1)	-.784	.199	15.593	1	.000	.457	.309	.674
mstatus(2)	-.252	.279	.816	1	.366	.777	.450	1.343
mstatus(3)	-.620	.266	5.442	1	.020	.538	.319	.906
MEG(1)	-.405	.479	.716	1	.397	.667	.261	1.705
econprim			37.107	8	.000			
econprim(1)	.125	.388	.103	1	.748	1.133	.529	2.425
econprim(2)	.682	.555	1.510	1	.219	1.978	.666	5.871
econprim(3)	.142	.498	.082	1	.775	1.153	.434	3.060
econprim(4)	1.227	1.050	1.366	1	.243	3.411	.436	26.703
econprim(5)	1.132	.419	7.304	1	.007	3.103	1.365	7.055
econprim(6)	-.370	1.034	.128	1	.720	.691	.091	5.237
econprim(7)	1.130	.271	17.377	1	.000	3.094	1.819	5.263
econprim(8)	1.444	.287	25.351	1	.000	4.238	2.416	7.436
owner_rent(1)	.911	.305	8.912	1	.003	2.487	1.367	4.523
econprim * owner_rent			5.291	8	.726			
econprim(1) by owner_rent(1)	-.937	.746	1.580	1	.209	.392	.091	1.689
econprim(2) by owner_rent(1)	-19.151	13233.978	.000	1	.999	.000	0.000	
econprim(3) by owner_rent(1)	-.659	1.173	.316	1	.574	.517	.052	5.156
econprim(4) by owner_rent(1)	-.805	1.521	.280	1	.596	.447	.023	8.812
econprim(5) by owner_rent(1)	-.450	.563	.640	1	.424	.637	.211	1.922
econprim(6) by owner_rent(1)	1.448	1.211	1.429	1	.232	4.253	.396	45.645
econprim(7) by owner_rent(1)	-.052	.365	.020	1	.887	.950	.465	1.941
econprim(8) by owner_rent(1)	-.414	.411	1.018	1	.313	.661	.295	1.478
Constant	-5.582	.336	276.594	1	.000	.004		

As we show in Table 6, our model 2 is adding the interaction term (1)=*ecoprim*owner_rented*. The additional variable reduced the -2 log likelihood to 1748.809 (df=23), and improved the explanatory power to between 11.8% (Cox & Snell R Square) and 26.6% (Nagelkerke R Square). Furthermore, the assumption of being unemployed and with house rented *[(ecoprim(5)*by owner_rent(1)*] has a positive effect on the response variable, contrary to our expectations. However, this effect has been considered not significant for the model estimated as we can note in variable in the equation. (See appendix B: Model summary).

Table 7: Variables Interaction term Model (2):: what would be happened if we have also family duties, for instance, be married?

Model Int(2)	B	S.E.	Wald	df	Sig.	Exp(B)	EXP(B) Lower	Upper
age	.059	.0	80.700	1	.000	1.1	1.047	1.075
sex(1)	.027	.2	.031	1	.861	1.0	.760	1.388
mstatus			14.614	3	.002			
mstatus(1)	-.720	.2	8.342	1	.004	0.5	.299	.794
mstatus(2)	.221	.4	.391	1	.532	1.2	.624	2.497
mstatus(3)	-.580	.3	2.892	1	.089	0.6	.287	1.092
MEG(1)	-.356	.5	.542	1	.462	0.7	.272	1.807
econprim			35.891	8	.000			
econprim(1)	.109	.4	.078	1	.780	1.1	.519	2.395
econprim(2)	.697	.6	1.568	1	.211	2.0	.674	5.977
econprim(3)	.162	.5	.105	1	.746	1.2	.442	3.123
econprim(4)	1.290	1.1	1.494	1	.222	3.6	.459	28.725
econprim(5)	1.156	.4	7.547	1	.006	3.2	1.393	7.252
econprim(6)	-.259	1.0	.062	1	.804	0.8	.101	5.920
econprim(7)	1.122	.3	16.446	1	.000	3.1	1.785	5.281
econprim(8)	1.435	.3	24.267	1	.000	4.2	2.373	7.432
owner_rent(1)	.911	.3	8.819	1	.003	2.5	1.363	4.540
econprim * owner_rent			3.014	8	.933			
econprim(1) by owner_rent(1)	-18.393	8950	.000	1	.998	0.0	0.000	
econprim(2) by owner_rent(1)	-19.137	1322	.000	1	.999	0.0	0.000	
econprim(3) by owner_rent(1)	.637	1.2	.273	1	.601	1.9	.174	20.574
econprim(4) by owner_rent(1)	.173	1.6	.012	1	.914	1.2	.052	27.094
econprim(5) by owner_rent(1)	.221	.6	.134	1	.715	1.2	.381	4.085
econprim(6) by owner_rent(1)	1.231	1.3	.925	1	.336	3.4	.279	42.052
econprim(7) by owner_rent(1)	-.644	.6	1.063	1	.303	0.5	.154	1.786
econprim(8) by owner_rent(1)	-.143	.6	.064	1	.800	0.9	.285	2.631
econprim * mstatus * owner_rent			15.314	18	.640			
econprim(1) by mstatus(1) by owner_rent(1)	17.298	89.0	.000	1	.998	32532113.8	0.000	
econprim(1) by mstatus(2) by owner_rent(1)	-1.248	132.0	.000	1	1.000	0.3	0.000	

(Continue)

Model Int(2)	B	S.E.	Wald	df	Sig.	Exp(B)	EXP(B) Lower	EXP(B) Upper
econprim(1) by mstatus(2) by owner_rent(1)	-1.248	13237.2	.000	1	1.000	0.3	0.000	
econprim(1) by mstatus(3) by owner_rent(1)	19.611	8950.4	.000	1	.998	0.0	0.000	
econprim(3) by mstatus(1) by owner_rent(1)	-18.704	11462.3	.000	1	.999	0.0	0.000	
econprim(3) by mstatus(2) by owner_rent(1)	-20.200	15657.5	.000	1	.999	0.0	0.000	
econprim(4) by mstatus(1) by owner_rent(1)	-19.665	17578.7	.000	1	.999	0.0	0.000	
econprim(4) by mstatus(2) by owner_rent(1)	-20.911	40193.0	.000	1	1.000	0.0	0.000	
econprim(5) by mstatus(1) by owner_rent(1)	-2.248	1.1	4.141	1	.042	0.1	.012	.920
econprim(5) by mstatus(2) by owner_rent(1)	-2.154	1.2	3.454	1	.063	0.1	.012	1.125
econprim(5) by mstatus(3) by owner_rent(1)	21.461	40193.0	.000	1	1.000		0.000	
econprim(6) by mstatus(1) by owner_rent(1)	1.228	1.4	.813	1	.367	3.4	.236	49.321
econprim(6) by mstatus(2) by owner_rent(1)	-19.507	40193.0	.000	1	1.000	0.0	0.000	
econprim(7) by mstatus(1) by owner_rent(1)	.764	.6	1.656	1	.198	2.1	.670	6.881
econprim(7) by mstatus(2) by owner_rent(1)	.474	.8	.334	1	.563	1.6	.322	8.005
econprim(7) by mstatus(3) by owner_rent(1)	.466	.7	.502	1	.478	1.6	.439	5.782
econprim(8) by mstatus(1) by owner_rent(1)	-.145	.6	.069	1	.793	0.9	.293	2.553
econprim(8) by mstatus(2) by owner_rent(1)	-1.326	.8	2.605	1	.107	0.3	.053	1.329
econprim(8) by mstatus(3) by owner_rent(1)	-.240	.7	.103	1	.748	0.8	.182	3.403
Constant	-5.753	.4	#####	1	.000	0.0		

As we note in table 7, turning to the interaction terms that were introduced in Model 3, concerning the interaction between=*maritalstatus* ecoprim*owner_rented*. Nonetheless, the coefficient for the variables obtained has a difficult understanding. Despite the interaction variables were not significant; there is a highlight that we wanted to point out. Also there are important changes to the coefficient values between model, although is not significant for either Models. Relating with the model 2, the same interaction term has been converted to be factor risk of be sick (see table) when we mixed these interaction with marital status. Indeed, the only significantly effect is when one respondent on the survey was unemployment, living in rent condition and divorced obtaining a confidence interval for exp(B) was .023, indicator that on this circumstances, are more chances(0.01 times) to be sick if you compared with a person with job, married and with owner house.

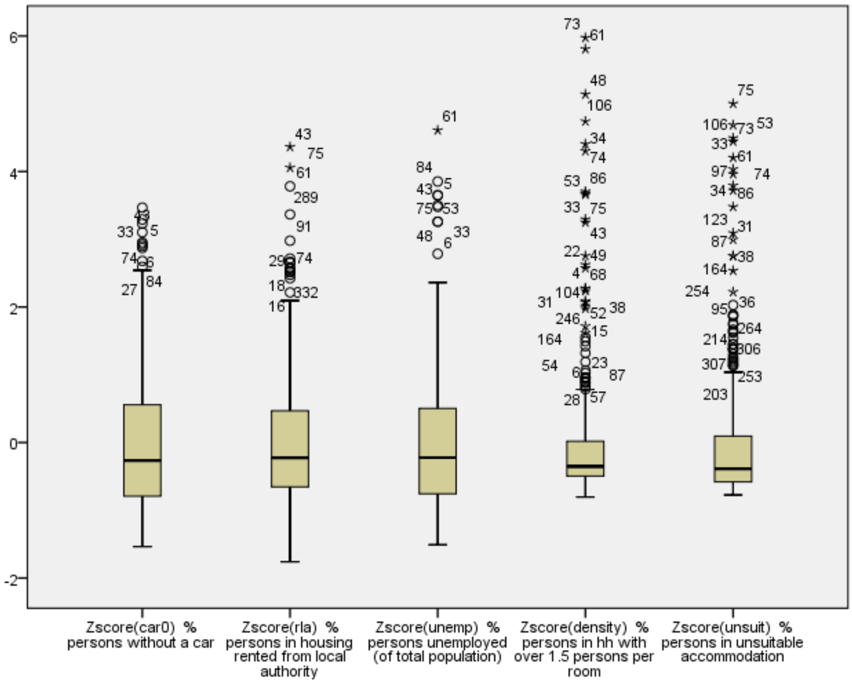
Finally, we add on the least model our forth condition (4) whether an individual is in London or not. Concerning the quality of the model, is seems that is not improving. In addition, the effect considered as healthy risk factor, were not met with the model since this effect is negative contribution of the equation. In fact, a person in London is .721 less likely to be illness than the rest of the individual in different geographical regions. Moreover, this effect was not significance. (see appendix B for further information).

'Material' Index Deprivation

Exploratory Analysis

They seem to exhibit very strong skew to the right. This fact is also apparent when box plots are drawn. Box plots also show that the selected variables have good degree of variability.

Figure 3: Box-Plot Z-score



Subsequently, bivariate correlation analysis was performed (see correlation matrix in the appendix C). This result suggests that the variables are strongly significantly correlated with each other. Despite there are a pair of variables with Pearson's correlation coefficient greater than 0.90 (unemployment and no car), we decided to keep them in our factor analysis due to importance of them. Furthermore, though such variables may they do not capture deprivation strongly when applied at the ward level. Wards contain a high proportion of people from an ethnic minority. Moreover, we noted a strongly correlated with household overcrowding. This illustrates a cultural preference for larger families rather than being indicative of deprivation. Thus, being a member of an ethnic minority may be a risk factor that can contribute to social deprivation, but since we attempt to measure Material Deprivation it is not a constituent part of it. Together these arguments are a strong case against including ethnicity as part of our deprivation index.

The selection of the variables are directly related to the concept of material deprivation and driven by the availability of data. The variables should be, "direct" measures of deprivation (Townsend et. al.1988). Variables reflecting area characteristic (such population, number of female) as well as long term illness adult were excluded as well as no earning in households which represented child not in age of work.

Thus, the variables that we have selected are as Table 9 shown below: i) *person without car*, ii) *housing rented by LA*, iii) *person without car* iv) *density over 1.5 pers per room* v) *persons in unsuitable accommodation.*

Table 8: Description variables

Variable	Domain	Minimum	Maximum	Mean	Std. Deviation
% persons without a car	Income	6.08	53.87	20.7542	9.55044
% persons in housing rented from local authority	Income	2.29	51.32	16.3766	8.00890
% persons unemployed (of total population)	Employment	1.81	10.59	3.9708	1.43575
% persons in hh with over 1.5 persons per room	Living env.	.11	5.15	.7093	.74278
% persons in unsuitable accommodation	Living env.	0.000	2.390	.32019	.413966

Furthermore, unemployment represents a state of lack of resources -and economical insecurity-; housing loan situation (rented) could be intended as a proxy for wealth, while overcrowding and persons in unsuitable accommodation has been inserted for its potential capacity to synthesize living conditions. (Cadum et al 1999).

Table 9: Summary of Component Principal Analysis

Variable	Initial Solution: Lack resources	Initial Solution: Living Env.	Rotated Solution: Lack resources	Rotated Solution: Living Env.
% persons without a car	.926		.930	
% persons in housing rented from local authority	.751		.905	
% persons unemployed (of total population)	.930		.933	
% persons in hh with over 1.5 persons per room	.804	.432		.879
% persons in unsuitable accommodation	.613	.718		.940
% variance explained by component	66.181	21.55	66.181	21.55
% variance explained by total	87.73		87.73	

As we can note in Table 9 above, initial component extraction shows that the four first variables are heavily loaded onto added in Factor Lack of resources. However, the only one variable adding in Living condition Factor is weighted, almost evenly on both factors. Therefore, in order to improve the interpretability and spread of variability between them, we ran oblique rotation, becoming the loading clearer as we have plotted in the figure 4 and figure 5 below:

Figure 4: Factor Plot

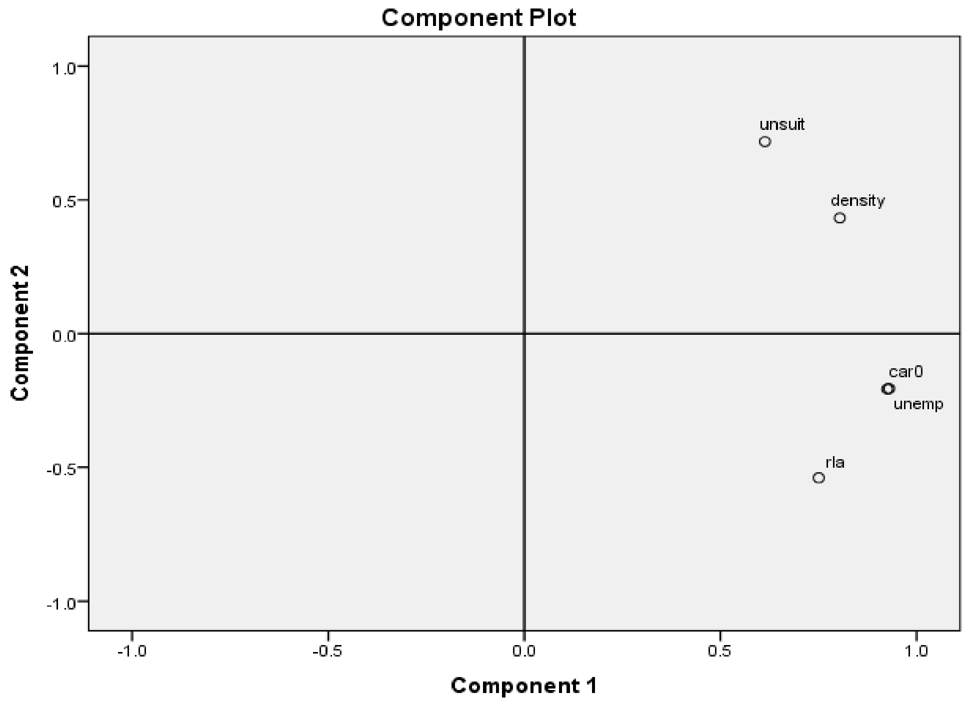
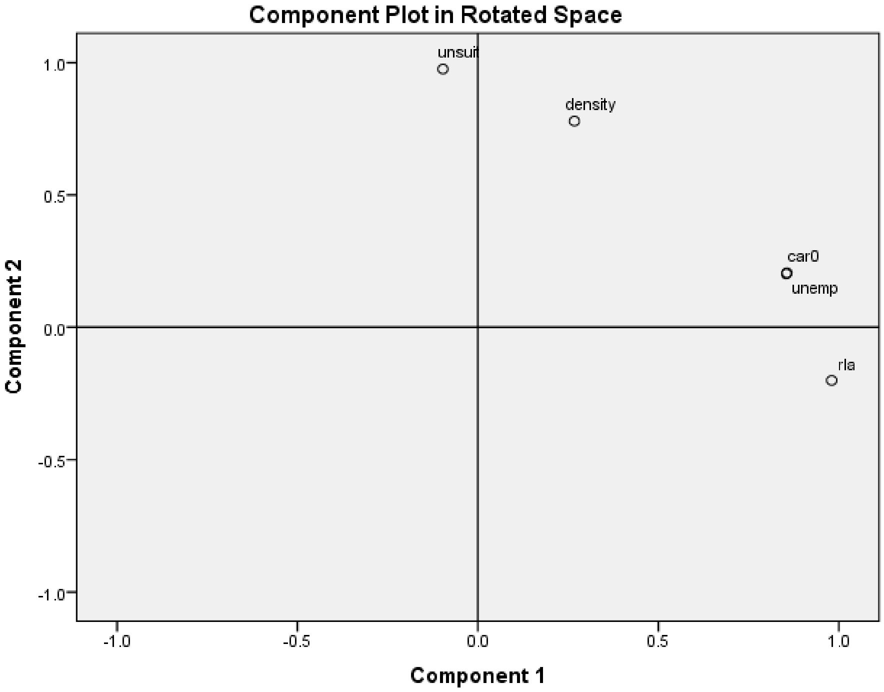


Figure 5: Factor Plot rotated



In figure 2, which is before rotation, the two factors are unlikely to be independent. It would be reasonable to expect that poor living conditions of the region is consequence of lack of income and therefore less likely to have owner house or car. On the figure 3, after rotation, using the direct oblimin method, we can observe how is minimizing such distance between factors. Thus, the screen plot was concise and showed inflexions that would justify retaining both components.

The Kaiser–Meyer–Olkin measure verified the sampling adequacy for the analysis, KMO = .766 ('moderate' according to Field, 2009), and all KMO values for individual items were > .754, which is well above the acceptable limit of .5 (Field,2009). Bartlett's test of sphericity χ^2 (10) = 1367.968, $p < .001$ (see appendix C), indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. Two components had eigenvalues over Kaiser's criterion of 1 and in combination explained 84,616% of the variance. The screeplot was concise and showed inflexions that would justify retaining both components.

Finally, we have decided the most appropriate factor to use is that which combines the variables measuring lack of resources. Whilst not measuring deprivation as such, a combination of these variables would be an indicator of the domains that affect living conditions. Then, the first component has been retained as our Index Deprivation. Therefore, component scores are then computed using the regression method as the following equation:

Component 1 *= .334*no_car + .411*LArented + .335*unemployed + .058*density + -102*unsuitable*

Based on 'Material' index of deprivation, the top 3 most deprived and least deprived wards in Great Britain is as the following in table 9:

Table 9: Top 3 Most/Least Deprived Areas

Rank	Most Deprived		Least Deprived	
1	Hackney	4.181268	Broadland	-1.47277
2	Southwark	4.06	East Dorset	-1.4626
3	Islington	3.535801	Chiltern	-1.39391

Area Classification

A number of appropriate cluster groups is subjective. We have fixed at final decision as balance between reduce number of clusters (heterogenic group) and excessive number of groups that may be difficult to understand. First, we ran a hierarchical cluster analysis (Ward Method) in order to decide how many cluster we needed, and consequently, we applied non-hierarchical (k-means). Dendogram of Ward Linktage at 0-5 for cluster analysis.(Euclidian distance) and Variance Ratio Criteria (VRC), have proved 3 cluster members as optimum to fix.

Table 10: VRC criteria

Variance Ratio Criteria		
WK=(VRCK+1 - VRCK) - (VRCK - VRCK-1)		
Num	WRC	WK
3	1513.171	-906.666
4	1059.838	91.97862
5	1105.827	-316.715
6	947.4695	

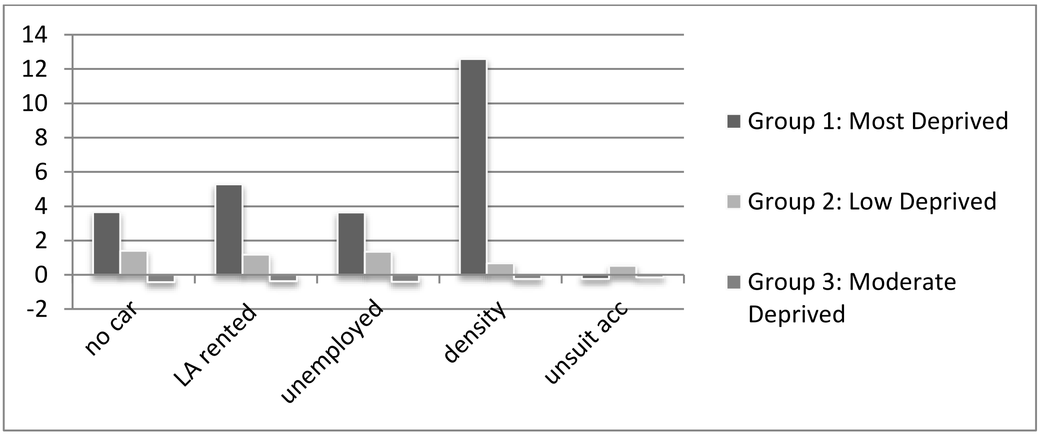
Thus, these are the final result of clusters:

Table 11. Descriptive Statistic Table (no standardized)

	Cluster Number of Case											
	1				2				3			
	Most Deprived				Low Deprived				Moderate Deprived			
VARIABLES	Mean	sd	N	N %	Mean	sd	N	N %	Mean	sd	N	N %
%persons without a car	57.22	7.59			16.83	5.57			34.85	7.61		
%persons in housing	60.93	8.51			13.49	4.85			26.56	8.61		
%persons unemployed	9.32	1.81			3.38	.77			5.96	1.36		
%persons in hh with over	13.15	2.37			.50	.30			1.42	1.21		
%persons in unsuitable	.220	.784			.255	.255			.534	.685		
			1	0.3%			281	76.8%			84	23.0%

Looking the descriptive statistics shows above in Table 11, we note that the first segment group comprises district with greater lack of income resources (Most Deprived). In contrast, the second contains low-level of material deprivation (Low Deprived). Finally, the last cluster could refer to those areas considered as moderate lack of material deprivation (Moderate Deprived). Nonetheless, in terms of precision, the volatility within the Moderate Deprived group is latent in the first two variables with a std. dv 7.61 and 8.61 respectively. Indeed, the composition of each group is quit disproportionate since only one area is being concerned on Most deprived and low deprived group contains 76,8% of the sample size.

Graph 1: Comparative Group standardized means (%population)



The graph 1 shows above that, the most distance between groups is in term of how many people are living in house, in other words, the living environmental. This corresponds to the only one district grouped in cluster 1, Tower Hamlets. However, on contrast, this do not affect heavily, since the relative weigh of this variable is not important.

Therefore, we show here the similar structure between the index deprivation and the clusters divisions. For instance, the equation of our Index Component described above, has a negative effect of unsuitable accommodation giving more importance the fact that people has LA rented their houses. Conversely, the low proportion of the density variable captured on our Deprivation Index, explain the fact that Tower Hamlets were not consider onto Top 3 Deprived Areas. Furthermore, the most discriminator variable is density as we show in final cluster centers (see appendix d), leaving permanent evidence of the significance of this area.

What these reflections all say is that cluster means produced by the k-means algorithm is sensitive to noise and outliers (Van Der Laan, 2003). Therefore, the extreme values detected on EDA before are affecting considerably our analysis. Subsequently, we should mention that most of the cluster analysis techniques is an important tool for outlier analysis. To judge validity, we should mention that we have violated some criteria as substantial or comprehensible clusters groups provided by (Dibb,1999). Specially interested is that eleven district of the cluster number 3, have been located on Inner Lond, demonstrating the evidence that most deprived areas within London are concentrated to the north and east of the City, from Newham to Islington and from Tower Hamlets north to Enfield and Waltham Forest. London is again by far the most deprived region on the living environment deprivation domain. However, if we decided to drop these areas considered as merely outliers, the conclusions would be change. There is some statistic evidence that demonstrate that this areas are outliers (see appendix C) as with labeling Rule. We developed with g'=1.5 and g'=2.0 as Tukey, J.W suggest and we identified these outliers that make this study with lot variability:

Table 12. Extreme values

n	District	Area
1	**Tower Hamlets**	**Inner Lond**
2	**Westminster,City of**	**Inner Lond**
3	**Hackney**	**Inner Lond**
4	**Newham**	**Inner Lond**
5	**Kensington and Chels**	**Inner Lond**
6	**Lambeth**	**Inner Lond**
7	**Haringey**	**Inner Lond**
8	**Southwark**	**Inner Lond**
9	**Islington**	**Inner Lond**
10	**Manchester**	**Greater Ma**
11	**Liverpool**	**Merseyside**
12	**Knowsley**	**Merseyside**

Conclusions

In this section, we discuss the results presented above in more detail, highlighting some of the key findings. It can be argued that the most influential variables in term of variability in health are socioeconomic status. According to our model we can argued that unemployment and tenure has highly and negative effect in healthy variation. Looking at the household tenure variables, it is interesting to note that ‘owner houses tenure’ seems to have a positive effect in be *no illness*, as we described in different models. However, the interaction term between covariates seems to be not significant in all this research.

Concerning district level modeling, it suggests that the data used had some extreme values that might be makes the results not significant. For instance, we can argue that district from Inner London was the most deprived area and the place with most percentage of the population likely to be illness. However, when we have a look at our Int(3) Binary Logistic Regression, we couldn’t inference this fact. (live or not in London were not a significant effect, even not negative for illness). As expected, the variation of illness that is attributable to district level was not reduced. For example, when we carried out a Cluster Analysis to attempt to cluster district, we noted that this technique is high sensible to intra variability among area. It’s the case of density variable which despite the weight of the contribution as illness determinant is not high; the variability has made the cluster classification not reliable and not relationship with our Material Deprivation. Hence, we suggest an alternative statistical modeling approach to study this type of research. Following of past quantitative research studies that attempts to measure this variability considered individual-district level, we release a healthy equation from econometric perspective where healthy variation is related to a set of explanatory variables (see Byrne, B.M., 2009):

$r = \mathrm{h}\big(\mathrm{u}(\mathrm{y}, \mathrm{z}, \mathrm{t})\big) + \mathrm{e}$ where ‘r’ is healthy measure, ‘h’ a continuous nondifferentiable function, ‘y’ related income and ‘z’ a set of demographic and personal characteristic in ‘t’ period on time with ‘e’ error term ‘y’ is continuous.

Thus, a multilevel modeling will deal with our requirement with (1) health differences between individuals (compositional effects). In other words, the variations in health status across individuals in a population with individuals ranged along a continuum from ‘best health to worst health’ (Murray et al., 1999: 537). Thus, the variability of health at individual level can be expected as ‘pure health inequalities’ because they relate to only one dimension rather than multidimensional (health and social status). (2) Health differences between population_groups (compositional effects)._This second approach pays attention to the social patterning of health and moves beyond the variations as individual level. However, the huge controversial along the meaning and measure of health variations within social groups, the lack of either standardize definition, or measure strategic, or index, has restricted the comparative analysis of such questions and (3) Health differences between groups occupying unequal positions in society by influencing environmental exposures, physical – poor housing condition – as well as psychosocial – stressful life event- and health-damaging behaviors.

APPENDIX A: Linear Regression

EDA:

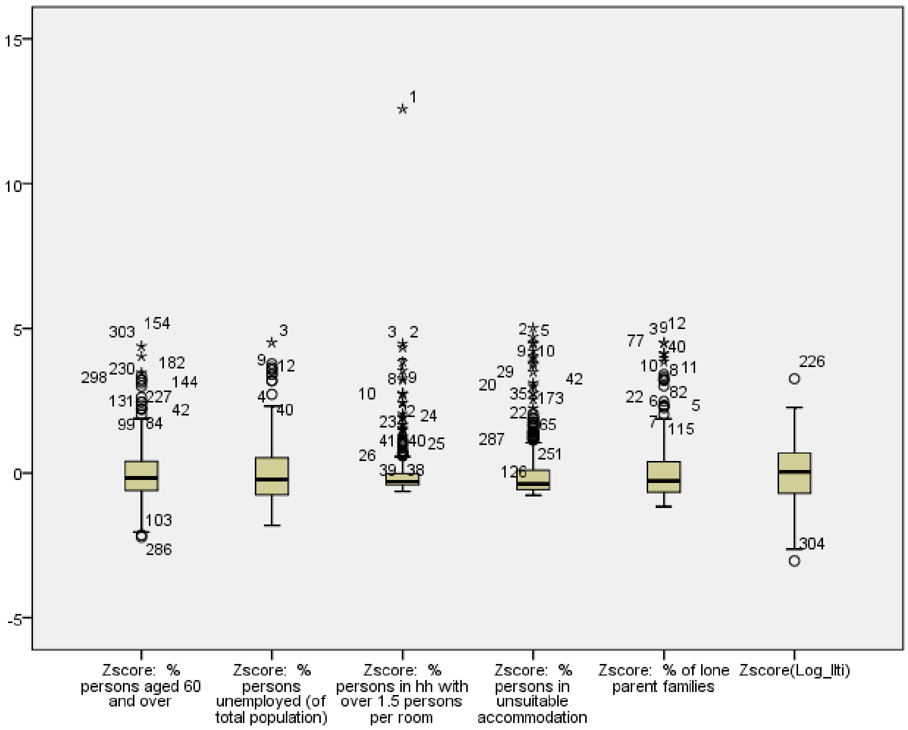
Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
%persons with llti in district	366	6.3	21.6	11.7	2.3
% persons without a car	366	6.1	57.2	21.1	9.9
% persons in owner occupied households	366	22.0	90.7	72.4	10.0
% persons in housing rented from local authority	366	2.3	60.9	16.6	8.4
% female in each district	366	46.7	54.6	51.3	0.8
% persons aged 0-17	366	7.8	28.6	22.5	2.0
% persons aged 20-29	366	9.3	25.1	14.8	2.4
% persons aged 45-59	366	13.3	26.2	17.3	1.5
% persons aged 60 and over	366	13.1	37.0	21.1	3.6
% persons unemployed (of total population)	366	1.3	10.6	4.0	1.5
% married	366	30.8	56.0	48.5	4.2
% persons in detached/semi-detached or terraced housing	366	1.9	97.7	87.7	13.1
% persons in hh with over 1.5 persons per room	366	0.1	13.1	0.7	1.0
% persons non-white	366	0.0	44.9	4.0	6.8
% persons born in the UK	366	57.9	99.0	93.9	6.4
total persons in households	366	1937.0	948415.0	126604.8	89358.0
% persons in unsuitable accommodation	366	0.0	2.4	0.3	0.4
% of hh with no earners	366	19.2	48.7	34.1	5.8
% of lone parent families	366	7.4	30.1	12.1	4.0
Valid N (listwise)	366				

TABLE 2: Correlations bivariate

Correlations

		ln_llti	age60p_centered	unemp_centered	density_centered	unsuit_centered	lpfam_centered	unemp_lpfam_centered
Pearson Correlation	ln_llti	1.000	.415	.572	.123	.087	.469	.168
	age60p_centered	.415	1.000	-.225	-.197	.100	-.224	-.188
	unemp_centered	.572	-.225	1.000	.579	.388	.907	.668
	density_centered	.123	-.197	.579	1.000	.502	.635	.590
	unsuit_centered	.087	.100	.388	.502	1.000	.499	.433
	lpfam_centered	.469	-.224	.907	.635	.499	1.000	.745
	unemp_lpfam_centered	.168	-.188	.668	.590	.433	.745	1.000
	unemp_unsuit_centered	-.028	-.167	.408	.485	.691	.495	.679
Sig. (1-tailed)	ln_llti		.000	.000	.009	.048	.000	.001
	age60p_centered	.000		.000	.000	.029	.000	.000
	unemp_centered	.000	.000		.000	.000	.000	.000
	density_centered	.009	.000	.000		.000	.000	.000
	unsuit_centered	.048	.029	.000	.000		.000	.000
	lpfam_centered	.000	.000	.000	.000	.000		.000
	unemp_lpfam_centered	.001	.000	.000	.000	.000	.000	
	unemp_unsuit_centered	.299	.001	.000	.000	.000	.000	.000

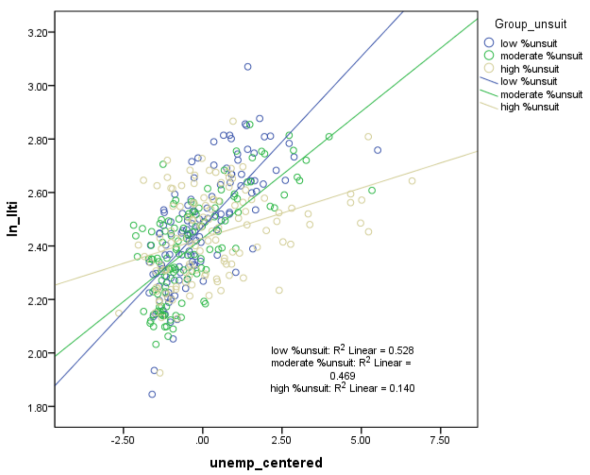
Figure Box Plot



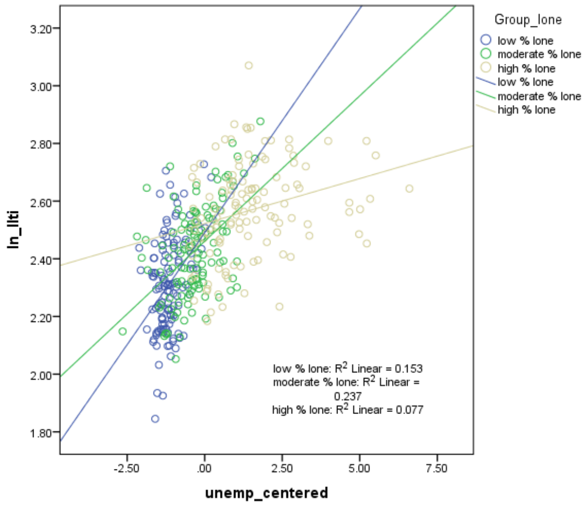
<u>Interaction Term Figure:</u>

We have categorized each regression in three degree/level (%) of: (1) Unsuitable accommodation and (2) Lone parents families. Each line represent one group from low (which represent low percentage of lone/unsuit on ward) to high (high level of percentage of lone/unsuit on ward).

Graph 1. INT 1.



Graph 2. INT 2.

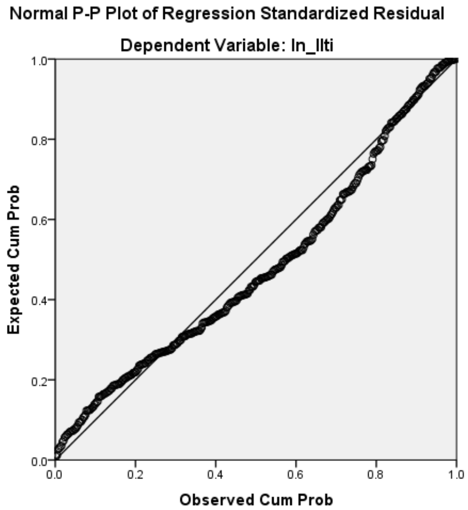


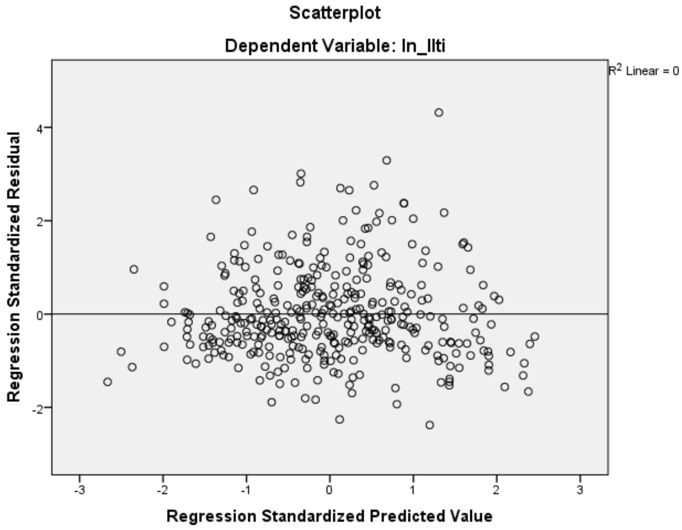
The same interpretation we can apply to graph 2: the nature of the relationship between llti and unemployed changes as a function of the % lone parental population of the district.

Descriptive Statistics

	Mean	Std. Deviation	N
ln_llti	2.4370	.19455	366
age60p_centered	.0000	3.62806	366
unemp_centered	.0000	1.46508	366
density_centered	.0000	.98567	366
unsuit_centered	.0000	.41299	366
lpfam_centered	.0000	4.00017	366
unemp_lpfam_centered	5.2990	13.48609	366
unemp_unsuit_centered	.2340	1.21819	366

Homoscedasticity: Scatterplot figure





showed R^2 is 0 with variance constant of the residual

APPENDIX B: Binary Logistic Regression.

Table Model Int(3). Variables in the equation

Model Int(3)	B	S.E.	Wald	df	Sig.	Exp(B)	EXP(B) Lower	Upper
age	.060	.008	51.110	1	.000	1.062	1.045	1.080
sex(1)	.030	.154	.038	1	.845	1.031	.763	1.393
mstatus			14.871	3	.002			
mstatus(1)	-.740	.251	8.720	1	.003	.477	.292	.780
mstatus(2)	.201	.355	.322	1	.570	1.223	.610	2.451
mstatus(3)	-.611	.344	3.164	1	.075	.543	.277	1.064
MEG(1)	-.277	.488	.323	1	.570	.758	.291	1.973
econprim			33.352	8	.000			
econprim(1)	.099	.392	.064	1	.800	1.104	.513	2.379
econprim(2)	.689	.557	1.533	1	.216	1.992	.669	5.933
econprim(3)	.155	.499	.096	1	.757	1.167	.439	3.104
econprim(4)	1.299	1.060	1.502	1	.220	3.666	.459	29.267
econprim(5)	1.164	.422	7.620	1	.006	3.201	1.401	7.313
econprim(6)	-.265	1.045	.064	1	.800	.767	.099	5.952
econprim(7)	1.089	.299	13.279	1	.000	2.972	1.654	5.338
econprim(8)	1.413	.297	22.667	1	.000	4.107	2.296	7.347
owner_rent(1)	-.002	.013	.027	1	.869	.998	.973	1.023
econprim * owner_rent	1.007	.633	2.532	1	.112	2.737	.792	9.457
econprim(1) by owner_rent(1)			2.234	8	.973			
econprim(2) by owner_rent(1)	-18.352	####	.000	1	.998	.000	0.000	
econprim(3) by owner_rent(1)	-19.155	####	.000	1	.999	.000	0.000	
econprim(4) by owner_rent(1)	.675	1.215	.308	1	.579	1.963	.181	21.253
econprim(5) by owner_rent(1)	.092	1.603	.003	1	.954	1.096	.047	25.376
econprim(6) by owner_rent(1)	.214	.611	.123	1	.726	1.239	.374	4.101
econprim(7) by owner_rent(1)	1.224	1.313	.869	1	.351	3.399	.260	44.524
econprim(8) by owner_rent(1)	-.600	.688	.761	1	.383	.549	.142	2.114
econprim * mstatus * owner_rent	-.127	.571	.050	1	.824	.881	.288	2.697
econprim(1) by mstatus(1) by owner_rent(1)			14.987	18	.663			
econprim(1) by mstatus(2) by owner_rent(1)	17.279	####	.000	1	.998	31936929.160	0.000	

(Continue)

Model 4 Lon	B	S.E.	Wald	df	Sig.	Exp(B)	EXP(B) Lower	EXP(B) Upper
econprim(1) by mstatus(2) by owner_rent(1)	-1.222	13231.125	.000	1	1.000	.295	0.000	
econprim(1) by mstatus(3) by owner_rent(1)	19.579	8918.294	.000	1	.998	#####	0.000	
econprim(3) by mstatus(1) by owner_rent(1)	-18.710	11438.350	.000	1	.999	.000	0.000	
econprim(3) by mstatus(2) by owner_rent(1)	-20.202	15605.054	.000	1	.999	.000	0.000	
econprim(4) by mstatus(1) by owner_rent(1)	-19.595	17516.172	.000	1	.999	.000	0.000	
econprim(4) by mstatus(2) by owner_rent(1)	-20.542	40192.970	.000	1	1.000	.000	0.000	
econprim(5) by mstatus(1) by owner_rent(1)	-2.215	1.114	3.954	1	.047	.109	.012	.969
econprim(5) by mstatus(2) by owner_rent(1)	-2.141	1.164	3.382	1	.066	.118	.012	1.151
econprim(5) by mstatus(3) by owner_rent(1)	21.462	40192.969	.000	1	1.000	#####	0.000	
econprim(6) by mstatus(1) by owner_rent(1)	1.185	1.369	.750	1	.386	3.272	.224	47.852
econprim(6) by mstatus(2) by owner_rent(1)	-19.535	40192.970	.000	1	1.000	.000	0.000	
econprim(7) by mstatus(1) by owner_rent(1)	.774	.594	1.695	1	.193	2.168	.676	6.949
econprim(7) by mstatus(2) by owner_rent(1)	.547	.820	.444	1	.505	1.727	.346	8.621
econprim(7) by mstatus(3) by owner_rent(1)	.499	.660	.572	1	.449	1.647	.452	6.004
econprim(8) by mstatus(1) by owner_rent(1)	-.154	.578	.071	1	.791	.858	.276	2.664
econprim(8) by mstatus(2) by owner_rent(1)	-1.319	.836	2.493	1	.114	.267	.052	1.375
econprim(8) by mstatus(3) by owner_rent(1)	-.239	.815	.086	1	.769	.787	.159	3.889
Lon_noLon(1)	-.328	.227	2.079	1	.149	.721	.462	1.125

Summary Model

Models	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1784.464[a]	.110	.247
Int (1)	1748.809[a]	.118	.266
Int (2)	1719.887[a]	.125	.281
Int (3)	1717.660[a]	.126	.282

Significance of the coefficient for the interaction term.

.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Model Int(1)	Step	35.654	9	.000
	Block	35.654	9	.000
	Model	475.096	23	.000
Model Int(2)	Step	28.923	18	.049
	Block	28.923	18	.049
	Model	504.019	41	.000

**Model Int(3): adding London_noLondon was not significantly different from 0 at 5% significance*

Multicollinearity assumption checking:

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficient	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-.041	.018		-2.206	.027		
	Marital status	.028	.005	.089	5.796	.000	.955	1.047
	Sex	-.039	.010	-.059	-3.820	.000	.932	1.073
	Tenure of household space	.014	.002	.097	6.379	.000	.982	1.019
	Economic position (primary)	.027	.001	.302	19.209	.000	.908	1.101

a. Dependent Variable: Limiting long-term illness

APPENDIX C: Index of Deprivation and Analysis Cluster

Correlations

	ns with llti in district	person s without	person s in owner	person s in housin	female in each district	person s unempl	% married	person s in detach	person s in hh with	person s non-white	person s born in the	person s in househ	person s in unsuita	% of hh with no earners	lone parent familie	AGE	person s aged 60 and	person s aged 0-17
%persons with llti in district		.657**	-.293**	.419**	.443**	.573**	-.260**	-.062	.111*	.011	.166**	.240**	.070	.892**	.466**	.637**	.387**	-.063
% persons without a car			-.743**	.735**	.367**	.886**	-.809**	-.566**	.572**	.504**	-.395**	.445**	.385**	.603**	.896**	.277**	-.092	.093
% persons in owner occupied				-.843**	-.175**	-.678**	.708**	.667**	-.597**	-.471**	.508**	-.231**	-.406**	-.309**	-.735**	-.097	.145**	-.002
% persons in housing rented					.054	.739**	-.615**	-.404**	.450**	.367**	-.245**	.308**	.112*	.344**	.707**	.064	-.270**	.274**
% female in each district						.283**	-.333**	-.320**	.151**	.109*	-.160**	.176**	.491**	.675**	.383**	.598**	.583**	-.338**
% persons unemployed (of							-.815**	-.525**	.579**	.578**	-.430**	.490**	.388**	.551**	.907**	.194**	-.225**	.246**
% married								.757**	-.713**	-.732**	.718**	-.472**	-.598**	-.274**	-.925**	-.023	.305**	-.092
% persons in detached/semi-									-.751**	-.616**	.781**	-.189**	-.713**	-.162**	-.643**	-.056	.017	.378**
% persons in hh with over 1.5										.783**	-.746**	.272**	.502**	.164**	.635**	.060	-.197**	.070
% persons non-white											-.875**	.438**	.457**	.043	.639**	-.043	-.326**	.154**
% persons born in the UK												-.289**	-.618**	.092	-.549**	.145**	.270**	.109*
total persons in households													.093	.227**	.468**	.072	-.188**	.203**
% persons in unsuitable														.205**	.499**	.145**	.100	-.349**
% of hh with no earners															.464**	.823**	.615**	-.178**
% of lone parent families																.156**	-.224**	.168**
AGE																	.817**	-.364**
% persons aged 60 and																		-.670**

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin		.766
Bartlett's Test of Sphericity	Approx. Chi-Square	1367.968
	df	10
	Sig.	0.000

%persons without a car							
Q1	Median	Q3	g	Lower	Upper	Highest	Lowest
13.1911		26.2286	2.20	**-15.4914**	**54.91113**	53.87	6.08
						52.19	6.51
	Q3-Q1	13.0375				51.61	7.70
	g'	28.68251				50.39	7.77
						48.95	8.82

llmit							
Q1	Median	Q3	g	Lower	Upper	Highest	Lowest
9.9621		13.0297	2.20	**3.213298**	**19.77848**	17.76	6.33
						17.58	6.86
	Q3-Q1	3.067627				17.40	6.92
	g'	6.748779				17.36	7.63
						17.29	7.79

%persons in housing rented from local authority							
Q1	Median	Q3	g	Lower	Upper	Highest	Lowest
11.1274		20.1491	1.50	**-2.40501**	**33.68149**	51.32	2.29
						48.86	2.97
	Q3-Q1	9.021625				46.66	3.07
	g'	13.53244				43.34	3.09
						40.24	3.09

%persons unemployed (of total population)							
Q1	Median	Q3	g	Lower	Upper	Highest	Lowest
2.8828		4.7444	2.20	**-1.21272**	**8.83992**	10.59	1.81
						9.50	1.88
	Q3-Q1	1.8616				9.21	1.93
	g'	4.09552				9.21	1.96
						9.21	2.09

%persons in hh with over 1.5 persons per room							
Q1	Median	Q3	g	Lower	Upper	Highest	Lowest
.3427		.7274	2.20	**-0.50364**	**1.57374**	13.15	.11
						5.15	.13
	Q3-Q1	0.3847				5.02	.14
	g'	0.84634				4.53	.15
						4.23	.16

%persons in unsuitable accommodation							
Q1	Median	Q3	g	Lower	Upper	Highest	Lowest
.08000		.36000	2.20	**-0.536**	**0.976**	2.390	0.000
						2.260	.010
	Q3-Q1	0.28				2.180	.010
	g'	0.616				2.160	.010
						2.060	.010[a]

lone parent						
Q1	Median	Q3	g	Lower	Upper	Highest
9.4475		13.6325	2.20	**0.240628**	**22.83941**	30.090
						30.060
	Q3-Q1	4.18496				28.670
	g'	9.206912				28.590
						27.960

Table Labeling identifying Outliers.

WARD'S METHOD CLUSTER ANALYSIS OUTPUTS

Cluster centroids

Descriptive Statistics	GROUP 1					GROUP 2					GROUP 3				
CLU4_1	N	max	min	Mean	Std. Dv.	N	max	min	Mean	Std. Dv.	N	max	min	Mean	Std. Dv.
% persons without a car	242	15.539	6.0825	**30.913**	4.4606	106	29.65	16.947	**48.954**	6.7285	13	45.301	34.003	**53.87**	6.3378
% persons in housing rented from local authority	242	12.466	2.2924	**29.06**	4.1108	106	23.473	8.8448	**43.342**	7.105	13	31.305	15.902	**51.318**	12.19
% persons unemployed (of total population)	242	3.2012	1.8057	**4.9222**	0.6305	106	5.2508	3.2369	**9.5025**	1.0199	13	7.8607	5.9901	**10.587**	1.3956
% persons in hh with over 1.5 persons per room	242	0.4743	0.1125	**1.9095**	0.2746	106	0.8814	0.1356	**2.7561**	0.6097	13	3.6801	2.1772	**5.1459**	0.9211
% persons in unsuitable accommodation	242	0.2521	0.01	**1.6**	0.2513	106	0.2947	0	**1.89**	0.3672	13	1.7946	0.86	**2.39**	0.4889

MULTILEVEL MODELLING REFERENCE:

Unhealthy people or Unhealthy place?

Study as area or place level, refer to the health effect of variables that describe something about the places, and not the people who inhabit them. Mark Trammer provides a useful distinction between types of place effect, referred to as collective and contextual place effects (Mark Trammer, Dimitris Ballas, 2012). Hence, place effects can be unpacked in three different levels. At first level, the way is to determine variation in health outcomes of individual who reside in them. Secondly, involves the analyzing of contextual heterogeneity due to group variability. Finally, to deal with individual-contextual interaction such as, social capital or income inequality cause effect in different population groups. In conclusion, the matter is 'who you are depends upon where you are' rather than 'who you are in relation to where you'. Multilevel methods will help us to solve the issues outlined above, anticipating determinants of health inequalities affect in different levels simultaneously, from individual, to the state.

References

Berkman L, Kawachi I (2000): eds. Social Epidemiology. New York, Oxford University Press

https://campus.fsu.edu/bbcswebdav/institution/academic/social_sciences/sociology/Reading%20Lists/Mental%20Health%20Readings/berkman%26kawachi_Historical%20Framework.pdf

Cadum E, Costa F, Biggeri A, Martuzzi M. (1999): Deprivazione e mortalità: un indice di deprivazione per l'analisi delle disuguaglianze su base geografica. Epid Prev; 23: 175-187.

Carstairs V, Morris R.(1991): Deprivation and Health in Scotland. Aberdeen University Press

C.J.L. Murray,1 E.E. Gakidou, & J. Frenk (1999): Health inequalities and social group differences: what should we measure? In: World Health Organization

http://www.who.int/bulletin/archives/77(7)537.pdf

Dibb, S. (1999), "Criteria guiding segmentation implementation: Reviewing the evidence", Journal of Strategic Marketing, Vol. 7, No. 2, pp. 107-129

Field, A. (2009): Discovering Statistics using IBM SPSS Statistics. London

Tukey J.(1950) Ann. Math. Statist.; 21, 607-611.

Forrest R, Gordon D.. People and Places: a 1991 Census atlas of England.1993 SAUS. University of Bristol

Hilary Graham, Institute for Health Research, Socioeconomic Inequalities in Health in the UK: Evidence on Patterns and Determinants, 2004)

http://disability-studies.leeds.ac.uk/files/library/graham-socioeconomic-inequalities.pdf

I Kawachi, S V Subramanian, N Almeida-Filho (2002): A Glossary for health inequalities. In: J Epidemiol Community Health 2002 ;56:647–652 http://jech.bmj.com/content/56/9/647.full.pdf+html

Jonathan S. Feinstein (1993): The Relationship between Socioeconomic Status and Health: A Review of the Literature. In The Milbank Quarterly, Vol. 71. No 2, pp 279-322.

http://www.jstor.org/stable/3350401

Kaplan GA (1996): People and places: contrasting perspectives on the association between social class and health. In: International Journal of Health Services, 26: 507-519

Leventhal T. & Brooks-Gunn J. (2000) The neighbourhoods they live in: the effects of neighborhood residence on child and adolescent outcomes Psychological Bulletin 126(2), 309-37.

Mark Tranmer, Dimitris Ballas(2012): Happy People or Happy Places? A Multilevel Modeling Approach to the Analysis of Happiness and Well-Being. In International Regional Science Review 2012 35: 70

http://irx.sagepub.com/content/35/1/70.full.pdf+html

Marmot M, Kogevinas M, Elston MA. Social/Economic Status and Disease. Annual Review of Public Health, 1987,8:111–35

Murray, Gakidou, & Frenk (1999): Health inequalities and social inequalities in health In vol. 77. In Bulletin.

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2560692/pdf/10743295.pdf

Townsed, Phillimore, Beattie (1988): Health and deprivation: inequality and the North. London Croom Helm

Van Der Laan, M., and J. Robins, (2003): Unified Methods for Censored Longitudinal Data and Causality, Springer, Berlin.