

Софийски университет „Св. Климент Охридски“

Стопански факултет

Магистърска теза

На тема:

**Приложение на финансови иконометрични
техники в моделиране на човешкото щастие**

Изготвена от:

Научен ръководител:

Петър Филипов

Доц. д-р Ралица Ганева

Дата 11.02.2014

СЪДЪРЖАНИЕ

Увод	5
1 Кратък преглед на теорията и емпириката	8
1.1 „Проучването на Световните Ценности" (World Values Survey) и мястото му в социално-икономическите анализи.	8
1.2 Емпирични анализи, изследващи от какво зависи щастието на отделните индивиди.	9
1.3 Статистически модели за измерване на вероятността за настъпване на дадено събитие. Скоринги. Приложение.....	10
1.4 Алтернативни методи за калкулиране на скоринги.....	11
1.4.1 Decision Trees („Дърво на решенията“).	12
1.4.2 Neural Networks („Невронни Мрежи“)	12
1.4.3 Random Forest („Случайни дървета“).	13
1.4.4 Monte Carlo („Монте Карло“).	13
1.4.5 Logistic regression („Логистична регресия“).	14
2 Методологическа рамка на емпиричния анализ.....	15
2.1 Избор на целева променлива.....	15
2.2 Алгоритъм за подбор на фактори, определящи щастието.	15
2.2.1 Първоначален подбор.	16
2.2.2 Основните параметри, използвани за сравняване на факторите.	16
2.2.3 Примери и правила за сравняване на фактори.	18
2.3 Спецификация на регресионния модел.	20
2.3.1 Логистична регресия.	20

2.3.2	Стъпков метод за избиране на оптимална регресия.	21
2.3.3	Изчисляване на показателите за оценка на модела.....	23
2.4	Корелационен анализ.....	25
2.5	Допълване на популацията с отхвърлените възможности на целевата променлива („ <i>Reject Inference</i> “).	25

3 Създаване на скоринг модел по подобрите данни27

3.1	Описание на данните, въз основа на които е проведено настоящото изследване.	27
3.1.1	Обхват на наблюдението.	27
3.1.2	Описание на извадката и статистически анализ за подбор на факторите, определящи щастието.....	27
3.1.3	Подбор на фактори.....	28
3.2	Описание на регресионния анализ.	30
3.2.1	Резултати от финалната регресия.	30
3.2.2	Променливи във финалната регресия. Точки за всяка променлива.	32
3.3	Описание на корелационния анализ.....	36
3.4	Основни резултати и изводи.	37
3.4.1	Резултати на база избраната извадка.	37
3.4.2	Резултати на база цялата популация.....	38
3.4.3	Сравнение с класацията на „Световното проучване на ценностите".....	40

Заклучение42

Библиография.....44

Приложения46

A1	РАЗПРЕДЕЛЕНИЕ НА УЧАСТНИЦИТЕ ПО ДЪРЖАВИ.	46
A2	ОСНОВНИ ПАРАМЕТРИ НА ФАКТОРИТЕ.....	48

Б1 ПЪЛЕН СПИСЪК НА ФАКТОРИТЕ.....	51
Б2 СЪКРАТЕН СПИСЪК НА ФАКТОРИТЕ.....	61
Б3 ФИНАЛЕН СПИСЪК НА ФАКТОРИТЕ.....	67
Б4 КРАЙНО ГРУПИРАНЕ НА ФАКТОРИТЕ.....	70
С1 КОРЕЛАЦИОННА МАТРИЦА НА ФАКТОРИТЕ.....	71

Увод

„Проучването на Световните Ценности¹“ представлява провеждане на анкети в редица държави в световен мащаб, чрез които се събира информация за това как живеят и мислят хората. За момента има проведени пет вълни от анкети, първата през 1981 година, а последната, от 2005 до 2008 година, съдържаща над 300 въпроса². Въпросите покриват личните разбирания на анкетираните по няколко основни теми - деца, общество, природа, равенство, свобода, наука, политика, религия, работа, и световни проблеми. Отговорите варират от няколкостепенни възможности до избор между „да“ и „не“. Богатото съдържание на тази публично достъпна база данни прави възможно анализирането на редица актуални за обществото проблеми.

Фокусът на настоящото изследване е свързан с включения в инструментариума на изследването въпрос, доколко хората се чувстват щастливи („*Feeling of happiness*“).

Основната цел на настоящата магистърска теза е да идентифицира кои са ключовите фактори, които в най-голяма степен определят дали хората се чувстват щастливи или нещастни. В допълнение, с помощта на статистически и иконометрични анализи, е потърсен отговор на въпроса „*Кои са факторите от които зависи човешкото щастие, и по какъв начин му влияят те?*“. Въз основа на емпиричните резултати ще се предложи подредба на страните, включени в последната вълна на изследването, в класация. В последствие ще се сравнят резултатите с други подобни изследвания.

¹ Превод от „*World Values Survey*“.

² Въпросите могат да бъдат подробно разгледани в „*Integrated Questionnaire*“, създадено за целта от организаторите на изследването

Така поставената цел ще се постигне с формулирането и решаването на следните изследователски задачи:

1. Преглед на теорията и проведените до момента проучвания, свързани с щастието.
2. Статистически анализ на емпиричните резултати от изследването.
3. Избор на методологическа рамка за провеждане на настоящото проучване.
4. Селекция на ключови фактори, повлияли нашата целева променлива.
5. Регресионен анализ и избор на подходящ модел.
6. Класиране на държавите посредством резултатите от нашия модел.
7. Сравнение с други подобни класации на човешкото щастие.

В нашите анализи ще приложим финансови методи, широко разпространени в кредитния бизнес при подбор и сегментация на клиентите, или така наречените скоринг модели. В практиката те се използват главно за оценка на риска на клиентите и изграждане на стратегии за събиране на вземания. Ще разгледаме подробно теорията зад тези модели и ще опишем нейното приложение върху проблема, който разглеждаме.

За изпълнението на целта, която си поставихме, ще разгледаме индивидуално всеки въпрос от анкетата и чрез статистически и иконометрични анализи ще се опитаме да систематизираме тяхното значение към избраната от нас целева променлива, да изберем най-значимите и емпирично да анализираме как точно влияят те.

В *Глава 1* ще разгледаме какви подходи са използвани до момента в изследвания, тематично свързани с нашето. Ще изброим и разгледаме и други подобни статии с данни извлечени от „Проучването на Световните Ценности“ и ще сравним методите използвани там, с останалите съществуващи в практиката.

В *Глава 2* ще представим подробно избраната от нас методология за емпиричен анализ, ще го разгледаме стъпка по стъпка и ще изброим неговите силни и слаби места. Ще съставим план за изпълнение на практическата част от изследването и боравенето с наличните данни.

В *Глава 3* ще пристъпим към същинския анализ на данните и ще видим доколко нашите очаквания, представени чрез избора ни на методи в предишните глави, отговарят на действителността. В заключение, ще извадим изводи от изследването и ще направим кратко сравнение с резултатите на подобни статии от предишни вълни на „Проучването на Световните Ценности“.

1 Кратък преглед на теорията и емпириката

1.1 Проучването на Световните Ценности („*World Values Survey*“) и мястото му в социално-икономическите анализи.

Това е глобален проект, който изследва разбиранията и ценностите на хората, как те се променят с времето и как влияят върху социалната и политическа обстановка. То се провежда от световна мрежа социолози, които от 1981 година са провели национални проучвания в над 100 страни. Световното проучване на ценностите е единственият източник на емпирични данни относно човешки възгледи, покриващо над 90% от световното население.

Изследването измерва, следи и анализира в световен мащаб подкрепата към демокрацията, толерантността към чужденци и етнически малцинства, подкрепата към половото равенство, ролята на религията и променящите се нива на вяра в хората, въздействието на глобализацията, отношението към околната среда, работата, семейството, политиката, националната идентичност, културата, икономическата несигурност и относителното удовлетворение от живота.

Анализите на база тези изследванията помагат да си обясним по-лесно местни и световни икономически и социални явления. Исторически те са били използвани от правителства, учени, студенти, журналисти и международни организации като Световната Банка и Обединените Нации. Данни от проучването на световните ценности са били използвани за обяснение на мотивите зад Арабската пролет, гражданските бунтове във Франция през 2005 и войната в Югославия през деветдесетте.

1.2 Емпирични анализи, изследващи от какво зависи щастието на отделните индивиди.

Съществуват редица проучвания, които анализират щастието под една или друга форма. Едно от по-ранните³, проведено върху данни от вълната до 1998 година образува така наречения „subjective well being“⁴ („относителното удовлетворение от живота“) като индекс за измерване на щастието. Той е образуван от комбинирането на два от въпросите в анкетата – „How satisfied are you with your life“ („Доколко сте доволни от живота си“) и „Feeling of happiness“ („Чувство за щастие“), като възможните отговори са претеглени според тяхната скала. Тъй като първият въпрос има скала от десет възможни отговора, а вторият четири, авторите са използвали следната формула за образуване на индекса.

$$SWB = \text{life satisfaction} - 2.5 * \text{happiness} \quad (1)$$

Чрез линейни регресии, използващи метода на най-малките квадрати („*Ordinary least squares*“) и вземайки образувания индекс за зависима променлива, те са доказали, че човешките гени и културни различия влияят върху това, колко щастлив се чувства човек, и доказват пряка връзка между нивото на щастие в дадена страна и нейната политическа стабилност.

В друго, по-скорошно изследване⁵ през 2008, авторите използват същия индекс и данни от всички проучвания събрани до момента, а именно от 1990 до 2008 година, и разглеждат промените на нивата на щастие през годините в участвалите държави. Те

³ Ronald Inglehart, Hans-Dieter Klingemann- „*Genes, Culture, Democracy, and Happiness. World values survey data.*“

⁴ Тъй като Световното изследване на ценностите и основните показатели, използвани в него са общоустановени в международната социално-икономическа литература, в случая са използвани английските наименования на термините, като при всяко въвеждане на нов термин в скоби е предложен български превод.

⁵ Ronald Inglehart, Roberto Foa, Christopher Peterson, Christian Welzel – „*Development, Freedom, and Rising Happiness.*“ World values survey data. 2008

доказват, че финансовото състояние не е единственият значим фактор, и обосновават преки връзки между щастието и религията, толерантността към различни малцинства, социалното ниво на демокрация и свободата на избор като цяло. Анализът отново е линейно регресионен, върху панелни данни, а за значимост на модела се разглежда коефициента „adjusted R²“.

Възможният избор на целева променлива, свързана със щастието, какво вече видяхме не е един. В друго проучване⁶ от 2008, авторите изследват странното намаляващо щастие на населението в Китай през периода 1990 - 2000, когато в страната се наблюдава икономически растеж. За зависима променлива те избират въпроса „How satisfied are you with your life“ („Доколко сте доволни от живота си“) и отново използват линейни регресии. Те доказват, че наблюдаваното явление може да се обясни от неравномерното разпределение на доходите между населението, нарастващото беззаконие, социалните лишения и политическото недоволство.

1.3 Статистически модели за измерване на вероятността за настъпване на дадено събитие. Скоринги. Приложение.

Често използван метод за контролиране на кредитния риск са така наречените кредитни скоринги или скоркарти. Скоркартата представлява статистически базиран модел, чрез който на всеки клиент се изчислява скор, под формата на число, който показва изчислената вероятност този клиент да следва определено поведение. При изчисляването на скората се използват редица източници на данни, сред които такива от формата попълнена от клиента, данни от кредитните бюра обобщаващи кредитната история на клиента, публични регистри, данни от работодател, ако има такъв, а в днешно време става все по-популярно използването на социалните мрежи и всякаква публично достъпна информация от интернет.

⁶ Hilke Brockmann, Jan Delhey, Christian Welzel, Hao Yuan - „*The China Puzzle: Falling Happiness in a Rising Economy*“. World Values Survey data.

Най-разпространеният вид скоринг в практиката е този, който се прилага при заявка за кредит („*application scorecard*“). Скоринга се опитва да определи вероятността от това клиента да спре да си плаща вноските, в даден интервал от време, което би довело до загуба за кредитната институция отпуснала кредита. Самата дефиниция за „лош“ клиент варира в различните кредитни институции на пазара, но най-популярна е тази при която клиент не е платил три вноски в последните 12 месеца. Самият скор, който ще получи клиента представлява натуралният логаритъм от вероятността клиента да е „лош“, според използвания скор. Нисък скор, обикновено се асоциира с висока вероятност клиента да е лош, а висок скор с по-добра категория клиенти.

Други разпространени скоринги са поведенческите („*behavioral scorecard*“), които предвиждат вероятността вече съществуващ клиент да спре плащанията си; скоринги изчисляващи вероятността един клиент да закупи нов продукт, ако му бъде предложен („*propensity scorecards*“); скоринги използвани за събиране на вземанията („*collection scorecards*“).

На база тези кредитни скоринги, се изграждат кредитни стратегии относно допускането на кредити за нови клиенти, образува се ниво под което кредити не се отпускат. Клиентите се сегментират и се третираат по най-подходящия за всеки сегмент начин.

Целият този процес допринася значително за автоматизацията на вземане на решения в кредитните институции и съответно бързината и точността на вземане на решения. Целта е ръчни решения да се взимат само в случаи, когато няма достатъчно информация или около границите поставени на база скоринга.

1.4 Алтернативни методи за калкулиране на скоринги.

От разгледаните изследвания в *Глава 1* се вижда, че **линейната регресия** е основният избор на аналитичен метод при така избраните целеви променливи. За основна задача на нашето проучване сме си поставили, не само изследване на връзките между отделните фактори върху щастieto, но и изграждане на точкова система, за сравнение на

отделните участници и държави. Нека разгледаме накратко някои от по-разпространените и използвани методи за оценяване, съществуващи в практиката.

1.4.1 Decision Trees („Дърво на решенията“).

Това е една от техниките са класифициране използване в изработването на скоринги.⁷ Тя е известна и като „*Recursive Partitioning*“ и „*Classification and Regression Trees*“ (CART). За пръв път е използвана в Харвардския университет през 1961 година. Това е един непараметричен метод за анализ на целевата променлива като функция от обясняващи характеристики. При тази техниката се изгражда дихотомно дърво от възли, като при всеки възел популацията се разделя на под-популации на базата на функцията на една от променливите. Системата взема предвид всички възможно деления и избира най-доброто, което се получава при минимална грешка на дискриминиране възможните стойности на целевата променлива. Процесът продължава докато във всеки възел не останат записи само с една от възможните стойности на целевата променлива, или докато по-нататъчно разделяне не е възможно. Тази техника се използва главно за намиране на най-силните променливи, които да се използват в регресионния анализ.

1.4.2 Neural Networks („Невронни Мрежи“).

Невронните мрежи са математическа репрезентация, вдъхновена от човешкия мозък и неговата способност да се адаптира на базата на получена нова информация. Това е нелинеен инструмент за оптимизиране. Представлява една многослойна система която съдържа „вход“, един или няколко скрити слоя за трениране и „изход“. Всеки от слоевете от своя страна съдържа неврони, които приемат информацията подадена през входа, преобразуват я, и я предават към следващия слой. Невроните във слоя „вход“ представляват стойността на всяка от потенциалните обясняващи характеристики. Във слоевете за трениране, тези начални стойности добиват тежести, изчислени чрез различни функции. Със всеки къс нова подадена информация, тези функции се променят, така, че

⁷ Martin Vojtek – „*Credit Scoring Methods*“, 2005

резултата при слоя „изход“ да е оптимален. Този метод обикновено дава по-добри резултати от класическите регресионни подходи но има и своите недостатъци. Обикновено резултатите са трудно обясними, защото функциите в слоевете за трениране остават черна кутия, и така логиката как и кога са били достигнати не е достъпна.

1.4.3 Random Forest („Случайни дървета“).

„Random Forest“⁸ е техника, базирана на така наречения „Bootstrapping“. „Bootstrapping“ представлява процес на генериране на извадка от цялата популация, като в нея се взимат само някои от всичките съществуващи записи и се повтарят n на брой пъти, за сметка на тези които не са били взети, така че да се постигне същата обща бройка записи. При „Random Forest“ се създават „ k “ на брой такива извадки, наречени дървета, като всяко от тях съдържа „ j “ на брой характеристики от всички възможни в извадката. След това се разглежда как всяко дърво поотделно дискриминира целевата характеристика, по подобен начин на вече описания при „Decision Trees“. За краен резултат се взимат предвид агрегираните стойности от всички дървета. Съответно, колкото повече дървета се създадат, толкова по-добър и силно дискриминиращ краен резултат ще получим. Оттук идва и основният недостатък на тази техника, а именно прекаленото нагаждане към началната ни популация („overfitting“). Целта тук е да се намери оптималният брой дървета, при който моделът същевременно да не е слаб, и да не е преувеличено добър.

1.4.4 Monte Carlo⁹ („Монте Карло“).

Това е друг метод който би могъл да се използва за подбирането на най-силните променливи. Той използва техника подобна на „Bootstrapping“, като генерира нови

⁸ Gerhard Arminger – „Expert Systems with Applications“. 2013

⁹ Christian P. Robert, George Casella – „Introducing Monte Carlo Method with R“. University Paris Dauphine, University of Florida. 2009

извадки записи от цялата популация, с разликата, че тук новите извадки съдържат всички променливи, но предварително задаваме разпределение, което искаме да следва всяка една от тях. Това означава, че те може напълно да се различават от началната извадка. Отново, колкото повече извадки се генерират, толкова по-точен резултат ще получим като дискриминационна сила на променливите. Естествено, съществува и граница, след която, приносът на нови извадки е пренебрежително малък.

1.4.5 Logistic regression („Логистична регресия“).

Основно се различава от линейната регресия по това, че целевата променлива е дихотомна, със стойности „0“ и „1“. Логистичната регресия трансформира крайния резултат, като го преобразува от стойности между $[-\infty, +\infty]$ на вероятности между $[0, 1]$. Съществуват и така наречените „*multinomial logistic regressions*“, при които целевата променлива може и да не е дихотомна. Сред предимствата на логистичната регресия са това, че характеристиките не е нужно да имат нормално разпределение, защото не се използва метода на най-малките квадрати, а този на максималната вероятност („*maximum likelihood*“) и, че резултатите са по-лесни за тълкуване.

Съществуват и редица други методи използвани при подобни анализи, като „*Genetic programming*“, „*k-nearest neighbor*“, „*expert systems*“, „*discriminant analysis*“, „*probit regressions*“.¹⁰

¹⁰ Hussein Abdou, John Pointon Credit Scoring – „*Statistical Techniques and Evaluation Criteria: A Review of the Literature.*“

2 Методологическа рамка на емпиричния анализ

2.1 Избор на целева променлива.

За максимално разграничаване на щастливите от нещастните, ще използваме само два от четирите възможни отговора на въпроса „Feeling of happiness“ („Чувство за щастие“), а именно „Not at all happy“ („Много нещастни“) и „Very happy“ („Много щастливи“). Частта от популацията отговорила по другите възможни начини „Quite happy“ („Щастливи“) и „Not very happy“ („Нещастни“) ще оставим настрана от регресионния анализ, но след като създадем оптимален модел, също ще включим в изследването, като се опитаме да изчислим как биха отговорили те, ако имаха само два отговора като възможност за избор.

2.2 Алгоритъм за подбор на фактори, определящи щастието.

Поради обема на наличните данни, за целта на анализа се налага да се направи подбор на въпросите, които да се използват в регресионния анализ. Това е важна стъпка, защото ще олесни значително избора ни на фактори.¹¹ Оттук нататък ще наричаме отделните въпроси „характеристики“ или „променливи“, защото реално те представляват белези по които се отличават участвалите в изследването.

¹¹ Практики за подбор на данни „Experian“.

2.2.1 Първоначален подбор.

За целта първо се преглежда списъка с променливите, за такива, които не биха свършили работа за нашето изследване, поради несъвършенства на данните. Това може да са например променливи които са с¹²:

- Малък брой отговорили хора, което би довело до практическа невъзможност да се извлекат статистически значими изводи за цялата популация.
- Голям процент неправилно попълнени полета.
- Прекалено много възможни отговори, съответно с малък процент отговорили на всеки от тях. Това би довело до неправилно тълкуване на резултатите.
- Въпроси, които изглеждат неадекватни за целите на нашето изследване. Трябва да се има предвид, че задаваните въпроси не са били подбрани за нашето изследване, и съответно може да има напълно незначими сред тях.

2.2.2 Основните параметри, използвани за сравняване на факторите.

След като приключим първоначалния оглед може да пристъпим към създаване на таблици с основните параметри, по които ще сравняваме наличните променливи и съответно техните възможни отговори, за по-нататъчен подбор.¹³ Самите таблици

¹² Gujarati – „*Basic Econometrics, Fourth edition*“, 2004, създаване на извадка за анализ.

¹³ Foster Provost, Tom Fawcett – „*Data Science for Business. What you need to know about data mining and data-analytic thinking*“. 2013

представяват кростабулации на всеки въпрос, с целевия въпрос, и допълнително изчислени показатели изброени по-долу. Всеки зададен въпрос в анкетите има определен брой възможни отговори, които отгук нататък ще наричаме „групи“. За целта на регресионния анализ, ние ще търсим групи с подобни параметри и допълнително ще ги групираме в по-общи групи, когато те са налице. В практиката съществуват редица показатели по които биха могли да се агрегират групите на всяка променлива. Тези, които ще вземем под внимание са:

- **„Unhappy Rate“** - Това представлява отношението на отговорилите „Много нещастен“ към тези отговорили „Много щастлив“, за всяка група на дадена характеристика.
- **„Odds“** - Това е отношението на отговорилите „Много щастлив“ към тези отговорили „Много нещастен“ за всяка група, или с други думи, обратното на „Unhappy rate“, дефиниран по горе.
- **„Index“** - Това е по-сложен показател, който се използва при агрегиране на групите, със следната формула:
(Odds на дадена група / Total Odds за характеристиката)*100, със знак „+“
ако Odds на дадена група > Total Odds за характеристиката
или
(Total Odds за характеристиката / Odds на дадена група)*100, със знак „-“
ако Odds на дадена група < Total Odds за характеристиката

Накратко, положителните стойности на този показател ще показват групи с по-голям дял на „нещастните“, а тези с отрицателен знак - с по-голям дял на „щастливите“.

- **„Weight of Evidence“** с формула:

$$\ln\left(\frac{EventRate}{NonEventRate}\right) \quad (2)$$

където „Event Rate“ е делът на отговорилите „Много щастлив“ за дадена група към общият брой отговорили „Много щастлив“, а „Non Event Rate“ е делът отговорили „Много нещастен“ за дадена група, към общият брой отговорили така за популацията.

„Index“ и „Weight of Evidence“ са алтернативни измервания на отношението щастливи и нещастни за всяка група.

- **„Information Value“** с формула:

$$\sum_{i=1}^n \left[\left(\frac{g_i}{G} - \frac{b_i}{B} \right) \cdot \log\left(\frac{g_i \cdot B}{b_i \cdot G} \right) \right] \quad (3)$$

където „g“ и „b“ са броят отговорили съответно „Много щастлив“ и „Много нещастен“ за всяка група, а „G“ и „B“ са общият брой отговорили съответно по такъв начин, за популацията.

„Information Value“ е показател който се използва за преценяване на цялостната сила на една характеристика спрямо друга.

2.2.3 Примери и правила за сравняване на фактори.

При първоначалното създаване на таблиците, характеристиките от тип „категория“, т.е. тези с не-числови отговори, обикновено се оставят с всичките си възможни групи използвани в анкетата, без те да се агрегират предварително. Например, за характеристиката „Образование“, може да имаме групи „Университет“, „Гимназия“, „Начално училище“ и т.н..

Характеристиките с числови отговори, предварително се групират в по-малък брой групи, за да се улесни изследването им. Обикновено популацията се разделя на 10 или 20

равни групи съответно с 10% или 5% от популацията във всяка група.¹⁴ Класовете в тези групи може да бъдат закръглени до стотици или хиляди, например ако става въпрос за заплата, за да бъдат по-разбираеми, независимо, че това би променило равните групи.

Следващото групиране в характеристиките става на база получените показатели за всяка от техните групи. Това се прави с цел да се избегне прекомерното нагаждане на модела, който искаме да създадем, към дадената популация, и да го направи валиден за анализ на бъдещи периоди. Освен това, така моделът ни ще бъде и по-гъвкав към малки промени в данните, които са възможни да настъпят с течение на времето, поради промени в живота и разбиранията на хората. Обикновено се следват следните общи правила при по-грубото групиране в характеристиките¹⁵:

- Интервалите трябва да представляват значителен процент от цялата популация, за да се избегне нагаждане. Важи за всички характеристики. Обикновено минимумът за всяка група е 5% от цялата популация.
- Интервали с близки стойности на „Unhappy Rate“ или „Index“ се групират заедно. Колко трябва да са близки стойностите, зависи от средните стойности за популацията.
- „Unhappy Rate“ и „Weight of Evidence“ трябва да следват логичен тренд през групите, за числовите характеристики. Би било нелогично да има обрати в средните интервали.
- Характеристики от тип „категория“ трябва да бъдат групирани единствено на логични принципи, независимо от техните показатели. Например, свързването на групи „Начално училище“ и „Университет“ за Образование, в една група би било

¹⁴Newsom – „Data Analysis“, 2012

¹⁵ Тези правила са следвани от редица водещи скоринг и кредитни компании като „UniCredit“ и „Experian“.

нелогично, защото това са профили на коренно различни хора. Ако все пак техните показатели са много близки, по-добър подход би бил, тази характеристики въобще да не се разглеждат по-нататък.

2.3 Спецификация на регресионния модел.

От възможните методи за регресионен анализ, за нашият модел, ще използваме логистична регресия. Целта на този анализ е да оптимизираме силата на потенциалните модели и да изберем най-подходящия. За да постигнем това, трябва да изберем модел, който най-силно да дискриминира между хора отговорили „Много нещастни“ и „Много щастливи“. След като изберем нашия модел, ще изчислим бета коефициентите на всяка характеристика, и чрез тях ще формираме скоринг на участниците, чрез който ще можем да правим сравнения и да извлечем изводи от проучването.

2.3.1 Логистична регресия.

При логистичната регресия, целевата променлива „Y“ следва Биномно разпределение с възможни стойности „0“ и „1“. Вероятността „Y“ да е „1“ зависи от избраните от нас регресори

x_i ($i=1, 2, \dots, n$) :

$$P(Y = 1 | X = x) = \pi(x). \quad (4)$$

Процедурата за изчисляване на вероятността е базирана на отношението между вероятността едно събитие да се случи и вероятността това събитие да не се случи:

$$odds(x) = \frac{P(Y = 1 | X = x)}{P(Y = 0 | X = x)} = \frac{P(Y = 1 | X = x)}{1 - P(Y = 1 | X = x)} = \frac{\pi(x)}{1 - \pi(x)}. \quad (5)$$

Натуралният логаритъм на отношението е линейна функция на регресорите „ x_i “ ($i=1, 2, \dots, n$):

$$\ln[\text{odds}(x)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n. \quad (6)$$

Като свържем формули (5) и (6) и приравним с „ $\pi(x)$ “, логистичната функция на вероятността едно събитие да се случи е:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}. \quad (7)$$

Логистичната регресия използва метода на максималната вероятност („*maximum likelihood*“) за изчисляването на бета коефициентите на регресорите.¹⁶

2.3.2 Стъпков метод за избиране на оптимална регресия.

За избирането на оптимална комбинация от обясняващи характеристики ще използваме постъпкова логистична регресия, с постепенно прибавяне на потенциалните регресори. Кратко описание на метода:

- Постъпковата логистична регресия е изработена, така, че да намира комбинацията от променливи, които най-ефективно да разграничават двете възможности в целевата променлива.
- Предварително избрани характеристики се добавят в регресията една по една, като се използва статистическият критерий от „*-2 Log Likelihood*“ грешка за добавена променлива.

¹⁶ Naeem Siddiqi – „*Credit Risk Scorecards. Developing and Implementing Intelligent Credit Scorings.*“,

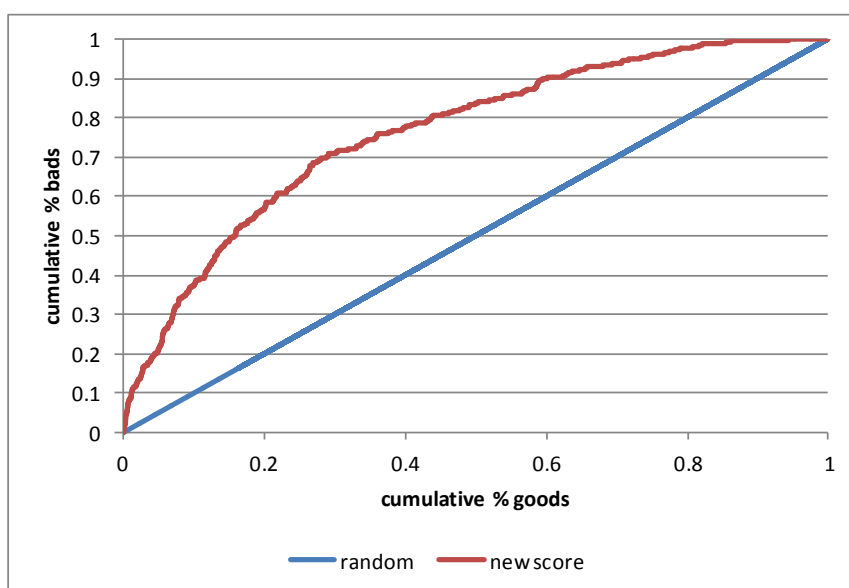
- След всяко добавяне на нова променлива в модела, се извършват автоматични тестове за проверка дали моделът не би бил по-добър ако някоя от вече влезлите в регресията променливи бъде изключена.
- Процесът на добавяне на променливи приключва когато всички налични характеристики са влезли в регресията, или когато не е възможно да бъде направено статистически значимо намаление на „-2 *Log Likelihood*“ грешката при добавяне на която и да е нова променлива.
- Нечисловите променливи (тези от тип категория) се добавят в логистичната регресия като групирани променливи с числа отбелязващи категорията.
- Редът по който влизат променливите в регресията може да бъде използван като приблизително подреждане на тяхната значимост за модела.

Стъпковата логистична регресия се използва за да се създадат модели, които са точни, защото тя премахва фактори, които не допринасят за по-дълбока дискриминация на зависимата променлива.

2.3.3 Изчисляване на показателите за оценка на модела.

Статистическите показатели които ще използваме за измерване на дискриминиращата сила на модела ни са няколко. Те са характерни за сравнението на финансови скоринги, и се използват като измерващи инструменти на добавената стойност от нов модел.

„GINI“ коефициент и „ROC“¹⁷ графика със следното обяснение¹⁸:



„ROC“ графиката представлява проста графика с кумулативните по определен скор отговорили „Много щастлив“ по абсцисата и съответно кумулативно отговорили „Много нещастен“ по ордината. „GINI“ коефициента представлява лицето на фигурата, която се получава под кривата на нашия модел, измерено в проценти. По-висок „GINI“ коефициент

¹⁷ „ROC“ е съкращение от „Receiver Operating Characteristic“.

¹⁸ David J. Hand – „Assessing and evaluation scorecards. Tools, methods, and consequences“, Imperial College and Winton Capital Management. 2013

показва по-добра дискриминираща способност на нашия модел. Модел без никаква дискриминираща сила би имал „GINI“ 0%, а перфектният модел ще бъде със „GINI“ 100%.

Формулата на самия коефициент е:

$$GINI = 1 - \sum_{i=S_1}^{S_n} (b_i - b_{i-1}) * (g_i + g_{i-1}) \quad (8)$$

, където

g_i = Кумулативният процент отговорили „Много щастлив“ за даден скор

b_i = Кумулативният процент отговорили „Много нещастен“ за даден скор

S_n =N-ти скор в сортираното разпределение на цялостния скор на популацията.

„KS“ статистика със следното обяснение:

„KS“ статистиката се използва при оценяването на силата на модела, като измерва най-голямото получено разграничаване между кумулативното разпределение на „Много нещастен“ и кумулативното разпределение на „Много щастлив“ получено от нашата регресия. Този коефициент също е лесно обясним чрез ROC графика. Неговата формула е:

$$D_n = \sup_x |F_n(x) - F(x)| \quad (9)$$

2.4 Корелационен анализ.

Корелационният анализ е нужен за да се подсигури статистическата стабилност на модела. Включването на високо корелирани характеристики в модела трябва да се избягва.¹⁹ За целта се създава корелационна матрица между избраните променливи в модела, и се прави проверка, доколко те са корелирани. Стойности над (0,5) - (0,6) и под (-0,5) - (-0,6) могат да бъдат интерпретирани като индикатор за висока корелация между две характеристики.

Анализът на корелационната матрица може също да обясни, защо променливи с висока обясняваща сила не са били избрани при постъпковата логистична регресия. Ако две или повече характеристики са високо корелирани, е малко-вероятно всички те да се окажат избрани във финалната регресионна стъпка. Ако все пак преценим, че предпочитаме някоя от тях, поради бизнес причини или проста логика, бихме могли предварително да не вкарваме в постъпковата регресия характеристики силно корелирани с нея.

2.5 Допълване на популацията с отхвърлените възможности на целевата променлива („*Reject Inference*“).

Този метод се използва в кредитирането за оценяване на хора, на които е бил отказан кредит, оттам идва и неговото име. Тъй като тази част от популацията няма реално поведение на плащане, трябва да се симулира такова. Това става, като използваме модела оценен хората със съществуващо поведение. Това допълване на популацията е нужно за да се модифицира модела, така, че да е адекватен за всички възможни профили хора.

В нашия случай, до този момент сме разгледали само частта от популацията, отговорила по избраните от нас два начина - „*Много нещастен*“ и „*Много щастлив*“. За да

¹⁹ Gujarati „*Basic Econometrics*“

може да извадим допълнителни изводи от избрания вече модел, ще видим какво биха отговорили хората според модела ни, ако съществуваха само тези две опции.²⁰

За целта ще минем през следните стъпки:

- Избираме оптимален модел с двата избрани варианта на целевата променлива.
- Прилагаме модела върху популацията, отговорила по различен начин. Изчисляваме техните вероятности да отговорят „Много нещастен“ и „Много щастлив“.

Този метод за задаване на стойност, която реално не е била такава, използваме, за да допълним извадката с профилите, които представляват хората отговорили по различен начин.²¹ Това е необходимо за да се обхване реалната картина на всички профили хора във всяка страна. Трябва да се има предвид, че метода може и да не даде очакваните резултати, защото е много зависим от дела на допълнително добавената популация, спрямо цялата популация. Принципно при по-висок дял, те биха повлияли повече на крайните резултати.

²⁰ Derek Montrichard- „*Reject Inference Methodologies in Credit Risk Modeling*.“ Canadian Imperial Bank of Commerce.

²¹ Jonathan Crook, John Banasik – „*Does Reject Inference Really Improve the Performance of Application Scoring Models?*“, 2011

3 Създаване на скоринг модел по подобрите данни

3.1 Описание на данните, въз основа на които е проведено настоящото изследване.

3.1.1 Обхват на наблюдението.

Данните в извадката, която ще използваме са от петата вълна анкети и обхващат периода от 2005 до 2008 година. Ще използваме извадката публикувана на интернет адреса на проучването, със страните, където е бил попълнен целият въпросник.²² Базата съдържа 364 полета с въпроси и допълнително извлечена от тях информация. Те могат да бъдат подробно разгледани в **приложение Б1**.

3.1.2 Описание на извадката и статистически анализ за подбор на факторите, определящи щастието.

Извадката съдържа 66 610 наблюдения от 48 държави, разпределени както е показано в **приложение А1**. Както може да се види, всяка държава е представена от поне 1 000 участника, което ще е достатъчно за нашия анализ.

²² Пълните данни са достъпни и могат да бъдат изтеглени от „www.wvsevsdb.com/wvs/WVSDData.jsp“

Избраната от нас целева променлива има следното разпределение:

Таблица 1: Разпределение на „Чувство за щастие“.

Feeling of happiness („Доколко лицето се чувства щастлив“)			
	Frequency	Percent	Cumulative Percent
Very happy („Много щастлив“)	19053	28.6	28.6
Quite happy („Щастлив“)	35947	54.0	82.6
Not very happy („Нещастен“)	9935	14.9	97.5
Not at all happy („Много нещастен“)	1675	2.5	100.0
Total	66610	100.0	

Отговорилите „Много щастлив“ и „Много нещастен“ са сумарно 20728. На базата на тези случаи ще изпълним анализа на променливи и регресионния анализ. Останалите 45882, отговорили „Щастлив“ и „Нещастен“ ще бъдат разгледани на по-късен етап, когато вече имаме избран модел.

3.1.3 Подбор на фактори.

Съгласно описаната методологическа рамка в *Глава 2*, емпиричният анализ включва подбор на факторите, които определят щастието. По-долу са представени емпиричните стъпки на проведения анализ.

3.1.3.1 Създаване на основните параметри за сравнение.

Създаваме кростаблицы на всяка променлива с целевата променлива. Резултатите могат да бъдат разгледани в **приложение А2**. Всяка страница съдържа по една променлива и изчислени за нея: „*Unhappy Rate*“, „*Odds*“, „*Index*“, „*Share*“, „*Weight of evidence*“ и „*Information Value*“, както и техни графики. В зелен цвят са отбелязани страниците на характеристики с агрегиран „*Information Value*“ коефициент по-голям от „0,1“, като потенциално по-силни характеристики.

На база на получените таблици преразглеждаме **приложение Б1**. В него имаме пълния списък променливи, с изчислени индивидуалното им влияние върху целевата променлива, показано с техните „*GINI*“ и „*Information value*“ коефициенти.

3.1.3.2 Изпълнение на първоначалния подбор.

Разглеждаме вече готовото **приложение Б1**. В жълто са отбелязани променливи които няма как да използваме в модела, поради прекалено много възможни отговори, с малък процент отговорили хора на всеки отговор. В жълто е поставена и характеристиката „*How satisfied are you with your life*“, защото тя е много близка до нашата целева променлива и няма мисъл да я разглеждаме като потенциален фактор.

В червено са отбелязани характеристики, които съществуват само за някои от страните или в определени региони. Тъй като не искам да изключваме страни от изследването, тези характеристики няма да бъдат разглеждани, защото няма да дадат нужната информация за цялата извадка.

В оранжево са отбелязани характеристики с „*GINI*“ коефициент под десет, като прекалено слаби. Тези три цвята се изключват, и така получаваме списъка в **приложение Б2**.

В **приложение Б2** имаме налични 193 характеристики, като ще се опитаме допълнително да намалим техния брой.

Подробно разглеждаме таблиците в **A2** и се опитваме да разберем дали са логични резултатите спрямо целевата променлива. В **Б2** с жълто са отбелязани променливите които са напълно нелогични спрямо нашите очаквания, за това кога би трябвало хората да се чувстват щастливи, или където липсва очевиден тренд свързан с целевата променлива. Отбелязани са и тези променливи с висок процент „*Not Asked*“ и също са изключени. Премахваме и всички останали променливи с „*GINI*“ коефициент под 15%, поради все още високия им брой. Така получаваме списък **Б3**.

3.1.3.3 Групиране по изчислените основни показатели.

Приложение Б3 съдържа всички променливи, чиито групи ще агрегираме допълнително, където е нужно, а след това ще включим в регресионния анализ. След подробно разглеждане на изчислените индикатори, и допълнително агрегиране, преминаваме към регресионния анализ. В приложение Б4 могат да бъдат намерени финалните групиране на променливите.

3.2 Описание на регресионния анализ.

За целта на регресионния анализ ще бъдат създадени набор от така наречените дъми променливи. На всяка група от променлива, участваща пряко в регресията, ще отговаря по една дъми променлива със стойности „0" и „1". Така регресията ще вземе първата група в променливата като нулева група с бета коефициент нула, и ще сравнява останалите групи спрямо нея. При финалното агрегиране на групите във всяка променлива избрахме групите с най-нешастни хора за нулеви, така че те да получават по нула точки, а групите с по-щастливи хора да получават нарастващи положителни точки.

3.2.1 Резултати от финалната регресия.

След включване на прегрупираните променливи в постъпковата логистична регресия получаваме следните резултати:

Таблица 2: Финална регресия на постъпковия метод:

	Estimate	Std. Error	z value	Pr(> z)	
Константа (<i>Intercept</i>)	-4.14	0.18	-23.52	< 2e-16	***
Удовлетвореност от финансовото положение на домакинството, група 2 / <i>Satisfaction with the f. sit. of household 2</i>	0.86	0.09	9.66	< 2e-16	***
Удовлетвореност от финансовото положение на домакинството, група 3 / <i>Satisfaction with the f. sit. of household 3</i>	1.69	0.09	18.45	< 2e-16	***

Удовлетвореност от финансовото положение на домакинството, група 4 / <i>Satisfaction with the f. sit. of household 4</i>	2.18	0.14	15.81	< 2e-16	***
Удовлетвореност от финансовото положение на домакинството, група 5 / <i>Satisfaction with the f. sit. of household 5</i>	2.49	0.11	22.14	< 2e-16	***
Състояние на здравето (субективно), група 2 / <i>State of health (subjective) 2</i>	1.56	0.09	16.84	< 2e-16	***
Състояние на здравето (субективно), група 3 / <i>State of health (subjective) 3</i>	2.48	0.10	25.74	< 2e-16	***
Състояние на здравето (субективно), група 4 / <i>State of health (subjective) 4</i>	3.33	0.12	28.85	< 2e-16	***
Чувство за свобода, група 2 / <i>How much freedom you feel 2</i>	0.65	0.10	6.30	2.90E-10	***
Чувство за свобода, група 3 / <i>How much freedom you feel 3</i>	0.99	0.11	9.39	< 2e-16	***
Чувство за свобода, група 4 / <i>How much freedom you feel 4</i>	1.28	0.10	13.46	< 2e-16	***
Уважение към индивидуалните човешки права, група 2 / <i>Respect for individual human rights 2</i>	0.38	0.10	3.85	0.00012	***
Уважение към индивидуалните човешки права, група 3 / <i>Respect for individual human rights 3</i>	0.93	0.10	9.17	< 2e-16	***
Спестявания на семейството през последната година, група 2 / <i>Family savings during past year 2</i>	0.39	0.09	4.40	1.06E-05	***
Спестявания на семейството през последната година, група 3 / <i>Family savings during past year 3</i>	0.82	0.12	6.83	8.65E-12	***
Свободно време, група 2 / <i>Leisure time 2</i>	0.41	0.08	5.09	3.66E-07	***
Свободно време, група 3 / <i>Leisure time 3</i>	0.73	0.09	8.42	< 2e-16	***
Същност на задачите: Независимост, група 2 / <i>Nature of tasks: independence 2</i>	0.40	0.09	4.52	6.14E-06	***
Значение на приятелството, група 2 / <i>Friends important 2</i>	0.39	0.09	4.37	1.23E-05	***
Значение на приятелството, група 3 / <i>Friends important 3</i>	0.63	0.09	7.15	8.56E-13	***

Significance codes:	‘***’ 0.001	‘**’ 0.01	‘*’ 0.05	‘.’ 0.1	‘ ’ 1

Променливите във финалната регресия са осем, със статистически значими коефициенти, което може да се види от „*P-value*“ на всеки коефициент. Стандартните грешки също са задоволително ниски.

Нашият модел има „*GINI*“ коефициент от 84% и „*KS*“ статистика от 69%, които са задоволително високи, имайки предвид огромния набор от данни с които разполагаме. За сравнение в кредитните скоринги добър „*GINI*“ коефициент варира в зависимост от целта на скоринга и се движи в рамките на 50-70%, а „*KS*“ статистиката в рамките на 40-60%.²³

3.2.2 Променливи във финалната регресия. Точки за всяка променлива.

Нека да разгледаме по-подробно какви бяха финалните групирания на променливите, влезли в избрания от нас модел.

Таблица 3:Променлива 1.

Удовлетвореност от финансовото положение на домакинството								
Група	Отговори	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
1	Неудовлетворен','2'	830	1226	5518	40.4	1.48	-768	11.4
2	3','4'	377	1942	9245	16.3	5.15	-221	17.4
3	'5','6'	265	4564	14530	5.5	17.22	151	29.1
4	'7'	75	3006	6907	2.4	40.08	353	15
5	8','9','Удовлетворен'	128	8315	9682	1.5	64.96	571	27.2

²³ „*Experian*“таблицы за сравнение на предиктивната способност за скоринги.

Таблица 4 :Променлива 2.

Състояние на здравето (субективно)								
Група	Отговори	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
1	Много лошо','Лошо'	673	472	3374	58.8	0.7	-1624	6.8
2	Средно'	535	2764	12960	16.2	5.17	-220	24.4
3	Добро'	325	7188	22168	4.3	22.12	195	44.6
4	Много добро'	142	8629	7380	1.6	60.77	534	24.2

Таблица 5:Променлива 3.

Чувство за свобода								
Група	Отговори	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
1	Никаква','2','3'	510	943	3910	35.1	1.85	-615	8.1
2	4','5'	400	2358	9405	14.5	5.89	-193	18.3
3	'6','7'	299	4293	13572	6.5	14.36	126	27.3
4	8','9','Много'	466	11459	18995	3.9	24.59	216	46.4

Таблица 6:Променлива 4.

Уважение към индивидуалните човешки права								
Група	Отговори	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
1	Никакво уважение'	336	1218	3036	21.6	3.62	-314	6.9
2	Не много'	806	5552	19214	12.7	6.89	-165	38.4
3	Малко','Много'	533	12283	23632	4.2	23.05	203	54.7

Таблица 7:Променлива 5.

Спестявания на семейството през последната година								
Група	Отговори	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
1	Похарчени спестявания и взет заем'	416	1581	5492	20.8	3.8	-299	11.2
2	Похарчени част от спестяванията'	1094	11039	30138	9	10.09	-113	63.5
3	Спестих пари'	165	6433	10252	2.5	38.99	343	25.3

Таблица 8:Променлива 6.

Свободно време								
Група	Отговори	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
1	Въобще не е важно', 'Не много важно'	613	3141	10707	16.3	5.12	-222	21.7
2	Доста важно'	648	7301	20998	8.2	11.27	-101	43.5
3	Изключително важно'	414	8611	14177	4.6	20.8	183	34.8

Таблица 9:Променлива 7.

Същност на задачите: Независимост								
Група	Отговори	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
1	Никаква независимост', '2', '3', '4'	337	2161	7140	13.5	6.41	-177	14.5
2	5', '6', '7', '8', '9', 'Пълна независимост',	1338	16892	38742	7.3	12.62	111	85.5

Таблица 10:Променлива 8.

Значение на приятелството								
Група	Отговори	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
1	Въобще не е важно', 'Не много важно',	455	2123	5712	17.6	4.67	-243	12.4
2	Доста важно'	612	6500	19601	8.6	10.62	-107	40.1
3	Изключително важно'	608	10430	20569	5.5	17.15	151	47.5

Нека разгледаме накратко и влиянието на обясняващите променливите върху целевата променлива:

- С подобряване на финансовата си ситуация, човек се чувства по-щастлив.
- Колкото по-здравословен живот води и съответно сегашното му здравословно състояние е по-добре, човек е по-щастлив.
- Колкото по-свободен се чувства човек да изразява мнението си и да управлява сам живота си, той е по-щастлив.

- Колкото повече уважава човешките права, човек е по-щастлив. Това е наистина интересен резултат. Той е може би по-скоро получен косвено от образованието и средата в която е израснал човек, но е избран като по-силен в обяснението дали човек е щастлив.
- По-високите спестявания през изминалата година правят човек по-щастлив. Това би спомогнало за по-оптимистичния поглед в бъдещето.
- Повече свободно време прави човек по-щастлив.
- Свободата човек да прави избор как да решава поставените му проблеми в работата го прави по-щастлив.
- И както вече беше отбелязано, колкото повече човек държи на приятелите си, толкова по-щастлив е той.

Нека разгледаме и точките, които получават променливите. За по-нагледен резултат сме образували точките като умножим бета коефициентите от регресията по сто, и закръглим получения резултат до стотици.

Таблица 11: Точки на променливите.

Удовлетвореност от финансовото положение на домакинството	
Отговори	Точки
'Неудовлетворен','2'	0.00
'3','4'	86.15
'5','6'	168.86
'7'	218.02
'8','9','Удовлетворен'	248.51
Състояние на здравето (субективно)	
Отговори	Точки
'Много лошо','Лошо'	0.00
'Средно'	156.08
'Добро'	248.31
'Много добро'	333.48
Чувство за свобода	
Отговори	Точки
'Никаква','2','3'	0.00
'4','5'	65.25
'6','7'	99.46
'8','9','Много'	128.32

Уважение към индивидуалните човешки права	
Отговори	Точки
'Никакво уважение'	0.00
'Не много'	38.12
'Малко', 'Много'	92.66
Спестявания на семейството през последната година	
Отговори	Точки
'Похарчени спестявания и взет заем'	0.00
'Похарчени част от спестяванията'	38.58
'Спестих пари'	81.71
Свободно време	
Отговори	Точки
'Въобще не е важно', 'Не много важно'	0.00
'Доста важно'	40.92
'Изключително важно'	73.26
Същност на задачите: Независимост	
Отговори	Точки
'Никаква независимост', '2', '3', '4'	0.00
'5', '6', '7', '8', '9', 'Пълна независимост',	39.90
Значение на приятелството	
Отговори	Точки
'Въобще не е важно', 'Не много важно',	0.00
'Доста важно'	38.81
'Изключително важно'	63.47
Константа	
Отговори	Точки
-	-413.71

3.3 Описание на корелационния анализ.

Ще образуваме корелационна матрица за да проверим дали постъпковия метод е подбрал високо корелирани характеристики, и ако е, да намерим начин да ги премахнем. Матрицата може да бъде намерена в **приложение С1**.

Както може да се види, прекалено силни корелации не са налице в комбинацията от характеристики, избрани в регресията. Една от по-високите налични корелации от 12% е между променливите „*Satisfaction with the financial situation of the household*” и „*How much*”

freedom you feel" което може да се обясни с това, че хората с повече средства могат да си позволят да се чувстват по-свободни.

Втората по-висока корелация, но отново в допустимите нива, посочени в *Глава 2*, е между променливите „*Leisure time*” и „*Friends importance*” и съответно колкото по-важни са приятелите за някого, толкова повече той цени своето свободно време.

3.4 Основни резултати и изводи.

3.4.1 Резултати на база избраната извадка.

След като вече имаме избран модел, можем да изчислим индивидуалните сумарни точки на всеки един участник в анкетите поотделно, като използваме таблицата с точките която вече получихме. Ще изчислим и средните точки на участниците от всяка страна и ще ги подредим в класация.

Таблица 12: Среден скор по държави участващи в петата вълна на проучването:

Ранг	Държава	Средно точки	Ранг	Държава	Средно точки
1	Норвегия	462.07	25	Тринидад и Тобаго	324.51
2	Швеция	444.92	26	Словения	324.32
3	Швейцария	440.07	27	Бразилия	323.18
4	Канада	427.82	28	Замбия	313.58
5	Финландия	409.99	29	Виетнам	309.34
6	Андора	408.04	30	Мали	309.04
7	Малайзия	407.85	31	Турция	305.33
8	Австралия	389.09	32	Япония	293.60
9	Йордания	382.67	33	Чили	290.76
10	САЩ	373.91	34	Китай	281.86
11	Кипър	373.10	35	Мароко	270.90
12	Уругвай	364.68	36	Индия	266.13
13	Тайланд	361.92	37	Етиопия	265.51
14	Тайван	360.37	38	Полша	259.51
15	Мексико	359.61	39	Буркина Фасо	253.69
16	Южна Африка	351.93	40	Перу	230.34
17	Аржентина	350.40	41	Сърбия	229.02

18	Испания	348.93	42	Румъния	205.24
19	Италия	345.99	43	Египет	197.03
20	Индонезия	340.73	44	Руанда	192.37
21	Гана	337.90	45	Украйна	190.80
22	Гватемала	331.96	46	Молдова	187.10
23	Южна Корея	330.39	47	България	159.09
24	Германия	327.24	48	Грузия	146.23

Както се вижда от таблицата, първи места заемат страните от Северна Европа и по-точно Скандинавския полуостров. Щатите, Австралия, и страни с голям дял чужди инвестиции също са в челните места.

Последните места са заети от бивши социалистически страни, в това число и България на предпоследно място, редом до страни от Африка и Южна Азия.

3.4.2 Резултати на база цялата популация.

Включваме първоначално отделените случаи с отговори „Щастлив“ и „Нещастен“. Като използваме метода, описан в Глава 2, „Допълване на популацията“, изчисляваме вероятността те да отговорят по друг начин, ако тези отговори не са възможни.

Таблица 13: Показатели при добавяне на предварително премахнатите случаи.

	Unhappy Rate	Odds	Share
Извадка на регресията	0.08	11.35	0.31
Извадка на премахнатите случаи	0.13	6.74	0.69
Общо	0.11	7.76	1.00
Коефициент	1.60	1.68	NA

Както се вижда, профилите, които предварително премахнахме са 69% от цялата популация, което означава, че тяхното добавяне може да предизвика значителни промени в крайното подреждане на държавите. Отношението на нещастни към щастливи в този сегмент е с около 5% по-голямо от това в извадката, която моделирахме. Вземайки предвид тези недостатъци, пресмятаме отново средните точки на всяка страна.

Крайното подреждане на държавите, след добавянето на хората отговорили „Щастлив“ и „Нещастен“ е посочено в таблица 14.

Таблица 14: Крайно подреждане на държавите.

Ранг	Държава	Средно точки	Ранг	Държава	Средно точки
1	Норвегия	452.33	25	Германия	319.62
2	Швеция	431.48	26	Словения	318.59
3	Швейцария	429.04	27	Бразилия	314.77
4	Канада	412.95	28	Тринидад и Тобаго	308.92
5	Финландия	401.45	29	Виетнам	301.08
6	Малайзия	401.40	30	Турция	297.86
7	Андора	399.45	31	Мали	292.76
8	Йордания	373.84	32	Япония	279.70
9	Австралия	373.00	33	Чили	279.45
10	Кипър	364.82	34	Китай	274.76
11	Уругвай	359.04	35	Мароко	253.62
12	САЩ	358.43	36	Етиопия	250.56
13	Тайван	355.30	37	Полша	249.69
14	Тайланд	347.69	38	Индия	248.37
15	Испания	343.59	39	Буркина Фасо	244.97
16	Мексико	343.24	40	Сърбия	229.09
17	Аржентина	341.54	41	Перу	224.23
18	Италия	340.20	42	Румъния	209.81
19	Южна Африка	335.11	43	Египет	192.63
20	Индонезия	332.79	44	Руанда	190.67
21	Замбия	325.23	45	Украйна	187.94
22	Южна Корея	324.36	46	Молдова	185.13
23	Гана	323.01	47	България	163.63
24	Гватемала	321.25	48	Грузия	144.73

В зелено са отбелязани страните, които запазват мястото си в класацията. В жълто са страните, които слизат надолу в класацията, а в синьо, тези, които се изкачват.

Като цяло се наблюдава спад в средните точки на държавите. Обяснимо е с това, че делът на нещастните в добавената популация е по-високо. Вижда се, че движения има

главно в средно класираните държави, което показва, че при тях е по-трудно да се определи къде точно се намират в класацията. Държавите заемащи първите и последните места остават същите.

3.4.3 Сравнение с класацията на „Световното проучване на ценностите“.

„Проучването на световните ценности“ е образувало собствена класация на държавите, подреждайки ги по различно формулиран индекс на щастие.²⁴ Класацията е направена на база информация до 2004 година, което означава, че не включва вълната от проучване на нашето изследване. Това означава, че тук има и страни, които не присъстват в нашата класация. Като целева променлива е използван индекс, изграден от двата въпроса („Доколко сте доволни от живота си“) и „Feeling of happiness“ („Чувство за щастие“).

Всички страни с високи доходи са отбелязани в по-тъмно. От тяхната група, всички са в класацията при „Висок“ или „Средно висок“ индекс на щастие. От десетте страни в Латинска Америка всички без Перу са поставени със „Средно висок“ индекс на щастие. Всички 25 бивши комунистически без Виетнам, Словения и Чехия са поставени със „Средно Нисък“ индекс. Всички бивши съветски страни са поставени с „Нисък“ индекс.

Таблица 15: Подреждане на държавите според класацията на „Световното проучване на ценностите“.

Висок		Средно Висок		Средно Нисък		Нисък	
Пуерто Рико	4.67	Саудитска Арабия	3.01	Южна Африка	1.86	Естония	0.24
Мексико	4.32	Сингапур	3	Хърватска	1.55	Сърбия	0.21
Дания	4.24	Англия	2.92	Гърция	1.45	Танзания	0.13
Ирландия	4.16	Западна Германия	2.67	Перу	1.32	Азербайджан	0.13
Исландия	4.15	Франция	2.61	Китай	1.2	Черна Гора	0.06
Швейцария	4	Аржентина	2.61	Южна Корея	1.12	Индия	0.03
Северна Ирландия	3.97	Виетнам	2.59	Иран	0.93	Литва	-
Колумбия	3.94	Чили	2.53	Полша	0.84	Македония	-

²⁴ World Values Survey- „Subjective well-being rankings of 82 societies.“

							0.14
Холандия	3.86	Филипини	2.32	Турция	0.84	Пакистан	-0.3
Канада	3.76	Тайван	2.25	<u>Босна</u>	<u>0.82</u>	<u>Латвия</u>	<u>-0.7</u>
Австрия	3.69	<u>Доминиканска Република</u>	<u>2.25</u>	Мароко	0.74	<u>Албания</u>	<u>-</u> <u>0.86</u>
<i>Ел Салвадор</i>	3.67	<i>Бразилия</i>	2.23	Уганда	0.67	<u>България</u>	<u>-</u> <u>0.87</u>
<i>Венецуела</i>	3.58	Испания	2.13	Алжир	0.57	<u>Беларус</u>	<u>-</u> <u>0.92</u>
Люксембург	3.52	Израел	2.08	Бангладеш	0.54	<u>Грузия</u>	<u>-</u> <u>1.11</u>
САЩ	3.47	Италия	2.06	Египет	0.52	<u>Румъния</u>	<u>-1.3</u>
Австралия	3.46	Източна Германия	2.02	<u>Унгария</u>	<u>0.41</u>	<u>Молдова</u>	<u>-</u> <u>1.63</u>
Нова Зеландия	3.39	Словения	2.02	<u>Словакия</u>	<u>0.4</u>	<u>Русия</u>	<u>-</u> <u>1.75</u>
Швеция	3.36	<i>Уругвай</i>	2.02	Йордания	0.39	<u>Армения</u>	<u>-1.8</u>
Нигерия	3.32	Португалия	1.99			<u>Украйна</u>	<u>-</u> <u>1.81</u>
Норвегия	3.25	Япония	1.96			Зимбабве	- 1.88
Белгия	3.23	<u>Чехия</u>	<u>1.94</u>			<u>Индонезия</u>	<u>-2.4</u>
Финландия	3.23						

Сравнявайки тази класация с нашата се вижда, че за бившите социалистически страни, резултатите от проучванията съвпадат. Нашето проучване подрежда страните от Северна Европа по-нагоре в класацията от тези в Южна Америка, за разлика от официалната класация. Също така Африканските държави в нашето проучване са по-нагоре от тези в Източна Европа, а в официалната класация наблюдаваме обратното.

Заклучение

В настоящата магистърска теза беше потърсен отговор на въпроса: „Кои са факторите от които зависи човешкото щастие, и по какъв начин му влияят те." Като основна база на изследването бяха използвани данни от „Проучването на Световните Ценности". За постигането на нашата цел бяха направени:

1. Преглед на някои проведени изследвания пряко свързани с щастието.
2. Подробен анализа на базата данни, с която разполагаме.
3. Избор на иконометрични техники за изпълнение на поставената задача.
4. Прилагане на избраните от нас техники върху наличните данни.
5. Анализирание на резултатите и сравнение с подобни проучвания.

В резултат на направения преглед на теорията и емпириката в икономическата литература може да заключим, че:

Разглежданият от нас въпрос относно човешкото щастие е актуален и значим в световен мащаб. Проведените до момента изследвания са разнообразни относно използваните техники и показват резултати, на базата на които стъпва и нашето изследване.

Избраният от нас подход за изчисляване на скоринг, посредством регресионен анализ, и по-конкретно логистична регресия, има широко приложение във финансовия сектор и дава значими резултати, приложен върху нашите социологически данни.

От изводите направени в разгледаните проучвания, може да се заключи, че факторите влияещи върху щастието са разнообразни по характер и включват както финансов, така и културен и локационен аспект.

След като обобщим резултатите от направеното проучване, може да се установи, че използваните нови методи, а именно логистичната регресия и методите за допълване на извадката, дават очакваните резултатите и са приложими при подобни изследвания.

Анализът на характеристиките показва, че фактори, които са бегло разглеждани до момента, като свободно време, приятелство и уважение на човешките права също допринасят към цялостното човешко щастие. Доходите остават основен фактор, но не са единствени в приноса за човешко щастие. Цялостната свобода на действие, както в работна, така и в извънработна среда, също се оказва значим фактор.

Получената от нас класация, независимо, че използва различна целева променлива и емпирични методи, е близка до официалната публикувана класация, която е била изградена върху данни от предишни вълни на социологическото проучване.

Една от основните предпоставки за развитие на анализа ни е още по-дълбокото сегментиране и допълване на въпросниците от „Проучване на Световните Ценности“. В идеалния случай, анкетите могат да бъдат проведени във всички съществуващи страни ежегодно, подобно на преброяване на населението.

В момента протича провеждането на шестата вълна от анкети, която ще събира данни от периода до 2014 година. Тази по-актуална база може да бъде изследвана за промени в човешките нагласи, настъпили в резултат на финансовата криза в последните години. Ще може да се наблюдава дали и как кризата измества фокуса на значими фактори спрямо щастieto. Сравненията със сегашната и по-ранните извадки, биха показали в каква насока може да се работи, така че дори във времена на криза, човешкото щастие да се поддържа на високи нива.

Библиография

- [1.] Christian P. Robert, George Casella - „*Introducing Monte Carlo Method with R*". University Paris Dauphine, University of Florida, 2009.
- [2.] D. Sharma - „*Guide to Credit Scoring in R*", 2009.
- [3.] David J. Hand - „*Assessing and evaluation scorecards. Tools, methods, and consequences* " Imperial College and Winton Capital Management, 2013.
- [4.] Derek Montrichard- „*Reject Inference Methodologies in Credit Risk Modeling.*" Canadian Imperial Bank of Commerce.
- [5.] Experian - „*Scoring Development Guide*", 2009.
- [6.] Foster Provost, Tom Fawcett - „*Data Science for Business. What you need to know about data mining and data-analytic thinking*", 2013.
- [7.] Gerhard Arminger - „*Expert Systems with Applications*", 2013.
- [8.] German Rodriguez - „*Introducing R*". Princeton University, 2012.
- [9.] Gujarati - „*Basic Econometrics, Fourth edition*", 2004.
- [10.] Hilke Brockmann, Jan Delhey, Christian Welzel, Hao Yuan - „*The China Puzzle: Falling Happiness in a rising Economy.*" World values survey data, 2008.
- [11.] Hussein Abdou, John Pointon Credit Scoring- „*Statistical Techniques and Evaluation Criteria: A Review of the Literature*"
- [12.] Jonathan Crook, John Banasik - „*Does Reject Inference Really Improve the Performance of Application Scoring Models?*," 2011.
- [13.] Martin Vojtek - „*Credit Scoring Methods*", 2005.
- [14.] Naeem Siddiqi - „*Credit Risk Scorecards. Developing and Implementing Intelligent Credit Scorings*", 2006.
- [15.] Newsom - „*Data Analysis*", 2012.
- [16.] Raymond Anderson - „*The Credit Scoring Toolkit*", 2007.
- [17.] Ronald Inglehart, Hans-Dieter Klingemann- „*Genes, Culture, Democracy, and Happiness*". World values survey data.

- [18.] Ronald Inglehart, Roberto Foa, Christopher Peterson, Christian Welzel -
„*Development, Freedom, and Rising Happiness.*“ World values survey data, 2008.
- [19.] UniCredit - „*Model Development Guidelines*“, 2010.
- [20.] World Values Survey - „*Integrated Questionnaire*“, 2011.
- [21.] World Values Survey- „*Subjective well-being rankings of 82 societies.*“

Ползвани интернет страници:

www.worldvaluessurvey.org

Приложения

A1 РАЗПРЕДЕЛЕНИЕ НА УЧАСТНИЦИТЕ ПО ДЪРЖАВИ.

	Участници	Процент	Кумулативен процент
Италия	1012	1.5	1.5
Испания	1200	1.8	3.3
САЩ	1249	1.9	5.1
Канада	2164	3.2	8.4
Япония	1096	1.6	10
Мексико	1560	2.3	12.3
Южна Африка	2988	4.4	16.8
Австралия	1421	2.1	18.9
Норвегия	1025	1.5	20.4
Швеция	1003	1.5	21.9
Аржентина	1002	1.5	23.4
Финландия	1014	1.5	24.9
Южна Корея	1200	1.8	26.7
Полша	1000	1.5	28.1
Швейцария	1241	1.8	30
Бразилия	1500	2.2	32.2
Чили	1000	1.5	33.7
Индия	2001	3	36.7
Словения	1037	1.5	38.2
България	1001	1.5	39.7
Румъния	1776	2.6	42.4
Китай	2015	3	45.3
Тайван	1227	1.8	47.2
Турция	1346	2	49.2
Украйна	1000	1.5	50.7
Перу	1500	2.2	52.9
Уругвай	1000	1.5	54.4

Гана	1534	2.3	56.7
Молдова	1046	1.6	58.2
Грузия	1500	2.2	60.4
Тайланд	1534	2.3	62.7
Индонезия	2015	3	65.7
Виетнам	1495	2.2	67.9
Сърбия	1220	1.8	69.8
Египет	3051	4.5	74.3
Мароко	1200	1.8	76.1
Йордания	1200	1.8	77.9
Кипър	1050	1.6	79.4
Гватемала	1000	1.5	80.9
Тринидад и Тобаго	1002	1.5	82.4
Андора	1003	1.5	83.9
Малайзия	1201	1.8	85.7
Буркина Фасо	1534	2.3	88
Етиопия	1500	2.2	90.2
Мали	1534	2.3	92.5
Руанда	1507	2.2	94.7
Замбия	1500	2.2	96.9
Германия	2064	3.1	100
Total	67268	100	

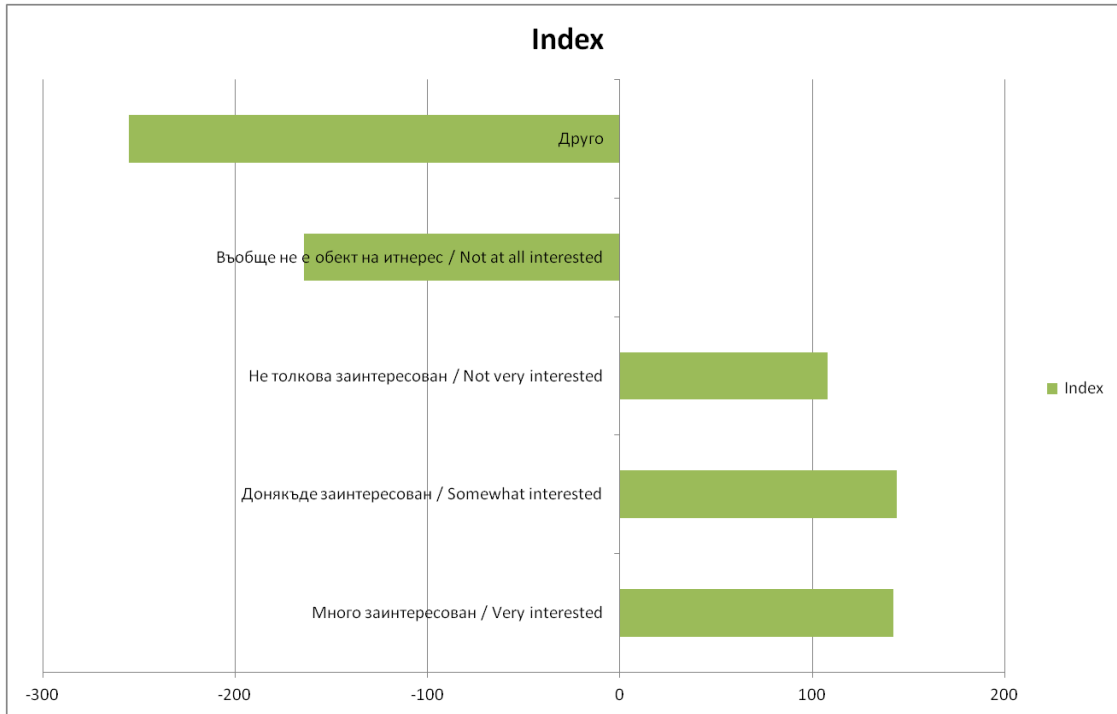
A2 ОСНОВНИ ПАРАМЕТРИ НА ФАКТОРИТЕ.

Поради големината на това приложение, първо е представен пример на таблиците и графиките, създадени за всяка характеристика, след което са представени прикачени Excel файлове, съдържащи всички таблици.

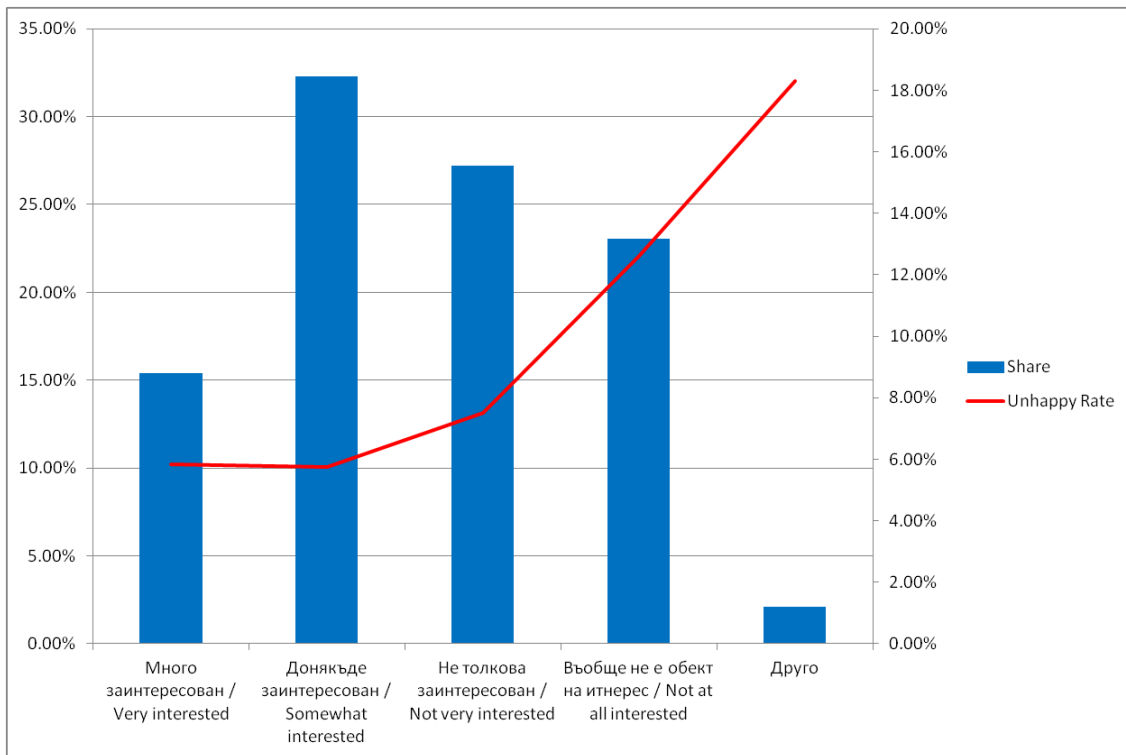
Примерни таблици:

Интерес към политиката											
	Много нещастен	Много щастлив	Total	Unhappy Rate	Odds	Index	Share	% Bads	% Goods	WoE	V-inf
Много заинтересован / <i>Very interested</i>	186	3006	3192	5.83%	16.16	142	15.40 %	11.10 %	15.78 %	0.35	0.02
Донякъде заинтересован / <i>Somewhat interested</i>	384	6306	6690	5.74%	16.42	144	32.28 %	22.93 %	33.10 %	0.37	0.04
Не толкова заинтересован / <i>Not very interested</i>	423	5212	5635	7.51%	12.32	108	27.19 %	25.25 %	27.36 %	0.08	0.00
Въобщо не е обект на интерес / <i>Not at all interested</i>	602	4172	4774	12.61%	6.93	-164	23.03 %	35.94 %	21.90 %	-0.50	0.07
Друго	80	357	437	18.31%	4.46	-255	2.11 %	4.78 %	1.87%	-0.94	0.03

Графика на „Index“:



Графика на „Share“ и „Unhappy Rate“:



Изчерпателни таблици:

Изчерпателните таблици от приложението са представени в три електронни документа (електронни таблици в Excel формат) с индивидуално описание на всички променливи и техните основни характеристики, според които ще бъдат подбирани. Разделени са на 3 групи по около 100 характеристиките във всеки файл поради ограниченията на Excel за брой страници във файл. Общо 363 таблици и 726 графики.



Основни параметри
на въпросите 1.xlsx



Основни параметри
на въпросите 2.xlsx



Основни параметри
на въпросите 3.xlsx

Б1 ПЪЛЕН СПИСЪК НА ФАКТОРИТЕ.

Първоначален списък на всички променливи, предоставени от „Световното изследване на ценностите“



Първоначален
списък с променливи

Променлива	GINI	IV
Feeling of happiness	100.0	0.0
split OECD		
split ballot		
How satisfied are you with your life	80.2	3.0
Country/region with Germany splitted		
Region where the interview was conducted	70.1	1.4
Satisfaction with the financial situation of household	67.5	1.8
Family important		
State of health (subjective)	65.9	1.9
Friends important		
Language at home	59.8	1.3
Leisure time		
Which party would you vote: first choice	55.8	0.9
Politics important		
Weight	52.9	0.5
Work important		
Party that would never vote	51.3	0.8
Religion important		
Language in which interview was conducted	51.2	1.0
Which party would you vote: second choice	48.9	0.6
Social class (subjective)	46.1	0.8
How much freedom you feel	45.5	0.8
Scale of incomes	45.1	0.7
Respect for individual human rights	37.7	0.5
Religious denomination	37.5	0.6
Ethnic group	35.1	0.6
Family savings during past year	34.1	0.5
Democraticness in own country	33.9	0.4
Do you think most people try to take advantage of you	33.8	0.4
How often use of PC	33.1	0.4

Leisure time	30.3	0.3
Highest educational level attained	29.8	0.3
Fate versus control	29.7	0.3
Profession/job	29.5	0.3
Confidence: The European Union	29.4	0.3
What age did you complete your education	28.8	0.3
Employment status	28.7	0.3
Nature of tasks: independence	27.1	0.3
I see myself as a world citizen	26.0	0.3
Government more responsibility	25.5	0.2
Ethnic diversity	25.2	0.2
Confidence: The Environmental Protection Movement	24.9	0.2
Schwartz: Important to this person to think up new ideas	24.9	0.3
Information source: Internet, Email	24.8	0.4
Confidence: Major Companies	24.4	0.2
How proud of nationality	24.2	0.3
Friends important	24.2	0.3
Confidence: The Police	24.2	0.2
Nature of tasks: manual vs. Cognitive	24.1	0.2
Information source: Daily newspaper	24.1	0.3
Confidence: Justice System	24.0	0.2
Trust: People you know personally	23.9	0.2
Neighbours: People with a criminal record		
Neighbours: Black people		
Neighbours: White people		
Neighbours: Coloured people		
Neighbours: Indians		
Neighbours: Kurds, Esids		
Neighbours: Indigenes; Aborigenes		
Neighbours: Maori		
Neighbours: Pacific Islanders		
Neighbours: Europeans/Pakeha		
Neighbours: Gipsies, Romani, Tziganes		
Neighbours: Americans		
Neighbours: Chaldean		
Neighbours: Mapuches		
Neighbours: Russians		
Neighbours: Emotionally unstable people		
Neighbours: Muslims		
Neighbours: Jews		
Neighbours: Evangelists		

Neighbours: People not from country of origin		
Neighbours: Political Extremists		
Neighbours: Christians		
Neighbours: Sunnis		
Neighbours: Shia		
Neighbours: French		
Neighbours: British		
Neighbours: Iranian		
Neighbours: Kuwaiti		
Neighbours: Turkish		
Neighbours: Jordanian		
Nature of tasks: routine vs. Creative	23.8	0.2
Trust: People of another religion	23.7	0.2
Schwartz: Important to this person to have a good time	23.6	0.2
Confidence: The Women's Movement	23.2	0.2
The world is better off, or worse off, because of science an	23.1	0.2
Confidence: The Government	23.0	0.2
Supervising someone	22.9	0.2
Age	22.8	0.2
Confidence: Charitable or humanitarian organizations	22.8	0.2
Trust: People of another nationality	22.5	0.2
Increase in taxes if extra money used to prevent environment	22.5	0.2
Politics important	22.2	0.2
Confidence: Labour Unions	22.2	0.2
Confidence: Parliament	22.2	0.2
Self positioning in political scale	22.0	0.2
Confidence: The United Nations	21.9	0.2
Wealth accumulation	21.9	0.2
Having experts make decisions	21.8	0.2
Year of birth	21.8	0.2
Chief wage earner profession/job	21.8	0.2
Justifiable: homosexuality	21.7	0.2
Democracy: People receive state aid for unemployment.	21.6	0.2
Membership of sport or recreation	21.2	0.3
Would give part of my income for environment	21.2	0.2
Political action: signing a petition	21.1	0.2
Most serious problem for own country: 1st choice	21.0	0.2
Aims of country: first choice (HK)		
Confidence: The Civil Services	20.9	0.2
Aims of country: second choice (HK)		
Marital status	20.8	0.3

Science and technology are making our lives healthier, easie	20.8	0.2
Interest during the interview	20.6	0.2
Most important: first choice (HK)		
We depend too much on science and not enough on faith	20.6	0.2
Most important: second choice (HK)		
Interested in politics	20.5	0.2
Schwartz: Important to this person adventure and taking risks	20.5	0.2
Membership of church or religious organization	20.3	0.2
Trust: People you meet for the first time	20.3	0.2
Confidence: The Political Parties	20.3	0.2
Democracy: Governments tax the rich and subsidize the poor.	20.1	0.2
First choice, if looking for a job	19.9	0.2
Amount of foreign aid of this country	19.7	0.3
Men make better business executives than women do	19.3	0.2
Democracy: The economy is prospering.	19.2	0.2
Because of science and technology, there will be more oport	19.1	0.2
Environmental problems in your community: Poor sewage and sa	19.0	0.1
Information source: Books	18.8	0.2
Most serious problem for own country: 2nd choice	18.8	0.1
Who should decide: international peacekeeping	18.7	0.2
Is the chief wage earner employed now	18.7	0.1
Science and technology make our way of life change too fast	18.7	0.2
Information source: Printed magazines	18.6	0.2
Democracy: People can change the laws in referendums.	18.6	0.2
Be willing to pay higher taxes in order to increase country'	18.5	0.2
Immigrant policy	18.4	0.1
Post-Materialist index 12-item	18.3	0.1
Environmental problems in the world: Global warming or the g	18.3	0.2
Information source: In depth reports on radio or TV	18.2	0.2
Men make better political leaders	18.2	0.1
Justifiable: euthanasia	18.2	0.1
Schwartz: Important to this person looking after environment	18.1	0.2
Politicians who don't believe in god unfit for public office	18.0	0.1
Government should reduce environmental pollution	17.9	0.1
Neighbours: People who have a AIDS	17.8	0.1
Neighbours: Homosexuals	17.7	0.1
Private ownership of business	17.7	0.1
Trust: Your neighborhood	17.7	0.1
Importance of democracy	17.6	0.2
Schwartz: Important to this person being very successful	17.6	0.2
How much do you trust your family	17.5	0.2

Having a strong leader	17.5	0.1
Institution of occupation	17.1	0.1
Post-Materialist index 4-item	17.1	0.1
Political action recently done: signing a petition	17.1	0.1
Aims of respondent: first choice	17.0	0.1
Democracy: Civil rights protect people's liberty against opp	16.8	0.1
Environmental vs. Economic growth	16.7	0.1
Incomes more equal	16.6	0.1
Democracy: The army takes over when government is incompeten	16.5	0.1
Being a housewife fulfilling	16.4	0.1
Schwartz: Important to this person to help the people	16.4	0.1
Work should always come first even if it means less spare ti	16.3	0.1
Requirements for citizenship: being born on my country's soi	16.3	0.2
Churches give answers: the social problems	16.3	0.1
Democracy: Criminals are severely punished.	16.2	0.1
Child qualities: hard work	16.2	0.1
Having a democratic political system	16.2	0.1
To develop talents you need to have a job	16.1	0.1
Work is a duty towards society	16.1	0.1
I see myself as an autonomous individual	16.1	0.1
Trust: Canadian people in general		
Trust: Shi'i		
Trust: Sunni		
Trust: Kurds		
Trust: Christians		
Trust: Relatives		
Trust: People in general		
Most people can be trusted	16.0	0.1
How often do you attend religious services	15.9	0.1
Hard work	15.8	0.2
Democracy: Religious authorities interpret the laws.	15.8	0.1
Opinion about scientific advance	15.5	0.2
Better if more people with strong religious beliefs	15.3	0.1
Environmental problems in your community: Poor water quality	15.2	0.1
Membership of art, music, educational	15.2	0.2
MDG: Reduce extreme poverty	15.1	0.1
How much more foreign aid this country should contribute	15.0	0.2
Religious leaders should not influence government	15.0	0.1
Schwartz: Important to this person to be rich	14.9	0.1
Having the army rule	14.8	0.1
Justifiable: prostitution	14.8	0.1

Information source: Talk with friends or colleagues	14.8	0.1
Priority: Global poverty versus National problems	14.7	0.1
Confidence: NATO		
Confidence: The Arab League		
Confidence: The Association of South East Asian Nations -ASEAN-		
Confidence: The Organisation for African Unity-OAU		
Confidence: The NAFTA		
Confidence: The Mercosur		
Confidence: The APEC		
Confidence: The United American States Organization (Organiz		
Confidence: Local/Regional Government		
Confidence: The Presidency		
Confidence: Confidence in CER with Australia		
Confidence: Education System		
Confidence: The non-Iraqi television		
Confidence: The American Forces		
Confidence: CIS		
Confidence: Banks		
Confidence: TV News		
Confidence: Non governmental Organizations (NGOs)		
Confidence: International Monetary Fund (IMF)		
Confidence: CARICOM		
Confidence: Religious leaders		
Confidence: Evangelical Church		
Confidence: The Press	14.7	0.1
Thinking about meaning and purpose of life	14.5	0.1
Having a strong leader (MX, 5 cats)		
Second choice if looking for a job	14.4	0.1
Having experts make decisions (MX, 5 cats)		
Neighbours: Militant minority	14.4	0.1
Having the army rule (MX, 5 cats)		
MDG: Improve housing conditions	14.3	0.1
Political system: Having an Islamic government, where religious authorities have absolute power		
Political system: Having an Arab socialist state		
Having a democratic political system (MX, 5 cats)		
Who should decide: human rights	14.2	0.1
I seek to be myself rather than to follow others	14.2	0.2
When jobs are scare men should have more right to a job	14.2	0.1
Membership of charitable/humanitarian organization	14.1	0.2
MDG: Reduce child mortality	14.1	0.1

Most serious problem of the world: 2nd choice	13.9	0.1
Aims of respondent: second choice	13.8	0.1
Who should decide: aid to developing countries	13.8	0.1
Competition is good	13.7	0.1
Requirements for citizenship: abiding by my country's laws	13.7	0.1
Religious leaders should not influence people vote	13.7	0.1
Environmental problems in the world: Loss of plant or animal	13.7	0.2
I decide my goals in life by myself	13.6	0.1
Membership of professional organization	13.6	0.2
Confidence: Armed Forces	13.6	0.1
Most serious problem of the world: 1st choice		
Requirements for citizenship: having ancestors from my count	13.5	0.1
Most serious problem of the world: 2nd choice		
Less importance placed on work	13.5	0.1
Most serious problem for own country: 1st choice		
Live up to what my friends expect	13.5	0.1
Most serious problem for own country: 2nd choice		
Who should decide: protection of the environment	13.5	0.1
MDG: Fight HIV	13.4	0.1
Who should decide: refugees	13.4	0.1
Aims of country: second choice	13.2	0.1
University is more important for a boy	13.1	0.1
MDG: Increase primary education	13.1	0.1
Greater respect for authority	13.1	0.1
More emphasis on technology	13.1	0.1
Confidence: Television	13.0	0.1
Family important	12.9	0.2
Schwartz: Important to this person tradition	12.6	0.1
Most important: second choice	12.3	0.1
Democracy: Women have the same rights as men.	12.3	0.1
Political action: Attending peaceful demonstrations	12.3	0.1
Size of town	12.1	0.0
Justifiable: suicide	12.1	0.1
Justifiable: divorce	11.8	0.1
Information source: News broadcasts on radio or TV	11.7	0.1
Membership of labour unions	11.6	0.1
How many children	11.5	0.1
Democracy: People choose their leaders in free elections.	11.5	0.1
I see myself as member of my local community	11.4	0.1
Environmental problems in your community: Poor air quality.	11.4	0.1
Environmental problems in the world: Pollution of rivers, la	11.3	0.2

Schwartz: Important to this person to always behave properly	11.2	0.1
Be willing to fight in war for your country	11.1	0.1
Justifiable: cheating on taxes	10.9	0.1
Justifiable: abortion	10.6	0.1
Most serious problem of the world: 1st choice	10.6	0.1
Child qualities: tolerance and respect for other people	10.6	0.1
One of main goals in life has been to make my parents proud	10.5	0.1
Justifiable: claiming government benefits	10.3	0.1
Schwartz: Important to this person living in secure surroundings	10.2	0.1
Confidence: Churches	10.1	0.1
Membership of political party	10.0	0.1
Churches give answers: the problems of family life	10.0	0.1
People who don't work turn lazy	9.9	0.0
Membership of any other organization	9.8	0.1
Child qualities: feeling of responsibility	9.7	0.0
Churches give answers: moral problems	9.7	0.1
Work important	9.6	0.1
Membership of environmental organization	9.5	0.1
Political action: joining in boycotts	9.4	0.0
I see myself as citizen of [Latin America]		
I see myself as citizen of [North America]		
I see myself as citizen of the [European Union]		
I see myself as citizen of [APEC]		
I see myself as citizen of [ASIA]		
I see myself as citizen of [Mercosur]		
I see myself as citizen of [my province or region]		
I see myself as citizen of [a country other than mine]		
I see myself as citizen of [CIS]		
I see myself as citizen of [The Caribbean]		
I see myself as citizen of [The African Union]		
I see myself as citizen of [Arab Maghreb Union]		
I see myself as citizen of [ASEAN]		
I see myself as citizen of [Arab Union]		
Requirements for citizenship: adopting the customs of my cou	9.4	0.1
Do you live with your parents	9.3	0.1
Employers should give priority to (nation) people than immig	9.0	0.1
Justifiable: avoiding a fare on public transport	9.0	0.1
Churches give answers: people's spiritual needs	8.9	0.1
Justifiable: someone accepting a bribe	8.5	0.1
I see myself as citizen of the [country] nation	8.5	0.1
Religion important	8.4	0.1

Marriage is an outdated institution	8.4	0.1
Neighbours: Immigrants/foreign workers	8.4	0.0
More emphasis on family life	8.4	0.1
Aims of country: first choice	8.3	0.1
Political action: Other	8.2	0.0
Heard of the Millennium Development Goals	8.1	0.0
Child qualities: imagination	8.0	0.0
Neighbours: People of a different race	8.0	0.0
Membership consumer organisation	7.9	0.1
It's humiliating to receive money without having to work for	7.9	0.1
Child qualities: independence	7.8	0.0
One secretary is paid more	7.8	0.1
General party preference		
Year survey	7.5	0.0
Justifiable: For a man to beat his wife	7.5	0.0
Neighbours: People of a different religion	7.2	0.0
Are you the chief wage earner in your house	7.1	0.0
Age of respondent - 3 intervals		
Voted in recent parliament elections	7.0	0.0
Education (country specific)		
Political action recently done: Other	7.0	0.0
Most important: first choice	6.7	0.1
Child needs a home with father and mother	6.5	0.0
Neighbours: People who speak a different language	6.4	0.0
Child qualities: determination perseverance	6.4	0.0
Woman as a single parent	6.2	0.0
Autonomy Index	5.8	0.0
How important is god in your life	5.8	0.0
Moments of prayer, meditation...	5.4	0.0
Political action recently done: Attending peaceful/lawful de	5.3	0.0
Child qualities: obedience	5.3	0.0
Neighbours: Unmarried couples living together	4.9	0.0
Child qualities: thrift saving money and things	4.8	0.0
Child qualities: unselfishness	4.5	0.0
Social class (subjective) with 6 categories		
Father immigrant	4.4	0.0
Income (country specific)		
Neighbours: Drug addicts	4.3	0.0
Mother immigrant	4.2	0.0
Size of town (country specific)		
Political action recently done: joining in boycotts	4.0	0.0

Neighbours: Heavy drinkers	3.2	0.0
Religious person	2.4	0.0
Sex	1.0	0.0
Weight [with split ups]		
Child qualities: religious faith	0.8	0.0
wave number		
Country/Region		
Original respondent number		

Б2 СЪКРАТЕН СПИСЪК НА ФАКТОРИТЕ.

Втори списък с намален брой променливи. Премахнати са тези въпроси, които са зададени само в някои страни. В него е описано кои въпроси са изключени поради нелогични трендове, голям процент неотговорили или недостатъчна дискриминираща сила.



Втори списък с
променливи.xlsx

Променлива	GINI	IV
Satisfaction with the financial situation of household	67.5	1.8
State of health (subjective)	65.9	1.9
Social class (subjective)	46.1	0.8
How much freedom you feel	45.5	0.8
Scale of incomes	45.1	0.7
Respect for individual human rights	37.7	0.5
Family savings during past year	34.1	0.5
Democraticness in own country	33.9	0.4
Do you think most people try to take advantage of you	33.8	0.4
How often use of PC	33.1	0.4
Leisure time	30.3	0.3
Highest educational level attained	29.8	0.3
Fate versus control	29.7	0.3
Profession/job	29.5	0.3
Confidence: The European Union	29.4	0.3
What age did you complete your education	28.8	0.3
Employment status	28.7	0.3
Nature of tasks: independence	27.1	0.3
I see myself as a world citizen	26.0	0.3
Government more responsibility	25.5	0.2
Ethnic diversity	25.2	0.2
Confidence: The Environmental Protection Movement	24.9	0.2
Schwartz: Important to this person to think up new ideas	24.9	0.3
Information source: Internet, Email	24.8	0.4
Confidence: Major Companies	24.4	0.2
How proud of nationality	24.2	0.3
Friends important	24.2	0.3
Confidence: The Police	24.2	0.2

Nature of tasks: manual vs. Cognitive	24.1	0.2
Information source: Daily newspaper	24.1	0.3
Confidence: Justice System	24.0	0.2
Trust: People you know personally	23.9	0.2
Nature of tasks: routine vs. Creative	23.8	0.2
Trust: People of another religion	23.7	0.2
Schwartz: Important to this person to have a good time	23.6	0.2
Confidence: The Women's Movement	23.2	0.2
The world is better off, or worse off, because of science an	23.1	0.2
Confidence: The Government	23.0	0.2
Supervising someone	22.9	0.2
Age	22.8	0.2
Confidence: Charitable or humanitarian organizations	22.8	0.2
Trust: People of another nationality	22.5	0.2
Increase in taxes if extra money used to prevent environment	22.5	0.2
Politics important	22.2	0.2
Confidence: Labour Unions	22.2	0.2
Confidence: Parliament	22.2	0.2
Self positioning in political scale	22.0	0.2
Confidence: The United Nations	21.9	0.2
Wealth accumulation	21.9	0.2
Having experts make decisions	21.8	0.2
Year of birth	21.8	0.2
Chief wage earner profession/job	21.8	0.2
Justifiable: homosexuality	21.7	0.2
Democracy: People receive state aid for unemployment.	21.6	0.2
Membership of sport or recreation	21.2	0.3
Would give part of my income for environment	21.2	0.2
Political action: signing a petition	21.1	0.2
Most serious problem for own country: 1st choice	21.0	0.2
Confidence: The Civil Services	20.9	0.2
Marital status	20.8	0.3
Science and technology are making our lives healthier, easie	20.8	0.2
Interest during the interview	20.6	0.2
We depend too much on science and not enough on faith	20.6	0.2
Interested in politics	20.5	0.2
Schwartz: Important to this person adventure and taking risks	20.5	0.2
Membership of church or religious organization	20.3	0.2
Trust: People you meet for the first time	20.3	0.2
Confidence: The Political Parties	20.3	0.2
Democracy: Governments tax the rich and subsidize the poor.	20.1	0.2

First choice, if looking for a job	19.9	0.2
Amount of foreign aid of this country	19.7	0.3
Men make better business executives than women do	19.3	0.2
Democracy: The economy is prospering.	19.2	0.2
Because of science and technology, there will be more oport	19.1	0.2
Environmental problems in your community: Poor sewage and sa	19.0	0.1
Information source: Books	18.8	0.2
Most serious problem for own country: 2nd choice	18.8	0.1
Who should decide: international peacekeeping	18.7	0.2
Is the chief wage earner employed now	18.7	0.1
Science and technology make our way of life change too fast	18.7	0.2
Information source: Printed magazines	18.6	0.2
Democracy: People can change the laws in referendums.	18.6	0.2
Be willing to pay higher taxes in order to increase country'	18.5	0.2
Immigrant policy	18.4	0.1
Post-Materialist index 12-item	18.3	0.1
Environmental problems in the world: Global warming or the g	18.3	0.2
Information source: In depth reports on radio or TV	18.2	0.2
Men make better political leaders	18.2	0.1
Justifiable: euthanasia	18.2	0.1
Schwartz: Important to this person looking after environment	18.1	0.2
Politicians who don't believe in god unfit for public office	18.0	0.1
Government should reduce environmental pollution	17.9	0.1
Neighbours: People who have a AIDS	17.8	0.1
Neighbours: Homosexuals	17.7	0.1
Private ownership of business	17.7	0.1
Trust: Your neighborhood	17.7	0.1
Importance of democracy	17.6	0.2
Schwartz: Important to this person being very successful	17.6	0.2
How much do you trust your family	17.5	0.2
Having a strong leader	17.5	0.1
Institution of occupation	17.1	0.1
Post-Materialist index 4-item	17.1	0.1
Political action recently done: signing a petition	17.1	0.1
Aims of respondent: first choice	17.0	0.1
Democracy: Civil rights protect people's liberty against opp	16.8	0.1
Environmental vs. Economic growth	16.7	0.1
Incomes more equal	16.6	0.1
Democracy: The army takes over when government is incompeten	16.5	0.1
Being a housewife fulfilling	16.4	0.1

Schwartz: Important to this person to help the people	16.4	0.1
Work should always come first even if it means less spare ti	16.3	0.1
Requirements for citizenship: being born on my country's soi	16.3	0.2
Churches give answers: the social problems	16.3	0.1
Democracy: Criminals are severely punished.	16.2	0.1
Child qualities: hard work	16.2	0.1
Having a democratic political system	16.2	0.1
To develop talents you need to have a job	16.1	0.1
Work is a duty towards society	16.1	0.1
I see myself as an autonomous individual	16.1	0.1
Most people can be trusted	16.0	0.1
How often do you attend religious services	15.9	0.1
Hard work	15.8	0.2
Democracy: Religious authorities interpret the laws.	15.8	0.1
Opinion about scientific advance	15.5	0.2
Better if more people with strong religious beliefs	15.3	0.1
Environmental problems in your community: Poor water quality	15.2	0.1
Membership of art, music, educational	15.2	0.2
MDG: Reduce extreme poverty	15.1	0.1
How much more foreign aid this country should contribute	15.0	0.2
Religious leaders should not influence government	15.0	0.1
Schwartz: Important to this person to be rich	14.9	0.1
Having the army rule	14.8	0.1
Justifiable: prostitution	14.8	0.1
Information source: Talk with friends or colleagues	14.8	0.1
Priority: Global poverty versus National problems	14.7	0.1
Confidence: The Press	14.7	0.1
Thinking about meaning and purpose of life	14.5	0.1
Second choice if looking for a job	14.4	0.1
Neighbours: Militant minority	14.4	0.1
MDG: Improve housing conditions	14.3	0.1
Who should decide: human rights	14.2	0.1
I seek to be myself rather than to follow others	14.2	0.2
When jobs are scare men should have more right to a job tha	14.2	0.1
Membership of charitable/humanitarian organization	14.1	0.2
MDG: Reduce child mortality	14.1	0.1
Most serious problem of the world: 2nd choice	13.9	0.1
Aims of respondent: second choice	13.8	0.1
Who should decide: aid to developing countries	13.8	0.1
Competition is good	13.7	0.1
Requirements for citizenship: abiding by my country's laws	13.7	0.1

Religious leaders should not influence people vote	13.7	0.1
Environmental problems in the world: Loss of plant or animal	13.7	0.2
I decide my goals in life by myself	13.6	0.1
Membership of professional organization	13.6	0.2
Confidence: Armed Forces	13.6	0.1
Requirements for citizenship: having ancestors from my count	13.5	0.1
Less importance placed on work	13.5	0.1
Live up to what my friends expect	13.5	0.1
Who should decide: protection of the environment	13.5	0.1
MDG: Fight HIV	13.4	0.1
Who should decide: refugees	13.4	0.1
Aims of country: second choice	13.2	0.1
University is more important for a boy	13.1	0.1
MDG: Increase primary education	13.1	0.1
Greater respect for authority	13.1	0.1
More emphasis on technology	13.1	0.1
Confidence: Television	13.0	0.1
Family important	12.9	0.2
Schwartz: Important to this person tradition	12.6	0.1
Most important: second choice	12.3	0.1
Democracy: Women have the same rights as men.	12.3	0.1
Political action: Attending peaceful demonstrations	12.3	0.1
Size of town	12.1	0.0
Justifiable: suicide	12.1	0.1
Justifiable: divorce	11.8	0.1
Information source: News broadcasts on radio or TV	11.7	0.1
Membership of labour unions	11.6	0.1
How many children	11.5	0.1
Democracy: People choose their leaders in free elections.	11.5	0.1
I see myself as member of my local community	11.4	0.1
Environmental problems in your community: Poor air quality.	11.4	0.1
Environmental problems in the world: Pollution of rivers, la	11.3	0.2
Schwartz: Important to this person to always behave properly	11.2	0.1
Be willing to fight in war for your country	11.1	0.1
Justifiable: cheating on taxes	10.9	0.1
Justifiable: abortion	10.6	0.1
Most serious problem of the world: 1st choice	10.6	0.1
Child qualities: tolerance and respect for other people	10.6	0.1
One of main goals in life has been to make my parents proud	10.5	0.1
Justifiable: claiming government benefits	10.3	0.1
Schwartz: Important to this person living in secure surroundings	10.2	0.1

Confidence: Churches	10.1	0.1
Membership of political party	10.0	0.1

БЗ ФИНАЛЕН СПИСЪК НА ФАКТОРИТЕ.

Трети списък с променливи. Това са променливите използвани в регресионния анализ.



Трети списък с
променливи за регр

Променлива	GINI	IV
Satisfaction with the financial situation of household	67.5	1.8
State of health (subjective)	65.9	1.9
Social class (subjective)	46.1	0.8
How much freedom you feel	45.5	0.8
Scale of incomes	45.1	0.7
Respect for individual human rights	37.7	0.5
Family savings during past year	34.1	0.5
Democraticness in own country	33.9	0.4
Do you think most people try to take advantage of you	33.8	0.4
How often use of PC	33.1	0.4
Leisure time	30.3	0.3
Highest educational level attained	29.8	0.3
Fate versus control	29.7	0.3
Employment status	28.7	0.3
Nature of tasks: independence	27.1	0.3
I see myself as a world citizen	26.0	0.3
Government more responsibility	25.5	0.2
Ethnic diversity	25.2	0.2
Confidence: The Environmental Protection Movement	24.9	0.2
Schwartz: Important to this person to think up new ideas	24.9	0.3
Information source: Internet, Email	24.8	0.4
Confidence: Major Companies	24.4	0.2
How proud of nationality	24.2	0.3
Friends important	24.2	0.3
Confidence: The Police	24.2	0.2
Nature of tasks: manual vs. Cognitive	24.1	0.2
Information source: Daily newspaper	24.1	0.3
Confidence: Justice System	24.0	0.2
Trust: People you know personally	23.9	0.2
Nature of tasks: routine vs. Creative	23.8	0.2

Trust: People of another religion	23.7	0.2
Schwartz: Important to this person to have a good time	23.6	0.2
Confidence: The Women's Movement	23.2	0.2
The world is better off, or worse off, because of science an	23.1	0.2
Confidence: The Government	23.0	0.2
Supervising someone	22.9	0.2
Confidence: Charitable or humanitarian organizations	22.8	0.2
Trust: People of another nationality	22.5	0.2
Increase in taxes if extra money used to prevent environment	22.5	0.2
Politics important	22.2	0.2
Confidence: Labour Unions	22.2	0.2
Confidence: Parliament	22.2	0.2
Self positioning in political scale	22.0	0.2
Confidence: The United Nations	21.9	0.2
Wealth accumulation	21.9	0.2
Having experts make decisions	21.8	0.2
Justifiable: homosexuality	21.7	0.2
Membership of sport or recreation	21.2	0.3
Would give part of my income for environment	21.2	0.2
Political action: signing a petition	21.1	0.2
Confidence: The Civil Services	20.9	0.2
Marital status	20.8	0.3
Interest during the interview	20.6	0.2
Interested in politics	20.5	0.2
Schwartz: Important to this person adventure and taking risks	20.5	0.2
Membership of church or religious organization	20.3	0.2
Trust: People you meet for the first time	20.3	0.2
Confidence: The Political Parties	20.3	0.2
First choice, if looking for a job	19.9	0.2
Men make better business executives than women do	19.3	0.2
Environmental problems in your community: Poor sewage and sa	19.0	0.1
Information source: Books	18.8	0.2
Is the chief wage earner employed now	18.7	0.1
Information source: Printed magazines	18.6	0.2
Immigrant policy	18.4	0.1
Post-Materialist index 12-item	18.3	0.1
Information source: In depth reports on radio or TV	18.2	0.2
Men make better political leaders	18.2	0.1
Schwartz: Important to this person looking after environment	18.1	0.2
Politicians who don't believe in god unfit for public office	18.0	0.1
Neighbours: People who have a AIDS	17.8	0.1

Neighbours: Homosexuals	17.7	0.1
Trust: Your neighborhood	17.7	0.1
Importance of democracy	17.6	0.2
Schwartz: Important to this person being very successful	17.6	0.2
How much do you trust your family	17.5	0.2
Having a strong leader	17.5	0.1
Political action recently done: signing a petition	17.1	0.1
Environmental vs. Economic growth	16.7	0.1
Incomes more equal	16.6	0.1
Being a housewife fulfilling	16.4	0.1
Schwartz: Important to this person to help the people	16.4	0.1
Work should always come first even if it means less spare ti	16.3	0.1
Churches give answers: the social problems	16.3	0.1
Child qualities: hard work	16.2	0.1
Having a democratic political system	16.2	0.1
I see myself as an autonomous individual	16.1	0.1
Most people can be trusted	16.0	0.1

Б4 КРАЙНО ГРУПИРАНЕ НА ФАКТОРИТЕ.

Крайно групиране на променливите, които ще се използват в регресионния анализ.

Примерни таблици:

Democraticness in own country							
Група	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
'Not at all democratic','2'	247	1126	3022	18	4.56	-249	6.6
'3','4','Just get by','Don't know','No answer','Not applicable','Not asked'	540	3168	12056	14.6	5.87	-194	23.7
'5','6'	366	3925	10807	8.5	10.72	-106	22.7
'7','8','9','Completely democratic'	522	10834	19997	4.6	20.75	182	47.1

Do you think most people try to take advantage of you							
Група	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
'Would take advantage','2','3'	689	3808	9834	15.3	5.53	-206	21.5
'4','5','Don't know','No answer','Not applicable','Not asked'	457	4562	13255	9.1	9.98	-114	27.4
'6','7','8','9','Try to be fair'	529	10683	22793	4.7	20.19	178	51.1

How often use of PC							
Група	Много нещастни	Много щастливи	Друго	Rate	Odds	Index	Share
'Don't know what a computer is','Never'	1271	9018	25591	12.4	7.1	-160	53.9
'Occasionally','Don't know','No answer','Not applicable','Not asked'	267	4670	10413	5.4	17.49	154	23
'Frequently'	137	5365	9878	2.5	39.16	344	23.1

Изчерпателни таблици:



Основни параметри
на групираните прои

C1 КОРЕЛАЦИОННА МАТРИЦА НА ФАКТОРИТЕ.

Корелационна матрица на групите на променливите във финалната регресия.



Корелационна
матрица.xlsx

V68	Satisfaction with the financial situation of household
V11	State of health (subjective)
V46	How much freedom you feel
V164	Respect for individual human rights
V251	Family savings during past year
V6	Leisure time
V246	Nature of tasks: independence
V5	Friends important

	(Intercept)	V68_fc2	V68_fc3	V68_fc4	V68_fc5	V11_fc2	V11_fc3	V11_fc4	V46_fc2	V46_fc3
V68_fc2	0.16									
V68_fc3	0.13	0.43								
V68_fc4	0.07	0.29	0.29							
V68_fc5	0.08	0.34	0.35	0.25						
V11_fc2	0.35	0	0	0.01	0					
V11_fc3	0.3	0.01	0.03	0.03	0.02	0.62				
V11_fc4	0.26	0.01	0.01	0	0.01	0.52	0.5			
V46_fc2	0.28	0.08	0.09	0.06	0.04	0.01	0.01	0		
V46_fc3	0.25	0.08	0.12	0.12	0.07	0.01	0.02	0.01	0.55	
V46_fc4	0.31	0.04	0.08	0.09	0.12	0.03	0.03	0.03	0.6	0.6
V164_fc2	0.39	0.05	0.05	0.03	0.03	0.01	0	0.01	0.01	0
V164_fc3	0.37	0.05	0.05	0.05	0.07	0.03	0.01	0.02	0.01	0.01
V251_fc2	0.33	0.05	0.08	0.05	0.07	0.01	0.01	0.03	0.03	0.02

V251_fc3	0.22	0.04	0.09	0.1	0.16	0.01	0.03	0.06	0	0.01
V6_fc2	0.2	0.04	0.02	0.01	0.03	0.03	0	0.03	0.01	0.02
V6_fc3	0.19	0.02	0	0.01	0.02	0.02	0.01	0	0	0.01
V246_fc2	0.38	0.04	0.04	0.03	0.01	0.01	0.01	0.01	0	0.04
V5_fc2	0.26	0.01	0.02	0.01	0.01	0.01	0.02	0	0	0.02
V5_fc3	0.28	0.01	0	0.01	0	0.03	0	0.01	0.01	0.01

	V46_fc4	V164_fc2	V164_fc3	V251_fc2	V251_fc3	V6_fc2	V6_fc3	V246_fc2	V5_fc2
V68_fc2									
V68_fc3									
V68_fc4									
V68_fc5									
V11_fc2									
V11_fc3									
V11_fc4									
V46_fc2									
V46_fc3									
V46_fc4									
V164_fc2	0.03								
V164_fc3	0.01	0.74							
V251_fc2	0	0.02	0.01						
V251_fc3	0.02	0.01	0.04	0.59					
V6_fc2	0.04	0.01	0	0.02	0.03				
V6_fc3	0.05	0.01	0.03	0.01	0.03	0.53			
V246_fc2	0.03	0.01	0.02	0.04	0.03	0.01	0.01		
V5_fc2	0.01	0.03	0.04	0	0.01	0.11	0.06	0.01	
V5_fc3	0.03	0	0.03	0.02	0.01	0.11	0.19	0.02	