FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Statistical Analysis for Post Traumatic Stress Disorder and Mental Health in Kosovo after War

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur/in

im Rahmen des Studiums

Software Engineering/Internet Computing

eingereicht von

Maryam Majlessi
Matrikelnummer 0225829

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung
Betreuer: Ao. Univ. Prof. Dipl. Ing. Dr. Techn. Karl Grill

Wien, 20.04.2011

(Unterschrift Verfasser/in) (Unterschrift Betreuer/in)

AKNOWLEDGMENTS

I would like to express my sincere gratitude to my direct thesis advisor, Prof. Karl Grill, for his supervision and support with this research as well as helpful discussions. His dedication and valuable suggestions have been conducive to its completion.

At the same time I wish to express my thanks to Prof. Dr. Thomas Wenzel, Chair, WPA, and Section on Psychological Consequences of Torture & Persecution, for providing me with a captivating topic as well as for his supervision and support in both contributing me to this research and helpful discussion. His dedication and valuable suggestions have been conducive to its completion.

ABSTRACT

This paper is a continuation of an epidemiologic research *"Long-term Sequels of War and Mental Health in Kosovo" – Kosovo*, August 2006 with the aim of analyzing the impacts of war on mental health in Kosovo. Because of the high levels of PTSD (Post Traumatic Stress Disorder), depression and emotional distress in a considerable amount of the general Kosovo population, it was decided to continue analyzing the aforementioned conditions. Therefore, predicting the PTSD symptoms from four questionnaires (HTQ, GHQ -28, MOS-20 and HSCL-25) and the determination of any relationship between PTSD symptoms, depression and anxiety are our main goals. The results represent the significant predictors of war related mental health disorders, focusing particularly on Post-traumatic Stress Disorder (PTSD), depression and emotional distress (anxiety) in 1116 participants from the age of 15 upwards. In total 256 out of 1159 (22%) of the participants suffered from PTSD (CDC-criteria) symptoms. By applying the logistic regression methods measured by HTQ (Harvard Trauma Questionnaire) and Mental Health Questionnaire, it was found that those who became sick but were unable to obtain health care were 3 times more likely to suffer from PTSD and Mental Health problems. Those whose money or property had been obtained by force or threat (Extortion) were a reliable predictor for the presence of mental health problems and increased the probability of their occurrence by 14%. The presence of PTSD symptoms in the unemployed was 35% higher than it was in the employed. The results obtained after the analysis of the MOS questionnaire, which assessed the participant's mental health, showed that the risk of psychological disorders and symptoms of nonspecific psychiatric morbidity increased by 39% in participants with a positive value of Roll functioning. Having analyzed the relationship between depression and PTSD, it was observed that 99% of the participants showing PTSD traits also suffered depression and anxiety symptoms.

Zusammenfassung

Diese Diplomarbeit ist die Fortsetzung der epidemiologischen Forschung "Long-term Sequels of War and Mental Health in Kosovo" – Kosovo, August 2006, mit dem Ziel ,die psychische Auswirkung des Kosovokrieges zu analysieren. Aufgrund der hohen Inzidenz von PTBS (posttraumatische Belastungsstörungen), Depressionen und akuten Belastungsstörungen erschien eine weitere vertiefte Analyse dieser Faktoren für angebracht, nämlich die Prognose über die PTBS aus vier Fragebögen (HTQ, GHQ -28, MOS-20 und HSCL-25) und Bestimmung der Beziehung zwischen PTBS- und Depressions-Symptomatik sowie Angst als unsere Hauptziele. Die Resultate stellen signifikante Prädiktoren psychischer Störungen wie PTBS, Depression und Angst im Zusammenhang mit dem Kosovokrieg bei 1116 Teilnehmern im Alter ab 15 Jahren dar. 256 der insgesamt 1159 (22%) Untersuchten weisen PTBS-Symptome (CDC-Criteria) vor. Anhand der Regressionsanalyse angewendet an Fragebögen wie HTQ (Harvard Trauma Questionnaire) und Mental Health Questionnaire. Durch die Auswertung der Fragebögen (HTQ -Harvard Trauma Questionair- und Mental Questionnaire) mit Regressionsanalyse konnte gezeigt werden, dass die an der Studie teilnehmenden Personen ohne Möglichkeit einer medizinischen Versorgung (als ein Prädiktor) drei Mal häufiger an psychischen Erkrankungen bzw. PTBS litten. Des Weiteren wiesen Personen, denen ihr Gut und Besitz durch Gewalt bzw. Erpressung weggenommen war (als ein weiterer Prädiktor), vermehrt psychische Erkrankungen mit einer14% höheren Wahrscheinlichkeit auf. Das Vorhandensein von PTBS-Symptomen war bei den Arbeitslosen um 35% höher. Die Auswertung der MOS-Fragebögen zeigte die Zunahme des Risikos von psychischen Erkrankungen bei Teilnehmern mit positivem Wert von ‚Roll functioning' um 39%. Die Analyse des Zusammenhangs zwischen Depression und PTBS zeigte, dass 99% der Personen mit PTBS ebenfalls an Depression und Angst litten.

Table of Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 The Need for Kosovo Data Analysis

The conflict between Serbs and Albanians in the former Federal Republic of Yugoslavia over control of the Southern Serbian Province of Kosovo, led to a full scaled war in February 1999. It marked the second major combat operation in NATO's history (Wikipedia). Resulting from the war, over 3500 people were killed and 800,000 Kosovar Albanians fled to neighboring countries (such as the former Yugoslav Republic of Macedonia, Albania, and Montenegro, as well as to other countries), while more than 260,000 people in Kosovo were estimated to be displaced within the country (Spiegel & Salama, War and mortality in Kosovo: an epidemiological testimony, 1998-1999).Once the war drew to an end, nearly 750 000 Albanians returned from Albanian, Macedonian and Montenegrin refugee camps to Kosovo (Spiegel & Salma, Kosovar Albanian Health Survey Report, 1999).They were confronted with extensive damages to their homes and belongings, had lost family members, as well as having been through extremely traumatic experiences (such as rape and torture). Due to this, the rate of psychiatric morbidity increased significantly. With the help of international united community, Kosovo rebuilt. However, fantasies of revenge killings and Post Traumatic Stress Disorder remain (Wenzel, 2006).

The Center for Disease Control and Prevention (CDC) in collaboration with the University of Prishtina, the Institute for Mental Health and Recovery of Kosovo, and Doctors of the World, conducted a mental health survey of the Kosovar Albanian population of age 15 and above, in August-October 1999, and a follow-up survey in May 2000. The results of these studies address the full range of Mental Health Problems that affected social functioning related to the recent war. Therefore in August 2006 the Kosovo Rehabilitation Center for Torture Victims (KRCT) conducted a survey “Long-term Sequels of War, Social Functioning and mental Health in Kosovo” which is a collaborative effort between the Ministry of Health of Kosovo, (MoH), KRCT, World Psychiatric Association (WPA), and Danish Refugee Council (DRC) to evaluate consequences of war on mental health in Kosovo. In this study (CDC August – 2006) the scope, the social and geographical distribution of mental health and social dysfunction problems in Kosovo were determined. The results indicated that PTSD, depression and emotional distress had become chronic in a considerable part of the general population. After analyzing the results of this study, Prof. Dr. Thomas Wenzel, Chair of the WPA Section on Psychological Consequences of Torture and Persecution, decided to continue this study for more assessing, PTSD symptoms, depression, anxiety and non specific emotional distress.

1.2 Literature Review

This paper is based on the epidemiologic research, *Long-term Sequels of War, Social Functioning and Mental Health in Kosovo –Kosovo, August 2006* (Wenzel, 2006)that was conducted in order to assess the impacts of war on social functioning and mental health in Kosovo as of October-November 2005. This kind of informative data on the Kosovo war had not been produced since CDC studies in 1999 and 2000. CDC (the Centers for Disease Control and Prevention) conducted a mental health survey on the Kosovar Albanian population age 15 and above in August-October 1999 and a follow-up survey in May 2000 (Cardozo, Vergara, Agani, & A. Gotway, 2000). The data analysis of Post Traumatic Stress Disorder and assessment of mental health in post-war Kosovo is a complex project for statisticians and psychological experts, and needs a series of carefully conducted phases in order to aid in the improvement of mental health and PTSD symptoms. To predict the absence or presence of PTSD symptoms, logistic regression is used. As the Kosovo data set contains a lot of predictors that need to be considered for our analysis, the first step was to select the most relevant of these predictors. Cristina G. Dascalu (2008) and her group conducted a study on Methods for Data Selection in Medical Databases: The Binary Logistic Regression Relations with Calculated Risks. They found that the binary logistic regression model is a sensitive method with which to identify sets of predictors that are related to each other (D.Duscalu, Carausu, & Manuc, 2008).

When we are discussing dichotomous dependent variables of interest in medical health researches, an appropriate and powerful statistical procedure for modeling such variables is logistic regression (F.Gillespie & Glisson, 1999). About 50 years ago, researchers discussed using an automatic procedure to select a statistical model when there is a large number of potential explanatory variables and no underlying theory on which to base the model selection (L.Pace, 2008). There are many variations of automatic variable selection, such as backward elimination or forward variable selection techniques, to identify independent predictors of morbidity or for developing parsimonious regression models (Austin & V.Tua, 2004). Omitting important prognostic factors results in a systematic mis-estimation of the regression coefficients and biased prediction, and including too many predictors, results in the loss of precision in the estimation of the regression coefficients and the predictions of new responses (Murtaugh, 1998). Most commonly, an automatic variable selection process adds (forward selection) and/or removes (backward elimination) covariates from the model at each step; thus, the term is “stepwise regression” (L.Pace, 2008). The method we used in our analysis is variable selection with stepwise logistic regression.

CHAO-Y. J. PENG&KUK L. LEE&GARY M. INGERSOLL (2000) demonstrate that logistic regression can be a powerful analytical technique for use when the outcome variable is dichotomous. The effectiveness of the logistic model was shown to be supported by; A) significance tests of the model against the null model; B) the significance test of each predictor; C) descriptive and inferential goodness-of-fit indices; D) and predicted probabilities (Peng, Lee, & Ingersoll, 2002).

As the use of logistic regression is becoming increasingly popular in medical research, there is a higher risk for researchers to incorrectly interpret the results of this analysis. Jason W. Osborne and his group conducted a study to highlight methods with which to successfully and correctly interpret the odds ratio, as well as to show how to transform it into an intuitive relative risk (RRs).They also suggested a method for handling odds ratios below 1 (Osborne, 2006).

1.3 Objective of Our Study

According to the *Long-term Sequels of War, Social Functioning and Mental Health in Kosovo, August 2006*; the total mean score of the GHQ-28 had not improved, and the prevalence of PTSD, depression and emotional distress remained high. Therefore predicting the PTSD symptoms and determination of the relationship between PTSD symptoms and depression, anxiety are our main goals.

Quantitative methods were used for analyzing and testing the intensity of Post Traumatic Stress Disorder and psychological distress six years after the 1999-Kosovo conflict. Statistical correlation was used to assess the relationship between PTSD, depression and emotional disorder (anxiety). The Direct Risks tables for HTQ, DEM, and GHQ questionnaires were calculated to reduce the number of parameters necessary for usage in the binary logistic regression. The binary logistic regression was conducted to predict the PTSD symptoms over HTQ Traumatic events, DEM and GHQ questionnaires. In this study the databases were created by using four self-reporting questionnaires concerning aspects of mental health, and relationship between PTSD, anxiety, depression and general health.

The four questionnaires are as follows:

General Health Questionnaire 28, (GHQ-28)

The Harvard Trauma Questionnaire (HTQ)

The Medical Outcome Study 20 (MSO -20)

The Hopkins Symptoms Checklist 25 (HSCL-25)

STATISTICAL THEORY UNDERLYING OUR ANALYSIS

2.1 Regression Methodology

Regression Analysis is a statistical tool for understanding the relationship between different variables. Usually, it leads to determining the causal effect of one variable on another variable. For exploring such problems, statisticians employ regression on the underlying variable of interest to estimate the quantitative effect of the causal variables over the predicted variable.

Formal Regression Analysis investigates the distribution of dependent scalar variable Y or some characteristics of its distribution (such as its mean) as a function of one or more independent variables ($X_1, \ldots, X_k$):

$$p(y|x_1, \ldots, x_k) = f(x_1, \ldots, x_k) \qquad \text{(1) (Fox, 1997)}$$

$p(y|x_1, \ldots, x_k)$ is the probability distribution of Y for these specific values X's. The relationship of Y to X is the chance that X's affect Y, or in other words the purpose is we want to use X's to predict Y. Generally this conditional probability distribution $p(y|x_1, \ldots, x_k)$ is assumed as a normal distribution pattern, therefore the variance of the dependent variable Y on X's $(x_1, \ldots, x_k)$ is the same. As the mean value of Y is a linear function of X's, it can be denoted by μ:

$$\mu \equiv E(y|x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \qquad (2)$$

The special form of equation (2) can be stated as follows:

$$Y_i = \beta_0 + \beta_i x_{i1} + \ldots + \beta_k x_{ik} + u_i \qquad (3)$$

Where:

Y_i is the value of dependent in the ith trial ,

β_0 , β_1 are parameters (coefficients regression),

x_i the value of the predictor variable in the ith trial,

u_i is the error with the mean value $E\{u_i\}$=0 and a constant variance δ^2 ; u_i and $u_j$are uncorrelated and their covariance is zero ({ u_i, u_j}=0 for $i \neq j$).

2.1.1 Estimation of Regression Coefficients with the Method of Least Squares: Fitted Regression Line

This model represents a simple linear regression with a single explanatory variable. The independent variable Y_i consists of the sum of two parts of (1) β_0+ $\beta_1 x_i$ which is the constant part and u_i the random part. Therefore it is concluded that Y_i is the random variable.

E $\{u_i\}$ = 0 (4),

E { Y_i} =E $\{\beta_0 + \beta_1 x_i + u_i\}$= E $\{\beta_0 + \beta_1 x_i\}$ + E$\{u_i\}$=$\beta_0 + \beta_1 x_i$ (5)

Since the relationship of the mentioned variables is statistical and not functional, the observations do not fall directly on the regression line. In order to find a correct estimation of regression line, the method of least squares is considered, which leads to an estimation of the parameters β_0 and β_1 by minimizing the sum of squares of each observation Y_i around its estimated expected value (Kutner, Nachtsheim, & Neter, 2004).

For each observation (x_i, y_i):

y_i - ($\beta_0 + \beta_1 x_i$) = 0 (6)

the sum of n squared deviations(residual) is denoted by Q :

$$Q = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \quad (7)$$

Regarding the definition of sum of squares the parameters β_0 and β_1 are those b_0 and b_1 which minimize Q.

$$\sum y_i = n\, b_0 + b_1 \sum x_i \quad (8)$$

$$\sum y_i \,.\, x_i = n\, b_0 . \sum x_i + b_1 \sum x_i^2 \quad (9)$$

Therefore the point estimators of β_0 and β_1 are as follows:

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad (10)$$

$b_0 = \frac{1}{n}(\sum y_i - b_1 \sum x_i) = \bar{y}$ - $b_1\ \bar{x}$, where $\bar{x}$ and $\bar{y}$ are the mean values of the x_i and y_i.

Since our regression model is: E { y_i} = β_0+ $\beta_1 x_i$, with the calculated estimator parameters b_0 and b_1 the Estimated Regression Function is:

$\hat{y}$ = b_0+$b_1\ x$, where $\hat{y}$ is the value of the estimated regression function. According to the Gauss-Markov theorem "Under the condition of regression model" equation1, the least squares estimators b_0 and b_1 are unbiased and have minimum variance among all unbiased linear estimators.

Hence,

E {b_0} = β_0 , E {b_1} = β_1 and therefore:

E { y_i} = $\hat{y}$ (11)

2.1.2 Estimation of Variance

At first the variance $\delta^2$of error u_i from equation 1 is required to be estimated. For this purpose (SSE) error sum of squares is calculated as follows:

From definition of the residual (Kutner, Nachtsheim, & Neter, 2004), the ith residual is:e_i = y_i- $\hat{y_i}$,

SSE = $\sum_{i=1}^{n}(y_i - \hat{y_i})^2 = \sum_{i=1}^{n} e_i{}^2$ (12)

and, MSE is the error mean of squares, with n-2 degree of freedom (2 reduced because β_0 and $\beta_1$are estimated in $\hat{y_i}$):

s^2=MSE =$\frac{SSE}{n-2} = \frac{\sum(y_i - \hat{y_i})^2}{n-2}$ (13) (Kutner, Nachtsheim, & Neter, 2004) .

2.1.3 Estimation of Regression Coefficients by Method of Maximum Likelihood

In statistics, the most powerful and commonly used method for parameter estimation is the Maximum Likelihood. This method works by selecting values of model parameters that maximize the likelihood function.

This method uses the results of the density function as the probe of the parameter value with the sample data and is denoted as *likelihood value* (L (µ)) of parameter µ. If the likelihood value is relatively large, it means this value is consistent with the sample data.

The normal density of an observation y_i , for regression model y_i= β_0+ $\beta_1 x_i$ + u_i with mean value , E { y_i} = β_0+ $\beta_1 x_i$ and variance $\delta^2\{y_i\}$ = δ^2 is denoted by:

$$f_i = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\left(\frac{y_i-\beta_0-\beta_1 x_i}{\sigma}\right)^2\right] \qquad (14)$$

And the likelihood function for n observations is:

$$L(\beta_0, \beta_1, \ \delta^2) = \prod_{i=1}^{n} \frac{1}{(2\pi\,\delta^2)^{1/2}} exp\left[-\frac{1}{2\,\delta^2}(y_i - \beta_0 - \beta_1 x_i)^2\right] \qquad (15)$$

$$L(\beta_0, \beta_1, \ \delta^2) = \frac{1}{(2\pi\,\delta^2)^{n/2}} exp\left[-\frac{1}{2\,\delta^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right]$$ **(16) (Kutner, Nachtsheim, & Neter, 2004)**

The maximum likelihood estimators are denoted by $\hat{\beta}_0$, $\hat{\beta}_1$,$\hat{\delta}^2$.

,$\hat{\beta}_1 = b_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$, the same as the least square estimator b_1 ,

$\hat{\beta}_0 = b_0 = \frac{1}{n}(\sum y_i - b_1 \sum x_i) = \bar{y} - b_1\,\bar{x}$, the same as the least square estimator b_0 and

$\delta^2 = \hat{\sigma}^2 = \frac{\sum(y_i-\hat{y}_i)^2}{n}$ **(17) (Kutner, Nachtsheim, & Neter, 2004)**.

2.1.4 t Statistic

The t statistic determines the significance of the regression coefficient by testing weather the observed parameter b_1 is significantly different from zero. Therefore if the result shows a non-zero b_1, it is clear that the dependent variable belongs to the regression model.

The t-Statistic is calculated by the following equation:

$t = \frac{b_1}{SE}$ **(18) (Kutner, Nachtsheim, & Neter, 2004)**.

(b_1 is the slope of sample regression line, SE is the standard error of the slope) .

2.1.5 R –Square

For describing the “degree of linear association” between X and y the computation of R-Squared is frequently used by researchers. The measure of R^2 is interpreted as the effect of x on reduction of variation y:

R^2 =1 - $\frac{sum\ of\ squared\ errors}{sum\ of\ deviation\ from\ mean}$ =1 - $\frac{\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2}{\sum(y_i-\bar{y})^2}$ **(19) (Kutner, Nachtsheim, & Neter, 2004)** .

Where $\bar{y}$ is the mean value of y.

The limitations of R^2:

If all observations fall on the regression line, it is concluded that "Sum of squared errors" is 0 and R^2=1. There is a perfect linear relationship between X and Y or all the predictors of sample X are on the linear regression Y.

If the regression line is a horizontal line with slope 0, hence $b_1 = 0$ and $y_i = \bar{y}$ then "$sum\ of\ squared\ errors$" and "$sum\ of\ deviation\ from\ mean$" are equal, then R^2=0. Therefore there is no linear relationship between X and Y.

When analyzing, the value of R^2 is between 0 and 1. The closer R^2 is to 1, the greater is the degree of linear association between X and Y.

2.1.6 Assumptions and Practical Consideration on Regression

The usage of regression analysis utilizes some requirements on the quality of data and the validity of assumption. The most important of these requirements are outlined here briefly:

Normal distribution: In multiple regression the dependent variable is assumed to follow normal distribution (Answers.com) .Both the X-values and the Y-values should be normally distributed. It is assumed that the residuals also follow normal distribution.

Assumption of linearity: The linear regression, as presented here, assumes the linearity of the relationship between the variables, i.e. linearity of the relationship should be at least approximately assumed. But in practice with real data this assumption can never be confirmed and there are always some minor deviations from linearity. However, this problem does not have a considerable affect on regression procedure (StatSoft(Electronic Statistics TextBook)).

The principal assumption for using linear regression models:

- Homoscedasticity: The variance of the dependent variable is constant for all the data. Therefore by increasing the values of the independent variables, the values of the dependent variable do not spread further.

- Preventing Autocorrelation: Examination of the temporal correlation between the residuals. Autocorrelation arises when the residuals correlate with each other in a period of time.
- Preventing Multicollinearity (Independence of Data and error e): Multicollinearity means that the independent variables are correlated. Concealed Multicollinearity means that two or more independent variables are linearly related to each other. All data should be independent (i.e. the cases should not correlate with each other). For example the value of X4 should therefore not easily be derived from X3.This is also valid for the errors or residuals (MESOSworld(Methodical Education for the Social Sciences)).

To detect this situation:

1. Tolerance: By determining R-square, tolerance is detected.
2. Calculating Variance Inflation Factor

Lower tolerance and higher VIF values indicate on multicollinearity (Baumgarth, 2008-2009).

2.2 The Binary Logistic Regression

In every study where data analysis is necessary to identify the relationship between a response variable and one or more explanatory variables, regression methods are requisite. As in the regression model previously mentioned, the purpose of analyzing data samples in such studies is to predict the response (dependent) variable Y regarding the independent data samples X or predictors. When such questions need answering, the regression method is utilized to find the solution. The response variable differs between the regression and logistic regression, whereby in logistic regression the variable is qualitative (binary or dichotomous) but in multiple regression it can be any numerical variable. Hence the general principle which is employed in logistic regression follows the linear regression methods. Additionally dependent variable in logistic regression is assumed to follow Bernoulli distribution (if dichotomous) but in multiple regression it follows normal distribution.

The goal of analysis with logistic regression is the same as linear regression; to find the best fit to describe the relationship between a dependent variable and one or more independent variables. In other words we use logistic regression to classify subjects and assess the quality of a classification rule with its sensitivity, specificity values.

In fact the method by which a statistical software or program fits a regression model is:

- The likelihood is specified by original sets of parameters.

- Then it is described in term of the new parameters in the regression equation.
- Finally most likely values of the regression equation parameters are found.

For analysis of a binary outcome variable, some distribution functions have been in use. Cox and Snell (1989) discuss some of these (Kutner, Nachtsheim, & Neter, 2004), (Cox, 2007). There are 8 options for the distribution of the X_i:

1. Binomial $[P(k) = n! / (k! \cdot (n - k)!)\ \pi^k (1 - \pi)^{N-k}$, where k is the number of successes ($X = 1$) in N trials of a Bernoulli process with probability of success π, $0 < \pi < 1$]
2. Exponential $[f(x) = (1/\lambda)e^{-1/\lambda}$, exponential distribution with parameter $\lambda > 0$]
3. Lognormal $[f(x) = 1/ (x\sigma\sqrt{(2\pi)})\ \exp [- (\ln x - \mu)^2 / (2\sigma^2)]$, lognormal distribution with parameters μ and $\sigma > 0$.]
4. Normal $[f(x)= 1/(\sigma\sqrt{(2\pi)})\ \exp[-(x - \mu)^2/(2\sigma^2)]$, normal distribution with parameters μ and $\sigma > 0$)
5. Poisson ($P(X = k) = (\lambda^k/k!)e^{-\lambda}$, poisson distribution with parameter $\lambda > 0$)
6. Uniform ($f(x) = 1/(b - a)$ for $a \leq x \leq b$, $f(x) = 0$ otherwise, continuous uniform distribution in the interval $[a, b]$, $a < b$)
7. Manual (allows specifying manually the variance of $\beta\hat{}$ under H_0 and H_1).
8. The Logistic function ($f(x;\mu,s) = \frac{e^{-(x-\mu)/s}}{s\left(1+e^{-(x-\mu)/s}\right)^2}$), x is random variable μ is mean and s stands for standard deviation.

As the response variable is a binary variable, it takes on the values 0 and 1. We suppose the probabilities of these values are π for 1 and 1-π for 0.

As in the previous chapter; the simple linear regression model:

$\boldsymbol{Y_i = \beta_0 + \beta_1 x_i + u_i \quad Y_i = 0, 1}$ **(20)**

Since E $\{u_i\}$=0 :

$\boldsymbol{E\{Y_i\} = \beta_0 + \beta_1 x_i}$ **(21)**

We assume that if Y_i is a Bernoulli random variable, the probability distribution function:

P (Y_i=1) =π_i

P (Y_i=0) =$1 - \pi_i$

From the definition of the expected value: E{Y} = $\sum_{s=1}^{k} Y_s f(Y_s)$ (Kutner, Nachtsheim, & Neter, 2004).

$$E\{Y_i\} = 1(\pi_i) + 0(1-\pi_i) = \pi_i = P(Y_i=1) \quad (22)$$

From equation 21 and 22:

$$E\{Y_i\} = \beta_0 + \beta_1 x_i = \pi_i \quad (23)$$

Hence the simple logistic regression is as follows:

$$Y_i = E\{Y_i\} + u_i \quad (24)$$

From equation 23 and 24:

$$E\{Y_i\} = \pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \quad (25)\ \text{(Kutner, Nachtsheim, \& Neter, 2004)}$$

By the definition a transformation of $\pi(x)$, the basic part of logistic regression is the logit information. The definition of logit information is as follows:

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] \quad (26)\ \text{(Kutner, Nachtsheim, \& Neter, 2004)}$$

$$= \beta_0 + \beta_1 x_i \quad (27)$$

From equation 27 it can be concluded that the logit has the properties of linear regression. The parameters of logit are linear and they may have the range from -∞ to +∞ depending on the range of the x. The important difference between linear and logistic regression is the error term. As mentioned before, the error term in linear regression follows normal distribution, but for logistic regression, as the response variable is a binary variable, it follows binomial distribution with conditional mean, $\pi(x)$. Therefore error term $u_i$follows a distribution with a mean 0 and variance $\pi(x)$[1- $\pi(x)$].The quantity of variable u_i, when y=1 u is equal to 1- $\pi(x)$ with probability $\pi(x)$, and if y=0, u is equal to - $\pi(x)$ with probability 1- $\pi(x)$.

2.2.1 Fitting the Logistic Regression Model and Estimating the Coefficient

In linear regression model (previous section) the least square method is used for estimating the unknown parameters of regression equation. However, for logistic regression because of the binary form of response variable, the estimators don't keep the same properties. Hence we consider the general form of maximum likelihood which calculates the value of unknown parameters through maximizing the probability of the observed data. Each Y_i observation is an ordinary Bernoulli variable:

Where, $P(Y_i=1) = \pi_i$ (28)

P $(Y_i=0) = 1 - \pi_i$ (29)

The probability distribution functions of the above Bernoulli variables are as follows:

$\boldsymbol{f_i(Y_i) = {\pi_i}^{Y_i}(1-\pi_i)^{1-Y_i} \quad , Y_i = 0, 1 \quad i = 1, \dots, n}$ **(30) (Kutner, Nachtsheim, & Neter, 2004)**

Because of the Y_i are independent their joint probability function is:

g $(Y_1, \dots, Y_n) = \prod_{i=1}^{n} f_i(Y_i) = \prod_{i=1}^{n} {\pi_i}^{Y_i}(1-\pi_i)^{1-Y_i}$ (31)

By applying the inverse of the cumulative distribution function F_L from equation (25):

$E\{Y_i\} = \pi_i = F_L(\beta_0 + \beta_1 X_i) = \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}$

$\boldsymbol{F_L^{-1}(\pi_i) = \beta_0 + \beta_1 x_i = \pi_i'}$ **(32)**

$\boldsymbol{F_L^{-1}(\pi_i) = log_e\left(\frac{\pi_i}{1-\pi_i}\right)}$ **(33)**

Since $f_i(1) = \pi_i$ and $f_i(0) = 1 - \pi_i$, $f_i(Y_i)$ stand for the probability that $Y_i = 0$ or 1.The likelihood function of equation 31 with the logarithmic form of the joint probability function is as follows:

$$log_e g(Y_1, \dots, Y_n) = l(\beta) = \prod_{i=1}^{n} {\pi_i}^{Y_i}(1-\pi_i)^{1-Y_i}$$

$$= log_e \prod_{i=1}^{n} {\pi_i}^{Y_i}(1-\pi_i)^{1-Y_i}$$

$$= \sum_{i=1}^{n}[Y_i log_e \pi_i + (1-Y_i) log_e(1-\pi_i)]$$

$$\boldsymbol{= \sum_{i=1}^{n}\left[Y_i log_e\left(\frac{\pi_i}{1-\pi_i}\right)\right] + \sum_{i=1}^{n} log_e(1-\pi_i)} \quad \textbf{(34)}$$

(Kutner, Nachtsheim, & Neter, 2004)

As $E\{Y_i\} = \pi_i$:

$\boldsymbol{1 - \pi_i = \left[1 + e^{\beta_0 + \beta_1 x_i}\right]^{-1}}$ **(35)**

Since $E\{Y_i\} = \pi_i = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$ and equation 35:

$\boldsymbol{log_e\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_i}$ **(36)**

Hence, from equation 36:

$$log_e L(\beta_0, \beta_1) = \sum_{i=1}^{n} Y_i\,(\beta_0 + \beta_1 x_i) - \sum_{i=1}^{n} log_e\left(1 + e^{\beta_0 + \beta_1 x_i}\right) \quad (37)$$

The maximum likelihood estimates of logistic regression model are those β_0, β_1 that maximize the log likelihood function which is calculated in equation 37. If we consider b_0, b_1 as the maximum likelihood found by computer – intensive procedures like SPSS from the fitted response function equation 37, the fitted logistic response function is as follows:

$$\hat{\pi} = \frac{e^{b_0 + b_1 x_i}}{1+e^{b_0 + b_1 x_i}} \quad (38)$$ **(Kutner, Nachtsheim, & Neter, 2004)**

By considering the inverse of the cumulative distribution function F_L from equation 33, the fitted response function equation 38 can be considered as the following equation:

$$\hat{\pi}_i' = b_0 + b_1 X \quad (39)$$

and from equation 33:

$$\hat{\pi}_i' = log_e\left(\frac{\pi_i}{1-\pi_i}\right) \quad (40)$$

The obtained function in equation 40 is called the fitted logit response function. In fact, the fitted value which computer statistic programs calculate, can be interpreted as the estimated probability that the response variable Y can be successfully predicted by the sample variable X which is the estimated mean response for *ith* X.

Interpretation of b_1:

We should notice that the interpretation of parameter b_1 (as a slope) is not as simple as it was when presented in simple linear regression. In the logistic regression the effect of increasing a unit in X is calculated by the estimated odds $\frac{\hat{\pi}}{1-\hat{\pi}}$ exp (b_1).

Assume the fitted logit response function for X=X_j and for X=$X_j + 1$ are:

$$\hat{\pi}'(X_j) = b_0 + b_1 X_j \quad (41)$$

$$\hat{\pi}'(X_j + 1) = b_0 + b_1(X_j + 1) \quad (42)$$

By subtracting these two fitted logit response values:

$$\hat{\pi}'(X_j + 1) - \hat{\pi}'(X_j) = b_1 \quad (43)$$

Regarding the equation 41:

$log_e(odds_2) - log_e(odds_1) = log_e(\frac{odds_2}{odds_1}) = b_1$, Thus by applying anti logarithmic of both sides we can obtain the *odds ratio*:

$$\widehat{OR} = \frac{odds_2}{odds_1} = \exp(b_1) \quad (44).$$

The logistic regression also generates the Odds Ratio for each predictor. OR are relative amounts by which odds of the outcome vary when the predictor variable is increased one unit. When exp (b_1) >1, then the probability of response variable increases and when exp (b_1) <1 the probability of response variable decreases. In the case of exp (b_1)=0 then there is no relationship between predictor and the dependent variable.

2.2.2 Testing the Significance of the Fitted Model in Logistic Regression

Let us now examine our fitted model to verify the coefficients of logistic parameters. This process contains some particular tests for assessing whether the independent variables are significantly related to the response variable in the model. There are several different measures for determining the *goodness of fit* or the *significance* of the logistic regression model. These practical measures are the *G statistic*, *Pearson statistic*, and Hosmer-Lemeshow statistic (Hosmer & Lemeshow, 2000) .

2.2.3 G Statistic

The approach to proving the significance of the model is rather common and follows testing whether the model with the included independent variable in the question reveals more information about the response variable. The logistic regression compares the observed value of a response variable (binary dependent variable) with the predicted value of a dependent variable in both cases (the model with the independent variable and the model without independent variable). If the predicted values of the model with a variable are better than those values in the model without a variable, it takes the independent variable as the significant variable.

In linear regression and logistic regression models, the principle of this comparison (of the observed to predicted values of response variables) is the same. With logistic regression the base function is the *likelihood function* whereas the method of the *sum of the squares* is used in linear regression. According to the likelihood function the likelihood Ratio is as follows:

$$D = -2\ln\frac{(likelihood\ of\ the\ fitted\ model)}{(likelihood\ of\ the\ saturated\ model)} \qquad \text{(45) (Hosmer \& Lemeshow, 2000)}$$

By performing equation 34 in equation 45 the *Likelihood Ratio Test* is calculated as follows:

$$D = -2\sum_{i=1}^{n}\left[Y_i ln\left(\frac{\widehat{\pi}_i}{Y_i}\right) + (1-Y_i) ln\left(\frac{1-\widehat{\pi}_i}{1-Y_i}\right)\right] \qquad \text{(46) (Hosmer \& Lemeshow, 2000)}$$

The Likelihood Ratio Statistic (D) in the logistic regression plays the role of the *residual, sum of squares* in the linear regression model.

Since the values of outcome variables are either 0 or 1 (binary), the likelihood of the saturated model is 1, hence the equation 45 becomes the following:

$D = 2ln(likelihood\ of\ the\ fitted\ model)$ **(47) (Hosmer & Lemeshow, 2000)**

In order to assess the significance of an independent variable the comparison of value D with and without the independent variable is used as follows:

$G = D(Model\ without\ variable) - D(Model\ with\ variable)$ **(48) (Hosmer & Lemeshow, 2000)**

Since the maximum likelihood $\beta_0 = ln({}^{k_1}/_{k_0})$ where $k_1 = \sum_{i=1}^{n} Y_i$ and $k_0 = \sum_{i=1}^{n}(1 - Y_i)$

then:

$G = 2\{\sum_{i=1}^{n}[Y_i ln(\widehat{\pi}_i) + (1 - Y_i)ln(1 - \widehat{\pi}_i)] - [k_1 ln\,(k_1) + k_0 ln\,(k_0) - nln(n)]\}$ **(49) (Hosmer & Lemeshow, 2000)**

If β_1=0, the G statistic follows a chi square distribution with 1 degree of freedom.

If $\chi^2(1) > G$,$P[\chi^2(1) > G] < 0.05$, this concludes that the associated independent variable is a significant variable for predicting the outcome.

2.2.4 Pearson Chi-Square Statistic

The aim is to analyze the deviation between observed values and predicted values $(y - \hat{y})$ (the residual prediction). The Pearson chi-square *goodness of fit* assumes that the predictor matrix rows are placed into the *J* groups in which identical categories are grouped together. O_j is an observed frequency for group j , π_j is the model expected value, m_j is the number of identical groups and P is Pearson's cumulative test statistic, which asymptotically approaches a χ^2 chi-square distribution with $J - p - 1$ degrees of freedom, and is obtained by summing J groups :

$P = \sum_{j=1}^{J} \frac{(O_j - m_j\widehat{\pi}_j)^2}{m_j\widehat{\pi}_j}$ **(50) (Rogue Wave web site, 2006)**

The large values obtained by the test statistic show that the logistic response function does not fit. If $X^2 \leq \chi^2$ (1-$\alpha; c - p$) then the null hypothesis (that says there is no difference between observed and model predicted values) is retained. It means that If $\chi^2 \geq \chi^2$(1-$\alpha; c - p$) then the null hypothesis is rejected.

2.2.5 Hosmer-Lemeshow Goodness of Fit Test

Hosmer and Lemeshow is an alternative significance test, which groups the predictions of the logistic regression model instead of grouping the predictor variables which is considered in the Pearson approach. The model prediction splits into K bins evenly. The test statistic is calculated as follows:

$$HL = \sum_{j=1}^{K} \frac{(O_j - n_j \bar{\pi}_j)^2}{n_j \bar{\pi}_j (1 - \bar{\pi}_j)} \qquad \textbf{(51) (Rogue Wave web site, 2006)}$$

Where O_j is the number of positive observation in group j, $\bar{\pi}_j$is the model's average predicted value in group j and $n_j$is the size of the group. The Hosmer –Lemeshow statistic follows a chi -square distribution with K-2 degrees of freedom.

2.2.6 Parameter Significance (Wald test)

The Wald test is a method of testing the significance of explanatory variables in the logistic regression model. For each of these explanatory variables in logistic regression an associated parameter is considered. It tests whether the parameters associated with the groups of the explanatory variables are zero. The Wald test is performed by analyzing the relation of the maximum likelihood estimate of the slope $\hat{\beta}_1$ to its standard error as follows:

$$W = \frac{\hat{\beta}_1{}^2}{(\widehat{SE}(\hat{\beta}_1))^2} \qquad \textbf{(52) (Rogue Wave web site, 2006)}$$

For large n is $W^2 \sim \chi^2$ with 1 degree of freedom. If $P[\chi^2(1) > W] < 0.05$, then it leads to the conclusion that the dependent variable is significant enough for predicting the outcome.

2.2.7 Model Selection Methods

The most important reason for minimizing the number of variables in a model in logistic regression is to make the result model more stable. The more variables contained in a model, the greater the estimated standard errors become and the more the model depends on the observed data (Rogue Wave web site, 2006).Generally the variable selection method varies from one problem to another, in this task the focus is on epidemiologic data analysis. The most important reason for including or excluding a covariant from a statistical analysis is a prior knowledge; however it is not always available for studies (Walter & Tiemeier, 2007). Thus statisticians tried to develop some other algorithms to achieve variable selection such as: change in the effect estimate, stepwise selection, modern techniques such as shrinkage and penalized regression, and others (Altman, 1990). A total of 59 (20%) of all reviewed publications used stepwise selection procedures with or without univariate pre-screening of potential covariates (Altman, 1990) .

2.2.7.1 Stepwise Method for Logistic Regression

When the number of predictors is large (n=30) the stepwise model is recommended .The stepwise model is adopted from multiple regression for use in logistic regression. This procedure determines the Wald statistical significant W (see equation 50)of the coefficient for every variable. In fact it decides whether a variable should be included to the model or should be excluded. Since in logistic regression errors follow a Binomial Distribution, the significant statistic measures via the likelihood ratio chi-square test. Hence this procedure selects the variable which provides the greatest influence in the likelihood ratio compared to the model without the mentioned variable.

2.2.8 Interpretation of the Fitted Logistic Regression Model

In the previous section of this chapter we discussed how to fit and test the significance of the model. In this section we will proceed to interpret these estimated coefficients. The assumptions are that the model fit the data and the variables in this model are significant. The aim of interpretation is to see what the estimated coefficients tell us about the questions that motivated the study (Hosmer & Lemeshow, 2000). Since the estimated coefficient is the slope of the regression equation, this means it is the slope of the function of a dependent variable (per unit of change) in the independent variable. In the following sections we will discuss the interpretation of possible measurement scales of an independent variable.

Dichotomous Independent variable: We assume that the independent variable is nominal and dichotomous, therefore the variable x is coded either as zero or one. Since in the logistic regression, the logit transformation is:

$$g(x) = \ln\{\pi(x)/[1-\pi(x)]\} = \beta_0 + \beta_1 x \quad \textbf{(53) (Hosmer \& Lemeshow, 2000)}$$

$g(1) - g(0) = [\beta_0 + \beta_1] - [\beta_0] = \beta_1$, then the logit differences are equal to β_1. In order to interpret β_1 , we calculate the Odds Ratio (OR) for x=1; $\pi(1)/[1-\pi(1)]$ and for x=0; $\pi(0)/[1-\pi(0)]$ are as follows:$OR = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}$ (Hosmer & Lemeshow, 2000) and by using $\pi(1)$ and $\pi(0)$, instead of keeping the standard definition equation 36 and simplifying it; OR is as follow:

$$OR = e^{\beta_1} \qquad \textbf{(54) (Hosmer \& Lemeshow, 2000)}$$

Hence the relationship between the odds ratio and the regression coefficient is e^{β_1} .

This simple relationship between the odds ratio and coefficient is a powerful analytical tool which has a wide usage in epidemiology (Hosmer & Lemeshow, 2000) . For example; if y denotes the presence or absence of depression and x denotes whether the person experienced a combat situation (such as rape) or not, and when OR = 2, we can estimate that depression is twice as likely to occur among the people who experienced a combat situation.

Polychotomous Independent variable: Categorical variables need special concerns in Regression Analysis because they cannot be entered to the regression equation easily as they are and they need to be recoded into a series of variables. Therefore a series of dummy or design variables must be created to represent the different levels of the categorical variable. There are different methods to create design variables for a polychotomous independent variable: i.e.; *Reference cell coding* and/or *deviation from means coding*

Reference cell coding- In this method one level of the classification variable is designated as the reference level to which parameter estimates for the remaining levels are directly comparable (Lewis, 2007). For the reference group, all of the design variables are set at zero and then a single design variable is set at 1 for each of the other groups. Under this coding scheme, the exponentiated parameter estimate of a level is interpreted as the odds ratio between that level and the reference level. Hence, it would make sense to assign to the reference level any particular level we wanted to pit against all others (Lewis, 2007), (Hosmer & Lemeshow, 2000) .

Deviation coding-This coding system signify the deviation of the “Group mean” from the “overall mean”. The “Group mean” is the logit for the group and the “overall mean” is the average logit over all groups (Hosmer & Lemeshow, 2000). This method is applied by setting all the value of design variables equal to -1 for one of the categories and then using 0 ,1 coding for the remainder of the categories. For interpreting the estimated coefficients the logit and its average for each of categories should be calculated. The interpretation of the logistic coefficients is not clear as in the *reference coding* method.

Continuous Independent variable: When the logistic regression contains a continuous independent variable; the interpretation of the estimated coefficient depends on the method that is used to enter the variable into the model.

We assume that the logit is linear in the continue covariate x; the equation for logit is $g(x) = \beta_0 + \beta_1 x$ (Hosmer & Lemeshow, 2000), If the x changes “1” unit, the slope changes as:

$\beta_1 = g(x+1) - g(x)$ for any value of x. We should consider the “1” is not clinically interesting. For example 1 year increasing in age variable may be too small and we consider the unit of changing age 10 year may be most meaningful. Hence for a realistic interpretation of continues variable a method is needed to define a meaningful change for the continue variable. If the log odds ratio changes “c” unit in x, the logit is as follows:

$g(x+c) - g(x) = c\beta_1$. Hence its odds ratio is obtained by exponentiating this logit difference, $OR(c) = OR(x+c, x) = \exp(c\beta_1)$ (Hosmer & Lemeshow, 2000). By replacing $\beta_1$with its maximum likelihood $\hat{\beta}_1$ the estimation is obtained. The standard error estimation for confidence interval is made by multiplying the estimated standard error of $\hat{\beta}_1$ by c as follows:

$\mathbf{exp}\ [c\widehat{\boldsymbol{\beta}}_1 \pm z_{1-\alpha/2} c\ \widehat{SE}(\widehat{\boldsymbol{\beta}})]$ **(Hosmer & Lemeshow, 2000) (55)**

The statement in equation 55 depends on the choice of c as I explained above, it is the changes of “1” unit in predictor x which should be specified in all calculations. Hence an important modeling consideration for continues variable is their scale in logit.

2.3 Discriminant Analysis

Discriminant analysis is another alternative method for studying the relationship between a dependent variable and one more explanatory variable, in this way that it determines which variables discriminate between two or more groups or dependent variables. The main task of discriminant analysis is to find the parameters of discriminant function optimally. It consists of two-step process: Determining the significance of a set discriminant functions parameters and classification. The first step is the same as MANOVA analysis which performs a F test in order to see which variables have significantly different means between the groups. If group means are found to be statistically significant then the second step "classification of variables" begins .The classification step is done by the canonical functions in this way that the discriminant function automatically finds some optimal combination of variables .The first function produces the most overall discrimination between groups, the second function produces the second most discrimination between groups and so on. The subjects are classified in the groups with the highest classification scores. The maximum numbers of discriminant functions are equal to the degree of freedom (Poulsen & French, 2003) .

For analyzing the grouping differences the Canonical Discriminant Function is considered. A sample is classified according to the sign of Y .It is a combination of discriminant variables that meets a certain condition as follows:

$$Y = b_1x_1 + b_2x_2 + \cdots + b_nx_n + c$$

Where Y is discriminant variable which is formed by discriminant function to predict the group variables, c is the constant, the b 's are discriminant coefficients and the x 's are discriminating variables.

2.3.1 The Discriminant Criterion

The discriminant value for two groups A and B with considering the mean variation between groups is as follows:

$$U = \frac{|\bar{Y}_A - \bar{Y}_B|}{s} \qquad (56)$$

, where s is standard deviation, is selected for presenting the mean variation of groups. The equivalent value of U is the following value:

$$U^2 = \frac{(\bar{Y}_A - \bar{Y}_B)^2}{s^2} \qquad (57)$$

The optimal discriminant function Y for groups A and B is measured by maximizing the discriminant value U or U^2.The assumption for equation 56 and 57 are:

1-Both of these equations are only for two groups valid.

2-The size of groups are equal.

3-The mean variation of groups are equal.

From holding the above promises the below Discriminant Criterion equation is obtained:

$\Gamma = \frac{the\ between\ groups\ sum\ of\ squares}{the\ within\ groups\ sum\ of\ squares}$, can be in detail presented as follow:

$$\boldsymbol{\Gamma = \frac{\sum_{g=1}^{G} I_g (\bar{Y}_g - \bar{Y})^2}{\sum_{g=1}^{G} \sum_{i=1}^{I_g} (Y_{gi} - \bar{Y}_g)^2} = \frac{SS_b}{SS_w}} \quad \textbf{(58)}$$

So the parameter should be chosen that by computing a discriminant score (Y_i) for each subject, the Ratio of the between groups of sum of squares to the within groups of sum of squares is as large as possible.

Wilks lambda is used to test the null hypothesis which is the populations have identical means on Y. *Wilks lambda* is $\lambda = \frac{SS_{within\_groups}}{SS_{total}}$ .The smaller Lambda is the more doubt on null hypothesis. For testing the significance of lambda, SPSS uses Chi distribution approximations.

2.3 Correlation

The Correlation presents how relations between variables can be measured. The relationship between variables helps researchers in prediction from one variable about the other variable. At first we consider the variance of a single variable:

$$\boldsymbol{Variance(S^2) = \frac{\sum(x_i - \bar{x})^2}{N-1}} \qquad \textbf{(59) (Field, 2005)}$$

where N is the number of observation, x_i is the data point and $\bar{x}$ represents the mean of the sample. The covariance which is the average of combined differences of two variables x and y can be written from equation 59 as follows:

$$\boldsymbol{cov(x, y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N-1}} \qquad \textbf{(60) (Field, 2005)}$$

Calculating of covariance is a way to determine whether two variables are related to each other. If covariance is positive it means both of the variables deviate from their mean in the same direction and the negative one means that they deviate their mean in opposite directions. As covariance depends on the scale of the measurement of variable, it is not comparable in an objective way. For overcoming this problem the researchers use the standardization form of covariance by dividing the covariance by the standard deviation. The standardization of covariance is noted as a *correlation coefficient* and recognized as the following equation:

$$r = \frac{cov_{xy}}{s_x s_y} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{(N-1)s_x s_y} \qquad \textbf{(61) (Field, 2005)}$$

The equation 61 was invented by Pearson and known as Pearson Correlation Coefficient.

The range of r is: $-1 \leq r \leq 1$

If the value of Pearson Correlation Coefficient is near to +1, it indicates a perfect linear association between x and y whereas value of r is near to -1, it indicates a strong negative linear association. The Pearson Correlation Coefficient 0 indicates no linear relationship between two variables x and y.

2.3.1 Non Parametric Correlation (Spearman Rank Correlation)

Another method for assessing the relationship between two variables is Spearman rank correlation developed by Charles Spearman (1904).It is a non parametric version of Pearson Correlation. When the distribution of data violated parametric rules such as non-normal distributed data, no appropriate transformation of data can be applied. For handling such data Spearman test works perfectly by applying two steps calculation on data such as first ranking the data and then applying Pearson equation 61 on those ranks as follows:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \qquad \textbf{(62) (Field, 2005)}$$

, where n raw scores X_i, Y_i are converted to ranks x_i, y_i and $d_i = x_i - y_i$; is the differences between the ranks of each observation.

The Pearson is more sensitive or in other word more powerful in advanced statistical work if the data are normally distributed than Spearman method but the Spearman method is more secure in case “this assumption is not correct”.

Chapter 3

DATA SET DESCRIPTIONS

3.1 Data Source

Data set for this analysis was taken from Prof. Dr. Thomas Wenzel (Chair of Section on Psychological Consequences of Torture & Persecution, World Psychiatric Association (WPA) and coordinator of the study "Long-Term Sequels of War, Social Functioning and Mental Health in Kosovo in August 2006 "). For this study the research team used random two-stage cluster sampling methodology to provide a representative sample which also reflected city/countryside and regional distributions. In order to achieve a 95% confidence interval, a total of 30 clusters with at least 40 adults over 15 years in each cluster were required. There were two Serbian clusters selected.

3.2 Data and Variable Description

In this study the databases are created by using four self-reporting questionnaires concerning maintained aspects of mental health, social functioning and relationships between PTSD, anxiety, depression, general health and social functioning.

The instruments used were the following:

General Health Questionnaire 28, (GHQ-28)

The Harvard Trauma Questionnaire (HTQ)

The Medical Outcome Study 20 (MSO -20)

The Hopkins Symptoms Checklist 25 (HSCL-25)

This report includes 1161 participants. Table 1 shows distribution of population according to trauma experiences from Kosovo_2006 Data set. Statistics provided by KRCT study(Kosova Rehabilitation Centre for Torture) indicated that a high proportion of the population had experienced traumatic events, and the majority of the population had been forced to evacuate under dangerous conditions (64.9%), forced to hide (57.4%), confined at home because of dangers outside (46.9%), and had been in a combat situation (39.7%) (Wenzel, 2006).

The important issues concerning post war depression and Post Traumatic Disorder (5 years later) in Kosovo for this analyze are taken from Kosovo_2006 Data set, are as follows:

I .Traumatic Experiences

Traumatic events that the surveyed population experienced are summarized in the Table 1. A high percentage of the sampling population experienced traumatic events. The majority of this population (64.9%) had been forced to evacuate under dangerous conditions, forced to hide (57.4%), confined at home because of outside dangers (46.9%), been in a combat situation (39.7%), had no shelter (36.1%), experienced extortion or been robbed (36.3%), or been deprived of water and food (34.2%). These high percentages of population are excellent sources for revealing more information on PTSD and depression. The main goal is to assess the HTQ for finding affective predictors for PTSD symptoms. In the next chapter with the use of binary logistic regression and SPSS analysis tools, these traumatic events are taken as input predictors.

Table 1 Distribution of Population by Exposure Traumatic Events (HTQ) (Wenzel, 2006)

TRAUMA EVENT	(Experienced +Witnessed)	Witnessed	Heard from others	NO
	N (%)	N (%)	N (%)	N (%)
Lack of shelter	419 (36.1)	143 (12.3)	241 (20.8)	358 (30.5)
Lack of food or water	397 (34.2)	80 (6.9)	251 (21.6)	439 (37.4)
Ill health without access to medical care	280 (24.1)	130 (11.2)	222 (19.1)	529 (45.6)
Confiscation or destruction of personal property	480 (41.3)	58 (5.0)	223 (19.2)	400 (34.5)
Combat situation	461 (39.7)	63 (5.4)	219 (18.9)	418 (3.3)
Used as a human shield	125 (10.8)	17 (1.5)	258 (22.2)	761 (65.5)
Exposure to sniper fire	144 (12.4)	83 (7.1)	245 (21.1)	689 (59.4)
Forced evacuation under dangerous conditions	754 (64.9)	21 (1.8)	72 (6.20	314 (26.9)
Beating to body	89 (7.7)	129 (11.1)	284 (24.5)	659 (56.7)
Rape or sexual abuse	10 (0.9)	16 (1.4)	318 (27.4)	817 (70.4)
Other type of sexual abuse or humiliation	6 (0.5)	15 (1.3)	300 (25.8)	840 (72.3)
Knifing or axing	15 (1.3)	40 (3.4)	295 (25.4)	811 (69.9)

Torture (physical or mental suffering)	124 (10.7)	82 (7.1)	311 (26.8)	644 (55.4)
Serious physical injury from combat	32 (2.8)	110 (9.5)	312 (26.9)	707 (60.9)
Imprisonment	24 (2.1)	56 (4.8)	311 (26.8)	770 (66.3)
Forced labor	14 (1.2)	22 (1.9)	248 (21.4)	877 (75.5)
Extortion or robbery	421 (36.3)	49 (4.2)	225 (19.4)	466 (40.1)
Brainwashing	64 (5.5)	18 (1.6)	225 (19.4)	854 (73.6)
Forced to hide	666 (57.4)	24 (2.1)	106 (9.1)	365 (31.4)
Kidnapped	44 (3.8)	54 (4.7)	277 (23.9)	786 (67.7)
Other forced separation from family	333 (28.7)	49 (4.2)	192 (16.5)	587 (50.5)
Forced to find and bury bodies	30 (2.6)	16 (1.4)	268 (23.1)	847 (72.9)
Forced isolation from family members	103 (8.9)	39 (3.4)	231 (19.9)	788 (67.9)
Present while someone searched for people or things	253 (21.8)	40 (3.4)	262 (22.6)	606 (52.2)
Forced to sing enemy songs	23 (2.0)	15 (1.3)	227 (19.6)	896 (77.10
Someone was forced to betray you	63 (5.40	10 (0.9)	204 (17.6)	884 (76.2)
Confined to home because danger outside	544 (46.9)	14 (1.2)	118 (10.2)	485 (41.7)
Prevent from burying someone	60 (5.2)	23 (2.0)	278 (23.9)	800 (68.8)
Forced to destroy the bodies or graves	6 (0.5)	8 (0.7)	233 (20.1)	914 (78.7)
Forced to physically harm family members or friend	9 (0.8)	9 (0.8)	229 (19.7)	914 (78.7)
Forced to physically harm someone who is not family or.	5 (0.4)	7 (0.60	227 (19.6)	922 (79.4)
Forced to destroy someone else’s property	10 (0.9)	7 (0.6)	194 (16.7)	950 (81.8)
Forced to betray family member or friend	7 (0.6)	6 (0.5)	169 (15.6)	979 (84.3)
Forced to betray someone who is not family member or.	10 (0.9)	5 (0.4)	173 (14.9)	973 (83.9)
Murder or death of spouse doe to the violence	19 (1.6)	36 (3.1)	336 (28.9)	770 (66.3)
Murder or death of son or daughter doe to the violence	18 (1.6)	24 (2.1)	338 (29.1)	781 (67.2)
Murder or death of family member or friend do to the.	116 (10.0)	31 (2.7)	323 (27.8)	691 (59.6)
Disappearance or kidnapping of spouse	20 (1.7)	25 (2.2)	330 (28.4)	786 (67.7)
Disappearance or kidnapping of son or daughter	21 (1.8)	25 (2.2)	336 (28.9)	779 (67.0)
Disappearance or kidnapping of other family member or.	89 (7.7)	31 (2.7)	314 (27.0)	727 (62.7)

Serious injury of family member or friend due to the.	79 (6.8)	43 (3.7)	341 (29.4)	698 (60.1)
Witness beating to head or body	151 (13.0)	124 (10.7)	355 (30.6)	531 (45.8)
Witness torture	148 (12.7)	125 (10.8)	395 (34.0)	493 (42.5)
Witness killing or murder	98 (8.4)	81 (7.0)	462 (39.8)	520 (44.8)
Witness rape or sexual abuse	13 (1.1)	17 (1.5)	404 (34.8)	727 (62.6)

II. Mental Health and Social Functioning

General Health Questionnaire consists of four main parts such as somatic symptoms, anxiety and insomnia, social dysfunction and symptoms of severe depression. Table 2 demonstrates the mean score of the General Health Questionnaire (GHQ-28).The mean total score for the GHQ-28 which has 28 variables (questions) is 7.9. The mean score for somatic symptoms is 2.6 and for anxiety and insomnia were 2.8 higher than social dysfunction and severe depression.

Table 2 The Mean Score of GHQ-28 (Wenzel, 2006)

The Mean Score of GHQ-28 (Wenzel, 2006)

Mental Health Status	
GHQ -28 (1-7 for all subscales)	Mean (SE)
Somatic symptoms	2.58 (0.07)
Anxiety and insomnia	2.80 (0.07)
Social dysfunction	1.54 (0.06)
Symptoms of severe depression	1.17 (0.06)
TOTAL (0-28)	7.91 (0.20)

III. Medical Outcomes Study (MOS-20)

The Medical Outcomes Study addresses health related quality of life issues that were developed for patients participating in the medical outcomes study. Table 3 demonstrates means scores of MOS-20 items. MOS-20 consists of 20 items on 6 different scales that assess physical functioning, bodily pain, role functioning, social functioning, mental health and self–perceived general health status. The MOS-20 was scored based on the user's manual; each row score transformed to fit a 0-to-100 scale using standard formula with the higher scores on this scale representing better functioning; a score of 75 or higher indicating normal social functioning, and for mental health status a cutoff score of lower than 52 representing the presence of psychiatric disorder (Wenzel, 2006).

Table 3 MOS-20 Mean Score (Wenzel, 2006)

MOS-20 (0-100 for all subscales)	Mean (SE)
General health perception	49.94 (0.76)
Mental health status	55.48 (0.66)
Bodily pain	63.47 (0.96)
Physical functioning status	72.68 (0.98)
Social functioning	47.15 (1.40)
Role functioning	51.27 (0.81)

IV. Hopkins Symptom Checklist-25 (HSCl-25)

Hopkins Symptom Checklist is a screening tool design to detect symptoms of anxiety and depression (Buchner) .In Kosovo-2006 Data set HSCl-25 consists of three groups of items: emotional distress, anxiety symptom and depression symptom. The Hopkins Symptoms Check List supplements the HTQ by assessing symptoms of depression and anxiety. It consists of 25 items: Part I of the HSCL-25 consists of 10 items for anxiety symptoms; Part II of 15 items for depressive symptoms. Two scores are calculated; the total score is the average of all 25 items, while the depression score is the average of the 15 depression items. The check list is scored by assigning the following numbers to the responses of each item:

1="not at all", 2="A little", 3="Quite a bit", and 4="Extremely". Total score = add up items 1–25 and divide by 25. If >1.75 participants are considered "checklist" positive" for some type of unspecified emotional distress. Depression score = add up items 11–25 and divide by 15. If >1.75 participants are generally considered "checklist positive" for major depression (Wenzel, 2006).

Table 4 HSCL Symptoms (Wenzel, 2006)

HSCL-25 Symptoms	%	(SE)
Total Depression prevalence (11-25) %	41.76	(0.01)
Total Emotional Distress prevalence (1-25) %	43.10	(0.01)

V. Post-traumatic Stress Disorder (PTSD)

HTQ combines a list of traumatic events with symptoms of Post Traumatic Disorder selected from the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). The occurrence of PTSD symptoms is therefore defined according to a scoring algorithm proposed by the Harvard Refugee Trauma Group. This scoring algorithm requires a score of 3 or 4 on at least one of four re-experiencing symptoms (criterion B), at least three of seven avoidance and numbing symptoms (criterion C), and at least two of five arousal symptoms (criterion D) (Wenzel, 2006).

Table 5 PTSD (Wenzel, 2006)

PTSD	Mean %	(SE)
PTSD Caseness according to CDC definition		22.05 (0.01)
Caseness according to DSM-IV PTSD cut off score of >2.5		17.2 (0.01)

VI. Displacement Characteristics

Table 6 Distribution of Population by Displacement and Refugee Status, Duration and Location of Displacement (Wenzel, 2006)

Characteristic		Number	Percentage (%)
Become Refugee			
	Yes	531	45.7
	No	630	54.3
Displaced within Kosovo			
	Yes	539	46
	No	622	54
Country went as a refugee			
	Albania	212	40
	Macedonia	191	36
	Montenegro	27	5
	Other	101	19
Refugee status duration outside Kosovo			
	0 – 7 days	18	3.5
	7 – 30 days	31	5.8
	More than 30 days	482	90.7
Displacement duration within Kosovo			
	0 – 7 days	132	24.5

	7 – 30 days	86	16
	More than 30 days	321	59.5
Currently displaced from home			
	Yes	125	10.8
	No	1,036	89.2
Since Sept. 1999, have you moved at all			
	Yes	234	20.2
	No	927	79.8
From rural to city (>10.000)			
	Yes	115	9.9
	No	1,046	90.1
From city to rural			
	Yes	68	5.9
	No	1,093	94.1
Same home as before the war			
	Yes	642	55.3
	No	519	44.7

Chapter 4

RESULTS OF STATISTICAL ANALYSIS

4.1 Results from Parameter Selection for Logistic Regression

Generally medical databases need different methods for parameter selection after databases are filled and designed with information. The most common step after database creation is the selection of variables by calculating risk factors in order to select the most relevant predictors. After applying mentioned method for finding predictors, the binary logistic regression procedure is used to obtain optimal predictors. The binary logistic regression is the generalization form of the linear regression model and is applicable for the analysis of the connections between one or more categorical independent variables (ordinal or binary) and a dependent categorical binary variable. The risk factors of the 45, HTQ trauma events are calculated over a binary variable (PTSD symptom).

As a dichotomous dependent variable, the DSM_IV_PTSD Score (Post-Traumatic Stress Disorder) variable is taken from Kosovo-2006 Dataset to perform the binary logistic regression. The risk factors, odds ratio and Pearson Chi-square values are calculated by crosstab s procedure of SPSS 18. (See Table 7). As the KRCT database contains a lot of missing values, the database needed filtering before the crosstabs procedures could be applied in order to prevent appearing errors. The Risk table is sorted by odds ratio. Then, the variables with risk factors greater than 1 are taken as relevant predictors and are suitable inputs for binary logistic regression analysis. Thus the colored predictors are selected for usage in the binary logistic regression procedure. (See table 7)

Table 7 The Direct Risks for Each Traumatic Event

	The risk for the cohort DSM_IV_PTSD=1	Odds Ratio (1/0)	lower	upper	Pearson Chi-square	Asimpt. sig. p
HTQ3	.276	5.263	3.622	7.648	85.649	0
HTQ2	.318	4.115	2.850	5.941	62.354	0
HTQ16	.397	3.663	.811	16.540	3.261	.071
HTQ27	.384	3.141	2.137	4.618	36.180	0
HTQ13	.466	2.709	1.679	4.371	17.730	0
HTQ5	.449	2.649	1.848	3.798	29.376	0
HTQ19	.453	2.549	1.705	3.812	21.775	0
HTQ7	.486	2.547	1.594	4.068	16.143	0
HTQ23	.495	2.508	1.443	4.360	11.253	.001
HTQ25	.508	2.455	.827	7.290	2.789	.095
HTQ11		2.445			.835	.361
HTQ12		2.445			1.889	.169
HTQ35	.509	2.445	.727	8.228	2.223	.136
HTQ39	.509	2.445	.727	8.228	2.223	.136
HTQ29	.514	2.416	.218	26.818	.550	.458
HTQ1	.491	2.387	1.671	3.410	23.690	0
HTQ4	.494	2.354	1.644	3.372	22.578	0
HTQ28	.553	2.155	1.070	4.342	4.821	.028
HTQ26	.582	2.006	1.002	4.015	4.005	.045
HTQ17	.570	1.990	1.393	2.841	14.657	0
HTQ20	.589	1.977	.854	4.578	2.626	0.105153587
HTQ8	.567	1.953	1.284	2.972	10.029	0.001540697
HTQ24	.603	1.881	1.265	2.796	9.964	.002
HTQ37	.611	1.869	1.070	3.264	4.958	.026
HTQ22	.621	1.839	.707	4.779	1.607	.205
HTQ33	.687	1.609	.166	15.570	.172	.679
HTQ14	.769	1.386	.450	4.269	.325	.568
HTQ21	.790	1.332	.914	1.942	2.238	.135
HTQ6	.813	1.291	.743	2.240	.825	0.363598131

HTQ18	.833	1.253	.589	2.665	.344	.557
HTQ36	.858	1.206	.254	5.737	.056	.814
HTQ10	.859	1.205	.134	10.855	.028	.868
HTQ30		1.205			.835	.361
HTQ31		1.205			.835	.361
HTQ34	.859	1.205	.134	10.855	.028	.868
HTQ32		1.148			1.255	.263
HTQ43	.893	1.148	.673	1.957	.256	.613
HTQ9	1.034	.960	.475	1.939	.013	.909
HTQ41	1.080	.912	.419	1.985	.054	.817
HTQ42	1.153	.844	.473	1.507	.330	.566
HTQ15	1.295	.737	.165	3.300	.160	.689
HTQ40	1.349	.703	.311	1.587	.726	.394
HTQ44	1.537	.606	.270	1.358	1.510	.219
HTQ38	1.554	.599	.074	4.825	.237	.626
HTQ45	1.554	.599	.074	4.825	.237	.626

4.2 Applying Logistic Regression

4.2.1 Results from Identifying PTSD's Predictors from HTQ Questionnaire

The goal is to make a prediction about the presence/absence of PTSD symptoms over the population who experienced traumatic events, and to determine whether the selected trauma event adequately describes a PTSD symptom. As in the previous chapter, the description of these traumatic occurrences is mentioned; 45 events from which the most relevant are selected for performing logistic regression. The summary of SPSS 18 output from running stepwise logistic regression is illustrated in tables 8, 9 and 10. The predictors were found in 4 steps, by adding them in the following order: HTQExp2, HTQExp3, HTQExp13 and HTQExp27. (SPSS 18 output file).

The first phase in this analysis is the Omnibus Test demonstrated in Table 8 to assess if the model describes sufficiently the observed data. This test gives the overall test for the model that includes the predictors one by one. The Chi-square value of$\chi^2(1) = 95$ with p-value <0.001 for the first added variable, $\chi^2(2) = 109.17$ and p-value< 0.001 for the second added variable, $\chi^2(3) = 121.32$ for the third added variable, $\chi^2(4) = 125.8$ and p-vale <0.001 for the final added variable, shows the model fits significantly better than a model without predictors.

Table 8 Chi-Square

Omnibus Tests of Model Coefficients				
		Chi-square	**df**	**Sig.**
Step 1	Step	95	1	0
	Block	95	1	0
	Model	95	1	0
Step 2	Step	14.17	1	0
	Block	109.17	2	0
	Model	109.17	2	0
Step 3	Step	12.15	1	0
	Block	121.32	3	0
	Model	121.32	3	0
Step 4	Step	4.48	1	0.03
	Block	125.8	4	0
	Model	125.8	4	0

Model Summery table 9 shows how three measures -2 Log likelihood, Cox & Snell R Square and Nagelkerke fit the data. Overall, the final model accounts for 12 – 20% of variance for PTSD Symptoms and it can be concluded that the predictors can only announce the PTSD symptom 83% percent correctly (see classification table 10).

Table 9 Variance for Overall Final Model

Model Summary			
Step	**-2 Log likelihood**	**Cox & Snell R Square**	**Nagelkerke R Square**
1	814.27	0.09	0.15
2	800.1	0.1	0.17

3	787.95	0.12	0.19
4	783.46	0.12	0.20

The next phase is to create the classification table which presents the results of using the regression model. In fact, the classification table is a way of assessing “Goodness of fit” and gives us the overall success rate of the model and finally determining the quality of classification rules (sensitivity, specificity). The overall success rate for this case is 83%. In this case from the 814 people who didn’t have PTSD symptoms, the model correctly predicted 799 would not, (therefore the model was very accurate). Similarly, of the 171 people who had PTSD symptoms, the model correctly predicted 19 of them but was inaccurate about 159 of them (see table 10 step 4). The sensitivity value in this case is $\frac{19}{152+19} = 0.111111$ .Hence the sensitivity value for this model is 11%. The specificity percentage for this model is $\frac{799}{799+15} = 0.98$. Hence the specificity value for this model is 98%. It means, of those who did not have PTSD symptoms, we predicted 98% correctly. False positive rate in this case is $\frac{15}{15+19} = 0.44$

This means of all those for whom we predicted PTSD symptoms, we were 44% wrong. False negative rate is $\frac{152}{152+799} = 0.159$. This means of all those for whom we predicted not to have PTSD symptoms 15.9% we predicted wrong.

Table 10 Quality of Model: Sensitivity, Specificity

Classification Table					
	Observed		**Predicted**		
			DSM-IV PTSD cut off score of >2.5		**Percentage Correct**
			0	1	
Step 1	DSM-IV PTSD cut off score of >2.5	0	814	0	100
		1	171	0	0
	Overall Percentage				82.6
Step 2	DSM-IV PTSD cut off score of >2.5	0	814	0	100
		1	171	0	0
	Overall Percentage				82.6
Step 3	DSM-IV PTSD cut off score of >2.5	0	814	0	100
		1	171	0	0
	Overall Percentage				82.6

Step 4	DSM-IV PTSD cut off score of >2.5	0	799	15	98.2
		1	152	19	11.1
	Overall Percentage				83

The next phase of logistic regression is to generate Variables in the Equation table 11. This table gives us information about the parameters of the model such as the coefficients, their standard errors, the Wald test statistic, p-values, and the exponentiated coefficients (also known as an odds ratio Exp (B)). According to Table 11, the significance values of the Wald statistics for each predictor indicate HTQExp3 “Ill health without access to medical care” (Wald=94.88, p<0.0001), HTQEXP2”Lack of food or water” (Wald=14.45, p<0.0001), HTQExp27“Confined to home because danger outside” (Wald=11.53, p<0.0001) and HTQ13 “Torture; physical or mental suffering “(Wald=10.02, p<0.0001). The aforementioned events significantly predict PTSD symptoms. Therefore we can reject the null hypothesis and say that the coefficients are significantly different from 0.

Table 11 Logistic Regression Coefficients, Wald Statistic, and Odds Ratio

Variables in the Equation									
		B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95% C.I. for EXP(B)**	
								Lower	Upper
Step 1^{a}	HTQExp3	1.745	.179	94.880	1	.000	5.728	4.032	8.139
	Constant	-2.165	.120	323.752	1	.000	.115		
Step 2^{b}	HTQExp2	.798	.210	14.445	1	.000	2.222	1.472	3.354
	HTQExp3	1.308	.210	38.789	1	.000	3.700	2.451	5.585
	Constant	-2.375	.139	292.796	1	.000	.093		
Step 3^{c}	HTQExp2	.725	.211	11.785	1	.001	2.065	1.365	3.125
	HTQExp3	1.194	.212	31.669	1	.000	3.301	2.178	5.004
	HTQExp27	.668	.194	11.845	1	.001	1.950	1.333	2.852

	Constant	-2.689	.175	237.332	1	.000	.068		
Step 4^d	HTQExp2	.695	.212	10.697	1	.001	2.003	1.321	3.038
	HTQExp3	1.161	.214	29.547	1	.000	3.192	2.100	4.850
	HTQExp13	.528	.245	4.642	1	.031	1.695	1.049	2.738
	HTQExp27	.620	.196	10.018	1	.002	1.859	1.266	2.729
	Constant	-2.711	.176	238.526	1	.000	.066		

Because the odds of HTQ3"Ill health without access to medical care" is 2, this indicates that if the score of "Ill health without access to medical care" predictor changes from 0 (heard, witnessed and none) to 1 (experienced) then the probability of having PTSD symptoms for these participants doubles. In other words, by experiencing the traumatic event "Ill health without access to medical care", participants are twice as likely to suffer from PTSD symptoms. The CI value of HTQ3; " Ill health without access to medical care " gives us a reliable predictor for PTSD symptoms .Similarly, if a participant experiences event HTQ2; "Lack of food or water " the chance of experiencing PTSD symptoms becomes 3 times higher. In the case that a participant experiences event HTQ13; "Torture; physical or mental suffering", the probability of PTSD symptoms in these participant becomes 1.7 times higher. By changing the score of event HTQ27; "Confined to home because danger outside" the probability of having PTSD symptoms increases 1.86 times more. As the CI values for all of the mentioned predictors are more than one, we can confidently say that all these predictors are reliable for predicting PTSD symptoms.

From Table 11, Step4 the fitted model is:

$$\text{Logit(p)} = \log\left({}^{p}/_{1-p}\right)$$
$$= -0.271 + 0.62\ HTQ27 + 0.53\ HTQ13 + 1.16\ HTQ3 + 0.695\ HTQ2$$

, where p is the probability of having PTSD symptoms.

4.2.2 Results from Identifying Mental Health's Predictors from HTQ Questionnaire

The purpose of this test is to make a prediction of presence/absence of Mental Health disorder for each participant that experienced HTQ traumatic events. In this section the main purpose is to test which one of these traumatic events most efficiently predicts the mental health disorder. The dichotomous dependent variable (rghq.t) for Mental Health status in this report is created from the GHQ.T (GHQ-28 Total Score) variable in the Kosovo-2006 data set. “rghq.t” is 0 for GHQ total score which has low GHQ.T scores (GHQ.T < 11), and 1 for high GHQ.T scores (GHQ.T > 12). As mentioned in section 4.1, a risk table of HTQ variables should be prepared to select the most relevant predictors but this time with respect to the rghq.t dichotomous variable.

Table 12 Risk & Odds Ratio

	The risk for the cohort DSM_IV_PTSD=1	Odds Ratio (1/0)	lower	upper	Pearson Chi-square	Asimpt. sig. p
HTQ3	.403	3.814	2.679	5.431	59.048	0
HTQ15	.470	3.440	1.142	10.360	5.431	.020
HTQ39	.508	2.933	.935	9.201	3.728	.054
HTQ38	.511	2.914	.722	11.761	2.472	.116
HTQ33	.514	2.894	.405	20.684	1.230	.267
HTQ34	.514	2.894	.405	20.684	1.230	.267
HTQ36	.576	2.327	.619	8.752	1.653	.199
HTQ26	.591	2.203	1.083	4.480	4.973	.026
HTQ2	.602	2.016	1.450	2.803	17.703	0
HTQ17	.607	1.989	1.433	2.761	17.208	0
HTQ27	.601	1.983	1.428	2.754	17.006	0
HTQ4	.609	1.963	1.419	2.716	16.865	0
HTQ35	.640	1.936	.541	6.931	1.067	.302
HTQ32	.643	1.926	.320	11.612	.530	.467
HTQ5	.649	1.801	1.302	2.491	12.814	.000
HTQ7	.665	1.795	1.130	2.852	6.274	.012
HTQ19	.647	1.782	1.267	2.508	11.160	.001
HTQ23	.677	1.756	.999	3.087	3.915	.048
HTQ18	.685	1.728	.876	3.410	2.543	.111
HTQ20	.701	1.670	.690	4.042	1.319	.251
HTQ13	.727	1.571	.964	2.558	3.333	.068

HTQ24	.727	1.558	1.069	2.272	5.369	.020
HTQ1	.741	1.506	1.086	2.090	6.046	.014
HTQ46	.731	1.504	.984	2.299	3.582	.058
HTQ29	.773	1.440	.130	15.966	.089	.765
HTQ25	.595	1.386	.827	7.290	2.166	.141
HTQ37	.796	1.375	.788	2.399	1.265	.261
HTQ28	.837	1.281	.618	2.652	.445	.505
HTQ14	.875	1.203	.419	3.459	.118	.731
HTQ16	.902	1.152	.222	5.984	.028	.866
HTQ42	.908	1.141	.697	1.868	.275	.600
HTQ6	.944	1.081	.644	1.815	.087	0.768
HTQ40	.950	1.073	.556	2.071	.044	.834
HTQ21	.954	1.065	.749	1.514	.123	.725
HTQ43	.963	1.052	.640	1.730	.040	.842
HTQ8	.994	1.008	.714	1.422	.002	.965
HTQ45	1.032	.958	.192	4.786	.003	.958
HTQ11		.741	.711	.772	1.047	.306
HTQ12		.739	.709	.771	2.811	.094
HTQ9	1.261	.739	.373	1.464	.757	.384
HTQ10	1.292	.718	.080	6.457	.088	.766
HTQ44	1.389	.654	.332	1.287	1.530	.216
HTQ41	1.409	.643	.293	1.411	1.232	.267
HTQ22	2.353	.353	.080	1.549	2.077	.150
HTQ30						
HTQ31						

Regarding table 12, those predictors found by the Risk Table above are entered into the logistic regression procedure as predictors. In this phase the Forward Stepwise model selection procedure for entering the variables into the method is used. In the Forward Stepwise model selection procedure variables are sequentially added to an empty model (intercept only model). The outputs of this analysis are as follows:

Table 13 Chi-Square

Omnibus Tests of Model Coefficients				
		Chi-square	**df**	**Sig.**
Step 1	Step	63.911	1	0
	Block	63.911	1	
	Model	63.911	1	0
Step 2	Step	7.683	1	.006
	Block	71.594	2	0
	Model	71.594	2	0

The Omnibus Tests Table 13 demonstrates the overall fit of the model in step one where $\chi^2(1) = 63.911$ and $p < 0.01$ show that the variable which is added fits the model significantly. In the second step $\chi^2(1) = 71.594$ and $p < 0.01$ indicate that the model with the added new variable is still highly significant and the new predictors predict the Mental Status well.

Table 14 Variance for Overall Final Model

Model Summary			
Step	**-2 Log likelihood**	**Cox & Snell R Square**	**Nagelkerke R Square**
1	992.829	.068	.099
2	985.146	.076	.110

Model Summary table 14 demonstrates that the overall final model accounts 7.6 – 11.5 % of variance for Mental Health Status.

Table 15 Quality of Model: Sensitivity, Specificity

Classification Table					
	Observed		Predicted		
			rghq.t		Percentage Correct
			0	1	
Step 1	rghq.t	0	668	0	100
		1	243	0	0
	Overall Percentage				73.3
Step 2	rghq.t	0	609	59	91.2
		1	186	57	23.5
	Overall Percentage				73.1

Table 15 describes how well the model fits the data. We can conclude that the predictors indicate the Mental Health Status of patients 73.1% correctly and all the selected variables predict the Mental Health Status efficiently. The sensitivity value in this case is $\frac{57}{186+57} = 0.235$.Hence the sensitivity value for this model is 23%. It means, of all those who had Mental Health Disorders, we predicted only 23% correctly (see table 15). The specificity percentage for this model is $\frac{609}{609+59} = 0.912$. Hence the specificity for this model is 91%. It means, of those who did not have Mental Health Disorders, we predicted 91% correctly. False positive rate in this case is $\frac{59}{59+57} = 0.509$. This means of all those for whom we predicted Mental Health Disorders, we were 50% wrong. False negative rate is $\frac{186}{186+609} = 0.234$. This means of all those for whom we predicted not to have Mental Health Disorders 23.4% we predicted wrong.

Table 16 Logistic Regression Coefficients, Wald Statistic and Odds Ratio

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1a	HTQExp3	1.316	.164	64.248	1	0	3.729	2.703	5.144
	Constant	-1.405	.096	213.864	1	0	.245		
Step 2b	HTQExp3	1.246	.166	56.107	1	0	3.477	2.510	4.818
	HTQExp17	.443	.159	7.754	1	.005	1.558	1.140	2.128
	Constant	-1.563	.114	186.477	1	0	.209		

This model indicates that there is a significant predictive power for variables HTQExp3 "Ill health without access to medical care "(p-value=0), and HTQExp17" Extortion or robbery" (p-value=0.005).The odds that a participant who has experienced event "Ill health without access to medical care", has mental health problems is 3.7 times greater than the corresponding odds for a participant who has not experienced this event. Finally, because the odds of HTQExp17" Extortion or robbery" is 1.140, this shows that the risk of having mental health problems for the participants who have experienced this event tend to be 14% more than those who haven't. The $C.I_{.95}$ of both variables are clearly reliable predictors for mental health status because they are above 1.

4.2.3 Results from Identifying PTSD Predictors from the Displacement and Refugee Status

The objective of this test is to predict the presence/absence of PTSD symptoms on Displacement and Refugee Status in the surveyed population to determine whether the selected Displacement and Refugee Status describes adequately PTSD symptoms. Table 6 demonstrates the description of Displacement and Refugee Status from Data set Kosovo_2006. The summary of SPSS 18 output from running stepwise logistic regression is illustrated in tables 17,18 and 19 .The predictor DEM6 "Are you currently employed?", was found in one step. In the first phase of this analysis Table 17 demonstrates the Omnibus Tests of model .The Chi-square values $\chi^2(1) = 10.037, p < 0.002$ show that the model with added predictor DEM6" Are you currently employed?" fits the data significantly better than without this predictor. Similarly, the Block chi-square shows that the effects of DEM6"Are you currently employed?" entered here significantly differs from 0.

Table 17 Chi-Square

Omnibus Tests of Model				
		Chi-Square	df	Sig.
Step1	Step	10.037	1	.002
	Block	10.037	1	.002
	Model	10.037	1	.002

Model Summary table 18 shows that the overall final model accounts for 2.9 – 4.4 % of variance for PTSD symptoms.

Table 18 Variance for Overall Final Model

Model Summery			
Step	-2Log Likelihood	Cox & Snell R Square	Nagelkerke R Square
1	361.117	.029	.044

The Classification Table 19 shows that 75.9% of cases can be correctly classified by using this DEM6 “Are you currently employed?” predictor.

Table 19 Quality of Model: Sensitivity, Specificity

Classification Table					
Observed			**Predicted**		
			PTSD Caseness According to CDC definition		**Percentage Correct**
			0	1	
Step 1	PTSD Caseness According to CDC definition	0	255	0	100
		1	81	0	0
	Overall Percentage				75.9

Table 20 Logistic Regression Coefficients, Wald Statistic and Odds Ratio

Variables in the Equation							
		B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**
Step 1	DEM6	1.335	.488	7.499	1	.006	3.8
	Constant	-3.657	.947	14.921	1	0	.026

The odds of having PTSD symptoms for unemployed participants are 33% higher than those who are in employment.

4.2.4 Results from Identifying Mental Health Symptoms (GHQ-28) from MOS Questionnaire

The purpose of this test is to make a prediction of presence/absence of Mental Health for every participant using their MOS items. As in the previous chapter described, MOS items are health related quality of life issues such as physical functioning, bodily pain, role functioning, social functioning and mental health. In this part the main aim is to test which one of these health items predicts Mental Health status. The dichotomous dependent variable (rghq.t) presents symptoms for nonspecific psychiatric morbidity in this report which is created from GHQ.T (GHQ-28 Total Score) variable in Kosovo-2006 dataset. rghq.t is 0 for a higher number of symptoms for nonspecific psychiatric morbidity (GHQ.T < 11) and 1 for lower GHQ.T (GHQ.T > 12)scores. Note that the higher the score for General Health issues, the worse the situation is.

The Chi-square values from Omnibus Tests Table 21 and their associated p-values < 0.05 show that each variable which is added to the model in each step fits the model significantly. We can hereby conclude that each MOS issue predicts Mental Health status 73% correctly.

Table 21 Chi-Square

Omnibus Tests of Model Coefficients				
		Chi-square	**df**	**Sig.**
Step 1	Step	135.743	1	.000
	Block	135.743	1	.000
	Model	135.743	1	.000
Step 2	Step	31.799	1	.000
	Block	167.542	2	.000
	Model	167.542	2	.000
Step 3	Step	12.054	1	.001
	Block	179.596	3	.000
	Model	179.596	3	.000
Step 4	Step	7.754	1	.005
	Block	187.350	4	.000
	Model	187.350	4	.000
Step 5	Step	5.732	1	.017
	Block	193.083	5	.000
	Model	193.083	5	.000

The Model Summary table 22 demonstrates that overall the final model accounts for 30.5-41% of the variance in Mental Health status.

Table 22 Variance for Overall Final Model

Model Summary			
Step	**-2 Log likelihood**	**Cox & Snell R Square**	**Nagelkerke R Square**
1	588.057	.226	.303
2	556.258	.271	.364
3	544.204	.287	.386
4	536.450	.298	.400
5	530.718	.305	.410

Table 23 describes how well the model fits the data. The predictors signify patients Mental Health Status 74% correctly and all the collective set of variables predicts Mental Health Status efficiently. The sensibility value for this model is 73.6% which shows of all who has mental health disorders we predict 73.6% correctly.

Table 23 Quality of Model: Sensitivity, Specificity

Classification Table					
	Observed		Predicted		
			rghq.t		Percentage Correct
			0	1	
Step 1	rghq.t	0	179	124	59.1
		1	26	201	88.5
	Overall Percentage				71.7
Step 2	rghq.t	0	179	124	59.1
		1	26	201	88.5
	Overall Percentage				71.7
Step 3	rghq.t	0	203	100	67.0
		1	41	186	81.9
	Overall Percentage				73.4
Step 4	rghq.t	0	240	63	79.2
		1	74	153	67.4
	Overall Percentage				74.2
Step 5	rghq.t	0	225	78	74.3
		1	60	167	73.6
	Overall Percentage				74.0

Table 24 Logistic Regression Coefficients, Wald Statistic, and Odds Ratio

Variables in the Equation									
		B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95% C.I. for EXP(B)**	
								Lower	**Upper**
Step 1^a	MOS_Body_Pain52	-2.412	.239	101.935	1	.000	.090	.056	.143
	Constant	.483	.114	17.892	1	.000	1.621		
Step 2^b	MOS_Body_Pain52	-2.178	.245	78.967	1	.000	.113	.070	.183
	MOS_Rolefunct52	1.247	.226	30.374	1	.000	3.479	2.233	5.420
	Constant	.020	.140	.021	1	.886	1.020		
Step 3^c	MOS_Body_Pain52	-2.014	.250	64.954	1	.000	.133	.082	.218
	MOS_General_Health52	-.964	.286	11.372	1	.001	.381	.218	.668
	MOS_Rolefunct52	1.082	.231	21.993	1	.000	2.952	1.878	4.640
	Constant	.220	.152	2.078	1	.149	1.245		
Step 4^d	MOS_Body_Pain52	-1.955	.251	60.465	1	.000	.142	.086	.232
	MOS_Mental52	-.601	.216	7.764	1	.005	.548	.359	.837
	MOS_General_Health52	-.911	.288	9.968	1	.002	.402	.229	.708
	MOS_Rolefunct52	1.003	.234	18.303	1	.000	2.726	1.722	4.315
	Constant	.486	.182	7.145	1	.008	1.625		
Step 5^e	MOS_Physical52	-.552	.229	5.784	1	.016	.576	.367	.903
	MOS_Body_Pain52	-1.911	.253	57.068	1	.000	.148	.090	.243
	MOS_Mental52	-.567	.218	6.786	1	.009	.567	.370	.869
	MOS_General_Health52	-.817	.292	7.825	1	.005	.442	.249	.783
	MOS_Rolefunct52	.815	.248	10.841	1	.001	2.260	1.391	3.671

	Constant	.795	.225	12.497	1	.000	2.215		

The Logit-coefficient of MOS_Physical is a negative value, therefore as the participants Physical Health status goes up, the chance of nonspecific-psychiatric-morbidity symptoms decrease. The odds ratio of Body_Pain is also negative, therefore as the value of MOS_Body_Pain improved by one unit, the risk of suffering from symptoms of nonspecific – psychiatric- morbidity decrease. Because the Logit-coefficient for MOS_Mental is -0.57, Mental Health status rate for the participants go over 50%. The participants tend to have fewer symptoms for nonspecific psychiatric morbidity. The Exp (B) value of General_Health is negative, therefore the probability of having symptoms for nonspecific psychiatric morbidity for higher General_Health score is less than the participants with a low General_Health score. In other words, the Exp (B) shows that the risks for those with ill health have an increased probability of developing Mental Health problems. The Exp (B) value of MOS_Rollfunct is positive which shows that participants have better roll functioning therefore they are likely to have 39% more symptoms for nonspecific psychiatric morbidity.

4.2.5 Results from Appling Block Logistic Regression Measured by HTQ, DEM and MOS

The purpose of this analysis is to determine the factors that influence (PTSD) Post Traumatic Stress Disorder symptoms regarding the HTQ events, DEM Characteristics and MOS items. The output measure is whether or not PTSD symptoms are present. The predictor variables are DEM6 “Are you currently employed?”; HTQExp2: ”Lack of food or water”; HTQExp3: “ill health without access to medical care”; HTQExp27: “Confined to home because danger outside”; HTQExp13: “Torture; physical or mental suffering”; MOS_Physical: “Physical Health”; MOS_Body_Pain: “Body pain”; MOS_Mental: “Mental Health”; MOS_Social: “Social Functioning”; MOS_General_Health: “General Health”; MOS_Rolefunc: “Role functioning”. The applied analysis for this purpose is to run Hierarchical Logistic Regression entering into the first block: HTQExp2, HTQExp3, HTQExp27, and HTQExp13. DEM6 is placed into the second block, while MOS items are put into the third block. SPSS creates a Regression Model for the HTQ variables specified in block 1, and creates the second model for DEM6 variable in block 2, while in the last model it creates the third block for MOS variables. The results of block 1 are shown in the following tables. The forced entry method is used on all 3 blocks.

I. Block 1

The following tables give us the information about block 1, such as the model after the HTQ variables have been added. The -2Log Likelihood in Table 21 has dropped to 604.668 which is a change of 65.591 (this value is given by the model Chi-square in Table 25). This value describes that the model is an improvement on not adding any variables. The Chi-square $\chi^2(4) = 65.59 \; , p < 0.0001$ shows the model is statistically highly significant and also using such trauma events as: HTQExp2: “Lack of food or water”, HTQExp3: “Ill health without access to medical care”; HTQExp27: “Confined to home because danger outside”; and HTQExp13: “Torture; physical or mental suffering” these predictors significantly improve the ability to predict PTSD symptoms. The Classification Table 28 shows that 71.4% of cases can be correctly classified by using these 4 predictors.

Table 25 Chi-Square

Omnibus Tests of Model Coefficients				
		Chi-square	**df**	**Sig.**
Step 1	Step	65.591	4	.000
	Block	65.591	4	.000
	Model	65.591	4	.000

Table 26 Variance for Overall Final Model

Model Summary			
Step	**-2 Log likelihood**	**Cox & Snell R Square**	**Nagelkerke R Square**
1	604.668	.111	.159

In Table 27, the “Hosmer and Lemeshow goodness-of-fit test” statistic results demonstrate whether the observed data is significantly different from the predicted values of the model.

It is expected that we arrive at a non-significant value for this test because the model interprets and predicts our real-world data well. The Chi-square is $\chi^2(5) = 2.60 \, , p = \; 0.761$ which shows statistically non-significant values.

Table 27 Chi-Square

Hosmer and Lemeshow Test			
Step	**Chi-square**	**df**	**Sig.**
1	2.601	5	.761

Table 28 Quality of Model: Sensitivity, Specificity

Classification Table					
	Observed		**Predicted**		
			Caseness according to DSM-IV PTSD cut off score of >2.5		**Percentage Correct**
			.00	1.00	
Step 1	Caseness according to DSM-IV PTSD cut off score of >2.5	.00	339	54	86.3
		1.00	105	57	35.2
	Overall Percentage				71.4

Table 29: Variable in the Equation shows the description of the parameters in the first block of the model. The significance values of the Wald statistic for predictors indicate that HTQExp3:"Ill health without access to medical care" ($Wald = 15.328, p < 0.001$ and HTQExp27: "Confined to home because danger outside" ($Wald = 8.673, p = 0.003$) both significantly predict PTSD symptoms. HTQExp2 "Lack of food or water" ($Wald = 2.539, p = 0.111$) and HTQExp13 "Torture (physical or mental suffering)" ($Wald = 2.948, p = 0.086$) however don't significantly predict PTSD symptoms.

Table 29 Logistic Regression Coefficients, Wald Statistic, and Odds Ratio

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1a	HTQExp2	.382	.240	2.539	1	.111	1.465	.916	2.345
	HTQExp3	.948	.242	15.328	1	.000	2.579	1.605	4.145
	HTQExp27	.621	.211	8.673	1	.003	1.860	1.231	2.812
	HTQExp13	.475	.277	2.948	1	.086	1.608	.935	2.765
	Constant	-1.845	.181	104.224	1	.000	.158		

The values Exp (B) for HTQExp2"Lack of food or water" (Exp (B) = 1.465,$CI_{0.95} = 0.916\ , 2.345)$ indicate that as "Lack of food or water" changes from 0 (heard, witnessed and none) to 1 (experienced), then the odds of having PTSD symptoms increases because Exp (B) is greater than 1. However, the interval value crosses 1, which limits the generalization of this result because the Exp (B) value in the population indicates either a positive (Exp (B) > 0) or a negative value (Exp (B) < 0) relationship. Therefore "Lack of food or water" is not a reliable predictor of PTSD symptoms.

The values of Exp (B) for HTQExp3 "Ill health without access to medical care" (Exp (B) = 2.579,$CI_{0.95} = 1.605\ , 4.145)$ indicate that if the value of "Ill health without access to medical care" changes from 0 (heard, witnessed and none) to 1 (experienced), then the odds of having PTSD symptoms also increase (because the value of Exp (B) is greater than 0). The ranges of confidence interval are $(1.605\ to\ \ 4.145)$, so we can be very confident that the value of Exp (B) in the population lies between these two values. Those participants who experienced "Ill health without access to medical care" also have PTSD symptoms.

The values of Exp (B) for HTQExp27"Confined to home because danger outside" (Exp (B) = 1.860,$CI_{0.95} = 1.231\ , 2.812)$ are similar to the aforementioned HTQExp3, whereby the participants who experience "Confined to home because danger outside" also have PTSD symptoms, therefore we can be confident of its reliability because the value Exp (B) lies between confidence interval$(1.605\ , 4.145)$.

The values of Exp (B) for HTQExp13 "Torture; physical or mental suffering "(Exp (B) = 1.608,$CI_{0.95} = 0.935\ , 2.765)$ are similar to HTQExp2 showing that the participants who experience "Torture; physical or mental suffering" have increased odds of suffering PTSD symptoms but because of the ranges of Confidence Interval $(0.935\ , 2.765)$ HTQ13 "Torture; physical or mental suffering" is not a completely reliable predictor for PTSD symptoms.

II. Block 2

The Output below demonstrates the new model after adding the new predictor DEM6"Are you currently employed?" to the model. The effect of adding this predictor to the model is that the -2 log-likelihood reduces to 587.904 (a reduction of 82.355 from the original mode) as demonstrated in the model Chi-square Table 30, and an additional reduction of 16.764 from the reduction caused by block 1 as demonstrated by the block statistics. This further improvement of block 2 is statistically highly significant ($\chi^2(1) = 16.746\ , p =\ 0$) which indicates that including the new predictor DEM6 "Are you currently employed?" in the model has significantly improved our ability to predict PTSD symptoms. Classification Table 32 shows that the model is now correctly classifying 72.3% of cases, which in block1 were 71.4% correctly classified. This extra 0.9% of cases now classified is not significantly higher, with only 5 cases being correctly added. The sensitivity value also is improved 3%.

Table 30 Chi-Square

Omnibus Tests of Model Coefficients				
		Chi-square	**df**	**Sig.**
Step 1	Step	16.764	1	.000
	Block	16.764	1	.000
	Model	82.355	5	.000

Table 31 Variance for Overall Final Model

Model Summary			
Step	**-2 Log likelihood**	**Cox & Snell R Square**	**Nagelkerke R Square**
1	587.904	.138	.197

Table 32 Quality of Model: Sensitivity, Specificity

Classification Table					
	Observed		**Predicted**		
			Caseness according to DSM-IV PTSD cut off score of >2.5		**Percentage Correct**
			.00	1.00	
Step 1	Caseness according to DSM-IV PTSD cut off score of >2.5	.00	339	54	86.3
		1.00	100	62	38.3
	Overall Percentage				72.3

Table 33 Logistic Regression Coefficients, Wald Statistic, and Odds Ratio

Variables in the Equation									
		B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95% C.I. for EXP(B)**	
								Lower	Upper
Step 1a	HTQExp2	.485	.245	3.921	1	.048	1.624	1.005	2.625
	HTQExp3	.778	.249	9.754	1	.002	2.177	1.336	3.547
	HTQExp27	.612	.214	8.194	1	.004	1.843	1.213	2.802
	HTQExp13	.633	.288	4.822	1	.028	1.883	1.070	3.311
	DEM6	1.595	.457	12.181	1	.000	4.929	2.012	12.073
	Constant	-4.897	.913	28.740	1	.000	.007		

As in the Table 33 demonstrates the Wald test of new model shows that all variables are statistically significant (all $p < 0.05$).The values of Exp (B) for all variables in bock 2 are greater than 1 (see Table 33) and the associated Confidence Intervals for all variables in block 2 are greater than one too. These values indicate that all predictors in the model are reliable predictors.

III. Block 3

In this block the MOS predictors are added to the model and the following tables analyzes the model with the MOS variables. The effect of adding MOS predictors to the model is to reduce the -2 log-likelihood to 486.07 (a reduction of 101.832 from block 2 as shown by the block statistics in Table 34). This additional improvement of block 3 is significant ($\chi^2(6) = 101.832\ , p = 0 < 0.01$) which indicate that including these MOS predictors in the model has significantly improved our ability to predict PTSD symptoms. The Classification Table 36 shows that the model is now correctly classifying 78.7% of cases which in block 2 it was only 72.3% correctly classified and so an extra 6.4% of cases are now classified.

Table 34 Chi-Square

Omnibus Tests of Model Coefficients				
		Chi-square	**df**	**Sig.**
Step 1	Step	101.832	6	0
	Block	101.832	6	0
	Model	184.187	11	0

Table 35 Variance for Overall Final Model

Model Summary			
Step1	**-2 Log likelihood**	**Cox & Snell R Square**	**Nagelkerke R Square**
	486.07	0.28	0.40
a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.			

Table 36 Quality of Model: Sensitivity, Specificity

Classification Table					
Observed			**Predicted**		
			Caseness according to DSM-IV PTSD cut off score of >2.5		**Percentage Correct**
			0	1	
Step 1	Caseness according to DSM-IV PTSD cut off score of >2.5	0	349	44	88.8
		1	74	88	54.3
	Overall Percentage				78.7
a. The cut value is .500					

Table 37 shows Variable in Equation contains all predictors and represents the details of the final model. The significant values of the Wald statistics for each predictor indicate that HTQExp27"Confined to home because danger outside" (Wald = 9.361, p = 0.002 <0.05), DEM6"Are you currently employed?" (Wald = 6.726, p =0.01 <0.05), MOS_Physical (Wald = 7.000, p = .0.008 <0.05), MOS_Body_Pain (Wald = 19.758, p = 0 <0.05) and MOS_General_Health (Wald = 11.000, p =0.001 <0.05) still significantly predict PTSD symptoms but HTQExp2 "Lack of food or water" (Wald = 3.573, p = 0.059 >0.05) HTQExp3 "Ill health without access to medical care" (Wald = 0.680, p = 0.410 >0.05), HTQExp13"Torture; physical or mental suffering "(Wald = 3.068, p = 0.080>0.05), MOS_Mental (Wald = 0.263, p =0.608>0.05), MOS_Social (Wald = 0.780, p =0.377>0.05) and MOS_Rolefunct (Wald = 0.992, p = 0.319>0.05) don't predict PTSD symptoms significantly.

Table 37 Logistic Regression Coefficients, Wald Statistic, and Odds Ratio

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1	HTQExp2	.513	.271	3.573	1	.059	1.670	.981	2.841
	HTQExp3	.229	.278	.680	1	.410	1.257	.730	2.167
	HTQExp13	.572	.327	3.068	1	.080	1.772	.934	3.361
	HTQExp27	.741	.242	9.361	1	.002	2.097	1.305	3.371
	DEM6	1.266	.488	6.726	1	.010	3.548	1.363	9.240
	MOS_Physical	-.011	.004	7.000	1	.008	.989	.981	.997
	MOS_Body_Pain	-.022	.005	19.758	1	.000	.978	.969	.988
	MOS_Mental	-.003	.007	.263	1	.608	.997	.984	1.010
	MOS_Social	.003	.004	.780	1	.377	1.003	.996	1.011
	MOS_General_Health	-.023	.007	11.000	1	.001	.978	.965	.991
	MOS_Rolefunct	.003	.003	.992	1	.319	1.003	.997	1.010
	Constant	-2.046	1.073	3.638	1	.056	.129		

By controlling the CI in the Table 37 for the significant predictors as in previous block, we can see HTQExp27"Confined to home because danger outside" has Exp (B) = 2.097 (which is greater than 1) and Confidence Interval (1.305, 3.371) indicate that if the event "Confined to home because danger outside "changes from 0 (heard, witnessed and none) to 1 (experienced) then the odds of suffering from PTSD symptoms increases and we can also be confident that this relationship in this sample is true of the whole population.

The value of Exp (B) for DEM6"Are you currently employed?" (Exp (B) = 3.548 and CI (1.363, 9.240) indicate if "Are you currently employed?" changes from 0 (no) to 1 (yes) then the odds of suffering from PTSD symptoms increases and we can also be confident that DEM6 is also a reliable predictor for PTSD symptoms.

The values of Exp (B) for MOS_Physical (Exp (B) = 0.989,$CI_{0.95} = 0.981, 0.997$) indicates that if the MOS_Physical decreases by one point, then the odds of suffering from PTSD symptoms increase. The CI does not cross 1 so we can be confident that the relationship between MOS_Physical and PTSD symptoms found in this sample would be found in 95% of samples from the same population. As MOS_Physical status improves by one unit the participants are about 1.9 times less suffering from PTSD symptoms.

For MOS_Body_Pain the values of (Exp (B) = .978,$CI_{0.95} = 0.969, 0.988$) indicates that if the MOS_Body_Pain improves then the odds of suffering from PTSD symptoms decreases but it is not a reliable predictor for PTSD symptoms.

4.3 Results from Applying the General Linear Model to assess the Effect of HSCL Symptoms on PTSD Symptoms

As mentioned in chapter 3, the Hopkins Symptom Checklist consists of a screening tool to detect symptoms of anxiety and depression. In this section we wish to assess symptoms of depression and anxiety regarding the people suffering from PTSD by using the General Linear Model procedures of SPSS 18.

I. Depression

For this purpose the GLM procedure of SPSS 18 is run over the HSCL_Depression mean score as a dependent variable and PTSD_CDC as a factor. The outputs of the GLM procedure are as follows:

Table 38 Factor Information

Between-Subjects Factors		
		N
PTSD_CDC PTSD Caseness according to CDC definition	0	903
PTSD_CDC PTSD Caseness according to CDC definition	1	256

The Between-Subjects factors information table 38 displays any value labels defined for PTSD_CDC factor in our Dataset.

Out of 1159 participants, 256 suffer from PTSD symptoms.

Table 39 Descriptive Statistics

Descriptive Statistics			
Dependent Variable: HSCL_Depression mean score (SD1-SD15)			
PTSD_CDC	Mean	Std. Deviation	N
0	1.5414	0.42464	903
1	2.6343	0.52377	256
Total	1.7828	0.63764	1159

Table 39 demonstrates descriptive statistics elements such as means and the standard deviation of PTSD factor.

Table 40 Tests of Between-Subjects Effects

Dependent Variable: HSCL_Depression mean score (SD1-SD15)					
Source	Type III Sum of Squares	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	238.223a	238.22	1184.96	0	0.51
Intercept	3477.72	3477.72	17298.72	0	0.94
PTSD_CDC	238.22	238.22	1184.96	0	0.51
Error	232.60	0.20			
Total	4154.51				
Corrected Total	470.83				
R Squared = .506 (Adjusted R Squared = .506)					

From the significant value 0 (which is lower than 0.05) of the Corrected Model, Table 40, shows that the overall model is significant and the effect size is $Partial\ Eta^2 = R^2 = 0.51$ which means that the model explains 51% of variance in the depression Symptom. The F-Test in this table is the overall test which shows whether the GLM model for our assumptions are acceptable. As the f ratio is significant with value 0, there are differences in the means hence the PTSD certainly has a significant effect on depression symptom.

II. Anxiety

In this part the selected factor for assessing PTSD Symptoms is DSM_IV_PTSD variable which is a combination of various trauma events and is taken from the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV see Chapter 3). The Output from the GLM procedure of SPSS 18 is as follows:

Table 41 Factor Information

Between-Subjects Factors		
		N
DSM_IV_PTSD Caseness according to DSM-IV PTSD cut off score of >2.5	0	952
	1	207

The Factor Information of the PTSD factor is demonstrated in the Between Subjects Factors table 41. It shows that from 1159 participants, 207 people suffer from PTSD symptoms and 952 people have no PTSD.

Table 42 Descriptive Statistics

Dependent Variable: HSCL Anxiety mean score (SA1-SA10)			
DSM_IV_PTSD Caseness according to DSM-IV PTSD cut off score of >2.5	**Mean**	**Std. Deviation**	**N**
0	1.52	0.44	952
1	2.93	0.44	207
Total	1.77	0.69	1159

In table 42 the Descriptive Statistics, such as mean and standard deviation, are demonstrated.

Table 43 Tests of Between-Subjects Effects

Dependent Variable: HSCL_Anxiety HSCL Anxiety mean score (SA1-SA10)						
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	335.012a	1	335.012	1693.249	0	0.594
Intercept	3378.584	1	3378.584	17076.34	0	0.937
DSM_IV_PTSD	335.012	1	335.012	1693.249	0	0.594
Error	228.915	1157	0.198			
Total	4226.546	1159				
Corrected Total	563.927	1158				
a. R Squared = .594 (Adjusted R Squared = .594)						

Due to the Corrected Model in table 43 having the significant value of 0, the overall model is significant and also gives us the effect size which is $Partial\ Eta^2 = R^2 = 0.594$. This means that people's existing PTSD symptoms explains 59.4% of the variance in anxiety Symptom. The main effect of PTSD symptoms is shown by the F-ratio in table 43 where the significant value is 0, which is lower than 0.05. Hence there are differences in group means and PTSD certainly has a significant effect on anxiety symptoms.

4.4 Results from Correlation and Nonparametric Correlation for PTSD Symptoms

The objective of data analyzing in this section is to assess the correlation between HTQ–Trauma Event Experiences and PTSD (Post Traumatic Stress Disorder) symptoms. The Bivariate Correlation procedure from SPSS 18 is used for this purpose. This procedure takes Pearson's r-method, which is the most standard type of correlation, and uses a value as a measure of an association which varies from -1 to 1, with 0 indicating no relationship, and 1 indicating perfect relationship. The three variables that are presented here are: PTSD_sum CDC; which is built from Diagnosis criteria for PTSD (see chapter 3) MEAN_HTQ; DSM_IV_PTSD Score.

As the output of this test shows (in Table 44), the measure of Pearson's r is 0.968 which indicates a perfect relationship between PTSD and HTQ_mean score.

This perfect correlation between PTSD and trauma events shows that the higher the value of trauma events experienced, the more people display PTSD symptoms.

Variables: PTSD_sum CDC, MEAN_HTQ, DSM_IV_PTSD Score

Methods: Bivariate Correlations

Table 44 Pearson Correlations

Correlations		
		DSM_IV_PTSD Score
PTSD_sum CDC	Pearson Correlation	0.882
	Sig. (2-tailed)	0
	N	1161
MEAN_HTQS Mean Score in HTQ	Pearson Correlation	0.968
	Sig. (2-tailed)	0
	N	1161

Variables: PTSD_sum CDC, MEAN_HTQS, DSM_IV_PTSD Score.

By applying nonparametric test the same results are obtained.

4.5 Results from Comparing Depression Symptom with PTSD Symptoms

The first phase for examining the influence of one variable over another variable is to create crosstabs tables. The outputs after performing the Crosstabs function between DSM-IV PTSD variable and HSCL_Depression are shown in the tables below:

The Case Processing Summary Table 45 shows the 1159 Valid cases, 2 Missing, and 1161 Total cases. The low percent of missing cases here reflects the people who were not asked this particular question in the survey.

Table 45 Factor Information

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
DSM-IV PTSD cut off score of >2.5 * HSCL Depression Cutoff	1159	99.8%	2	0.02%	1161	100%
PTSD_CDC definition * HSCL Depression Cutoff	1159	99.8%	2	0.02	1161	100%

The Crosstabs Table 46 shows that out of 1159 valid cases in this test, 82.1% (951 people) do not have PTSD symptoms while in 99.7% of cases no depression is recorded. In only 2 cases are PTSD symptoms recorded but they do not have depression. Hence only 1% of participants have positive PTSD and negative depression. In other words 99% of the participants who have PTSD symptoms also have depression.

Table 46 Crosstabs

Crosstab					
			HSCL Depression Cutoff		Total
			0	1	
DSM-IV PTSD cut off score >2.5	0	Count	673	278	951
		% within DSM-IV PTSD cut off score >2.5	70.8%	29.2%	100%
		% within HSCL Depression Cutoff	99.7%	57.4%	82.1%
		% of Total	58.1%	24%	82.1%
	1	Count	2	206	208
		% within DSM-IV PTSD cut off score of >2.5	1%	99%	100%
		% within HSCL Depression Cutoff	0.3%	42.6%	17.9%
		% of Total	.0.2%	17.8%	17.9%
Total		Count	675	484	1159
		% within DSM-IV PTSD cut off score of >2.5	58.2%	41.8%	100%
		% within HSCL Depression Cutoff	100%	100%	100%
		% of Total	58.2%	41.8%	100%

Once again the Chi-square test from subcommand of the Crosstabs procedure is selected to obtain the test statistic and its associated P-value. The Table 47 presents the output of this function. The results indicate that there is a statistically significant relationship between depression and PTSD symptoms (Criteria DSM-IV) with the Chi-Square value 341.95 and p-value <0.05.

Table 47 Chi-Square Test

Chi-Square Tests f DSM-IV PTSD		
	Value	**Exact Sig. (2-sided)**
Pearson Chi-Square	341.95	.000
N of Valid Cases	1159	

Table 48 Chi-Square Test

Chi-Square Tests f PTSD CDC		
	Value	**Exact Sig. (2-sided)**
Pearson Chi-Square	410.38	.000
N of Valid Cases	1159	

Table 48 shows the same results for CDC criteria of PTSD variable. It‘s mean there is also a statistically significant relationship between PTSD-CDC criteria and depression with Chi-square value 410.38 and P-value 0.

Table 49 Reliability Coefficient

Symmetric Measures					
		Value	**Asymp. Std. Error**	**Approx. Tb**	**Approx. Sig.**
Nominal by Nominal	Contingency Coefficient	.477			.000
Measure of Agreement	Kappa	.460	.024	18.492	.000
N of Valid Cases		1159			

Symmetric Measure Table 49 demonstrates the reliability coefficient for two variables, PTSD symptoms (DSM-IV Criteria) and depression cutoff score. As the Kappa value is 0.460 (which in order to be reliable should be more than 0.7 and positive), in this case the relationship is statistically of high significance but not of high reliability.

Table 50 Crosstabs

Crosstab					
			HSCL Depression Cutoff		**Total**
			0	1	
PTSD CDC definition	.00	Count	667	236	903
		% within PTSD CDC definition	73.9%	26.1%	100%
		% within HSCL Depression Cutoff	99.8%	48.8%	77.9%
		% of Total	57.5%	20.4%	.77.9%
	1.00	Count	8	248	256
		% within PTSD CDC definition	3.1%	96.9%	100%
		% within HSCL Depression Cutoff	1.2%	51.2%	22.1%
		% of Total	0.7%	21.4%	22.1%
Total		Count	675	484	1159
		% within PTSD CDC definition	58.2%	41.8%	100%
		% within HSCL Depression Cutoff	100%	100%	100%
		% of Total	58.2%	41.8%	100%

The results of the Crosstabs procedure over PTSD (CDC Criteria) and depression cut off score are demonstrated in Table 50. Out of 1159 participants, 667 (57.5%) valid cases have neither depression nor PTSD symptoms. 20.4% of valid cases have both PTSD symptoms as well as depression. There are only 8 (0.7%) participants displaying PTSD symptoms, however they don't have depression which supports a strong relationship between depression and PTSD (CDC Criteria).

Table 51 Pearson Correlation

Correlations		
		HSCL_Depression mean score (SD1-SD15)
PTSD_sum CDC compute HTQ cut Recurrence and Avoidance and Arousal	Pearson Correlation	0.804
	Sig. (2-tailed)	0
	N	1159
DSM_IV_PTSD Score DSM-IV PTSD Score According to Bosnian manual (Items 1-16)	Pearson Correlation	0.914
	Sig. (2-tailed)	0
	N	1159

In Correlation Table 51 the Pearson Correlation has a value of 0.914 for variable PTSD(CDC Criteria) indicating a perfect relationship between depression symptoms and PTSD symptoms, but we also see the Pearson Correlation value reduces to 0.804 for variable PTSD(DSM_IV_PTS criteria).

Figure 1 Scatter plot of PTSD and Depression

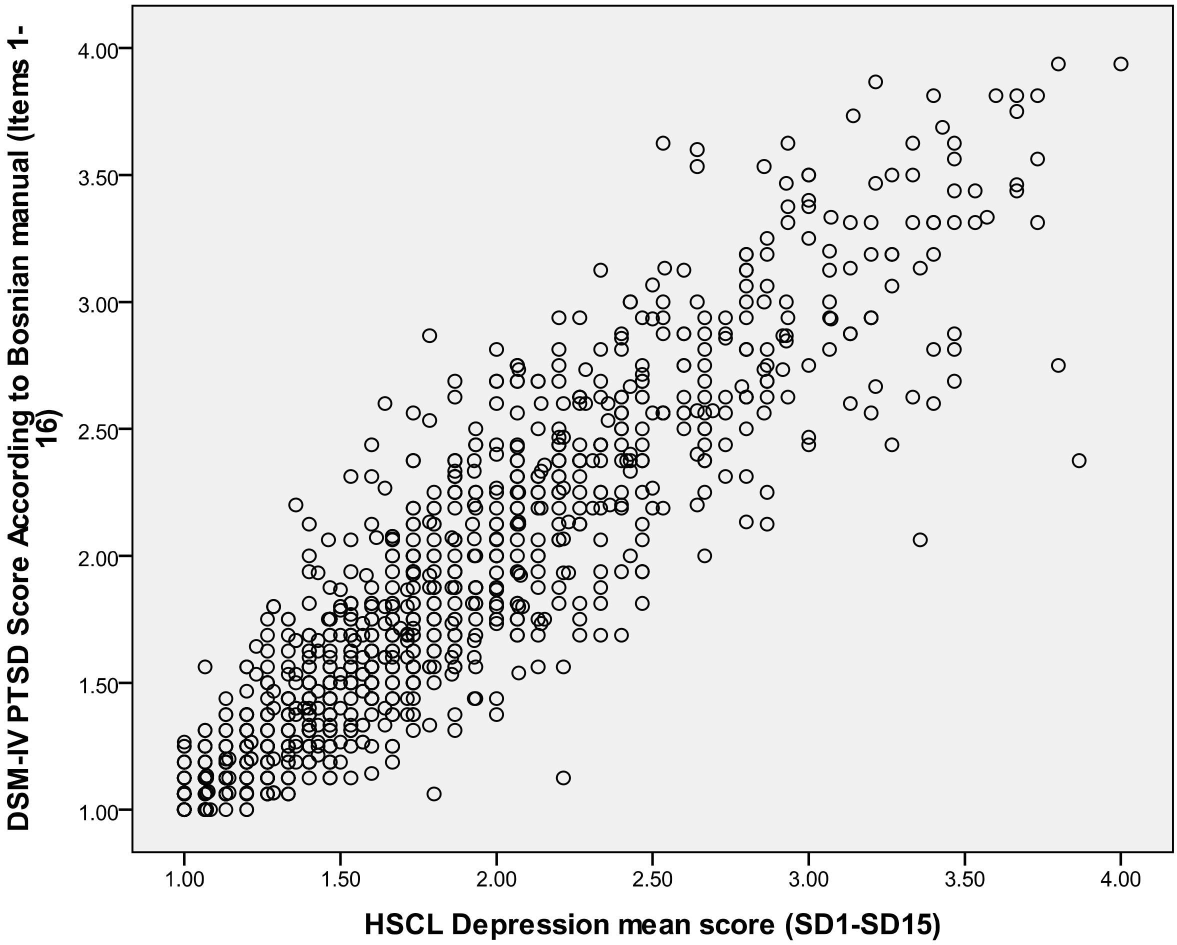


Figure 1 illustrates the perfect relationship between PTSD symptoms variable and depression.

4.6 Results from Comparing Anxiety Symptoms with PTSD Symptoms

For representing the relationship between anxiety symptoms with DSM_IV_PTSD more accurately, the crosstab procedure is considered. As shown in Table 52, 100% of the respondents with positive DSM_IV_PTSD also had anxiety symptoms.

Table 52 Crosstabs

Crosstab					
			HSCL Anxiety Cut off score		Total
			0	1	
DSM_IV_PTSD Caseness according to DSM-IV PTSD cut off score of >2.5	0	Count	676	276	952
		% within DSM_IV_PTSD Caseness according to DSM-IV PTSD cut off score of >2.5	71%	29%	100%
		% within HSCL_Anxiety Cut HSCL Anxiety Cut	100%	57.10%	82.10%
		% of Total	58.30%	23.80%	82.10%
DSM_IV_PTSD Caseness according to DSM-IV PTSD cut off score of >2.5	1	Count	0	207	207
		% within DSM_IV_PTSD Caseness according to DSM-IV PTSD cut off score of >2.5	0%	100%	100%
		% within HSCL_Anxiety Cut HSCL Anxiety Cut	0%	42.90%	17.90%
		% of Total	0%	17.90%	17.90%
Total		Count	676	483	1159
		% within DSM_IV_PTSD Caseness according to DSM-IV PTSD cut off score of >2.5	58.30%	41.70%	100%
		% within HSCL Anxiety Cut	100%	100%	100%
		% of Total	58.30%	41.70%	100%

For the Chi-Square test in Table 53, the McNemar Test is chosen which is a nonparametric test and is suitable for psychological studies. The null hypothesis assumes that the anxiety symptom has no affect on Post Traumatic Disorder (PTSD). In the above table the significant value for McNemar's Chi-Square is 0, so the null hypothesis is rejected. It is concluded that there is certainly a relationship between anxiety symptoms and PTSD.

Table 53 Chi-Sqaure

Chi-Square Tests		
	Value	**Exact Sig. (2-sided)**
McNemar Test		0
N of Valid Cases	1159	
Binomial distribution used.		

Table 54 Reliability Coefficient

Symmetric Measures					
		Value	**Asymp. Std. Error**	**Approx. Tb**	**Approx. Sig.**
Nominal by Nominal	Contingency Coefficient	0.53			0
Measure of Agreement	Kappa	0.56	0.02	21.10	0
N of Valid Cases		1159			
a. Not assuming the null hypothesis.					
b. Using the asymptotic standard error assuming the null hypothesis.					

Symmetric Measures are presented in table 54, which shows the reliability coefficient for PTSD symptoms (DSM-IV criteria) and anxiety cutoff score. The Kappa value is 0.56 which is lower than 0.7 and shows the reliability rate is not high but is much better than the reliability rate of depression cutoff score (0.460).

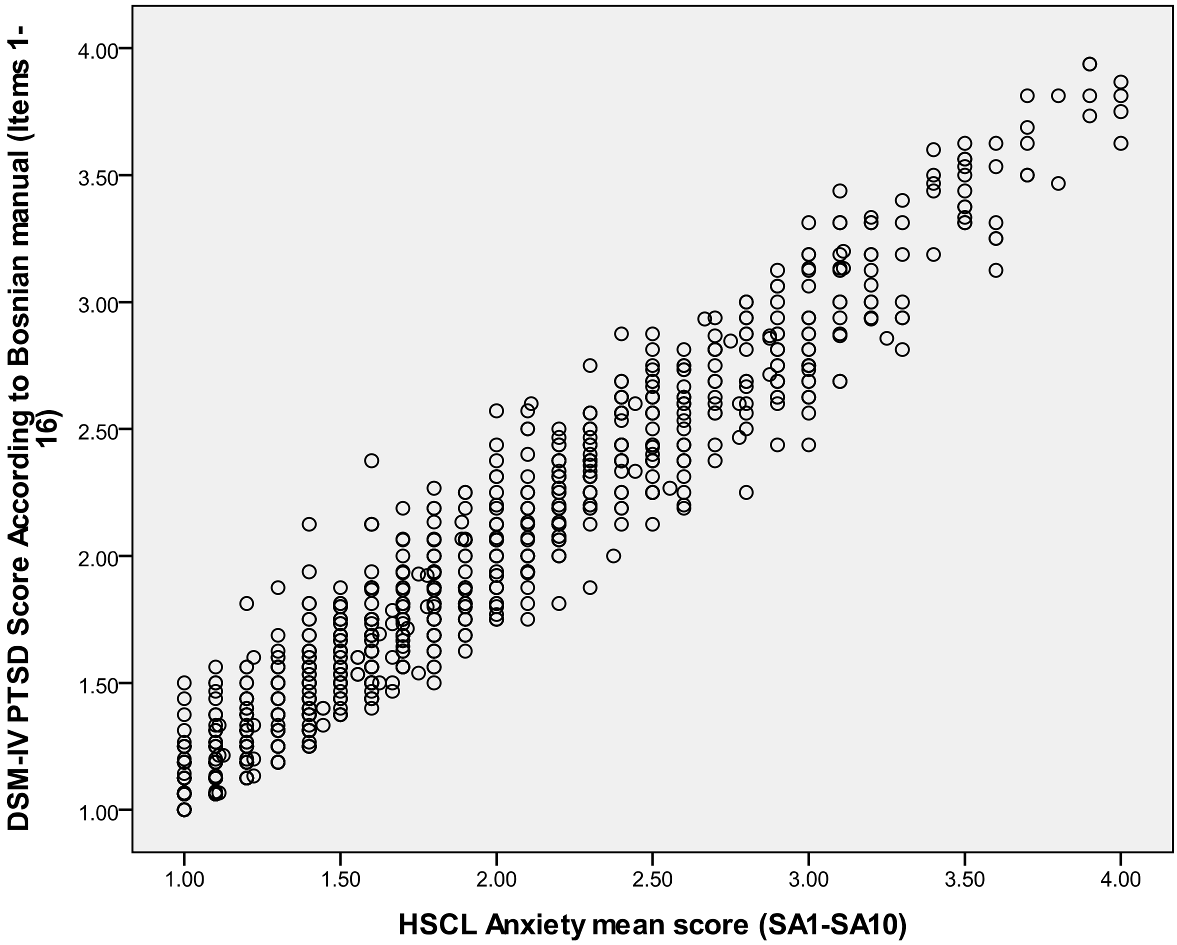
Table 55 Spearman's Correlations

Correlations			
			HSCL_Anxiety mean score (SA1-SA10)
Spearman's rho	PTSD_sum CDC	Correlation Coefficient	0.831
		Sig. (2-tailed)	0
		N	1159
	DSM_IV_PTSD Score According to Bosnian manual (Items 1-16)	Correlation Coefficient	0.965
		Sig. (2-tailed)	0
		N	1159

The Correlation Table 55 calculated the Pearson Correlation value 0.831 for PTSD (CDC Criteria) which indicates a perfect relationship between anxiety und PTSD symptoms and the Pearson correlation value 0.965 for PTSD (for DSM-IV Criteria). From the obtained values, it is clear that the relationship rate between anxiety and PTSD (DSM-IV) is higher than anxiety and PTSD (CDC Criteria). But in comparison to depression the Pearson Correlation value reduces for variable PTSD (DSM_IV Criteria).

To illustrate the mentioned relationship, Scatter plot is selected. It is a representative graph to depict the interconnection between anxiety symptom and DSM_IV_PTSD. It shows a relative intensive relationship.

Figure 2 Scatter plot of PTSD and Anxiety



4.6 Results of Applying Crosstab Correlation for Comparing MOS Questionnaire and PTSD Symptoms

As previously mentioned in chapter 3, MOS-20 consists of 20 items on 6 different scales that assess physical functioning, bodily pain, role functioning, social functioning, mental health and self–perceived general health status. To better understand the relationship between these MOS-20 items and anxiety with PTSD symptoms Pearson Correlation is constructed. The Correlation Table 56 demonstrates the summary of correlation coefficients over mentioned MOS-20 variables. As in Table 56 the Pearson value for anxiety and PTSD (Sum-CDC criteria it contains three Arousal, Avoidance and Recurrence Symptoms) is 0.86 and is statistically significant (p-value = 0 < 0.05), therefore it shows a high positive correlation between anxiety and PTSD which means that those with higher anxiety tend to have higher PTSD symptoms. Meanwhile those with lower anxiety had lower PTSD symptoms and those in between had PTSD that were neither especially high nor especially low. The Pearson Correlation value for anxiety and PTSD (DSM-IV criteria) is 0.97, showing an even stronger relationship.

The Pearson Correlation value for Physical Functioning is -0.433 for PTSD (Sum-CDC criteria) and -0.435 for PTSD (DSM-IV criteria) and both of them are statistically significant. Because the Pearson Correlation values lie between -0.25 and -0.75, we can conclude there to be a negative moderate degree of correlation. In other words, there is also a tendency for the people who have high Physical functioning to have low PTSD.

The Pearson Correlation values for Body Pain are -0.59 for PTSD (Sum-CDC criteria) and -0.64 for PTSD (DSM-IV) and both of them are statistically significant. These Pearson Correlation values show a negative moderate degree of correlation too. Thus there is also a tendency for the participants who have high body Pain to have low PTSD symptoms.

The Pearson Correlation values for Mental Health Status are -0.41 for PTSD (Sum-CDC criteria) and -0.45 for PTSD (DSM-IV) and again both of them are statistically significant. These Correlation values show a negative moderate degree of correlation too. Therefore there is a tendency for the participants with low Mental Health Status to have high PTSD symptoms and for high Mental Health Status to have low PTSD symptoms.

The Pearson Correlation values for Social Functioning are -0.186 for PTSD (Sum-CDC criteria) and -0.183 for PTSD (DSM-IV) both of them are statistically significant. The Pearson Correlation values lie between -0.3 to +0.3 therefore it is concluded that there is little or no association. It means low levels of social functioning are associated with high or low levels of PTSD symptoms or high levels of social functioning are associated with high or low levels of PTSD symptoms.

The Pearson Correlation values for General Health Status are -0.56for PTSD (Sum-CDC criteria) and -0.59 for PTSD (DSM-IV) and both of them are statistically significant. These degrees show negative moderate degrees of correlation. Therefore there is a tendency for the participants with low General Health Status to have high PTSD symptoms and for those with high General Health Status to have low PTSD symptoms.

The Correlation values for Role Functioning are 0.19 PTSD (Sum-CDC criteria) and 0.18 for PTSD (DSM-IV criteria) and both of them are statistically significant. It shows Role Functioning and PTSD symptoms have little or no association.

Table 56 Pearson Correlation

		HSCL Anxiety mean score (SA1-SA10)	**Physical function status**	**Bodily pain**	**Mental health status**
HTQ cut Recurrence and Avoidance and Arousal	**Pearson Correlation**	.861	-.433	-.598	-.411
	Sig. (2-tailed)	.000	.000	.000	.000
	N	1159	1152	1150	1144
DSM-IV PTSD Score	**Pearson Correlation**	.971	-.435	-.641	-.458
	Sig. (2-tailed)	.000	.000	.000	.000
	N	1159	1152	1150	1144

Table 57 Pearson Correlation

		Social Functioning	General health perception	Role functioning	MOS_20
compute HTQ cut Recurrence and Avoidance and Arousal	Pearson Correlation	-.186	-.560	.199	-.595
	Sig. (2-tailed)	.000	.000	.000	.000
	N	607	1161	1158	1161
DSM-IV PTSD Score According to Bosnian manual (Items 1-16)	Pearson Correlation	-.184	-.596	.181	-.640
	Sig. (2-tailed)	.000	.000	.000	.000
	N	607	1161	1158	1161

CONCLUSION

Logistic regression provides a useful means for modeling PTSD symptoms on Harvard Trauma Questionnaire, General Health Questionnaire and Displacement Refugee status. The Population at increased risk for PTSD symptoms, as measured by HTQ scores, was those who experienced illness despite being unable to obtain health care, those who experienced torture as physical or mental suffering, those confined to home because of danger outside and those with lack of food or water. The aforementioned traumatic events play a significant role in predicting PTSD symptoms. The fitted model for PTSD symptoms measured by HTQ scores is:

$$\text{Logit(p)} = \log\left({}^{p}/_{1-p}\right)$$
$$= -0.271 + 0.62\,HTQ27 + 0.53\,HTQ13 + 1.16\,HTQ3 + 0.695\,HTQ2$$

Where p is the probability of having PTSD symptoms.HTQ27 is “Confined to home because danger outside” event, $HTQ13$ is “Torture; physical or mental suffering”, HTQ3 “Ill health without access to medical care” and HTQ2”Lack of food or water”. We can confidently say that all these predictors are reliable for predicting PTSD symptoms.

The population at increased risk for Mental Health problems as measured by HTQ scores were those whose money or property were obtained by force or threats (Extortion) and those who experienced illness but were unable to obtain medical health care .

The fitted model for identifying Mental Health Disorder is as follows:

$$\text{Logit(p)} = \log\left({}^{p}/_{1-p}\right) = -1.563 + 1.246\,HTQ3 + .443\,HTQ17$$

where p is the probability of having Mental Health Disorder, HTQExp17 is;” Extortion or robbery” and HTQ3 is “Ill health without access to medical care”. In other words those who experienced illness but were unable to obtain medical health care are more than twice as likely to suffer from Mental Health problems as those who did not. Similarly, the risk of Mental Health problems occurring in participants whose money or property was obtained by using force or threats tended to be 0.14% higher than those who did not.

The statistical analysis reveals PTSD symptoms certainly have a significant effect on depression and anxiety.

Bibliography

Altman, D. G. (1990). *Practical statistics for medical research, London1990.* London: Chapman & Hall/CRC.

Answers.com. (n.d.). Retrieved from http://wiki.answers.com/Q/What_is_the_difference_between_multiple_regression_and_logistic_regression

Austin, P., & V.Tua, J. (2004, November). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *Journal of Clinical Epidemiology* , p. http://www.ncbi.nlm.nih.gov/pubmed/15567629.

Baumgarth, C. (2008-2009). *Multiple Regressionanalyse.* Marmara University.

Buchner, A. (n.d.). *Institute für Ecperimentelle Pscychologie*. Retrieved from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/user-guide-by-distribution/z/logistic_regression

Cardozo, B. L., Vergara, A., Agani, F., & A. Gotway, C. (2000). Social Functioning, and Attitudes of Kosovar Albanians Following the War in Kosovo. *Journal of American Medical Association* , pp. http://jama.ama-assn.org/content/284/5/569.abstract.

Cox, D. R. (2007). Regression Models and Life-Tables. *Journal of the Royal Statistical Society* , p. http://www.stat.rutgers.edu/home/rebecka/Stat687/cox.pdf.

D.Duscalu, C., Carausu, E. M., & Manuc, D. (2008). Methods for Data Selection in Medical Databases: The Binary Logistic Regression -Relations with the Calculated Risks, World Academy of Science, Engineering and Technology. *World Academy of Science, Engineering and Technology* , pp. http://www.waset.org/journals/waset/v38/v38-49.pdf.

F.Gillespie, D., & Glisson, C. (1999). *Quantitative methods in social work: state of the art.* Charles Glisson Editors.

Field, A. (2005). *Discovering Statistics Using SPSS.* Second Edition.

Fox, V. J. (1997). *Applied regression analysis, linear models, and related methods.* Sage Publications.

Hosmer, D. W., & Lemeshow, S. (2000). *David W. Hosmer, Stanley Lemeshow, Applied Logistic Regression, A Wiley- Interscience Publication, Second Edition.* A Wiley- Interscience Publication, Second Edition.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied Linear Regression Models.* Fourth Edition.

L.Pace, N. (2008, December). Independent Predictors from Stepwise Logistic Regression May Be Nothing More than Publishable P Values. *Anesthesia & Analgesia* , pp. http://www.anesthesia-analgesia.org/content/107/6/1775.full.

Lewis, T. (2007). PROC LOGISTIC: The Logistics Behind Interpreting Categorical Variable Effects. *U.S. Office of Personnel Management* , p. http://www.nesug.org/proceedings/nesug07/sa/sa11.pdf.

MESOSworld(Methodical Education for the Social Sciences). (n.d.). Retrieved from http://www.mesosworld.ch/lerninhalte/Biv_Regression/de/html/unit_VorausRegress.html

Murtaugh, P. A. (1998). Paul A. Murtaugh, Methods of variable selection in regression modeling. *Communications in Statistic- Simulation Computation* , p. http://www.informaworld.com/smpp/content~db=all~content=a780037033.

Osborne, J. W. (2006, October). Practical assessment Research and Evaluation. *North Carolina State University* .

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational* , p. http://bit.csc.lsu.edu/~jianhua/emrah.pdf.

Poulsen, J., & French, A. (2003). DISCRIMINANT FUNCTION ANALYSIS (DA). Available at http://userwww.sfsu.edu/~efc/classes/biol710/discrim/discrim.pdf. p. http://userwww.sfsu.edu/~efc/classes/biol710/discrim/discrim.pdf.

Rogue Wave web site. (2006). Retrieved from http://www2.roguewave.com/support/docs/sourcepro/edition9/html/analyticsug/3-3.html

Spiegel, P. B., & Salama, P. (1998-1999). War and mortality in Kosovo: an epidemiological testimony. p. http://conflict.lshtm.ac.uk/media/Kosovo_Mortality_in_Kosovo_Spiegel_Salama_Lancet_2001.pdf.

Spiegel, P. B., & Salma, P. (1999, September). Kosovar Albanian Health Survey Report. *International Emergencies and Refugee Health Branch* , pp. http://www.docstoc.com/docs/44348169/Kosovar-Albanian-Health-Survey-Report-September-1999.

StatSoft(Electronic Statistics TextBook). (n.d.). Retrieved from http://www.statsoft.com/textbook/multiple-regression/#assumptions

Walter, S., & Tiemeier, H. (2007, December). Variable selection: current practice in epidemiological studies. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2791468/. *European Journal of Epidemiology* , p. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2791468/.

Wenzel, T. (2006, August). , Long-term Sequels of War, Social Functioning and Mental Health in Kosovo. p. http://www.proasyl.de/fileadmin/proasyl/fm_redakteure/Newsletter_Anhaenge/122/Longtime.pdf.

Wikipedia. (n.d.).