Available at www.ijcasonline.com ISSN 2349 – 0594

International Journal of Modern Chemistry and Applied Science

*International Journal of Modern Chemistry and Applied Science 2015, 2(3), 153-163*

# Ordinal logistic regression analysis to assess the factors that affect health status of students in Ambo University: a case of natural and computational sciences college, Ambo University

**O.ChandraSekhara Reddy*, Endale Alemayehu**

*Statistics Department, Ambo University, Ambo, Ethiopia.*

**Abstract:** This study was intended to investigate the factors that affect the health status of the students in Ambo University in case of College of Natural and computational Science. The self-health assessment was applied as an indicator of health status in this study. The sample for this study was drawn from the students of College of Natural and computational Science Ambo University. The information was collected by well-designed questionnaire from the students. The dependent variables were health status of students. The independent variables included sex, income, lack of hygiene of café, usage of medicine, smoking cigarette, alcohol consumption, physical exercise, Environmental factor and bad smelling around the cafe. The statistical method of analysis applied for testing the association and the factors affecting health status in the students was Ordinal Logistic regression and chi-square test statistics. The results revealed that sex, income, lack of hygiene of café, smoking cigarette, bad smelling around the café, Alcohol consumption and Environmental factor are factors that affect health status of the students. In general most of the variables included in the regression model had significant effects like sex, lack of hygiene of café, usage of medicine, smoking cigarette, bad smelling around the café and Environmental factor are found to be more significantly associated with health status of students.

**Key Words:** Factors that Affect Health Status of Student, Stratified Random Sampling, Ordinal Logistic Regression, SPSS and Ambo University.

## 1. Introduction

### 1.1 Background of the Study

Health is the level of functional or metabolic efficiency of a living being. In humans, it is the general condition of a person's mind, body, usually meaning to be free from illness, injury or pain (as in "good health" or "healthy"). Health is the presence of a positive capacity to lead energetic satisfying and productive life. Society as a whole benefits from peoples good health just as individual do. For that reason many government and voluntary agencies stood preserve and improve the health of all the people [1].

World Health Organization is an agency of the United Nations work to promote better health through the world Health services include all services dealing with the diagnosis and treatment of disease, or the promotion, maintenance and restoration of health. Mental illness is described as “the spectrum of cognitive, emotional, and behavioral conditions, which interfere with social and emotional well-being and the lives and productivity of people”. Having a mental illness can seriously impair, temporarily or permanently, the mental functioning of a person. Other terms include: 'mental health problem', 'illnesses, ‘disorder’, and ‘dysfunction’[2]

Ethiopia experiences a heavy burden of disease mainly attributed to communicable infectious diseases and nutritional deficiencies. However, there has been encouraging improvements in the coverage and utilization of the health services over the periods of implementation of the Health Sector Development Plan (HSDP).HSDP constitutes the health chapter of the national poverty reduction strategy and aims to increase immunization coverage and decrease under-five mortality at large [3].

The health service currently reaches about 72% of the population and The Federal Ministry of Health aims to reach 85% of the population by 2009 through the Health Extension Program(HEP). The HEP is designed to deliver health promotion, immunization and other disease prevention measures along with a limited number of high-impact curative interventions. (Health sector development plan) health facilities in Ethiopia are too few because of large population size and incorrect distribution of resources, lack of money, basic drugs and equipment. The major health problems of the country remain largely preventable communicable diseases and nutritional disorders. The health care system represents a social determinant of health as well as it influences other determining factors [4].

Health standard of Ethiopia is too few when compared to other because of several reasons like in sufficient economy and improper use of materials in hand. Ethiopia faces the difficult problems that

improved health status in both vital for growth and development and expensive as well, this and many other problems placed very sever burden on development in general and health resources in particular in more ways than one.[5]

The absence of regulatory systems to monitor the quality, safety, and efficacy of Medicines can compromise the overall effectiveness of health care services and endanger the public health. A strong regulatory system is considered an essential component of a health system. In Ethiopia, the Food, Medicine and Health Care Administration and Authority (FMHACA), formerly known as the Drug Administration and Control Authority, regulates the country's pharmaceutical sector in an environment vulnerable to drug smuggling and the circulation of substandard medicines.[6]

### 1.2 Statement of the problem

Health service forms part of the basic need of the society. Health was factor in the development of endeavor of country hence it determine the future of nation by affecting the wellbeing of its population for these realization of its human potential. Ethiopia one of the developing countries have a shortage of major public health service and facilities such as clean water supply, education and others are adequate.

**To get better utilization service the below question are raised:-**

- Does ordinal logistic regression model suitable for health status data?
- What are the determinant factors of health status at University level in particular Ambo University?
- Are the students satisfied in the service provided by the university clinic?

### 1.3 Objective of the Study

#### 1.3.1 General Objective

- The main objective of this study was to investigate the factors affecting health status of 3rd year natural and computational science students in Ambo University.

#### 1.3.2 Specific Objectives

Under the heading of the main objective this study tried to address the following specific objectives. These are:-

- To apply ordinal logistic regression model for health status data
- To determine the most prevalent factor(s) of health status in Ambo University.
- To know which sex group most affected by health status.
- To determine the percentage of people who are satisfied with services provided by the clinic of Ambo University?

### 1.4 Scope of the Study

The study focus on health status and investigation of the major factors related to health status of students in Ambo University 3rd year Natural and computational Science. The study deals about the major factors that affect health status students.

### 1.5 Significance of the Study

This study was expected to show clear the possible of obtaining better utilization of the clinic by identifying the key factors that affected the health status of students. Besides, the finding of this study was expected to solve the factors that affected the health of student in this clinic and it enables the students to get best health service.

### 1.6 Limitation of the Research

This study purposes to target all students of Ambo University who study in College of Natural and computational Science. There are a number of problems and difficulties that should be suggested during the accomplishment of this paper. Some of the limitations are:

- Ambo University was not flexible to give relevant data as much as required.
- The students were not volunteer to give full information about health status.
- Lack of detailed information in Ambo University clinic about health status

## 2. Literature Review

### 2.1 Modern Public Health

Literature search applications for healthcare professionals facilitate searching biomedical literature databases to find and display medical reference information. With onset of the epidemiological transition and as the prevalence of infection disease decreased through the 20thcentery. Public health began to put more focus on chronic disease such as cancer and heart disease. For instance in the United states public health work established many programs to help the poor in New York keep their infants healthy, leading teams of nurses in to the crowded neighborhoods of hell Kitchen and teaching mothers how to dress, feed and bath their babies. The WHO reports lack of exclusive breast feeding during the first six months of life contributed as to over a million avoidable child deaths each year. There is recognition that our health is affected by many factors including where we live, Genetics, our income, our education states and our social relationships. There are known as "social determinants of health."The new public health seeks to address those health inequalities by adducting for population, based policies that improve health in an equitable manner (7).

### 2.2 Applications in Health Status

The main stakeholders in the healthcare process were health status consumers (patients).

Consumer oriented care, where patients directly involved in the process of care, will greatly improve the healthcare process. Technology can play key role in consumer-oriented healthcare (for example, making information accessible to consumers, integrating consumers' preferences into HISs, remote monitoring, communication, etc.), which studied in branch of medical informatics called Consumer Health Informatics (CHI). The management of diseases with chronic conditions is very costly[8]

**2.3 Health Sectors in Ethiopia**

Matrices of health in Ethiopia were among the world's worst. According to the US government Ethiopians health care system is wholly inadequate, even often improvements. Ethiopian experiences a heavily border of disease mainly attributed to communicable infectious disease nutritional deficiencies, shortage and high turnover of human resources and in adequacy of essential drugs and supplies have also contributed to the Burden. However there has been encouraging improvements in the coverage and utilization of the health services over the period of implementation of health sector development plan.

**2.4 Factors that Affect Health Status**

According to Probst, Samuels, Jespersen, et al (2002)34% of African and Native Americans along with 25% of Hispanics communities are poor, or live below the low income level. This poverty causes minorities to delay necessary medical care and affects the amount of health status resources available within community. Probst, Samuels, Jespersen, et al (2002) state that 70% of African Americans, 84% of Hispanics and 67% of Native Americans live in communities that are Health Professions Shortage of water. In addition, Ellickson *et al.* found that early smokers were affected at least three times more than non-smokers to regularly use marijuana and hard drugs. Students who binge drank were affected than both non-drinkers and drinkers who did not binge to report involvement in other risky health behaviors according to Cutler.

Most health determinants like lifestyle background, economic and social conditions are cause for the illness of Ambo people. Disease like typhoid and food hygiene are usual disease that affect Ambo peoples as well as Ambo University students, because of factors like whether change, food hygiene ,sanitation problem dusts are usually cause common called most students.

**3. Data and Methodology**

**3.1 Description of Study Area**

The study was conducted at Ambo University main campus which was one of the higher education institutions in Ethiopia. It was found in which are located western shoa of Oromia. It was created as agricultural college in 1939. Ambo University was situated about 114 km for from west of Addis Ababa. It becomes one of the higher institutions in 1999 E.C and training with the help of ministry of education. Now a day, the University has three additional branches, namely Awaro, wolliso and guder campus.

**3.2 Data**

The Method of Data Collection in this study was primary source using questionnaire as instruments of data collection. The questionnaire was administered and distributed to the students under the direct supervision of the researcher.

**3.3 Method of Data Collection**

**3.3.1 Sampling design**

Sampling design was a system of taking small ratio of observation from a large population to get information of that large population from the sample observation by using some statistical design. In this study, stratified random sampling were used to determine the sample size of the target population the study population were heterogeneous.

**3.3.2 Sample Size Determination**

To determine the sample size, degree of freedom and margin of errors would be considered.

In this study, stratified random sampling was applied to determine the sample size of the students from the whole population. The total number of samples 'n' was calculated as follows:-

Where, $N_h$=total number of students from each department.

$N_1$ = 50 Number of students in biology, $N_2$= 43 Number of students in chemistry,

$N_3$ = 32Number of students in physics, $N_4$ = 44 Number of students in sport science,

$N_5$ = 41Number of students in mathematics and $N_6$ = 57 Number of students in statistics.

$N=N_1+N_2+N_3+N_4+N_5+N_6$= 50+43+32+44+41+57 = 267

Now, the sample size was determined by using the formula for sample size determination for the population. This is.

$$n_o = \frac{(Z_{\alpha_{f_2}})^2 \ \mathbf{p(1-p)}}{\mathbf{d^2}}$$

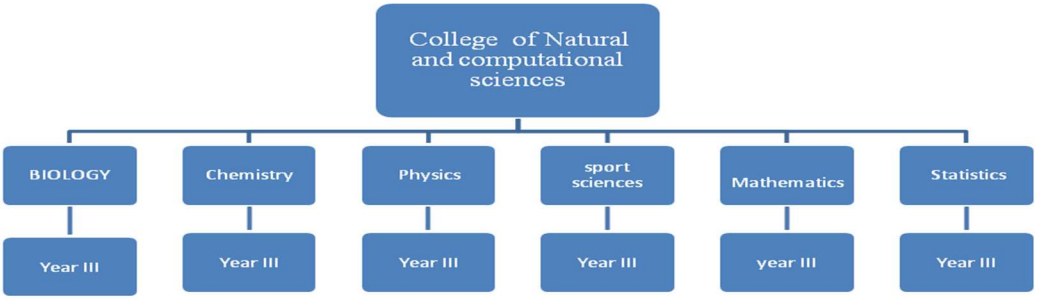
$Z_{\alpha_{f_2}}$ - Critical value at 95% confidence level of certainty (1.96)

P = probability of success, q = probability of failure.

Based on the previous study p=0.75

$$n_o = \frac{(1.96)^2 \ 0.75\left(1-0.75\right)}{0.05^2} = 289$$

Note, if a Finite population correlation (FPC) is >5%, there is a satisfactory sample size approximation to




the required, so we used the correlation formula of sample size. FPC = $\frac{n_o}{N}$ = 289/267=1.082>0.05, so the sample fraction was greater than 0.05 so, we can do a sample size $n = \frac{n_o}{1 + \frac{n_o}{N}}$

$$n = \frac{289}{1 + \frac{289}{267}} = 139$$

Where, n=total number of samples, α=the significance level=0.05.
d = margin of error =0.05 and= $Z_{0.025}$=1.96.
For stratum h (each department), number of sample was calculated by using proportional allocation; based on this the following results were obtained.

$n_h = \frac{Nh}{N} n, h = 1, 2, 3, 4, 5, 6$ , where $n_h$ is sample size of the $h^{th}$ stratum (department)

$$n_1 = \frac{N1}{N} n = \frac{50}{267} 139 = 26$$

$$n_2 = \frac{N2}{N} n = \frac{43}{267} 139 = 22$$

$$n_3 = \frac{N3}{N} n = \frac{32}{267} 139 = 17$$

$$n_4 = \frac{N4}{N} n = \frac{44}{267} 139 = 23$$

$$n_5 = \frac{N5}{N} n = \frac{41}{267} 139 = 21$$

$$n_6 = \frac{N6}{N} n = \frac{57}{267} 139 = 30$$

Therefore, sample size (n) = $n_1 + n_2 + n_3 + n_4 + n_5 + n_6$= 26+22+17+23+21+30 = 139 is calculated.

### 3.4 Variable of the study

The dependent variable was Health status of the students, which was categorized as Unhealthy (illness often), Moderate and Healthy (strong)).

Table 1 : Nature of the dependent variable with the corresponding code

| Categories | How often did you visit Ambo university clinic due to sickness for the last 4 months? | Code |
|---|---|---|
| Unhealthy | >=5 times | 0 |
| Moderate | 1-4 times | 1 |
| Healthy | None | 2 |

The independent variables that used in this study are sex, income, and hygiene of café, usage of medicine, smoking cigarette, alcohol consumption, physical exercise, environmental factor &bad smelling around the cafe.

Table 2:List of independent variables with respective name and category codes

| Variable designation | Description | Value labels |
|---|---|---|
| X1 | Sex | 0=”Female”, 1=”Male” |
| X2 | Average income of student | 0=”<200”, 1=”200-600”, 2=”600-1000”, 3=”>1000” |
| X3 | Lack of hygiene of café | 0=”Good”, 1=”fair”,2=”bad” |
| X4 | Usage of medicine | 0=”No”, 1=”Yes” |
| X5 | Smoking cigarette | 0=”No”, 1=”Yes” |
| X6 | Alcohol consumption | 0=”No”, 1=”Yes” |
| X7 | Physical exercise | 0=”never”,1=”1-3days”,2=” >3 days” |
| X8 | Environmental factor | 0=”have not influence” ,1=”Not such much”,2=”Strict factor” |
| X9 | Bad smelling around the café | 0=”No”, 1=”Yes” |

### 3.5 Statistical Model

#### 3.5.1 Descriptive Statistics

Statistical analysis that would use in this study was descriptive statistics such as a table, bar-charts to describe the frequency distribution and percentage.

#### 3.5.2 Inferential Statistics

Inferential statistics is the procedure by which we reach a conclusion about a population based on the information contain in the sample drawn from that population. Therefore among inferential statistics, we would use chi-square test and ordinal logistic regression.

### 3.5.3 Chi-square Test Statistics

It was used to analysis the association between dependent and independent variables. Hypothesis testing: Ho: There was significant association between dependent variable and independent variables Vs H1: Not Ho

To test the null hypothesis the we can compare $X^2_{cal}$ with $X^2_{tab}$ is given by

$$X^2_{cal} = \sum\sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$ (i=1, 2.....n, j=1, 2,3,…m) Where, $E_{ij}$ is expected frequency corresponding to $(ij)^{th}$ $and$ $O_{ij}$ be the observed frequency. From this calculated and tabulated values with degree freedom (n-1) (m-1) we decide about the rejection of Ho that say there was no significant association between two variables.

We reject Ho if $X^2_{cal}$ greater than $X^2_{tab}$ or if p-value less than level of significance, otherwise we do not reject Ho.

| | |
|---|---|
| **Demographic and socioeconomic characteristics**<br>• Sex<br>• Average income<br>• Lack of hygiene of café<br>• Environmental factor<br>• Bad smelling around the café<br>**Lifestyle of the students**<br>• Usage of medicine<br>• Smoking cigarette<br>• Alcohol consumption<br>• Physical exercise | **Health status of the Students** (healthy, moderate and unhealthy) |

**Assumptions of chi-squares**

- The sample must be randomly selected from the population.
- The population must be normally distributed for the variable under study.
- The observation must be independent of each other.
- Each cell (each data entry) should contain at least five observations.

### 3.5.4 Ordinal Logistic Regression

Logistic regression may be useful when we are trying to model a categorical dependent variable as a function of one or more independent variables being the dependent variable has two outcomes. Ordinal logistic regression (OLR) is a type of logistic regression analysis when the response variable has more than two categorizes with having natural order or rank.

In statistics, the ordered logit model (also ordered logistic regression or proportional odds model), is a regression model for ordinal dependent variable. It is natural to consider methods for more categorical responses having more than two possible values. The most well-known of these ordinal logistic regression methods is called the proportional odds model. The

**Steps of chi-squares:-**

1. State hypothesis (i.e. null and alternative hypothesis)
2. Select desired level of significance (alpha level of significance)
3. Check the statistics
4. Compute the calculated value( $X^2_{cal}$ )
5. Obtain the critical (tabulated) value ($X^2_{tab}$ )
6. Interpret the test

basic idea underlying the proportional odds model is re-expressing the categorical variable in terms of a number of tertiary variables based on internal cut-points in the ordinal scale.

We can consider the 3 tertiary logistic models corresponding to regressing each of the

$$logit[p(y \geq 3)] = \alpha_j + \beta_1 X_1 + \beta_2 X_2 \ +.\ldots\ldots+\beta_k\ X$$

Separately against the $X's.$ many variables of interest are ordinal. That is, we can rank the values, but the real distance between categories is unknown. Diseases are grade on scales from unhealthy to healthy**.**

**Assumptions of ordinal regression**

- The dependent variable should be measured at the ordinal level.
- Ordinal independent variables must be either continuous or categorical.
- There is no multicollinearity.
- Each independent variable has an identical effect at each cumulative split of the ordinal dependent variable.
- The relationship between each pair of outcome groups is the same.
- The effects of any explanatory variables are consistent or proportional across the different thresholds.

.

### 3.5.5 The Model of Ordinal Logistic regression

$$ln\left(\theta_j\right) = \alpha_j\ - \beta_1 X_1 + \beta_2 X_2 \ +.\ldots\ldots+\beta_9\ X_9$$

Where j goes from 1 to the number of categories minus 1.

This was the part we really want to find out. We have been thinking of attitudinal responses to abortion as a set of three unordered responses, but there is a very clear and intentional ordering to these responses like as unhealthy, moderate, and healthy. If we know that a category is ordinal then there are special models that tell us how independent variables relate to someone being higher or lower on the scale.

We could use the ordered logit model so that we could use the categorized directly as our dependent variable. Different links lead to proportional odds models or ordered probit models. The model cannot be consistently estimated using ordinary least squares; it is usually estimated using maximum likelihood.

The /proportional odds model and the /partial proportional odds model are special cases of the cumulative logit model. This is a cumulative logit model that assumes that the odds of response below a given response level are constant regardless of which level we picked. This model allows separated intercepts for the cumulative logit, but restricted the parameter sets for the predictors to be the same across all logits. A proportional odds model that constrains some predictors to have common parameters and leaves other predictors free to have separate parameters is called a partial proportional odds model.

The basic form of the generalized linear model is:

$$\mathbf{Link}(\gamma_j) = \frac{\theta_j - [\beta_1 X_1 + \beta_2 X_2 + \cdots . + \beta_k X_k]}{\exp(\tau_1 Z_1 + \tau_2 Z_2 + \cdots + \tau_m Z_m)}$$

Where, $\gamma_j$ is the cumulative probability for the $j^{th}$ category, $\theta_j$ is the threshold for the $j^{th}$ category, $\beta_1, \ldots, \beta_k$ are the regression coefficients, $X_1, \ldots, X_k$ are the predictor variables, and k is the number of predictors. The numerator on the right side determined the location of the model. The denominator of the equation specifies the scale. The $\tau_1,\ldots,\tau_m$ are coefficients for the scale component and $Z_1,\ldots,Z_m$ are m predictor variables for the scale component (chosen from the same set of variables as the X's).

**3.5.7 Fitting an Ordinal Logit Model**

The quantity to the left of the equal sign is called a logit. If the spacing between levels of the ordinal response scale is known, so that numerical scores can reasonably be assigned to the response levels, then a /mean response model can be fit. When we have more than two events, we can extend the ordinal logistic regression model. Suppose the underlying relationship to be characterized is: $y_i = x'\beta+\varepsilon$, ( Agresti, 2002; Greene, 2003), where $y_i$ is the exact but un observed dependent variable (perhaps the exact level of improvement by the patient); x is the vector of independent variables, and β is the vector of regression coefficients which we wish to estimate.

The Model fitting Information gives the -2 log-likelihood (-2LL) values for the baseline and the final model and a chi-square to test the difference between the -2LL for the two models.

The goodness of fit statistics indicated that the model fits much better than the location only model. From the observed and expected frequencies, we could compute the usual Pearson and Deviance goodness-of-fit measures.

The Pearson goodness of fit statistic is $\chi^2 = \Sigma\Sigma\left(\frac{\mathbf{Oij} - \mathbf{Eij}}{\mathbf{Eij}}\right)^2$.

The deviance measure is $D = 2\Sigma\Sigma O_{ij}\ln\left(\frac{\mathbf{Oij}}{\mathbf{Eij}}\right)$.

Both of the goodness-of-fit statistics should be used only for models that have reasonably large expected values in each cell. If we have a continuous independent variable or many categorical predictors or some predictors with many values, we may have many cells with small expected values. There are several $R^2$like statistics that can be used to measure the strength of the association between the dependent variable and the predictor variables. They are not as useful as the $R^2$ statistic in regression, since their interpretation is not straight forward. Three commonly used statistics are:-

1) Cox and Snell $R^2 = 1 - \left(\frac{(L(\beta^{\dagger}((0)))}{(L(\beta))}\right)^{\frac{2}{n}}$

2) Nagelkerke's $R^2 = \frac{\textbf{Cox and Snell R2}}{1 - \frac{\mathbf{L}\left(\boldsymbol{\beta}^{(0)}\right)\mathbf{2}}{\mathbf{n}}}$ and

3) McFadden's $R^2 = 1 - \left(\frac{L(\beta)}{L\left(\beta^{(0)}\right)}\right)$. Where, L (β) is the log-likelihood function for the model with the estimated parameters and L ($\beta^{(0)}$) is the log-likelihood with just the thresholds, and n is the number of cases (sum of all weights).

**3.5.8 Parameter Estimates of ordinal logistic regression**

The model is usually estimates using maximum likelihood**.** The method of maximum likelihood corresponds too many well-known estimation methods in statistics. The proportional odds model assumes that the true $\beta$ values are the same in all three models, so that the only difference in models is the intercept terms, $\alpha_C$ c=1, 2, 3. This means that the estimates from the three ordinal models could be pooled to provide just one set of $\beta$ estimates. By exponentiation the pooled estimate relative to a given predictor, i.e. taking, $e^{\beta j}$ we obtained an estimate of the common odds ratio (OR) that described the relative odds for $y \geq 3$ for values of $X_j$ differing by one unit.

The maximum likelihood estimate (MLE) of θ is that value of θ that maximizedlik (θ): it is the value that made the observed data the "most probable". If the $X_i$ are iid, then the likelihood simplified to: lik (θ) $= \prod_{i=1}^{n} \mathbf{f(xi|\theta)}$.

Rather than maximizing this product which could be quite tedious, we often used the fact that the

logarithm is an increasing function so it would be equivalent to maximized the log likelihood:- $l(\theta) = \sum_{i=1}^{n} \log\left(\mathbf{f(xi|\theta)}\right)$.

**Checking Model Adequacy**

**Goodness of Fit and the Deviance**

A most general way to detected lack of fit searches for any way the model fails. A goodness of fit test compared the model fit with the data. This approach regards the data as representing the fit of the most complex model possible the saturated model, which has a separated parameter for each observation. In certain cases, this test statistic has a large-sample chi-squared null distribution.

The likelihood-ratio statistic for this test is the deviance of the model and Pearson of the model

The Pearson statistics for testing goodness of fit is $\chi^2 = \Sigma\Sigma(\frac{\mathbf{Oij - Eij}}{\mathbf{Eij}})^2$ .

**4 Results and Discussions**

The corresponding likelihood-ratio (deviance) statistics is = $2\Sigma\Sigma O_{ij}\ln(\frac{\mathbf{Oij}}{\mathbf{Eij}})$. Then, $\chi^2$ and $G^2$ have approximated chi-squared null distributions. $G^2$ follows $\chi^2$ distribution with k degree of freedom. Where, k is the number predicated variable.

If $G^2$ is larger that is $L_1 > L_0$, then reject $H_0$ and conclude that the model is good fit or variables are significantly affected the respondent or dependent variable.

Pseudo-$R^2$ statistics would be use in ordinal regression to estimate the variance explained by the independent variable.

This would be made through the following:-

Nagelkerke's $R^2 = \frac{\text{Cox and Snell R2}}{1 - \frac{\mathbf{L}\left(\boldsymbol{\beta}^{(0)}\right)^2}{\mathbf{n}}}$

**4.1 Descriptive Statistics**

Table 3: Frequency (Percentage distribution) of Health status of students due to Unhealthy, moderate & Healthy with their frequencies in each categories are described as table instead:

| Variables | category | Health Status of Students %(n) | | | |
|---|---|---|---|---|---|
| | | Unhealthy | Moderate | Healthy | Total%(n) |
| Sex | Male | 35 (14) | 62.5(35) | 79.1 (34) | 59.71(83) |
| | Female | 65(26) | 37.5 (21) | 20.9 (9) | 40.29 (56) |
| Income | <200 | 72.5(29) | 46.4 (26) | 41.9 (18) | 52.52(73) |
| | 200-600 | 17.5 (7) | 41.1 (23) | 34.9 (15) | 32.37 (45) |
| | 600-1000 | 10 (4) | 8.9 (5) | 20.9 (9) | 12.95 (18) |
| | >1000 | | 3.6 (2) | 2.3 (1) | 2.16 (3) |
| hygiene of cafe | good | 42.5 (17) | 25 (14) | 20.9 (9) | 28.78 (40) |
| | fair | 55 (22) | 53.6 (30) | 37.2 (16) | 48.92 (68) |
| | bad | 2.5 (1) | 21.4 (12) | 41.9 (18) | 22.30 (31) |
| medicine | No | 40 (16) | 41.1(23) | 46.5(20) | 42.45(59) |
| | Yes | 60(24) | 58.9 (33) | 53.5(23) | 57.55 (80) |
| smoking | No | 100 (40) | 98.2(55) | 81.4 (35) | 93.52(130) |
| | Yes | | 1.8 (1) | 18.6 (8) | 6.48 (9) |
| Alcohol | No | 52.5 (21) | 82.1(46) | 69.8(30) | 69.78(97) |
| | Yes | 47.5(19) | 17.9 (10) | 30.2 (13) | 30.22 (42) |
| Exercise | Never | 35 (14) | 25 (14) | 32.6 (14) | 30.22 (42) |
| | 1-3 days | 45 (18) | 53.6 (30) | 46.5 (20) | 48.92(68) |
| | >3 days | 20 (8) | 21.4 (12) | 20.9 (9) | 20.86(29) |
| Environmental | have no influence | | 33.9 (19) | 25.6 (11) | 21.58 (30) |
| | not such much | 85 (34) | 33.9 (19) | 27.9 (12) | 46.76 (65) |
| | strict factor | 15 (6) | 32.1 (18) | 46.5 (20) | 31.66(44) |
| smelling | No | 57.5 (23) | 39.3(22) | 20.9 (9) | 38.85(54) |
| | Yes | 42.5(17) | 60.7 (34) | 79.1 (34) | 61.15 (85) |

Table 3 shows the percentage of students those are not such much by environmental factor (85%) are more than that of students affects strictly with environmental factor (15%). Also the students that getting lower income (72.5%) per-month are more affected by health status compare to the student those have monthly income of 600-100(10%).

The number of female students (65%) shows, most of students that are belongs to unhealthy female students compare to the number of male students (35%).

Status of health problem of students because of bad smelling around café (|) are exceed the students in which their health status are not determined by those problem() .

We can illustrate this sex group difference by using bar-chart as below:

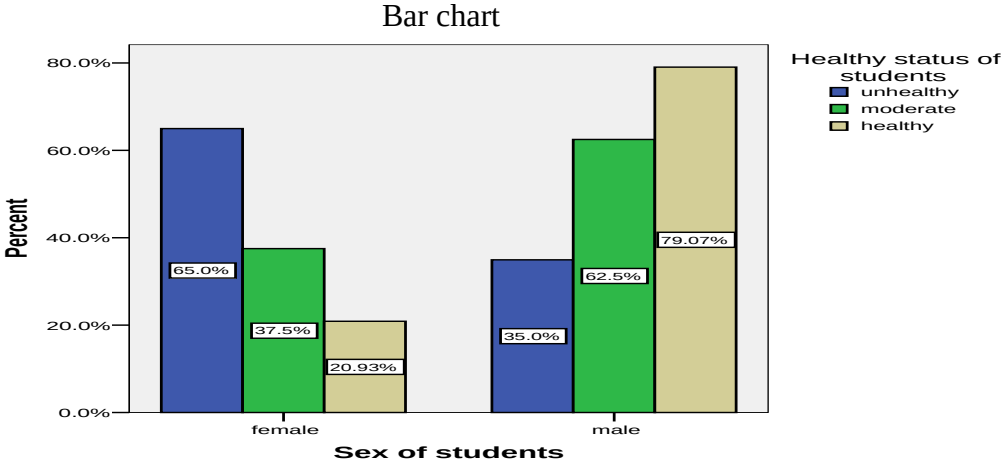



Figure 1 Health Status of male and female students

In the figure above demonstrates one can conclude that the number of female that are more exposed to healthy problems are larger than males of the same status. I.e. female students are affected by factors that affects health status than male students (i.e. there is a difference on sex when we compare depending on their percents female (65%) and male (35%) from the above figure).

**4.2 Bivariate Analysis**

Bivariate analysis is one of the simplest forms of the quantitative (statistical) analysis. It involves the analysis of two variables (often denoted as X, Y), for the purpose of determining the empirical relationship between them. Let as we see by taking Chi-Square Test association instead. Chi-square is simply an extension of a cross-tabulation that gives us more information about the relationship. However, it provides no information about the direction of the relationship (positive or negative) between the two variables. It deals the associations/relationships among two variables and test hypotheses about relationships between two or more ordinal level variables.

Table 4: Chi-Square Test Statistics

| **Variables** | Pearson Chi-Square | df | Sig. |
|---|---|---|---|
| Sex | 17.033 | 1 | 0.000 |
| Income | 12.962 | 3 | 0.044 |
| Lack of hygiene of café | 19.998 | 2 | 0.000 |
| Usage of medicine | 0.432 | 1 | 0.806 |
| Smoking cigarette | 15.250 | 1 | 0.000 |
| Alcohol consumption | 9.724 | 1 | 0.008 |
| Physical exercise | 1.333 | 2 | 0.856 |
| Environmental factor | 37.162 | 2 | 0.000 |
| Bad smelling around the café | 11.673 | 1 | 0.003 |

Table 4.2 shows that Health status of a student have a significant association with sex, bad smelling around the café, income, Lack of hygiene of café, smoking cigarette, Alcohol consumption and Environmental factor **.however,** Physical exercise and Usage of medicine have no significant association with Health status of a student.

The Pearson chi-square with p-value < 0.05 means there is significant association between health status and the aforementioned explanatory variables those have a p-value of less than 0.05. The chi-square is highly likely to be significant when our sample size is large, as it is certainly with our sample of 139 cases. In such condition we may want to set a lower p-value for rejecting the assumption of a good fit, may be $p<0.05$.

**4.3 Ordinal Logistic Regression Analysis**

Ordinal logistic regression was a type of logistic regression analysis that when the response variable is categorized more than two with having natural order or rank. That is, we can rank the values, but the real distance between categories is unknown. Under **Ordinal** Logistic Regression Analysis we can deal Model Fitting Information, Goodness-of-Fit, Pseudo R-Square, Parameter Estimates and Test of parallel lines.

Logit link function is used in the analysis because it is evenly distributed categories and is reasonable choices when the changes in the cumulative probabilities are gradual and logit involves all levels of the response and dichotomizes the response scale.

Table 4.3 Model Fitting Information

| Model | -2 Log Likelihood | Chi-Square | df | Sig. |
|---|---|---|---|---|
| Intercept Only | 295.439 | | | |
| Final | 201.745 | 93.695 | 14 | .000 |

Link function: Logit.

The Model Fitting Information table, which gives the -2 log likelihood for the intercept only and final models can be used in comparisons of nested models. The statistically significant chi-square statistic (***p<0.05***) indicates that the Final model gives a significant improvement over the baseline intercept-only model. This tells us that the model gives better predictions than if we just guessed based on the marginal probabilities for the outcome categories. Therefore, the Full model (with factors that affect health status as a predictor) is significantly better than the 'health status' model.

Table 4.4 Goodness-of-Fit Statistics

| | Chi-Square | df | Sig. |
|---|---|---|---|
| Pearson | 241.204 | 246 | .574 |
| Deviance | 194.813 | 246 | .993 |

Link function: Logit

From the above table the results for our analysis suggest the model does fit very well (***p>0.05) (i.e. fail to reject the*** null hypothesis depending on the observed data). Also the model fits adequately.

Table 4.5 Pseudo R-Square

For logistic and ordinal regression models, it not possible to compute the same $R^2$ statistic as in linear regression so three approximations are computed instead.

| | |
|---|---|
| Cox and Snell | .490 |
| Nagelkerke | .553 |
| McFadden | .310 |

Link function: Logit.

What constitutes a "good" $R^2$ value depends upon the nature of the outcome and the explanatory variables Here, the pseudo $R^2$ values (e.g. Nagelkerke = 55.3%) indicates that there is relatively small proportion of the variation in health status between students . This is just as we would expect because there are numerous factors that affect health status of student.

Table 4.6 Parameter Estimates of Ordinal Logistic Regression

| | | Estimate | Std. Error | Wald | df | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower Bound | Upper Bound |
| Threshold | [Healthy = 0 ] | -7.659 | 1.925 | 15.831 | 1 | .000 | -11.431 | -3.886 |
| | [Healthy = 1 ] | -4.727 | 1.841 | 6.592 | 1 | .010 | -8.336 | -1.119 |
| Location | [Sex=0 ] | -1.561 | .415 | 14.180 | 1 | .000 | -2.374 | -.749 |
| | [Sex=1 ] | 0(a) | . | . | 0 | . | . | . |
| | [Income=0] | -2.336 | 1.268 | 3.392 | 1 | .066 | -4.822 | .150 |
| | [Income=1] | -.957 | 1.252 | .584 | 1 | .445 | -3.411 | 1.497 |
| | [Income=2] | -.156 | 1.316 | .014 | 1 | .905 | -2.736 | 2.423 |
| | [Income=3] | 0(a) | . | . | 0 | . | . | . |
| | [cafe=0 ] | -1.731 | .615 | 7.925 | 1 | .005 | -2.936 | -.526 |
| | [cafe=1 ] | -1.580 | .535 | 8.702 | 1 | .003 | -2.629 | -.530 |
| | [cafe=2 ] | 0(a) | . | . | 0 | . | . | . |
| | [medicine=0 ] | .447 | .418 | 1.142 | 1 | .285 | -.373 | 1.266 |
| | [medicine=1 ] | 0(a) | . | . | 0 | . | . | . |
| | [smoking=0 ] | -2.900 | 1.208 | 5.759 | 1 | .016 | -5.268 | -.532 |
| | [smoking=1 ] | 0(a) | . | . | 0 | . | . | . |
| | [Alcohol=0 ] | .962 | .456 | 4.451 | 1 | .035 | .068 | 1.856 |
| | [Alcohol=1 ] | 0(a) | . | . | 0 | . | . | . |
| | [Exercise=0 ] | -.423 | .560 | .571 | 1 | .450 | -1.522 | .675 |
| | [Exercise=1 ] | -.057 | .520 | .012 | 1 | .913 | -1.075 | .961 |
| | [Exercise=2 ] | 0(a) | . | . | 0 | . | . | . |
| | [Factor=0 ] | 1.153 | .551 | 4.377 | 1 | .036 | .073 | 2.233 |
| | [Factor=1 ] | -1.002 | .458 | 4.781 | 1 | .029 | -1.901 | -.104 |
| | [Factor=2 ] | 0(a) | . | . | 0 | . | . | . |
| | [smelling=0 ] | -1.160 | .423 | 7.520 | 1 | .006 | -1.989 | -.331 |
| | [smelling=1 ] | 0(a) | . | . | 0 | . | . | . |

Link function: Logit. a This parameter is set to zero because it is redundant.

There is also a strong association between factors that affect health status, even when p-values are less than alpha level. We can see significant for Environmental factor, sex, bad smelling around the café**,** Lack of hygiene of café, smoking cigarette, Alcohol consumption and health status of students.

In the Parameter Estimates table we see the coefficients, their standard errors, the Wald test and associated p-values (Sig.), the 95% confidence interval of the coefficients and odds ratios. Since p-values less than alpha level they are statistically significant; otherwise not. The thresholds are shown at the top of the parameter estimates output, and they indicate where the latent variable is cut to make the three groups that we observe in our data. The threshold coefficients are representing the intercepts, specifically the point (in terms of a logit) where health status might be predicted into the three categories.

$$ln(\theta_1) = -7.659 - 1.561X_1 - 2.336X_2 - 1.731X_3 + 0.447X_4 - 2.9X_5 + 0.962X_6 - 0.423X_7 + 1.153X_8 - 1.160X_9$$

$$ln(\theta_2) = -4.727 - 1.561X_1 - 2.336X_2 - 1.731X_3 + 0.447X_4 - 2.9X_5 + 0.962X_6 - 0.423X_7 + 1.153X_8 - 1.160X_9$$

In this coefficients the negative sign indicates that those variables have negative effects on health status from table 4.6 sex, income, lack of hygiene of café , physical exercise, smoking cigarette, bad smelling around café are the factors that have negative effects on health status. The estimates labeled location are the coefficients for the predictor variables. There appears to be relationship between health status and the factors that affect health status. From the observed significance levels in above table, we can see that all explanatory variables are factors that affect health status of students. Based on the small observed significance level, we can reject the null hypothesis that it is zero. The estimates labeled Location are the ones we are interested in. They are the coefficients for the predictor variables. The Wald statistic is the square of the ratio of the coefficient to its standard error. The significance of the Wald statistic in the column with heading sig (< 0.05) indicates the importance of the predictor variables in the model (we reject the Null hypothesis Ho: $\beta_i$ =0) and high values of the Wald statistic shows that the corresponding predictor variable is significant.

Conclusion, the findings indicate that Health status of a student is associated with sex, income, alcohol consumption, lack of hygiene of cafe, bad smelling around the café, smoking cigarette, usage of medicine, Environmental factor and physical exercise. From these the independent variables that have no significance association with Health status of a student are income, usage of medicine and physical exercise. And so sex, bad smelling around the café, lack of hygiene of café, smoking cigarette, Alcohol consumption, Environmental factor are found to be more significantly associated with health status of students.

By taking the exponent of the pooled estimate relative to a given predictor, i.e. taking $e^{\beta j}$, we obtain an estimate of the common odds ratio that describes the relative odds for $X_j$ differing by one unit. Values greater than one indicate that the variable in question increases the odds of being Unhealthy student and values between 0 and 1 indicate a decrease in the odds of being Unhealthy student. Thus, the above table demonstrates the lack of hygiene of café, shortage of water and hygine of toilet, smoking cigarette and alcohol consumption increases the odds of an individual student being Unhealthy is just the complement of the odds of being level healthy. $\exp(\hat{\beta})$, is a factor by which the odds of being Unhealthy of individual change when with independent variable increases by one unit (i.e. from 0 to 2). For instance from Table 4.6 the value of the odd ratio $\exp(\hat{\beta}) =$ 0.145, for low income and $\exp(\hat{\beta}) =$ 0.103, for high income indicates that students having income < 200 have 0.145 times chances to be Unhealthy than those having income between 200 & 600 and $\exp(\hat{\beta})$ = 0.360, for sex indicates that female students have 0.360 times chances to be Unhealthy than male student. Similarly, the values odds ratio $\exp(\hat{\beta}) =$ 0.287, for change of weather condition and environment indicates that the students those said change of weather condition and environment can affect health status have 0.287 times chances to be Unhealthy than those said not.

Table 4.7 Test of parallel lines

| Model | -2 Log Likelihood | Chi-Square | df | Sig. |
|---|---|---|---|---|
| Null Hypothesis | 201.745 | | | |
| General | 186.531(a) | 15.214(b) | 14 | .364 |

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

A. The log-likelihood value cannot be further increased after maximum number of step-halving.

B. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.

C Link function: Logit.

One of the assumptions underlying ordinal logistic regression is that the relationship between each pair of outcome groups is the same. This is commonly referred to as the test of parallel lines because the null hypothesis states that the slope coefficients in the model are the same across response categories (and lines of the same slope are parallel). If we ***fail to reject*** the null hypothesis, we conclude that the assumption holds.

From the above table shows parallel line test for general model with chi square value 15.214 and p-value=0.364 which is greater than the 5% level of significance, ***fail to reject*** the null hypothesis. Therefore, there is no enough evidence to reject the null hypothesis for general model. Thus, the proportional odds assumption appears to have held for general model.

## 5 Conclusions and Recommendations

### 5.1 Conclusions

The main objective of this study was to investigate the factors affecting health status of students in Ambo University. As we conclude this study was an attempt to examine the impact of some factors on the health status of students in this university. The finding shows that the student's health status is affected by some factors during their life in this university. Targeting this sex group (female) would greatly reduce the risk Health status when compared to males regarding to the observed data.

In this analysis we have looked at regression models that can be applied when our outcome is represented by an ordinal variable. Therefore we have seen how to evaluate the ordinal Proportional odds model by completing a series of tertiary logistic regressions at each of the cumulative splits in the data, and how this can allow us to directly evaluate the consistency in Odds ratios across an ordinal outcome. Where the Proportional odds assumption is justified ordinal regression models can be a powerful means of summarizing relationships that utilizes all the information present in the ordinal outcome.

Furthermore the findings indicate that Health status of a student is associated with sex, income, alcohol consumption, lack of hygiene of cafe, bad smelling around the café, smoking cigarette, usage of medicine, Environmental factor, physical exercise.

The independent variables that have no significance association with Health status of a student are lack of usage of medicine, income and physical exercise.

The Health status of a student is strongly associated with sex, smoking cigarette, lack of hygiene of café, Environmental factor, alcohol consumption and bad smelling around café. Perhaps due to the biological factors as well as the prevailing low socio economic status females for maintaining Health status is mandatory.

In conclusion, sex, lack of hygiene of café, usage of medicine, bad smelling around café, smoking cigarette and Environmental factor are found to be more significantly associated with health status of students in Ambo University students.

### 5.2 Recommendations

Based on the results of this study, we recommend a number of interventions to be put in place in Ambo University to reduce health status of students. These include:-

- ✓ The university should give emphasis on factors that affect health status of students.
- ✓ Female students have some differences on health status when compared to male students. In order to address this deference the university should give special treatment for female.
- ✓ The concerned bodies should work on minimizing lack of hygiene of café and bad smelling around café in the university.
- ✓ The Dean of Ambo University should give attention to problems and report for the concerned body to solve those problems.

## 6 .References


1. T. Croft **(1991)** Date Editing and Imputation. Demographic and health surveys world conference proceedings, II: 1337–1356,
2. A. Agresti, **(2002)** Categorical Data Analysis, Second Edition. Hoboken, New Jersey: John Wiley & Sons, Inc.
3. AM. Karim, W. Betemariam, S. Yalew, M. Carnell, Ethiopian Journal of Health Development **(2010)**24,1: 92–99.
4. J S. Long and J. Freese**,(2006)** Regression Models for Categorical and Limited Dependent Variables Using Stata, Second Edition. College Station, Texas: Stata Press.
5. T. F.Liao, **(1994)** Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models. Thousand Oaks, CA: Sage Publications, Inc.
6. G Angeles , DK Guilky DK, Sociological Methods & Research **(2005)** 34: 76–121
7. CAMH Healthy Aging Project. **(2008).** Improving Our Response to Older Adults with Substance Use, Mental Health and Gambling Problems: A Guide for Supervisors, Managers, and Clinical Staff. Toronto: Centre for Addiction and Mental Health.
8. A Admassiea, D Abebawa Journal of Development Effectiveness**(2009)** 1(4): 430–449