CHAPTER ONE
INTRODUCTION

1.1Bakckground:

People who complete suicide are depressed at the time of their deaths; depression that is untreated, undiagnosed, or ineffectively treated is the number one cause of suicide.

Suicide is the 3rd leading cause of death for 15 to 24-year-olds and 2nd for 24 to 35-year-olds, on average, 1 person commits suicide every 16.2 minutes.

From the above statistics it become clear it's huge problem so as statisticians even if we can't prevent this we can at least help by analysis predict if there is any hope of the survive of the person who commits suicide .

1.2Problem Identification:

Committing suicide, still not appreciated as a public health problem by the majority of health professionals and public policy experts and received only limited attention. Limitation of knowledge, absence of realistic to admit that there is problem to deal with and to figure out a solution for it.

1.3Problem definition:

After the person commits suicide by taking any kind of poisons and we receive him in ER we need to know what are the main factors (predictor or explanatory variables) that increase the likelihood of been survive and leave the system or not (the dependent variable)?

1.4Research Objectives:

The main objective is to study the probit analysis and to find suitable data for it, to make it clearer.

The specific objective is to estimate the inequalities in risk of being involve in over dose or taking poisons intentionally and whether he will survive or not so:

Describing frequencies, rate, contributing factors to the person who commits suicide.

Identify the major factor that would predict which patients most are at risk by constructing probit regression models to examine the predictor.

1.5 Research Hypothesis:

1. probit models do not fit to the data of committing suicide.
2. The liver function does not represent the main contributing factors of casualty.
3. Bilirubin test does not determine whether the suicidal person will survive or not.
4. Liver function does not determine whether the suicidal person will has neck stiffness.

1.6 Research Methodology:

1.6.1 Data source and type:

Data is pertaining to the different suicidal person who commits suicide from medical case.

1.6.2 Statistical method:

1. Descriptive statistics is employed in estimating factor.
2. Fitting probit regression model to examine the independent association of each casualty and the explanatory variables.

1.7 Research Important:

1. The application of the probit analysis is the most important aspect of this study because there is lack of study of it,
2. Analyze critical data like data of suicidal people and try to figure out a quick solution to help doctor by probabilities and predictions that whether each case they get will survive or not , that also help to be attention about the risk of poison and how it affect in our body at all.

1.8 Organization of the study:

The study organization is summarized such that chapter one including background, problem identification, problem definition, research objectives, research hypothesis, research methodology, importance of the study and organization.

Chapter two presented overall understanding of the concept of suicide and poisoning.

Chapter three reviewed the research method and the basic ideas of probit regression.

Chapter four is the analysis and discussion which includes the model building results.

Chapter five is the results and recommendations of the study.

CHAPTER TWO
LITERATURE REVIEW

2.1Commiting Suicide:

Suicide is a tragic and potentially preventable public health problem. In 2000, suicide was the 11th leading cause of death in the U.S. Specifically, 10.6 out of every 100,000 persons died by suicide. The total number of suicides was 29,350, or 1.2 percent of all deaths. Suicide deaths outnumber homicide deaths by five to three. It has been estimated that there may be from eight to 25 attempted suicides per every one suicide death. The alarming numbers of suicide deaths and attempts emphasize the need for carefully designed prevention efforts.

Suicidal behavior is complex. Some risk factors vary with age, gender and ethnic group and may even change over time. The risk factors for suicide frequently occur in combination. Research has shown that more than 90 percent of people who kill themselves have depression or another diagnosable mental or substance abuse disorder, often in combination with other mental disorders. Also, research indicates that alterations in neurotransmitters such as serotonin are associated with the risk for suicide. Diminished levels of this brain chemical have been found in patients with depression, impulsive disorders, a history of violent suicide attempts, and also in postmortem brains of suicide victims.

Adverse life events in combination with other risk factors such as depression may lead to suicide. However, suicide and suicidal behavior are not normal responses to stress. Many people have one or more risk factors and are not suicidal. Other risk factors include: prior suicide attempt; family history of mental disorder or substance abuse; family history of suicide; family violence, including physical or sexual abuse;

firearms in the home; incarceration; and exposure to the suicidal behavior of others, including family members, peers, or even in the media.

2.2 Gender Differences:

Suicide was the 8th leading cause of death for males and the 19th leading cause of death for females in 2000.1 More than four times as many men as women die by suicide,1 although women report attempting suicide during their lifetime about three times as often as men.5 Suicide by firearm is the most common method for both men and women, accounting for 57 percent of all suicides in 2000. White men accounted for 73 percent of all suicides and 80 percent of all firearm suicides.

2.3 Children, Adolescents, and Young Adults:

In 2000, suicide was the 3rd leading cause of death among 15- to 24-year-olds -- 10.4 of every 100,000 persons in this age group -- following unintentional injuries and homicide. Suicide was also the 3rd leading cause of death among children ages 10 to 14, with a rate of 1.5 per 100,000 children in this age group. The suicide rate for adolescents ages 15 to 19 was 8.2 deaths per 100,000 teenagers, including five times as many males as females. Among people 20 to 24 years of age, the suicide rate was 12.8 per 100,000 young adults, with seven times as many deaths among men as among women.

2.4 Older Adults:

Older adults are disproportionately likely to die by suicide. Comprising only 13 percent of the U.S. population, individuals age 65 and older accounted for 18 percent of all suicide deaths in 2000. Among the highest rates (when categorized by gender and race) were white men age 85 and older: 59 deaths per 100,000 persons, more than five times the national U.S. rate of 10.6 per 100,000.

2.5 Attempted Suicides:

Overall, there may be between eight and 25 attempted suicides for every suicide death; the ratio is higher in women and youth and lower in men and the elderly.2 Risk factors for attempted suicide in adults include depression, alcohol abuse, cocaine use, and separation or divorce.7,8 Risk factors for attempted suicide in youth include depression, alcohol or other drug use disorder, physical or sexual abuse, and disruptive behavior.8,9 As with people who die by suicide, many

people who make serious suicide attempts have co-occurring mental or substance abuse disorders. The majority of suicide attempts are expressions of extreme distress and not just harmless bids for attention. A suicidal person should not be left alone and needs immediate mental health treatment.

2.6 Prevention:

Preventive efforts to reduce suicide should be based on research that shows which risk and protective factors can be modified, as well as which groups of people are appropriate for the intervention. In addition, prevention programs must be carefully tested to determine if they are safe, truly effective, and worth the considerable cost and effort needed to implement and sustain them.

Many interventions designed to reduce suicidality also include the treatment of mental and substance abuse disorders. Because older adults, as well as women who die by suicide, are likely to have seen a primary care provider in the year prior to their suicide, improving the recognition and treatment of mental disorders and other suicide risk factors in primary care settings may be one avenue to prevent suicides among these groups.11 Improving outreach to men at risk for suicide is a major challenge in need of investigation.

Recently, the manufacturer of the medication clozapine received the first ever Food and Drug Administration indication for effectiveness in preventing suicide attempts among persons with schizophrenia.12 Additional promising pharmacologic and psychosocial treatments for suicidal individuals are currently being tested.

If someone is suicidal, he or she must not be left alone. Try to get the person to seek help immediately from his or her doctor or the nearest hospital emergency room, or call 911. It is also important to limit the person's access to firearms, medications, or other lethal methods for suicide.

2.7 Common Questions and Answers about Suicide:

What should you do if someone tells you they are thinking about suicide?

If someone tells you they are thinking about suicide, you should take their distress seriously, listen nonjudgmentally, and help them get to a professional for evaluation and treatment. People consider suicide when they are hopeless and unable to see alternative solutions to problems. Suicidal behavior is most often related to a mental disorder (depression) or to alcohol or other substance abuse. Suicidal behavior is also more likely to occur when people experience stressful events (major losses, incarceration). If someone is in imminent danger of harming himself or herself, do not leave the person alone. You may need to take emergency steps to get help, such as calling 911. When someone is in a suicidal crisis, it is important to limit access to firearms or other lethal means of committing suicide.

What are the most common methods of suicide?

Firearms are the most commonly used method of suicide for men and women, accounting for 60 percent of all suicides. Nearly 80 percent of all firearm suicides are committed by white males. The second most common method for men is hanging; for women, the second most common method is self-poisoning including

drug overdose. The presence of a firearm in the home has been found to be an independent, additional risk factor for suicide. Thus, when a family member or health care provider is faced with an individual at risk for suicide, they should make sure that firearms are removed from the home.

How do suicide rates compare between men and women?

More than four times as many men as women die by suicide; but women attempt suicide more often during their lives than do men, and women report higher rates of depression.

2.8 Suicide by poison:

Poison is anything that kills or injures through its chemical actions. Most poisons are swallowed (ingested). The word poison comes from the Latin word - potare - meaning to drink. But poisons can also enter the body in other ways:

- By breathing
- Through the skin
- By IV injection
- From exposure to radiation
- Venom from a snake bite or insect bite

2.9 Poisoning Causes:

Poisons include highly toxic chemicals not meant for human ingestion or contact, such as cyanide, paint thinners, or household cleaning products.

Many poisons, however, are substances meant for humans to eat, including foods and medicines.

Foods:

- Some mushrooms are poisonous
- Drinking water contaminated by agricultural or industrial chemicals
- Food that has not been properly prepared or handled

Drugs:

Drugs that are helpful in therapeutic doses may be deadly when taken in excess.

Examples include:

- Beta blockers: Beta blockers are a class of drugs used to treat heart conditions (for example, angina, abnormal heart rhythms) and other conditions, (for example, high blood pressure, migraine headache prevention, social phobia, and certain types of tremors). In excess, they can cause difficulty breathing, coma, and heart failure.
- Warfarin(Coumadin): Coumadin is a blood thinner used to prevent blood clots. It is the active ingredient in many rat poisons and may cause heavy bleeding and death if too much is taken.
- Vitamins: Vitamins, especially A and D, if taken in large amounts can cause liver problems and death.

CHAPTER THREE
Regression & PROBIT ANALYSIS

3.1 Introduction:

Linear regression analysis is the most widely used of all statistical techniques: it is the study of linear, additive relationships between variables, usually under the assumption of independently and identically normally distributed errors. Let Y denote the "dependent" variable whose values you wish to predict, and let X_1, …, X_k denote the "independent" variables from which you wish to predict it. Then the equation for predicting the value of Y at time t (or in row t of the data set), which will be denoted here by $\tilde{Y}(t)$, is of the form:

$$\tilde{Y}(t) = b_0 + b_1 X_1(t) + b_2 X_2(t) + \ldots + b_k X_k(t).$$

This formula has the property that the prediction for Y is a straight-line function of each of the X variables, holding the others fixed, and the contributions of different X variables to the predictions are additive. The slopes of their individual straight-line relationships with Y are the constants b_1, b_2, …, b_k, the so-called coefficients of the variables. That is, bi is the change in the predicted value of Y per unit of change in X_i, other things being equal, for i=1, …, k. The additional constant b0, the so-called intercept, is the prediction that the model would make if all the X's were zero (if that is possible). The coefficients and intercept are estimated by least squares, i.e., setting them equal to the unique values that minimize the sum of squared errors within the sample of data to which the model is fitted.

Galton termed this phenomenon a regression towards mediocrity, which in modern terms is a regression to the mean. To a naïve observer this might suggest that later generations are going to exhibit less variability--literally more mediocrity--than

earlier ones, but that is not case. It is a purely statistical phenomenon. Unless every child is exactly as the same size as the parent in relative terms (i.e., unless the correlation is exactly equal to 1), the predictions must regress to the mean regardless of biology.

Regression to the mean is an inescapable fact of life. Your children can be expected to be less exceptional (for better or worse) than you are. Your score on a final exam in a course can be expected to be less good (or bad) than your score on the midterm exam. A baseball player's batting average in the second half of the season can be expected to be closer to the mean (for all players) than his batting average in the first half of the season. And so on. The key word here is "expected." This does not mean it's certain that regression to the mean will occur, but that's the way to bet! More precisely, that's the way to bet if you wish to minimize squared error.

We have already seen a suggestion of regression-to-the-mean in some of the time series forecasting models we have studied: plots of forecasts tend to be smoother--i.e., they exhibit less variability--than the plots of the original data. This is not true of random walk models, but it is generally true of moving-average models and other models that base their forecasts on more than one past observation.

The intuitive explanation for the regression effect is simple: the thing we are trying to predict usually consists of a predictable component ("signal") and a statistically independent unpredictable component ("noise"). The best we can hope to do is to predict (only) that part of the variability which is due to the signal. Hence our forecasts will tend to exhibit less variability than the actual values, which implies a regression to the mean.

Another way to think of the regression effect is in terms of selection bias. In general a player's performance over any given period of time can be attributed to a combination of skill and luck. Suppose that we select a sample of professional athletes whose performance was much better than average (or students whose grades were much better than average) in the first half of the year. The fact that they did so well in the first half of the year makes it probable that both their skill and their luck were better than average during that period. In the second half of the year we may expect them to be equally skillful, but we should not expect them to be equally lucky. So we should predict that in the second half their performance will be closer to the mean. Meanwhile, players whose performance was merely average in the first half probably had skill and luck working in opposite directions for them. We should therefore expect their performance in the second half to move away from the mean in one direction or another, as we get another independent test of their skill. We don't know which direction they will move, though, so even for them we should predict that their second half performance will be closer to the mean than their first half performance. However, the actual performance of the players should be expected to have an equally large variance in the second half of the year as in the first half, because it merely results from a redistribution of independently random luck among players with the same distribution of skill as before.

3.2 Justification for regression assumptions:

Why should we assume that relationships between variables are linear?

1. Because linear relationships are the simplest non-trivial relationships that can be imagined (hence the easiest to work with).

2. Because the "true" relationships between our variables are often at least approximately linear over the range of values that are of interest to us.

3. Even if they're not, we can often transform the variables in such a way as to linearize the relationships.

This is a strong assumption, and the first step in regression modeling should be to look at scatter plots of the variables (and in the case of time series data, plots of the variables vs. time), to make sure it is reasonable a priori. And after fitting a model, plots of the errors should be studied to see if there are unexplained nonlinear patterns. This is especially important when the goal is to make predictions for scenarios outside the range of the historical data, where departures from perfect linearity are likely to have the biggest effect. If you see evidence of nonlinear relationships, it is possible (though not guaranteed) that transformations of variables will straighten them out in a way that will yield useful inferences and predictions via linear regression.

And why should we assume that the effects of different independent variables on the expected value of the dependent variable are additive? This is a very strong assumption, stronger than most people realize. It implies that the marginal effect of one independent variable (i.e., its slope coefficient) does not depend on the current values of other independent variables. But… why shouldn't it? It's conceivable that one independent variable could amplify the effect of another. In a multiple regression model, the estimated coefficient of a given independent variable supposedly measures its effect while "controlling" for the presence of the others. However, the way in which controlling is performed is extremely simplistic: multiples of other variables are merely added or subtracted.

Many users just throw a lot of independent variables into the model without thinking carefully about this issue, as if their software will automatically figure out exactly how they are related. It won't! Even "automatic" model-selection methods (e.g., stepwise regression) require you to have a good understanding of your own data and

to use a guiding hand in the analysis. They work only with the variables they are given, in the form that they are given, and then they look only for linear, additive patterns among them in the context of each other. You need to collect the relevant data, clean it up if necessary, perform descriptive analysis to look for patterns before fitting any models, and study the diagnostic tests of model assumptions afterward, especially statistics and plots of the errors. You should also try to apply the appropriate economic or physical reasoning. Here too, it is possible (but not guaranteed) that transformations of variables or the inclusion of interaction terms might separate their effects into an additive form, if they do not have such a form to begin with, but this requires some thought and effort on your part.

And why should we assume the errors of linear models are independently and identically normally distributed?

1. This assumption is often justified by appeal to the Central Limit Theorem of statistics, which states that the sum or average of a sufficiently large numbers of independent random variables--whatever their individual distributions--approaches a normal distribution. Much data in business and economics and engineering and the natural sciences is obtained by adding or averaging numerical measurements performed on many different persons or products or locations or time intervals. Insofar as the activities that generate the measurements may occur somewhat randomly and somewhat independently, we might expect the variations in the totals or averages to be somewhat normally distributed.
2. It is (again) mathematically convenient: it implies that the optimal coefficient estimates for a linear model are those that minimize the mean squared error (which are easily calculated), and it justifies the use of a host of statistical

tests based on the normal family of distributions. (This family includes the t distribution, the F distribution, and the Chi-square distribution.)

3. Even if the "true" error process is not normal in terms of the original units of the data, it may be possible to transform the data so that your model's prediction errors are approximately normal.

But here too caution must be exercised. Even if the unexplained variations in the dependent variable are approximately normally distributed, it is not guaranteed that they will also be identically normally distributed for all values of the independent variables. Perhaps the unexplained variations are larger under some conditions than others, a condition known as "heteroscedasticity". For example, if the dependent variable consists of daily or monthly total sales, there are probably significant day-of-week patterns or seasonal patterns. In such cases the variance of the total will be larger on days or in seasons with greater business activity--another consequence of the central limit theorem. (Variable transformations such as logging and/or seasonal adjustment are often used to deal with this problem.) It is also not guaranteed that the random variations will be statistically independent. This is an especially important question when the data consists of time series: if the model is not correctly specified, it is possible that consecutive errors (or errors separated by some other number of periods) will have a systematic tendency to have the same sign or a systematic tendency to have opposite signs, a phenomenon known as "autocorrelation" or "serial correlation".

A very important special case is that of stock price data, in which percentage changes rather than absolute changes tend to be normally distributed. This implies that over moderate to large time scales, movements in stock prices are lognormally distributed rather than normally distributed. A log transformation is typically applied to historical stock price data when studying growth and

volatility. Caution: although simple regression models are often fitted to historical stock returns to estimate "betas", which are indicators of relative risk in the context of a diversified portfolio, I do not recommend that you use regression to try to predict future stock returns.

You still might think that variations in the values of portfolios of stocks would tend to be normally distributed, by virtue of the central limit theorem, but the central limit theorem is actually rather slow to bite on the lognormal distribution because it is so asymmetrically long-tailed. A sum of 10 or 20 independently and identically lognormally distributed variables has a distribution that is still quite close to lognormal. If you don't believe this, try testing it by Monte Carlo simulation: you'll be surprised. (I was.)

Because the assumptions of linear regression (linear, additive relationships with i.i.d. normally distributed errors) are so strong, it is very important to test their validity when fitting models, a topic discussed in more detail on the testing-model-assumptions page, and be alert to the possibility that you may need more or better data to accomplish your objectives. You can't get something from nothing. All too often, naïve users of regression analysis view it as a black box that can automatically predict any given variable from any other variables that are fed into it, when in fact a regression model is a very special and very transparent kind of prediction box. Its output contains no more information than is provided by its inputs and its inner mechanism needs to be compared with reality in each situation where it is applied.

3.3 Correlation and simple regression formulas:

A variable is, by definition, a quantity that may vary from one measurement to another in situations where different samples are taken from a population or observations are made at different points in time. In fitting statistical models in

which some variables are used to predict others, what we hope to find is that the different variables do not vary independently (in a statistical sense), but that they tend to vary together.

In particular, when fitting linear models, we hope to find that one variable (say, Y) is varying as a straight-line function of another variable (say, X). In other words, if all other possibly-relevant variables could be held fixed, we would hope to find the graph of Y versus X to be a straight line (apart from the inevitable random errors or "noise").

A measure of the absolute amount of "variability" in a variable is (naturally) its variance, which is defined as its average squared deviation from its own mean. Equivalently, we can measure variability in terms of the standard deviation, which is defined as the square root of the variance. The standard deviation has the advantage that it is measured in the same units as the original variable, rather than squared units.

Our task in predicting Y might be described as that of "explaining" some or all of its variance--i.e., why, or under what conditions, does it deviate from its mean? Why is it not constant? That is, we would like to be able to improve on the "naive" predictive model: Ý(t) = CONSTANT, in which the best value for the constant is presumably the historical mean of Y. More precisely, we hope to find a model whose prediction errors are smaller, in a mean square sense, than the deviations of the original variable from its mean.

In using linear models for prediction, it turns out very conveniently that the only statistics of interest (at least for purposes of estimating coefficients to minimize squared error) are the mean and variance of each variable and

the correlation coefficient between each pair of variables. The coefficient of correlation between X and Y is commonly denoted by r_{XY}.

The correlation coefficient between two variables is a statistic that measures the strength of the linear relationship between them, on a relative (i.e., unit less) scale of -1 to +1. That is, it measures the extent to which a linear model can be used to predict the deviation of one variable from its mean given knowledge of the other's deviation from its mean at the same point in time.

The correlation coefficient is most easily computed if we first standardize each of the variables--i.e., express it in units of standard deviations from its own mean--using the population standard deviation (the one whose formula has n rather than n-1 in the denominator), rather than the sample standard deviation. The standardized value of X will be denoted here by XSTD, and the value of XSTD in period t is defined in Excel notation as:

XSTD(t) = (X(t) - AVERAGE(X))/STDEV.P(X)

(I am going to be a bit sloppy and use Excel functions rather than conventional math symbols in some places to avoid problems with fonts, as well as to illustrate how the calculations would be done on a spreadsheet.) For example, suppose that AVERAGE(X) = 20 and STDEV.P(X) = 5. If X(t) = 25, then XSTD(t) = 1, if X(t) = 10, then XSTD(t) = -2, and so on. YSTD will denote the similarly standardized value of Y.)

Now, the correlation coefficient is equal to the average product of the standardized values of the two variables:

$$r_{XY} = \text{AVERAGE(XSTDYSTD)} = (\text{XSTD(1)YSTD(1)} + \text{XSTD(2)YSTD(2)} + \ldots + \text{XSTD(n)YSTD(n)})/n$$

... Where n is the sample size. Thus, for example, if X and Y are stored in columns on a spreadsheet, you can use the AVERAGE and STDEV.P functions to compute their averages and population standard deviations, then you can create two new columns in which the values of XSTD and YSTD in each row are computed according to the formula above. Then create a third new column in which XSTD is multiplied by YSTD in every row. The average of the values in the last column is the correlation between X and Y. Of course, in Excel, you can just use the formula =CORREL(X,Y) to calculate a correlation coefficient, where X and Y denote the cell ranges of the data for the variables. (Note: in some situations it might be of interest to standardize the data relative to the sample standard deviation, but the population statistic is the correct one to use in the formula above.)

If the two variables tend to vary on the same sides of their respective means at the same time, then the average product of their deviations (and hence the correlation between them) will be positive, since the product of two numbers with the same sign is positive. Conversely, if they tend to vary on opposite sides of their respective means at the same time, their correlation will be negative. If they vary independently with respect to their means--that is, if one is equally likely to be above or below its mean regardless of what the other is doing--then the correlation will be zero.

The correlation coefficient can be said to measure the strength of the linear relationship between Y and X for the following reason. The linear equation for predicting YSTD from XSTD that minimizes mean squared error is simply:

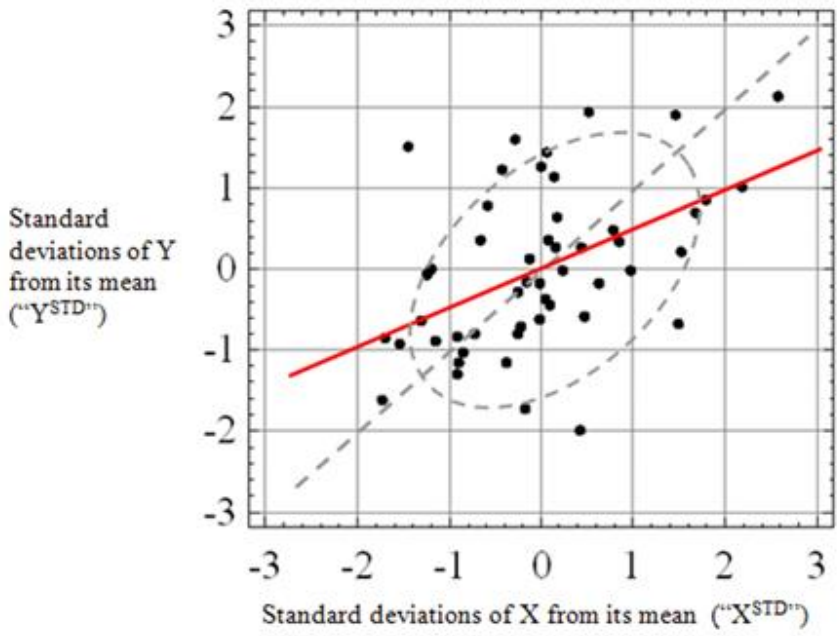
$$\tilde{Y}_{STD}(t) = r_{XY} X_{STD}(t)$$

Where $\tilde{Y}_{STD}$ denotes the prediction for Y_{STD} . Thus, if X is observed to be 1 standard deviation above its own mean, then we should predict that Y will be rXY standard deviations above its own mean; if X is 2 standard deviations below its own mean, then we should be predict that Y will be 2rXY standard deviations below its own mean, and so on.

In graphical terms, this means that, on a scatter plot of Y_{STD} versus X_{STD}, the line for predicting Y_{STD} from X_{STD} so as to minimize mean squared error is the line that passes through the origin and has slope r_{XY}. This fact is not supposed to be obvious, but it is easily proved by elementary differential calculus of several variables.

Here is an example: on a scatter plot of YSTD versus XSTD, the visual axis of symmetry is a line that passes through the origin and whose slope is equal to 1 (i.e., a 45-degree line), which is the gray dashed line on the plot below. It passes through the origin because the means of both standardized variables are zero, and its slope is equal to 1 because their standard deviations are both equal to 1. (The latter fact means that the points are equally spread out horizontally and vertically in terms of mean squared deviations from zero, which forces their pattern to appear roughly symmetric around the 45-degree line if the relationship between the variables really is linear.) However, the gray dashed line is the not the best line to use for predicting the value of YSTD for a given value of XSTD. The best line for predicting YSTD from XSTD has a slope of less than 1: it regresses toward the X axis. The regression line is shown in red, and its slope is the correlation between X and Y, which is 0.46 in this case. Why is this true? Because, that's the way to bet if you want to minimize the mean squared error measured in the Y direction . If instead you wanted to predict X_{STD} from Y_{STD} so as to minimize mean squared error

measured in the X direction, the line would regress in the other direction relative to the 45-degree line, and by exactly the same amount.



If we want to obtain the linear regression equation for predicting Y from X in unstandardized terms, we just need to substitute the formulas for the standardized values in the preceding equation, which then becomes:

$$(\tilde{Y}(t) - \text{AVERAGE}(Y))/\text{STDEV.P}(Y) = r_{XY}\,(X(t) - \text{AVERAGE}(X))/\text{STDEV.P}(X).$$

If we now rearrange this equation and collect constant terms, we obtain:

$$\tilde{Y}(t) = b_0 + b_1\,X(t)$$

where:

$b_1 = r_{XY}\,(\text{STDEV.P}(Y)/\text{STDEV.P}(X))$ is the estimated slope of the regression line, and

$b_0 = \text{AVERAGE}(Y) - b_1\,(\text{AVERAGE}(X))$ is the estimated Y-intercept of the line.

Notice that, as we claimed earlier, the coefficients in the linear equation for predicting Y from X depend only on the means and standard deviations of X and Y and on their coefficient of correlation.

The additional formulas that are needed to compute standard errors, t-statistics, and P-values (statistics that measure the precision and significance of the estimated coefficients) are given in this set of notes and also illustrated in this spreadsheet file. Perfect positive correlation ($r_{XY} = +1$) or perfect negative correlation ($r_{XY} = -1$) is only obtained if one variable is an exact linear function of the other, without error. In such a case, one variable is merely a linear transformation of the other--they aren't really "different" variables at all!

In general, therefore, we find less-than-perfect correlation, which is to say, we find that r_{XY} is less than 1 in absolute value. Therefore our prediction for Y_{STD} will typically be smaller in absolute value than our observed value for X_{STD}. That is, we will always predict Y to be closer to its own mean, in units of its own standard deviation, than X was observed to be, which is Galton's phenomenon of regression to the mean.

So, the technical explanation of the regression-to-the-mean effect hinges on two mathematical facts: (i) the correlation coefficient, calculated in the manner described above, happens to be the coefficient that minimizes the squared error in predicting Y_{STD} from X_{STD}, and (ii) the correlation coefficient is never larger than 1 in absolute value, and it is only equal to 1 when YSTD is an exact (noiseless) linear function of XSTD.

The term "regression" has stuck and has even mutated from an intransitive verb into a transitive one since Galton's time. We don't merely say that the predictions for Y "regress to the mean"--we now say that we are "regressing Y on X" when we

estimate a linear equation for predicting Y from X, and we refer to X as a "regressor" in this case.

When we have fitted a linear model, we can compute its mean squared prediction error and compare this to the variance of the original variable. As noted above, we hope to find that the MSE is less than the original variance. The relative amount by which the mean squared error is less than the variance of the original variable is referred to as the fraction of the variance that was explained by the model. For example, if the MSE is 20% less than the original variance, we say we have "explained 20% of the variance."

It turns out that in a simple regression model (a linear model with only one "X" variable), the fraction of variance explained is precisely the square of the correlation coefficient--i.e., the square of r. Hence, the fraction-of-variance-explained has come to be known as "R-squared". The interpretation and use of R-squared are discussed in more detail .

In a multiple regression model (a linear model with two or more "X" variables), there are many correlation coefficients that must be computed, in addition to all the means and variances. For example, we must consider the correlation between each X variable and the Y variable, and also the correlation between each pair of X variables. In this case, it still turns out that the model coefficients and the fraction-of-variance-explained statistic can be computed entirely from knowledge of the means, standard deviations, and correlation coefficients among the variables--but the computations are no longer easy. We will leave those details to the computer.

3.4 Probit analysis:

• Probit analysis is a type of regression used to analyze binomial response variables.

• It transforms the sigmoid dose-response curve to a straight line that can then be analyzed by regression either through least squares or maximum likelihood.

• Probit analysis can be conducted by one of three techniques:

Using tables to estimate the probits and fitting the relationship by eye,

Hand calculating the probits, regression coefficient, and confidence intervals, or Having a statistical package such as SPSS do it all for you.

In probability theory and statistics, the probit function is the quantile function associated with the standard normal distribution . It has applications in exploratory statistical graphics and specialized regression modeling of binary response variables.

The standard normal distribution is commonly denoted as N(0,1) and its cumulative distribution function as $\Phi(z)$. As an example, consider the familiar fact that the standard normal distribution places 95% of probability between −1.96 and 1.96, and is symmetric around zero. It follows that:

$$\Phi(-1.96) = 0.025 = 1 - \Phi(1.96).$$

The probit function gives the 'inverse' computation, generating a value of an N(0,1) random variable, associated with specified cumulative probability. Formally, the probit function is the inverse of $\Phi(z)$, denoted $\Phi^{-1}(p)$. Continuing the example,

$$\text{probit}(0.025) = -1.96 = -\text{probit}(0.975).$$

In general,

$$\Phi(\text{probit}(p)) = p$$

and

$$\text{probit}(\Phi(z)) = z.$$

3.5 POBIT REGRESSION MODEL:

Probit Analysis is a method of analyzing the relationship between a stimulus (dose) and the quantal (all or nothing) response. Quantitative responses are almost always preferred, but in many situations they are not practical. In these cases, it is only possible to determine if a certain response (such as death) has occurred. In a typical quantal response experiment, groups of animals are given different doses of a drug. The percent dying at each dose level is recorded. These data may then be analyzed using Probit Analysis.

The Probit Model assumes that the percent response is related to the log dose as the cumulative normal distribution. That is, the log doses may be used as variables to read the percent dying from the cumulative normal. Using the normal distribution, rather than other probability distributions, influences the predicted response rate at the high and low ends of possible doses, but has little influence near the middle. Hence, much of the comparison of different drugs is done using response rates of fifty percent.

The popularity of the method is due in large part to the work of Finney (1971), in his book Probit Analysis. He explains the proper use and analysis of quantal response data. In NCSS, we have coded the algorithms given in his book, and we refer you to it for further information and background.

3.6 Background:

The idea of probit analysis was originally published in Science by Chester Ittner Bliss in 1934. He worked as an entomologist for the Connecticut agricultural experiment station and was primarily concerned with finding an effective pesticide

to control insects that fed on grape leaves (Greenberg 1980). By plotting the response of the insects to various concentrations of pesticides, he could visually see that each pesticide affected the insects at different concentrations, i.e. one was more effective than the other. However, he didn't have a statistically sound method to compare this difference. The most logical approach would be to fit a regression of the response versus the concentration, or dose and compare between the different pesticides. Yet, the relationship of response to dose was sigmoid in nature and at the time regression was only used on linear data. Therefore, Bliss developed the idea of transforming the sigmoid dose-response curve to a straight line. In 1952, a professor of statistics at the University of Edinburgh by the name of David Finney took Bliss' idea and wrote a book called Probit Analysis (Finney 1952). Today, probit analysis is still the preferred statistical method in understanding dose-response relationships.

3.7 The Basics:

Probit Analysis is a specialized regression model of binomial response variables.

Remember that regression is a method of fitting a line to your data to compare the relationship of the response variable or dependent variable (Y) to the independent variable (X).

$$\mathbf{Y = a + b\,X + e}$$

Where

a = y-intercept

b = the slope of the line

e = error term

Also remember that a binomial response variable refers to a response variable with only two outcomes. For example:

Flipping a coin: Heads or tails.

Testing beauty products: Rash/no rash.

The effectiveness or toxicity of pesticides: Death/no death.

3.8 Applications:

Probit analysis is used to analyze many kinds of dose-response or binomial response experiments in a variety of fields. However, because my background knowledge of probit analysis stems only from toxicology, the examples from this webpage will only be of toxicology.

Probit Analysis is commonly used in toxicology to determine the relative toxicity of chemicals to living organisms. This is done by testing the response of an organism under various concentrations of each of the chemicals in question and then comparing the concentrations at which one encounters a response. As discussed above, the response is always binomial (e.g. death/no death) and the relationship between the response and the various concentrations is always sigmoid. Probit analysis acts as a transformation from sigmoid to linear and then runs a regression on the relationship.

Once a regression is run, the researcher can use the output of the probit analysis to compare the amount of chemical required to create the same response in each of the various chemicals. There are many endpoints used to compare the differing toxicities of chemicals, but the LC50 (liquids) or LD50 (solids) are the most widely used outcomes of the modern dose-response experiments. The LC50/LD50 represent the concentration (LC50) or dose (LD50) at which 50% of the population responds.

For example, consider comparing the toxicity of two different pesticides to aphids, pesticide A and pesticide B. If the LC50 of pesticide A is 50ug/L and the LC50 of pesticide B is 10ug/L, pesticide B is more toxic than A because it only takes 10ug/L to kill 50% of the aphids, versus 50ug/L of pesticide B.

3.9 Logit vs. Probit:

Logit is another form of transforming binomial data into linearity and is very similar to probit. Logit functions by taking the log of the odds: logit(P) = log P/ (1-P). Yet, the relationship between logit and probit is almost indistinguishable: Logit ≈ ($\pi/\sqrt{3}$) x probit. In general, if response vs dose data are not normally distributed, Finney suggests using the logit over the probit transformation (Finney, 1952). Although the multivariate usage of probit analysis is beyond the content of this webpage, it is worth noting that the similarity between probit and logit doesn't hold in a multivariate realm (Hahn and Soyer date unknown). Hahn and Soyer suggest that logit provides a better fit in the presence of extreme independent variable levels and conversely that probit better fit random effects models with moderate data sets (Hahn and Soyer date unknown).

3.10 Summary:

Probit Analysis is a type of regression used with binomial response variables. It is very similar to logit, but is preferred when data are normally distributed.

Most common outcome of a dose-response experiment in which probit analysis is used is the LC50/LD50.

Probit analysis can be done by eye, through hand calculations, or by using a statistical program.

CHAPTER FOUR
STATISTICAL ANALYSIS

4.1Probit Analysis: survive versus amylase (liver function)

Distribution: Normal

Table (4.1): Response Information:

Variable	Value	Count
survive	1(survive)	8
	0(not survive)	65
	Total	73

From above table we find 8 survive and 65 don't survive from total observations which are target of the study**.**

Estimation Method: Maximum Likelihood

Table (4.2): Regression Table:

Variable	Coef	Standard Error	Z	P
Constant	-4.25000	1.49525	-2.84	0.004
amylase	0.682734	0.325858	2.10	0.036
Natural Response	0 (the probability that a unit fails without being exposed to any of the stress).			

Estimated regression model estimates of the coefficients in the regression model, with standard errors.

The p-value (0.004 ,0.036) for the model of survive versus the liver function shows the addition of the predictor variables(liver function) significantly reduces the

deviance compared to a model containing only a constant term. A small p-value (less than 0.05 if operating at the 5% significant level) indicates that the model has significantly reduced the deviance and is thus useful for predicting the probability of the studied outcome

Log-Likelihood = -22.691

Table (4.3) Goodness-of-Fit Tests:

Method	Chi-Square	DF	P
Pearson	28.0401	17	0.044
Deviance	22.5000	17	0.166

The p-value for the residual term tests whether there is significant a sign of a better model may be possible. A small p-value indicates that significant deviance remains in the residuals, so that a better model might be possible.

Tolerance Distribution:

Table (4.4): Parameter Estimates:

Parameter	Estimate	Standard Error	95.0% Normal CI	
			Lower	Upper
Mean	6.22497	0.857212	4.54487	7.90508
StDev	1.46470	0.699079	0.574755	3.73262

Table (4.5) of Percentiles:

Percent	Percentile	Standard Error	95.0% Fiducial CI	
			Lower	Upper
1	2.81757	0.878096	-22.3235	3.76074
2	3.21685	0.702282	-16.1692	3.99924
3	3.47017	0.595461	-12.2765	4.16265
4	3.66074	0.519192	-9.36002	4.29740
5	3.81576	0.461098	-7.00091	4.42024
6	3.94770	0.415648	-5.00904	4.54088
7	4.06338	0.379935	-3.28351	4.66761
8	4.16696	0.352256	-1.76740	4.80999
9	4.26117	0.331507	-0.430597	4.98151
10	4.34788	0.316878	0.736402	5.20293
20	4.99225	0.370460	4.37923	11.8772
30	5.45688	0.530262	4.81283	18.8829
40	5.85389	0.694330	5.07848	24.9739
50	6.22497	0.857212	5.30138	30.6923
60	6.99306	1.02483	5.51313	36.4220
80	7.45769	1.20716	5.73301	42.5587
90	8.10206	1.42293	5.98526	49.7458
91	8.18878	1.72468	6.32984	59.7183
92	8.28298	1.76544	6.37589	61.0607
93	8.38656	1.80975	6.42586	62.5190
94	8.50225	1.85851	6.48074	64.1226
95	8.63419	1.91301	6.54193	65.9137
96	8.78920	1.97521	6.61163	67.9565
97	8.97977	2.04836	6.69340	70.3567
98	9.23310	2.13836	6.79376	73.3075
99	9.63237	2.25811	6.92694	77.2304

4.2 Probit Analysis: neck versus amylase

Distribution: Normal

Table (4.6) Response Information:

Variable	Value	Count
neck	1	47 (Event)
	0	26
	Total	73

Estimation Method: Maximum Likelihood:

Table (4.7) Regression Table:

Variable	Coef	Standard Error	Z	P
Constant	2.62513	0.999752	2.63	0.009
amylase	-0.527725	0.229247	-2.30	0.021
Natural Response	0 (the probability that a unit fails without being exposed to any of the stress).			

Estimated regression model estimates of the coefficients in the regression model, with standard errors.

The p-value (0.009, 0.021) for the model of neck versus the liver function shows the addition of the predictor variables(liver function) significantly reduces the deviance compared to a model containing only a constant term. A small p-value (less than 0.05 if operating at the 5% significant level) indicates that the model has significantly reduced the deviance and is thus useful for predicting the probability of the studied outcome.

Log-Likelihood = -44.704

Table (4.8) Goodness-of-Fit Tests:

Method	Chi-Square	DF	P
Pearson	43.6097	17	0.000
Deviance	51.3973	17	0.000

The p-value for the residual term tests whether there is significant a sign of a better model may be possible. A small p-value indicates that significant deviance remains in the residuals, so that a better model might be possible.

Tolerance Distribution:

Table (4.9) Parameter Estimates:

Parameter	Estimate	Standard Error	95.0% Normal CI	
			Lower	Upper
Mean	4.97444	0.411337	4.16823	5.78064
StDev	1.89493	0.823167	0.808771	4.43976

Table (4.10) of Percentiles:

percent	percentiles	Standard error	95.0% Fiducial CI	
			Lower	Upper
1	9.38270	2.22349	7.01131	38.4963
2	8.86614	2.00126	6.72801	35.0240
3	8.53840	1.86053	6.54773	32.8214
4	8.29186	1.75485	6.41175	31.1649
5	8.09132	1.66901	6.30088	29.8177
6	7.92062	1.59606	6.20629	28.6713
7	7.77095	1.53219	6.12316	27.6663
8	7.63695	1.47509	6.04855	26.7666
9	7.51507	1.42324	5.98054	25.9485
10	7.40288	1.37558	5.91778	25.1956
20	6.56925	1.02463	5.44482	19.6077
30	5.96814	0.777948	5.08998	15.5922
40	5.45451	0.577329	4.76294	12.1849
50	4.97444	0.411337	4.39628	9.06119
60	4.49436	0.303434	3.74862	6.21848
70	3.98074	0.325378	1.62458	4.60822
90	3.37963	0.498466	-2.14522	4.00766
91	2.54599	0.819715	-7.66354	3.46506
92	2.31193	0.865415	-8.41222	3.39809
93	2.43381	0.915417	-9.22635	3.32614
94	2.17792	0.970757	-10.1223	3.24781
95	2.02826	1.03294	-11.1238	3.16113
96	1.85756	1.10425	-12.2668	3.06314
97	1.65702	1.18848	-13.6107	2.94895
98	1.41047	1.29256	-15.2639	2.80966
99	1.08273	1.43160	-17.4630	2.62595

4.3 Probit Analysis: passing versus amylase:

Distribution: Normal

Table (4.11) Response Information:

Variable	Value	Count
passing	2	33 (event)
	1	40
	Total	73

Estimation Method: Maximum Likelihood

Table (4.12) Regression Table:

Variable	Coef	Standard Error	Z	P
Constant	0.338460	0.870565	0.39	0.697
amylase	-0.108915	0.203805	-0.53	0.593
Natural Response	0 (the probability that a unit fails without being exposed to any of the stress).			

Estimated regression model estimates of the coefficients in the regression model, with standard errors.

The p-value (0.697, 0.593) for the model of passing versus the liver function shows the addition of the predictor variables (liver function) isn't significantly reduces the deviance compared to a model containing only a constant term. A large p-value more than 0.05 if operating at the 5% significant level indicates that the model hasn't significantly reduced the deviance and isn't useful for predicting the probability of the studied outcome.

Means there is no association between the liver function and passing which is sociological problem.

So I use this analysis just to show how probit technique.

Log-Likelihood = -50.121

Table (4.13) Goodness-of-Fit Tests:

Method	Chi-Square	DF	P
Pearson	43.1778	17	0.000
Deviance	56.8249	17	0.000

The p-value for the residual term tests whether there is significant a sign of a better model may be possible. A small p-value indicates that significant deviance remains in the residuals, so that a better model might be possible.

Tolerance Distribution:

Tabl (4.14) Parameter Estimates:

Parameter	Estimate	Standard Error	95.0% Normal CI	
			Lower	Upper
Mean	3.10755	2.46647	-1.72663	7.94174
StDev	9.18145	17.1806	0.234488	359.503

Table (4.15) of Percentiles:

percent	percentiles	Standard error	95.0% Fiducial CI	
			Lower	Upper
1	24.4668	37.9291	*	*
2	21.9639	33.2491	*	*
3	20.3760	30.2803	*	*
4	19.1814	28.0474	*	*
5	18.2097	26.2314	*	*
6	17.3826	24.6860	*	*
7	16.6575	23.3312	*	*
8	16.0081	22.1183	*	*
9	15.4176	21.0155	*	*
10	14.8741	20.0006	*	*
20	10.8349	12.4701	*	*
30	7.92231	7.07687	*	*
40	5.43365	2.65901	*	*
50	3.10755	2.46647	*	*
60	0.781460	6.55650	*	*
70	-1.70720	11.1547	*	*
90	-4.61975	16.5777	*	*
91	-8.65895	24.1187	*	*
92	-9.20252	25.1343	*	*
93	-9.79304	26.2377	*	*
94	-10.4423	27.4512	*	*
95	-11.1675	28.8066	*	*
96	-11.9946	30.3526	*	*
97	-12.9663	32.1692	*	*
98	-14.1609	34.4026	*	*
99	-15.7488	37.3720	*	*

We notes here when there is no significance association between the response variable and the explanatory factor we can't calculate the 95.0% Fiducial CI for the Percentiles.

4.4 Probit Analysis: passing versus bilirubin

Distribution:Normal

Table (4.16) Response Information:

Variable	Value	Count
passing	2	33 (event)
	1	40
	Total	**73**

Estimation Method: Maximum Likelihood

Table (4.17): Regression

Variable	Coef	Standard Error	Z	P
Constant	0.0329984	0.285819	0.12	0.908
amylase	-0.0018518	0.0029679	-0.62	0.533
Natural Response	0 (the probability that a unit fails without being exposed to any of the stress).			

Estimated regression model estimates of the coefficients in the regression model, with standard errors.

The p-value (0.908, 0.533) for the model of passing versus the bilirubin test

Shows the addition of the predictor variables (bilirubin) isn't significantly reduces the deviance compared to a model containing only a constant term. A large p-value more than 0.05 if operating at the 5% significant level indicates that the model has not significantly reduced the deviance and isn't thus useful for predicting the probability of the studied outcome

Log-Likelihood = -50.065

Table (4.18) Goodness-of-Fit Tests:

Method	Chi-Square	DF	P
Pearson	68.2560	30	0.000
Deviance	93.5375	30	0.000

Tolerance Distribution:

Table (4.19): Parameter Estimates:

Parameter	Estimate	Standard Error	95.0% Normal CI	
			Lower	Upper
Mean	17.8196	130.704	-238.355	273.994
StDev	540.014	865.488	23.3439	12492.2

Table (4.20): of Percentiles:

percent	percentiles	Standard error	95.0% Fiducial CI	
			Lower	Upper
1	1126.87	1911.38	*	*
2	1274.08	1675.68	*	*
3	1033.47	1526.17	*	*
4	963.215	1413.73	*	*
5	906.064	1322.29	*	*
6	857.420	1244.48	*	*
7	814.768	1176.27	*	*
8	776.578	1115.21	*	*
9	741.846	1059.69	*	*
10	709.876	1008.60	*	*
20	472.307	629.752	*	*
30	301.003	359.079	*	*
40	154.631	140.297	*	*
50	17.8196	130.704	*	*
60	-118.991	332.628	*	*
70	-265.364	563.215	*	*
80	-436.668	835.914	*	*
90	-674.237	1215.48	*	*
91	-706.207	1266.61	*	*
92	-740.939	1322.17	*	*
93	-779.129	1383.28	*	*
94	-821.780	1451.53	*	*
95	-870.425	1529.38	*	*
96	-927.576	1620.86	*	*
97	-997.836	1733.34	*	*
98	-1091.23	1882.88	*	*
99	-1238.44	2118.63	*	*

4.5 Probit Analysis: survive versus bilirubin

Distribution: Normal

Table (4.21) Response Information:

Variable	Value	Count
survive	1	8 (event)
	0	65
	Total	73

Estimation Method: Maximum Likelihood

Table (4.22): Regression

Variable	Coef	Standard Error	Z	P
Constant	-1.57906	0.386154	-4.09	0.000
bilirubi	0.0039042	0.003569	1.09	0.274
Natural Response	0 (the probability that a unit fails without being exposed to any of the stress).			

Estimated regression model estimates of the coefficients in the regression model, with standard errors.

The p-value (0.000, 0.274) for the model of survive versus the bilirubin test

A show the addition of the predictor variables (bilirubin) isn't significantly reduces the deviance compared to a model containing only a constant term. A large p-value more than 0.05 if operating at the 5% significant level indicates that the model has not significantly reduced the deviance and isn't thus useful for predicting the probability of the studied outcome

But medically there is strong relationship between bilirubin tests and survive of the suicidal so if we increase the sample it might change the result.

Log-Likelihood = -24.641

Table (4.23) Goodness-of-Fit Tests:

Method	Chi-Square	DF	P
Pearson	43.4861	30	0.053
Deviance	32.2797	30	0.355

The p-value for the residual term tests whether there is significant a sign of a better model may be possible. A small p-value indicates that significant deviance remains in the residuals, so that a better model might be possible.

Tolerance Distribution:

Table (4.24) Parameter Estimates:

Parameter	Estimate	Standard Error	95.0% Normal CI	
			Lower	Upper
Mean	404.454	289.209	-162.386	971.294
StDev	256.135	234.166	42.6851	1536.96

Table (4.25) of Percentiles:

percent	percentiles	Standard error	95.0% Fiducial CI	
			Lower	Upper
1	-191.406	264.867	*	*
2	-121.584	202.575	*	*
3	-77.2836	163.668	*	*
4	-43.9585	135.020	*	*
5	-16.8511	112.425	*	*
6	6.22161	94.0641	*	*
7	26.4518	79.0953	*	*
8	44.5656	67.1994	*	*
9	61.0393	58.3888	*	*
10	76.2034	52.8366	*	*
20	188.885	101.210	*	*
30	270.137	169.667	*	*
40	339.563	231.031	*	*
50	404.454	289.209	*	*
60	469.345	347.776	*	*
70	538.772	410.677	*	*
80	620.023	484.480	*	*
90	732.705	587.033	*	*
91	747.869	600.846	*	*
92	764.343	615.855	*	*
93	782.456	632.361	*	*
94	802.687	650.798	*	*
95	825.759	671.830	*	*
96	852.867	696.545	*	*
97	886.192	726.935	*	*
98	930.492	767.342	*	*
99	1000.31	831.047	*	*

4.6 Probit analysis (output):

The default output consists of:

- The response information.
- The factor information.
- The regression table, which include the estimated coefficients and their standard errors.

- Z-value and p values: The Z. test tests that the coefficient is significantly different than 0; in other words, is it a significant predictor?
- Natural response rate: the probability that a unit fails without being exposed to any of the stress.
- The test for equal slopes, which tests that the slopes associated with the factor levels are significantly different.
- The log. Likelihood from the last iteration of the algorithm.
- The goodness-of fit tests, which evaluate how well the model fits the data. The null hypothesis is that the model fits the data adequately. Therefore, the higher the p-value the better the model fits the data.
- The parameter estimates for the distribution and their standard errors and 95% confidence intervals. The parameter estimates are transformations of the estimated coefficients in the regression table.
- The table percentiles, which includes the estimated percentiles, standard errors and 95% fiducially confidence intervals.
- The probability plot, which helps you to assess whether the chosen distribution fits your data.
- The relative potency- compares the potency of stress for two levels of a factor to get this output; you must have a factor and choose a weibull, lognormal or log logistic distribution.

Figure (4.1)

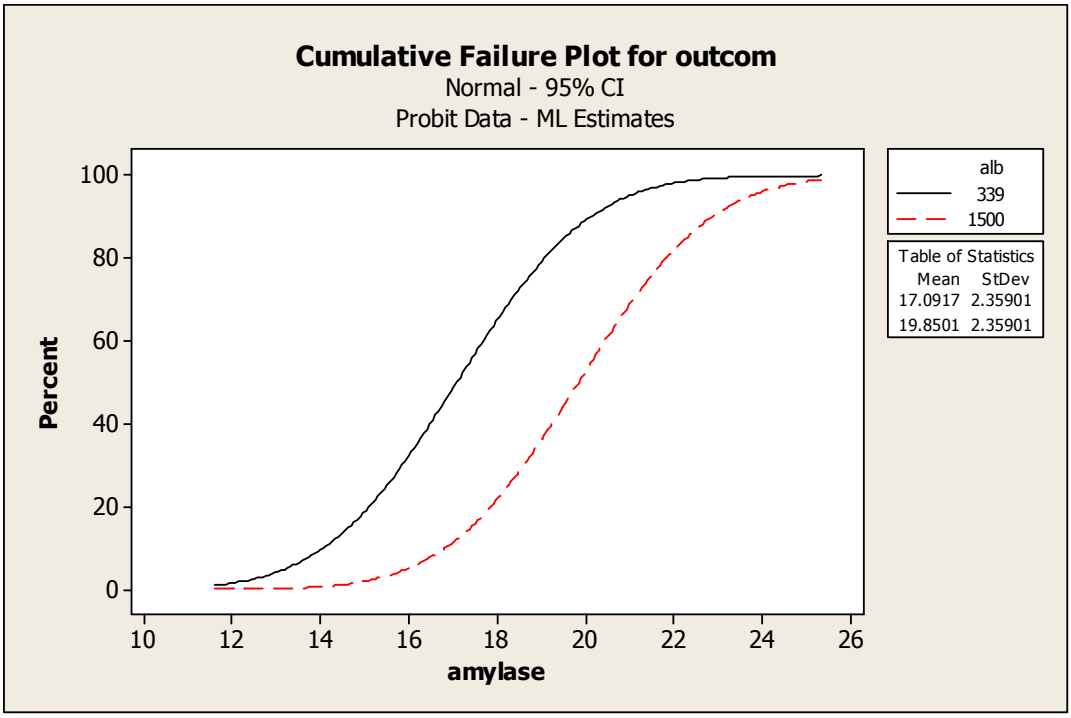


Figure (4.2)

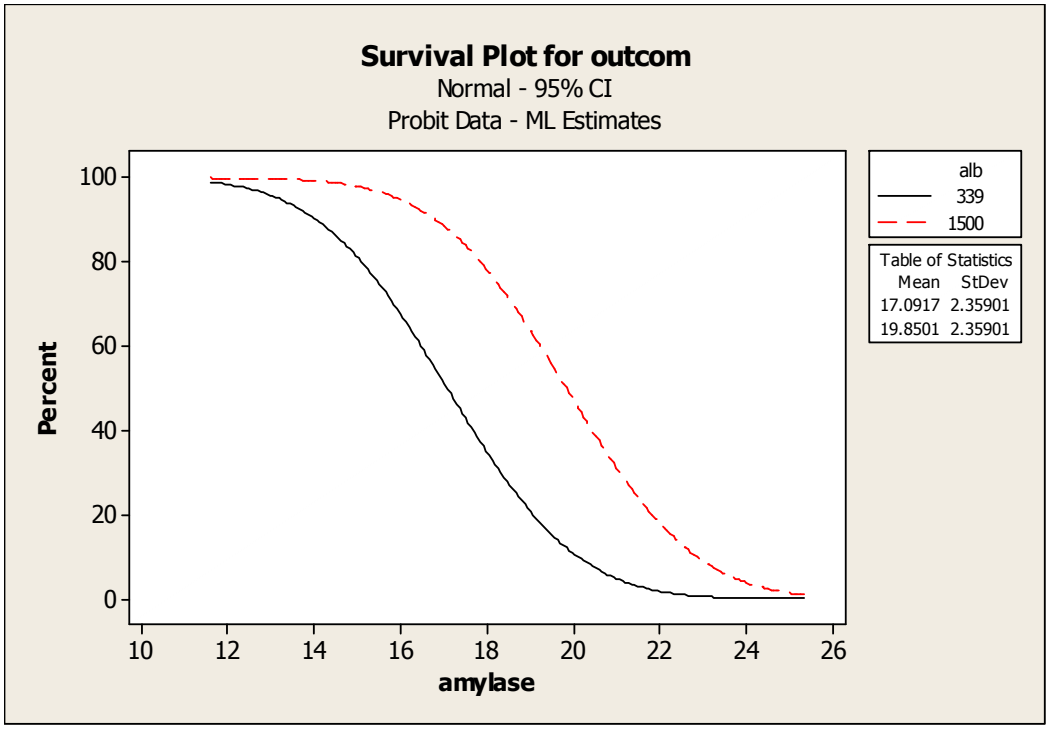
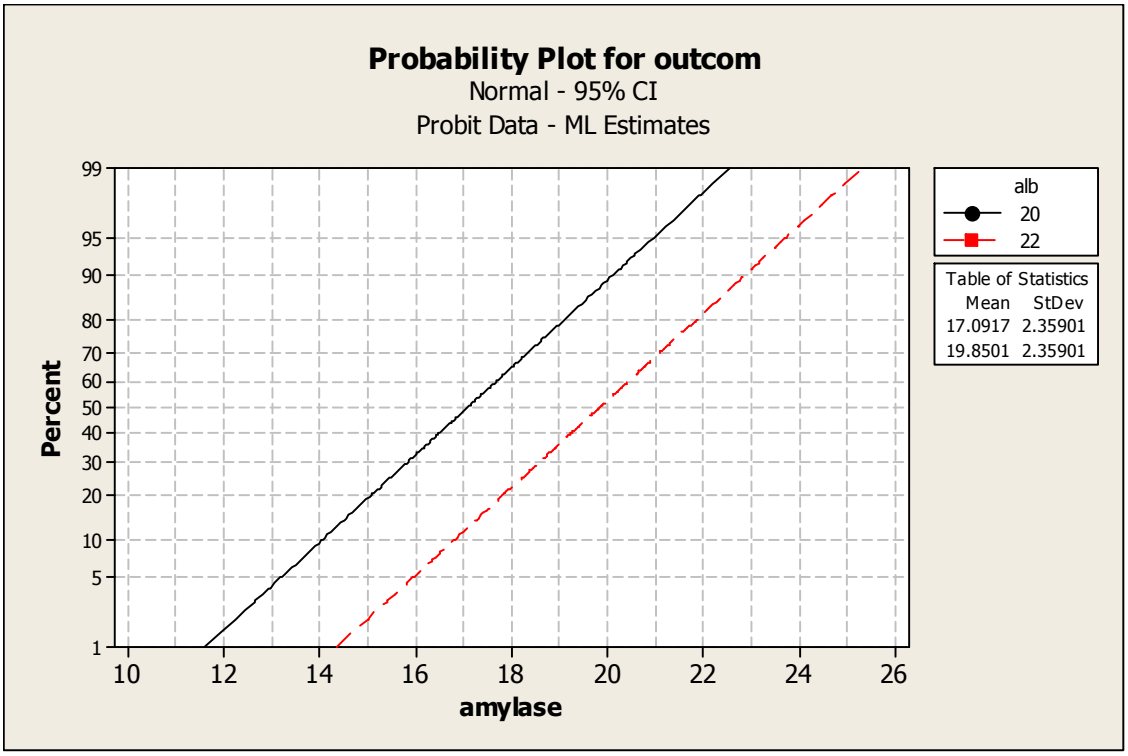


Figure (4.3)


Probability Plot for outcom
Normal - 95% CI
Probit Data - ML Estimates
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
10
12
14
16
18
20
22
24
26
amylase
alb
20
22
Table of Statistics
Mean StDev
17.0917 2.35901
19.8501 2.35901

CHAPTER FIVE
RESULTS AND RECOMMENDATION

5.1 The Study Result:

1. When the person lost hope and deside to put the end of his life and committing suicide the most easy way to take poisons and we receive him in any medical section the most important factor that will determine whether we will help him to leave the system alive the liver function (amylase) which with the analysis in this study we find this variable significantly the most contributing factor.
2. Albumin is significantly explaining the dependent variable so it's very important test we should do for the suicidal as first as he inter the system.
3. Bilirubin the second test we do but in this study it isn't very significant meaning isn't determine the release of the suicidal but if we increase the sample size maybe the result will change because this variable is very important medically and the doctor order to do it when the person suspected drug toxicity.
4. Passing throws the system before or not doesn't affect the survival of the suicidal or not.

5.2 The Recommendation:

This study should not be the last for probit analysis which is huge world itself. The committing suicide increases each year, so I hope this study helps to improve the medical system in our country to improve centers to those people who are suicidal and treats them and admit that they have a problem to solve.

References

Finney, D. J., Ed. (1952). Probit Analysis. Cambridge, England, Cambridge University Press.

Finney, D. J. and W. L. Stevens (1948). "A table for the calculation of working probits and weights in probit analysis." Biometrika **35**(1-2): 191-201.

Greenberg, B. G. (1980). "Chester I. Bliss, 1899-1979." International Statistical Review / Revue Internationale de Statistique **8**(1): 135-136.

Hahn, E. D. and R. Soyer. (date unknown). "Probit and Logit Models: Differences in a Multivariate Realm." Retrieved May 28, 2008, from