LOGISTIC REGRESSION EXAMPLES USING THE SAS SYSTEM, VERSION 6, FIRST pdf

1: Proc Logistic and Logistic Regression Models

Logistic Regression Examples Using the SAS System, Version 6, First Edition www.enganchecubano.com[1/14/14 PM] Logistic Regression Examples Using the SAS System.

In this lesson we focused on Binary Logistic Regression. Below is a brief summary and link to Log-Linear and Probit models. Summary Points for Logistic Regression Cases are independent Does NOT assume a linear relationship between the dependent variable and the independent variables, but it does assume linear relationship between the logit of the explanatory variables and the response. Independent variables can be even the power terms or some other nonlinear transformations of the original independent variables The dependent variable does NOT need to be normally distributed, but it typically assumes a distribution from an exponential family e. When there are continuous predictors, the G2 and X2 are not the best statistics for assessing the overall fit of the model. Usually some grouping of the data is needed. The most commonly, use the Hosmer-Lemeshow statistic, and influence values and plots. As with any other model you can take into consideration sample size and power. For more details see Agresti , Section 5. For a more detailed discussion refer to Agresti , Ch. They are related in a sense that the loglinear models are more general than logit models, and some logit models are equivalent to certain loglinear models e. Y is Binomial Systematic component: The logistic regression model uses the logistic cumulative distribution function cdf. For example, probit 0. Fitted values between these two models are often very similar. Rarely does one of these models fit substantially better or worse than the other, although more difference can be observed with sparse data. Why does this work? Think back to intro statistics classes and approximating binomial distribution with normal. More on probit models see Agresti , Section 3. Analysis of Binary Data. American Statistician, 56, Annals of Statistics, 9, The Many faces of logistic regression.

2: NHANES Tutorials - Module 9 - Logistic Regression

Packed with step-by-step examples, this book shows you how to use the SAS System to perform logistic, probit, and conditional logistic regression analyses. This book enables statisticians, researchers, and new students to learn from the set of examples so that they can perform their own analyses and produce and understand the output.

Many people equate odds with probability and thus equate odds ratios with risk ratios. When the outcome of interest is uncommon i. When the outcome is more common, however, the odds ratio increasingly overstates the risk ratio. So, to avoid confusion, when event rates are high, odds ratios should be converted to risk ratios. N Engl J Med ; Zhang J, Yu KF. A method of correcting the odds ratio in cohort studies of common outcomes. When can odds ratios mislead? Allison By Leon Gordis Setting Up a Logistic Regression in NHANES Simple logistic regression is used for univariate analyses when there is one dependent variable and one independent variable, while multiple logistic regression model contains one dependent variable and multiple independent variables. Correct weight model statement IMPORTANT NOTE Simple logistic regression is used for univariate analyses when there is one dependent variable and one independent variable, while multiple logistic regression model contains one dependent variable and multiple independent variables. Determine the appropriate weight for the data used It is always important to check all the variables in the model, and use the weight of the smallest common denominator. In the example of univariate analysis, the 4-year MEC weight is used, because the hypertension variable is from the MEC examination. In the multivariate analysis example, the 4-year MEC morning subsample weight is used, because the fasting triglycerides variable is from the morning fasting subsample from the lab component, which is the smallest common denominator for all variables in the model. Examples Simple logistic regressions for gender, age, cholesterol, and BMI: Because these analyses use 4 years of data and includes variables that come from the household interview and the MEC e. Simple logistic regression for fasting triglyceride: Because this analysis uses 4 years of data and fasting triglycerides were only done on the morning subsample, the MEC morning fasting subsample 4-year weight - wtsaf4yr is the right one. Because this analysis uses 4 years of data and includes variables from the household interview, MEC and morning subsample of the MEC, the weight for the smallest group - the morning fasting subsample 4 -year weight - wtsaf4yr is the right one. See the Weighting module for more information on weighting and combining weights. You need to use the correct command for the software that you are using. Make sure that you are using the correct commands for the version of software on your computer. Provide a model statement Remember that when you run logistic regression analyses, you must provide a model statement to specify the dependent variable and independent variable s , and you can have only one model statement each time you run a logistic regression analysis. How to Use SUDAAN Code to Perform Logistic Regression In this module, you will use simple logistic regression to analyze NHANES data to assess the association between gender riagendr â€" the exposure or independent variable â€" and the likelihood of having hypertension based on bpxsar, bpxdar â€" the outcome or dependent variable, among participants 20 years old and older. You will then use multiple logistic regression to assess the relationship after controlling for selected covariates. The covariates include gender riagendr , age ridageyr , cholesterol lbxtc , body mass index bmxbmi and fasting triglycerides lbxtr. Create dependent dichotomous variable For continuous variables, you have a choice of using the variable in its original form continuous or changing it into a categorical variable e. The categorical variables should reflect the underlying distribution of the continuous variable and not create categories where there are only a few observations. For the dependent variable, you will create a dichotomous variable, hyper, which defines people as having or not having hypertension. Specifically, a person is said to have hypertension if their systolic blood pressure measured in the MEC exceeds or their diastolic blood pressure exceeds 90 or if they are taking blood pressure medication. Remember for logistic regression to work in SUDAAN, this variable needs to be defined as 0 meaning outcome did not occur, here person does not have hypertension or 1 outcome occurs, here person has hypertension. The code to create this variable is below:

Create independent categorical variables In addition to creating the dependent dichotomous variable, this example will also create additional independent categorical variables age, hichol, bmigrp from the age, cholesterol, and BMI categorical variables to use in this analysis. Transform highly skewed variables Because the triglycerides variable lbxtr is highly skewed, you will use a log transformation to create new variable to use in this analysis. Create eligibility variable Because not every participant in NHANES responded to every question asked, there may be a different level of item non-response to each variable. To ensure that your analyses are done on the same number of respondents, create a variable called eligible which is 1 for individuals who have a non-blank value for each of the variables used in the analyses, and 0 otherwise. Although this is a univariate analysis using only exam variables, the fasting subsample weight wtsaf4yr is included in determining the eligible variable. This is because you will be conducting a multivariate analysis using the triglycerides variable later and will limit the sample to persons included in both analyses. The SAS code defining eligible is: You can read the explanations in the summary table below. You may need to format the variables in your dataset the same way to reproduce results presented in the tutorial. In this example, the MEC weight for four years of data is used. Because only a subpopulation is of interest, use the subpopn statement to select this subgroup. Please note that for accurate estimates, it is preferable to use subpopn in SUDAAN to select a subgroup for analysis, rather than select the study subgroup in the SAS program while preparing the data file. In earlier versions, you need a subgroup and levels statement. Men are less likely to have hypertension than women. Their odds of hypertension are 0. Assuming a p-value less than 0. The Satterthwaite adjusted F gives the most conservative estimate of the test statistics. The p-value of 0. You can follow the steps outlined below to perform a multivariate logistic regression. Not all respondents were tested on triglycerides. Please note that for accurate estimates, it is preferable to use subpopn in SUDAAN to select a subgroup for analysis, rather than select the study subgroup in the SAS program while preparing the dataset. Use a class statement for categorical variables in version 9. Odds ratios should be interpreted as adjusted odds ratios because there are multiple covariates in the model. The adjusted odds of hypertension are 1. How to Use SAS 9. The dependent variable Y is hypertension, and the independent variables Xj, or covariates, are age, gender, high cholesterol, body mass index, and fasting triglycerides. In this task , you will only be reviewing the Multivariate Logistic Procedure. You should not use a where clause or by-group processing in order to analyze a subpopulation with the SAS Survey Procedures. In this example, the sel variable is set to 1 if the sample person is 20 years or older, and 2 if the sample person is younger than 20 years. Then this variable is used in the domain statement to specify the population of interest those 20 years and older. There is a summary table of the SAS program below. Use the nomcar option to read all observations. WEIGHT wtsafyr; se the weight statement to account for the unequal probability of sampling and non-response. In this example, the 4-year fasting weight variable is used. Use the param and ref options to choose your reference group for the categorical variables. The vadjust option specifies whether or not to use variance adjustment. You can compare your results with the sample output, which you can download from the Sample Code and Datasets page. Or, you can view an animated version of the results with narration by clicking the link below. In the narration, the highlighted elements show that: How to Use Stata Code to Perform Logistic Regression In this module, you will use simple logistic regression to analyze NHANES data to assess the association between gender riagendr â€” the exposure or independent variable â€” and the likelihood of having hypertension based on bpxsar, bpxdar â€” the outcome or dependent variable, among participants 20 years old and older. The covariates include age ridageyr , cholesterol lbxtc , body mass index bmxbmi and fasting triglycerides lbxtr. Please see the Stata Tips page to review them before continuing. The general format of this command is below: The vce option specifies the method for calculating the variance and the default is "linearized" which is Taylor linearization. Here is the svyset command for fur years of MEC data: Remember for logistic regression to work in Stata, this variable needs to be defined as 0 meaning outcome did not occur, here person does not have hypertension or 1 outcome occurs, here person has hypertension. Create independent categorical variables In addition to creating the dichotomous dependent variable, this example will also create additional independent categorical variables age, hichol, bmigrp from the age, cholesterol, and BMI categorical variables to use in this analysis. Choose reference groups for categorical variables For all categorical variables, you need to decide which category to use as the reference group. If you do not specify the reference group options, Stata will choose the lowest numbered group by default. You can use the following general command to tell Stata the reference group: Code to specify reference groups Variable Code to specify reference group Reference group.

3: CiteSeerX â€” Citation Query Logistic Regression Examples Using the

Generally speaking, logistic regression is a statistical technique that tries to explain or predict a dichotomous outcome (e.g., two levels, as in yes/no, succeed/fail, heal/don't heal) from a set of independent variables.

SAS will create dummy variables for a categorical variable on-the-fly. There are various coding schemes from which to choose. The default coding for all the categorical variables in proc logistic is the effect coding. There are other coding schemes available, such as orthogonal polynomial coding scheme and reference cell coding. We can double check what coding scheme is used and which group is the reference group by looking at the Class Level Information part of the output. If we want to compare level 2 vs. Usually, contrast is done using less than full rank, reference cell coding as used in proc glm. We also used estimate option at the end of contrast statement to get the estimate of the difference between group 1 and group 2. It is always a good idea to check the Class Level Information to see how the variable is coded so we know that the contrast statement gives us the expected contrast among groups. We can also test the combined effect of multiple parameters using the test statement. In the example below, we first tested on the joint effect of read and math. Next we tested on the hypothesis that the effect of read and math are the same. This test divides subjects into deciles based on predicted probabilities, then computes a chi-square from observed and expected frequencies. It tests the null hypothesis that there is no difference between the observed and predicted values of the response variable. Therefore, when the test is not significant, as in this example, we can not reject the null hypothesis and say that the model fits the data well. We can also request the generalized R-square measure for the model by using rsquare option after the model statement. SAS gives the likelihood-based pseudo R-square measure and its rescaled measure. Koch offers more details on how the generalized R-square measures that you can request are constructed and how to interpret them. Proc logistic can generate a lot of diagnostic measures for detecting outliers and influential data points for a binary outcome variable. These diagnostic measures can be requested by using the output statement. We can then plot these variables against the predicted values to investigate the influence of each point on the model. For example, we may want to know the predicted probabilities for groups defined by female and prog when math and read are held at their grand means. Notice that the score procedure does not care what model we have run. It uses the estimated parameters to generate linear predictions. In our logistic regression case, the predicted values are therefore in the logit scale. In the output data set created by proc score, we have a variable called hiwrite. This is the new variable that proc score created for predicted values. This process will be simplified with SAS 9. The syntax one will use looks like the following: Exact logistic regression provides a way to get around these difficulties. What it does is to enumerate the exact distributions of the parameters of interest, conditional on the remaining parameters. The data set has very small cells, with each cell having only 3 observations. This is the syntax used for grouped data. That is we have frequencies of the events for each of the cells. This type of syntax works for both the maximum likelihood logistic regression and exact logistic regression. Generalized Logits Model for Multinomial Logistic Models Proc logistic also perform analysis on nominal response variables. Since the response variable no longer has the ordering, we can no longer fit a proportional odds model to our data. But we can fit a generalized logits model. This analysis can be done using proc catmod and that is how it is used to be done. We will illustrate what a generalized logits model is and how to perform an analysis using proc logistic. School children in experimental learning settings were surveyed to determine which teaching styles they preferred. The response variable style takes three values: We want to determine the preference of students by their schools and programs. The programs are regular and after-school programs with 1 being regular and 2 being after-school. In a generalized logit model, we will pick a particular category of responses as the baseline reference and compare every other category with the baseline response. In our example, we will choose team as the baseline category. This means that we allow two different sets of regression parameters, one for each logit. We can calculate the generalized odds from the frequency table, similar to what we have done in the case of proportional odds model. That is, for each of the preference choices there are possible six cell counts. If we use both school and program and also include their interaction, we will use up all the degrees of freedom. That is we have a saturated model. This is the best model we can get, fitting each cell with its own parameter. Number of unique profiles: Because our model is saturated, the goodness-of-fit statistics are zero with zero degree of freedom. We also see that the default type of coding scheme, e. We also see that the overall effect of the interaction of school and program is not significant. This leads us to a simpler model with only the main effect. For example the odds ratio of class to team for program1 versus program 2 is. We can say that the odds for students in program 1 to choose class over team is. Or we can say that the odds for students in program 1 to choose class over team is. Or we can say that the odds for students in school 1 to choose class over team is. It is oftentimes easier to describe in terms of probabilities. We can use the output statement to generate these probabilities as shown below. It extends logistic regression to handle ordinal response variables. In this section, we are going to use SAS data set https: Each subject in the data set was asked to evaluate the following statement: The response is recoded in a variable called warm. It has four levels: This will be the response variable in our analysis. Other variables in the data set include age, education level, gender of the subject, and other subject related variables. Thus we allow the intercept to be different for different cumulative logit functions, but the effect of the explanatory variables will be the same across different logit functions. This is the proportionality assumption and this is why this type model is called proportional odds model. Also notice that although this is a model in terms of cumulative odds, we can always recover the probabilities of each response category as follows. The other way of getting the same result is to run a proportional odds model with only the intercept as a predictor. Here clogit stands for cumulative logit. The formula for the odds is shown in the table below. Race Gender SD vs.

4: SSA Logistic Regression Model Using the SAS System

Logistic Regression Examples Using The Sasr System Version 6 First Edition Author Sas Business Solutions Mar Document for Logistic Regression Examples Using The.

For example, subjects are followed over time, are repeatedly treated under different experimental conditions, or are observed in logical units e. For discrete responses, however, we have to face a greater mathematical complexity and statistical analysis is not that straightforward any longer. Reasons for this are: A crucial point in standard logistic regression analysis is that observations are independent of one another and it is known that violations of this assumption result in invalid statistical inference. However, many study designs in applied sciences give rise to correlated data. For example, subjects are followed over time and responses are assessed at different time points, are repeatedly treated under different experimental conditions, or are observed in logical units e. We show that each of them estimates parameters from one of two different statistical models and comment on the interpretation of parameters and the statistical properties of the methods involved. The models and the estimation procedures are illustrated by an example of a multicenter randomized controlled clinical trial. The data have been collected in a multicenter randomized controlled clinical trial conducted in eight different clinics. The purpose of the study was to assess the effect of a topical cream treatment compared to no treatment on curing nonspecific infections. A suitable measure for treatment effect in this case is the odds ratio and the FREQ procedure can be used to achieve an estimator. Some small data manipulation actually writing each person with its response cure in a single line has to precede the analysis the variable status will be needed in a later analysis: However, proceeding this way completely ignores the fact that the study was undertaken in several clinics and we might suspect that different features of them personnel, environment, typical population of patients might influence the treatment effect in the individual clinic. Note that this also implies that there is a correlation of patients within individual clinics. In a clinic with high treatment effect many patients will be cured and a cure in a single patient will be accompanied with a higher probability by a cure of another patient from the same clinic. Because both, fixed effects here: First, we introduce some notation: Thus, this procedure can be used to fit our data set in the context of marginal models and the following call of the GENMOD procedure realizes this: In a marginal model the effect of treatment is modelled separately from the within-clinic correlation. A marginal logistic regression model for our data set is given by: The interpretation of the parameters is analogous to the standard logistic regression model. The transformed regression coefficient exp btreat is the odds for cure for a treated patient divided by the odds for cure in a patient from the control group. However, in this model we adjusted for the correlation between patients from the same clinic and we assumed that this correlation is identical for every two patients from the same clinic. Of course, patients from different clinics are considered to be independent. The interpretation of the parameter does not depend on the respective clinic but rather is valid for the whole population of clinics in the study and actually averages the treatment effect across the clinics. This is why the parameters from marginal models are sometimes called population-averaged parameters. The macro is designed for the analysis of Generalized Linear Mixed Models GLMM , and as our random effects logistic regression model is a special case of that model it fits our needs. An overview about the macro and the theory behind is given in Chapter 11 of Littell et al. Briefly, the estimating algorithm uses the principle of quasi-likelihood and an approximation to the likelihood function of the model that results in an iterative procedure repeatedly fitting a linear mixed model to a pseudo response. In a random effects model it is assumed that there is natural heterogeneity across the clinics and that this heterogeneity can be modelled by a probability distribution in our case the normal distribution which means that the regression coefficients vary from one clinic to another. By using this approach the correlation between patients from the same clinic arises from their sharing specific but unobserved properties of the respective clinic. A random effects logistic regression model more specifically: We assume further that, given ui, the responses from the same clinic are mutually independent, that is the correlation between patients from the same clinic is completely explained by them having been treated in the same clinic. By including only the random intercept ui but keeping the treatment effect fixed we assume that there is a single individual cure probability in each clinic but the effect of treatment is identical across the clinics. This single individual cure probability is the reason why parameters from random effects models are also called subject-specific parameters The interpretation of parameters is also analogous to the standard logistic regression model. The transformed regression coefficient exp btreat is the odds for cure for a treated patient compared to a control group patient. The first two estimation methods are based on a Taylor series expansion around values for the random effect parameters where the first method expands around the zero vector and the second around the best empirical linear unbiased predictor EBLUP of the random effects. Both approximations result in algorithms that iterate linear mixed models for suitably defined pseudo responses. The third method, finally, is a refined GEE estimation, that is, it actually fits a marginal model. To be concrete, it extends the fitting algorithm used in the GENMOD procedure by using a quadratic instead of a simple method of moment estimation equation for the correlation parameters Davidian, The following SAS code fits the described models: The macro code is rather technical, and the instructions how to define the correct model are given in the header of the macro. This is necessary because the likelihood of the model incorporates untractable integrals that can be calculated directly only in special cases or with considerable effort. It attempts to maximize the likelihood directly by numerical integration methods, more precisely by adaptive Gaussian quadrature. At least theoretically, it delivers exact maximum likelihood ML estimates of the parameters if the number of quadrature points is large enough. In the first step the treatment effect here the log odds ratio, logor is estimated for each single study and in the second step an overall treatment estimator is calculated as a weighted average of the study estimates where the weights are the inverse of the estimated variances varlogor of the single study treatment effects. To fit the model with the MIXED procedure we consider the estimated study treatment effects as the response and include no further covariates. If we want to estimate a random effects meta-analysis model we have to include a RANDOM-Statement to declare the study effect as a random effect and an additional parameter in the PARMS-Statement which estimates the variance of the treament effect between studies. Please note that we added 0. It uses conditional maximum likelihood estimation, that is, it maximizes the conditional likelihood function of the model, given the sufficient statistics of the parameters. In fact, we trick the PHREG procedure by considering a degenerated survival model with only two survival times status and the actual response cure as the censoring indicator, that is to say we define a patient to be censored if he was not cured. Estimation of the overall treatment; An advantage of this method is that it completely removes the random intercept from the likelihood equation and thus does not rely on the assumption of normality of the random effects distribution and this is also why we get no estimate for the random intercept. Here we merely report the syntax and do not go too much into the details of exact conditional analysis, a good explanation of the model is given by Derr, As such, we can also use the traditional meta-analytic methods for analysis. In a reply to this paper, Stijnen, , demonstrates how the standard meta-analytic random effect model can be estimated by the MIXED procedure. In the USA and other countries. The following table compares the results of the different procedures for the infection data set. It can be shown that this is necessarily the case and one can even calculate a multiplicative factor that converts the estimates from the marginal to the random effect model Zeger, Liang, Albert, The Meta-Analysis estimates from the fixed and the random effects model are identical. That means that the random effect model considers the between-clinic heterogeneity as statistically nonsignificant and automatically sets it to zero. It is difficult to give definite general recommendations which of the methods to use because this depends on the data at hand and on the desired interpretation of parameters population-averaged vs.

5: References :: SAS/STAT(R) User's Guide

Logistic Regression Examples Using the SAS(R) System, Version 6, First Edition Packed with step-by-s moreÂ» tep examples, this book shows you how to use the SAS System to perform logistic, probit, and conditional logistic regression analyses.

Computing Corner Logistic Regression Model Using the SAS System Logistic Regression is commonly used to predict the probability that a unit under analysis will acquire the event of interest as a linear function of changes in values of one or more continuous-level variables, dichotomous binary variables , or a combination of both continuous and binary independent variables. The dependent variable is dichotomous and is coded either zero event did not occur or one event did occur. The logistic function is used to estimate, as a function of unit changes in the independent variable s the probability that the event of interest will occur. Logistic regression, when properly used, develops a model which attempts to predict the probability of an event of interest occurring in the population from which the data under analysis are assumed to have been randomly sampled. The SAS routine is as follows: The independent variable is the age of patient in years AGE. A UNITS option can also be used when single unit changes in the values of the independent variable may not be substantively relevant to the analysis at hand. The impact of changes or more than one unit in the independent variable can be obtained by using the UNITS option, which is available in Version 6. An example of this option is as follows: This option provides odds ratios for 5, 10 and 20 increments in patient age. In other words, this procedure will work in version 6. The simplest program was on data taken from Cox and Snell , pp , consisting of the number of ingots not ready for rolling R out of N tested, for a number of combinations of heating time and soaking time. The SAS program is below. The SAS output is shown in Table 1. With these options, you can compute confidence limits from regression parameters and odds ratios. Estimated odds ratios are computed by exponentiating the parameter estimates for a logistic regression model when the following conditions are met: Similarly, confidence limits for odds ratios are computed by exponentiating the confidence limits for the logistic regression parameters. There are two available methods of computing confidence limits for logistic regression parameters: The likelihood ratio model is an iterative process based on the profile likelihood function. The Wald method is a simpler method based on the asymptotic normality of the parameter estimator. These two methods should produce approximately the same results for large samples, but may produce different results for small samples. When the parameter estimate is very large, however, these two methods may produce different results even for large sample sizes. It also shows how to use an option to adjust the confidence coefficient for the confidence limits. First, you must create your SAS data set. Combine the chemical diabetics and overt diabetics into one group-the event group. The normals are the nonevent group. You set up this SAS data set as follows: The explanation of this output is as follows: The confidence limits are labeled Profile Likelihood Confidence Limits. The construction of these confidence intervals is derived from the asymptotic chi-square distribution of the likelihood ratio test. The confidence limits are labeled Wald Confidence Limits. Wald confidence limits are computed by assuming a normal distribution for each parameter estimator. This computation method is less time consuming than the one based on the profile likelihood function because it does not involve an iterative process. However, it is considered to be less accurate, especially for small sample sizes. When you examine the confidence intervals for the parameter estimates, you can see that the Wald confidence intervals are symmetric about the point estimate, but the profile likelihood confidence intervals are asymmetrical. This is because the upper and lower profile likelihood confidence limits are computed separately using an iterative process, and the distribution of a parameter estimate is not symmetric for small sample sizes. Profile likelihood confidence limits for odds ratios are a transformation of the confidence limits that you can produce with the PLCL option for the corresponding regression parameters. It requests confidence intervals for the odds ratios of all explanatory variables. Computation of these confidence intervals is based on the asymptotic normality of the parameter estimators.

LOGISTIC REGRESSION EXAMPLES USING THE SAS SYSTEM, VERSION 6, FIRST pdf

SAS Program and information from example 2 pp.

LOGISTIC REGRESSION EXAMPLES USING THE SAS SYSTEM, VERSION 6, FIRST pdf

6: - Summary Points for Logistic Regression | STAT

In the logit model the log odds of the outcome is modeled as a linear combination of the predictor variables. The purpose of this page is to show how to use various data analysis commands. It does not cover all aspects of the research process which researchers are expected to do. In particular, it does not cover data cleaning and checking, verification of assumptions, model diagnostics and potential follow-up analyses. Suppose that we are interested in the factors that influence whether a political candidate wins an election. The predictor variables of interest are the amount of money spent on the campaign, the amount of time spent campaigning negatively, and whether the candidate is an incumbent. A researcher is interested in how variables, such as GRE Graduate Record Exam scores , GPA grade point average and prestige of the undergraduate institution, effect admission into graduate school. Description of the data For our data analysis below, we are going to expand on Example 2 about getting into graduate school. We have generated hypothetical data, which can be obtained from our website by clicking on https: You can store this anywhere you like, but the syntax below assumes it has been stored in the directory c: This data set has a binary response outcome, dependent variable called admit, which is equal to 1 if the individual was admitted to graduate school, and 0 otherwise. There are three predictor variables: We will treat the variables gre and gpa as continuous. The variable rank takes on the values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest. We start out by looking at some descriptive statistics. Some of the methods listed are quite reasonable while others have either fallen out of favor or have limitations. Logistic regression, the focus of this page. Probit analysis will produce results similar tologistic regression. The choice of probit versus logit depends largely onindividual preferences. When used with a binary response variable, this model is knownas a linear probability model and can be used as a way todescribe conditional probabilities. However, the errors i. Fora more thorough discussion of these and other problems with the linearprobability model, see Long , p. Two-group discriminant function analysis. A multivariate method for dichotomous outcome variables. To model 1s rather than 0s, we use the descending option. The class statement tells SAS that rank is a categorical variable. For more information on dummy versus effects coding in proc logistic, see our FAQ page: The first part of the above output tells us the file being analyzed c: We see that all observations in our data set were used in the analysis fewer observations would have been used if any of our variables had missing values. If we omitted the descending option, SAS would model admit being 0 and our results would be completely reversed. The -2 Log L In the next section of output, the likelihood ratio chi-square of The Score and Wald tests are asymptotically equivalent tests of the same hypothesis tested by the likelihood ratio test, not surprisingly, these tests also indicate that the model is statistically significant. The section labeled Type 3 Analysis of Effects, shows the hypothesis tests for each of the variables in the model individually. The chi-square test statistics and associated p-values shown in the table indicate that each of the three variables in the model significantly improve the model fit. For gre and gpa, this test duplicates the test of the coefficients shown below. However, for class variables e. The logistic regression coefficients give the change in the log odds of the outcome for a one unit increase in the predictor variable. For every one unit change in gre, the log odds of admission versus non-admission increases by 0. For a one unit increase in gpa, the log odds of being admitted to graduate school increases by 0. The coefficients for the categories of rank have a slightly different interpretation. For example, having attended an undergraduate institution with a rank of 1, versus an institution with a rank of 4, increases the log odds of admission by 1. An odds ratio is the exponentiated coefficient, and can be interpreted as the multiplicative change in the odds for a one unit change in the predictor variable. For example, for a one unit increase in gpa, the odds of being admitted to graduate school versus not being admitted increase by a factor of 2. For more information on interpreting odds ratios see our FAQ page: How do I interpret odds ratios in logistic regression? We can also test for differences between the other levels of rank. We can test this type of hypothesis by adding a contrast statement to the code for proc logistic. The syntax shown below is the same as that shown above, except that it includes a contrast statement. Following the word contrast, is the label that will appear in the output, enclosed in single quotes i. This is followed by the name of the variable we wish to test hypotheses about i. After the slash i. For more information on use of the contrast statement, see our FAQ page: How can I create contrasts with proc logistic? The only difference is the additional output produced by the contrast statement. Under the heading Contrast Test Results we see the label for the contrast rank 2 versus 3 along with its degrees of freedom, Wald chi-square statistic, and p-value. We can see that the estimated difference was 0. You can also use predicted probabilities to help you understand the model. In the syntax below we use multiple contrast statements to estimate the predicted probability of admission as gre changes from to in increments of When estimating the predicted probabilities we hold gpa constant at 3. The term intercept followed by a 1 indicates that the intercept for the model is to be included in estimate. The predicted probabilities are included in the column labeled Estimate in the second table shown above. Looking at the estimates, we can see that the predicted probability of being admitted is only 0. Things to consider Empty cells or small cells: You should check for empty or smallcells by doing a crosstab between categorical predictors and the outcome variable. If a cell has very few cases a small cell , the model may become unstable or it might not run at all. Separation or quasi-separation also called perfect prediction: A condition in which the outcome does not vary at some levels of the independent variables. See our page FAQ: Both logit and probit models require more cases than OLS regression because they use maximum likelihood estimation techniques. It is sometimes possible to estimate models for binary outcomes in datasets with only a small number of cases using exact logistic regression available with the exact option in proc logistic. For more information see our data analysis example for exact logistic regression. It is also important to keep in mind that when the outcome is rare, even if the overall dataset is large, it can be difficult to estimate a logit model. Many different measures of psuedo-R-squared exist. They all attempt to provide information similar to that provided by R-squared in OLS regression; however, none of them can be interpreted exactly as R-squared in OLS regression is interpreted. The diagnostics for logistic regression are different from those for OLS regression. For a discussion of model diagnostics for logistic regression, see Hosmer and Lemeshow , Chapter 5. Note that diagnostics done for logistic regression are similar to those done for probit regression. Applied Logistic Regression Second Edition. John Wiley and Sons, Inc.

7: Logistic Regression Using the SAS System: Theory and Application

SAS Institute (). Logistic Regression Examples Using the SAS System, Version 6. Strauss, David (). The Many faces of logistic regression.

Walker, Statistics Branch, John J. Walker , " This paper reports the results of research and analysis by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This paper is released to inform interested parties of research and encourage the discussion. Those Who Left Versus T This paper compares individuals who received FS over the period â€" with those who left the FS rolls. To investigate and understand the relationship concerning continued participation in the FS program data from waves one, four, seven, and ten of the Survey of Income and Program Participation SIPP are used to derive descriptive statistics and a logit regression model. Schinus terebinthifolius, known as Brazilian pepper, is an exotic, invasive plant species in Florida that displaces native plant species and disrupts wildlife habitat. Aerial surveys typically used to monitor ecosystem change may be aug-mented with texture analyses to improve the speed and consisten Aerial surveys typically used to monitor ecosystem change may be aug-mented with texture analyses to improve the speed and consistency with which S. Image processing using high-resolution imagery can take advantage of high spectral variability in adjacent pixels of the same cover type by measuring spatial patterns of tex-ture in neighborhoods of pixels. Texture features derived from first and second-order sta-tistics and edge components in high-resolution digital color infrared images were tested for their ability to discriminate S. Multiple linear logistic regressions found a best subset combination of texture features that consistently identified core areas of S. Misclassification of other cover types as S. Germain A, Eddie Bevilacqua A , " This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues. Other uses, including reproduction and distribution, or sel Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited. In most cases authors are permitted to post their version of the article e. Pearlstine , "

8: How to Use SAS for Logistic Regression with - www.enganchecubano.com

Informal and nontechnical, this book both explains the theory behind logistic regression and looks at all the practical details involved in its implementation using the SAS System. Several social science real-world examples are included in full detail.

9: SAS Documentation Example Code and Data

These data sets were used in the examples of multinomial logistic regression modeling techniques. Statistical analysis was conducted using the SAS System for Windows.

LOGISTIC REGRESSION EXAMPLES USING THE SAS SYSTEM, VERSION 6, FIRST pdf

Progress in Transplantation Bc kuo 9th edition Art as style/style as art, and the problem with that Biological networks : rainforests, coral reefs, and the Galapagos Islands Sonia Kleindorfer and James G. French Music, Culture, and National Identity, 1870-1939 (Eastman Studies in Music) International laws of war Child abuse psychology journal article Bound in shallows Mixed Plate And Noodles Dnd dungeon campaigns Nora roberts born in fire Black Bear Cub (Read and Discover) Life, E-Book, MCAT Full Length Practice Test Enterprise cloud strategy 2nd edition Asthma Diana Grootendorst Hydrology and water chemistry of shallow aquifers along the upper Clark Fork, western Montana His eye on the sparrow : teaching and learning in an African American church Wendy L. Haight and Janet Ca The Didache and Justins first apology Sas management console 9.4 user guide Instructors resource guide to accompany Fit and well Class social studies book Skills in counseling women The Rise of the Rappites Cleft Lip and or Palate Sayings and doings. Commercialism in schools The little mailman of Bayberry Lane Cultural anthropology 2nd edition Imperial Guptas and their times Horsemen on the hills. Aboard the Ship Great Republic to New Orleans Blunder zachary shore Last stage to Sula Honda aero 50 manual Part one : The inward disciplines. The enchanted umbrella Genius and Eminence (International Series in Social Psychology) Blown Away (Hardy Boys (All New Undercover Brothers) Vlad Dracula the Impaler A Technical Manual for Church Planters