regression models for categorical dependent variables using stata

Regression Models for Categorical Dependent Variables Using Stata

regression models for categorical dependent variables using stata are an essential tool for researchers and data analysts working with outcome variables that are not continuous. Whether your dependent variable is binary, ordinal, or nominal, understanding how to appropriately model it in Stata can unlock valuable insights from your data. This article will guide you through the essentials of using regression models tailored for categorical outcomes in Stata, showcasing the types of models available, how to implement them, and tips for interpreting results effectively.

Understanding Categorical Dependent Variables and Their Challenges

When working with statistical data, not all dependent variables are numeric or continuous. Categorical dependent variables represent outcomes divided into distinct groups or categories. For example, a survey response might be “yes” or “no” (binary), a rating might be “poor,” “average,” or “excellent” (ordinal), or a choice among multiple brands (nominal). Traditional linear regression models, such as ordinary least squares (OLS), are ill-suited for these types of data because they assume continuous outcomes and constant variance, leading to biased or nonsensical predictions.

This is where specialized regression models for categorical dependent variables come in. Stata, a popular statistical software, provides a suite of commands designed to handle various types of categorical outcomes, allowing you to model relationships between predictors and categorical responses accurately.

Types of Regression Models for Categorical Dependent Variables in Stata

Choosing the right regression model depends primarily on the nature of your dependent variable. Here, we’ll explore the most commonly used models in Stata for different categorical outcomes.

1. Logistic Regression for Binary Outcomes

If your dependent variable has only two possible categories (e.g., success/failure, yes/no), logistic regression is generally the go-to method. In Stata, you can fit a logistic regression model using the command:

```stata
logit dependent_variable independent_variables
```

or the equivalent

```stata
logistic dependent_variable independent_variables
```

The logistic model estimates the log-odds of the event occurring as a linear function of the predictors. This model is powerful for estimating probabilities and odds ratios and is widely used in fields like epidemiology, social sciences, and marketing.

2. Multinomial Logistic Regression for Nominal Outcomes

When your dependent variable has more than two unordered categories (nominal), such as types of transportation (car, bike, bus), multinomial logistic regression is appropriate. Stata implements this through the `mlogit` command:

```stata
mlogit dependent_variable independent_variables
```

This model compares each category to a baseline category, estimating relative risk ratios for predictors. Interpreting these results requires understanding that coefficients reflect the effect on the relative risk of belonging to a specific category compared to the reference.

3. Ordinal Logistic Regression for Ordered Categories

If your categorical dependent variable has a natural order but unknown spacing between categories (e.g., satisfaction ratings), ordinal logistic regression is an excellent choice. Stata’s `ologit` command fits this model:

```stata
```

```
ologit dependent_variable independent_variables
```

The ordinal logistic model assumes proportional odds, meaning the relationship between predictors and the odds of being in a higher category is constant across thresholds. Checking this assumption is crucial to ensure model validity.

4. Probit Models as Alternatives

Probit regression is similar to logistic regression but assumes a normal distribution of the error term. Stata supports probit models via the `probit` command for binary outcomes and `oprobit` for ordinal outcomes:

```stata
probit dependent_variable independent_variables
oprobit dependent_variable independent_variables
```

Choosing between logit and probit often depends on theoretical considerations or convention, as their results are generally similar.

Implementing Regression Models for Categorical Dependent Variables Using Stata: A Step-by-Step Guide

Navigating Stata’s commands for categorical regression can seem daunting at first. Below is a general workflow to help you approach such analyses confidently.

Step 1: Prepare and Explore Your Data

Before fitting models, ensure your categorical dependent variable and predictors are correctly coded. Use commands like `tabulate` and `summarize` to inspect your data distributions. For example:

```stata
tabulate dependent_variable
summarize independent_variable1 independent_variable2
```

Check for missing values and consider recoding variables if necessary.

Step 2: Choose the Appropriate Model

Based on the structure of your dependent variable, select the right regression type. For example, use `logit` for binary outcomes or `mlogit` for nominal categories.

Step 3: Fit the Model

Run the regression model using the appropriate Stata command. For instance, a logistic regression might look like:

```stata
logit outcome_var predictor1 predictor2
```

Stata will provide coefficients, standard errors, z-values, and p-values.

Step 4: Interpret the Results

Interpreting coefficients depends on the model type. For logistic regression, exponentiate coefficients to obtain odds ratios:

```stata
logistic outcome_var predictor1 predictor2
```

or use:

```stata
estat or
```

to display odds ratios after fitting a `logit` model.

For multinomial logistic regression, interpret relative risk ratios similarly by exponentiating coefficients.

Step 5: Evaluate Model Fit and Assumptions

Assessing model adequacy is vital. Use goodness-of-fit tests, pseudo R-squared values, and classification tables where available. For ordinal logistic regression, test the proportional odds assumption using the Brant test (via the `brant` command after installing the package):

```stata
```

```
brant
```

If assumptions are violated, consider alternative modeling approaches.

Advanced Tips for Working with Categorical Regression Models in Stata

Handling Interaction Terms and Nonlinear Effects

Stata allows inclusion of interaction terms to explore how the effect of one predictor varies with another. Use the `#` operator to specify interactions, for example:

```stata
logit outcome_var c.age##i.gender
```

This models the interaction between continuous age and categorical gender. Additionally, consider nonlinear effects by including polynomial terms or splines if theory or data suggest nonlinearity.

Post-Estimation Commands for Deeper Insights

Stata’s post-estimation tools are invaluable. For example, use `margins` to compute predicted probabilities or marginal effects, which often provide more intuitive interpretations than raw coefficients:

```stata
margins, at(predictor1=(value1 value2))
marginsplot
```

Plotting predicted probabilities across values of a predictor can visually communicate your model’s findings effectively.

Dealing with Multicollinearity and Model Selection

Multicollinearity among predictors can inflate standard errors and obscure true effects. Use the `vif` command after fitting linear models (though not directly after logistic models) or examine correlation matrices before modeling. Stepwise selection procedures (`stepwise` command) can help

identify parsimonious models but use with caution to avoid overfitting.

Practical Example: Modeling Employment Status with Multinomial Logistic Regression

Suppose you have survey data on employment status with three categories: employed, unemployed, and not in labor force. You want to understand how education level and age influence employment status.

Load your data and run:

```stata
mlogit employment_status education age
```

Stata treats one category as the base (default is the first alphabetically unless specified). To change the base category:

```stata
mlogit employment_status education age, baseoutcome(unemployed)
```

After fitting, interpret coefficients as the effect of predictors on the relative risk of being employed or not in labor force compared to unemployed.

Use `margins` to estimate predicted probabilities:

```stata
margins education
marginsplot
```

This visualization can highlight how employment probabilities change with education level.

Final Thoughts on Harnessing Stata for Categorical Outcome Modeling

Regression models for categorical dependent variables using Stata open up a realm of analytical possibilities beyond traditional linear regression. By selecting the right model type—be it logistic, multinomial, or ordinal—researchers can rigorously explore relationships involving categorical outcomes. Stata’s rich command set, combined with intuitive post-estimation tools, makes it a powerful platform for these analyses.

As you delve into modeling categorical data, remember that understanding the nature of your dependent variable is key. Always complement your statistical modeling with careful data exploration, thoughtful interpretation, and validation checks. With these practices, you can confidently leverage regression models for categorical dependent variables using Stata to extract meaningful insights from complex datasets.

Frequently Asked Questions

What types of regression models can be used for categorical dependent variables in Stata?

In Stata, common regression models for categorical dependent variables include logistic regression (logit) for binary outcomes, multinomial logistic regression (mlogit) for nominal variables with more than two categories, and ordered logistic/probit regression (ologit, oprobit) for ordinal dependent variables.

How do I run a binary logistic regression in Stata for a categorical dependent variable?

You can run a binary logistic regression using the command: logistic depvar indepvars, where depvar is the binary categorical dependent variable and indepvars are the independent variables. For example: logistic outcome age gender.

What command is used for multinomial logistic regression in Stata?

The command to run multinomial logistic regression in Stata is mlogit. The syntax is: mlogit depvar indepvars. This is used when the dependent variable is nominal with more than two categories.

How do I interpret the output of an ordered logistic regression in Stata?

Ordered logistic regression output in Stata provides coefficients (log-odds), standard errors, z-values, and p-values. Positive coefficients indicate higher likelihood of being in a higher category of the ordinal dependent variable. You can also compute odds ratios with the command: ologit depvar indepvars, or use the option or after the command to get odds ratios.

Can Stata handle multinomial probit models for

categorical dependent variables?

Yes, Stata can estimate multinomial probit models using the command mprobit. However, it is computationally intensive and may require more time and memory compared to multinomial logit models.

How can I check model fit for regression models with categorical dependent variables in Stata?

Model fit can be assessed using likelihood ratio tests, pseudo R-squared values reported in the output, and information criteria such as AIC and BIC. Additionally, post-estimation commands like estat classification or lrtest can be used to evaluate fit.

What post-estimation commands are useful after running logistic regression in Stata?

Useful post-estimation commands include margins for predicted probabilities, estat classification for classification tables, and lroc for the ROC curve. For example, after logistic regression, using margins can help interpret the effect of predictors on the probability of the outcome.

How do I include factor variables in regression models for categorical dependent variables in Stata?

In Stata, you can include factor variables using i.variable notation. For example, logistic depvar i.gender i.education includes gender and education as categorical predictors and automatically creates dummy variables within the model.

Is it possible to run mixed-effects logistic regression models in Stata for categorical dependent variables?

Yes, Stata supports mixed-effects logistic regression models using the melogit command for binary outcomes and meologit for ordinal outcomes. These models account for clustering or hierarchical data structures.

How do I handle imbalanced categories in categorical dependent variables when running regression models in Stata?

To address imbalanced categories, you can use options like the fweight or pweight to apply sampling weights, or use techniques such as oversampling the minority class outside Stata. Additionally, consider using robust standard errors or alternative modeling approaches like penalized regression if imbalance affects model performance.

Additional Resources

Regression Models for Categorical Dependent Variables Using Stata: An In-Depth Review

regression models for categorical dependent variables using stata represent a critical toolset for researchers and analysts dealing with non-continuous outcome variables. In empirical research, dependent variables often take on categories rather than numeric values, such as binary outcomes (yes/no), nominal groups (types), or ordered levels (ratings). Stata, a widely used statistical software, offers comprehensive capabilities to model such categorical data accurately, thereby enabling nuanced interpretations and robust inferences. This article explores the methodologies, applications, strengths, and limitations of regression models tailored for categorical dependent variables within the Stata environment.

Understanding the Need for Specialized Regression Models

Traditional linear regression assumes a continuous dependent variable and normally distributed errors. When the outcome variable is categorical, these assumptions break down, making ordinary least squares (OLS) inappropriate. For instance, predicting whether a patient has a disease (yes/no) or categorizing voter preference (party A, party B, or independent) demands models that respect the discrete nature of the dependent variable. Using OLS in these contexts can lead to biased estimates, predicted probabilities outside the [0,1] range, and misleading conclusions.

Stata addresses these challenges by providing a suite of regression models designed explicitly for categorical dependent variables. These include logistic regression for binary outcomes, multinomial logistic regression for nominal categories, and ordered logistic or probit regression for ordinal responses. By leveraging maximum likelihood estimation and link functions suited to discrete outcomes, Stata’s models ensure valid probability predictions and interpretable coefficients.

Key Regression Models for Categorical Dependent Variables in Stata

1. Binary Logistic Regression

Binary logistic regression is the most common technique for modeling dichotomous dependent variables. In Stata, the command `logit` or `logistic`

facilitates this analysis, estimating the log-odds of the event occurring as a function of explanatory variables.

- **Features:** Models the probability of a binary outcome; outputs odds ratios; handles continuous and categorical predictors.
- **Use Cases:** Medical diagnosis (disease/no disease), marketing response (purchase/no purchase), employment status (employed/unemployed).
- **Example:** `logit disease_age disease_sex` estimates how age and sex influence disease presence.

While binary logistic regression is robust, it assumes linearity in the log-odds, independence of observations, and absence of multicollinearity among predictors.

2. Multinomial Logistic Regression

When the dependent variable has more than two nominal categories without intrinsic order, multinomial logistic regression is appropriate. Stata implements this via the `mlogit` command.

- **Features:** Extends binary logistic regression to multiple categories; estimates relative risk ratios compared to a base category.
- **Use Cases:** Modeling choice behavior among multiple brands, modes of transportation, or political party affiliation.
- **Example:** `mlogit transport_mode age income` models how age and income affect the choice among car, bus, or bike.

A notable consideration is model complexity: as the number of categories and predictors increases, interpretation and computation become more demanding.

3. Ordinal Logistic and Probit Regression

When the dependent variable is categorical with a natural order but unknown spacing (e.g., satisfaction ratings from “poor” to “excellent”), ordered logistic (`ologit`) or ordered probit (`oprobit`) regression are preferred.

- **Features:** Accounts for ordered nature; estimates cumulative

probabilities; provides threshold parameters alongside regression coefficients.

- **Use Cases:** Customer satisfaction surveys, educational attainment levels, pain severity scales.
- **Example:** `ologit satisfaction age income` assesses how demographics relate to satisfaction levels.

These models assume proportional odds or parallel regression lines, meaning the effect of predictors is consistent across outcome thresholds—a condition that should be tested using Stata's `brant` test.

Implementing Regression Models for Categorical Outcomes in Stata

Stata's syntax is user-friendly, and its comprehensive documentation supports applied researchers. A typical workflow involves:

1. **Data Preparation:** Ensuring the dependent variable is correctly coded as categorical, handling missing values, and creating dummy variables if necessary.
2. **Model Selection:** Choosing the appropriate regression model based on the nature of the dependent variable (binary, nominal, ordinal).
3. **Estimation:** Running the model using commands like `logit`, `mlogit`, or `ologit`.
4. **Diagnostics:** Assessing model fit through likelihood ratio tests, pseudo R-squared values, and testing assumptions (e.g., proportional odds).
5. **Interpretation:** Extracting odds ratios, relative risk ratios, or marginal effects using post-estimation commands such as `margins`.

For example, to model employment status (employed/unemployed) using age and education, a binary logistic regression might be specified as:

```stata
logit employed age education
margins, at(age=(20(10)60)) vsquish
marginsplot
```

This sequence estimates the model and visualizes how predicted probabilities vary with age.

Advantages of Using Stata for Categorical Regression

Stata’s integration of regression models for categorical dependent variables offers several benefits:

- **Comprehensive Model Suite:** From simple binary logistic to complex multinomial and ordered models, Stata covers diverse analytical needs.
- **Robust Post-Estimation Tools:** Commands like `margins` and `predict` enable nuanced interpretation and visualization of predicted probabilities.
- **Diagnostic Testing:** Tools for checking assumptions (e.g., proportional odds), multicollinearity, and goodness-of-fit enhance model reliability.
- **User Community and Documentation:** Extensive manuals, online forums, and tutorials facilitate learning and troubleshooting.

Potential Challenges and Considerations

Despite its strengths, users must be mindful of certain limitations when modeling categorical dependent variables in Stata:

- **Sample Size Requirements:** More complex models, especially multinomial logistic regression with many categories, require larger datasets to avoid overfitting.
- **Interpretation Complexity:** Relative risk ratios and odds ratios can be unintuitive, necessitating careful explanation and use of marginal effects for clarity.
- **Assumption Testing:** Violations of proportional odds or independence assumptions can bias results; these require diagnostic checks and possibly alternative modeling strategies.
- **Computational Intensity:** High-dimensional categorical models may demand significant computational resources, potentially limiting feasibility in some environments.

Comparing Stata with Other Statistical Software for Categorical Regression

While Stata is lauded for its user-friendly interface and powerful commands, alternatives like R, SAS, and SPSS also provide regression models for categorical dependent variables. Compared to R, Stata offers a more streamlined experience with less coding overhead, though R's flexibility and open-source nature appeal to advanced users. SAS is preferred in large-scale corporate environments for its enterprise features, while SPSS is often favored by social scientists for its graphical interface.

In terms of model capabilities, Stata consistently ranks among the top for its balance of power, ease of use, and extensive post-estimation tools. This makes it especially attractive for applied researchers who require rigorous yet accessible categorical data analysis.

Emerging Trends and Extensions

Recent developments in categorical regression modeling within Stata include:

- **Generalized Structural Equation Modeling (GSEM):** Expands regression models to include latent variables and complex path structures involving categorical outcomes.
- **Multilevel and Mixed Models:** Allows modeling of hierarchical data with categorical dependent variables, accommodating clustering and random effects.
- **Machine Learning Integration:** Incorporating classification trees and ensemble methods alongside traditional regression to improve predictive accuracy.

These advancements underscore Stata’s commitment to evolving alongside statistical methodology, providing users with cutting-edge tools for categorical data analysis.

The landscape of regression models for categorical dependent variables using Stata is both rich and dynamic. By understanding the theoretical underpinnings, practical implementation, and potential pitfalls, analysts can harness Stata’s full potential to uncover meaningful insights from discrete outcome data. Whether dealing with binary decisions, nominal choices, or ordered preferences, Stata equips researchers with the precision and flexibility necessary for rigorous categorical regression modeling.

Regression Models For Categorical Dependent Variables Using Stata

regression models for categorical dependent variables using stata: *Regression Models for Categorical Dependent Variables Using Stata, Third Edition* J. Scott Long, Jeremy Freese, 2014-09-10 Regression Models for Categorical Dependent Variables Using Stata, Third Edition shows how to use Stata to fit and interpret regression models for categorical data. The third edition is a complete rewrite of the book. Factor variables and the margins command changed how the effects of variables can be estimated and interpreted. In addition, the authors' views on interpretation have evolved. The changes to Stata and to the authors' views inspired the authors to completely rewrite their popular SPost commands to take advantage of the power of the margins command and the flexibility of factor-variable notation. The new edition will interest readers of a previous edition as well as new readers. Even though about 150 pages of appendixes were removed, the third edition is about 60 pages longer than the second. Although regression models for categorical dependent variables are common, few texts explain how to interpret such models; this text fills the void. With the book, Long and Freese provide a suite of commands for model interpretation, hypothesis testing, and model diagnostics. The new commands that accompany the third edition make it easy to include powers or interactions of covariates in regression models and work seamlessly with models estimated with complex survey data. The authors' new commands greatly simplify the use of margins, in the same way that the marginsplot command harnesses the power of margins for plotting predictions. The authors discuss how to use margins and their new mchange, mtable, and mgen commands to compute tables and to plot predictions. They also discuss how to use these commands to estimate marginal effects, averaged either over the sample or at fixed values of the regressors. The authors introduce and advocate a variety of new methods that use predictions to interpret the effect of variables in regression models. The third edition begins with an excellent introduction to Stata and follows with general treatments of the estimation, testing, fit, and interpretation of this class of models. New to the third edition is an entire chapter about how to interpret regression models using predictions—a chapter that is expanded upon in later chapters that focus on models for binary, ordinal, nominal, and count outcomes. Long and Freese use many concrete examples in their third edition. All the examples, datasets, and author-written commands are available on the authors' website, so readers can easily replicate the examples with Stata. This book is ideal for students or applied researchers who want to learn how to fit and interpret models for categorical data.

regression models for categorical dependent variables using stata: *Regression Models for Categorical Dependent Variables Using Stata, 3rd Edition* Scott Long, 2014

regression models for categorical dependent variables using stata: *Regression Models for Categorical Dependent Variables Using Stata, Second Edition* J. Scott Long, Jeremy Freese, 2006 The goal of the book is to make easier to carry out the computations necessary for the full interpretation of regression nonlinear models for categorical outcomes usign Stata.

regression models for categorical dependent variables using stata: A Gentle Introduction to Stata, Second Edition Alan C. Acock, 2008-09-03 A Gentle Introduction to Stata, Second Edition is aimed at new Stata users who want to become proficient in Stata. After reading this introductory text, new users will not only be able to use Stata well but also learn new aspects of

Stata easily. Acock assumes that the user is not familiar with any statistical software. This assumption of a blank slate is central to the structure and contents of the book. Acock starts with the basics; for example, the portion of the book that deals with data management begins with a careful and detailed example of turning survey data on paper into a Stata-ready dataset on the computer. When explaining how to go about basic exploratory statistical procedures, Acock includes notes that should help the reader develop good work habits. This mixture of explaining good Stata habits and good statistical habits continues throughout the book. Acock is quite careful to teach the reader all aspects of using Stata. He covers data management, good work habits (including the use of basic do-files), basic exploratory statistics (including graphical displays), and analyses using the standard array of basic statistical tools (correlation, linear and logistic regression, and parametric and nonparametric tests of location and dispersion). Acock teaches Stata commands by using the menus and dialog boxes while still stressing the value of do-files. In this way, he ensures that all types of users can build good work habits. Each chapter has exercises that the motivated reader can use to reinforce the material. The tone of the book is friendly and conversational without ever being glib or condescending. Important asides and notes about terminology are set off in boxes, which makes the text easy to read without any convoluted twists or forward-referencing. Rather than splitting topics by their Stata implementation, Acock chose to arrange the topics as they would be in a basic statistics textbook; graphics and postestimation are woven into the material in a natural fashion. Real datasets, such as the General Social Surveys from 2002 and 2006, are used throughout the book. The focus of the book is especially helpful for those in psychology and the social sciences, because the presentation of basic statistical modeling is supplemented with discussions of effect sizes and standardized coefficients. Various selection criteria, such as semipartial correlations, are discussed for model selection. The second edition of the book has been updated to reflect new features in Stata 10 and includes a new chapter on the use of factor analysis to develop valid, reliable scale measures.--Publisher's website.

regression models for categorical dependent variables using stata: *Categorical Data Analysis and Multilevel Modeling Using R* Xing Liu, 2022-02-24 Categorical Data Analysis and Multilevel Modeling Using R provides a practical guide to regression techniques for analyzing binary, ordinal, nominal, and count response variables using the R software. Author Xing Liu offers a unified framework for both single-level and multilevel modeling of categorical and count response variables with both frequentist and Bayesian approaches. Each chapter demonstrates how to conduct the analysis using R, how to interpret the models, and how to present the results for publication. A companion website for this book contains datasets and R commands used in the book for students, and solutions for the end-of-chapter exercises on the instructor site.

regression models for categorical dependent variables using stata: The SAGE Handbook of Regression Analysis and Causal Inference Henning Best, Christof Wolf, 2013-12-20 ‘The editors of the new SAGE Handbook of Regression Analysis and Causal Inference have assembled a wide-ranging, high-quality, and timely collection of articles on topics of central importance to quantitative social research, many written by leaders in the field. Everyone engaged in statistical analysis of social-science data will find something of interest in this book.’ - John Fox, Professor, Department of Sociology, McMaster University ‘The authors do a great job in explaining the various statistical methods in a clear and simple way - focussing on fundamental understanding, interpretation of results, and practical application - yet being precise in their exposition.’ - Ben Jann, Executive Director, Institute of Sociology, University of Bern ‘Best and Wolf have put together a powerful collection, especially valuable in its separate discussions of uses for both cross-sectional and panel data analysis.’ -Tom Smith, Senior Fellow, NORC, University of Chicago Edited and written by a team of leading international social scientists, this Handbook provides a comprehensive introduction to multivariate methods. The Handbook focuses on regression analysis of cross-sectional and longitudinal data with an emphasis on causal analysis, thereby covering a large number of different techniques including selection models, complex samples, and regression discontinuities. Each Part starts with a non-mathematical introduction to the method covered in that

section, giving readers a basic knowledge of the method's logic, scope and unique features. Next, the mathematical and statistical basis of each method is presented along with advanced aspects. Using real-world data from the European Social Survey (ESS) and the Socio-Economic Panel (GSOEP), the book provides a comprehensive discussion of each method's application, making this an ideal text for PhD students and researchers embarking on their own data analysis.

regression models for categorical dependent variables using stata: *No Substitute for Competence* Simon Lanz, 2020-01-08 Issue ownership theory is a tale of two actors. On the one hand, it theorizes how parties compete with each other in their struggle for votes. On the other hand, issue ownership is about the citizen. It claims that voters are more likely to support a party if they think it is competent to handle issues they care about. This book provides unique insights into the undertheorized and understudied links between party competence and the vote. It argues that issue ownership voting (or competence-based voting) consists of three assumptions: First, voters are primarily interested in having issues handled by a competent party. Unlike in other issue voting models this implies that voters are reluctant (or unable) to deal with the specificities of the exact solution to a political problem. Though positional considerations feed into evaluations of party competence, other factors are important, too. This is reflected by the second assumption, following which issue handling competence is a subjective preference with various sources. Third, competence is more decisive in the decision-making process if the voter cares deeply about the issue. These three assumptions yield the key formula of issue ownership voting: Voters support the most competent party on the most important issue.

regression models for categorical dependent variables using stata: Informatics in Control, Automation and Robotics Oleg Gusikhin, Kurosh Madani, Henk Nijmeijer, 2023-02-24 The book focuses the latest endeavors relating researches and developments conducted in fields of control, robotics, and automation. Through more than ten revised and extended articles, the present book aims to provide the most up-to-date state of the art of the aforementioned fields allowing researcher, Ph.D. students, and engineers not only updating their knowledge but also benefiting from the source of inspiration that represents the set of selected articles of the book. The deliberate intention of editors to cover as well theoretical facets of those fields as their practical accomplishments and implementations offers the benefit of gathering in a same book a factual and well-balanced prospect of nowadays research in those topics. A special attention toward "Intelligent Robots and Control" may characterize another benefit of this book.

regression models for categorical dependent variables using stata: *Computational and Statistical Methods for Analysing Big Data with Applications* Shen Liu, James Mcgree, Zongyuan Ge, Yang Xie, 2015-11-20 Due to the scale and complexity of data sets currently being collected in areas such as health, transportation, environmental science, engineering, information technology, business and finance, modern quantitative analysts are seeking improved and appropriate computational and statistical methods to explore, model and draw inferences from big data. This book aims to introduce suitable approaches for such endeavours, providing applications and case studies for the purpose of demonstration. Computational and Statistical Methods for Analysing Big Data with Applications starts with an overview of the era of big data. It then goes onto explain the computational and statistical methods which have been commonly applied in the big data revolution. For each of these methods, an example is provided as a guide to its application. Five case studies are presented next, focusing on computer vision with massive training data, spatial data analysis, advanced experimental design methods for big data, big data in clinical medicine, and analysing data collected from mobile devices, respectively. The book concludes with some final thoughts and suggested areas for future research in big data. - Advanced computational and statistical methodologies for analysing big data are developed - Experimental design methodologies are described and implemented to make the analysis of big data more computationally tractable - Case studies are discussed to demonstrate the implementation of the developed methods - Five high-impact areas of application are studied: computer vision, geosciences, commerce, healthcare and transportation - Computing code/programs are provided where appropriate

regression models for categorical dependent variables using stata: Data Analysis in Medicine and Health using R Kamarul Imran Musa, Wan Nor Arifin Wan Mansor, Tengku Muhammad Hanis, 2023-09-12 Data analysis plays a vital role in guiding medical treatment plans, patient care, and the formulation of control and prevention policies in the field of healthcare. In today's era, researchers in these domains require a firm grasp of data, statistical concepts, and programming skills due to the increasing complexity of data. Reproducible analyses and cutting-edge statistical methods are becoming increasingly necessary. This book, which is both comprehensive and highly practical, addresses these challenges by laying a solid foundation of data and statistical theory for readers. Subsequently, it equips them with practical skills to conduct analyses using the powerful R programming language, widely used by statisticians. The book takes a gentle approach to help readers navigate data and statistical analysis using R, minimizing the learning curve. RStudio is used as the integrated development environment (IDE) for enhanced productivity for readers to run their R codes. Following a logical sequence commonly applied in medical and health research, the book covers fundamental concepts of data analysis and statistical modeling techniques. It provides readers, including those with limited statistical knowledge and programming skills, with hands-on experience through R programming. The online version of this book is available on bookdown.org, a publishing platform provided by RStudio, PBC specifically designed to host books written using the bookdown package in R. Additionally, all R codes and datasets in this book can be found on the author's GitHub repository.

regression models for categorical dependent variables using stata: *International Handbook of the Demography of Poverty and Inequality* Joachim Singelmann, Dudley L. Poston Jr., 2025-05-27 This handbook provides a comprehensive and comparative review of poverty and inequality in developing and developed countries from demographic, sociological, and political perspectives. It discusses, among others, the basics of poverty and inequality, issues of poverty and inequality among race/ethnic and sexual minority subpopulations, social processes, country and comparative perspectives, child poverty, and anti-poverty policies. It also touches on regional disparities. There has been rapid decline in poverty and inequality in China, East Asian and Pacific countries. Declines have also occurred in South Asia. But in Sub-Saharan Africa, poverty and inequality have risen in some countries, while in other regions, such as transition economies in Europe and Central Asia, only modest changes have occurred. This handbook provides a great reference for demographers, social scientists, policy makers and NGOs.

regression models for categorical dependent variables using stata: Politicking Online Costas Panagopoulos, Vassia Gueorguieva, Allison Slotnick, Girish Gulati, Christine Williams, 2009-05-01 Of the many groundbreaking developments in the 2008 presidential election, the most important may well be the use of the Internet. In Politicking Online contributors explorethe impact of technology for electioneering purposes, from running campaigns andincreasing representation to ultimately strengthening democracy. The book reveals how social networking sites such as MySpace and Facebook are used in campaigns along withe-mail, SMS text messaging, and mobile phones to help inform, target, mobilize, and communicate with voters. While the Internet may have transformed the landscape of modern political campaigns throughout the world, Costas Panagopoulos reminds readers that officials and campaign workers need to adapt to changing circumstances, know the limits of their methods, and combine new technologies with more traditional techniques to achieve an overall balance.

regression models for categorical dependent variables using stata: Analysis of Incidence Rates Peter Cummings, 2019-04-16 Incidence rates are counts divided by person-time; mortality rates are a well-known example. Analysis of Incidence Rates offers a detailed discussion of the practical aspects of analyzing incidence rates. Important pitfalls and areas of controversy are discussed. The text is aimed at graduate students, researchers, and analysts in the disciplines of epidemiology, biostatistics, social sciences, economics, and psychology. Features: Compares and contrasts incidence rates with risks, odds, and hazards. Shows stratified methods, including standardization, inverse-variance weighting, and Mantel-Haenszel methods Describes Poisson

regression methods for adjusted rate ratios and rate differences. Examines linear regression for rate differences with an emphasis on common problems. Gives methods for correcting confidence intervals. Illustrates problems related to collapsibility. Explores extensions of count models for rates, including negative binomial regression, methods for clustered data, and the analysis of longitudinal data. Also, reviews controversies and limitations. Presents matched cohort methods in detail. Gives marginal methods for converting adjusted rate ratios to rate differences, and vice versa. Demonstrates instrumental variable methods. Compares Poisson regression with the Cox proportional hazards model. Also, introduces Royston-Parmar models. All data and analyses are in online Stata files which readers can download. Peter Cummings is Professor Emeritus, Department of Epidemiology, School of Public Health, University of Washington, Seattle WA. His research was primarily in the field of injuries. He used matched cohort methods to estimate how the use of seat belts and presence of airbags were related to death in a traffic crash. He is author or co-author of over 100 peer-reviewed articles.

regression models for categorical dependent variables using stata: Signaling Effects of Crowdfunding on Venture Investors' Decision Making Michael Mödl, 2020-08-31 Michael Mödl examines the impact and signaling effects of crowd-based start-up financing on subsequent venture capital funding rounds. The digital era has substantially expanded the entrepreneurial financing landscape and crowdfunding is emerging as a novel way for young innovative firms to secure scarce early-stage funding. As the evaluations of venture capitalists and business angels remain to be consequential for securing critical resources in later stages, questions on the interactions between new and traditional forms of venture financing arise. Drawing on choice experimental research designs the author provides causal empirical evidence that while "the crowd" is generally seen as a negative signal, it can generate certain positive signals which increase the likelihood of professional venture investors to consider crowdfunded start-ups for investment. The presented findings yield important implications for capital-seeking entrepreneurs, investors and public policy. Dr. Michael Maximilian Mödl completed his dissertation under supervision of Prof. Dietmar Harhoff, Ph.D. at Ludwig-Maximilians-Universität München and at the Max Planck Institute for Innovation and Competition.

regression models for categorical dependent variables using stata: Transnational Activities and Immigrant Integration in Germany Reinhard Schunck, 2014-02-17 This book investigates both the causes and effects of transnational activities among immigrants in relation to their integration into the receiving society. It uses large scale, representative data about first and second generation immigrants in Germany. It develops a formal theoretical model, which explains both transnational involvement and paths of immigrant integration. Important questions are answered: What consequences does transnational involvement have on integration? Is transnational involvement a distinct form of integration? Is it an alternative to assimilation? Does it hinder or facilitate assimilation? Longitudinal analyses are presented which show that immigrant integration and transnational involvement do not necessarily oppose each other. The book shows that although low levels of integration may coincide with strong transnational ties, the relationship is not causal. This book shows how immigrant integration and transnational involvement are related to each other and how a joint examination of both processes may advance our understanding of the general dynamics of migration and integration.

regression models for categorical dependent variables using stata: Applied Statistics for the Social and Health Sciences Rachel A. Gordon, 2023-10-31 Covering basic univariate and bivariate statistics and regression models for nominal, ordinal, and interval outcomes, Applied Statistics for the Social and Health Sciences provides graduate students in the social and health sciences with fundamental skills to estimate, interpret, and publish quantitative research using contemporary standards. Reflecting the growing importance of Big Data in the social and health sciences, this thoroughly revised and streamlined new edition covers best practice in the use of statistics in social and health sciences, draws upon new literatures and empirical examples, and highlights the importance of statistical programming, including coding, reproducibility,

transparency, and open science. Key features of the book include: interweaving the teaching of statistical concepts with examples from publicly available social and health science data and literature excerpts; thoroughly integrating the teaching of statistical theory with the teaching of data access, processing, and analysis in Stata; recognizing debates and critiques of the origins and uses of quantitative methods.

regression models for categorical dependent variables using stata: *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists* Scott M. Lynch, 2007-06-30 Introduction to Applied Bayesian Statistics and Estimation for Social Scientists covers the complete process of Bayesian statistical analysis in great detail from the development of a model through the process of making statistical inference. The key feature of this book is that it covers models that are most commonly used in social science research - including the linear regression model, generalized linear models, hierarchical models, and multivariate regression models - and it thoroughly develops each real-data example in painstaking detail. The first part of the book provides a detailed introduction to mathematical statistics and the Bayesian approach to statistics, as well as a thorough explanation of the rationale for using simulation methods to construct summaries of posterior distributions. Markov chain Monte Carlo (MCMC) methods - including the Gibbs sampler and the Metropolis-Hastings algorithm - are then introduced as general methods for simulating samples from distributions. Extensive discussion of programming MCMC algorithms, monitoring their performance, and improving them is provided before turning to the larger examples involving real social science models and data.

regression models for categorical dependent variables using stata: Confirmatory Factor Analysis J. Micah Roos, Shawn Bauldry, 2021-10-06 Measurement connects theoretical concepts to what is observable in the empirical world, and is fundamental to all social and behavioral research. In this volume, J. Micah Roos and Shawn Bauldry introduce a popular approach to measurement: confirmatory factor analysis, with examples in every chapter draw from national survey data. Data to replicate the examples are available on a companion website, along with code in R, Stata, and Mplus.

regression models for categorical dependent variables using stata: Quantitative Data Analysis Donald J. Treiman, 2014-01-30 This book is an accessible introduction to quantitative dataanalysis, concentrating on the key issues facing those new toresearch, such as how to decide which statistical procedure issuitable, and how to interpret the subsequent results. Each chapterincludes illustrative examples and a set of exercises that allowsreaders to test their understanding of the topic. The book, writtenfor graduate students in the social sciences, public health, andeducation, offers a practical approach to making sociological senseout of a body of quantitative data. The book also will be useful tomore experienced researchers who need a readily accessible handbookon quantitative methods. The author has posted stata files, updates and data sets athis websitehttp://tinyurl.com/Treiman-stata-files-data-sets.

regression models for categorical dependent variables using stata: Computer Analysis of Images and Patterns George Azzopardi, Nicolai Petkov, 2015-08-25 The two volume set LNCS 9256 and 9257 constitutes the refereed proceedings of the 16th International Conference on Computer Analysis of Images and Patterns, CAIP 2015, held in Valletta, Malta, in September 2015. The 138 papers presented were carefully reviewed and selected from numerous submissions. CAIP 2015 is the sixteenth in the CAIP series of biennial international conferences devoted to all aspects of computer vision, image analysis and processing, pattern recognition, and related fields.

Related to regression models for categorical dependent variables using stata

Regression with multiple dependent variables? - Cross Validated Is it possible to have a (multiple) regression equation with two or more dependent variables? Sure, you could run two separate regression equations, one for each DV, but that

regression - When is R squared negative? - Cross Validated Also, for OLS regression, R^2 is the squared correlation between the predicted and the observed values. Hence, it must be non-negative. For simple OLS regression with one predictor, this is

regression - Trying to understand the fitted vs residual plot? - Cross A good residual vs fitted plot has three characteristics: The residuals "bounce randomly" around the 0 line. This suggests that the assumption that the relationship is linear is

How should outliers be dealt with in linear regression analysis? What statistical tests or rules of thumb can be used as a basis for excluding outliers in linear regression analysis? Are there any special considerations for multilinear regression?

regression - Why do we say the outcome variable "is regressed on" The word "regressed" is used instead of "dependent" because we want to emphasise that we are using a regression technique to represent this dependency between x

Interpreting Z-Scores of Linear Regression Coefficients Well, under the hypothetical scenario that the true regression coefficient is equal to 0, statisticians have figured out how likely a given Z-score is (using the normal distribution

correlation - What is the difference between linear regression on y The Pearson correlation coefficient of x and y is the same, whether you compute pearson(x, y) or pearson(y, x). This suggests that doing a linear regression of y given x or x given y should be

regression - Difference between forecast and prediction - Cross I was wondering what difference and relation are between forecast and prediction? Especially in time series and regression? For example, am I correct that: In time series, forecasting seems

regression - Linear model with both additive and multiplicative In a log-level regression, the independent variables have an additive effect on the log-transformed response and a multiplicative effect on the original untransformed response

How to derive the standard error of linear regression coefficient How to derive the standard error of linear regression coefficient Ask Question Asked 11 years, 7 months ago Modified 1 year ago

Regression with multiple dependent variables? - Cross Validated Is it possible to have a (multiple) regression equation with two or more dependent variables? Sure, you could run two separate regression equations, one for each DV, but that

regression - When is R squared negative? - Cross Validated Also, for OLS regression, R^2 is the squared correlation between the predicted and the observed values. Hence, it must be non-negative. For simple OLS regression with one predictor, this is

regression - Trying to understand the fitted vs residual plot? A good residual vs fitted plot has three characteristics: The residuals "bounce randomly" around the 0 line. This suggests that the assumption that the relationship is linear is

How should outliers be dealt with in linear regression analysis? What statistical tests or rules of thumb can be used as a basis for excluding outliers in linear regression analysis? Are there any special considerations for multilinear regression?

regression - Why do we say the outcome variable "is regressed The word "regressed" is used instead of "dependent" because we want to emphasise that we are using a regression technique to represent this dependency between x

Interpreting Z-Scores of Linear Regression Coefficients Well, under the hypothetical scenario that the true regression coefficient is equal to 0, statisticians have figured out how likely a given Z-score is (using the normal distribution

correlation - What is the difference between linear regression on y The Pearson correlation coefficient of x and y is the same, whether you compute pearson(x, y) or pearson(y, x). This suggests that doing a linear regression of y given x or x given y should be

regression - Difference between forecast and prediction - Cross I was wondering what difference and relation are between forecast and prediction? Especially in time series and regression? For example, am I correct that: In time series, forecasting seems

regression - Linear model with both additive and multiplicative In a log-level regression, the independent variables have an additive effect on the log-transformed response and a multiplicative effect on the original untransformed response

How to derive the standard error of linear regression coefficient How to derive the standard error of linear regression coefficient Ask Question Asked 11 years, 7 months ago Modified 1 year ago

Regression with multiple dependent variables? - Cross Validated Is it possible to have a (multiple) regression equation with two or more dependent variables? Sure, you could run two separate regression equations, one for each DV, but that

regression - When is R squared negative? - Cross Validated Also, for OLS regression, R^2 is the squared correlation between the predicted and the observed values. Hence, it must be non-negative. For simple OLS regression with one predictor, this is

regression - Trying to understand the fitted vs residual plot? A good residual vs fitted plot has three characteristics: The residuals "bounce randomly" around the 0 line. This suggests that the assumption that the relationship is linear is

How should outliers be dealt with in linear regression analysis? What statistical tests or rules of thumb can be used as a basis for excluding outliers in linear regression analysis? Are there any special considerations for multilinear regression?

regression - Why do we say the outcome variable "is regressed The word "regressed" is used instead of "dependent" because we want to emphasise that we are using a regression technique to represent this dependency between x

Interpreting Z-Scores of Linear Regression Coefficients Well, under the hypothetical scenario that the true regression coefficient is equal to 0, statisticians have figured out how likely a given Z-score is (using the normal distribution

correlation - What is the difference between linear regression on y The Pearson correlation coefficient of x and y is the same, whether you compute pearson(x, y) or pearson(y, x). This suggests that doing a linear regression of y given x or x given y should be

regression - Difference between forecast and prediction - Cross I was wondering what difference and relation are between forecast and prediction? Especially in time series and regression? For example, am I correct that: In time series, forecasting seems

regression - Linear model with both additive and multiplicative In a log-level regression, the independent variables have an additive effect on the log-transformed response and a multiplicative effect on the original untransformed response

How to derive the standard error of linear regression coefficient How to derive the standard error of linear regression coefficient Ask Question Asked 11 years, 7 months ago Modified 1 year ago

Regression with multiple dependent variables? - Cross Validated Is it possible to have a (multiple) regression equation with two or more dependent variables? Sure, you could run two separate regression equations, one for each DV, but that

regression - When is R squared negative? - Cross Validated Also, for OLS regression, R^2 is the squared correlation between the predicted and the observed values. Hence, it must be non-negative. For simple OLS regression with one predictor, this is

regression - Trying to understand the fitted vs residual plot? A good residual vs fitted plot has three characteristics: The residuals "bounce randomly" around the 0 line. This suggests that the assumption that the relationship is linear is

How should outliers be dealt with in linear regression analysis? What statistical tests or rules of thumb can be used as a basis for excluding outliers in linear regression analysis? Are there any special considerations for multilinear regression?

regression - Why do we say the outcome variable "is regressed The word "regressed" is used instead of "dependent" because we want to emphasise that we are using a regression technique to represent this dependency between x

Interpreting Z-Scores of Linear Regression Coefficients Well, under the hypothetical scenario that the true regression coefficient is equal to 0, statisticians have figured out how likely a given Z-score is (using the normal distribution

correlation - What is the difference between linear regression on y The Pearson correlation coefficient of x and y is the same, whether you compute pearson(x, y) or pearson(y, x). This suggests that doing a linear regression of y given x or x given y should be

regression - Difference between forecast and prediction - Cross I was wondering what difference and relation are between forecast and prediction? Especially in time series and regression? For example, am I correct that: In time series, forecasting seems

regression - Linear model with both additive and multiplicative In a log-level regression, the independent variables have an additive effect on the log-transformed response and a multiplicative effect on the original untransformed response

How to derive the standard error of linear regression coefficient How to derive the standard error of linear regression coefficient Ask Question Asked 11 years, 7 months ago Modified 1 year ago

Regression with multiple dependent variables? - Cross Validated Is it possible to have a (multiple) regression equation with two or more dependent variables? Sure, you could run two separate regression equations, one for each DV, but that

regression - When is R squared negative? - Cross Validated Also, for OLS regression, R^2 is the squared correlation between the predicted and the observed values. Hence, it must be non-negative. For simple OLS regression with one predictor, this is

regression - Trying to understand the fitted vs residual plot? - Cross A good residual vs fitted plot has three characteristics: The residuals "bounce randomly" around the 0 line. This suggests that the assumption that the relationship is linear is

How should outliers be dealt with in linear regression analysis? What statistical tests or rules of thumb can be used as a basis for excluding outliers in linear regression analysis? Are there any special considerations for multilinear regression?

regression - Why do we say the outcome variable "is regressed on" The word "regressed" is used instead of "dependent" because we want to emphasise that we are using a regression technique to represent this dependency between x

Interpreting Z-Scores of Linear Regression Coefficients Well, under the hypothetical scenario that the true regression coefficient is equal to 0, statisticians have figured out how likely a given Z-score is (using the normal distribution

correlation - What is the difference between linear regression on y The Pearson correlation coefficient of x and y is the same, whether you compute pearson(x, y) or pearson(y, x). This suggests that doing a linear regression of y given x or x given y should be

regression - Difference between forecast and prediction - Cross I was wondering what difference and relation are between forecast and prediction? Especially in time series and regression? For example, am I correct that: In time series, forecasting seems

regression - Linear model with both additive and multiplicative In a log-level regression, the independent variables have an additive effect on the log-transformed response and a multiplicative effect on the original untransformed response

How to derive the standard error of linear regression coefficient How to derive the standard error of linear regression coefficient Ask Question Asked 11 years, 7 months ago Modified 1 year ago

Back to Home: https://old.rga.ca