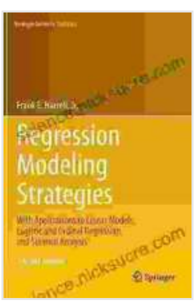
Generalized Linear Models: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival

Generalized linear models (GLMs) are a powerful class of statistical models that extend the ordinary linear model to accommodate a wider range of response variables. They are used in a variety of applications, including linear models, logistic and ordinal regression, and survival analysis.

GLMs are based on the idea that the response variable follows a distribution from the exponential family. The exponential family is a family of probability distributions that includes the normal, binomial, Poisson, and gamma distributions, among others. The exponential family is characterized by its probability density function, which takes the form:

$$f(y; \theta, \phi) = \exp[(y\theta - b(\theta)) / \phi + c(y, \phi)]$$



Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis (Springer Series in Statistics)

by Frank E. Harrell Jr.

4.6 out of 5

Language	: English
File size	: 13893 KB
Screen Reader	: Supported
Print length	: 607 pages

where y is the response variable, θ is the natural parameter, ϕ is the dispersion parameter, and b(θ) and c(y, ϕ) are known functions.

The natural parameter θ is related to the mean of the distribution. The dispersion parameter ϕ is related to the variance of the distribution.

GLMs are fitted by maximizing the log-likelihood function. The log-likelihood function is the logarithm of the joint probability density function of the observed data. Maximizing the log-likelihood function produces parameter estimates that maximize the likelihood of observing the data.

Once a GLM has been fitted, it can be used to make predictions about the response variable. Predictions are made by using the fitted model to estimate the mean of the response variable for a given set of predictor variables.

GLMs are used in a variety of applications, including:

- **Linear models:** GLMs can be used to fit linear models, which are models in which the response variable is a continuous variable. Linear models are used in a variety of applications, including regression analysis and analysis of variance.
- **Logistic regression:** GLMs can be used to fit logistic regression models, which are models in which the response variable is a binary variable. Logistic regression models are used in a variety of applications, including classification and prediction.
- **Ordinal regression:** GLMs can be used to fit ordinal regression models, which are models in which the response variable is an ordinal

variable. Ordinal regression models are used in a variety of applications, including ranking and rating.

- **Survival analysis:** GLMs can be used to fit survival analysis models, which are models in which the response variable is the time until an event occurs. Survival analysis models are used in a variety of applications, including medical research and public health.

The theory of GLMs is based on the idea that the response variable follows a distribution from the exponential family. The exponential family is a family of probability distributions that includes the normal, binomial, Poisson, and gamma distributions, among others.

The exponential family is characterized by its probability density function, which takes the form:

$$f(y; \theta, \phi) = \exp[(y\theta - b(\theta)) / \phi + c(y, \phi)]$$

where y is the response variable, θ is the natural parameter, ϕ is the dispersion parameter, and $b(\theta)$ and $c(y, \phi)$ are known functions.

The natural parameter θ is related to the mean of the distribution. The dispersion parameter ϕ is related to the variance of the distribution.

GLMs are fitted by maximizing the log-likelihood function. The log-likelihood function is the logarithm of the joint probability density function of the observed data. Maximizing the log-likelihood function produces parameter estimates that maximize the likelihood of observing the data.

Once a GLM has been fitted, it can be used to make predictions about the response variable. Predictions are made by using the fitted model to estimate the mean of the response variable for a given set of predictor variables.

GLMs are fitted by maximizing the log-likelihood function. The log-likelihood function is the logarithm of the joint probability density function of the observed data. Maximizing the log-likelihood function produces parameter estimates that maximize the likelihood of observing the data.

There are a variety of algorithms that can be used to maximize the log-likelihood function. One common algorithm is the iterative weighted least squares (IWLS) algorithm. The IWLS algorithm iteratively updates the parameter estimates until they converge to a maximum.

Once the parameter estimates have converged, the fitted GLM can be used to make predictions about the response variable. Predictions are made by using the fitted model to estimate the mean of the response variable for a given set of predictor variables.

The interpretation of a GLM depends on the type of response variable. For linear models, the interpretation is straightforward. The slope of the regression line represents the change in the mean of the response variable for a one-unit increase in the predictor variable. The intercept of the regression line represents the mean of the response variable when the predictor variable is zero.

For logistic regression models, the interpretation is a bit more complex. The log-odds of the response variable is a linear function of the predictor variables. The log-odds of the response variable is the logarithm of the odds of the response variable occurring. The odds of the response variable occurring is the ratio of the probability of the response variable occurring to the probability of the response variable not occurring.

For ordinal regression models, the interpretation is even more complex. The cumulative log-odds of the response variable is a linear function of the predictor variables. The cumulative log-odds of the response variable is the logarithm of the odds of the response variable being less than or equal to a given category. The odds of the response variable being less than or equal to a given category is the ratio of the probability of the response variable being less than or equal to that category to the probability of the response variable being greater than that category.

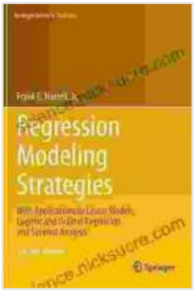
For survival analysis models, the interpretation is again complex. The hazard function is a linear function of the predictor variables. The hazard function is the instantaneous rate of failure. The instantaneous rate of failure is the probability of failure at a given time, given that the subject has survived up to that time.

GLMs are a powerful class of statistical models that can be used to model a wide variety of response variables. They are used in a variety of applications, including linear models, logistic and ordinal regression, and survival analysis. GLMs are relatively easy to fit and interpret, making them a valuable tool for data analysis.

Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis (Springer Series in Statistics)

by Frank E. Harrell Jr.

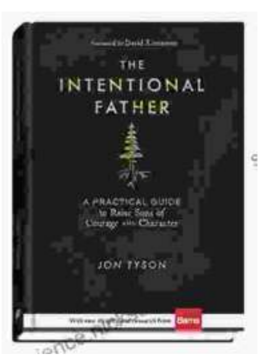
4.6 out of 5



Language : English
File size : 13893 KB
Screen Reader : Supported
Print length : 607 pages

Compilation of Short Stories on Mental Illness and Ways to Handle Them

Mental illness is a serious issue that affects millions of people around the world. It can be a debilitating condition that can make it difficult to live a normal life....



The Practical Guide to Raising Courageous and Characterful Sons

As parents, we all want our sons to grow up to be good men. We want them to be kind, compassionate, and brave. We want them to stand up for what they...