Graphical user interface, application

Description automatically generated**STAT 608 (After MT#2 Cheat Sheet)**

**STAT 608 Chapter 8: Logistic Regression**

Linear Models: Models with transformed response variables are **NOT** linear models:

* **Logistic** regression models are **NOT** linear models

Logistic Regression: Why not use a linear regression when y is a categorical variable? **(1)** possible predictions out of bounds **(2)** Nonconstant variance; If y~Bernoulli 🡪

**(3)** When the response Y~Bernoulli, the logistic regression model correctly models the mean.

* Logit(p) = logodds(p) = : Model: ; **NOTE: (1)** log odds is **linear** in x; **(2)**

How B0 Affects Logit Model: Larger values of shift curve up and to the left;

How B1 Affects Logit Model: Larger values of steeper slope; if , slope is positive

**Interpreting Slope:** The odds of (being classified as positive value of Y) are predicted to be multiplied by approximately when x is increased by 1 unit

* Text

  Description automatically generated with low confidence**(1)** Slope is multiplicative not additive **(2)** Talking about odds, not the chance or probability (or mean, i.e. we don’t say on average anymore) **(3)** We must exponentiate before interpreting. **(4)** Still a statistic, not a parameter **(5)** Still controlling for other variables in the model.

**Logistic Regression: Multiple Values of Y at every X (A sample of size mi at every obs. value of xi)**

* (note that the prob. of success is a function of x.
* Can't use prop. of successes instead of odds b.c. (Var. isn’t constant)

Finding Parameter Estimates: The parameters for the logistic regression model are found by maximizing the log

likelihood. This is equivalent to minimizing the deviance -2\*log(L). **NOTE:** for Lin Reg Models, minimizing

RSS had a closed form solution, for logistic regression models, we need an iterative method to find estimates.

Hypothesis Tests and CIs: NOTE that Wald test uses Z-values not T-values:

* To construct CI:

Residuals/Deviance:

* In logistic regression the concept of residual sums of squares is replaced by deviance.
* The saturated model is one with a separate proportion of successes for every value of .
* Deviance measures the difference between the log likelihood from the saturated model and the log likelihood from our fitted model
* where n = # of levels of X
* When each mi is large enough, the deviance statistic can be used as a chi-squared goodness-of-fit test for the logistic model.
* Note: if sample sizes are too large then tiny differences btwn. the saturated and fitted models would be considered significant 🡪 reject H0: logistic Model is appropriate
* In model output: Null Deviance 🡪 used to test if model is significant ( Similar to the overall F-test
* In model output: Residual Deviance is used to test goodness of fit.
* We can also use deviance to test whether two nested models are significantly different: The difference between the Null Deviances for the two models is compared to a with df = the difference between the df for the two models. NOTE: in general this doesn’t give the same result as the Wald z-test

Pearson Goodness of fit Statistic: Alternative measure to null deviance; If mi is large enough, deviance and pearson will be similar; o.w. we prefer deviance.

Residuals for Logistic Regression: **(1)** Response Residuals: . Problem is that the variance is not constant, so resp. resids. are difficult to interpret.

* **(2)** Pearson Residuals: ; fixes nonconstant error variance (the resids. are the sqrt of the obs. contribution to the pearson chi-squared statistic) Pearson Residuals still don’t account for the var. of the model estimate for so we correct for that with the std. Pearson resids
* **(3)** Deviance Residuals: ; std. dev. resids. **(Dev. Residuals are preferred over others)**

**Logistic Regression: One Value of Y at every X (mi =1)**

* If we fit data with the assumption that all the mi=1, the parameter estimates, their sds, z-scores and p-values all stay the same as before
* The difference is in the deviance (AIC values are also different); deviance is constant

Binary Deviance: when mi=1, the deviance between the saturated model and the current model only depends on the log likelihood of our model.

* Deviance doesn’t provide an assessment of the goodness of fit (it also doesn’t have a chi-squared distribution)
* However, we can still use deviance to compare two models; the difference between two deviances still has an approximate chi-squared distribution

Binary Residuals: Instead of examining residual plots, compare the fitted model to a nonparametric fit.

**Transformations (of the x-vars in particular), Marginal Model Plots, Outliers:**

* Transform Y when: **(1)** We have outliers, **(2)** The relationship between log(odds) and x-vars is nonlinear
* Transform X when: **(1)** log(odds) is a function of log(x) for right skewed distributions of (x|y) and log(odds) is a function of x^2 when (x|y) ~N **(2)** To make our variance constant.

Transforming Predictors for Binary Data: Normal Predictor: When and

* If then we have nonconstant variance 🡪 add x2 term into model
* If then the log(odds) is a linear function of x with

Transforming Predictors for Binary Data: Multivariate Normal; When we have p predictors that are multivariate normal with different cov. matrices for Y=0 and Y=1 then:

* The log(odds) are a function of
* If , add a quadratic term in
* If the regression of on has a different slope for y=0 and y=1, add the interaction

Marginal Model Plots: Residual plots are difficult to interpret for logistic regression models… instead use marginal model plots (same concept as before)

**Homework 6 Problems:**

NOTE: if marginal model plot for x doesn't look good, for variable x, draw density plots for (X|Y=1) and (X|Y=0). If they look normally distributed but with different variances, then add X2 to the model. If they look right skewed, add log(X) to the model.

**STAT 608 Chapter 9: Serially Correlated Errors**

Fitting a LS model when autocorrelation is present: If we ignore correlation, we get unbiased coefficient estimates, but the standard errors will be wrong. If we have

**positive correlation** 🡪 < it should be 🡪 **p-values** for hypo tests about coefficients will be t**oo small** and **CI's** for the coefficients will be **too narrow**

* We can transform GLS models to LS models in order to access the usual residual diagnostic plots

Autocorrelation: To identify autocorrelation: **(1)** Look at lag plot (x=Yt-1, y=Yt) to see evidence of autocorrelation **(2)** Look for patterns in time series plot (Y vs time variable) **(3)** Look at acf plot of standardized residuals from fitting a linear regression model (without correlation structure specified)

* ; Autocorrelations bigger than in absolute value are statistically significantly different from 0.

Generalized Least Squares:

GLS when errors are AR(1) **Model:** where

* In general
* Covariance matrix for AR(1) model:

Transforming an AR(1) model into a model with iid errors: Model:

* **(1)** **(2)** Subtracting (1) from the model we get: which; since we have **(3)** .
* Thus, setting and setting
* we now have a model with iid errors
* Equivalently we can define
* The first observation will be influential when we do these sorts of transformations

A General Approach to Transforming GLS to LS:

* Recall: . B/c is symmetric and positive definite, it can be written as where is a lower triangular matrix with positive diagonals
* Multiplying each side of our model by 🡪
* Apply the transformations: and fit the linear model to obtain the GLS estimate of using Least Squares