BIST P8110: Applied Regression II

19. Intro to PROC GENMOD

Qixuan Chen

Department of Biostatistics
Columbia University

PROC GENMOD

► The GENMOD procedure fits generalized linear models (GLM), such as
  ► linear regression
  ► logistic regression
  ► ordinal logistic regression
  ► Poisson regression
  ► negative binomial regression
  ► zero-inflated Poisson regression

► The GENMOD procedure can also fit generalized linear models to correlated responses by the GEE method

SAS Syntax

The following statements are often used in the GENMOD procedure. Items within the < > are optional.

```
PROC GENMOD < options > ;
    CLASS variables ;
    MODEL response = < effects > < /options > ;
    REPEATED SUBJECT = subject-effect < /options > ;
    RUN ;
```

The REPEATED statement is used in GEE analysis.

PROC GENMOD Statement

- ► The PROC GENMOD statement invokes the procedure.
- ► Two most commonly used options are
  - ► DATA=SAS data-set, to specify the SAS data set containing the data to be analyzed.
  - ► PLOTS, to specify the plots to be created.
    - ► Before requesting plots, enable ODS Graphics
      ```
      ODS GRAPHICS ON;
      ODS GRAPHICS OFF;
      ```
    - ► If to produce all plots available, use options
      ```
      PROC GENMOD PLOTS=all;
      ```
    - ► If to request one of the available plots, use options
      ```
      PROC GENMOD PLOTS=PREDICTED;
      ```
    - ► The default is to display related multiple plots in a panel. If to display multiple plots individually, use options
      ```
      PROC GENMOD PLOTS(UNPACK)=DFBETA;
      ```

CLASS Statement

▶ The CLASS statement names the categorical variables to be used in the analysis.

▶ The CLASS statement must precede the MODEL statement.

▶ The options can be specified for each variable by enclosing them in parentheses after the variable name

```
CLASS treatment (REF='4') age (REF='1') / PARAM = REF ;
```

▶ The global options can be also specified for the CLASS statement by placing them after a slash (/).

```
CLASS treatment age / PARAM = REF REF =first;
```

MODEL Statement

▶ The MODEL statement specifies the response and the explanatory variables.

```
MODEL response = <effects> </options> ;
```

▶ The response can be specified in the form of a single variable or in the form of a ratio of two variables denoted events/trials.

   ▶ The first form is applicable to all responses. The second form is applicable only to grouped binomial response data.

▶ The options often use

   ▶ COVB, DIST, LINK, OFFSET, SCALE, and TYPE3.

MODEL Statement Options

- ▶ COVB – requests that the parameter estimate covariance matrix be displayed.
- ▶ TYPE3 – requests that statistics for Type 3 contrasts be computed for each effect specified in the MODEL statement.
- ▶ DIST – specifies the built-in probability distribution to use in the model.

DIST=	Distribution
NOR	normal
BIN	binomial
MULT	multinomial
POI	Poisson
NB	negative binomial
ZIP	zero-inflated Poisson
ZINB	zero-inflated negative binomial

MODEL Statement Options

▶ LINK – specifies the link function to be used in the model.

LINK=	Link Function
ID	identity
LOG	log
LOGIT	logit

▶ OFFSET – specifies a variable in the input data set to be used as an offset variable.

▶ SCALE – specifies an overdispersed model.

  ▶ EX: In Poisson distribution, we specify $var(Y) = \phi\lambda$. The dispersion parameter ϕ can be estimated using Pearson's chi-square statistic or Deviance divided by the degrees of freedom

```
MODEL y = x / SCALE=P;
MODEL y = x / SCALE=D;
```