1 Introduction

Many popular statistical methods are based on mathematical models that assume data follow a normal distribution. The most obvious among these are the analysis of variance for planned experiments and multiple regression for general analyses of independent and dependent variables. In many situations, the normality assumption is not plausible. Consequently, use of methods that assume normality may perform unsatisfactorily. In these cases, other alternatives that do not require data to have a normal distribution are attractive.

The collection of models called Generalized Linear Models (GLIMs) have become important, and practical, statistical tools. The basic idea of GLIMs is an adaption of standard regression to quite different kinds of data. The variables may be dichotomous (agree/disagree), categorical (as with a 5-point Likert scale), counts (number of arrest records), or nominal (choose among six candidates for mayor). The motivation is to tailor the regression relationship connecting the outcome to relevant independent variables so that it is appropriate to the properties of the dependent variable. The payoff is an analysis that often is more justifiable for the particular problem than a standard regression model would be.

The statistical theory and methods for fitting Generalized Linear Models (GLIMs) to simple random sample data are described in, amongst others, McCullach & Nelder (1989) and Agresti (2002). However, researchers from the social and economic sciences are often applying these methods to data from complex survey designs. Consequently, inappropriate results are obtained if these methods are applied to complex samples. For quite some time, these methods were extended to include the use of frequency and probability weights in an effort to deal with complex samples. Although this approach yields the appropriate estimates for complex samples, the corresponding standard error estimates are not appropriate. Using a result of Fuller (1975), Binder (1983) proposed methods to obtain the appropriate standard error estimates of the parameters of linear and nonlinear models as well as those of general estimating functions in the case of complex survey designs. These methods are implemented in, amongst others, SAS PROC SURVEYLOGISTIC (SAS Institute 2004) and AM (American Institutes for Research & Cohen 2004).

LISREL (Jöreskog & Sörbom 2006) includes a SurveyGLIM module, which implements the methods in Agresti (2002) and Binder (1983) to fit GLIMs to complex survey data and simple random sample data. Unlike other statistical software for generalized linear modeling for complex survey data such as SAS PROC SURVEYLOGISTIC and AM, LISREL allows for a wide variety of sampling distributions and link functions.

In this note, we provided illustrative examples of generalized linear modeling for counts, binary responses and ordinal responses and nominal responses.

2 GLIMs for counts

Variables measured in scientific studies come in a wide assortment. When statisticians refer to a "count" variable, they mean a variable that is ordinal, typically scored 0, 1, 2, ..., without fractional values such as 2.4 or 6.75. They also mean that the variable is a tally that records how often some behavior occurred, or of how many incidents of a particular kind were observed in each subject of a study.

In many situations, count variables are skewed. The percentage of subjects with a score of zero or 1 is very large, those with a score of 4 or 5 or 6 considerably less common, and those with a score of 11 or 12 rare. For example, the number of delinquent acts committed by a teenager is a count variable. It
is zero for the great majority. A young person who commits 1 or 2 or 3 delinquent acts is relatively rare compared to those who have no offenses. The frequencies of 1 or 2 or 3 decrease rapidly compared to those with no offenses. Juveniles who commit as many as 9 or 10 delinquent acts are very rare. As another example, the number of visits that a person makes to his or her primary care physician in a year is a count. The great majority visit the doctor not at all or once or twice in a year. Some may seek help 5, 6, or 7 times. A very few chronically ill may visit on as many as 15 occasions.

Count variables are often analyzed in exactly the same way that a continuous variable is handled, most often with a method that incorrectly assumes the count is a bell-shaped normal distribution. But counts are ordinal variables, usually skewed with a small range. They have none of the characteristics of a continuous variable. While in many instances there are few practical problems treating them as if they were continuous variables, it is easy to find examples where an inappropriate analysis of a count variable loses important information that a better approach would convey. GLIMs for counts are a special kind of model that is designed to represent the unique features of count variables in a statistically optimal way.

GLIMs for counts usually assume a Poisson distribution for the response variable. In this section, we illustrate the use of the SurveyGLIM module of LISREL 8.80 by using some practical examples based on health-related count data. More specifically, a Poisson-log model is fitted to substance abuse data. A description of the data follows.

3 The data

The data set forms part of the data library of the Alcohol and Drug Services Study (ADSS). The ADSS is a national study of substance abuse treatment facilities and clients. Background data and data on the substance abuse of a sample of 1752 clients were obtained. The sample was stratified by census region and within each stratum a sample was obtained for each of three facility treatment types within each of the four census regions. The specific data set is provided in the location C:\LISREL Student\SGLIMEX as the LSF cntdiag.lsf. The first portion of this file is shown in the following LSF window.

A brief description of the variables to be used in the subsequent GLIM analyses follows.

- CENREG is the census region of the client (1 for Northeast, 2 for Midwest, 3 for South and 4 for West).
FACTYPE is the facility treatment type of the client (1 for residential treatment, 2 for outpatient methadone treatment, 3 for outpatient non-methadone treatment and 4 for more than one type of treatment).

A2TWA0 is the design weight of the client.

cntdiag is the number of abuse diagnoses of the client (0, 1, 2 or 3).

sex is the value of a dummy variable for the gender (0 for male and 1 for female) of the client.

race_d is the value of a dummy variable for the race (0 for nonwhite and 1 for white) of the client.

More information on the ADSS and the data are available at http://www.icpsr.umich.edu.

4 The model

For this specific example, the mean model may be expressed as

\[ E[\text{cntdiag}_k] = \exp(\alpha + \beta_1 \cdot \text{sex}_k + \beta_2 \cdot \text{race}_d_k) \]

where \( E[\text{cntdiag}_k] \) denotes the mean number of diagnoses for client \( k \), \( \text{sex}_k \) and \( \text{race}_d_k \) denotes the values of the variables sex and race_d respectively and \( \alpha, \beta_1 \) and \( \beta_2 \) denote unknown parameters. From this model, it follows that the ratio of the mean numbers of diagnoses for female (\( \text{sex}_k = 1 \)) and male (\( \text{sex}_k = 0 \)) clients may be expressed as

\[ \frac{\exp(\alpha + \beta_1 + \beta_2 \cdot \text{race}_d)}{\exp(\alpha + \beta_1 \cdot \text{race}_d)} = \exp(\beta_1) \]

Similarly, it follows that \( \exp(\beta_2) \) is the ratio of the mean numbers of diagnoses for white and nonwhite clients. The model fitted value is a mean number of diagnoses for client \( k \) and is given by

\[ \hat{E}[\text{cntdiag}_k] = \exp(\hat{\alpha} + \hat{\beta}_1 \cdot \text{sex}_k + \hat{\beta}_2 \cdot \text{race}_d_k) \]

where \( \hat{\alpha}, \hat{\beta}_1 \) and \( \hat{\beta}_2 \) denote the maximum likelihood estimates of \( \alpha, \beta_1 \) and \( \beta_2 \) respectively.

5 Analyzing counts from a complex sampling design

A question that a researcher may want to address is whether ethnicity and gender effects are associated with the number of substance abuse diagnoses. An appropriate statistical model for this type of count variable is a GLIM with a Poisson distribution and a log link function.

5.1 Setting up the analysis

* Select the Open option on the File menu of the main window of LISWIN to load the Open dialog box.
* Select the Lisrel Data (*.lsf) option from the Files of type drop-down list box.
* Browse for the location C:\LISREL Student Examples\SGLIMEX.
* Select the file cntdiag.lsf by clicking on it.
* Click on the Open button to open the file cntdiag.lsf in a LSF window.
* Select the Title and Options option on the SurveyGLIM menu to load the Title and Options dialog box.
* Enter the string Poisson-Log Model for ADSS Data into the Title string box.
* Click the Next button to load the Distributions and Links dialog box.
* Select the Poisson option from the Distribution type drop-down list box.
* Select the Pearson option from the Estimate scale? drop-down list box.
* Click the Next button to load the Dependent and Independent Variables dialog box.
Specify the response variable cntdiag by selecting it from the Variables in data list box and clicking on the Add button of the Dependent variable section.

In a similar fashion, add the covariates sex and race_d to the Independent variables list box.

Click the Next button to load the Survey Design dialog box.

Select the variable CENREG by clicking on it.

Click on the Add button of the Stratification variable section.

Select the variable FACTYPE by clicking on it.

Click on the Add button of the Cluster variable section.

Select the variable A2TWA0 by clicking on it.

Click on the Finish button to open the following text editor window for cntdiag.pr2.

Click on the Run Prelis toolbar icon to produce the text editor window for cntdiag.out.

5.2 Discussion of results – Poisson-log model

A portion of the results of the Poisson-log GLIM analysis is shown in the following text editor window.

---

Note: The Wald F Test and Chi-square Statistics are statistics to test the null hypothesis that all the regression weights are equal to zero.

Estimated Regression Weights

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.3302</td>
<td>0.0597</td>
<td>5.9246</td>
<td>0.0000</td>
</tr>
<tr>
<td>sex</td>
<td>0.0619</td>
<td>0.0703</td>
<td>0.8728</td>
<td>0.3829</td>
</tr>
<tr>
<td>race_d</td>
<td>0.1167</td>
<td>0.0620</td>
<td>1.8918</td>
<td>0.0599</td>
</tr>
<tr>
<td>SCALE</td>
<td>0.7479</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The scale parameter estimate is based on the Pearson Chi-square value phi = Square Root of (The Pearson Chi-square value/degrees of freedom)
Both the values of the Wald $F$ and $\chi^2$ test statistics are not statistically significant if a significance level of 5% is used. Hence, there is insufficient evidence to conclude that both gender and race influence the number of diagnoses of a client. This finding is supported by the non-significant $z$ test statistic values for the significance of the individual parameters.

The scale parameter estimate is less than unity which indicates under-dispersion for the response variable. In other words, the sample variance of the variable cntdiag is less than its mean.

5.3 Estimated outcomes for different groups

The fitted model follows from the output file above as

$$\hat{E}[\text{cntdiag}] = \exp(0.33 + 0.06*\text{sex}_k + 0.12*\text{race}_{dk})$$

Although gender and race did not significantly affect the number of diagnoses, the following examples illustrate how the fitted model can be used to calculate the mean of number of diagnoses for various subgroups when there are statistically significant differences among them. This fitted model implies that the mean number of diagnoses for a white female client ($\text{sex}_k = 1$ and $\text{race}_k = 1$) is given by

$$\exp(0.33 + 0.06 + 0.12) = \exp(0.51) = 1.67$$

Similarly, the mean number of diagnoses for a nonwhite female client ($\text{sex}_k = 1$ and $\text{race}_k = 0$) is 1.48. It also follows from the output above that $\exp(\hat{\beta}_1) = \exp(0.06) = 1.06$ is the multiplicative effect of gender on the fitted number of diagnoses for a client. This implies that, on the average, female clients have a 6% higher estimated mean number of diagnoses than male clients. Similarly, it follows that $\exp(\hat{\beta}_2) = \exp(0.12) = 1.13$ which implies that, on the average, the fitted number of diagnoses is 13% higher for white clients than for nonwhite clients.

5.4 Ignoring stratification and clustering in the sample

Setting up the analysis

- Select the Survey Design option on the SurveyGLIM menu to load the Survey Design dialog box.
- Select the variable CENREG by clicking on it.
- Click on the Remove button of the Stratification variable section.
- Select the variable FACTYPE by clicking on it.
- Click on the Remove button of the Cluster variable section.
- Click on the Finish button to open the following text editor window for cntdiag.pr2.

$$\exp(0.06) = 1.06$$

Click on the Run Prelis toolbar icon to produce the text editor window for cntdiag.out.
5.5 Discussion of results

A portion of the text editor window for `cntdiag.out` is shown below.

![Estimated Regression Weights](cntdiag.OUT)

The results above indicate that although the parameter estimates are identical to those obtained when the design of the complex survey was taken into account, the standard error estimates are significantly smaller (cf. Brogan, 1998). As a consequence, both gender and race appear to have a statistically significant effect on the number of substance abuse diagnoses at a $p < 0.05$ level of confidence. This is a reversal of the results obtained when the complex sampling design was taken into account. As this example indicates, inferences based on an analysis that does not correct for the reduced precision of a complex sampling design can be very misleading.

6 GLIMs for binary response variables

Binary response variables are often the focus of empirical studies. Examples of binary response variables are diagnosis of breast cancer (absent or present), heart disease (yes or no), damage to solid rocket booster joints (damage or no damage), and depression in substance abuse clients (yes or no), credit risk (good or bad), etc. The analysis of GLIMs with binary response variables with SurveyGLIM is illustrated in this section. More specifically, a Bernoulli-logit model is fitted to substance abuse data.

SurveyGLIM can also fit models with binary response variables to either simple random sample or complex sample data. This feature is illustrated in this section by fitting Bernoulli-logit model to substance abuse data.

6.1 The data

The data set forms part of the data library of the Alcohol and Drug Services Study and is described in section 3.4.1. The data set to be analyzed consists of the complete cases for a selection of variables and is provided as the LSF `abuse1.lsf` in the location `C:\LISREL Student\SGLIMEX`. The first portion of this data set is shown in the following LSF window.
The variables to be used in the subsequent GLIM analyses are

- **CENREG** is the census region of the client (1 for Northeast, 2 for Midwest, 3 for South and 4 for West).
- **FACTYPE** is the facility treatment type of the client (1 for residential treatment, 2 for outpatient methadone treatment, 3 for outpatient non-methadone treatment and 4 for more than one type of treatment).
- **A2TWA0** is the design weight of the client.
- **depr** is the value of a dummy variable for the depression status (0 for no depression history and 1 for a history of depression) of the client.
- **sex** is the value of a dummy variable for the gender (0 for male and 1 for female) of the client.
- **race_d** is the value of a dummy variable for the race (0 for nonwhite and 1 for white) of the client.
6.2 The model

The probability model for the specific Bernoulli-logit GLIM is given by

$$P(\text{depr}_k = 1) = \frac{\exp(\alpha + \beta_1 \cdot \text{sex}_k + \beta_2 \cdot \text{race}_d)}{1 + \exp(\alpha + \beta_1 \cdot \text{sex}_k + \beta_2 \cdot \text{race}_d)}$$

where $P(\text{depr}_k = 1)$ denotes the probability that client $k$ has a history of depression and $\alpha$, $\beta_1$ and $\beta_2$ denote unknown parameters. The ratio of the probabilities that a female client ($\text{sex}_k = 1$) and a male client ($\text{sex}_k = 0$) has a history of depression respectively follows as

$$\frac{\exp(\alpha + \beta_1 + \beta_2 \cdot \text{race}_d)}{1 + \exp(\alpha + \beta_2 \cdot \text{race}_d)} = \exp(\beta_1)$$

In a similar fashion, it follows that $\exp(\beta_2)$ is the ratio of the probabilities that a white client and a nonwhite client have a history of depression respectively. The corresponding estimated model follows as

$$\hat{P}(\text{depr}_k = 1) = \frac{\exp(\hat{\alpha} + \hat{\beta}_1 \cdot \text{sex}_k + \hat{\beta}_2 \cdot \text{race}_d)}{1 + \exp(\hat{\alpha} + \hat{\beta}_1 \cdot \text{sex}_k + \hat{\beta}_2 \cdot \text{race}_d)}$$

where $\hat{P}(\text{depr}_k = 1)$ denotes the estimated probability that client $k$ has a history of depression and $\hat{\alpha}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ denote the maximum likelihood estimates of $\alpha$, $\beta_1$ and $\beta_2$ respectively.

7 Analyzing binary outcomes from complex survey designs

To explore a potential link between depression and a respondent's gender and ethnicity, a GLIM with Bernoulli distribution and logit link function is fitted to the data described above. The Bernoulli distribution is used since the outcome variable, depr, is dichotomous (0 for no depression history and 1 for a history of depression).

7.1 Setting up the analysis

- Select the Open option on the File menu of the main window of LISWIN to load the Open dialog box.
- Select the Lisrel Data (*.lsl) option from the Files of type drop-down list box.
- Select the file abuse1.lsf by clicking on it.
- Click on the Open button to open the file abuse1.lsf in a LSF window.
- Select the Title and Options option on the SurveyGLIM menu to load the Title and Options dialog box.
- Enter the title A Bernoulli-Logit Model for ADSS Data into the Title string box.
- Select the Bernoulli option from the Distribution type drop-down list box.
- Click on the Next button to load the Dependent and Independent Variables dialog box.
- Select the variable depr by clicking on it.
- Click on the Add button in the Dependent variable section.
- Select the variables sex and race_d.
- Click on the Continuous button of the Independent variables section.
- Select the variable FACYPE by clicking on it.
- Click on the Add button of the Cluster variable section.
- Select the variable SEX by clicking on it.
- Select the variable RACE by clicking on it.
- Click on the Add button of the Stratification variables section.
- Select the variable CENREG by clicking on it.
Select the variable A2TWA0 by clicking on it.

Click on the Add button of the Weight variable section.

Click on the Finish button to open the following text editor window for abuse1.pr2.

Click on the Run Prelis toolbar icon to produce the text editor window for abuse1.out.

7.2 Discussion of results – Bernoulli-logit model

A portion of the output file abuse1.out is shown in the following text editor window.

The results above indicate that both the gender and the race of clients have a statistically significant influence on their depression status if a significance level if 5% is used. There is sufficient evidence to conclude that female clients (sex = 1) are more likely than male clients to have a depression history and that white clients (race_d = 1) are less likely than nonwhite clients to have a history of depression.

7.3 Estimated outcomes for different groups

The estimated model is obtained from the results above as

\[
\hat{P}(\text{depr}_i = 1) = \frac{\exp(-0.14 + 0.70*\text{sex}_i - 0.57*\text{race}_d_i)}{1 + \exp(-0.14 + 0.70*\text{sex}_i - 0.57*\text{race}_d_i)}.
\]
The estimated probability that a nonwhite female client \((\text{sex}_k = 1 \text{ and } \text{race}_d = 0)\) has a history of depression follows from this fitted model as

\[
\frac{\exp(-0.14 + 0.70)}{1 + \exp(-0.14 + 0.70)} = \frac{\exp(0.56)}{1 + \exp(0.56)} = 0.64
\]

Similarly, the estimated probability that a nonwhite male client has a history of depression follows as 0.47. From the results above, it follows that \(\exp(\hat{\beta}_1) = \exp(0.70) = 2.01\) which implies that female clients are twice as likely as male clients to have a history of depression. Similarly, \(\exp(\hat{\beta}_2) = \exp(-0.57) = 0.57\) implies that whites are 43% less likely than nonwhites to have a history of depression.

### 8 GLIMs for ordinal response variables

Researchers are often involved in studying ordinal response variables such as mental impairment (well, mild symptom formation, moderate symptom formation or impaired), patient satisfaction measured on a 5-point Likert scale, severity of lower back pain (none, mild, moderate or severe), arthritis improvement (none, some or marked), etc. In this section, we illustrate generalized linear modeling for ordinal response variables with SurveyGLIM. A cumulative logit model is fitted to substance abuse data.

#### 8.1 The data

The data set comes from part of the data library of the Alcohol and Drug Services Study and is described in section 3.4.1. The data set to be analyzed consists of the complete cases for a selection of variables and is provided as the PSF `cntdiag.lsf` in the location `C:\LISREL Student\SGLIMEX`. The first portion of this data set is shown in the following PSF window.

![PSF Window](cntdiag.PSF)

A brief description of the variables to be used in the subsequent GLIM analyses follows.

- **CENREG** is the census region of the client (1 for Northeast, 2 for Midwest, 3 for South and 4 for West).
- **FACTYPE** is the facility treatment type of the client (1 for residential treatment, 2 for outpatient methadone treatment, 3 for outpatient non-methadone treatment and 4 for more than one type of treatment).
- **A2TWA0** is the design weight of the client.
cntdiag is the number of abuse diagnoses of the client (0, 1, 2 or 3).
- sex is the value of a dummy variable for the gender (0 for male and 1 for female) of the client.
- race_d is the value of a dummy variable for the race (0 for nonwhite and 1 for white) of the client.

8.2 The model
The probability model for the specific cumulative logit model is given by

\[
P(cntdiag_k \leq l) = \frac{\exp(\alpha_l + \beta_1^l \cdot sex_k + \beta_2^l \cdot race_d_k)}{1 + \exp(\alpha_l + \beta_1^l \cdot sex_k + \beta_2^l \cdot race_d_k)} \quad l = 1, 2, 3
\]

where \( P(cntdiag_k \leq l) \) denotes the cumulative probability that category \( l \) was recorded for client \( k \) and \( \alpha_1, \alpha_2, \alpha_3, \beta_1 \) and \( \beta_2 \) denote unknown parameters. The specific probabilities for the each response category for client \( k \) for both these models may be obtained from the following expressions.

\[
P(cntdiag_k = 1) = P(cntdiag_k \leq 1)
P(cntdiag_k = 2) = P(cntdiag_k \leq 2) - P(cntdiag_k \leq 1)
P(cntdiag_k = 3) = P(cntdiag_k \leq 3) - P(cntdiag_k \leq 2).
\]

In the case of the cumulative logit model, the ratio of the odds in the first \( l \) categories for a female client \( (sex_k = 1) \) and a male client \( (sex_k = 0) \) respectively follows as

\[
\frac{\exp(\alpha_l + \beta_1^l \cdot race_d)}{\exp(\alpha_l + \beta_2^l \cdot race_d)} = \exp(\beta_l)
\]

Similarly, it follows that \( \exp(\beta_2) \) is the ratio of the odds for a white client and a nonwhite client respectively. The corresponding estimated probability models are given by

\[
\hat{P}(cntdiag_k \leq l) = \frac{\exp(\hat{\alpha}_l + \hat{\beta}_1^l \cdot sex_k + \hat{\beta}_2^l \cdot race_d_k)}{1 + \exp(\hat{\alpha}_l + \hat{\beta}_1^l \cdot sex_k + \hat{\beta}_2^l \cdot race_d_k)} \quad l = 1, 2, 3
\]

where \( \hat{P}(cntdiag_k \leq l) \) denotes the estimated cumulative probability that at most the number of diagnoses listed in the first \( l \) categories are recorded for client \( k \) and \( \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\beta}_1 \) and \( \hat{\beta}_2 \) denote the maximum likelihood estimates of \( \alpha_1, \alpha_2, \alpha_3, \beta_1 \) and \( \beta_2 \) respectively.

9 Analyzing ordinal outcomes from complex survey designs
In a previous example, a GLIM with a Poisson distribution and a log link function was used to examine the possible association between ethnicity and gender effects and the number of substance abuse diagnoses (cntdiag). Since this variable assumes values between 0 and 3 in the sample data, an alternative approach is to examine the strength of the relationship between the predictors and the cumulative number of diagnoses. A GLIM with a multinomial distribution and a cumulative logit link function may be used for this purpose.

9.1 Setting up the analysis
- Select the Open option on the File menu of the main window of LISWIN to load the Open dialog box.
- Select the Lisrel Data (*.lis) option from the Files of type drop-down list box.
- Browse for the location C:\LISREL Student Examples\SGLIMEX.
- Select the file cntdiag.lsf by clicking on it.
Click on the **Open** button to open the file `cntdiag.lsf` in a LSF window.

**Select the Title and Options option on the SurveyGLIM menu to load the Title and Options dialog box.**

**Enter the title A cumulative logit model into the Title string box.**

**Click on the Next button to load the Distributions and Links dialog box.**

**Select the Multinomial option from the Distribution type drop-down list box.**

**Select the Ordinal logit option from the Link function drop-down list box.**

**Click on the Next button to load the Dependent and Independent Variables dialog box.**

**Select the variable `cntdiag` by clicking on it.**

**Click on the Add button in the Dependent variable section.**

**Select the variables `sex` and `race_d`.**

**Click on the Continuous button of the Independent variables section.**

**Click on the Next button to load the Survey Design dialog box.**

**Select the variable CENREG by clicking on it.**

**Click on the Add button of the Stratification variable section.**

**Select the variable FACTYPE by clicking on it.**

**Click on the Add button of the Cluster variable section.**

**Select the variable A2TWA0 by clicking on it.**

**Click on the Add button of the Weight variable section.**

**Click on the Finish button to open the following text editor window for `cntdiag.pr2`.**

![Image](cntdiag.pr2)

Click on the **Run Prelis** toolbar icon to produce the text editor window for `cntdiag.out`.

### 9.2 Discussion of results – Cumulative-logit model

A portion of the output file `cntdiag.out` is shown in the following text editor window.
At a 5% level of significance the results above indicate that there is insufficient evidence that gender and race affect the cumulative probabilities of the number of diagnoses of clients. Although the results for race_d border on statistical significance, interpreting the test of the parameter estimate precisely is consistent with the non-significance of the omnibus test of the model (see the Wald F-test and Wald $\chi^2$-statistic).

### 9.3 Estimated outcomes for different groups

Since $\hat{\alpha}_1 = -1.69$, the estimated probability that a white female client ($race_k = 1$ and $sex_k = 1$) has no diagnoses follows from the results above as

$$P(cntdiag_k = 1) = \frac{\exp(-1.69 - 0.20 - 0.39)}{1 + \exp(-1.69 - 0.20 - 0.39)} = 0.09$$

Similarly, the estimated probabilities that a white female client has at most 1 diagnosis and 2 diagnoses follow as 0.44 and 0.79 respectively. These estimated cumulative probabilities imply that the estimated probabilities that a white female client has 1 diagnosis, 2 diagnoses and 3 diagnoses are 0.44 - 0.09 = 0.35, 0.79 - 0.44 = 0.35 and 1 - 0.09 - 0.35 = 0.21 respectively. The effect estimates, $\hat{\beta}_1 = -0.20$ and $\hat{\beta}_2 = -0.39$, suggest that the cumulative probability starting at the no diagnoses end of the scale decreases for both females and whites. Given the race of a client, the estimated probability of a number of diagnoses below any level for a female client is $\exp(-0.20) = 0.82$ times the estimated probability for a male client. Similarly, given the gender of a client, the estimated probability of a number of diagnoses below any level for a white client is $\exp(-0.39) = 0.68$ times the estimated probability for a nonwhite client.

### 10 References


