3 Generalized Linear Models (GLMs)

0 Introduction

• In a simple linear regression model for continuous $Y$:

$$Y = \alpha + \beta x + \varepsilon,$$

usually $\varepsilon \sim iid \; N(0, \sigma^2)$.

$Y$ = response

$x$ = (numeric) covariate, indep or explanatory variable

$\beta = E(Y|x + 1) - E(Y|x)$

$2\beta = E(Y|x + 2) - E(Y|x)$, etc.

$\beta$ catches the linear relationship between $X$ and $Y$.

When $\beta = 0$, there is no linear relationship between $X$ and $Y$. 
• Given data \((x_i, y_i), i = 1, 2, \ldots, n\), we can estimate \(\alpha, \beta\), and hence \(E(Y|x)\). A common method to estimate \(\alpha, \beta\) is least squares (LS) by minimizing the following sum of squares (SS)

\[
\sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2.
\]

• Minimizing \(\sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \Rightarrow \)

\[
\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \\
\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}
\]

where \(\bar{x}\) is the sample mean of \(\{x_i\}\)'s, \(\bar{y}\) is the sample mean of \(\{y_i\}\)'s.

• \(\hat{\alpha}, \hat{\beta}\) have good statistical properties.

• Normality is **Not** required for the LS estimation.
A Description of the Linear Model
• Under $\varepsilon \overset{iid}{\sim} N(0, \sigma^2)$ (so $Y$ is also normal), the above model can be re-written as

$$Y|\mathbf{x} \overset{ind}{\sim} N(\alpha + \beta \mathbf{x}, \sigma^2),$$

or equivalently

$$Y|\mathbf{x} \overset{ind}{\sim} N(\mu(\mathbf{x}), \sigma^2), \quad \mu(\mathbf{x}) = \alpha + \beta \mathbf{x}$$

• MLE of $(\alpha, \beta) = \text{LSE of } (\alpha, \beta)$.

• Simple linear regression model can be extended to more than 1 covariate:

$$Y|\mathbf{x} \overset{ind}{\sim} N(\mu(\mathbf{x}), \sigma^2)$$

$$\mu(\mathbf{x}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$  

$\beta_k$: average change in $Y$ with one unit increase in $x_k$ while holding other covariates fixed (if $x_k$’s are unrelated variables)

• The above model can be easily extended to non-normal data $Y$.  

Slide 125
I Three Components of a GLM

- Data: \((x_i, y_i), i = 1, 2, \ldots, n\)

\[
y_i = \text{response}
\]

\[
x_i = (x_{1i}, x_{2i}, \ldots, x_{pi}) \text{ covariate, indep or explanatory variable}
\]

- A GLM has 3 components: random component, systematic component and the link function.

I.1 Random component

- Response \(Y\) is the random component of a GLM. We need to specify a distribution for \(Y\), such as normal, Bernoulli/Binomial or Poisson.

For the normal GLM, we specify the normal distribution for \(Y\).
I.2 Systematic component

- For covariates $x_1, x_2, \cdots, x_p$, form linear combination:

$$\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$ 

This linear combination is called the *systematic component* of a GLM. In a regression setting, the covariate values are viewed as fixed, hence the name of *systematic component*.

*Note*: we allow interactions such as $x_3 = x_1 x_2$, power functions such as $x_2 = x_1^2$ and other transformation for the covariates (*e.g.*, $x_2 = e^{x_1}$). In this case, we have to be careful in interpreting $\beta_k$'s.
I.3 Link function

- Denote $\mu = E(Y|x)$.

- With a smooth and monotone function $g(\mu)$, we relate $\mu$ and the systematic component via the formula:

$$g(\mu) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$ 

This function $g(\mu)$ is called the link function of a GLM.

- **Note:** Since both $\mu$ and the systematic component are both fixed quantities, there is NO error term in the above formula!

- Obviously, a normal GLM assumes

$$g(\mu) = \mu.$$ 

This link function is called the identity link.
• **Note:** In modelling the relationship between continuous response $Y$ and a covariate $x$, often time we would try to apply a transformation function $g(\cdot)$ to $Y$ so that $g(Y)$ may have a distribution closer to normal (even though normality is not necessary) and then fit

$$g(Y) = \alpha + \beta x + \epsilon.$$  

This is a *transformation model*.

A GLM with link function $g(\mu)$ ($\mu = E(Y | x)$)

$$g(\mu) = \alpha + \beta x$$

is **NOT** the same as the above transformation model, and we don’t apply the link function to the response $Y$!

Will see more later ...
I.4 Fitting and inference of a GLM

- Since we specify the distribution of $Y$, given data we use *Maximum Likelihood* (instead of *Least squares*) approach for *estimation* and *inference* on effect parameters $\beta_1, \ldots, \beta_p$.

- There is a unified algorithm for *estimation* and *inference*.

- Using Proc Genmod of SAS, we get the estimate, SE and p-value for testing $H_0 : \beta_k = 0$, etc.

```plaintext
proc genmod data=; * if y=1/0, then we need "descending" here;
   model y = x / dist= link=;
run;
```

The default distribution is normal with identity link. Common distributions are:

<table>
<thead>
<tr>
<th>Dist=</th>
<th>Distribution</th>
<th>Default Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binomial</td>
<td>Bin</td>
<td>B</td>
</tr>
<tr>
<td>Gamma</td>
<td>Gam</td>
<td>G</td>
</tr>
<tr>
<td>NegBin</td>
<td>NB</td>
<td>negative binomial</td>
</tr>
<tr>
<td>Normal</td>
<td>Nor</td>
<td>N</td>
</tr>
<tr>
<td>Poisson</td>
<td>Poi</td>
<td>P</td>
</tr>
</tbody>
</table>
If $y$ is binary (1/0) with 1 being the success (that is, we would like to model $P[Y = 1]$), we should use descending option in Proc Genmod.

For binomial response $y$ (of course, we should have $n$ – # of Bernoulli trials to get $y$), we have to use:
```sas
proc genmod data=;
   model y/n = x / dist=bin link=;
run;
```

Note: $y$ and $n$ are two variables in the data set. We don’t define a new variable $p = y/n$ and use “model $p = x$”. The / in y/n is just a symbol.

Data is organized in the same way as for Proc Reg of SAS.
II GLMs for Binary Response $Y$

- When the response $Y$ is binary (1/0, 1=success, 0=failure):

  $$\mu = E(Y) = 1 \times P[Y = 1] + 0 \times P[Y = 0] = P[Y = 1] = \pi$$

  is the success probability.

- A GLM for binary $Y$ with link function $g(\cdot)$ relates $\pi$ to the systematic component in the following:

  $$g(\pi) = \alpha + \beta x.$$

- Different choice of the link function $g(\pi)$ leads to a different binary GLM.
II.1 Linear probability model

- If we choose the link function $g(\cdot)$ to be the identity link $g(\pi) = \pi$, then we have a linear probability model:

$$\pi = \alpha + \beta x.$$ 

- Linear probability model is reasonable only if $\alpha + \beta x$ yields values in $(0,1)$ for valid values of $x$.

- $\beta$ has a nice interpretation:

$$\beta = \pi(x + 1) - \pi(x)$$

risk difference when $x$ increases by one unit.

- When the linear probability fits the data well, we can also use LS to make inference on $\beta$. The LS & ML estimation and inference will be similar.

Testing $H_0 : \beta = 0$ under this model is basically the same as the Cochran-Armitage trend test.
• Inference for the risk difference in a $2 \times 2$ table can be achieved using the linear probability model:

\[
\begin{array}{cccc}
Y & 1 & 0 \\
X & 1 & y_1 & n_1 - y_1 & n_1 \\
0 & y_2 & n_2 - y_2 & n_2 \\
\end{array}
\]

Let $\pi_1 = P[Y = 1|X = 1]$, $\pi_0 = P[Y = 1|X = 0]$, and we would like to make inference in $\phi = \pi_1 - \pi_0$, the risk difference between row 1 ($X = 1$) and row 2 ($X = 0$).

We can fit the following linear probability model to the above table

$$\pi = \alpha + \beta x.$$ 

Then $\beta$ is the same as $\phi$. 
• SAS program for making inference on risk difference for a 2 × 2 table:

```sas
data main;
  input x y n;
  1 * *
  0 * *
;
proc genmod;
  model y/n = x / dist=bin link=identity;
run;
```

• Output would look like:

```
Analysis Of Maximum Likelihood Parameter Estimates

Parameter DF Estimate Standard Error Wald 95% Confidence Limits Wald Chi-Square
Intercept 1 * * * *
X 1 * * * *
Scale 0 1.0000 0.0000 1.0000 1.0000 *
```
• Snoring and Heart Disease Example (Table 3.1 on p. 69)

<table>
<thead>
<tr>
<th>x</th>
<th>Heart Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes ($y_i$)</td>
</tr>
<tr>
<td>0 Never</td>
<td>24</td>
</tr>
<tr>
<td>Snoring</td>
<td></td>
</tr>
<tr>
<td>2 Occasionally</td>
<td>35</td>
</tr>
<tr>
<td>4 Nearly every night</td>
<td>21</td>
</tr>
<tr>
<td>5 Every night</td>
<td>30</td>
</tr>
</tbody>
</table>

• After assigning scores $x_i$: 0, 2, 4, 5 to snoring, we can calculate the sample proportions $p_i$ for each snoring level and plot $p_i$ against $x_i$ to see if linear probability model is reasonable.
SAS program and Part of its output:

```sas
data table3_1;
  input snoring score y y0;
  n = y+y0;
  p = y/n;
  logitp = log(p/(1-p));
datalines;
  0 0 24 1355
  1 2 35 603
  2 4 21 192
  3 5 30 224;

title "Snoring and heart disease data using class variable with identity link";
proc genmod;
class snoring;
model y/n = snoring / dist=bin link=identity noint;
estimate "level 1 - level 0" snoring -1 1 0 0;
estimate "level 2 - level 1" snoring 0 -1 1 0;
estimate "level 3 - level 2" snoring 0 0 -1 1;
run;

title "Sample proportion vs score";
proc plot;
  plot p*score;
run;

title "Sample logit vs score";
proc plot;
  plot logitp*score;
run;
```
The GENMOD Procedure

Contrast Estimate Results

<table>
<thead>
<tr>
<th>Label</th>
<th>Mean Estimate</th>
<th>Mean Confidence Limits</th>
<th>L’Beta Estimate</th>
<th>Standard Error</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>level 1 - level 0</td>
<td>0.0375</td>
<td>0.0185</td>
<td>0.0564</td>
<td>0.0375</td>
<td>0.0097</td>
</tr>
<tr>
<td>level 2 - level 1</td>
<td>0.0437</td>
<td>-0.0000</td>
<td>0.0875</td>
<td>0.0437</td>
<td>0.0223</td>
</tr>
<tr>
<td>level 3 - level 2</td>
<td>0.0195</td>
<td>-0.0369</td>
<td>0.0759</td>
<td>0.0195</td>
<td>0.0288</td>
</tr>
</tbody>
</table>

Sample proportion vs score 11

Plot of p*score. Legend: A = 1 obs, B = 2 obs, etc.
• The plots indicates linear probability model with the chosen scores for snoring may fit the data well (good choice of snoring scores).

• Consider linear probability model:

\[
\pi = \alpha + \beta x,
\]

where \( x \) is the snoring score.

• SAS program:

```sas
title "Snoring and heart disease data using score with identity link";
proc genmod;
  model y/n = score / dist=bin link=identity;
run;
```
• SAS output:

**************************************************************************
Snoring and heart disease data using score with identity link  13

The GENMOD Procedure

Model Information

Data Set WORK.TABLE3_1
Distribution Binomial
Link Function Identity
Response Variable (Events) y
Response Variable (Trials) n

Number of Observations Read  4
Number of Observations Used  4
Number of Events  110
Number of Trials  2484

Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>Binary Value</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Event</td>
<td>110</td>
</tr>
<tr>
<td>2</td>
<td>Nonevent</td>
<td>2374</td>
</tr>
</tbody>
</table>
### Criteria For Assessing Goodness Of Fit

<table>
<thead>
<tr>
<th>Criteria</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>2</td>
<td>0.0692</td>
<td>0.0346</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>2</td>
<td>0.0692</td>
<td>0.0346</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>2</td>
<td>0.0688</td>
<td>0.0344</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>2</td>
<td>0.0688</td>
<td>0.0344</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>-417.4960</td>
<td></td>
</tr>
<tr>
<td>Full Log Likelihood</td>
<td></td>
<td>-10.1609</td>
<td></td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>2</td>
<td>24.3217</td>
<td></td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>2</td>
<td>36.3217</td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>2</td>
<td>23.0943</td>
<td></td>
</tr>
</tbody>
</table>

### Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.0172</td>
<td>0.0034</td>
<td>0.0105, 0.0240</td>
<td>25.18</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.0198</td>
<td>0.0028</td>
<td>0.0143, 0.0253</td>
<td>49.97</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000, 1.0000</td>
<td></td>
</tr>
</tbody>
</table>

- The fitted model is

\[
\hat{\pi} = 0.017 + 0.0198x, x = 0, 2, 4, 5
\]
• From the fitted model, we can calculate the estimated heart disease probability for each level of snoring:

<table>
<thead>
<tr>
<th>Snoring($x$)</th>
<th>Heart Disease</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes ($y_i$)</td>
<td>No</td>
</tr>
<tr>
<td>0 Never</td>
<td>24</td>
<td>1355</td>
</tr>
<tr>
<td>2 Occasionally</td>
<td>35</td>
<td>605</td>
</tr>
<tr>
<td>4 Nearly every night</td>
<td>21</td>
<td>192</td>
</tr>
<tr>
<td>5 Every night</td>
<td>30</td>
<td>224</td>
</tr>
</tbody>
</table>

Since the fitted values $\hat{\pi} \approx p_i$, the linear probability model fits the data well.

• The model has a nice interpretation: For non-snorers, the heart disease prob is 0.017 (the intercept). For occasional snorers, the HD prob increases 0.04 (more than double), etc.
• **Note**: We can recover the original binary data (1/0 – called \(hd\) in the new data set) with 1 for heart disease, and use the following program to get exactly the same results:

```plaintext
title "Snoring and binary heart disease in proc genmod";
proc genmod descending;
    model hd = score / dist=bin link=identity;
run;
```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.0172</td>
<td>0.0034</td>
<td>0.0105 0.0240</td>
<td>25.18</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.0198</td>
<td>0.0028</td>
<td>0.0143 0.0253</td>
<td>49.97</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td></td>
</tr>
</tbody>
</table>

Without the option descending, Proc Genmod models \(P[Y = 0] = 1 - \pi\):

\[
1 - \pi = 1 - \alpha - \beta x.
\]

Therefore, if we don’t use the option descending, the intercept estimate will be equal to \(1 - 0.0172 = 0.9828\), and the estimate for the coefficient of snoring score \((x)\) will be \(-0.0198\).
We can also fit a linear regression model to the binary data and will get similar results.

```r
proc reg;
  model hd = score;
run;
```

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t|
|----------|----|--------------------|----------------|---------|------|
| Intercept| 1  | 0.01687            | 0.00516        | 3.27    | 0.0011 |
| score    | 1  | 0.02004            | 0.00232        | 8.65    | <.0001 |

**Note:** Since `proc reg` models $E(Y) = \pi$, the above results should be similar to the linear prob model with the option `descending` (if binary response data is used).
II.2 Logistic regression model

- For binary response $Y$, if we take the link function $g(\pi)$ in the GLM as

\[ g(\pi) = \logit(\pi) = \log \left( \frac{\pi}{1 - \pi} \right), \]

then we have a logistic regression model:

\[ \logit(\pi) = \alpha + \beta x. \]

Here the function $g(\pi) = \logit(\pi) = \log\{\pi/(1 - \pi)\} = \log(\text{odds})$ is called the logit function of $\pi$. Note that with this link, any $x$ and $\alpha, \beta$ will yield a valid $\pi$:

\[ \pi(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}. \]

- With a fitted logistic regression, the estimated prob at $x$ is given by

\[ \hat{\pi}(x) = \frac{e^{\hat{\alpha} + \hat{\beta} x}}{1 + e^{\hat{\alpha} + \hat{\beta} x}}. \]
Figure 3.2. Logistic regression functions.
• Interpretation of $\beta$:

$\pi$ at $x$:

$$\log \frac{\pi(x)}{1 - \pi(x)} = \alpha + \beta x$$

$\pi$ at $x + 1$:

$$\log \frac{\pi(x + 1)}{1 - \pi(x + 1)} = \alpha + \beta(x + 1)$$

$$\log \frac{\pi(x + 1)}{1 - \pi(x + 1)} - \log \frac{\pi(x)}{1 - \pi(x)} = \beta$$

$$\beta = \log \left\{ \frac{\pi(x + 1)/\{1 - \pi(x + 1)\}}{\pi(x)/\{1 - \pi(x)\}} \right\}$$

$$e^\beta = \frac{\pi(x + 1)/\{1 - \pi(x + 1)\}}{\pi(x)/\{1 - \pi(x)\}}$$

odds-ratio with one unit increase in $x$

$$\Rightarrow 2\beta = \log \left\{ \frac{\pi(x + 2)/\{1 - \pi(x + 2)\}}{\pi(x)/\{1 - \pi(x)\}} \right\}$$

log odds-ratio with two unit increase in $x$, etc.
Inference for the odds-ratio in a $2 \times 2$ table can be achieved using the logistic regression model:

$$
\begin{array}{c|cc|c}
 & 1 & 0 & \ \ \\
\hline
X & 1 & y_1 & n_1 - y_1 & n_1 \\
   & 0 & y_2 & n_2 - y_2 & n_2 \\
\end{array}
$$

Let $\pi_1 = P[Y = 1|x = 1]$, $\pi_0 = P[Y = 1|x = 0]$, and we would like to make inference on $\theta = \frac{\pi_1/(1-\pi_1)}{\pi_0/(1-\pi_0)}$, the odds-ratio between row 1 and row 2.

We can fit the following logistic regression model:

$$\text{logit}(\pi) = \alpha + \beta x.$$ 

Since $x$ can only take 0 and 1, $e^\beta = \theta$ is the odds-ratio of interest.

Testing $H_0 : \beta = 0 \iff H_0 : X \perp Y$. 

---

Slide 148
• SAS program for making inference on odds ratio for a $2 \times 2$ table:

```sas
data main;
  input x y n;
  1 * *
  0 * *
;
proc genmod;
  model y/n = x / dist=bin link=logit;
run;
```

• Output would look like:

```
Analysis Of Maximum Likelihood Parameter Estimates

Parameter  DF  Estimate  Standard Error  Wald 95% Confidence Limits  Wald Chi-Square
Intercept  1    *        *            *        *            *            *
X          1    *        *            *        *            *            *
Scale      0    1.0000   0.0000       1.0000   1.0000
```
• Logistic regression model for Snoring and Heart Disease Example.

If there is a nearly straight line in the plot of sample logit against $x$ indicates a good fit of the logistic regression:

$$\text{sample logit} = \log \frac{p_i}{1 - p_i}.$$

Sample logit vs score

Plot of logitp*score. Legend: A = 1 obs, B = 2 obs, etc.
title "Snoring and heart disease data using score with logit link";
proc genmod;
    model y/n = score / dist=bin link=logit;
run;
**************************************************************************
Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-3.8662</td>
<td>0.1662</td>
<td>-4.1920, -3.5405</td>
<td>541.06</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.3973</td>
<td>0.0500</td>
<td>0.2993, 0.4954</td>
<td>63.12</td>
</tr>
</tbody>
</table>

- Comparison of estimated probs:

<table>
<thead>
<tr>
<th>Snoring($x$)</th>
<th>Heart Disease</th>
<th>Linear</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes ($y_i$)</td>
<td>No</td>
<td>$n_i$</td>
</tr>
<tr>
<td>0 Never</td>
<td>24</td>
<td>1355</td>
<td>1379</td>
</tr>
<tr>
<td>2 Occasionally</td>
<td>35</td>
<td>605</td>
<td>638</td>
</tr>
<tr>
<td>4 Nearly every night</td>
<td>21</td>
<td>192</td>
<td>213</td>
</tr>
<tr>
<td>5 Every night</td>
<td>30</td>
<td>224</td>
<td>254</td>
</tr>
</tbody>
</table>

⇒ Linear prob model is better than the logistic model.
• We can also use the original binary response $hd$ and use the following SAS program with descending option and will get the same results.

```
title "Snoring and heart disease data using score with logit link";
proc genmod descending;
    model hd = score / dist=bin link=logit;
run;
```

****************************************************************************

**Analysis Of Maximum Likelihood Parameter Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-3.8662</td>
<td>0.1662</td>
<td>-4.1920, -3.5405</td>
<td>541.06</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.3973</td>
<td>0.0500</td>
<td>0.2993, 0.4954</td>
<td>63.12</td>
</tr>
</tbody>
</table>

• **Note**: if we don’t use the option descending, then we are modeling $P[Y = 0] = 1 - \pi = \tau$. If the original logistic model for $\pi$ is true, then we also have a logistic model for $\tau$:

$$\log \left( \frac{\tau}{1-\tau} \right) = \log \left( \frac{1-\pi}{\pi} \right) = -\log \left( \frac{\pi}{1-\pi} \right) = -\alpha - \beta x.$$ 

Therefore, all estimates will be the mirror image of those from the previous logistic model.
II.3 Log linear probability model

- For binary response $Y$, if we take the link function $g(\pi)$ in the GLM as the log function, then we have a log-linear probability model:

$$\log(\pi) = \alpha + \beta x.$$ 

- Given $x$ and $\alpha, \beta$, solving for $\pi$ we have:

$$\pi = e^{\alpha + \beta x}.$$ 

Of course, the model is only reasonable if the model produces valid $\pi$’s in (0,1) for $x$ in the valid range.
• Interpretation of $\beta$:

$$\log \pi(x) = \alpha + \beta x$$

$$\log \pi(x + 1) = \alpha + \beta(x + 1)$$

$$\log \pi(x + 1) - \log \pi(x) = \beta$$

$$\beta = \log \left\{ \frac{\pi(x + 1)}{\pi(x)} \right\}$$

$$e^\beta = \frac{\pi(x + 1)}{\pi(x)}$$

RR with one unit increase in $x$

$$\Rightarrow e^{2\beta} = \frac{\pi(x + 2)}{\pi(x)}$$

RR with two unit increase in $x$
• Inference for the RR in a $2 \times 2$ table can be achieved using the log-linear probability model:

\[
\begin{array}{cccc}
 & Y & \\
X & 1 & 0 & \\
1 & y_1 & n_1 - y_1 & n_1 \\
0 & y_2 & n_2 - y_2 & n_2 \\
\end{array}
\]

Let $\pi_1 = P[Y = 1|x = 1]$, $\pi_0 = P[Y = 1|x = 0]$, and we would like to make inference on $RR = \frac{\pi_1}{\pi_0}$, the relative risk between row 1 and row 2. We can fit the following log-linear probability model:

\[
\log(\pi) = \alpha + \beta x.
\]

Since $x$ can only take 0 and 1, $e^\beta$ is the RR of interest.

Testing $H_0 : \beta = 0 \Leftrightarrow H_0 : X \perp Y$. 
• SAS program for making inference on relative risk for a $2 \times 2$ table:

```sas
data main;
  input x y n;
  1 * *;
  0 * *
;
proc genmod;
  model y/n = x / dist=bin link=log;
run;
```

• Output would look like:

```
Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>X</td>
<td>1</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
```
II.4 Probit regression model

- For binary response $Y$, if we take the link function in the GLM as 
  $g(\pi) = \Phi^{-1}(\pi)$, the inverse of the cumulative distribution function (cdf) of N(0,1), then we have a probit regression model

  $$\Phi^{-1}(\pi) = \alpha + \beta x.$$ 

- For any $x$ and $\alpha, \beta$, the model yields valid $\pi$:

  $$\pi = \Phi(\alpha + \beta x).$$

- A probit model is very similar to a logistic regression. That is, if 
  $$\Phi^{-1}\{\pi(x)\} = \alpha + \beta x$$
  is true, then 
  $$\text{logit}\{\pi(x)\} \approx \alpha^* + \beta^* x$$
  with $\alpha^* = 1.7\alpha$ and $\beta^* = 1.7\beta$. However, the fitted probs from these 2 models will be similar.
For the Snoring/Heart Disease example, the fitted results:

```plaintext
title "Snoring and heart disease data using score with probit link";
proc genmod;
   model y/n = score / dist=bin link=probit;
run;
```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-2.0606</td>
<td>0.0704</td>
<td>-2.1986 -1.9225</td>
<td>855.49</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.1878</td>
<td>0.0236</td>
<td>0.1415 0.2341</td>
<td>63.14</td>
</tr>
</tbody>
</table>

\[ \hat{\pi}(x) = \Phi(-2.0606 + 0.1878x) \].

For example, when \( x = 2 \) (occasional snorers), \( \hat{\pi}(x) \) is:

\[ \hat{\pi}(2) = \Phi(-2.0606+0.1878 \times 2) = \Phi(-1.685) = P[Z \leq -1.685] = 0.046. \]

**Note:** \( 1.7 \times (-2.0606) = -3.5, 1.7 \times 0.1878 = 0.32 \), very close to the estimates from the logistic model.
We can also use the original binary response hd and use the following SAS program with descending option and will get the same results.

title "Snoring and heart disease data using score with logit link";
proc genmod descending;
   model hd = score / dist=bin link=probit;
run;

**************************************************************************
Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-2.0606</td>
<td>0.0704</td>
<td>-2.1986 -1.9225</td>
<td>855.49</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.1878</td>
<td>0.0236</td>
<td>0.1415 0.2341</td>
<td>63.14</td>
</tr>
</tbody>
</table>

Note: if we don’t use the descending option, then we are modeling \( P[Y = 0] = 1 - \pi = \tau \). If the original probit model for \( \pi \) is true, then we also have a probit model for \( \tau \):

\[
\Phi^{-1}(\tau) = \Phi^{-1}(1 - \pi) = -\Phi^{-1}(\pi) = -\alpha - \beta x.
\]

Therefore, all estimates will be the mirror image of those from the previous probit model.
• Comparison of estimated probs from 3 models:

<table>
<thead>
<tr>
<th>Snoring ((x))</th>
<th>Heart Disease</th>
<th></th>
<th></th>
<th>Linear</th>
<th>Logit</th>
<th>Probit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes ((y_i))</td>
<td>24</td>
<td>1355</td>
<td>1379</td>
<td>0.017</td>
<td>0.017</td>
<td>0.021</td>
</tr>
<tr>
<td>No</td>
<td>2</td>
<td>Occasionally</td>
<td>35</td>
<td>605</td>
<td>638</td>
<td>0.055</td>
</tr>
<tr>
<td>Nearly every night</td>
<td>21</td>
<td>192</td>
<td>213</td>
<td>0.099</td>
<td>0.096</td>
<td>0.099</td>
</tr>
<tr>
<td>Every night</td>
<td>30</td>
<td>224</td>
<td>254</td>
<td>0.118</td>
<td>0.116</td>
<td>0.132</td>
</tr>
</tbody>
</table>

⇒

1. Logistic model and probit model give very close predicted \(\pi\)'s.
2. Linear prob model is better than the logistic model.
Sample proportions and fitted $\pi$'s from 3 models

Figure 3.1. Fit of models for snoring and heart disease data.
III GLMs for Count Data

- In many applications, the response $Y$ is count data:
  1. Monthly # of car accidents on a particular highway.
  2. Yearly # of new cases of certain disease in counties over US, etc.

- For count data $Y$, a common distributional assumption is $Y \sim \text{Poisson}(\mu)$:
  $$E(Y) = \text{var}(Y) = \mu.$$

- A GLM for count data $Y$ usually uses log as the link function:
  $$\log(\mu) = \alpha + \beta x.$$

  $$\Rightarrow \mu(x) = e^{\alpha + \beta x}.$$

  Of course, other link functions, such as identity link, are also possible.

- Interpretation of $\beta$:
  $$e^\beta = \frac{\mu(x + 1)}{\mu(x)}, e^\beta - 1 = \text{percentage increase in } \mu \text{ with 1 unit increase in } x$$
III.1 Example: Female horseshoe crabs and their satellites (Table 3.2, page 76-77)
• Data (a subset):

    data crab;
    input color spine width satell weight;
        weight=weight/1000; color=color-1;
    datalines;
3 3 28.3 8 3050
4 3 22.5 0 1550
2 1 26.0 9 2300
4 3 24.8 0 2100
4 3 26.0 4 2600
3 3 23.8 0 2100
2 1 26.5 0 2350
4 2 24.7 0 1900
.
.
.

$y_i = \# \text{ of satellites (male crabs) for female crab } i$

$x_i = \text{ carapace width of female crab } i$

• Model the relationship between $\mu_i = E(Y_i|x_i)$ and $x_i$ using the log-linear model

    \[ \log(\mu_i) = \alpha + \beta x_i \]

assuming $Y_i \sim \text{Poisson}(\mu_i)$. 
• SAS Program and output:

```sas
title "Analysis of crab data using Poisson distribution";
title2 "(without overdispersion) with log link";
proc genmod data=crab;
   model satell = width / dist=poi link=log;
run;
```

**************************************************************************

Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-3.3048</td>
<td>0.5422</td>
<td>-4.3675 -2.2420</td>
<td>37.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.1640</td>
<td>0.0200</td>
<td>0.1249 0.2032</td>
<td>67.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{\mu}(x) = e^{-3.3048+0.1640x} \]

\[ \hat{\beta} = 0.1640 \text{ with } \hat{SE}(\hat{\beta}_1) = 0.02, \text{ p-value } < 0.0001. \]

However, the inference may not be valid since the count data \( Y \) often has an over-dispersion issue:

\[ \text{var}(Y) > \text{E}(Y). \]
### III.2 Over-dispersion in count data

- **Empirical check of over-dispersion:**

<table>
<thead>
<tr>
<th>Carapace width ((x))</th>
<th>Num. of Obs.</th>
<th>(\bar{y})</th>
<th>(S^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\leq 23.25)</td>
<td>14</td>
<td>1</td>
<td>2.77</td>
</tr>
<tr>
<td>23.25 - 24.25</td>
<td>14</td>
<td>1.43</td>
<td>8.88</td>
</tr>
<tr>
<td>24.25 - 25.25</td>
<td>28</td>
<td>2.39</td>
<td>6.54</td>
</tr>
<tr>
<td>25.25 - 26.25</td>
<td>39</td>
<td>2.69</td>
<td>11.38</td>
</tr>
<tr>
<td>26.25 - 27.25</td>
<td>22</td>
<td>2.86</td>
<td>6.88</td>
</tr>
<tr>
<td>27.25 - 28.25</td>
<td>24</td>
<td>3.87</td>
<td>8.81</td>
</tr>
<tr>
<td>28.25 - 29.25</td>
<td>18</td>
<td>3.94</td>
<td>16.88</td>
</tr>
<tr>
<td>(&gt; 29.25)</td>
<td>14</td>
<td>5.14</td>
<td>8.29</td>
</tr>
</tbody>
</table>

**Observation:** \(S^2 \gg \bar{y} \implies \text{var}(Y_i|x_i) > E(Y_i|x_i)\), over-dispersion!
• A common approach to take into account over-dispersion in inference is to assume the following variance-mean relationship for count data $Y$:

$$\text{var}(Y) = \phi \text{E}(Y),$$

$\phi$ — over-dispersion parameter.

• Estimation of $\phi$ using the *Pearson* statistic

$$\hat{\phi}_P = \frac{1}{df} \sum \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$$

This can be specified by `scale=pearson` or `scale=p` in Proc Genmod. A common choice.

• Estimation of $\phi$ using the Deviance statistic:

$$\hat{\phi}_D = \frac{2[\log(L_S) - \log(L_M)]}{df}$$

This can be specified by `scale=deviance` or `scale=d` in Proc Genmod.
- SAS program and output:

```sas
title "Analysis of crab data using overdispersed Poisson";
title2 "distribution with log link";
proc genmod data=crab;
  model satell = width / dist=poi link=log scale=pearson;
run;
```

********************************************************************************

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Conf. Limits</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-3.3048</td>
<td>0.9673</td>
<td>-5.2006 -1.4089</td>
<td>11.67</td>
<td>0.0006</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.1640</td>
<td>0.0356</td>
<td>0.0942 0.2339</td>
<td>21.22</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.7839</td>
<td>0.0000</td>
<td>1.7839 1.7839</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The scale parameter was estimated by the square root of Pearson Chi-Square/DOF.
• With the option scale=pearson, the Pearson estimate 
\[ \sqrt{\widehat{\phi}_P} = 1.7839, \] indicating a lot of over-dispersion.

• From the output, we see that we got the same estimates of \( \alpha \) and \( \beta \). However, their standard errors are inflated by \[ \sqrt{\widehat{\phi}} = 1.7839 \] (larger SE's).

• Based on the estimated model:

\[ \log(\mu) = -3.3048 + 0.1640x \]

\[ \Rightarrow \text{With } 1\text{cm increase in carapace width, the average # of satellites will increase by } e^{0.1640} - 1 = 0.18 = 18\%. \]
III.3 GLM for count data with other links

- Plot of smoothing of raw data indicates the identity link function:

![Graph showing smoothing of horseshoe crab counts.](image)

**Figure 3.5.** Smoothings of horseshoe crab counts.
• Consider the GLM with the identity link:

\[ \mu = \alpha + \beta x. \]

• SAS program and output:

```sas
title "Analysis of crab data using overdispersed Poisson";
title2 "distribution with identity link";
proc genmod data=crab;
   model satell = width / dist=poi link=identity scale=pearson;
run;
```

Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-11.5321</td>
<td>2.6902</td>
<td>-16.8048 -6.2593</td>
<td>18.38</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.5495</td>
<td>0.1056</td>
<td>0.3425 0.7565</td>
<td>27.07</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.7811</td>
<td>0.0000</td>
<td>1.7811 1.7811</td>
<td></td>
</tr>
</tbody>
</table>

⇒

1. A lot of over-dispersion: \( \hat{\phi}_P^{1/2} = 1.7811. \)
2. Significant evidence against \( H_0 : \beta = 0. \)
3. Fitted model: \( \hat{\mu} = -11.5321 + 0.5495x. \)
Comparison of GLMs with log and identity links

Figure 3.6. Estimated mean number of satellites for log and identity links.
III.4 Negative binomial for over-dispersed count data

- We can assume a negative-binomial distribution for count response $Y$ to automatically handle over-dispersion:

$$E(Y) = \mu, \quad \text{var}(Y) = \mu + D\mu^2,$$

where $D > 0$ is a positive parameter.

- Note: Suppose we have a Bernoulli process with success probability $\pi$ and we would continue the trial until we obtain $r$ successes. Let $Y =$ extra # of trial in order to achieve our goal, then the distribution of $Y$ is called a negative binomial with pmf

$$f(y) = \binom{y + r - 1}{r - 1} \pi^r (1 - \pi)^y, \quad y = 0, 1, 2, ...$$

$$\Rightarrow \quad E(Y) = \frac{r(1 - \pi)}{\pi}, \quad \text{var}(Y) = \frac{r(1 - \pi)}{\pi^2} = \mu + \frac{1}{r}\mu^2$$

In this case $D = 1/r$. 
• In the general negative binomial distribution, we can allow $r$ to be a non-integer. If $r \to \infty$, we have the Poisson distribution.

• The above distribution can be specified in SAS using `dist=negbin`.

• SAS program and output for the crab data:

```sas
title "Analysis of crab data using Negative Binomial distribution with log link";
proc genmod data=crab;
   model satell = width / dist=negbin link=log; * other links are possible;
run;
```

Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-4.0525</td>
<td>1.2642</td>
<td>-6.5303 -1.5747</td>
<td>10.28</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.1921</td>
<td>0.0476</td>
<td>0.0987 0.2854</td>
<td>16.27</td>
</tr>
<tr>
<td>Dispersion</td>
<td>1</td>
<td>1.1055</td>
<td>0.1971</td>
<td>0.7795 1.5679</td>
<td></td>
</tr>
</tbody>
</table>

⇒ 1. $\hat{D} = 1.1$.

2. Fitted model: $\log(\hat{\mu}) = -4.0525 + 0.1921x$. similar fit.

• **Note:** We don’t use the option `scale=`. There may be some computational issue with neg. bin. dist.
III.5 GLMs for rate data

- When the response $Y$ represents the number of events occurred over a time window with length $T$ or over a population with size $T$, etc, it may be more meaningful to model the rate data $R = Y/T$.

- Let $\mu = \mathbb{E}(Y)$. Then the expected rate $r = \mathbb{E}(R)$ is

  $$ r = \frac{\mu}{T}. $$

- If we assume a log-linear model for the rate $r$:

  $$ \log(r) = \alpha + \beta x, $$

  then the model for $\mu$ is

  $$ \log(\mu) = \log(T) + \alpha + \beta x. $$

The term $\log(T)$ is called an offset and can be specified using offset=logt if we define the variable $logt = \log(T)$. 

Slide 175
Example: British train accidents over time (Table 3.4, page 83):

<table>
<thead>
<tr>
<th>Year</th>
<th>Train-km</th>
<th>Train Collisions</th>
<th>Train-road Collisions</th>
<th>Year</th>
<th>Train-km</th>
<th>Train Collisions</th>
<th>Train-road Collisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>518</td>
<td>0</td>
<td>3</td>
<td>1988</td>
<td>443</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2002</td>
<td>516</td>
<td>1</td>
<td>3</td>
<td>1987</td>
<td>397</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2001</td>
<td>508</td>
<td>0</td>
<td>4</td>
<td>1986</td>
<td>414</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>2000</td>
<td>503</td>
<td>1</td>
<td>3</td>
<td>1985</td>
<td>418</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1999</td>
<td>505</td>
<td>1</td>
<td>2</td>
<td>1984</td>
<td>389</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>1998</td>
<td>487</td>
<td>0</td>
<td>4</td>
<td>1983</td>
<td>401</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>1997</td>
<td>463</td>
<td>1</td>
<td>1</td>
<td>1982</td>
<td>372</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1996</td>
<td>437</td>
<td>2</td>
<td>2</td>
<td>1981</td>
<td>417</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1995</td>
<td>423</td>
<td>1</td>
<td>2</td>
<td>1980</td>
<td>430</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1994</td>
<td>415</td>
<td>2</td>
<td>4</td>
<td>1979</td>
<td>426</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1993</td>
<td>425</td>
<td>0</td>
<td>4</td>
<td>1978</td>
<td>430</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1992</td>
<td>430</td>
<td>1</td>
<td>4</td>
<td>1977</td>
<td>425</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>1991</td>
<td>439</td>
<td>2</td>
<td>6</td>
<td>1976</td>
<td>426</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>1990</td>
<td>431</td>
<td>1</td>
<td>2</td>
<td>1975</td>
<td>436</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>1989</td>
<td>436</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: British Department of Transport.
\* \( y \) = yearly \# of train accidents with road vehicles from 1975-2003.
\* \( T \) = \# of train-KM’s.
\* \( x \) = \# of years since 1975.
\* Consider log-rate GLM:

\[
\log(\mu) = \log(T) + \alpha + \beta x.
\]

```
title "Analysis of British train accident data";
proc genmod data=train;
  model y = x / dist=poi link=log offset=logt scale=pearson;
run;
```

```
Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-4.2114</td>
<td>0.1987</td>
<td>-4.6008 -3.8221</td>
<td>449.41</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>year</td>
<td>1</td>
<td>-0.0329</td>
<td>0.0134</td>
<td>-0.0593 -0.0066</td>
<td>5.99</td>
<td>0.0144</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.2501</td>
<td>0.0000</td>
<td>1.2501 1.2501</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

\( \Rightarrow \log(\text{rate}) = -4.21 - 0.0329x \). Accidents decline over time.
\* \textbf{Note}: If we assume a different model for the expected rate $r$, we will have a different model for $\mu = \mathbb{E}(Y)$. The thing that matters is to find a model for $\mu = \mathbb{E}(Y)$.

For example, if we assume

$$\frac{1}{r} = \alpha + \beta x, \quad \Rightarrow \frac{T}{\mu} = \alpha + \beta x$$

$$\Rightarrow \quad \frac{1}{\mu} = \alpha(1/T) + \beta(x/T).$$

So the link function is $g(\mu) = \mu^{-1}$. If we define $t1$ for $1/T$ and $x1$ for $x/T$ in our data set, then we can use the following program to fit the above model:

```plaintext
proc genmod data=mydata;
  model y = t1 x1 / noint dist=poi link=power(-1) scale=pearson;
run;
```
IV Inference for GLM and Model Checking

IV.1 Inference for $\beta$ in a GLM

- After we fit a GLM, we can make inference on $\beta$ such as:

  - Wald test for $H_0 : \beta = 0$ v.s. $H_a : \beta \neq 0$:

    $$Z = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

    Compare $Z$ to $N(0,1)$ to get p-value (Note: $SE(\hat{\beta})$ has to be the correct SE, e.g. needs to account for over-dispersion).

  - LRT test for $H_0 : \beta = 0$ v.s. $H_a : \beta \neq 0$ with NO over-dispersion:

    $$G^2 = 2(\log L_1 - \log L_0),$$

    where $L_0$ is the maximum likelihood of model under $H_0$, $L_1$ is the maximum likelihood of model under $H_0 \cup H_a$.

    Compare $G^2$ to $\chi^2_1$.

    In order to construct the LRT, we need to fit two models, one
under $H_0$, one under $H_0 \cup H_a$.

**LRT test for $H_0 : \beta = 0$ v.s. $H_a : \beta \neq 0$ with over-dispersion:**

$$G^2 = \frac{2(\log L_1 - \log L_0)}{\hat{\phi}},$$

where $\hat{\phi}$ is the estimate $\phi$ under $H_0 \cup H_a$. Compare $G^2$ to $\chi^2_1$.

For the crab data:

```plaintext
proc genmod data=crab;
  model satell = / dist=poi link=log;
run;
```

Criteria For Assessing Goodness Of Fit

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>172</td>
<td>632.7917</td>
<td>3.6790</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>172</td>
<td>632.7917</td>
<td>3.6790</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>172</td>
<td>584.0436</td>
<td>3.3956</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>172</td>
<td>584.0436</td>
<td>3.3956</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>359</td>
<td>35.9898</td>
<td></td>
</tr>
<tr>
<td>Full Log Likelihood</td>
<td>-</td>
<td>-494.0447</td>
<td></td>
</tr>
</tbody>
</table>
proc genmod data=crab;
model satell = width / dist=poi link=log;
run;

*****************************************************************************

Criteria For Assessing Goodness Of Fit

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>171</td>
<td>567.8786</td>
<td>3.3209</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>171</td>
<td>567.8786</td>
<td>3.3209</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>171</td>
<td>544.1570</td>
<td>3.1822</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>171</td>
<td>544.1570</td>
<td>3.1822</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>68.4463</td>
<td></td>
</tr>
<tr>
<td>Full Log Likelihood</td>
<td></td>
<td>-461.5881</td>
<td></td>
</tr>
</tbody>
</table>

\[ G^2 = \frac{2(68.4463 - 35.9898)}{3.1822} = 20.2, \text{ compared to } \chi^2_1. \]
Construct a $(1 - \alpha)$ CI for $\beta$:

$$[\hat{\beta} - z_{\alpha/2}SE(\hat{\beta}), \hat{\beta} + z_{\alpha/2}SE(\hat{\beta})] = [\hat{\beta}_L, \hat{\beta}_U]$$

$\Rightarrow$ We can get a CI for functions of $\beta$.

For example, in a logistic regression, $e^{\beta}$ is the odds-ratio ($\theta$) of success with one unit increase of $x$. Then a $(1 - \alpha)$ CI for $e^{\beta} = \theta$:

$$[e^{\hat{\beta}_L}, e^{\hat{\beta}_U}]$$.
IV.2 Model checking

- In some situations, we can check to see if a GLM

\[ g(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p \]

fits the data well.

- **Conditions**: No over-dispersion (e.g. binary/binomial data), \# of unique values of \( x \) is fixed, \( n_i \to \infty \).

- Snoring/Heart disease example:

<table>
<thead>
<tr>
<th>( x )</th>
<th>Heart Disease</th>
<th>No</th>
<th>( n_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Never</td>
<td>Yes (( y_i ))</td>
<td>24</td>
<td>1355</td>
</tr>
<tr>
<td>Snoring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Occasionally</td>
<td></td>
<td>35</td>
<td>605</td>
</tr>
<tr>
<td>4 Nearly every night</td>
<td></td>
<td>21</td>
<td>192</td>
</tr>
<tr>
<td>5 Every night</td>
<td></td>
<td>30</td>
<td>224</td>
</tr>
</tbody>
</table>
If we consider the data as $y_i | n_i \sim Bin(n_i, \pi_i), i = 1, 2, 3, 4 = I$, we have $I = 4$ data points.

Consider a model such as the logistic regression:

$$\text{logit}\{\pi(x)\} = \alpha + \beta x.$$  

$\Rightarrow$ ML $L_M$.

A Saturated model has a separate $\pi_i$ for each value of $x$ (perfect fit).

$\Rightarrow$ ML $L_S$.

Deviance is the LRT comparing current model to the saturated model:

$$\text{Dev} = 2[\log(L_S) - \log(L_M)].$$

If the current model is good, then $\text{Dev} \sim \chi^2_{I-(p+1)}$. A smaller Dev indicates a better fit.
SAS proc genmod automatically presents the Deviance for a model:

```sas
proc genmod;
  model y/n = score / dist=bin link=logit;
run;
```

Criteria For Assessing Goodness Of Fit

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>2</td>
<td>2.8089</td>
<td>1.4045</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>2</td>
<td>2.8089</td>
<td>1.4045</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>2</td>
<td>2.8743</td>
<td>1.4372</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>2</td>
<td>2.8743</td>
<td>1.4372</td>
</tr>
</tbody>
</table>

Linear probability model is better than the logistic model using deviance!
★ Note: We can also use the following *Pearson* $\chi^2$ statistic for model checking in this situation:

$$\chi^2 = \sum \frac{(y_i - \hat{E}(y_i)_{\text{model}})^2}{\hat{\text{var}}(y_i)_{\text{model}}}$$

where $\hat{E}(y_i)_{\text{model}}$ is the est. mean of $y_i$ under current model, $\hat{\text{var}}(y_i)_{\text{model}}$ is the est. variance of $y_i$ under current model.

★ If the model fits the data well, $\chi^2 \sim \chi^2_{I-(p+1)}$. A small $\chi^2$ indicates a better fit.

★ If we use the Pearson $\chi^2$, we get the same conclusion:
Linear probability model is better than the logistic model!

★ Note: If $Y$ is binary, we should use option aggregate= in the model statement:

```r
proc genmod descending;
   model hd = score / dist=bin link=logit aggregate=score;
run;
```
IV.3 Residuals

- We can obtain Deviance residuals or Pearson $\chi^2$ residuals after fitting a GLM.

- Deviance residuals:

$$\text{Dev} = 2[\log(L_S) - \log(L_M)] = \sum d_i,$$

$$r_{Di} = d_i^{1/2} \text{sign}(y_i - \mu_i)$$ is the deviance residual.

- Standardized Deviance residuals is the standardized version of $r_{Di}$. Standardized deviance residuals can be used to identify outliers.

- Pearson residuals:

$$e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{var}(y_i)}}.$$

$$\mathbb{E}(e_i) \approx 0, \text{ var}(e_i) < 1.$$
• Standardized Pearson residual:

\[ r_i = \frac{y_i - \hat{\mu}_i}{SE}. \]

E(\(r_i\)) \(\approx\) 0, \(\text{var}(r_i)\) \(\approx\) 1, \(r_i\) behaves like a N(0,1) variable. Standardized Pearson residuals can be used to identify outliers.

• Use residuals in the model Statement of Proc Genmod to obtain these residuals.