

Chapter 4: Generalized Linear Models-I

Dipankar Bandyopadhyay

Department of Biostatistics,
Virginia Commonwealth University

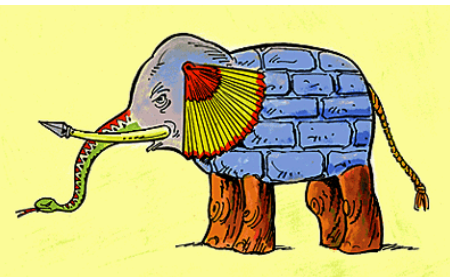
BIOS 625: Categorical Data & GLM

[Acknowledgements to Tim Hanson and Haitao Chu]

The Blind Men and The Elephant

A group of blind men gathered around an elephant, trying to find out what elephant look like.

- The 1st one touched the tusk, and said: 'It is just like a spear.'
- The 2nd one touched the ear, and said: 'It is just like a big fan.'
- The 3rd one touched the knee, and said: 'It is just like a tree.'
- The 4th one touched the body, and said: 'It is just like a wall.'
- The 5th one touched the trunk, and said: 'It is just like a snake.'
- The last one touched the tail, and said: 'It is just like a rope.'



- Generalized linear models (GLMs) form a very large class that include many highly used models as special cases: ANOVA, ANCOVA, regression, logistic regression, Poisson regression, log-linear models, etc.
- By developing the GLM in the abstract, we can consider many components that are similar across models (fitting techniques, deviance, residuals, etc).
- Each GLM is completely specified by three components:
 - (a) the distribution of the outcome Y_i ,
 - (b) the linear predictor η_i , and
 - (c) the link function $g(\cdot)$.

4.1.1 Model components

- ① *Random* component is response Y with independent realizations $\mathbf{Y} = (Y_1, \dots, Y_N)$ from a distribution in a (one parameter) exponential family:

$$f(y_i|\theta_i) = a(\theta_i)b(y_i) \exp[y_i Q(\theta_i)].$$

- ▶ Members include chi-square, binomial, Poisson, and geometric distributions.
- ▶ $Q(\theta_i)$ is called the *natural parameter*.
- ▶ θ_i may depend on explanatory variables $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$.

Two parameter exponential families include gamma, Weibull, normal, beta, and negative binomial distributions.

- ② The *systematic components* are $\boldsymbol{\eta} = (\eta_1, \dots, \eta_N)$ where $\eta_i = \sum_{j=1}^p \beta_j x_{ij} = \boldsymbol{\beta}' \mathbf{x}_i$.
- ▶ Called the *linear predictor*.
 - ▶ Relates \mathbf{x}_i to θ_i via link function.
 - ▶ Most models have an intercept and so often $x_{i1} = 1$ and there are $p - 1$ actual predictors.
- ③ The *link function* $g(\cdot)$ connects the random Y_i and systemic η_i components. Let $\mu_i = E(Y_i)$. Then $\eta_i = \mathbf{x}_i' \boldsymbol{\beta} = g(\mu_i)$.
- ▶ $g(\cdot)$ is monotone and smooth.
 - ▶ $g(m) = m$ is “identity link.”
 - ▶ The $g(\cdot)$ such that $g(\mu_i) = Q(\theta_i)$ is called the *canonical link*.

The model is

$$E(Y_i) = g^{-1}(x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p),$$

for $i = 1, \dots, N$, where Y_i is distributed according to a 1-parameter exponential family (for now).

$g^{-1}(\cdot)$ is called the *inverse link function*. Common choices are

- ❶ $g(x) = x$ so $g^{-1}(x) = x$ (identity link)
- ❷ $g(x) = \log x$ so $g^{-1}(x) = e^x$ (log-link)
- ❸ $g(x) = \log\{x/(1-x)\}$ so $g^{-1}(x) = e^x/(1+e^x)$ (logit link)
- ❹ $g(x) = F^{-1}(x)$ so $g^{-1}(x) = F(x)$ where $F(\cdot)$ is a CDF (inverse-CDF link)

4.1.2 Bernoulli response

- Let $Y \sim \text{Bern}(\pi) = \text{bin}(1, \pi)$. Then

$$p(y) = \pi^y (1 - \pi)^{1-y} = (1 - \pi) \exp\{y \log(\pi/(1 - \pi))\}.$$

- So $a(\pi) = 1 - \pi$, $b(y) = 1$, $Q(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$.
- So $g(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$ is the canonical link.
- $g(\pi)$ is the log-odds of $Y_i = 1$, also called the logit of π :
 $\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$.
- Using the canonical link we have the GLM relating Y_i to $\mathbf{x}_i = (1, x_{i1}, \dots, x_{i,p-1})$:

$$Y_i \sim \text{Bern}(\pi_i), \quad \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + x_{i1}\beta_1 + \dots + x_{i,p-1}\beta_{p-1} = \mathbf{x}_i' \boldsymbol{\beta},$$

the **logistic regression** model.

4.1.3 Poisson response

- Let $Y_i \sim \text{Pois}(\mu_i)$. Then

$$p(y) = e^{-\mu} \mu^y / y! = e^{-\mu} (1/y!) e^{y \log \mu}.$$

- So $a(\mu) = e^{-\mu}$, $b(y) = 1/y!$, $Q(\mu) = \log \mu$.
- So $g(\mu) = \log \mu$ is the canonical link.
- Using the canonical link we have the GLM relating Y_i to \mathbf{x}_i :

$$Y_i \sim \text{Pois}(\mu_i), \quad \log \mu_i = \beta_0 + x_{i1}\beta_1 + \cdots + x_{i,p-1}\beta_{p-1},$$

the **Poisson regression** model.

4.1.5 Deviance

- For a GLM, let $\mu_i = E(Y_i)$ for $i = 1, \dots, N$. The GLM places *structure* on the means $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)$; instead of N parameters in $\boldsymbol{\mu}$ we really only have p : β_1, \dots, β_p determines $\boldsymbol{\mu}$ and data reduction is obtained. So really, $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$ in a GLM through $\mu_i = g^{-1}(\mathbf{x}_i\boldsymbol{\beta})$.
- Here's the log likelihood in terms of (μ_1, \dots, μ_N) :

$$L(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^N \log p(y_i; \mu_i).$$

- If we forget about the model (with parameter $\boldsymbol{\beta}$) and just “fit” $\hat{\mu}_i = y_i$, the observed data, we obtain the largest the likelihood can be when the $\boldsymbol{\mu}$ have no structure at all; we get $L(\hat{\boldsymbol{\mu}}; \mathbf{y}) = L(\mathbf{y}; \mathbf{y})$. This is the largest the log-likelihood can be, when $\boldsymbol{\mu}$ is unstructured and estimated by plugging in \mathbf{y} .

- This terrible “model,” called the *saturated* model, is not useful for succinctly explaining data or prediction, but rather serves as a reference point for real models with $\mu_i = g^{-1}(\beta' \mathbf{x}_i)$.
- We can compare the fit of a real GLM to the saturated model, or to other GLMs with additional or fewer predictors, through the *drop in deviance*.
- Let $L(\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}); \mathbf{y})$ be the log likelihood evaluated at the MLE of $\boldsymbol{\beta}$. The *deviance* of the model is $D = -2[L(\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}); \mathbf{y}) - L(\mathbf{y}; \mathbf{y})]$.
- Here we are plugging in $\hat{\mu}_i = g^{-1}(\hat{\boldsymbol{\beta}}' \mathbf{x}_i)$ for the first part and $\hat{\mu}_i = y_i$ for the second.

Using D for goodness-of-fit

- If the sample size N is fixed, but the data are recorded in such a way that each y_i gets more observations (this can happen with Poisson and binomial data), then $D \overset{\bullet}{\sim} \chi^2_{N-p}$ tests $H_0 : \mu_i = g^{-1}(\beta' \mathbf{x}_i)$ versus $H_1 : \mu_i$ arbitrary. GOF statistic.
- For example, let Y_i is the number of diabetics y_i out of n_i at three different BMI levels. A more realistic scenario is that as more data are collected, N increases. In this case, a rule-of-thumb is to look at $D/(N - p)$; $D/(N - p) > 2$ indicates some lack-of-fit.
- Then, we can try modeling the mean more flexibly. If this does not help, then including random effects or the use of quasi-likelihood (a variance fix) can help.
- Alternatively, one can resort to a different sampling model, i.e., negative binomial instead of Poisson.

4.2 Binary response regression

- Let $Y_i \sim \text{Bern}(\pi_i)$. Y_i might indicate the presence/absence of a disease, whether someone has obtained their drivers license or not, etc.
- Through a GLM we wish to relate the probability of “success” to explanatory covariates $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ through $\pi_i = \pi(\mathbf{x}_i) = g^{-1}(\mathbf{x}_i' \boldsymbol{\beta})$.
- So then,

$$Y_i \sim \text{Bern}(\pi(\mathbf{x}_i)),$$

$$\text{and } E(Y_i) = \pi(\mathbf{x}_i) \text{ and } \text{var}(Y_i) = \pi(\mathbf{x}_i)[1 - \pi(\mathbf{x}_i)].$$

4.2.1 Simplest link, $g(x) = x$

When $g(x) = x$, the identity link, we have $\pi(\mathbf{x}_i) = \beta' \mathbf{x}_i$. When $\mathbf{x}_i = x_i$ is one-dimensional, this reduces to

$$Y_i \sim \text{Bern}(\alpha + \beta x_i).$$

- When x_i large or small, $\pi(x_i)$ can be less than zero or greater than one.
- Appropriate for a restricted range of x_i values.
- Can of course be extended to $\pi(\mathbf{x}_i) = \beta' \mathbf{x}_i$ where $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})$.
- Can be fit in SAS proc `genmod`.

Example

Association between snoring (as measured by a snoring score) and heart disease. Let s be someone's snoring score, $s \in \{0, 2, 4, 5\}$ (see text, p. 118).

Snoring	s	Heart disease		Proportion yes	Linear fit	Logit fit
		yes	no			
Never	0	24	1355	0.017	0.017	0.021
Occasionally	2	35	603	0.055	0.057	0.044
Nearly every night	4	21	192	0.099	0.096	0.093
Every night	5	30	224	0.118	0.116	0.132

This is fit in proc genmod:

```
data glm;  
  input snoring disease total @@;  
  datalines ;  
  0 24 1379 2 35 638 4 21 213 5 30 254  
  ;  
proc genmod;  
  model disease/total = snoring / dist=bin link=identity;  
run;
```

The GENMOD Procedure

Model Information

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

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	110
2	Nonevent	2374

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	0.0692	0.0346
Scaled Deviance	2	0.0692	0.0346
Pearson Chi-Square	2	0.0688	0.0344
Scaled Pearson X2	2	0.0688	0.0344
Log Likelihood		-417.4960	
Full Log Likelihood		-10.1609	
AIC (smaller is better)		24.3217	
AICC (smaller is better)		36.3217	
BIC (smaller is better)		23.0943	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	.0172466	.0034369	.0105104	.0239829	25.18	<.0001
snoring	1	.0197778	.0027978	.0142942	.0252615	49.97	<.0001
Scale	0	1.000000	.0000000	1.000000	1.000000		

NOTE: The scale parameter was held fixed.

Interpretation?

- The fitted model is

$$\hat{\pi}(s) = 0.0172 + 0.0198s.$$

- For every unit increase in snoring score s , the probability of heart disease increases by about 2%.
- The p -values test $H_0 : \alpha = 0$ and $H_0 : \beta = 0$. The latter is more interesting and we reject at the $\alpha = 0.001$ level. The probability of heart disease is strongly, *linearly* related to the snoring score.
- What do you think that SCALE term is in the output? Note:
 $P(\chi^2_2 > 0.0692) \approx 0.966$.

4.2.3 Logistic regression

- Often a fixed change in x has less impact when $\pi(x)$ is near zero or one.

Example: Let $\pi(x)$ be probability of getting an A in a statistics class and x is the number of hours a week you work on homework. When $x = 0$, increasing x by 1 will change your (very small) probability of an A very little. When $x = 4$, adding an hour will change your probability quite a bit. When $x = 20$, that additional hour probably won't improve your chances of getting an A much. You were at essentially $\pi(x) \approx 1$ at $x = 10$. Of course, this is a *mean* model. Individuals will vary.

- The most widely used nonlinear function to model probabilities is the canonical, logit link:

$$\text{logit}(\pi_i) = \alpha + \beta x_i.$$

- Solving for π_i and then dropping the subscripts we get the probability of success ($Y = 1$) as a function of x :

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}.$$

- When $\beta > 0$ the function increases from 0 to 1; when $\beta < 0$ it decreases. When $\beta = 0$ the function is constant for all values of x and Y is unrelated to x .
- The logistic (or anti-logit) function is $\text{logit}^{-1}(x) = e^x / (1 + e^x)$.
- To fit the snoring data to the logistic regression model we use the same SAS code as before (`proc genmod`) except specify `LINK=LOGIT` and obtain $\hat{\alpha} = -3.87$ and $\hat{\beta} = 0.40$ as maximum likelihood estimates.

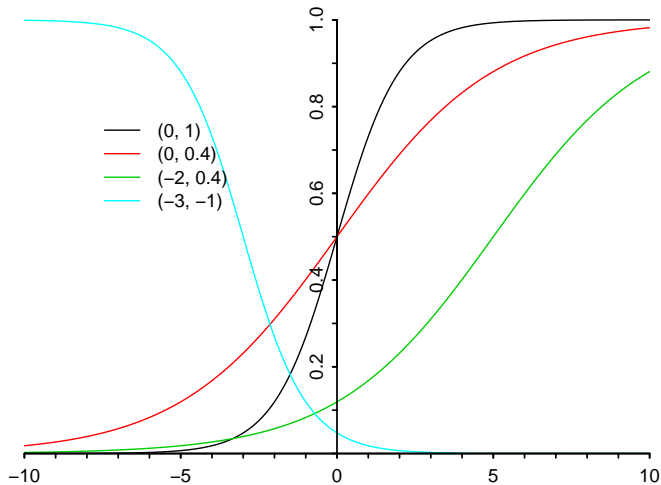


Figure : Logistic curves $\pi(x) = e^{\alpha + \beta x} / (1 + e^{\alpha + \beta x})$ with $(\alpha, \beta) = (0, 1), (0, 0.4), (-2, 0.4), (-3, -1)$. What about $(\alpha, \beta) = (\log 2, 0)$?

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	2.8089	1.4045
Scaled Deviance	2	2.8089	1.4045
Pearson Chi-Square	2	2.8743	1.4372
Scaled Pearson X2	2	2.8743	1.4372
Log Likelihood		-418.8658	
Full Log Likelihood		-11.5307	
AIC (smaller is better)		27.0615	
AICC (smaller is better)		39.0615	
BIC (smaller is better)		25.8341	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.86625	.1662144	-4.19202	-3.54047	541.06	<.0001
snoring	1	.3973366	.0500107	.2993175	.4953557	63.12	<.0001
Scale	0	1.000000	.0000000	1.000000	1.000000		

NOTE: The scale parameter was held fixed.

You can also use `proc logistic` to fit binary regression models.

```
proc logistic ; model disease/total = snoring; run;
```

The LOGISTIC Procedure

Model Information

Data Set	WORK.GLM
Response Variable (Events)	disease
Response Variable (Trials)	total
Model	binary logit
Optimization Technique	Fisher's scoring
Number of Observations Read	4
Number of Observations Used	4
Sum of Frequencies Read	2484
Sum of Frequencies Used	2484

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	110
2	Nonevent	2374

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	902.827	841.732
SC	908.645	853.367
-2 Log L	900.827	837.732

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	63.0956	1	<.0001
Score	72.6881	1	<.0001
Wald	63.1238	1	<.0001

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.8662	0.1662	541.0562	<.0001
snoring	1	0.3973	0.0500	63.1238	<.0001

Effect	Odds Ratio Estimates		
	Point Estimate	95% Wald Confidence Limits	
snoring	1.488	1.349	1.641

Association of Predicted Probabilities and Observed Responses

Percent Concordant	58.6	Somers' D	0.419
Percent Discordant	16.7	Gamma	0.556
Percent Tied	24.7	Tau-a	0.035
Pairs	261140	c	0.709

- The fitted model is then

$$\hat{\pi}(x) = \frac{\exp(-3.87 + 0.40x)}{1 + \exp(-3.87 + 0.40x)}.$$

- As before, we reject $H_0 : \beta = 0$; there is a strong, positive association between snoring score and developing heart disease.
- Which model provides better fit? (Fits at the 4 s values are in the original data table with raw proportions.)

Note: $P(\chi_2^2 > 2.8089) \approx 0.246$.

4.2.4 What is β when $x = 0$ or 1 ?

Consider a general link $g\{\pi(x)\} = \alpha + \beta x$.

Say $x = 0, 1$. Then we have a 2×2 contingency table.

	$Y = 1$	$Y = 0$
$X = 1$	$\pi(1)$	$1 - \pi(1)$
$X = 0$	$\pi(0)$	$1 - \pi(0)$

- Identity link, $\pi(x) = \alpha + \beta x$: $\beta = \pi(1) - \pi(0)$, the difference in proportions.
- Log link, $\pi(x) = e^{\alpha + \beta x}$: $e^{\beta} = \pi(1)/\pi(0)$ is the relative risk.
- Logit link, $\pi(x) = e^{\alpha + \beta x} / (1 + e^{\alpha + \beta x})$:
 $e^{\beta} = [\pi(1)/(1 - \pi(1))]/[\pi(0)/(1 - \pi(0))]$ is the odds ratio.

4.2.5 Inverse CDF links*

- The logistic regression model can be rewritten as

$$\pi(x) = F(\alpha + \beta x),$$

where $F(x) = e^x / (1 + e^x)$ is the CDF of a standard logistic random variable L with PDF

$$L \sim f(x) = e^x / (1 + e^x)^2.$$

- In practice, any CDF $F(\cdot)$ can be used as $g^{-1}(\cdot)$. Common choices are $g^{-1}(x) = \Phi(x) = \int_{-\infty}^x (2\pi)^{-0.5} e^{-0.5z^2} dz$, yielding a *probit* regression model (LINK=PROBIT) and $g^{-1}(x) = 1 - \exp(-\exp(x))$ (LINK=CLL), the complimentary log-log link.
- Alternatively, $F(\cdot)$ may be left unspecified and estimated from data using nonparametric methods. Bayesian approaches include using the Dirichlet process and Polya trees. Q: How is β interpreted?

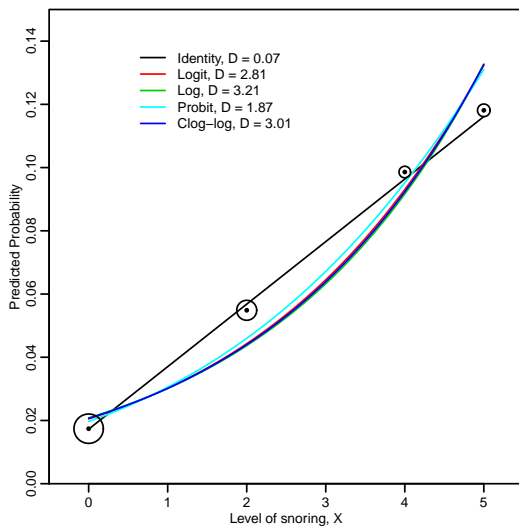


Figure : Predicted probabilities from 5 regression models

Comments:

- There's several links we can consider; we can also toss in quadratic terms in x_i , etc. How to choose? Diagnostics? Model fit statistics?
- We haven't discussed much of the output from PROC LOGISTIC; what do you think those statistics are? Gamma? AIC?
- For snoring data, $D = 0.07$ for identity versus $D = 2.81$ for logit links. Which model fits better? The $df = 4 - 2 = 2$ here. What is the 4? What is the 2? The corresponding p -values are 0.97 and 0.25.
- The log link yields $D = 3.21$ and $p = 0.2$, the probit link yields $D = 1.87$ and $p = 0.4$, and CLL link yields $D = 3.01$ and $p = 0.22$. Which link would you pick?
- How would you interpret β ? Are any links significantly inadequate?

Comments (continued):

- Identity, Logit and Log links give nice interpretations, i.e. regression coefficients correspond to RD, $\log(\text{OR})$ and $\log(\text{RR})$;
- Predicted probabilities and 95% CI from Logit, Probit and C-log-log links are within the natural boundary of $[0, 1]$.
- When using Identity and Log Links, one may consider Bayesian approaches to incorporate the natural boundary constraints, particularly if maximum likelihood methods fails to converge or if the predicted probabilities or confidence limits are outside the natural boundary (See Chu and Cole 2010, Epidemiology, 21(6), 855-862)

4.3.1 Poisson loglinear model

- We have

$$Y_i \sim \text{Pois}(\mu_i).$$

The log link $\log(\mu_i) = \mathbf{x}_i' \boldsymbol{\beta}$ is most common, with one predictor \mathbf{x} we have

$$Y_i \sim \text{Pois}(\mu_i), \quad \mu_i = e^{\alpha + \beta x_i},$$

or simply $Y_i \sim \text{Pois}(e^{\alpha + \beta x_i})$.

- The mean satisfies

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

Then

$$\mu(x+1) = e^{\alpha + \beta(x+1)} = e^{\alpha + \beta x} e^{\beta} = \mu(x) e^{\beta}.$$

- Increasing x by one increases the mean by a factor of e^{β} .
- Note that the log maps the positive rate μ into the real numbers \mathbb{R} , where $\alpha + \beta x$ lives. This is also the case for the logit link for binary regression, which maps π into the real numbers \mathbb{R} .

Example: Crab mating

Table 4.3 (p. 123) has data on 173 female horseshoe crabs.

- C = color (1,2,3,4=light medium, medium, dark medium, dark).
- S = spine condition (1,2,3=both good, one worn or broken, both worn or broken).
- W = carapace width (cm).
- Wt = weight (kg).
- Sa = number of satellites (additional male crabs besides her nest-mate husband) nearby.

We initially examine width as a predictor for the number of satellites. Figure 4.3 doesn't tell us much. Aggregating over width categories in Figure 4.4 helps & shows an approximately linear trend. We'll fit three models using `proc genmod`.

$$Sa_i \sim \text{Pois}(e^{\alpha + \beta W_i}),$$

$$Sa_i \sim \text{Pois}(\alpha + \beta W_i),$$

and

$$Sa_i \sim \text{Pois}(e^{\alpha + \beta_1 W_i + \beta_2 W_i^2}).$$

SAS code:

```
data crab; input color spine width satell weight;
    weight=weight/1000; color=color-1;
    width_sq=width*width;
datalines ;
3 3 28.3 8 3050
4 3 22.5 0 1550
... et cetera ...
5 3 27.0 0 2625
3 2 24.5 0 2000
;
proc genmod;
    model satell = width / dist=poi link=log ;
proc genmod;
    model satell = width / dist=poi link=identity ;
proc genmod;
    model satell = width width_sq / dist=poi link=log ;
run;
```

Output from 3 models

```

The GENMOD Procedure
Model Information
Data Set          WORK.CRAB
Distribution       Poisson
Link Function      Log
Dependent Variable  satell
Number of Observations Read      173
Number of Observations Used      173

```

```

Criteria For Assessing Goodness Of Fit

Criterion          DF          Value          Value/DF
Deviance            171          567.8786          3.3209
Scaled Deviance     171          567.8786          3.3209
Pearson Chi-Square   171          544.1570          3.1822
Scaled Pearson X2    171          544.1570          3.1822
Log Likelihood              68.4463
Full Log Likelihood    -461.5881
AIC (smaller is better)  927.1762
AICC (smaller is better)  927.2468
BIC (smaller is better)  933.4828

Algorithm converged.

```

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	−3.30476	.5422416	−4.36753	−2.24198	37.14	<.0001
width	1	.1640451	.0199653	.1249137	.2031764	67.51	<.0001
Scale	0	1.000000	.0000000	1.000000	1.000000		

NOTE: The scale parameter was held fixed.

The GENMOD Procedure				
Model Information				
Data Set			WORK.CRAB	
Distribution			Poisson	
Link Function			Identity	
Dependent Variable			satell	
Number of Observations Read			173	
Number of Observations Used			173	
Criteria For Assessing Goodness Of Fit				
Criterion	DF	Value	Value/DF	
Deviance	171	557.7083	3.2615	
Scaled Deviance	171	557.7083	3.2615	
Pearson Chi-Square	171	542.4854	3.1724	
Scaled Pearson X2	171	542.4854	3.1724	
Log Likelihood		73.5314		
Full Log Likelihood		-456.5030		
AIC (smaller is better)		917.0060		
AICC (smaller is better)		917.0766		
BIC (smaller is better)		923.3126		
Algorithm converged.				
Analysis Of Maximum Likelihood Parameter Estimates				
	Standard	Wald 95% Confidence	Wald	
DF	Estimate	Error	Limits	Chi-Square Pr > ChiSq
1	-11.5321	1.510400	-14.4924 -8.57173	58.29 <.0001
1	.5494968	.0592926	.4332855 .6657082	85.89 <.0001
0	1.000000	.0000000	1.000000 1.000000	

NOTE: The scale parameter was held fixed.

The GENMOD Procedure

Model Information

Data Set	WORK.CRAB
Distribution	Poisson
Link Function	Log
Dependent Variable	satell

Number of Observations Read	173
Number of Observations Used	173

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	170	558.2359	3.2837
Scaled Deviance	170	558.2359	3.2837
Pearson Chi-Square	170	539.1413	3.1714
Scaled Pearson X2	170	539.1413	3.1714
Log Likelihood		73.2676	
Full Log Likelihood		-456.7668	
AIC (smaller is better)		919.5336	
AICC (smaller is better)		919.6756	
BIC (smaller is better)		928.9935	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	-19.6525	5.637435	-30.7017	-8.60335	12.15	0.0005
width	1	1.365990	.4134378	.5556670	2.176313	10.92	0.0010
width_sq	1	-.021958	.0075600	-.036776	-.007141	8.44	0.0037
Scale	0	1.000000	.0000000	1.000000	1.000000		

NOTE: The scale parameter was held fixed.

- Write down the fitted equation for the Poisson mean from each model.
- How are the regression effects interpreted in each case?
- How would you pick among models?
- Are there any potential problems with any of the models? How about prediction?