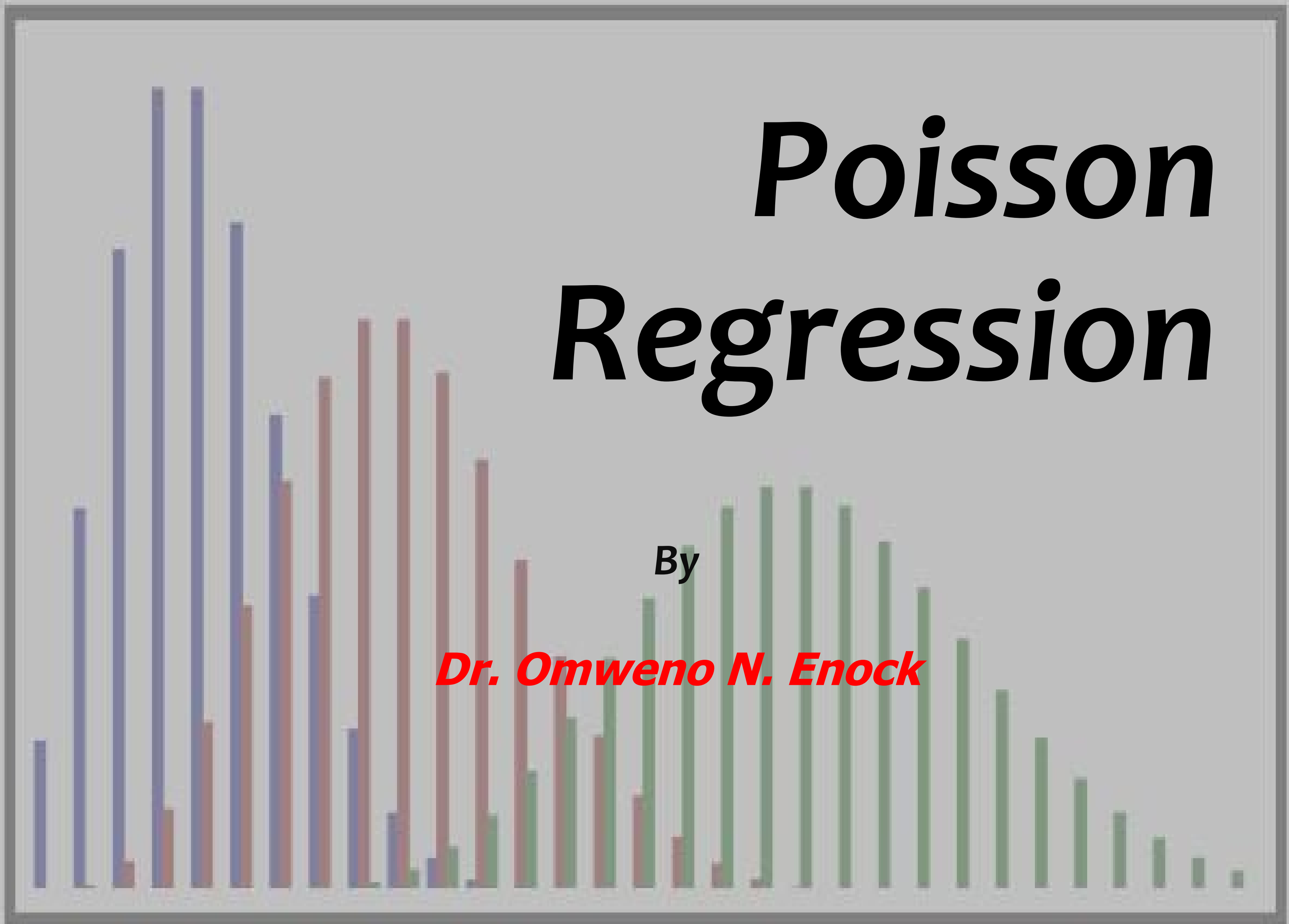


Poisson Regression

By

Dr. Omweno N. Enock

Outline

1. *Definition of the Poisson distribution*
2. *Poisson Regression*
   - *Poisson Regression Model*
   - *Parameter Estimation*
   - *Measures of Goodness-of-Fit*
   - *Confidence interval*
   - *Residual Analysis*

3. *Numerical Example*

Poisson Distribution

Poisson Distribution is the discrete probability of count of the events which occur randomly in a given interval of time.

In Poisson distribution, the trials must be very large while the probability of occurrences of outcome under observation must be small. In addition, the independence of trials and consistency of probability from trial to trial properties are required.

Determining The Poisson Distribution

The Poisson random variable satisfies the following conditions:

- *The number of successes in two disjoint time intervals is independent.*
- *The probability of a success during a small time interval is proportional to the entire length of the time interval.*

Definition

A random variable Y is said have a Poisson distribution with parameter μ if it takes integer values y=0,1,2… with probability

$$Pr(Y = y) = \frac{e^{-\mu}\mu^{y}}{y!} \quad (1)$$

for $\mu > 0$. The mean and variance of this distribution is

$$E(y) = \mu \; var(y) = \mu \quad (2)$$

Since the mean is equal to variance, any factor that affects one will affect the other.

Example

- *Number of accidents on a highway in a certain area in a specified time*

Example

- *Number of accidents on a highway in a certain area in a specified time*
- *Number of telephone calls received at small business in an one-hour period.*

Example

- *Number of accidents on a highway in a certain area in a specified time*
- *Number of telephone calls received at small business in an one-hour period.*
- *Number of customers that enter a bank in an one -hour period.*

Poisson Regression Model

Poisson regression is a form of regression analysis used to model count data.

We write Poisson regression model in terms of the mean response. We assume that there exists a function, g, that relatives the mean of the response to linear predictor

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n = x_i`\beta \ (2)$$

The function g is usually called the Link function. The relationship between the mean and the linear predictor η is,

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(x_i`\beta) \ (3)$$

Identity link function : $\boldsymbol{g(\mu_i) = \eta_i = x_i`\beta}$

When this link is used, $E(y_i) = \mu_i = x_i`\beta$ since $\mu_i = g^{-1}(x_i`\beta) = x_i`\beta$.

Log-link function : $\boldsymbol{g(\mu_i) = ln(\mu_i) = x_i`\beta}$

The relationship between the mean of the response variable and the linear predictor is

$$\mu_i = g^{-1}(x_i`\beta) = e^{x_i`\beta}$$

The log link is particularly attractive for Poisson distribution because it ensures that all of the predicted values for response variable will be nonnegative.

Parameter Estimation

Maximum likehood method

The method of maximum likelihood is used to estimate the parameters in Poisson Regression. If we have a random sample of n observations on the response y and predictors x, than the likehood function is

$$L(\beta; y) = \prod_{i=1}^{n} f_i(y_i) = \prod_{i=1}^{n} \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

$$= \frac{\prod_{i=1}^{n} \mu_i^{y_i} e^{-\sum_{i=1}^{n} \mu_i}}{\prod_{i=1}^{n} y_i!} \quad (4)$$

Where $\mu_i = g^{-1}(x_i`\beta)$.

We find parameter estimates by maximizing the log likehood function

$$l(\beta; y) = lnL(\beta; y) = \sum_{i=1}^{n} y_i \, ln(\mu_i) - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} ln(y_i!) \quad (5)$$

Because the derivative of $l(\beta; y)$ respect to β results in $U(\beta, y) = 0$ is nonlinear function of β, β cannot be directly solved. We use some other methods to find it.

Newton-Raphson method

Newton – Raphson method is based on Taylor series expansion around some given point.

The Newton – Raphson iteration for solving $U(\beta, y) = 0$ is

$$\beta_{1+k} = \beta_k + \frac{U(\beta_k . y)}{I(\beta_k . y)} \quad (6)$$

where $I(\beta_k, y) = -\frac{\partial^2 l(\beta, y)}{\partial \beta^2}$ and starts with an initial value β_0.

Fisher`s Scoring method

Assume that we have $J(\beta) = E(I(\beta, y))$.

In many cases calculating $J(\beta)$ is much more easier than $I(\beta, y)$. For this reason the iteration is modified to use $J(\beta)$ instead of $I(\beta, y)$ as follows

$$\beta_{1+k} = \beta_k + \frac{U(\beta_k, y)}{J(\beta_k)} \quad (7)$$

This method is calling Fisher`s scoring method.

Measures Goodness-of-Fit

In Poisson regression there are two common measures for Goodness-of-Fit

- *Pearson`s Chi-squared*
- *Deviance*

Both measures have approximate Chi-square distributions under hypothesis that the corrent model is appropriate for fixed number of combinations of independent variables and large counts.

Test Statistics for measures of Goodness-of-Fit

Pearson`s Chi-squared: $\chi_p^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$ (8)

Deviance: $D = 2 \sum_{i=1}^{n} \left\{ y_i \, log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i) \right\}$ (9)

Pseudo R^2

The R-squared statistic does not extend to Poisson regression models. Various pseudo R-squared tests have been proposed. These pseudo measures have the property that, when applied to the linear model, they match the interpretation of the linear model R-squared. The Pseudo R^2 is defined as

$$R^2 = \frac{\ell(\hat{\beta}_0) - \ell(\hat{\beta})}{\ell(\hat{\beta}_0)} = 1 - \frac{-2\ell(\hat{\beta})}{-2\ell(\hat{\beta}_0)} \qquad (10)$$

where $l(\hat{\beta}_0)$ is the log likelihood of the model when only the intercept is included.

The pseudo R^2 goes from 0 to 1 with 1 being a perfect fit.

Confidence Interval

The confidence interval for the mean of a Poisson distribution can be expressed using the relationship between the cumulative distribution functions of the Poisson and chi-squared distributions. The chi-squared distribution is itself closely related to the gamma distribution, and this leads to an alternative expression. Given an observation k from a Poisson distribution with mean μ_i, a confidence interval for μ_i with confidence level $1 - \alpha$ is

$$\frac{1}{2}\chi^2(\frac{\alpha}{2}; 2k) \le \mu_i \le \frac{1}{2}\chi^2(1-\alpha; 2k+2)$$

where $\chi^2(p;n)$ is the quantile function of the chi-squared distribution with n degrees of freedom.

When quantiles of the Gamma distribution are not available, an accurate approximation to this exact interval has been proposed

$$k\left(1-\frac{1}{9k}-\frac{z_{\alpha/2}}{3\sqrt{k}}\right)^3 \leq \mu_i \leq (k+1)\left(1-\frac{1}{9(k+1)}-\frac{z_{\alpha/2}}{3\sqrt{k+1}}\right)^3$$

where $z_{\alpha/2}$ *denotes the standard normal deviate with upper tail area* $\alpha / 2$*.*

An approximate large sample 95% confidence interval for e^{β} *is calculated as*

$$exp\left[\hat{\beta} \pm 1.96\left(se_{\hat{\beta}}\right)\right]$$

where $se_{\hat{\beta}}$ *is a standart error.*

Residual Analysis

As in any regression analysis, a complete residual analysis should be employed. This involves plotting the residuals against various other quantities such as the regressor variables and the response variable. Various residuals may be of interest.

We used some types of residuals in Poisson regression

- *Raw residual*
- *Pearson residuals*
- *Deviance residual*

Raw Residual

The raw residual is the different between the actual response and the estimated value from the model. The formula for the raw residual is

$$r_i = y_i - \hat{\mu}_i$$

Pearson Residual

The Pearson residual corrects for the unequal variance in the raw residuals by dividing by the standard deviation. The formula for the Pearson residual is

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}$$

Deviance Residual

The deviance residual is another popular residual. It is popular because the sum of squares of these residuals is the deviance statistic. The formula for the deviance residual is

$$r_i^D = \sqrt{2(y_i\, log(y_i/\hat{\mu}_i) - y_i + \hat{\mu}_i}$$

Measures of Influence

DFFITS *– Measure of how much an observation has effected its fitted value from the regression model. Values larger than 2*sqrt((k+1)/n) in absolute value are considered highly influential.*

DFBETAS *– Measure of how much an observation has effected the estimate of a regression. Values larger than 2/sqrt(n) in absolute value are considered highly influential.*

Measures of Influence

DFFITS – *Measure of how much an observation has effected its fitted value from the regression model. Values larger than 2*sqrt((k+1)/n) in absolute value are considered highly influential.*

BETAS – *Measure of how much an observation has effected the estimate of a regression. Values larger than 2/sqrt(n) in absolute value are considered highly influential.*

Measures of Influence

DFFITS – *Measure of how much an observation has effected its fitted value from the regression model. Values larger than 2*sqrt((k+1)/n) in absolute value are considered highly influential.*

DFBETAS – *Measure of how much an observation has effected the estimate of a regression. Values larger than 2/sqrt(n) in absolute value are considered highly influential.*

Leverage Values (Hat Diag) - *Measure of how far an observation is from the others in terms of the levels of the independent variables (not the dependent variable). Observations with values larger than 2(k+1)/n are considered to be potentially highly influential, where k is the number of predictors and n is the sample size.*

Cook's D - *Measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values. Values larger than 4/n are considered highly influential.*

Measures of Influence

DFFITS – *Measure of how much an observation has effected its fitted value from the regression model. Values larger than 2*sqrt((k+1)/n) in absolute value are considered highly influential.*

DFBETAS – *Measure of how much an observation has effected the estimate of a regression. Values larger than 2/sqrt(n) in absolute value are considered highly influential.*

Leverage Values (Hat Diag) - *Measure of how far an observation is from the others in terms of the levels of the independent variables (not the dependent variable). Observations with values larger than 2(k+1)/n are considered to be potentially highly influential, where k is the number of predictors and n is the sample size.*

Cook's D - *Measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values. Values larger than 4/n are considered highly influential.*

Numerical Example

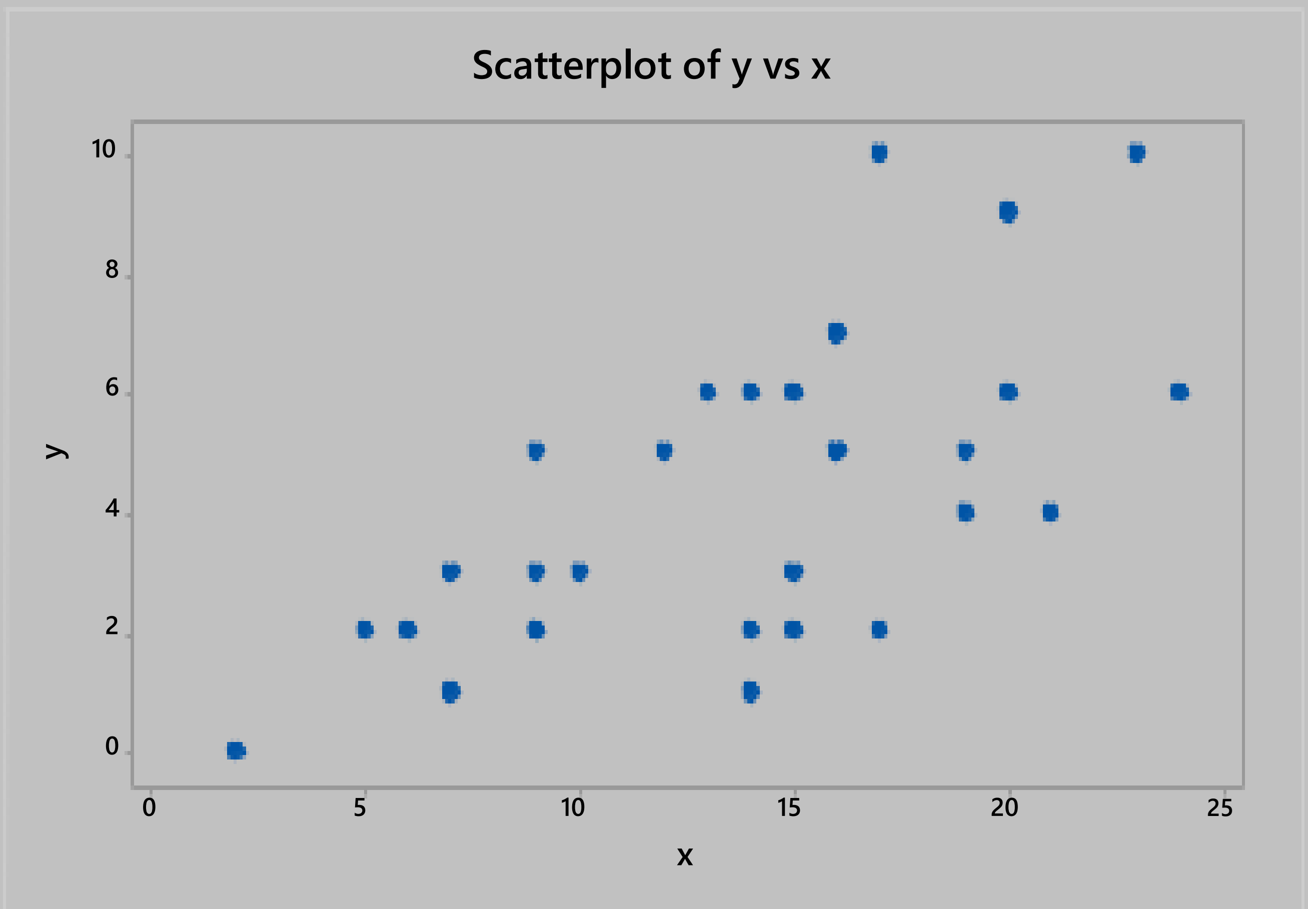
Data from publication of Statistic Department of Penn State collage.

i	x	y	i	x	y
1	2	0	16	16	7
2	15	6	17	13	6
3	19	4	18	6	2
4	14	1	19	16	5
5	16	5	20	19	5
6	15	2	21	24	6
7	9	2	22	9	2
8	17	10	23	12	5
9	10	3	24	7	1
10	23	10	25	9	3
11	14	2	26	7	3
12	14	6	27	15	3
13	9	5	28	21	4
14	5	2	29	20	6
15	17	2	30	20	9

One-Sample Kolmogorov-Smirnov Test

		y
N		30
Poisson Parameter(a,b)	Mean	4,23
Most Extreme Differences	Absolute	,127
	Positive	,127
	Negative	-,071
Kolmogorov-Smirnov Z		,698
Asymp. Sig. (2-tailed)		,714

a Test distribution is Poisson.
b Calculated from data.


Scatterplot of y vs x
10
8
6
4
2
0
y
0
5
10
15
20
25
x

Coefficients

Term	Coef	SE Coef	95% CI	Z-Value	P-Value	VIF
Constant	0,308	0,289	(-0,259; 0,875)	1,06	0,287	
x	0,0764	0,0173	(0,0424; 0,1103)	4,41	0,000	1,00

Regression Equation

```
y  =  exp(Y')
Y' = 0,308 + 0,0764 x
```

Regression Model :

y = exp(0,308 + 0,0764 x)

The p value for x is 0.000. İt means the predictor is highly significant

The 95% confidence interval:

for e^{β_0} is $exp[0.308 \pm 1.96(0.289)]$—*[-0.289;0.875]*

for e^{β_1} is $exp[0.764 \pm 1.96(0.0173)]$—*[0.0424;0.1103]*

```
Deviance at Each Iterative Step

Step   Deviance
   1  28,118609
   2  27,842234
   3  27,842092
   4  27,842092
```

Number of iterations is 4.

Deviance Table

Source	DF	Seq Dev	Contribution	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	1	20,47	42,37%	20,47	20,4677	20,47	0,000
x	1	20,47	42,37%	20,47	20,4677	20,47	0,000
Error	28	27,84	57,63%	27,84	0,9944		
Total	29	48,31	100,00%				

Deviance Table

Source	DF	Seq Dev	Contribution	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	1	20,47	42,37%	20,47	20,4677	20,47	0,000
x	1	20,47	42,37%	20,47	20,4677	20,47	0,000
Error	28	27,84	57,63%	27,84	0,9944		
Total	29	48,31	100,00%				

The Deviance table includes the following:

- *To test the null hypothesis which has no predictors. Minus two times the log likehood for the reduced model is* $-2\ell(\widehat{\beta}_0) = 48.31$ *("Total" row in the Deviance Table)*
- *Minus two times the log likehood for the fitted model is* $-2\ell(\widehat{\beta}) = 27.84$ *("Error" row in the Deviance Table)*
- *The deviance test statistic is D=48.31−27.84=20.47.*
- *The p-value comes from a χ2 distribution with 2−1=1 degrees of freedom.*

Goodness-of-Fit

Goodness-of-Fit Tests

Test	DF	Estimate	Mean	Chi-Square	P-Value
Deviance	28	27,84209	0,99436	27,84	0,473
Pearson	28	26,09324	0,93190	26,09	0,568

The high p values are indicate that there is no evidence of lack of fit

Pseudo R²

```
Model Summary

Deviance   Deviance
    R-Sq  R-Sq(adj)      AIC
  42,37%     40,30%  124,50
```

$$R^2 = 1 - \frac{-2l(\widehat{\beta})}{-2l(\widehat{\beta_0})} = 1 - \frac{27.84}{48.31} = 0.4237$$

i	FITS1	DEVRES1	HI1	COOK1	DFIT1	DBET1
1	1,584995	-1,78045	0,104459	0,103222	-0,45436	0,184879
2	4,276891	0,784971	0,034763	0,012951	0,160944	0,025003
3	5,804628	-0,79396	0,062181	0,019833	-0,19916	0,037199
4	3,962475	-1,78079	0,03558	0,042363	-0,29108	0,081712
5	4,616254	0,176215	0,036357	0,000624	0,035341	0,001204
6	4,276891	-1,23023	0,034763	0,022614	-0,21267	0,043656
7	2,704947	-0,44961	0,060099	0,006249	-0,1118	0,011747
8	4,982546	1,974329	0,040969	0,112531	0,474408	0,215842
9	2,91958	0,046852	0,05364	6,63E-05	0,011518	0,000126
10	7,878085	0,725379	0,180237	0,076643	0,391517	0,125658
11	3,962475	-1,0909	0,03558	0,01859	-0,19282	0,035858
12	3,962475	0,950598	0,03558	0,020039	0,200197	0,038653
13	2,704947	1,246375	0,060099	0,066237	0,363969	0,124512
14	1,993024	0,004939	0,086873	1,27E-06	0,001595	2,32E-06
15	4,982546	-1,52116	0,040969	0,039763	-0,282	0,076268
16	4,616254	1,030087	0,036357	0,024097	0,21953	0,046441
17	3,671174	1,11235	0,038288	0,030578	0,247298	0,058815
18	2,151167	-0,10431	0,080341	0,000505	-0,03177	0,000928
19	4,616254	0,176215	0,036357	0,000624	0,035341	0,001204
20	5,804628	-0,34217	0,062181	0,003943	-0,0888	0,007395
21	8,503197	-0,90675	0,233132	0,146062	-0,54049	0,224021
22	2,704947	-0,44961	0,060099	0,006249	-0,1118	0,011747
23	3,401288	0,809576	0,042442	0,017391	0,186501	0,033306
24	2,321858	-0,97928	0,073616	0,032277	-0,25408	0,059802
25	2,704947	0,176278	0,060099	0,001095	0,046792	0,002058
26	2,321858	0,425656	0,073616	0,008495	0,130346	0,015739
27	4,276891	-0,65276	0,034763	0,007112	-0,11927	0,01373
28	6,762348	-1,15069	0,105476	0,07437	-0,38567	0,133052
29	6,265215	-0,10672	0,080542	0,000535	-0,0327	0,000983
30	6,265215	1,024812	0,080542	0,056864	0,337237	0,104569

Fits and Diagnostics for Unusual Observations

Obs	y	Fit	SE Fit	95% CI	Resid	Std Resid	Del Resid	HI	Cook's D
8	10,000	4,983	0,452	(4,171; 5,952)	1,974	2,02	2,03	0,040969	0,11
21	6,000	8,503	1,408	(6,147; 11,763)	-0,907	-1,04	-1,02	0,233132	0,15

Obs	DFITS	
8	0,474408	R
21	-0,540485	

X

R Large residual
X Unusual X

Fits and Diagnostics for Unusual Observations

Obs	y	Fit	SE Fit	95% CI	Resid	Std Resid	Del Resid	HI	Cook's D
8	10,000	4,983	0,452	(4,171; 5,952)	1,974	2,02	2,03	0,040969	0,11
21	6,000	8,503	1,408	(6,147; 11,763)	-0,907	-1,04	-1,02	0,233132	0,15

Obs	DFITS	
8	0,474408	R
21	-0,540485	

X

R Large residual
X Unusual X

The default residuals in this output are deviance residuals, so observation 8 has a deviance residual of 1.974 and a studentized deviance residual of 2.02, while observation 21 has a leverage of 0.233132.

References

Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining. Introduction to Linear Regression Analysis, 5th Edition. John Wiley & Sons, 2012

Rodríguez, G. Lecture Notes on Generalized Linear Models. Princton Univercity. 2007

McCullagh P., Nelder J.A. Generalized Linear Models. 2nd Edition. Chapman and Hall.1989

Kleinbaum D., Kupper L., Nizam A., Muller K. Applied Regression Analysis and Other Multivariable Methods. 3rd Edition. Duxbury Press.1998

Raymond H. Myers. Classical and Modern Regression with Applications. 2nd Edition. Dubuxury Classic.2000

Simple guiding questions will be provided to guide you how to respond to the about slides