# رگرسیون پواسون

زمانی که با داده‌های شمارشی و گسسته به عنوان متغیر پاسخ سر و کار داریم، روش و مدل رگرسیونی با شیوه‌های رگرسیون خطی ساده (OLS) تفاوت دارد. رگرسیون پواسون(Poisson Regression) ، یک روش در «مدل‌های خطی تعمیم یافته (Generalized Linear Models)» محسوب می‌شود که در آن تابع احتمال برای متغیر پاسخ توزیع پواسن در نظر گرفته می‌شود. در نتیجه این مدل رگرسیونی برای داده‌های شمارشی مناسب است

## 12.3 - Poisson Regression

The Poisson distribution for a random variable $Y$ has the following probability mass function for a given value $Y = y$:

$$\mathrm{P}(Y = y|\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!},$$

for $y = 0, 1, 2, \ldots$. Notice that the Poisson distribution is characterized by the single parameter $\lambda$, which is the mean rate of occurrence for the event being measured. For the Poisson distribution, it is assumed that large counts (with respect to the value of $\lambda$) are rare.

In **Poisson regression** the dependent variable ($Y$) is an observed count that follows the Poisson distribution. The rate $\lambda$ is determined by a set of $k$ predictors $\mathbf{X} = (X_1, \ldots, X_k)$. The expression relating these quantities is

$$\lambda = \exp\{\mathbf{X}\beta\}.$$

Thus, the fundamental Poisson regression model for observation $i$ is given by

$$\mathrm{P}(Y_i = y_i|\mathbf{X}_i, \beta) = \frac{e^{-\exp\{\mathbf{X}_i\beta\}}\exp\{\mathbf{X}_i\beta\}^{y_i}}{y_i!}.$$

That is, for a given set of predictors, the categorical outcome follows a Poisson distribution with rate $\exp\{\mathbf{X}\beta\}$. For a sample of size $n$, the likelihood for a Poisson regression is given by:

$$L(\beta; \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{n}\frac{e^{-\exp\{\mathbf{X}_i\beta\}}\exp\{\mathbf{X}_i\beta\}^{y_i}}{y_i!}.$$

This yields the log likelihood:

$$\ell(\beta) = \sum_{i=1}^{n} y_i\mathbf{X}_i\beta - \sum_{i=1}^{n}\exp\{\mathbf{X}_i\beta\} - \sum_{i=1}^{n}\log(y_i!).$$

Maximizing the likelihood (or log likelihood) has no closed-form solution, so a technique like iteratively reweighted least squares is used to find an estimate of the regression coefficients, $\hat{\beta}$. Once this value of $\hat{\beta}$ has been obtained, we may proceed to define various goodness-of-fit measures and calculated residuals. For the residuals we present, they serve the same purpose as in linear regression. When plotted versus the response, they will help identify suspect data points.

### Goodness-of-Fit

Overall performance of the fitted model can be measured by two different chi-square tests. There is the **Pearson statistic**

$$X^2 = \sum_{i=1}^{n}\frac{(y_i - \exp\{\mathbf{X}_i\hat{\beta}\})^2}{\exp\{\mathbf{X}_i\hat{\beta}\}}$$

and the **deviance statistic**

$$D = 2\sum_{i=1}^{n}\left[y_i\log\left(\frac{y_i}{\exp\{\mathbf{X}_i\hat{\beta}\}}\right) - (y_i - \exp\{\mathbf{X}_i\hat{\beta}\})\right].$$

Both of these statistics are approximately chi-square distributed with $n - k - 1$ degrees of freedom. When a test is rejected, there is a statistically significant lack of fit. Otherwise, there is no evidence of lack-of-fit.

To illustrate, the relevant software output from the simulated example is:

## Pseudo $R^2$

The value of $R^2$ used in linear regression also does not extend to Poisson regression. One commonly used measure is the **pseudo $R^2$**, defined as

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

where $\ell(\hat{\beta}_0)$ 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.

### Raw Residual

The **raw residual** is the difference between the actual response and the estimated value from the model. Remember that the variance is equal to the mean for a Poisson random variable. Therefore, we expect that the variances of the residuals are unequal. This can lead to difficulties in the interpretation of the raw residuals, yet it is still used. The formula for the raw residual is

$$r_i = y_i - \exp\{\mathbf{X}_i\beta\}.$$

### 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 residuals is

$$p_i = \frac{r_i}{\sqrt{\hat{\phi}\exp\{\mathbf{X}_i\beta\}}},$$

where

$$\hat{\phi} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i - \exp\{\mathbf{X}_i\hat{\beta}\})^2}{\exp\{\mathbf{X}_i\hat{\beta}\}}.$$

$\hat{\phi}$ is a dispersion parameter to help control overdispersion.

### Deviance Residuals

**Deviance residuals** are also popular because the sum of squares of these residuals is the deviance statistic. The formula for the deviance residual is

$$d_i = \mathrm{sgn}(y_i - \exp\{\mathbf{X}_i\hat{\beta}\})\sqrt{2\left\{y_i\log\left(\frac{y_i}{\exp\{\mathbf{X}_i\hat{\beta}\}}\right) - (y_i - \exp\{\mathbf{X}_i\hat{\beta}\})\right\}}.$$

The plots below show the Pearson residuals and deviance residuals versus the fitted values for the simulated example.

## Hat Values

The hat matrix serves the same purpose as in the case of linear regression - to measure the influence of each observation on the overall fit of the model. The hat values, $h_{i,i}$, are the diagonal entries of the Hat matrix

$$H = \mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}\mathbf{W}^{1/2},$$

where $\mathbf{W}$ is an $n \times n$ diagonal matrix with the values of $\exp\{\mathbf{X}_i\hat{\beta}\}$ on the diagonal. As before, a hat value (leverage) is large if $h_{i,i} > 2p/n$.

## Studentized Residuals

Finally, we can also report Studentized versions of some of the earlier residuals. The **Studentized Pearson residuals** are given by

$$sp_i = \frac{p_i}{\sqrt{1 - h_{i,i}}}$$

and the **Studentized deviance residuals** are given by

$$sd_i = \frac{d_i}{\sqrt{1 - h_{i,i}}}.$$

مثال ۱

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

## مثال ۲

.**Example of Fit Poisson Model**A quality engineer is concerned about two types of defects in molded resin parts: discoloration and clumping. Discolored streaks in the final product can result from contamination in hoses and from abrasions to resin pellets. Clumping can occur when the process is run at higher temperatures and faster rates of transfer. The engineer identifies three possible predictor variables for the responses (defects). The engineer records the number of each type of defect in hour long sessions, while varying the predictor levels.The engineer wants to study how several predictors affect discoloration defects in resin parts. Because the response variable describes the number of times that an event occurs in a finite observation space, the engineer fits a Poisson model.

Y: *Discoloration*

تعداد نقص (تغییر رنگ)

X1: '*Hours Since Cleanse*'

چند ساعت از زمان پاکسازی

X2: *Temperature*

دما

*X3:* '*Size of Screw*'

اندازه پیچ

1. Enter the sample data, ResinDefects.MTW.
2. Choose **Stat** > **Regression** > **Poisson Regression** > **Fit Poisson Model**.
3. In **Response**, enter '*Discoloration Defects*'.
4. In **Continuous predictors**, enter '*Hours Since Cleanse*' *Temperature*.
5. In **Categorical predictors**, enter '*Size of Screw*'.
6. Click **Graphs**.
7. In **Residuals for plots**, select **Standardized**.
8. Under **Residuals plots**, select **Four in one**.
9. Click **OK** in each dialog box.

# Interpret the results

The plot of the standardized deviance residuals versus the fitted values shows a distinct curve. In the plot of the residuals versus order, the residuals in the middle tend to be higher than the residuals at the beginning and end of the data set. For these data, both patterns are because of a missing interaction term between the size of the screw and the temperature. The pattern is visible on the residuals versus order plot because the engineer did not collect the data in random order. The engineer refits the model with the interaction between temperature and the size of the screw to model the defects more accurately.

**Poisson Regression Analysis: Discoloratio versus Hours Since , Temperature, ...**

```
Method
```

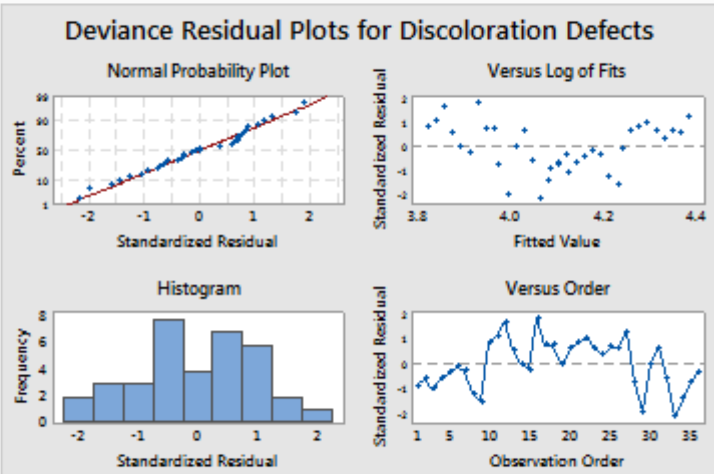
```
Link function                  Natural log
Categorical predictor coding  (1, 0)
Rows used                      36

Deviance Table

Source                  DF  Adj Dev  Adj Mean  Chi-Square  P-Value
Regression               3   56.670   18.8900       56.67    0.000
  Hours Since Cleanse    1    4.744    4.7444        4.74    0.029
  Temperature            1   38.800   38.8000       38.80    0.000
  Size of Screw          1   13.126   13.1256       13.13    0.000
Error                   32   31.607    0.9877
Total                   35   88.277

Model Summary

Deviance   Deviance
    R-Sq  R-Sq(adj)     AIC
  64.20%     60.80%  253.29

Coefficients

Term                      Coef   SE Coef   VIF
Constant                4.3982    0.0628
Hours Since Cleanse    0.01798   0.00826  1.00
Temperature          -0.001974  0.000318  1.00
Size of Screw
  small                -0.1546    0.0427  1.00

Regression Equation

Discoloration Defects  =  exp(Y')

Size of
Screw
large    Y' = 4.398 + 0.01798 Hours Since Cleanse - 0.001974 Temperature

small    Y' = 4.244 + 0.01798 Hours Since Cleanse - 0.001974 Temperature

Goodness-of-Fit Tests

Test      DF  Estimate     Mean  Chi-Square  P-Value
```

| | | | | | |
|---|---|---|---|---|---|
| Deviance | 32 | 31.60722 | 0.98773 | 31.61 | 0.486 |
| Pearson | 32 | 31.26713 | 0.97710 | 31.27 | 0.503 |

Fits and Diagnostics for Unusual Observations

| Obs | Discoloration Defects | Fit | Resid | Std Resid | |
|---|---|---|---|---|---|
| 33 | 43.00 | 58.18 | -2.09 | -2.18 | R |

R  Large residual




For the model with the interaction, the AIC is approximately 236, which is lower than the model without the interaction. The AIC criterion indicates that the model with the interaction is better than the model without the interaction. The curvature in the residuals versus fits plot is gone. The engineer decides to interpret this model rather than the model without the interaction.

**Poisson Regression Analysis: Discoloratio versus Hours Since , Temperature, ...**

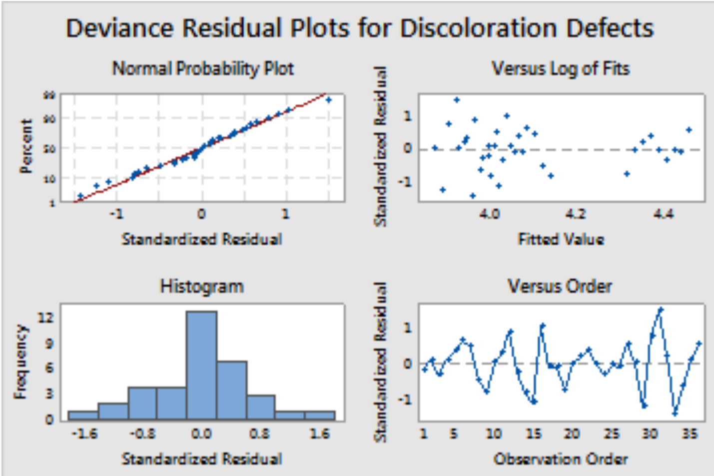
Method

| | |
|---|---|
| Link function | Natural log |
| Categorical predictor coding | (1, 0) |
| Rows used | 36 |

Deviance Table

| Source | DF | Adj Dev | Adj Mean | Chi-Square | P-Value |
|---|---|---|---|---|---|
| Regression | 4 | 75.911 | 18.9778 | 75.91 | 0.000 |
| Hours Since Cleanse | 1 | 4.744 | 4.7444 | 4.74 | 0.029 |
| Temperature | 1 | 56.970 | 56.9703 | 56.97 | 0.000 |
| Size of Screw | 1 | 30.518 | 30.5182 | 30.52 | 0.000 |

```
  Temperature*Size of Screw   1   19.241   19.2412       19.24    0.000
Error                        31   12.366    0.3989
Total                        35   88.277

Model Summary

Deviance   Deviance
    R-Sq  R-Sq(adj)     AIC
  85.99%     81.46%  236.05

Coefficients

Term                             Coef   SE Coef   VIF
Constant                       4.5760    0.0736
Hours Since Cleanse           0.01798   0.00826  1.00
Temperature                 -0.003285  0.000441  1.92
Size of Screw
  small                       -0.5444    0.0990  5.37
Temperature*Size of Screw
  small                      0.002804  0.000640  6.64

Regression Equation

Discoloration Defects  =  exp(Y')

Size of
Screw
large    Y' = 4.576 + 0.01798 Hours Since Cleanse - 0.003285 Temperature

small    Y' = 4.032 + 0.01798 Hours Since Cleanse - 0.000481 Temperature

Goodness-of-Fit Tests

Test      DF  Estimate     Mean  Chi-Square  P-Value
Deviance  31  12.36598  0.39890       12.37    0.999
Pearson   31  12.31611  0.39729       12.32    0.999
```




## Example of Predict with a Poisson regression model

AChoose **Stat** > **Regression** > **Poisson Regression** > **Predict**.

1. From **Response**, select *Discoloration Defects*.
2. In the table, enter *6* for *Hours Since Cleanse*, *115* for *Temperature*, and *large* for *Size of Screw*.
3. Click **OK**.

## Interpret the results

Minitab uses the stored model to calculate that the predicted number of discoloration defects is 72.1682. The prediction interval indicates that the engineer can be 95% confident that the mean number of discoloration defects will fall within the range of 67.5477 to 77.1047.

**Prediction for Discoloration Defects**

```
Regression Equation

Discoloration Defects  =  exp(Y')


Y' = 4.3982 + 0.01798 Hours Since Cleanse - 0.001974 Temperature
     + 0.000000 Size of Screw_large - 0.1546 Size of Screw_small

Settings

Variable              Setting
Hours Since Cleanse         6
Temperature               115
```

```
Size of Screw          large

Prediction


    Fit   SE Fit         95% CI
72.1682  2.43628  (67.5477, 77.1047)
```

| I | x1 | x2 | x3 | y |
|---|---|---|---|---|
| 1 | 0 | small | 80 | 53 |
| 2 | 1 | small | 80 | 56 |
| 3 | 2 | small | 80 | 54 |
| 4 | 3 | small | 80 | 58 |
| 5 | 4 | small | 80 | 61 |
| 6 | 5 | small | 80 | 64 |
| 7 | 6 | small | 80 | 64 |
| 8 | 7 | small | 80 | 58 |
| 9 | 8 | small | 80 | 57 |
| 10 | 0 | small | 215 | 51 |
| 11 | 1 | small | 215 | 54 |
| 12 | 2 | small | 215 | 59 |
| 13 | 3 | small | 215 | 52 |
| 14 | 4 | small | 215 | 49 |
| 15 | 5 | small | 215 | 48 |
| 16 | 6 | small | 215 | 64 |
| 17 | 7 | small | 215 | 57 |
| 18 | 8 | small | 215 | 58 |
| 19 | 0 | large | 80 | 69 |
| 20 | 1 | large | 80 | 76 |
| 21 | 2 | large | 80 | 79 |
| 22 | 3 | large | 80 | 82 |
| 23 | 4 | large | 80 | 80 |
| 24 | 5 | large | 80 | 79 |
| 25 | 6 | large | 80 | 83 |
| 26 | 7 | large | 80 | 84 |
| 27 | 8 | large | 80 | 91 |
| 28 | 0 | large | 215 | 48 |
| 29 | 1 | large | 215 | 41 |
| 30 | 2 | large | 215 | 55 |
| 31 | 3 | large | 215 | 61 |
| 32 | 4 | large | 215 | 53 |
| 33 | 5 | large | 215 | 43 |
| 34 | 6 | large | 215 | 49 |
| 35 | 7 | large | 215 | 55 |
| 36 | 8 | large | 215 | 59 |