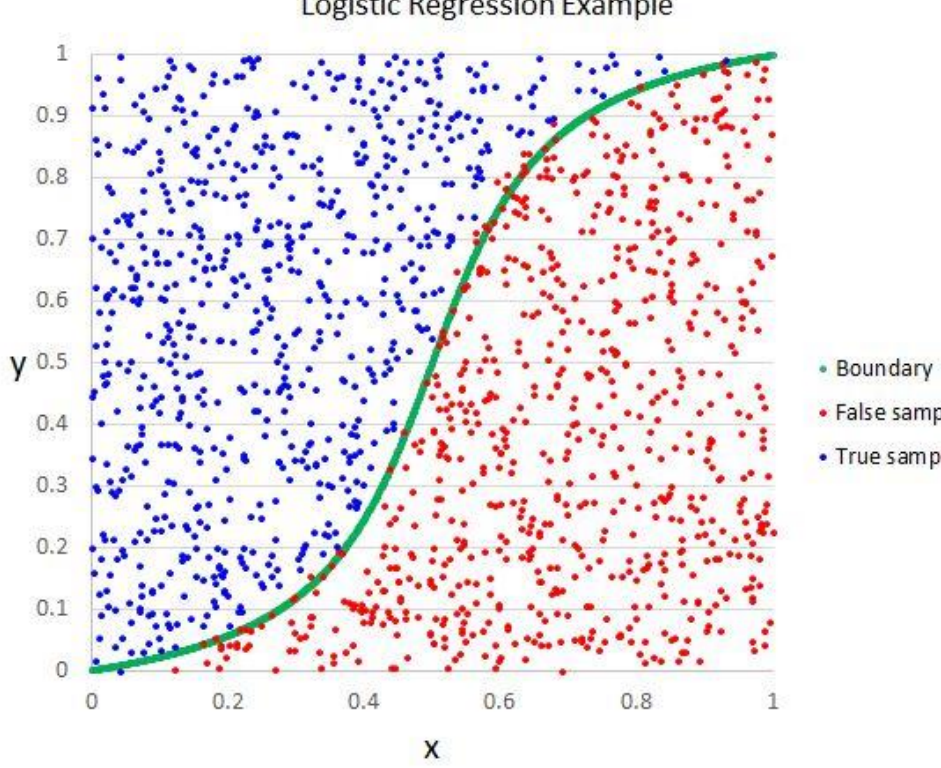


Interpreting Output

```
model = smf.glm(formula='Response~Age',data=data,family=sm.families.Binomial())
result = model.fit()
print(result.summary())
```

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                Response   No. Observations:                   92
Model:                            GLM   Df Residuals:                       90
Model Family:                Binomial   Df Model:                            1
Link Function:                  logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -24.968
Date:                Wed, 26 Dec 2018   Deviance:                       49.937
Time:                        01:07:16   Pearson chi2:                     46.3
No. Iterations:                     7   Covariance Type:                nonrobust
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -20.4078      4.523     -4.512      0.000     -29.273     -11.542
Age            0.4259      0.095      4.492      0.000       0.240       0.612
==============================================================================
```

```
result.null_deviance
```

```
123.15634524584677
```

Interpreting Output- Deviance

Deviance or Residual Deviance is similar to SSE in the sense it measures how much remains unexplained by the model built with predictors included.

Null Deviance shows how well the model predicts the response with only the intercept as a parameter. The intercept is the logarithm of the ratio of cases with y=1 to the number of cases with y=0. This is similar to SST, which gives total variation when all coefficients are zero (null hypothesis).

Interpreting Output- Testing the overall Model

The *z-values and the* associated *p-values provide* significance of individual predictor variables.

Generalized Linear Model Regression Results

Dep. Variable:	Response	No. Observations:	92
Model:	GLM	Df Residuals:	90
Model Family:	Binomial	Df Model:	1
Link Function:	logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-24.968
Date:	Wed, 26 Dec 2018	Deviance:	49.937
Time:	01:07:16	Pearson chi2:	46.3
No. Iterations:	7	Covariance Type:	nonrobust

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-20.4078	4.523	-4.512	0.000	-29.273	-11.542
Age	0.4259	0.095	4.492	0.000	0.240	0.612

```
1 result.null_deviance
```

```
123.15634524584677
```

```
1 result.aic
```

```
53.936628910751196
```

Python outputs AIC (Akaike's Information Criterion) and you need to pick the model with the lowest AIC.

Interpreting Output- Testing the Overall Model

- AIC provides a means for model selection.
- ***AIC = D + 2k, where k is the # of parameters in the model*** including the intercept.
- AIC is *similar to Adjusted R2*in the sense it penalizes for adding more parameters to the model.
- It offers a relative estimate of the information lost when a model is used to represent the process that generated the data.
- It does not test a model in the sense of null hypothesis and hence doesn't tell anything about the quality of the model. It is only a relative measure between multiple models.
- AIC = n Log(SSE/n) + 2k for Ordinary Least Squares

Logistic Regression – Pseudo R^2

- Note that R^2 is not defined in Logistic Regression

- McFadden's Pseudo R^2

- Pseudo R^2 = 1 - $\frac{Residual\ Dev}{Null\ Dev}$

- = 1 - $\frac{49.937}{123.156} = 0.59$

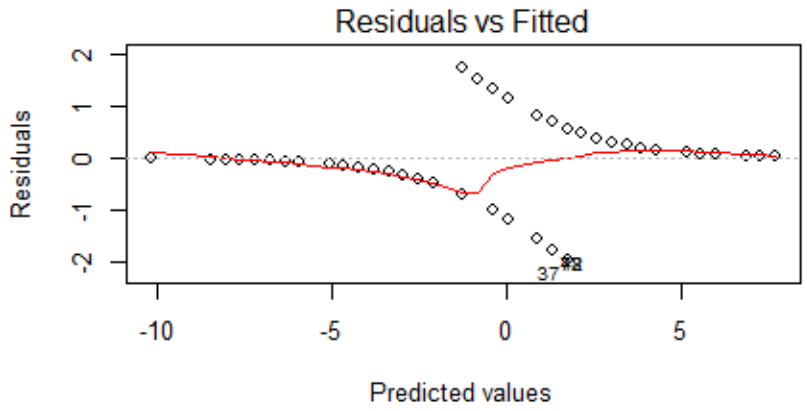
```
1 model = smf.glm(formula='Response~Age',data=data,family=sm.families.Binomial(
2 result = model.fit()
3 print(result.summary())
```

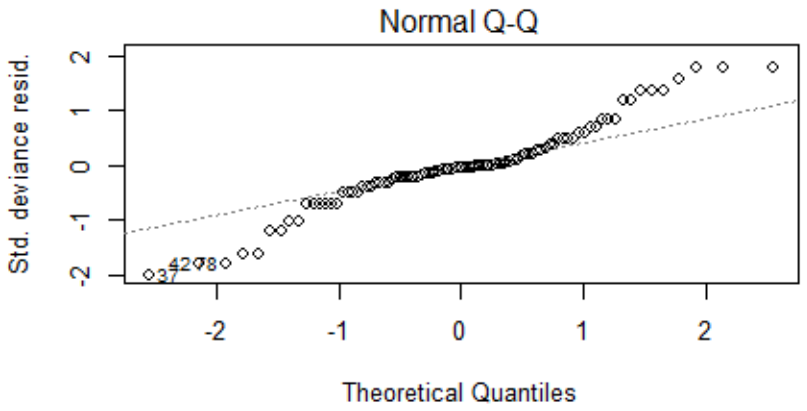
```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                Response   No. Observations:                   92
Model:                            GLM   Df Residuals:                       90
Model Family:                Binomial   Df Model:                            1
Link Function:                  logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -24.968
Date:                Wed, 26 Dec 2018   Deviance:                       49.937
Time:                        01:07:16   Pearson chi2:                     46.3
No. Iterations:                     7   Covariance Type:                nonrobust
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -20.4078      4.523     -4.512      0.000     -29.273     -11.542
Age            0.4259      0.095      4.492      0.000       0.240       0.612
==============================================================================
```

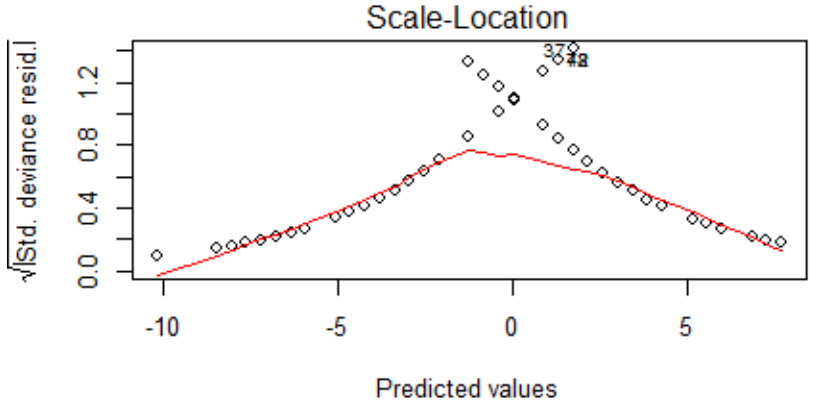
```
1 result.null_deviance
```

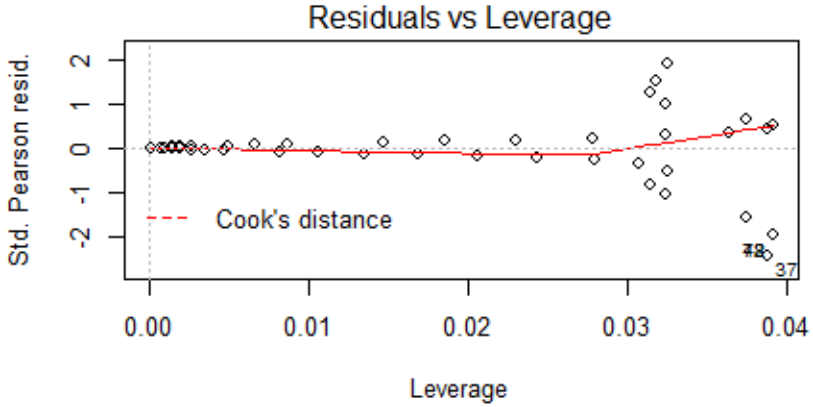
```
123.15634524584677
```

Residual Plot

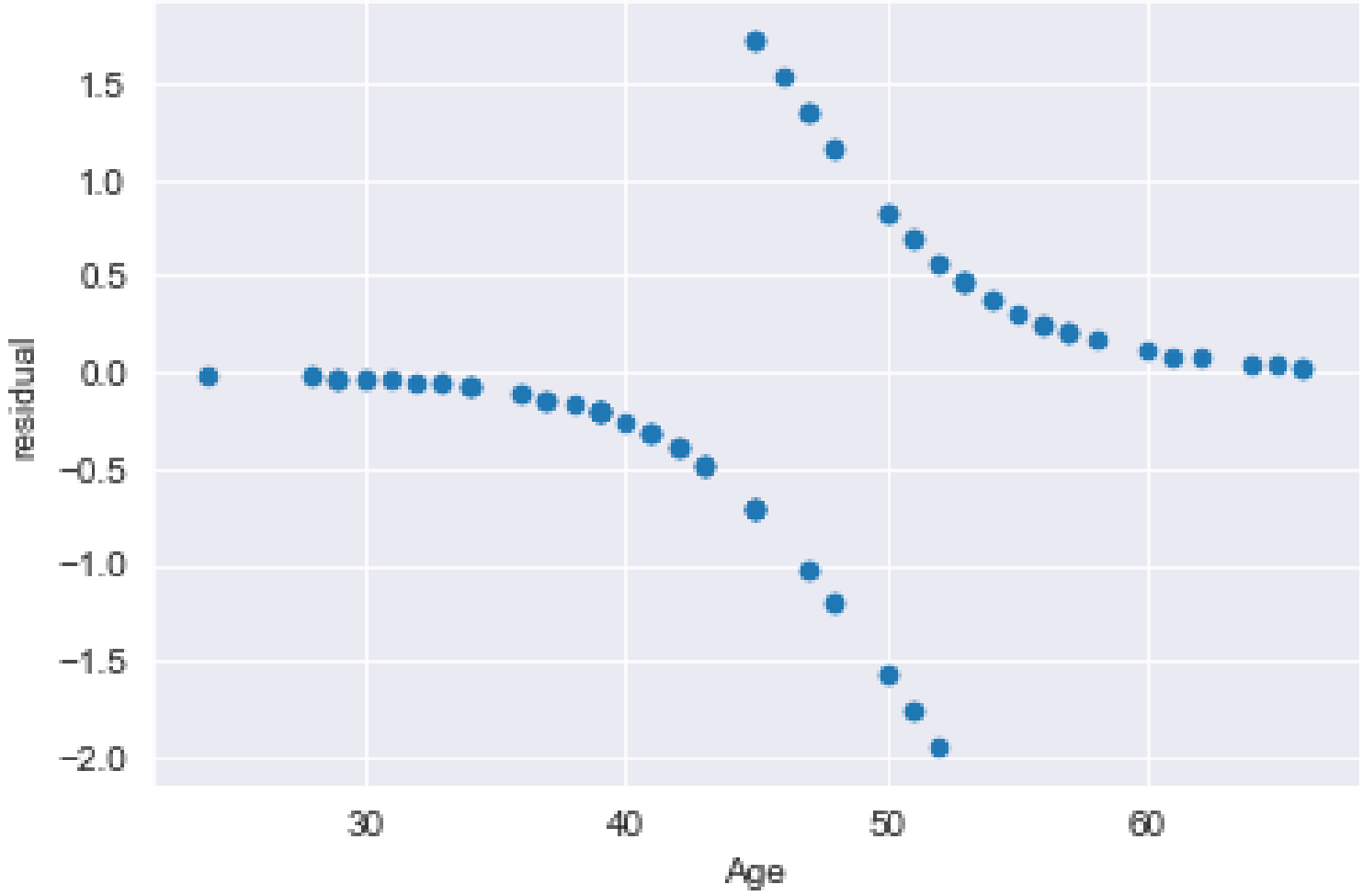








Understanding Residual Plot



Reference

Head First Statistics