

## 19msms26(assignment 6)

**Description:** The 'birthwt' data frame has 189 rows and 10 columns. The data were collected at Baystate Medical Center, Springfield, Mass during 1986. This data is taken from Library Mass of R.

**Format:**

This data frame contains the following columns:

'low' indicator of birth weight less than 2.5kg '

age' mother's age in years

'lwt' mother's weight in pounds at last menstrual period

'race' mother's race ('1' = white, '2' = black, '3' = other)

'smoke' smoking status during pregnancy

'ptl' number of previous premature labors

'ht' history of hypertension

'ui' presence of uterine irritability

'ftv' number of physician visits during the first trimester

'bwt' birth weight in grams

**Objective:-**The data contains risk factors associated with low infant birth weight. The question of interest here is how health and actions of a mother during pregnancy affects the birth weight of a baby. The variable of interest "low" is an indicator of birth weight less than 2.5 kg (5 lbs 8 ounces) in which "1" an unhealthy birth weight less than 2.5 kg and "0" represented a healthy birth weight above 2.5 kg.

### **Analysis:-**

Since our response variable is a factor variable with binary outcomes, so fitting Binomial glm with logit link function might be a suitable model.

General model :-

**The Model:** 
$$\text{Ln}\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

$$\text{Odds}(Y = 1) = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}$$

Where  $e^\beta$  is the ratio of the odds of two values of X one unit apart.

Here, Low is our response variable (Y) .

**Model 1:**

$$\text{logit}(P(Y = 1)) = \alpha + \beta \text{ smoke}(\text{for level 1})$$

In this we will first look at whether a child is born with a below normal birth weight is affected by smoking habit of mothers during pregnancy.

```
Call:
glm(formula = low ~ smoke, family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0197  -0.7623  -0.7623   1.3438   1.6599

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.0871     0.2147  -5.062 4.14e-07 ***
smoke1        0.7041     0.3196   2.203  0.0276 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 234.67  on 188  degrees of freedom
Residual deviance: 229.80  on 187  degrees of freedom
AIC: 233.8

Number of Fisher Scoring iterations: 4
```

Looking at the results from this first model we can see that being smoker does affect on whether a child is born below normal birth weight and both intercept as well as smoking habit are significant and have p-values less than 0.05.

## Model 2:-

In this model we will see the effect of race and add on race in the model 1 and refit the model.

$$\text{logit}(P(Y = 1)) = \alpha + \beta \text{smoke}(\text{for level 1}) + \gamma \text{race}(\text{white})$$

**Note:** - I have converted race in two level factor variables i.e. white and non-white.

```
glm(formula = dt2$low ~ dt2$smoke + dt2$race, family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3402  -0.8840  -0.5433   1.4968   1.9930

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.7382     0.2379  -3.103  0.00191 **
dt2$smoke1    1.1130     0.3643   3.056  0.00225 **
dt2$racewhite -1.1003     0.3645  -3.019  0.00254 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 234.67  on 188  degrees of freedom
Residual deviance: 219.98  on 186  degrees of freedom
AIC: 225.98

Number of Fisher Scoring iterations: 4
```

Here we observed that white race mothers are less likely to give birth to a low weight baby. The possible reasons might be the depending on the race of mothers; they may get different treatment and care. And all the terms in the model is significant. Also aic and residual deviance is less than the model 1.

## Model3:-

$$\text{logit}(P(Y = 1)) = \alpha + \beta \text{smoke}(\text{for level 1}) + \gamma \text{race}(\text{white}) + c \text{hypertension}(\text{level 1}) + d \text{ui}(\text{level 1})$$

In this model we will see the if hypertension and presence of uterine irritability will affect the on whether baby is born underweight in presence of smoking habits and affect of race. Because from the figure 1 and figure 2 in appendix we observed that the birth weight for babies whose mothers have history of hypertension and uterine irritability present tend to lower compare to those whose mothers have these two present.

```
glm(formula = low ~ race + smoke + hypertension + ui, family = binomial,
    data = dt2)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6588	-0.7920	-0.4829	1.1445	2.1009

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.0136	0.2637	-3.844	0.000121 ***
racewhite	-1.0766	0.3749	-2.872	0.004082 **
smoke1	1.0915	0.3742	2.917	0.003532 **
hypertension	1.3582	0.6290	2.159	0.030822 *
ui1	1.0067	0.4378	2.299	0.021485 *

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 234.67 on 188 degrees of freedom

Residual deviance: 211.17 on 184 degrees of freedom

AIC: 221.17

Number of Fisher Scoring iterations: 4

We observed that this model is better fit to data compare to both model1 & 2 in terms of aic and residual deviance both.

**Conclusion:** - From the analysis of deviance below, we observed the residual deviance and aic is lowest for the model.

Analysis of Deviance Table

```
Model 1: low ~ 1
Model 2: low ~ smoke
Model 3: low ~ smoke + race
Model 4: low ~ race + smoke + hypertension + ui
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      188      234.67
2      187      229.81  1    4.8674 0.027369 *
3      186      219.98  1    9.8274 0.001719 **
4      184      211.17  2    8.8097 0.012218 *
---
```

**Options available in R to implement Analysis of deviance in R:-**

1. Using the anova command we can get the analysis of deviance table for one or more models.
2. Using the deviance command in R specifying the type of deviance we can get the different deviances but the con is we need to specify the specific type of deviance, where as in anova command we get the full table of null, residual deviance and Deviance with the degree of freedoms.

**Appendix:-**

Figure 1 and Figure 2

