

# **STAT 226 Lecture 19**

## **Residuals for Logit Models**

### **Section 5.2.4-5.2.5**

---

Yibi Huang

# **Residuals for Binomial Response Models**

---

## Pearson Residuals & Standardized Pearson Residuals

When goodness-of-fit test suggests a GLM fits poorly, residuals can highlight where the fit is poor.

$$\text{Pearson Residual } e_i = \frac{y_i - n_i \widehat{\pi}_i}{\sqrt{n_i \widehat{\pi}_i (1 - \widehat{\pi}_i)}}$$

$$\text{Standardized (Pearson) Residual } r_i = \frac{e_i}{\sqrt{1 - h_i}}$$

- $h_i = \textit{leverage}$  of the observation  $i$  (details are skipped).  
The greater an observation's leverage, the greater its influence on the model fit.
- Note  $\sum_i e_i^2 = X^2$  (Pearson chi-square)
- When model holds and  $n_i \widehat{\pi}_i$  are large,
  - $e_i$  is approx.  $N(0, \nu)$  but  $\nu < 1$ ,  $r_i$  is approx.  $N(0, 1)$
  - $|r_i| > 2$  or  $3$  means lack of fit
- Useful for grouped data only

## Deviance Residuals for Binomial Response Models

The **deviance residual** is defined as

$$d_i = \text{sign}(y_i - \widehat{\mu}_i) \sqrt{2 \left[ y_i \log \left( \frac{y_i}{\widehat{\mu}_i} \right) + (n_i - y_i) \log \left( \frac{n_i - y_i}{n_i - \widehat{\mu}_i} \right) \right]}$$

where  $\widehat{\mu}_i = n_i \widehat{\pi}(\mathbf{x}_i)$ .

**Standardized deviance residual** =  $\frac{d_i}{\sqrt{1 - h_i}}$  where  $h_i$  is leverage.

- Observe that  $\sum_i d_i^2 = \text{Deviance}$
- When model holds and  $n_i \widehat{\pi}_i$  large
  - $d_i$  approx.  $N(0, \nu)$  but  $\nu < 1$ , should use standardized  $d_i$
- Useful for grouped data only.

## Example (Berkeley Graduate Admissions)

Dept	Men			Women		
	Number Admitted	Number Rejected	Percent Admitted	Number Admitted	Number Rejected	Percent Admitted
A	512	313	62%	89	19	82%
B	353	207	63%	17	8	68%
C	120	205	37%	202	391	34%
D	138	279	33%	131	244	35%
E	53	138	28%	94	299	24%
F	22	351	6%	24	317	7%

```
UCB = read.table(  
  "http://www.stat.uchicago.edu/~yibi/s226/UCBadmissions.txt", h=T)  
UCB.fit1 = glm(cbind(Admitted, Rejected) ~ Dept + Gender,  
  family=binomial, data=UCB)
```

```
summary(UCB.fit1)
```

```
Deviance Residuals:
```

```
      1      2      3      4      5      6      7      8
-1.2487 -0.0560  1.2533  0.0826  1.2205 -0.2076  3.7189  0.2706
      9     10     11     12
-0.9243 -0.0858 -0.8509  0.2052
```

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.68192    0.09911   6.880 5.97e-12 ***
DeptB        -0.04340    0.10984  -0.395   0.693
DeptC        -1.26260    0.10663 -11.841 < 2e-16 ***
DeptD        -1.29461    0.10582 -12.234 < 2e-16 ***
DeptE        -1.73931    0.12611 -13.792 < 2e-16 ***
DeptF        -3.30648    0.16998 -19.452 < 2e-16 ***
GenderMale   -0.09987    0.08085  -1.235   0.217
```

```
---
```

```
Null deviance: 877.056 on 11 degrees of freedom
Residual deviance: 20.204 on 5 degrees of freedom
AIC: 103.14
```

LRT indicates strong Dept effect, but little Gender effect ( $P$ -value  $\approx 0.22$ ).  $\Rightarrow$  little evidence of gender bias in UCB graduate admissions.

```
drop1(UCB.fit1, test="Chisq")
```

Single term deletions

Model:

```
cbind(Admitted, Rejected) ~ Dept + Gender
```

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		20	103		
Dept	5	784	857	763	<2e-16
Gender	1	22	103	2	0.22

However, ...

However, goodness of fit test shows the main effect model fits poorly. The Deviance = 20.204 can be obtained from the summary output, or from the commands below

```
UCB.fit1$deviance  
[1] 20.204
```

The  $P$ -value for goodness of fit test  $\approx 0.00114$  is computed as follows.

```
pchisq(20.204, df=5, lower.tail=F)  
[1] 0.001144
```

Apparently there is Gender\*Dept interaction (because the saturated model is the two-way interaction model).



R function `residuals()` gives deviance residuals by default, and Pearson residuals with option `type="pearson"`.

```
residuals(UCB.fit1) # deviance residuals
  1      2      3      4      5      6      7      8
-1.24867 -0.05601  1.25334  0.08257  1.22051 -0.20756  3.71892  0.27061
  9     10     11     12
-0.92434 -0.08577 -0.85093  0.20518
residuals(UCB.fit1, type="pearson") # Pearson residuals
  1      2      3      4      5      6      7      8
-1.25381 -0.05602  1.26287  0.08261  1.24151 -0.20620  3.51867  0.26895
  9     10     11     12
-0.92078 -0.08573 -0.84403  0.20648
```

By default, R function `rstandard()` gives standardized deviance residuals.

```
rstandard(UCB.fit1)
```

```
      1      2      3      4      5      6      7      8      9
-4.0108 -0.2797  1.8666  0.1412  1.6059 -0.3046  4.2565  0.2814 -1.8881
     11     12
-1.6468  0.3007
```

With option `type="pearson"`, `rstandard()` reports standardized Pearson residuals.

```
rstandard(UCB.fit1, type="pearson")
```

```
      1      2      3      4      5      6      7      8      9
-4.0273 -0.2797  1.8808  0.1413  1.6335 -0.3026  4.0273  0.2797 -1.8808
     11     12
-1.6335  0.3026
```

```

pearson.res = round(residuals(UCB.fit1, type="pearson"),2)
std.res = round(rstandard(UCB.fit1,type="pearson"), 2)
cbind(UCB, pearson.res, std.res)

```

	Gender	Dept	Admitted	Rejected	pearson.res	std.res	
1	Male	A	512	313	-1.25	-4.03	<--
2	Male	B	353	207	-0.06	-0.28	
3	Male	C	120	205	1.26	1.88	
4	Male	D	138	279	0.08	0.14	
5	Male	E	53	138	1.24	1.63	
6	Male	F	22	351	-0.21	-0.30	
7	Female	A	89	19	3.52	4.03	<--
8	Female	B	17	8	0.27	0.28	
9	Female	C	202	391	-0.92	-1.88	
10	Female	D	131	244	-0.09	-0.14	
11	Female	E	94	299	-0.84	-1.63	
12	Female	F	24	317	0.21	0.30	

Standardized Pearson residuals suggest Dept. A as main source of lack of fit ( $r_i = -4.03$  and  $4.03$ ), while Pearson residuals fail to catch the lack of fit of the first observation (Gender = Male, Dept = A).

Leaving out Dept. A, the model with Dept main effects and gender main effects fits well (Deviance = 2.556, df = 4,  $P$ -value  $\approx$  0.63.)

```
UCB.fit2 = glm(cbind(Admitted,Rejected) ~ Dept + Gender,  
              family=binomial, data=UCB, subset=(Dept != "A"))  
UCB.fit2$deviance  
[1] 2.556  
UCB.fit2$df.residual  
[1] 4  
pchisq(2.556, df=4, lower.tail=F)  
[1] 0.6346
```

Knowing the main effect model fits the data well when leaving out Dept. A, we can use it to do inference.

LRT shows Gender effect is not significant ( $P$ -value = 0.72), little evidence of Gender bias in admissions in Dept. B-F.

```
drop1(UCB.fit2, test="Chisq")
Single term deletions

Model:
cbind(Admitted, Rejected) ~ Dept + Gender
      Df Deviance AIC LRT Pr(>Chi)
<none>      3  72
Dept      4    501 562 498  <2e-16
Gender    1      3  70  0    0.72
```

In Dept. A, odds of admission for men are  $\frac{512 \times 19}{313 \times 89} = 0.35$  times the odds for women.

Dept A	Admitted	Rejected
Male	512	313
Female	89	19

## Conclusion:

- In Dept. A, women are more likely to be admitted
- In Dept. B-F, no significant diff. in admission rates of men and women.

## Conclusion:

- In Dept. A, women are more likely to be admitted
- In Dept. B-F, no significant diff. in admission rates of men and women.

However, if we ignore Dept, Gender is significant but in the opposite direction — odds of admission for men are  $e^{0.61} = 1.84$  times the odds for women, Men were significantly more likely to be admitted. Why?

```
UCB3 = glm(cbind(Admitted,Rejected) ~ Gender, family=binomial, data=UCB)
summary(UCB3)$coef
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.8305	0.05077	-16.357	3.868e-60
GenderMale	0.6104	0.06389	9.553	1.263e-21

## Conclusion:

- In Dept. A, women are more likely to be admitted
- In Dept. B-F, no significant diff. in admission rates of men and women.

However, if we ignore Dept, Gender is significant but in the opposite direction — odds of admission for men are  $e^{0.61} = 1.84$  times the odds for women, Men were significantly more likely to be admitted. Why?

```
UCB3 = glm(cbind(Admitted,Rejected) ~ Gender, family=binomial, data=UCB)
summary(UCB3)$coef
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.8305	0.05077	-16.357	3.868e-60
GenderMale	0.6104	0.06389	9.553	1.263e-21

- This is an example of Simpson's paradox.



## **Sparse Data**

---

**Caution:** Parameter Estimates in Logistic Models Can be *Infinite!*

Example 1:

	<i>S</i>	<i>F</i>
<i>X</i> = 1	8	2
<i>X</i> = 2	10	0

Model:

$$P(S) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

$$e^{\hat{\beta}} = \text{odds-ratio} = \frac{8 \times 0}{2 \times 10} = 0 \quad \Rightarrow \quad \hat{\beta} = \text{log-odds-ratio} = -\infty$$

Empty cells in multi-way contingency table can cause infinite estimates.

Software may not realize this, and gives a finite estimate!

- Large **Number of Fisher Scoring iterations** is a warning sign of infinite parameter estimate(s)
- Large values of SEs for coefficients are also warning signs

```
S = c(8,10)
F = c(2,0)
X = c(1,2)
glm1 = glm(cbind(S,F) ~ X, family = binomial)
summary(glm1)
```

Call:

```
glm(formula = cbind(S, F) ~ X, family = binomial)
```

Deviance Residuals:

```
[1] 0 0
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-22.35	54605.92	0	1
X	23.73	54605.92	0	1

Null deviance: 2.9953e+00 on 1 degrees of freedom

Residual deviance: 2.4675e-10 on 0 degrees of freedom

AIC: 6.3947

Number of Fisher Scoring iterations: 22

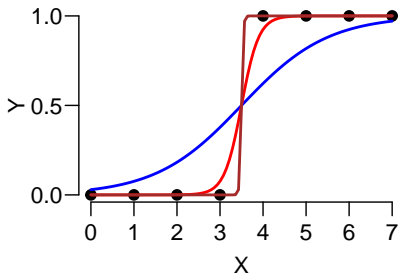
Infinite estimates exist when  $x$ -values with  $y = 1$  can be *perfectly separated* from  $x$ -values with  $y = 0$ .

Example 2:

$X = c(0, 1, 2, 3, 4, 5, 6, 7)$

$Y = c(0, 0, 0, 0, 1, 1, 1, 1)$

Model:  $\pi(x) = P(Y = 1|x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$



$$\pi(x) = \begin{cases} \frac{\exp(x - 3.5)}{1 + \exp(x - 3.5)} & \text{blue} \\ \frac{\exp(5(x - 3.5))}{1 + \exp(5(x - 3.5))} & \text{red} \\ \frac{\exp(50(x - 3.5))}{1 + \exp(50(x - 3.5))} & \text{brown} \end{cases}$$

The higher the  $\beta$  value, the closer the logistic curve to the data points.

```
glm2 = glm(Y ~ X, family = binomial)
```

```
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(glm2)
```

```
Call:
```

```
glm(formula = Y ~ X, family = binomial)
```

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-1.504e-05	-2.110e-08	0.000e+00	2.110e-08	1.504e-05

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-160.3	285119.4	-0.001	1
X	45.8	80643.9	0.001	1

```
Null deviance: 1.1090e+01 on 7 degrees of freedom
```

```
Residual deviance: 4.5253e-10 on 6 degrees of freedom
```

```
AIC: 4
```

```
Number of Fisher Scoring iterations: 25
```

### Example 3.

```
X1 = c(0.5, 2, 10, 6, 8, 1, 8, 4, 0, 7, 5, 1.1)
```

```
X2 = c(7, 9.5, 6, 0, 8.1, 9, 10, 0, 8, 5, 3, 4)
```

```
Y = c(0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0)
```

```
summary(glm(Y ~ X1+X2, family = binomial))
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-0.000018180	-0.000000021	0.000000000	0.000000021	0.000018302

Coefficients:

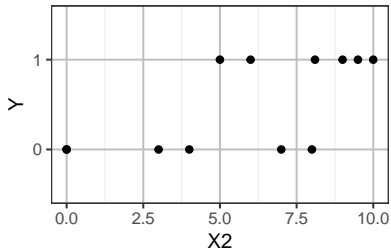
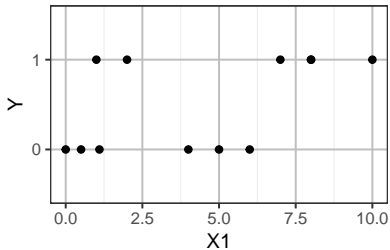
	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-208.0	1127014.6	0	1
X1	21.8	141530.0	0	1
X2	23.2	139959.6	0	1

Null deviance: 16.63553233343869 on 11 degrees of freedom

Residual deviance: 0.00000000066586 on 9 degrees of freedom

AIC: 6

Number of Fisher Scoring iterations: 25



Observe points with  $Y = 1$  and  $Y = 0$  are completely separated in the  $X_1$ - $X_2$  plane.

