

Likelihood and Deviance

Goals for today

- + Summarize the fit of a logistic regression model
- + Compare fit for nested models

$$SSE = \sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2$$

Logistic regression

Data: Grad application data

- + admit: accepted to grad school? (0 = no, 1 = yes)
- + gre: GRE score
- + gpa: undergrad GPA
- + rank: prestige of undergrad institution

$$Admit_i \sim Bernoulli(\pi_i) \quad \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 GPA_i$$

Goal: Summarize how well our model fits the data

How do we summarize model fit in *linear* regression?

$$R^2 = 1 - \frac{SSE}{SST_{\text{total}}}$$

$$R^2_{\text{adj}} = 1 - \frac{SSE / (n-p)}{SST_{\text{total}} / (n-1)}$$

observations
parameters

$$R^2 = 1 - \frac{SSE}{SST_{\text{total}}}$$

Summarizing linear regression model fit

Example model:

$$\begin{aligned} SSE &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ \hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i + \dots \end{aligned}$$

```
linear_mod <- lm(gre ~ gpa, data = grad_app) with a binary response, we
summary(linear_mod) don't use SSE to fit model
```

...

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	meaningful
## (Intercept)	192.30	47.92	4.013	7.15e-05	***
## gpa	116.64	14.05	8.304	1.60e-15	***
## Residual standard error:	106.8	on 398 degrees of freedom			
## Multiple R-squared:	0.1477,	Adjusted R-squared:	0.1		

...

R^2

R^2_{adj}

Is R^2_{adj} appropriate for logistic regression? No!

Summarizing logistic regression

```
logistic_mod <- glm(admit ~ gpa, data = grad_app,  
                      family = binomial)  
summary(logistic_mod)
```

```
...  
##             Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -4.3576     1.0353  -4.209 2.57e-05 ***  
## gpa         1.0511     0.2989   3.517 0.000437 ***  
## ---  
## Null deviance: 499.98 on 399 degrees of freedom  
## Residual deviance: 486.97 on 398 degrees of freedom
```

...

R reports deviance instead of R^2 , R^2_{adj}
for logistic regression

$$-2(\log \text{likelihood for a model w/ GPA}) = 486.97$$

Deviance

```
logistic_mod <- glm(admit ~ gpa, data = grad_app,  
                      family = binomial)  
summary(logistic_mod)
```

```
...  
##             Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -4.3576     1.0353  -4.209 2.57e-05 ***  
## gpa          1.0511     0.2989   3.517 0.000437 ***  
## ---  
## Null deviance: 499.98 on 399 degrees of freedom  
## Residual deviance: 486.97 on 398 degrees of freedom
```

For logistic regression

$$\text{deviance} = -2(\log \text{likelihood})$$

Like SSE in linear reg., want to minimize deviance

- larger likelihood \Rightarrow smaller deviance
- makes #S more manageable

Class activity

https://sta214-s23.github.io/class_activities/ca_lecture_6.html

Class activity: deviance

```
glm(admit ~ gre, family = binomial, data = grad_app)
```

```
...
## Degrees of Freedom: 399 Total (i.e. Null); 398 Residual
## Null Deviance: 500
## Residual Deviance: 486.1          AIC: 490.1
...
...
```

What is the deviance of my fitted model?

$$-2 \log L = 486.1$$

$$\Rightarrow \log L = -\frac{1}{2} (486.1) = -243.05$$

$$\Rightarrow L = e^{-243.05}$$

Class activity: deviance

```
glm(admit ~ gre, family = binomial, data = grad_app)
```

```
...
## Degrees of Freedom: 399 Total (i.e. Null); 398 Residual
## Null Deviance: 500
## Residual Deviance: 486.1      AIC: 490.1
```

...

486.1 487



Which predictor (GRE or GPA) gives a model with a better fit?

smaller deviance \Rightarrow higher likelihood!

Class activity: deviance

```
glm(admit ~ gre, family = binomial, data = grad_app)
```

```
...
## Degrees of Freedom: 399 Total (i.e. Null); 398 Residual
## Null Deviance:      500
## Residual Deviance: 486.1      AIC: 490.1
...
```

Which predictor (GRE or GPA) gives a model with a better fit?

GRE has a slightly smaller deviance (486.1 vs. 487), so GRE gives a slightly better fit.

Class activity: hypotheses

$$Y_i \sim \text{Bernoulli}(\pi_i) \quad \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 \text{GRE}_i$$

We want to know if there is actually a relationship between GRE score and grad school admission.

How can I express this as null and alternative hypotheses about one or more model parameters?

$$H_0: \beta_1 = 0 \quad (\text{no relationship})$$

$$H_A: \beta_1 \neq 0 \quad (\text{some relationship})$$

Steps in hypothesis testing

- specify hypotheses

e.g.

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

- calculate a test statistic

e.g.: t statistic, F statistic, z statistic

- calculate a p-value

- draw a conclusion

intuition: how much does deviance change when we add GRE to the model?

Test statistic: drop in deviance

deviance of the intercept-only model (just β_0)

```
...  
##  
## Null deviance: 499.98 on 399 degrees of freedom  
## Residual deviance: 486.06 on 398 degrees of freedom  
...  
deviance of the full model
```

499.98 = deviance for intercept-only model

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 \quad \text{corresponds to } H_0: \beta_1 = 0$$

486.06 = deviance for full model

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 \text{GRE}_i \quad \text{corresponds to } H_A: \beta_1 \neq 0$$

drop-in-deviance:

$$G = \text{deviance of reduced model} - \text{deviance of full model}$$
$$= 499.98 - 486.06 = 13.92$$

Steps in hypothesis testing

$$Admit_i \sim Bernoulli(\pi_i) \quad \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 \text{GRE}_i$$

- + Specify hypotheses

$$H_0 : \beta_1 = 0 \quad H_A : \beta_1 \neq 0$$

- + Calculate a test statistic

G = deviance for reduced model - deviance for full model

(H_0)

(H_A)

- + Calculate a p-value

Calculating p-values

What is a p-value?

p-value = probability of "our data or more extreme",
if H_0 is true

$$= P(G \geq 13.92 \mid H_0 \text{ is true})$$

(how unusual is the observed, e.g. $G=13.92$,
if H_0 is actually true)

smaller p-values are stronger evidence against H_0

Calculating p-values

What is a p-value?

p-value: How unusual the data is if H_0 is true. e.g.,

$$P(G \geq 13.92 | H_0)$$

G is a random variable \Rightarrow I need a distribution for G under H_0

How do we calculate a p-value?

null distribution: distribution of G under H_0

we compare the observed G to the null distribution to see if it is unusual

Calculating p-values

What *is* a p-value?

p-value: How unusual the data is *if H_0* is true. e.g.,

$$P(G \geq 13.92 | H_0)$$

How do we calculate a p-value?

Compare the observed test statistic to the null distribution

How do we get the null distribution?

Exploring the null distribution with simulation

- + Want to know how G (drop in deviance) behaves if H_0 is true
- + So, need data for which we *know* H_0 is true!

See
hw 1

```
x <- runif(1000, 0, 5)
p <- exp(-3 + 0*x)/(1 + exp(-3 + 0*x))
y <- rbinom(1000, 1, p)
```

$$t_i \sim \text{Bernoulli}(\pi_i)$$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = -3 + 0x_i$$

```
m1 <- glm(y ~ x, family = binomial)
summary(m1)
```

$$H_0: \beta_1 = 0$$

...

##

```
## Null deviance: 442.86 on 999 degrees of freedom
## Residual deviance: 442.80 on 998 degrees of freedom
```

...

$$G = 442.86 - 442.80 = 0.06$$

next time: repeat many times!

Exploring the null distribution with simulation

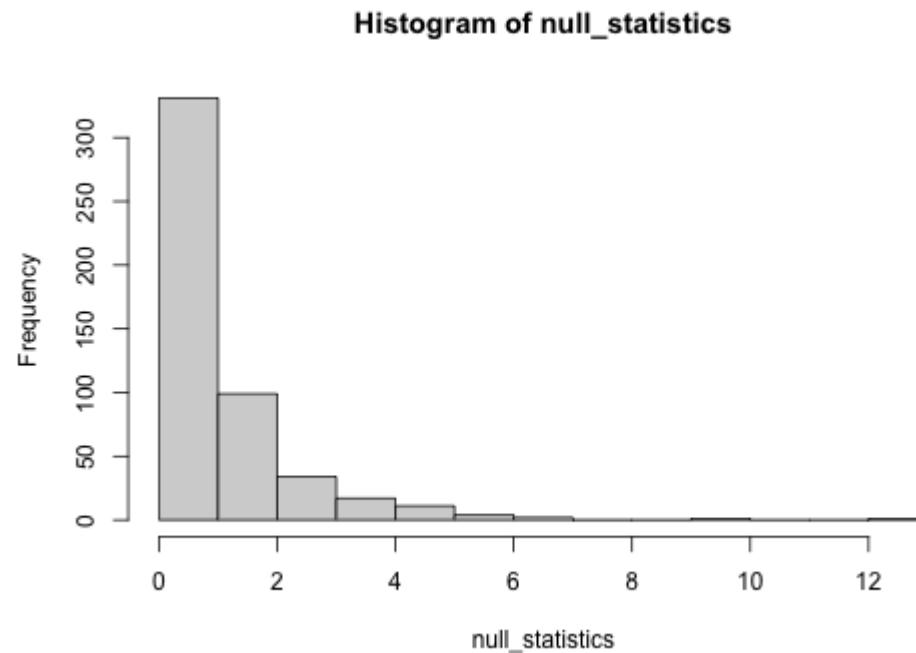
- + Simulating one set of data only gives us one statistic under H_0
- + We need to simulate many datasets to explore the full distribution

```
null_statistics <- c()
nsim <- 500
for(i in 1:nsim){
  x <- runif(1000, 0, 5)
  p <- exp(-3 + 0*x)/(1 + exp(-3 + 0*x))
  y <- rbinom(1000, 1, p)
  m1 <- glm(y ~ x, family = binomial)

  null_statistics[i] <- m1$null.deviance -
    m1$deviance
}
```

Exploring the null distribution with simulation

```
hist(null_statistics)
```



Are there any issues with this approach to approximating the null distribution?