Hypothesis testing for logistic regression

Modeling grad admissions

Data: Grad application data

- `admit`: accepted to grad school? (0 = no, 1 = yes)
- `gre`: GRE score
- `gpa`: undergrad GPA
- `rank`: prestige of undergrad institution ← categorical, 4 levels
  1 = most prestige,
  4 = least prestige

> We want to know whether there is a relationship between the prestige of a student's undergrad institution and the probability they are admitted to graduate school, *after accounting for their GRE score and GPA.*
> (put these variables in the model)
>
> How could we use hypothesis testing to investigate this research question?

(rank 1 is the reference category)

Modeling grad admissions

=1 rank=2

=0 else

$$Admit_i \sim Bernoulli(\pi_i)$$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 GRE_i + \beta_2 GPA_i + \beta_3 Rank2_i$$
$$+ \beta_4 Rank3_i + \beta_5 Rank4_i$$

> How would I interpret the coefficients in this model?

β_1: a unit increase in GRE is associated with a change in odds of admission by a factor of $\exp\{\beta_1\}$, holding GPA & rank fixed

β_3: Going to a rank 2 school is associated with a change in the odds of admission by a factor of $\exp(\beta_3)$ vs. a rank 1 school, holding GRE & GPA fixed

Hypotheses

$$Admit_i \sim Bernoulli(\pi_i)$$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 GRE_i + \beta_2 GPA_i + \beta_3 Rank2_i + \beta_4 Rank3_i + \beta_5 Rank4_i$$

Research question: Is there a relationship between the prestige of a student's undergrad institution and the probability they are admitted to graduate school, after accounting for their GRE score and GPA?

> What are my null and alternative hypotheses?

$H_0: \beta_3 = \beta_4 = \beta_5 = 0$

$H_A:$ at least one of $\beta_3, \beta_4, \beta_5 \neq 0$

Testing hypotheses

$$Admit_i \sim Bernoulli(\pi_i)$$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 GRE_i + \beta_2 GPA_i + \beta_3 Rank2_i + \beta_4 Rank3_i + \beta_5 Rank4_i$$

Hypotheses:

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_A : \text{at least one of } \beta_3, \beta_4, \beta_5 \neq 0$$

> How should we test these hypotheses and calculate a p-value?

- Calculate G = change in deviance between full & reduced models
- Need null distribution (distribution of G under H_0)

Class activity

https://sta214-s23.github.io/class_activities/ca_lecture_8.html

Class activity

Full model:

```
m1 <- glm(admit ~ gre + gpa + as.factor(rank),
          data = grad_app, family = binomial)
summary(m1)
```

deviance of intercept-only model

our reduced model is not intercept-only

```
...
##
##     Null deviance: 499.98  on 399  degrees of freedom
## Residual deviance: 458.52  on 394  degrees of freedom
...
```

Deviance of full model = 458.52

Class activity

$$\log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = -4.95 + 0.003\, GRE_i + 0.755\, GPA_i$$

Reduced model:

```
m2 <- glm(admit ~ gre + gpa,
          data = grad_app, family = binomial)
summary(m2)
```

```
...
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.949378   1.075093  -4.604 4.15e-06 ***
## gre          0.002691   0.001057   2.544   0.0109 *
## gpa          0.754687   0.319586   2.361   0.0182 *
## ---
##
##     Null deviance: 499.98  on 399  degrees of freedom
## Residual deviance: 480.34  on 397  degrees of freedom
...
```

Deviance for reduced model

$$G = 480.34 - 458.52 = 21.82$$

Class activity

```
nsim <- 500
null_statistics <- rep(0, nsim)
for(i in 1:nsim){
  x <- grad_app$gre
  p <- exp(-0.77 + 0*x)/(1 + exp(-0.77 + 0*x))
  y <- rbinom(length(x), 1, p)
  m1 <- glm(y ~ x, family = binomial)
  null_statistics[i] <- m1$null.deviance -
    m1$deviance
}
hist(null_statistics)
```

need to change (need GPA, GRE, rank)

need to change

need to change (include all variables)

need deviance of reduced model

also need to fit a reduced model

How do I modify this code to simulate from our reduced model?

Class activity

```
nsim <- 500
null_statistics <- rep(0, nsim)
for(i in 1:nsim){
  x1 <- grad_app$gre
  x2 <- grad_app$gpa
  x3 <- grad_app$rank
  p <- exp(-4.95 + 0.003*x1 + 0.755*x2)/(
    1 + exp(-4.95 + 0.003*x1 + 0.755*x2))
  y <- rbinom(length(x1), 1, p)

  m1 <- glm(y ~ x1 + x2 + as.factor(x3),
            family = binomial)
  m2 <- glm(y ~ x1 + x2,
            family = binomial)

  null_statistics[i] <- m2$deviance - m1$deviance
}
```

explanatory variables

simulate from reduced model

← full model

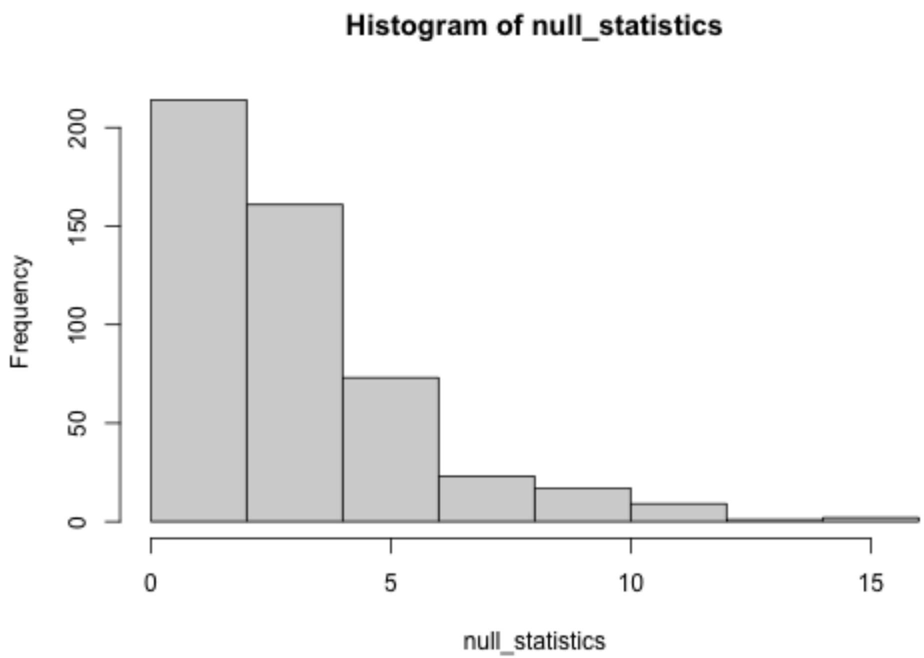
← reduced model

drop in deviance

Class activity

```
hist(null_statistics)
```

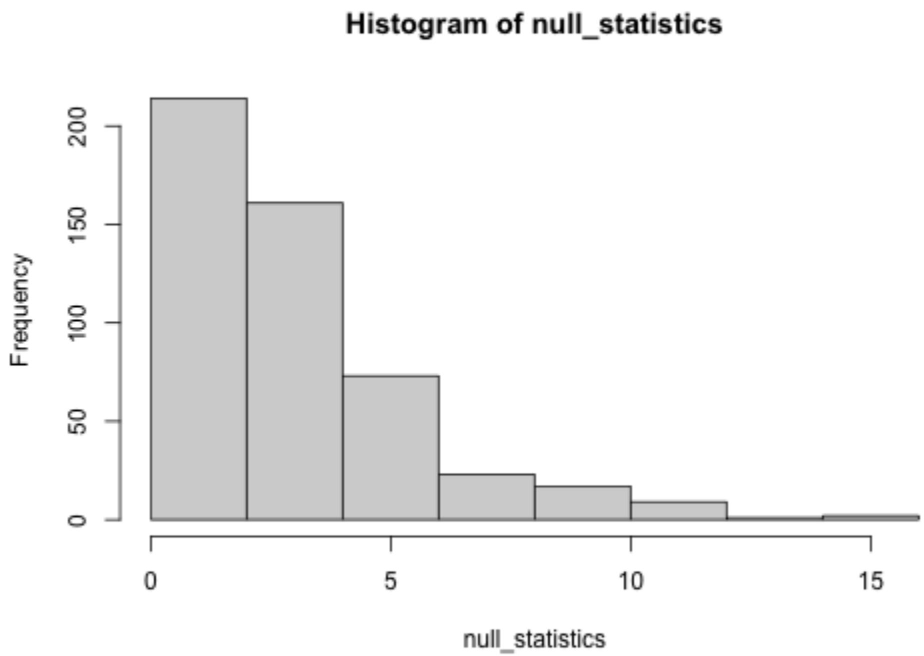
observed $G = 21.82$



Calculating a p-value

p -value = $P(G \geq 21.82 \mid H_0)$

$\approx$ fraction of simulations where $G \geq 21.82$




```
mean(null_statistics >= 21.82)
```

```
## [1] 0
```

p -value ≈ 0 $\Rightarrow$ strong evidence against H_0

Parametric bootstrapping

Goal: Test hypotheses comparing full and reduced models

Step 1: Fit the models and calculate a test statistic

Step 2: Repeat the following many times:

- Simulate from the reduced model
- Fit full and reduced models on simulated data
- Calculate a test statistic with the simulated data

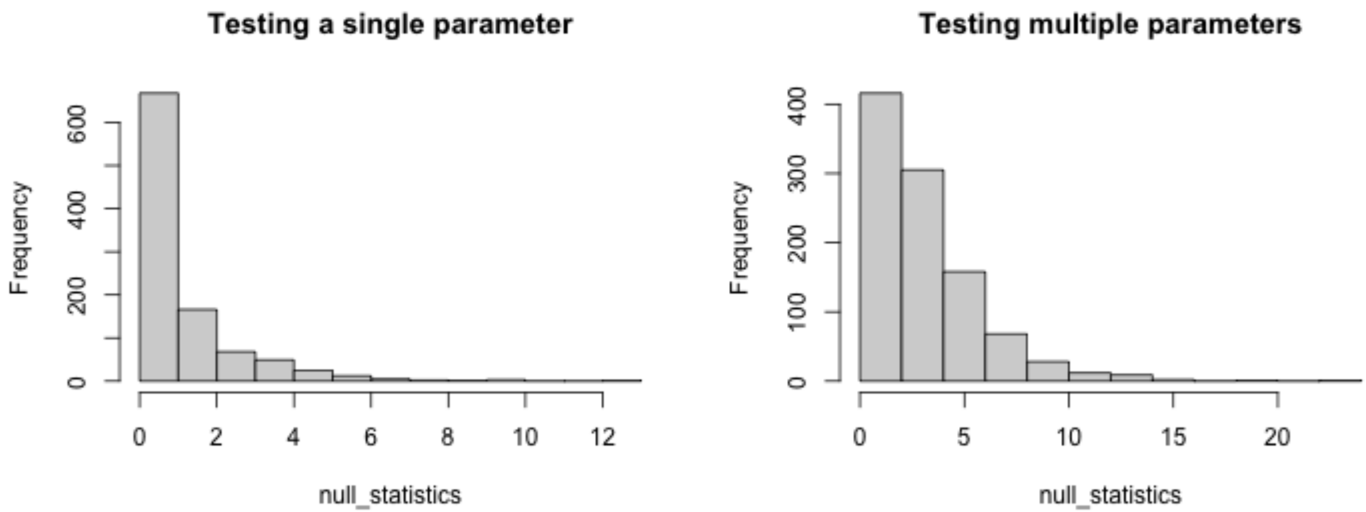
Step 3: Compare observed test statistic (step 1) with simulated test statistics (step 3) to calculate p-value

> Are there any disadvantages to this method?

- finite # of repetitions
- requires a lot of work to get null distribution

Examining the null distributions

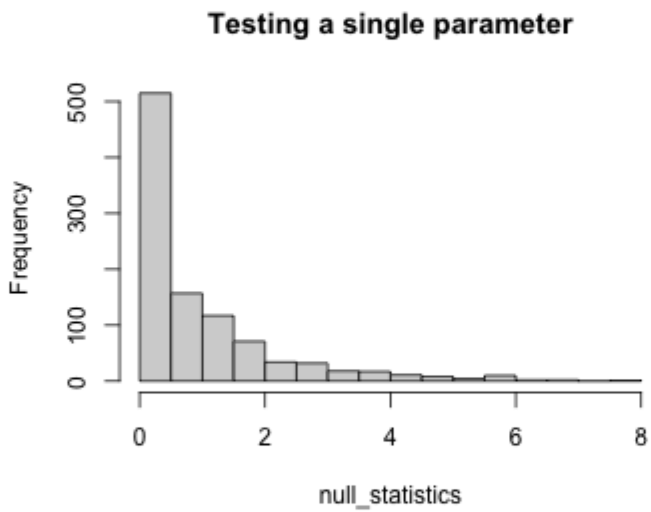
$H_0: \beta_1 = 0$ $H_0: \beta_3 = \beta_4 = \beta_5 = 0$

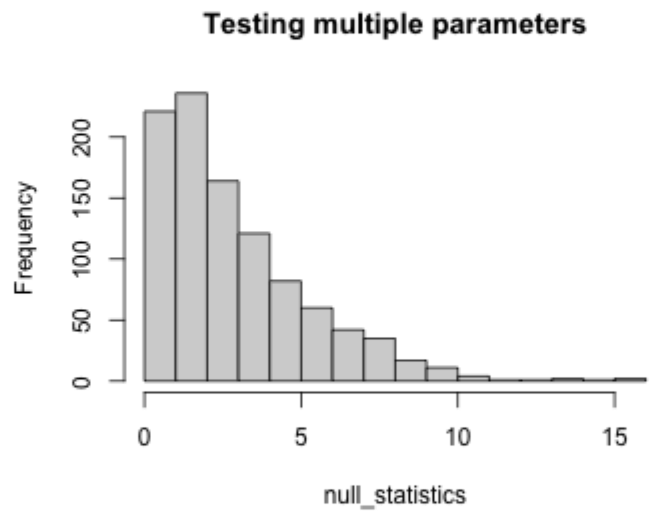


> What do you notice about these distributions?

- when testing multiple parameters, G tends to be larger
- right-skewed

Examining the null distributions





Why do we always have $G \geq 0$?

G = deviance reduced − deviance full

we minimize deviance (equiv. to maximizing likelihood) when we fit our model

χ^2 distribution

\# obs

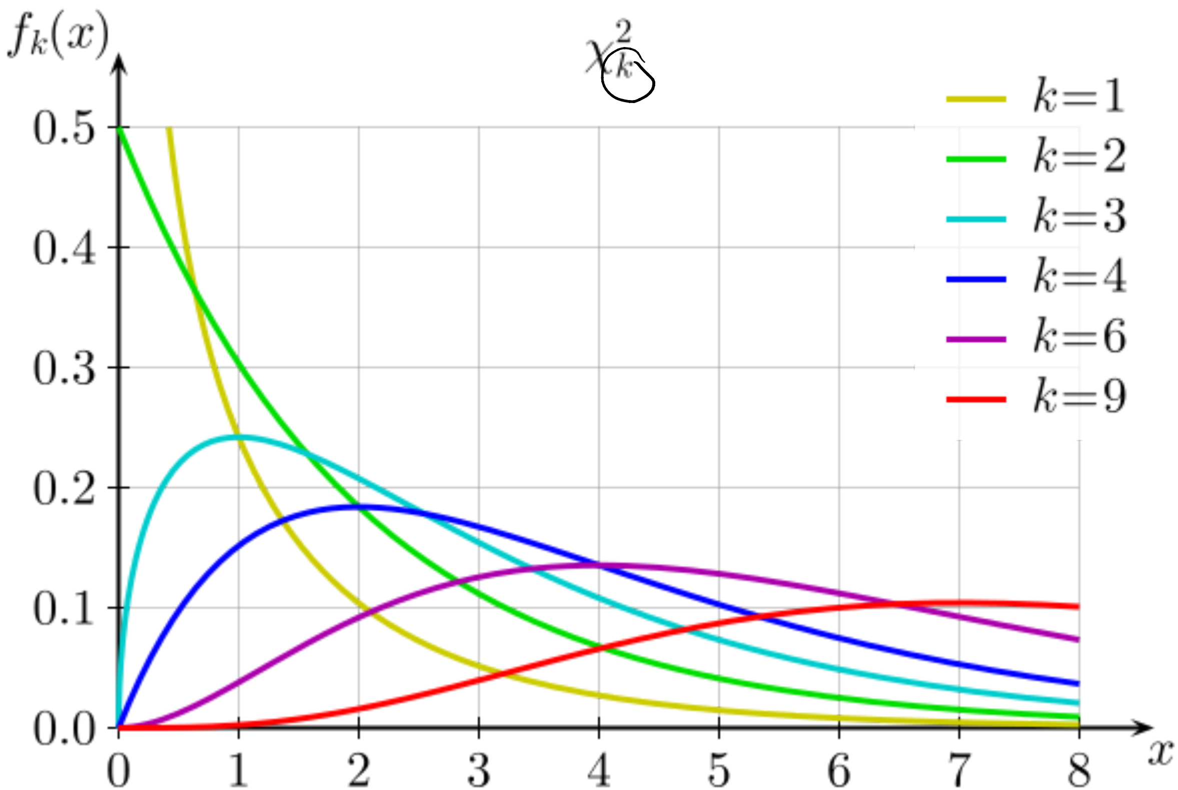
\# parameters

$n - p_{\text{reduced}}$

$n - p_{\text{full}} \leftarrow$ \# parameters

Under $H_0, G \sim \chi^2_{\underbrace{df_{\text{reduced}} - df_{\text{full}}}_{\text{\# parameters tested}}}$ (i.e. # βs in H_0)

χ^2_k distribution: parameterized by degrees of freedom k



Computing a p-value

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

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

$G =$ deviance for reduced model - deviance for full model = 13.92
$\sim \chi^2_1$

"probability for a χ^2"

```
pchisq(13.92, df = 1, lower.tail=FALSE)
```

```
## [1] 0.0001907579
```

p-value

Computing a p-value

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 GRE_i + \beta_2 GPA_i + \beta_3 Rank2_i$$
$$+ \beta_4 Rank3_i + \beta_5 Rank4_i$$

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_A : \text{at least one of } \beta_3, \beta_4, \beta_5 \neq 0$$

$G =$ deviance for reduced model - deviance for full model = 21.82

```
pchisq(..., df = ..., lower.tail=FALSE)
```

21.82 3

Computing a p-value

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 GRE_i + \beta_2 GPA_i + \beta_3 Rank2_i + \beta_4 Rank3_i + \beta_5 Rank4_i$$

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_A : \text{at least one of } \beta_3, \beta_4, \beta_5 \neq 0$$

$G =$ deviance for reduced model - deviance for full model = 21.82

```
pchisq(21.82, df = 3, lower.tail=FALSE)
```

```
## [1] 7.110521e-05
```

close to 0

Likelihood ratio test for nested models

Goal: Compare full and reduced models

Steps:

- Calculate deviance for full and reduced models
- G = deviance for reduced - deviance for full
- p-value: $G \sim \chi^2_q$

Alternative: Wald tests for single parameters

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

Hypotheses:

Test statistic:

$$z =$$

Example

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

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

```
...
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.901344   0.606038  -4.787 1.69e-06 ***
## gre          0.003582   0.000986   3.633  0.00028 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
...
```

$z =$

Wald tests vs. likelihood ratio tests

Wald test

- like t-tests
- test a single parameter
- some example hypotheses:
  - $H_0 : \beta_1 = 0$ vs.
    $H_A : \beta_1 \neq 0$
  - $H_0 : \beta_1 = 1$ vs.
    $H_A : \beta_1 > 1$

Likelihood ratio test

- like nested F-tests
- test one or more parameters
- some example hypotheses:
  - $H_0 : \beta_1 = 0$ vs.
    $H_A : \beta_1 \neq 0$

p-values are different, because test statistics and distributions are different