Inference with logistic regression

Last time: comparing deviances

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

Reduced model: $$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0$$

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

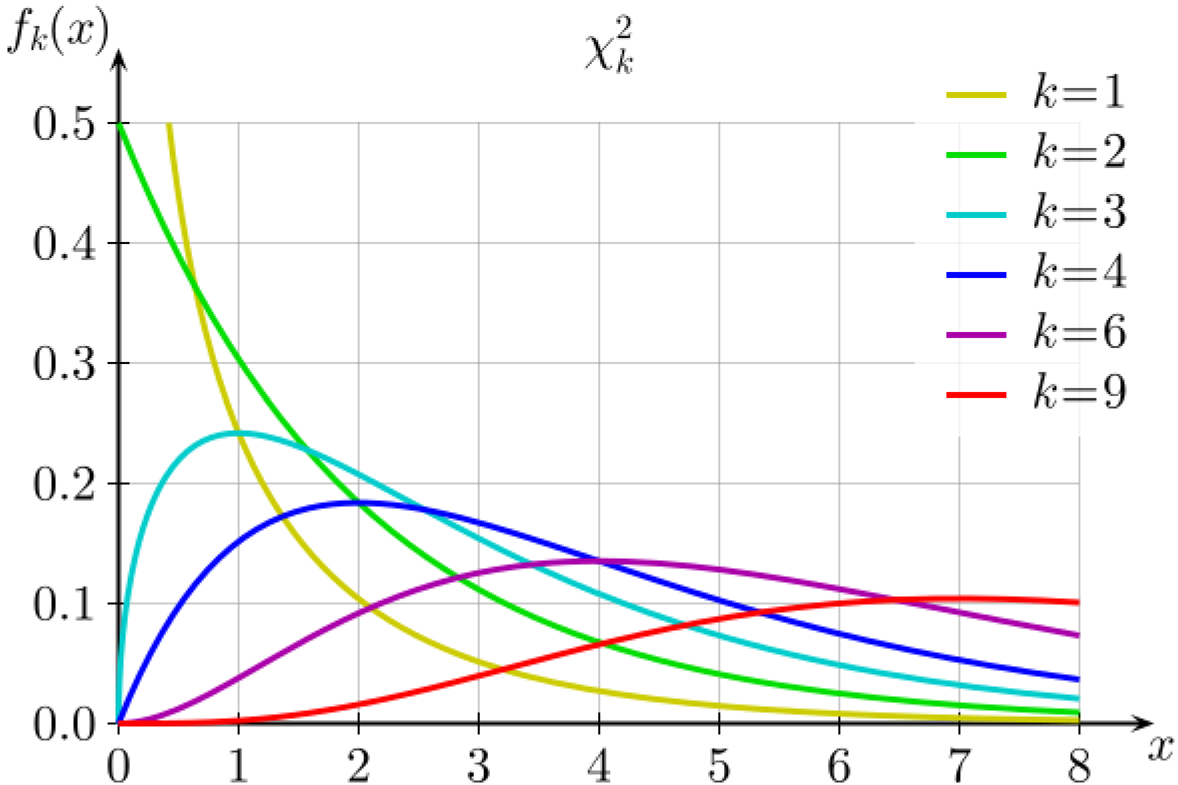
drop-in-deviance: $G =$ deviance for reduced model - deviance for full model = 13.92

> If H_0 is true, how unusual is $G = 13.92$?

χ^2 distribution

Under $H_0, G \sim \chi^2_{df_{\text{reduced}}-df_{\text{full}}}$

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 $\sim \chi^2_1$

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

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

Concept check

> Our p-value is 0.0002. What is the most appropriate conclusion? Go to https://pollev.com/ciaranevans637 to respond.

(A) We reject the null hypothesis, since $p < 0.05$.

(B) We fail to reject the null hypothesis, since $p < 0.05$.

(C) The data provide strong evidence of a relationship between GRE score and the probability of admission to graduate school.

(D) The data do not provide strong evidence of a relationship between GRE score and the probability of admission to graduate school.

Likelihood ratio test for nested models

Likelihood ratio test: strengths and weaknesses

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

Class activity

https://sta214-f22.github.io/class_activities/ca_lecture_8.html

Class activity

$Y_i =$ dengue status (0 = no, 1 = yes)

$$Y_i \sim Bernoulli(\pi_i) \qquad \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 WBC_i$$

> Researchers want to test whether there is any relationship between WBC and dengue status. What are H_0 and H_A?

Class activity

```
...
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.73743    0.08499   20.44   <2e-16 ***
## WBC         -0.36085    0.01243  -29.03   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6955.8  on 5719  degrees of freedom
## Residual deviance: 5529.8  on 5718  degrees of freedom
...
```

Wald test:

Class activity

```
...
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.73743    0.08499   20.44   <2e-16 ***
## WBC           -0.36085    0.01243  -29.03   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6955.8  on 5719  degrees of freedom
## Residual deviance: 5529.8  on 5718  degrees of freedom
...
```

Likelihood ratio test:

Class activity

$Y_i =$ dengue status (0 = no, 1 = yes)

$$Y_i \sim Bernoulli(\pi_i) \qquad \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 WBC_i$$

> Researchers want to test whether patients with lower WBC are more likely to have dengue. What are H_0 and H_A?

Class activity

```
...
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.73743    0.08499   20.44   <2e-16 ***
## WBC         -0.36085    0.01243  -29.03   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6955.8  on 5719  degrees of freedom
## Residual deviance: 5529.8  on 5718  degrees of freedom
...
```

Test: