Likelihood ratio tests

Last time

Data on the RMS *Titanic* disaster. We have data on 891 passengers on the ship, with the following variables:

- `Passenger`: A unique ID number for each passenger.
- `Survived`: An indicator for whether the passenger survived (1) or perished (0) during the disaster.
- `Pclass`: Indicator for the class of the ticket held by this passengers; 1 = 1st class, 2 = 2nd class, 3 = 3rd class.
- `Sex`: Binary Indicator for the biological sex of the passenger.
- `Age`: Age of the passenger in years; Age is fractional if the passenger was less than 1 year old.
- `Fare`: How much the ticket cost in US dollars.
- + others

Last time

Is there a relationship between passenger age and their probability of survival, after accounting for sex, passenger class, and the cost of their ticket?

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 Age_i \cdot Sex_i + \beta_4 \log(Fare_i + 1)$$

> What hypotheses should we test to investigate this research question?

Likelihood ratio tests

```
...
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.40695    0.44682  -3.149  0.00164 **
## Age             0.01107    0.01107   1.000  0.31730
## Sexmale        -1.27467    0.41654  -3.060  0.00221 **
## log(Fare + 1)   0.69449    0.11065   6.276 3.47e-10 ***
## Age:Sexmale    -0.03638    0.01378  -2.639  0.00831 **
##
##     Null deviance: 964.52  on 713  degrees of freedom
## Residual deviance: 697.21  on 709  degrees of freedom
...
```

> What information replaces R^2 and R^2_{adj} in the GLM output?

Deviance

Definition: The *deviance* of a fitted model with parameter estimates $\widehat{\beta}$ is given by

$$2\ell(\text{saturated model}) - 2\ell(\widehat{\beta})$$

Residual and null deviance

```
m1 <- glm(Survived ~ Age*Sex + log(Fare + 1),
          data = titanic, family = binomial)
summary(m1)
```

```
...
##     Null deviance: 964.52  on 713  degrees of freedom
## Residual deviance: 697.21  on 709  degrees of freedom
...
```

Comparing deviances

```
m1 <- glm(Survived ~ Age*Sex + log(Fare + 1),
          data = titanic, family = binomial)
summary(m1)
```

```
...
##     Null deviance: 964.52  on 713  degrees of freedom
## Residual deviance: 697.21  on 709  degrees of freedom
...
```

```
m2 <- glm(Survived ~ Sex + log(Fare + 1),
          data = titanic, family = binomial)
summary(m2)
```

```
...
##     Null deviance: 964.52  on 713  degrees of freedom
## Residual deviance: 708.04  on 711  degrees of freedom
...
```

Comparing deviances

Full model:

Hypotheses:

Reduced model:

Test statistic:

Comparing deviances

Full model:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Sex_i + \beta_2 \log(Fare_i + 1) + \beta_3 Age_i + \beta_4 Age_i \cdot Sex_i$$

Reduced model:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Sex_i + \beta_2 \log(Fare_i + 1)$$

$$G = 2\ell(\widehat{\beta}) - 2\ell(\widehat{\beta}^0)$$

Why is G always ≥ 0?

Comparing deviances

Full model:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Sex_i + \beta_2 \log(Fare_i + 1) + \beta_3 Age_i + \beta_4 Age_i \cdot Sex_i$$

Reduced model:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Sex_i + \beta_2 \log(Fare_i + 1)$$

$$G = 2\ell(\widehat{\beta}) - 2\ell(\widehat{\beta}^0) = 10.83$$

> If the reduced model is correct, how unusual is $G = 10.83$?

Likelihood ratio test