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?

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:

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

Last time: Deviance

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

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

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:

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

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

A different research question...

Suppose we include passenger class in the model instead of Fare:

$$
\begin{aligned}
\log\left(\frac{p_i}{1-p_i}\right) &= \beta_0 + \beta_1 Sex_i + \beta_2 FirstClass_i + \beta_3 SecondClass_i \\
&\quad + \beta_4 Age_i + \beta_5 Age_i \cdot Sex_i
\end{aligned}
$$

> How would I test the hypothesis that second class passengers have the same chance of survival as third class passengers (after accounting for Sex and Age)?

A different research question...

Suppose we include passenger class in the model instead of Fare:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Sex_i + \beta_2 FirstClass_i + \beta_3 SecondClass_i \\ + \beta_4 Age_i + \beta_5 Age_i \cdot Sex_i$$

> How would I test the hypothesis that second class passengers have the same chance of survival as *first* class passengers (after accounting for Sex and Age)?

Contrasts