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:

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

$H_0: \beta_3 = \beta_4 = 0$

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

Last time: Deviance

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

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

$\hat{p}_i = Y_i$

uses $\hat{p}_i = \dfrac{e^{\hat{\beta}^T X_i}}{1 + e^{\hat{\beta}^T X_i}}$

G = deviance of reduced model − deviance of full model

$= -2\ell(\hat\beta^0) - (-2\ell(\hat\beta)) = 2\ell(\hat\beta) - 2\ell(\hat\beta^0)$

Comparing deviances

estimates for full model

estimates for reduced model

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

Full model

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

$G = 708.04 - 697.21 = \boxed{10.83}$

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

Reduced model

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

Analogy in linear regression: SSE $\downarrow$ when we add parameters

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?

$\ell(\hat{\beta}) \geq \ell(\hat{\beta}^0)$

$\hat{\beta}$ = MLE $\Rightarrow$ $\hat{\beta}$ maximizes ℓ

$\hat{\beta}^0$ = MLE for reduced (restricted model) pchisq(10.83, df = 2, lower.tail = F)

(pretty small)

Comparing deviances

$H_0: \beta_3 = \beta_4 = 0$

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:

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

difference in # parameters between full & reduced models

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

If the reduced model is "just as good" as the full model)

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

Need a distribution for G!

If H_0 (reduced model) is true, $G \sim \chi^2_q$

q = # parameters test

Likelihood ratio test

Model : $\log\left(\frac{p_i}{1-p_i}\right) = \beta^T X_i$

1) Hypotheses : $H_0: \beta = \beta^0$ (reduced model) $H_A: \beta \neq \beta^0$ (full model)

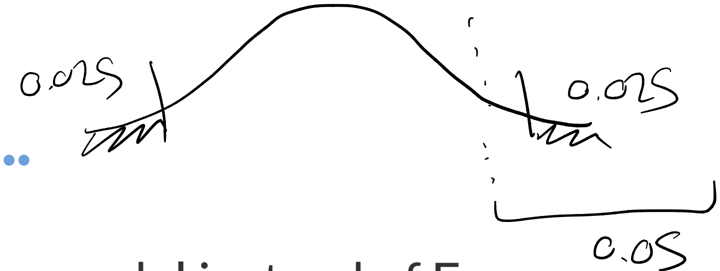
2) Fit full and reduced models, $\Rightarrow$ $\hat{\beta}$ (full model)
$\hat{\beta}^0$ (reduced model)

3) $G = 2\ell(\hat{\beta}) - 2\ell(\hat{\beta}^0) = 2\log\left(\frac{L(\hat{\beta})}{L(\hat{\beta}^0)}\right)$ (likelihood ratio)

4) If H_0 is true, $G \sim \chi^2_q$ q = # parameters tested

5) Compute a p-value

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 third class passengers (after accounting for Sex and Age)?

$H_0: \beta_3 = 0$

$H_A: \beta_3 \neq 0$

LRT: $G \sim \chi^2_1$

Wald: $W \sim \chi^2_1$

Question: Should we include class?

$H_0: \beta_2 = \beta_3 = 0$

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)?

H_0: $\beta_2 = \beta_3$ $\qquad \beta_2 - \beta_3 = 0$

H_A: $\beta_2 \neq \beta_3$ $\qquad \beta_2 - \beta_3 \neq 0$

option 1: Re-code the indicator variables (e.g., make First class the reference)

Contrasts

Intuition: $\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\,\mathrm{Cov}(X_1, X_2)$

$\mathrm{Var}(X_1 - X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) - 2\,\mathrm{Cov}(X_1, X_2)$

$H_0: \beta_2 - \beta_3 = 0$

$H_A: \beta_2 - \beta_3 \neq 0$

<u>Estimate $\beta_2 - \beta_3$</u>: $\hat\beta_2 - \hat\beta_3$ ← is this "close" to 0?

$\Rightarrow$ Need a distribution for $\hat\beta_2 - \hat\beta_3$!

$$\hat\beta_2 - \hat\beta_3 = \underbrace{\begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}}_{c^T} \begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \\ \hat\beta_2 \\ \hat\beta_3 \\ \hat\beta_4 \\ \hat\beta_5 \end{bmatrix} = c^T \hat\beta$$

Know that $\hat\beta \approx N(\beta, \mathcal{I}^{-1}(\beta))$

$$\Rightarrow \quad c^T \hat\beta \approx N(c^T\beta, c^T \mathcal{I}^{-1}(\beta) c)$$

$$c^T\hat{\beta} \overset{\sim}{\sim} N\left(c^T\beta,\ c^T \mathcal{I}^{-1}(\beta)\, c\right)$$

Under H_0: $c^T\beta = 0$ $\qquad$ H_A: $c^T\beta \neq 0$

Test statistic:

$$Z = \frac{c^T\hat{\beta} - 0}{\sqrt{c^T \mathcal{I}^{-1}(\hat{\beta})\, c}} \ \overset{\sim}{\sim} N(0,1)$$

equivalently:

$$W = \frac{\left(c^T\hat{\beta} - 0\right)^2}{c^T \mathcal{I}^{-1}(\hat{\beta})\, c} \ \overset{\sim}{\sim} \chi^2_1$$

(uses fact that if $Z \sim N(0,1)$ then $Z^2 \sim \chi^2_1$)

- extension of Wald test
- does <u>not</u> require refitting the model (for combos of > 2 coefficients, hard to re-code model anyway)