Lecture 10: Inference with logistic regression models

Recall: the Titanic data

Data on 891 passengers on the *Titanic*. Variables include:

- `Survived`
- `Pclass`
- `Sex`
- `Age`

Logistic regression model

$$\text{Survived}_i \sim \text{Bernoulli}(p_i)$$

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 \text{Class2}_i + \beta_3 \text{Class3}_i + \beta_4 \text{Male}_i + \beta_5 \text{Age}_i$$

Fitting the model in R

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.77701265	0.401123305	9.416089	4.682044e-21
as.factor(Pclass)2	-1.30979927	0.278065527	-4.710398	2.472337e-06
as.factor(Pclass)3	-2.58062532	0.281442020	-9.169296	4.761161e-20
Sexmale	-2.52278092	0.207390924	-12.164375	4.811152e-34
Age	-0.03698527	0.007655948	-4.830919	1.359041e-06

Suppose I want to know whether there is a relation between age and the probability of survival. What hypotheses would I test?

Wald tests for single coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.77701265	0.401123305	9.416089	4.682044e-21
as.factor(Pclass)2	-1.30979927	0.278065527	-4.710398	2.472337e-06
as.factor(Pclass)3	-2.58062532	0.281442020	-9.169296	4.761161e-20
Sexmale	-2.52278092	0.207390924	-12.164375	4.811152e-34
Age	-0.03698527	0.007655948	-4.830919	1.359041e-06

Another question

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.77701265	0.401123305	9.416089	4.682044e-21
as.factor(Pclass)2	-1.30979927	0.278065527	-4.710398	2.472337e-06
as.factor(Pclass)3	-2.58062532	0.281442020	-9.169296	4.761161e-20
Sexmale	-2.52278092	0.207390924	-12.164375	4.811152e-34
Age	-0.03698527	0.007655948	-4.830919	1.359041e-06

Suppose I want to know whether there is a relation between *passenger class* and the probability of survival. What hypotheses would I test?

Recall: nested tests for linear regression

Logistic regression model performance

```
Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         3.777013   0.401123   9.416  < 2e-16 ***
as.factor(Pclass)2 -1.309799   0.278066  -4.710 2.47e-06 ***
as.factor(Pclass)3 -2.580625   0.281442  -9.169  < 2e-16 ***
Sexmale            -2.522781   0.207391 -12.164  < 2e-16 ***
Age                -0.036985   0.007656  -4.831 1.36e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 964.52  on 713  degrees of freedom
Residual deviance: 647.28  on 709  degrees of freedom
```

Nested logistic regression models

```r
m1 <- glm(Survived ~ as.factor(Pclass) + Sex + Age,
          family = binomial, data = titanic)

m1$deviance
```

```
[1] 647.2831
```

```r
m2 <- glm(Survived ~ Sex + Age,
          family = binomial, data = titanic)
m2$deviance
```

```
[1] 749.9569
```

Preview: likelihood ratio test

Preview: likelihood ratio test

```r
m1 <- glm(Survived ~ as.factor(Pclass) + Sex + Age,
          family = binomial, data = titanic)


m2 <- glm(Survived ~ Sex + Age,
          family = binomial, data = titanic)

pchisq(m2$deviance - m1$deviance, df=2, lower.tail=F)
```

```
[1] 5.06597e-23
```