Logistic Regression Notes

Troy Edwards

2023-03-16

Response: Categorical & binary (success or fail)
Predictors: Quantitative or categorical
The odds of something happening, given probability p, are:

$$odds = \frac{p}{1-p}$$

Example:

$$P(making\ free\ throw) = 0.8$$
$$odds = \frac{0.8}{0.2} = 4$$

The transformation from p to $ln(odds)$ is called the logistic or logit transformation.

$$ln(\frac{p}{1-p}) = \beta_0 + \beta_1 x$$
$$\frac{p}{1-p} = e^{\beta_0+\beta_1 x}$$
$$p = (1-p)e^{\beta_0+\beta_1 x} = e^{\beta_0+\beta_1 x} - pe^{\beta_0+\beta_1 x}$$
$$p + pe^{\beta_0\beta_1 x} = e^{\beta_0+\beta_1 x}$$
$$p(1+e^{\beta_0+\beta_1 x}) = e^{\beta_0+\beta_1 x}$$
$$p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

```
donner <- read.csv("donner.csv")
regres1 <- glm(survived ~ age, data = donner, family = binomial("logit"))
summary(regres1)
```

```
##
## Call:
## glm(formula = survived ~ age, family = binomial("logit"), data = donner)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5401  -1.1594  -0.4651   1.0842   1.7283
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.81852    0.99937   1.820   0.0688 .
## age         -0.06647    0.03222  -2.063   0.0391 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 56.291  on 43  degrees of freedom
## AIC: 60.291
##
## Number of Fisher Scoring iterations: 4
```

$$\widehat{odds} = e^{1.81852 - 0.06647 age}$$
$$\hat{p} = \frac{e^{1.81852 - 0.06647 age}}{1 + e^{1.81852 - 0.06647 age}}$$

```
ggplot(donner, aes(age, survived)) +
  geom_point() +
  geom_smooth(method = "glm", method.args = list(family = "binomial"), se = F)
```

```
## `geom_smooth()` using formula 'y ~ x'
```