Lesson 27. Logistic Regression and Odds Ratios – Part 1

1 Last time...

- Suppose π is the probability of success
- The **odds** of success is a ratio of probabilities:

$$\text{odds}(\pi) = \frac{\pi}{1-\pi}$$

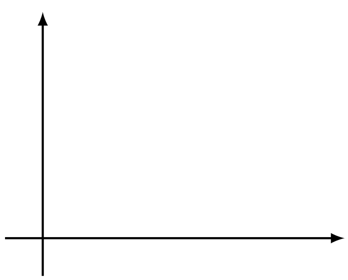
- The logistic regression model:
  - Logit form:

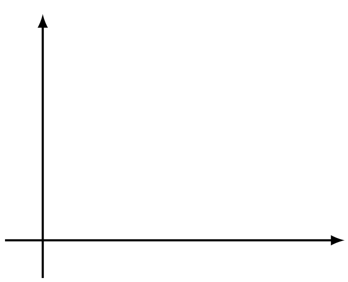
$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X$$

  - Probability form:

$$\pi = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$$

2 How parameter values affect the shape of the curve in probability form





- The "midpoint" on the y-axis, where $\pi = 0.5$, occurs at ____
- The slope of the curve at this "midpoint" is ____
  - If $\beta_1 < 0$, the curve has a ____ slope
  - If $\beta_1 > 0$, the curve has a ____ slope
  - A larger $|\beta_1|$ corresponds to a ____ slope
- Increasing the value of β_0 shifts the curve to the ____

3 The odds ratio

- The **odds ratio** is a ratio of odds:

- Interpretation: an odds ratio of 2 means the odds of success are 2 times as high under condition A versus condition B

Example 1. A study investigated whether a handheld device that sends a magnetic pulse into a person's head might be an effective treatment for migraine headaches. Reserarchers recruited 200 subjects who suffered from migraines and randomly assigned them to receive either the TMS (transcranial magnetic stimulation) treatment or a placebo. Whether or not the subject was pain free two hours after treatment, as well as which treatment was received, is recorded below.

	TMS	Placebo	Total
Pain-free	39	22	61
Not pain-free	61	78	139
Total	100	100	200

a. Using the raw data above, estimate the odds ratio of being pain-free with TMS versus with the placebo.
b. Interpret the odds ratio in the context of the problem.

4 The odds ratio in logistic regression with a binary predictor

- Consider a fitted logistic regression model in logit form:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 X$$

- How can we interpret $\hat{\beta}_1$ when X is binary?

- Exponentiating yields:

Example 2. Continuing with the TMS data in Example 1...

In Part 2 of this lesson, we fit a logistic regression model that uses treatment status to predict the probability of being pain-free. The data resides in a CSV file called `data/tms.csv`. The R code looks like this:

```
tms.data <- read.csv('data/tms.csv')
fit <- glm(PainFree ~ TMS, data = tms.data, family = binomial)
summary(fit)
```

The output looks like this:

```
Call:
glm(formula = PainFree ~ TMS, family = binomial, data = tms.data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9943  -0.9943  -0.7049   1.3723   1.7402

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.2657     0.2414  -5.243 1.58e-07 ***
TMS           0.8184     0.3167   2.584  0.00977 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 246.02  on 199  degrees of freedom
Residual deviance: 239.13  on 198  degrees of freedom
AIC: 243.13

Number of Fisher Scoring iterations: 4
```

a. State the fitted model in logit form.

b. Using the fitted slope from the logistic regression model, estimate the odds ratio of being pain free with TMS versus with the placebo.

- Note that the odds ratio we estimated directly from the data in Example 1 and the odds ratio we estimated from the logistic regression model in Example 2 are equal
- This is not a coinidence! These will always match for a binary predictor

5 The odds ratio in logistic regression with a quantitative predictor

- Again, consider a fitted logistic regression model in logit form:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 X$$

- How can we interpret $\hat{\beta}_1$ when X is quantitative?

	At x^* → At $x^* + 1$
logit	⟶
odds	⟶
probability	⟶

Example 3. Continuing with the `MedGPA` data from Lesson 26...

We looked at a binary response variable (*Acceptance* = 1 if accepted, 0 if not) and a quantitative predictor (*GPA*) for 55 medical school applicants from a college in the Midwest.

In Part 2 of this lesson, we fit a logistic regression model that predicts the probability of being accepted into medical school based on GPA. The R code looks like this:

```
library(Stat2Data)
data(MedGPA)

fit <- glm(Acceptance ~ GPA, data = MedGPA, family = binomial)
summary(fit)
```

The output looks like this:

```
Call:
glm(formula = Acceptance ~ GPA, family = binomial, data = MedGPA)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7805  -0.8522   0.4407   0.7819   2.0967

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -19.207      5.629  -3.412 0.000644 ***
GPA            5.454      1.579   3.454 0.000553 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 75.791  on 54  degrees of freedom
Residual deviance: 56.839  on 53  degrees of freedom
AIC: 60.839

Number of Fisher Scoring iterations: 4
```

a. State the fitted model.
b. Calculate the odds of acceptance with a GPA of 3.3.
c. Estimate the probability of acceptance with a GPA of 3.3.

d. Using the estimated slope, calculate the odds ratio of acceptance for a 4.0 GPA versus a 3.0 GPA.
e. Calculate the odds ratio comparing the odds of acceptance for a 3.4 GPA versus a 3.3 GPA.
f. Interpret the odds ratio from part e in the context of the problem.