Lecture 17: Simple Logistic Regression
STAT 632, Spring 2020

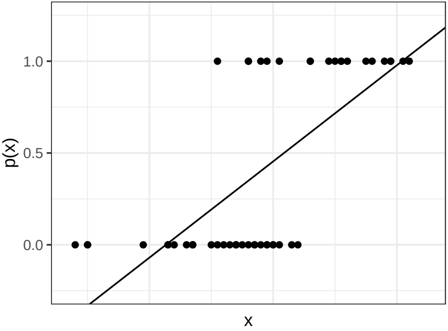
Simple Logistic Regression

- Simple logistic regression is a method to model a binary response variable, $Y \in \{0, 1\}$, using a single predictor variable x.
- Specifically, the method models $p(x) = Pr(Y = 1|x)$, the probability $Y = 1$ given predictor x.

Simple Logistic Regression

Why not use linear regression to represent these probabilities?

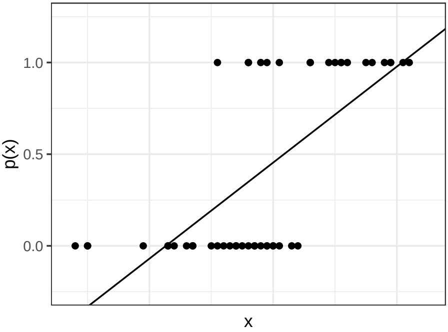
$$p(x) = Pr(Y = 1|x) = \beta_0 + \beta_1 x$$



Simple Logistic Regression

Why not use linear regression to represent these probabilities?

$$p(x) = Pr(Y = 1|x) = \beta_0 + \beta_1 x$$

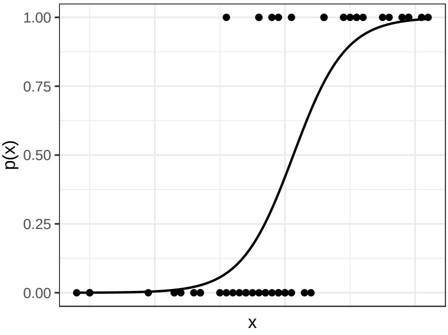


Answer: Fitting a linear regression model can result in estimating probabilities that are less than 0 or greater than 1.

Simple Logistic Regression

The **logistic function** is commonly used to model $p(x)$ since it always gives outputs between 0 and 1.

$$p(x) = Pr(Y = 1|x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}}$$



Simple Logistic Regression

Two ways to express the simple logistic regression model:

Probability form:

$$p(x) = Pr(Y = 1|x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}} = \frac{1}{1 + e^{-\beta_0-\beta_1 x}}$$

which can be interpreted as the probability $Y = 1$ for a given value x of the predictor.

Logit form:

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

The left-hand side is called the *logit* or *log-odds*. Logistic regression expressed in terms of the logit is linear in its parameters.

Simple Logistic Regression

Some algebraic manipulation can be used to show that the two representations are equivalent:

$$
\begin{aligned}
p &= \frac{1}{1 + e^{-\beta_0 - \beta_1 x}} \\
\frac{1}{p} &= 1 + e^{-\beta_0 - \beta_1 x} \\
\frac{1-p}{p} &= e^{-\beta_0 - \beta_1 x} \\
\frac{p}{1-p} &= e^{\beta_0 + \beta_1 x} \\
\log\left(\frac{p}{1-p}\right) &= \beta_0 + \beta_1 x
\end{aligned}
$$

Here we are letting $p = p(x)$ to simplify notation.

Inference

Hypothesis test for β_1:
$H_0 : \beta_1 = 0$
$H_A : \beta_1 \neq 0$

Test statistic:

$$z = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

This is sometimes referred to as the Wald z-statistic.

A $1 - \alpha$ confidence interval for β_1:

$$\hat{\beta}_1 \pm z_{\alpha/2} se(\hat{\beta}_1)$$

Example: 2016 US Presidential Election

- Data set called `Election16` from the `Stat2Data` library. The data contain results from the 2016 presidential election and demographic information from all 50 states.

- The binary response variable is `TrumpWin`, whether Trump won the state (1=yes, 0=no).

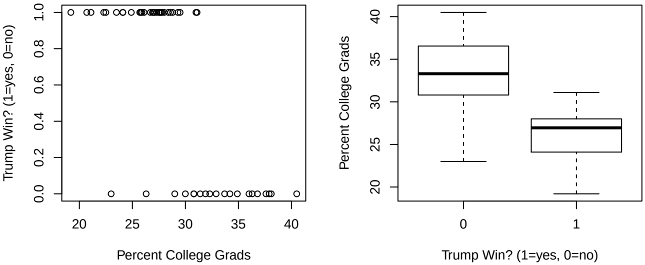
- The predictors are
  - `HS`: Percent of high school graduates in the state
  - `BA`: Percent of college graduates in the state
  - `Adv`: Percent with advanced degrees in the state
  - `Dem.Rep`: Percent Democratic - Percent Republican
  - `Income`: Per capita income in the state

Example

```
> library(Stat2Data)
> data("Election16")
> head(Election16, n=10)
         State Abr Income   HS   BA  Adv Dem.Rep TrumpWin
1      Alabama  AL  43623 84.3 23.5  8.7     -17        1
2       Alaska  AK  72515 92.1 28.0 10.1     -17        1
3      Arizona  AZ  50255 86.0 27.5 10.2      -1        1
4     Arkansas  AR  41371 84.8 21.1  7.5      -7        1
5   California  CA  61818 81.8 31.4 11.6      16        0
6     Colorado  CO  60629 90.7 38.1 14.0      -1        0
7  Connecticut  CT  70331 89.9 37.6 16.6      11        0
8     Delaware  DE  60509 88.4 30.0 12.2       6        0
9      Florida  FL  47507 86.9 27.3  9.8       1        1
10     Georgia  GA  49620 85.4 28.8 10.7      -4        1
```

Example

To demonstrate simple logistic regression, we will fit a model with `TrumpWin` as the response, and `BA`, percent of college graduates in the state, as the predictor.



Example

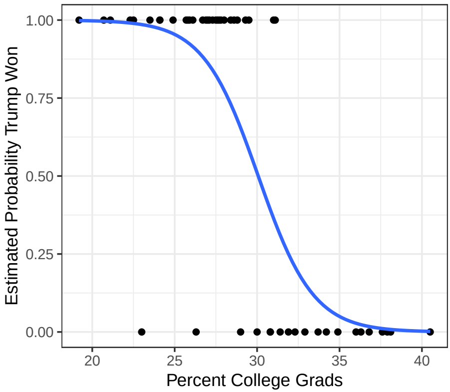
```
> glm1 <- glm(TrumpWin ~ BA, data=Election16, family=binomial)
> summary(glm1)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  17.9973     5.1098   3.522 0.000428 ***
BA           -0.5985     0.1735  -3.449 0.000562 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> confint(glm1)
                2.5 %     97.5 %
(Intercept)  9.809403 30.2884563
BA          -1.016162 -0.3211666
```

Example

```
ggplot(Election16, aes(BA, TrumpWin)) + geom_point()  +
  geom_smooth(method = "glm", method.args = list(family = "binomial"), se=F) +
  xlab("Percent College Grads") +
  ylab("Estimated Probability Trump Won") + theme_bw()
```



Example

The fitted logistic regression model in terms of the logit:

$$\log\left(\frac{\hat{p}(x)}{1-\hat{p}(x)}\right) = \hat{\beta}_0 + \hat{\beta}_1 x = 17.9973 - 0.5985x$$

In California, 31.4% of the population has a BA, so the estimate for the logit is

$$17.9973 - 0.5985(31.4) = -0.7956$$

Example

The fitted logistic regression model in probability from:

$$\hat{p}(x) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x}} = \frac{e^{17.9973 - 0.5985x}}{1 + e^{17.9973 - 0.5985x}}$$

In California, 31.4% of the population has a BA, so the estimate for the probability that Trump won is

$$\hat{p}(31.4) = \frac{e^{17.9973 - 0.5985(31.4)}}{1 + e^{17.9973 - 0.5985(31.4)}} = \frac{e^{-0.7956}}{1 + e^{-0.7956}} = 0.31097$$

Example

In R, the estimate for the logit can be obtained with the command

```
> new_x <- data.frame(BA = 31.4)
> predict(glm1, newdata = new_x)
         1
-0.7970077
```

The estimate for the probability can be obtained with the command

```
> new_x <- data.frame(BA = 31.4)
> predict(glm1, newdata = new_x, type="response")
       1
0.310666
```

Any difference from the manual calculations are due to rounding.

Interpreting the Coefficients

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

In terms of the *logit* we have the following interpretation:

An one unit increase in x is associated with a change in the log-odds, or logit, by β_1.

Going back to the example, a one unit increase in `BA` is associated with a $\hat{\beta}_1 = -0.5985$ change in the log-odds.

Interpreting the Coefficients

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

In terms of the *odds* we have the following interpretation:

An increase in x by 1 is associated with a *multiplicative* change in the odds by e^{β_1}. In other words, a unit increase in x multiplies the odds by e^{β_1}.

Going back to the example, a one unit increase in `BA` is associated with a multiplicative change of $e^{\hat{\beta}_1} = e^{-0.5985} = 0.55$ in the odds that Trump wins (for example, changing the odds from 4 to $0.55(4) = 2.2$).

We can also use this interpretation for different increments. For instance, an increase in `BA` by 0.1 is associated with a multiplicative change of $e^{0.1(\hat{\beta}_1)} = e^{0.1(-0.5985)} = 0.9419$ in the odds that Trump wins.

Interpreting Coefficients

The sign of β_1 also has meaningful interpretation:

- If $\beta_1 > 0$, then increasing x will be associated with increasing the probability $p(x)$.
- If $\beta_1 < 0$, then increasing x will be associated with decreasing the probability $p(x)$.

Estimation

The parameters β_0 and β_1 for logistic regression can be estimated using maximum likelihood.

Let $\{(x_i, y_i)\}$ be a sample of n independent observations with $y_i \in \{0, 1\}$ a binary response. Assume that y_i follows a Bernoulli distribution with probability p_i. We can write this compactly as:

$y_i \sim \text{Bern}(p_i)$
$Pr(y_i = 1) = p_i$
$Pr(y_i = 0) = 1 - p_i$
where p_i is given by the logistic function:

$$p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

Estimation

The likelihood function gives the probability of the observed zeros and ones in the data, expressed as function of the parameters.

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i}$$

The log-likelihood is given by

$$l(\beta_0, \beta_1) = \log(L(\beta_0, \beta_1)) = \sum_{i=1}^{n}[y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

The maximum likelihood estimates (MLEs), denoted $\hat{\beta}_0$ and $\hat{\beta}_1$, are the values of the parameters that maximize the log-likelihood function.

Estimation

To help understand the expression for the likelihood on the previous slide note that:

$$p_i^{y_i}(1-p_i)^{1-y_i} = \begin{cases} p_i, & \text{if } y_i = 1 \\ 1-p_i, & \text{if } y_i = 0 \end{cases}$$

Estimation

The MLEs are computed by evaluating the partial derivatives of $l(\beta_0, \beta_1)$ with respect to each parameter and setting the equations equal to 0:

$$\frac{\partial l(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^{n}(y_i - p_i) = 0 \qquad \frac{\partial l(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^{n} x_i(y_i - p_i) = 0$$

- Unlike multiple linear regression, there is no closed form for $\hat{\beta}_0$ and $\hat{\beta}_1$ that solve these equations.
- Instead, iterative methods are used to solve these equations and perform the optimization.
- Popular approaches are gradient descent and the Newton-Raphson method.