Regular Linear Regression

\[ y = X \beta, \quad y \in [-\infty, \infty] \]

What if \( y \) is categorical?

\[ y \in \{0, 1\} \]

For independent decision variables, no change in method.

For dependent response variables, requires very different framework.

\[ \Rightarrow \text{logistic regression} \]

\[ \frac{x}{\text{linear model}} \rightarrow \leftarrow \text{t} \rightarrow \text{link function} \rightarrow (0, 1) \]

Interpreted as a probability.

Applications
- Pass/fail test
- Tumor/normal tissue
- Disease/not
What is our link function?

Logistic function (logistic regression) (one of many link functions)

\[ L(t) = \frac{1}{1 + e^{-t}} \]

\[ L(0) = \frac{1}{1 + e^0} = \frac{1}{1+1} = \frac{1}{2} \]

\[ L(-\infty) = \frac{1}{1 + e^{\infty}} \rightarrow 0 \]

\[ L(\infty) = \frac{1}{1 + e^{-\infty}} \rightarrow \frac{1}{1+0} = 1 \]

\[ t = \text{(linear model)} \]

\[ \Rightarrow \ P(t) = \frac{1}{2} \]

\[ (\text{linear model}) > 0 \Rightarrow y = 1 \]

\[ (\text{linear model}) < 0 \Rightarrow y = 0 \]
Interpreting coefficients of Logistic Regression
output is Probability
Input $\rightarrow$ link function $\rightarrow$ $P$
$\Rightarrow$ coefficients tell us about $\Delta t$, not $\Delta P$!

Let's talk about odds:

$P$: long run fraction of successes
$P = 0.15$, ~15% of many trials will succeed.

odds: $\frac{P}{1-P} \leftarrow P$ of success $P + (1-P) = 1$

$1-P \leftarrow P$ of failure

Team has $P = 0.2$ of winning
they have $\frac{0.2}{1-0.2} = \frac{0.2}{0.8} = \frac{1}{4}$ odds
What are the odds for a logistic function

\[ P = \frac{1}{1 + e^{-t}} \]

\[ \text{odds} = \frac{P}{1 - P} = \frac{1 + e^{-t}}{1 - (1 + e^{-t})} \]

\[ = \frac{1}{e^{-t}} = e^t \]

Linear model \( \beta_0 + \beta_1 x \)

odds of response = 1 are \( e^{\beta_0 + \beta_1 x} \)

For Huntington’s Example:

\[ \text{odds}(\text{CAGs}) \]

\[ = e^{-10 + 0.4\text{[CAGs]}} \]

\[ = 0.14 : 1 \]

\[ \text{odds}(35 \text{ CAGs}) \]

\[ = 55 : 1 \]
Odds Ratio

\[
\frac{\text{odds}(x+1)}{\text{odds}(x)}
\]

fold increase in odds with a unit increase in the input.

For Huntington's, fold increase of getting disease given 1 more CAG.

\[
\text{OR} = \frac{e^{\beta_0 + \beta_1(x+1)}}{e^{\beta_0 + \beta_1 x}} = \frac{e^{\beta_0 + \beta_1 x}}{e^{\beta_0 + \beta_1 x}} = e^{\beta_1}
\]

For Huntington's:

\[
\text{OR} = e^{\beta_1} = e^{0.4} = 1.5
\]

For Huntington's:

\[
t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots
\]

\[
\text{OR}_{x_1} = e^{\beta_1} \quad \text{unit } \Delta \text{ in } x_1
\]

\[
\text{OR}_{x_2} = e^{\beta_2} \quad \text{unit } \Delta \text{ in } x_2
\]