# 10. Logistic regression

## Logistic regression

- Logistic regression can be used when the dependent variable has two outcomes: yes/no, 0/1.
- Predict/describe $E(Y_i|x_i)$.
- Why can't we use linear regression?
- Testing with logistic regression

## Example: Chilean plebiscite data

- Some history:
  - 1973: Coup $\Rightarrow$ military government of Pinochet
  - 1988: Plebiscite to decide the future of the government:
    Yes-vote = keep military government for 8 more years,
    No-vote = change to civilian government.
- Six months before plebiscite, national survey of 2700 randomly selected Chilean voters:
  - 868 planned to vote yes
  - 889 planned to vote no
  - 558 were undecide
  - 187 planned to abstain
  - 168 did not answer
- We only look at yes/no votes

## Why is the linear model not good for these data?

- Problems:
  - The model is only reasonable for a limited range. Outside this range we get fitted values that are smaller than zero or larger than one.
  - Nonparametric regression shows S-shaped fit, not a linear fit.
  - $Y_i$ can only take values 0 and 1. Errors are not normally distributed. However, for large sample sizes, the central limit theorem will save us.
  - The variance of the statistical errors is not constant.
- Why don't we have similar problems with 0-1 independent variables?

## Possible solutions

- Cut off the graph at zero and one.
    - Sometimes OK, if relationship is approximately linear in a certain range
- Use logistic regression:
    - logit$(u) = \log(u/(1-u))$.
    - If $u \in (0,1)$, then logit$(u) \in (-\infty, \infty)$
    - In principle, one could use logit transformation on the $y$-values, but one has to perturb them a little bit (how much?) since logit$(0)$ and logit$(1)$ are not defined
    - We perform the logit transformation on $E(Y_i|x_i)$:

$$\begin{aligned} \mathrm{logit} E(Y_i|x_i) &= \alpha + \beta x_i \\ \mathrm{logit} P_\theta(Y_i = 1|x_i) &= \alpha + \beta x_i \end{aligned}$$

## Interpretation in terms of hidden variables

See board

## Computation of the estimator

- Write logit$P_\theta(Y_i = 1|x_i) = \mathbf{x}_i^T \theta$, where $\theta = (\alpha, \beta)^T$
- Density of one observation:

$$\begin{aligned} P_\theta[Y_i = y_i|\mathbf{x}_i] &= \left( \frac{P_\theta[Y_i = 1|\mathbf{x}_i]}{P_\theta[Y_i = 0|\mathbf{x}_i]} \right)^{y_i} P_\theta[Y_i = 0|\mathbf{x}_i] \\ &= \exp[y_i \mathbf{x}_i^T \theta - \log(1 + \exp(\mathbf{x_i}^T \theta))] \end{aligned}$$

- Log likelihood:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^n \log P_\theta(Y_i = y_i|x_i) \\ &= \sum_{i=1}^n \left[ y_i \mathbf{x}_i^T \theta - \log(1 + \exp(\mathbf{x}_i^T \theta)) \right] \end{aligned}$$

## Computation of the estimator (2)

- Maximizer is given by solution of:

$$\sum_{i=1}^{n}(y_i - P_{\hat{\theta}}[Y_i = 1|\mathbf{x}_i])\mathbf{x}_i = \mathbf{0}$$

- Solve with iterative methods

## Interpretation: $\pi_i$, odds and log odds

- Let $\pi_i = P(Y = 1|x_i)$ be the conditional probability that $Y = 1$ given that $X = x_i$.
- Note that $E(Y|x_i) = \pi_i$ (derivation on board).
- $\pi_i/(1-\pi_i)$ are the *odds* that $Y = 1$ given $X = x_i$.
- $\log(\pi_i/(1-\pi_i))$ are the *log odds*.
- See table for log odds.

## Interpretation

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta x_i.$$

- Logistic regression is an additive model for the log odds. This gives one interpretation for $\beta$: If $X$ is increased by one, then the *log odds* are *increased* by $\beta$.
- Logistic regression is a multiplicative model for the odds:

  $$\frac{\pi_i}{1-\pi_i} = \exp(\alpha + \beta X_i) = \exp(\alpha)[\exp(\beta)]^{X_i}$$

  This gives another interpretation for $\beta$: If $X$ is increased by one, then the *odds* are *multiplied* by $\exp(\beta)$.
- Note that:

  $$\pi_i = \frac{1}{1 + \exp[-(\alpha + \beta X_i)]}$$

## Interpretation (2)

- $$\pi_i = \frac{1}{1 + \exp[-(\alpha + \beta X_i)]}$$
- Differentiating with respect to $X_i$ (see derivation on board) gives that the slope at $X_i$ is $\pi_i(1-\pi_i)\beta$.
- Hence, the derivative of the fitted graph is $\pi_i(1-\pi_i)\beta$. This gives a third interpretation for $\beta$. If $X = x_i$, and $X$ is increased by $\epsilon$ (small), then $\pi_i$ will increase by $\epsilon\pi_i(1-\pi_i)\beta$.
- See table of slopes. Note that the slopes are quite constant between $\pi = 0.2$ and $\pi = 0.8$. In this range the S-curve is close to linear.
- We don't interpret $\alpha$.
- How does all this work for the Chile data?

## Multiple logistic regression

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

$$\begin{aligned}\frac{\pi_i}{1-\pi_i} &= \exp(\alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}) \\ &= \exp(\alpha)\exp(\beta_1 X_{i1})\ldots\exp(\beta_k X_{ik}) \\ &= \exp(\alpha)[\exp(\beta_1)]^{X_{i1}}\ldots[\exp(\beta_k)]^{X_{ik}}\end{aligned}$$

$$\pi_i = \frac{1}{1 + \exp[-(\alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})]}$$

## Types of independent variables

- The $X$'s can be as general as in linear regression:
  - quantitative variables
  - transformations of quantitative variables
  - dummy regressors for qualitative variables
  - interaction regressors

## Testing with logistic regression

- Wald test (analogous to t-test)
- Likelihood ratio test (analogous to F-test)
  - Full model $m_1$
  - Null model $m_0$ (special case of full model)
  - Compute likelihood for both models: $L_1$ and $L_0$. $L_1 \geq L_0$. Why?