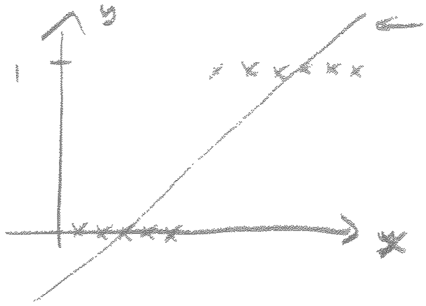
Logistic Regression

Categorical outcome ; $y \in \{0, 1\}$

Smoker (for 1)

Non-smoker (for 0)

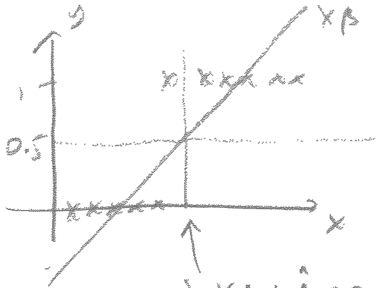
Can we still run regression?



← regression line $E(y) = X\beta$

Now, taking $y \in \{0,1\} \Rightarrow E(y) = P(y=1)$

So, regression line now loosely means $P(y=1 | x)$ linear in x.



$\{x : x\hat{\beta} = 0.5\}$ = "The decision boundary"

We predict $y=1$ if $x >$ this boundary, and $y=0$ if $x <$ this boundary

This easily generalizes to multiple x-variables where the decision boundary is the vector $\{\underset{\sim}{x} : \underset{\sim}{x}'\hat{\beta} = 0.5\}$

Is there a problem with this approach?

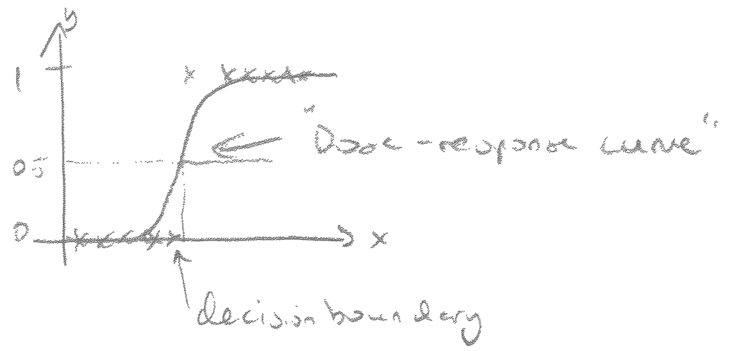
Well, the assumption $p(y=1|x) = x\beta$ means we can have $p(y=1|x) > 1$ or < 0! So the regression line cannot really represent a probability

<u>Fix</u>: logistic regression

Assume a <u>transformation</u> of $p(y=1|x)$ is linear in x.

logit: $\log \dfrac{p(y=1|x)}{1-p(y=1|x)} = \log \dfrac{p(y=1|x)}{p(y=0|x)} \equiv$ "log-odds" $= x\beta$

$$\Rightarrow \quad p(y=1|x) = \frac{e^{x\beta}}{1+e^{x\beta}}$$ which is in $[0,1]$ always.



Working with Logistic regression

Assumption : $y_i \sim Bin(m, \pi_i)$

$$\text{logit}(\pi_i) = \log \frac{\pi_i}{1-\pi_i} = \underset{\sim}{x}_i' \beta$$

If $y \in \{0,1\}$, $m=1$, if y a proportion or % m can be >1.

- Assume y_i's are independent
- Note, $V(y_i) = m\pi_i(1-\pi_i)$ so variance is non-constant
- Assume - no outliers
  - logit( ) is the correct transformation
    s.t. $p(y=1|x) = \text{logit}^{-1}(x\beta)$
    (may need to transform x's for this to hold
    - or perhaps consider other transformations)

<u>Fitting</u> Problems — non-constant error-variance
— $p(y=1|x)$ nonlinear in β

$\Rightarrow$ Solution Iterative Weighted Least Squares (IWLS)

Solve a sequence of linear approximations to the problem (nonlinear regression)

Use weights to account for non-constant variance

Weighted Least Squares WLS

$V(y_i) = \sigma_i^2$ (non-constant)

→ general $V(\underset{\sim}{y}) = \sigma^2 V$

↑ a matrix with variance differences

- if V is diagonal, just non-constant error variance
- if V is non-diagonal, allow for correlated errors.

① What happens if we use LS?

$$\hat{\beta} = (X'X)^{-1}X'y \rightarrow V(\hat{\beta}) = \sigma^2 (X'X)^{-1}(X'VX)(X'X)^{-1}$$

② WLS

$$\min_{\beta} \sum_{i=1}^{n} w_i (y_i - x_i'\beta)^2 = \sum_{i=1}^{n} (w_i^{1/2} y_i - w_i^{1/2} \underset{\sim}{x}_i'\beta)^2$$

$$= \| W^{1/2} y - W^{1/2} X\beta \|^2 = \| \tilde{y} - \tilde{X}\beta \|^2 \text{ where transformed}$$

variables $\tilde{y} = W^{1/2} y$

$\tilde{X} = W^{1/2} X$

$$\Rightarrow \hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}\tilde{y} = (X'WX)^{-1}X'Wy$$

$$V(\hat{\beta}) = (X'WX)^{-1}(X'WVWX)(X'WX)^{-1}\sigma^2$$

If we use $W = V^{-1}$ $\Rightarrow V(\hat{\beta}) = \sigma^2(X'V^{-1}X)^{-1}$

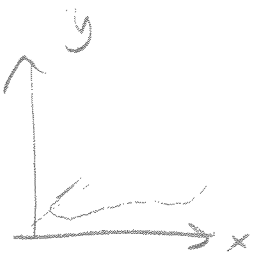
which one can show is optimal!

So, we should use weights $w_i \sim \frac{1}{V(y_i)}$ optimally. 5

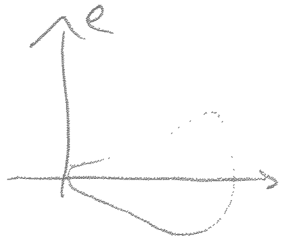
But we don't know $V(y_i)$ usually $\rightarrow$ estimate it

Iterative procedure

① LS



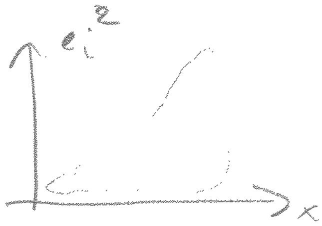
$\Rightarrow$ residuals



② Now $E(e_i) = 0$ since unbiased fit using LS

$$E(e_i^2) = V(e_i) = \sigma_i^2$$

Square the residuals and fit a regression model to e_i^2



$\Rightarrow$ fitted values $= \hat{\sigma}_i^2$ $\Rightarrow$ weights in WLS problem ③

You can regress e_i^2 on x's or $\hat{y}$'s from ①.

Logistic regression - Inference

Linear Model

↓

$\varepsilon_i \sim N(0, \sigma^2)$ assumption

↓

RSS, t-test, F-test

Logit model, $y_i \sim Bin(n, \pi_i)$

↓

Binomial model

↓

?

In logistic regression, RSS is replaced by the

<u>residual deviance</u> $= -2 \times \text{log-likelihood}(y, \hat{\beta})$

In linear models with normal errors, RSS $\sim \chi^2$ distributed.

That is approximately true for residual deviance.

(when n is large)

F-test is replaced with χ^2-test

Compare the residual deviance $D(M)$ (model M)

to $\chi^2_{\underbrace{n-p}_{\text{degrees-of-freedom}}}$

If $D(M) > \chi^2_{n-p}(1-\alpha)$ - reject model M

as a good fit.

Comparing models

F-test is now replaced w. χ^2-test

complex Model M , simple Model M_0

p parameters $\geq$ p_0 parameter

$D(M) \leq D(M_0)$ residual deviance

Test : $D(M_0) - D(M)$ compared with $\chi^2_{p-p_0}$

If $D(M_0) - D(M) > \chi^2_{p-p_0}(1-\alpha)$, reject the simple model M_0

You can also use AIC to select models

Coefficient test

t-test replaced with z-test - a large-n approximation, so be careful

Diagnostics

Residual plots can look weird - make sure residuals on average ≈ 0 - e.g. examine by plotting a local average in residual plot (see Demo)

CART & categorical outcomes

CART works just as in regression case.

MSE criterion now replaced with error rate

$$\frac{\#\ y\text{-obs} \neq \text{majority category}}{\text{total}\ \#\ \text{observations}}$$

You grow the CART tree to minimize the "impurity" in each region