Harvard-MIT Division of Health Sciences and Technology
HST.951J: Medical Decision Support, Fall 2005
Instructors: Professor Lucila Ohno-Machado and Professor Staal Vinterbo

**6.873/HST.951 Medical Decision Support**
**Fall 2005**

# *Decision Analysis*

(part 2 of 2)

## *Review Linear Regression*

Lucila Ohno-Machado

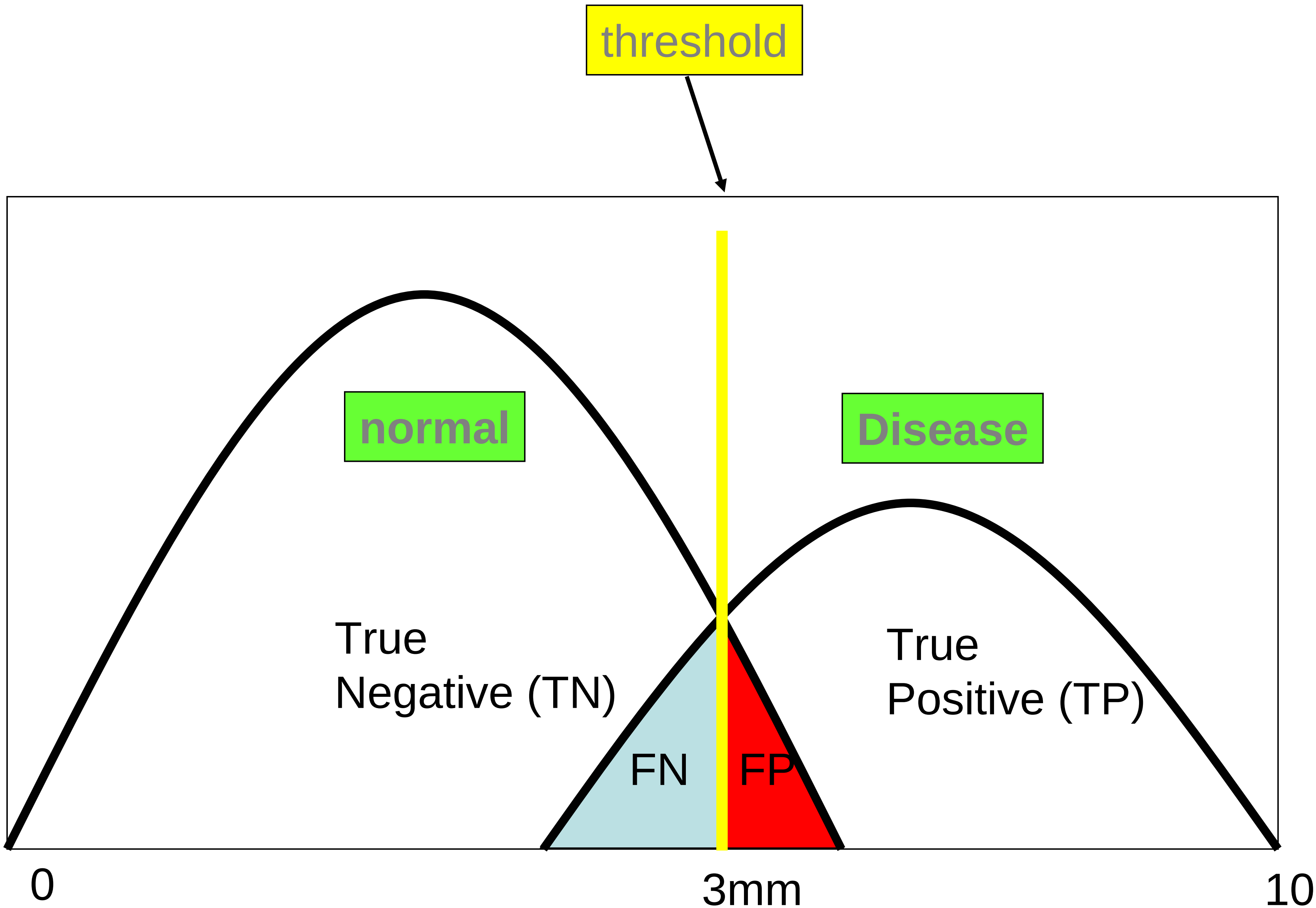
# Outline

- Homework clarification
- Sensitivity, specificity, prevalence
- Cost-effectiveness analysis
- Discounting cost and utilities
- Review of Linear Regression

# 2 x 2 table
# (contingency table)

| | PPD+ | PPD- | |
|---|---|---|---|
| TB | 8 | 2 | 10 |
| no TB | 3 | 87 | 90 |
| | 11 | 89 | 100 |

**Prevalence of TB = 10/100**
**Sensitivity of test = 8/11**
**Specificity of test = 87/89**


threshold
normal
Disease
True
Negative (TN)
True
Positive (TP)
FN
FP
0
3mm
10

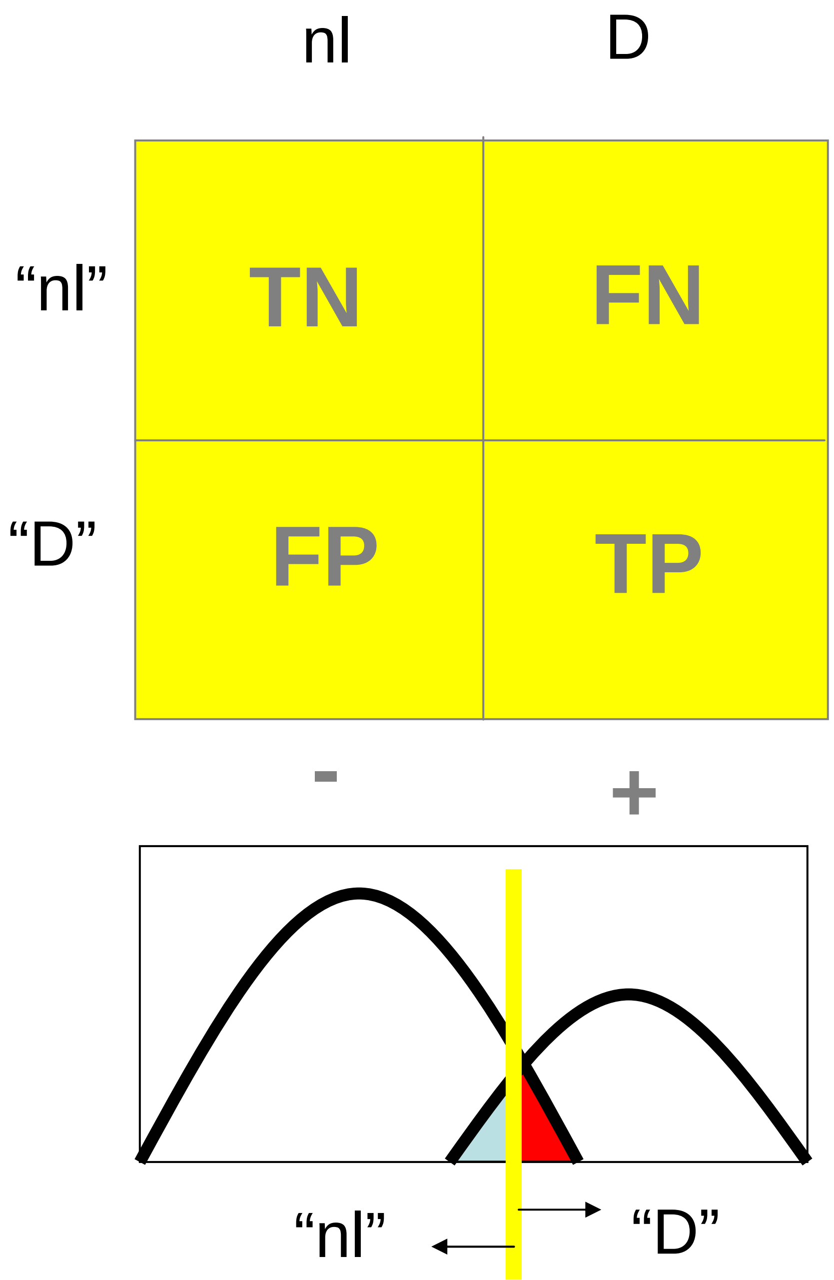
Sens = TP/TP+FN

Spec = TN/TN+FP

PPV = TP/TP+FP

NPV = TN/TN+FN

Accuracy = TN +TP



Sens = TP/TP+FN

40/50 = .8

Spec = TN/TN+FP

45/50 = .9

PPV = TP/TP+FP

40/45 = .89

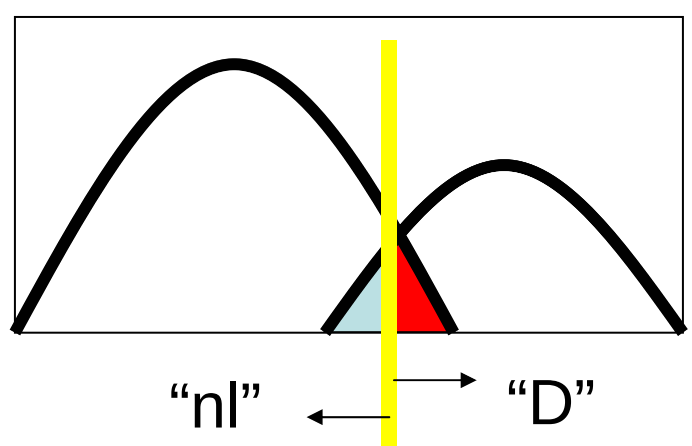
NPV = TN/TN+FN

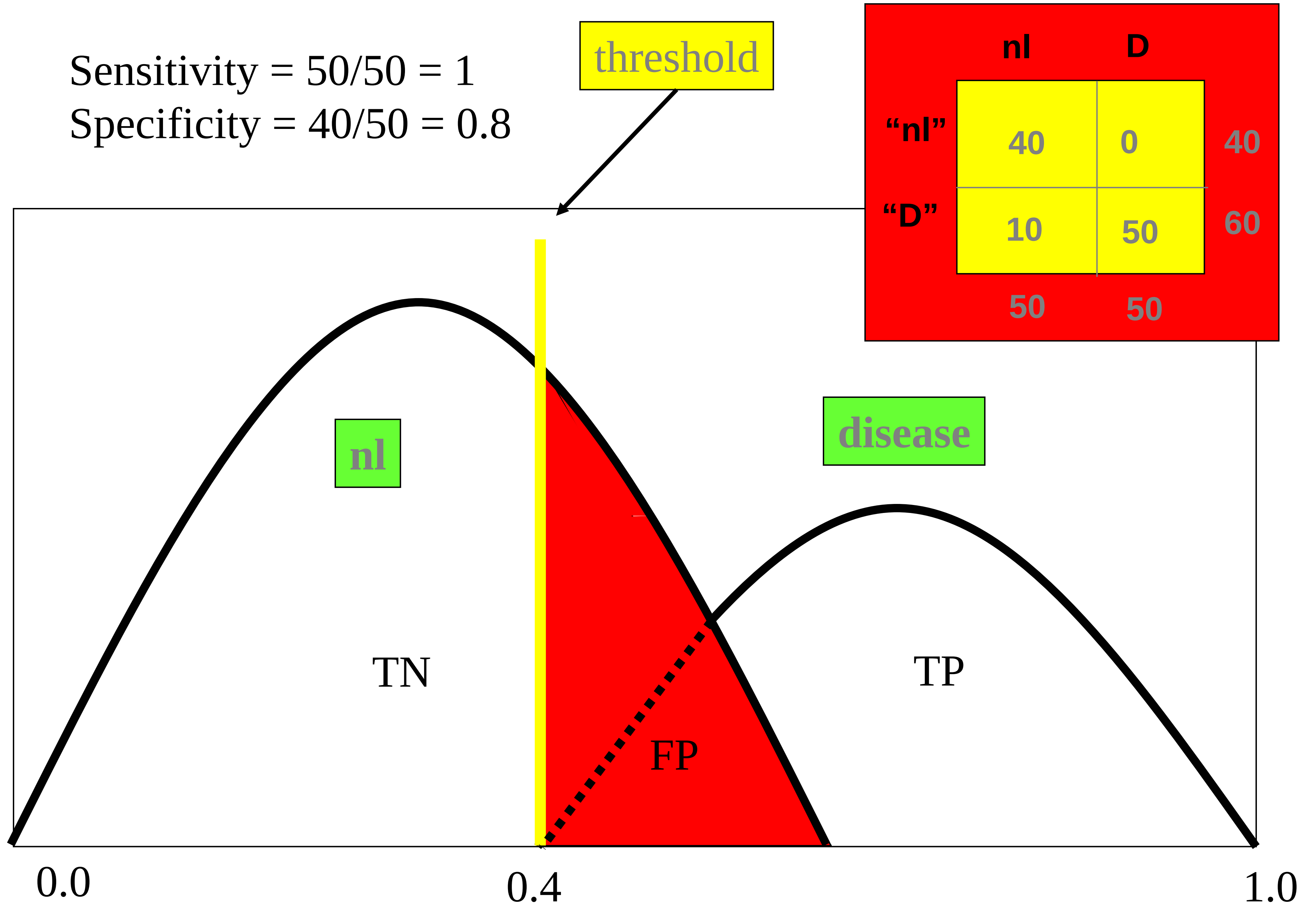
45/55 = .81

Accuracy = TN +TP

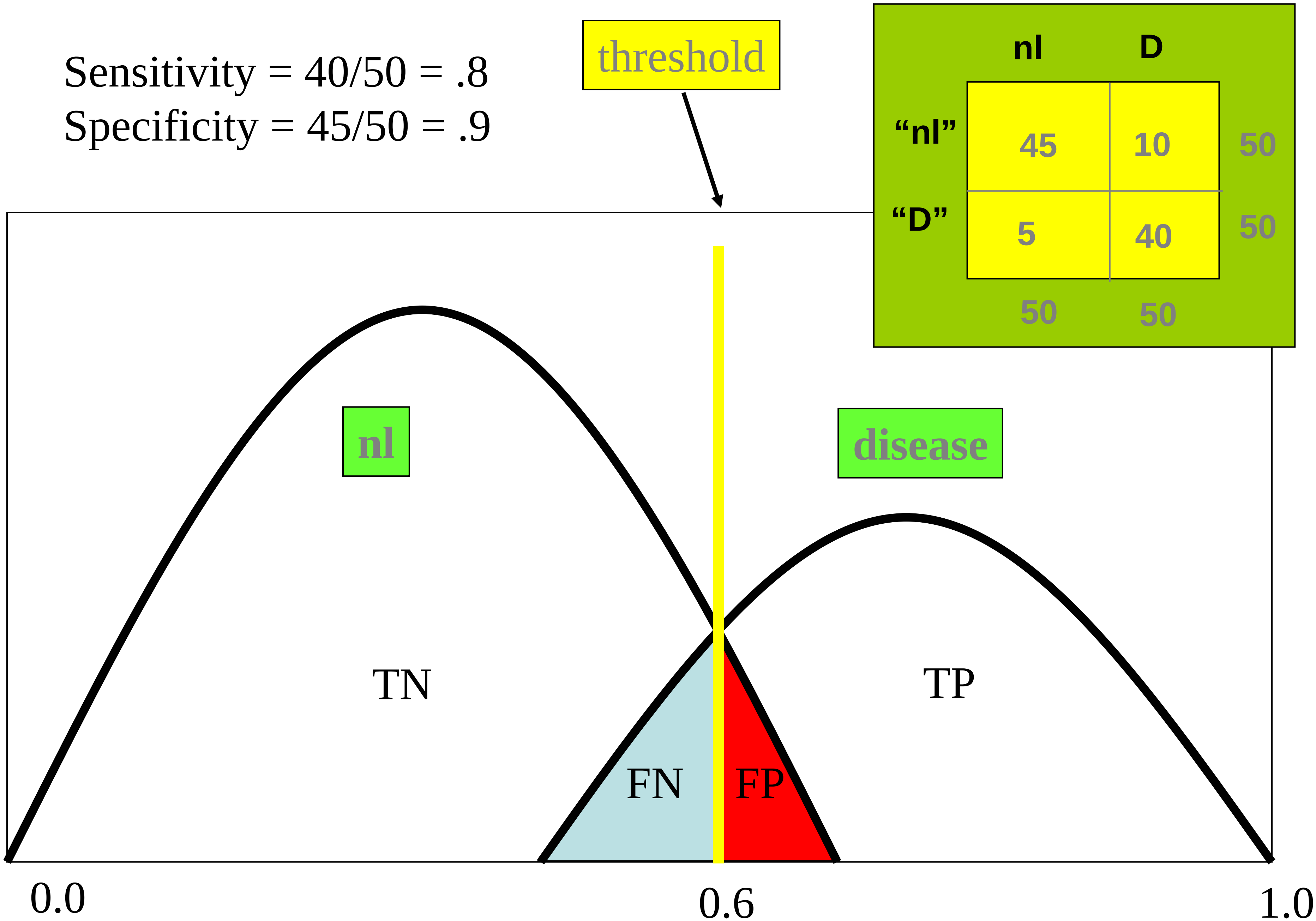
85/100 = .85

| | nl | D |
|---|---|---|
| “nl” | 45 | 10 |
| “D” | 5 | 40 |

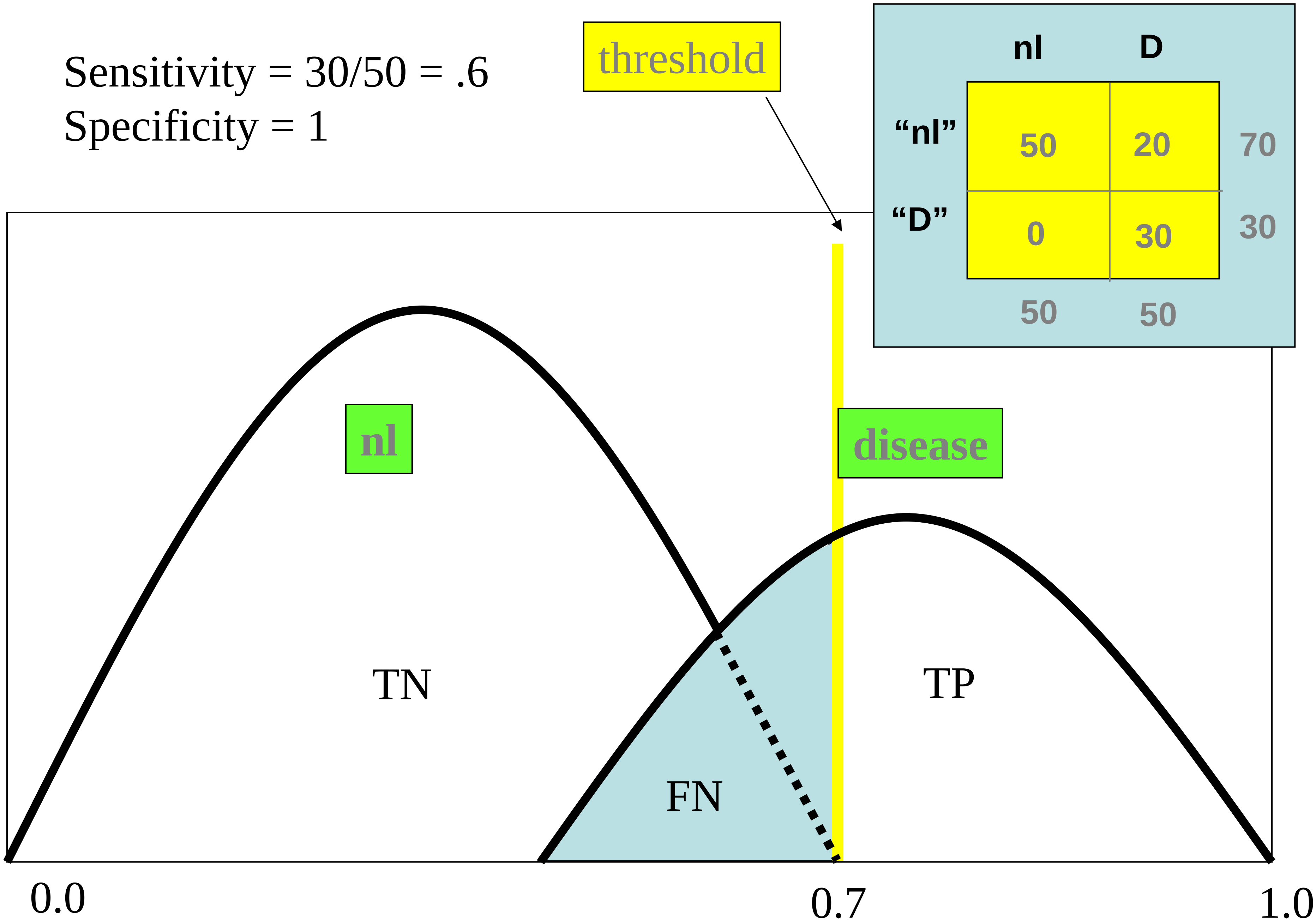



Sensitivity = 50/50 = 1
Specificity = 40/50 = 0.8
threshold
nl
D
"nl"
"D"
40
0
40
10
50
60
50
50
nl
disease
TN
FP
TP
0.0
0.4
1.0

Sensitivity = 40/50 = .8
Specificity = 45/50 = .9



Sensitivity = 30/50 = .6
Specificity = 1



# Cost-effectiveness analysis

- Comparison of costs with health effects
  - Cost per Down case syndrome averted
  - Cost per year of life saved
- Perspectives (society, insurer, patient)
- Comparators
  - Comparison with doing nothing
  - Comparison with “standard of care”

# Discounting costs

- It is better to spend \$10 next year than today (its value will be only \$9.52, assuming 5% rate)
- Even better to spend it 2 years from now (\$9.07)
- For cost-effectiveness analysis spanning multiple years, recommended rate is usually 5%

- $C = C_0 + C_1/(1-r)^1 + C_2/(1-r)^2 + \ldots$
- $C_0$ are costs at time 0

# Discounting utilities

- Value for full mobility is 10 today (is it only 9.52 next year?)
- Should the discount rate be the same as for costs?
- If smaller, then it would always be better to wait one more year…

# Levels of Evidence in Evidence-Based Medicine

US Task Force

- Level 1: at least 1 randomized controlled trial (RCT)
- Level 2-I: controlled trials (CT)
- Level 2-II: cohort or case-control study
- Level 2-III: multiple time series with or without the intervention
- Level 3: expert opinions

# Examples

- Cost per year of life saved, Life years/US$1M
- By pass surgery middle-aged man
  - $11k/year, 93
- CCU for low risk patients
  - $435k/year, 2

# Importance of good stratification

- Bypass surgery
  - Left main disease 93
  - One vessel disease 12

- CCU for chest pain
  - 5% risk of MI 2
  - 20% risk of MI 10

# Intro to Modeling

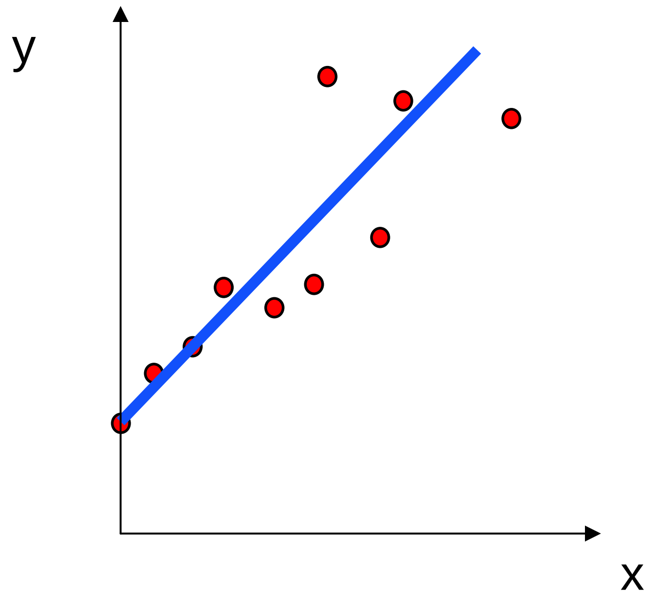
# Univariate Linear Model

- Y is what we want to predict (dependent variable)
- X are the predictors (independent variables)
- Y=f(X), where f is a linear function

$$y_1 = 1\beta_0 + x_1\beta_1$$

$$y_2 = 1\beta_0 + x_2\beta_1$$

…



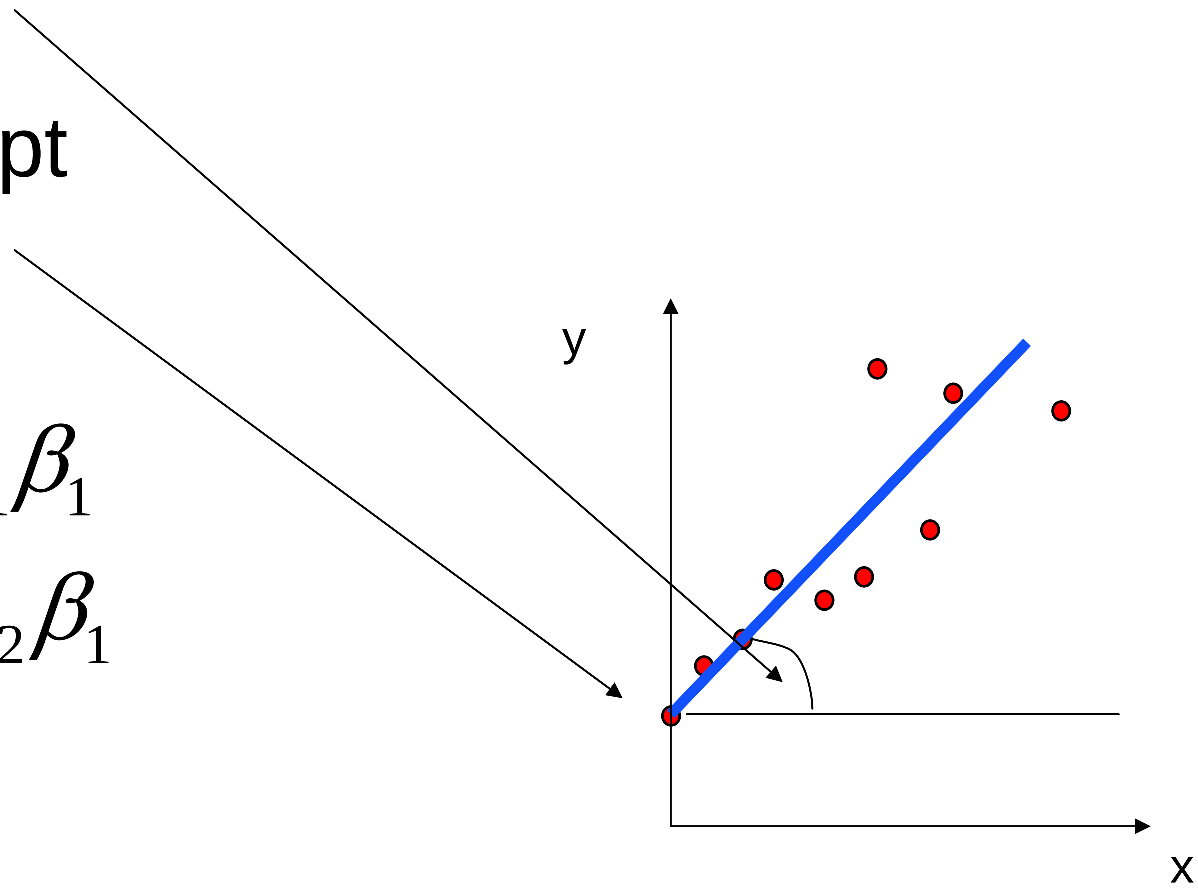
# Univariate Linear Model

$\beta_1$ is slope
$\beta_0$ is intercept

$$y_1 = 1\beta_0 + x_1\beta_1$$

$$y_2 = 1\beta_0 + x_2\beta_1$$

…



# Multivariate Model

- Simple model: structure and parameters
- 3 predictors, 4 parameters $\beta$
- one of the parameters ($\beta_0$) is a constant

$$y_1 = 1\beta_0 + x_{11}\beta_1 + x_{12}\beta_2 + x_{13}\beta_3$$

$$y_2 = 1\beta_0 + x_{21}\beta_1 + x_{22}\beta_2 + x_{23}\beta_3$$

$$\begin{bmatrix} 1 & \mathbf{1} \\ x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

# Notation and Terminology

- $x_i$ is vector of j *inputs*, *covariates*, *independent variables*, or *predictors* for case i (i.e., what we $_{x_1}$ know for all cases)

$$\begin{bmatrix} age \\ salt \\ smoke \end{bmatrix} = \begin{bmatrix} 30 \\ 10 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

- X is matrix of *j* x n
  $x_i$ column vectors (input data for each case)

$$\begin{bmatrix} x_{11} x_{21} \\ x_{12} x_{22} \\ x_{13} x_{23} \end{bmatrix}^T = \begin{bmatrix} x_{11} x_{12} x_{13} \\ x_{21} x_{22} x_{23} \end{bmatrix}$$

# Prediction

- $y_i$ is scalar: *output*, *dependent variable* (i.e., what we want to predict)

e.g., mean blood pressure

$$\begin{bmatrix} pred_{pat1} \\ pred_{pat2} \end{bmatrix} = \begin{bmatrix} 100 \\ 98 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

- Y is vector of $y_i$

# Multivariate Linear Model

$$y_1 = 1\beta_0 + x_{11}\beta_1 + x_{12}\beta_2 + x_{13}\beta_3$$
$$y_2 = 1\beta_0 + x_{21}\beta_1 + x_{22}\beta_2 + x_{23}\beta_3$$

Y = $X^T\beta$, where each $x_i$ includes a term for 1 (constant) ($x_{10}$=1, $x_{20}$=1, etc.) to be multiplied by the intercept $\beta_0$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_{10} x_{11} x_{12} x_{13} \\ x_{20} x_{21} x_{22} x_{23} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

# Regression and Classification

Regression: continuous outcome

- E.g., blood pressure

$$Y = f(X)$$

Classification: categorical outcome

- E.g., death (binary)

$$G = \hat{G}(X)$$

# Loss function

- Y and X random variables
- f(x) is the model
- L(Y,f(X)) is the loss function (penalty for being wrong)
- It is a function of how much to pay for discrepancies between Y (real observation) and f(X) estimated value for an observation)
- In several cases, we use only the error and leave the cost for the decision analysis model

# Regression Problems

Let's concentrate on simple errors for now:

- Expected Prediction Error (EPE):

  $[Y-f(X)]^2$

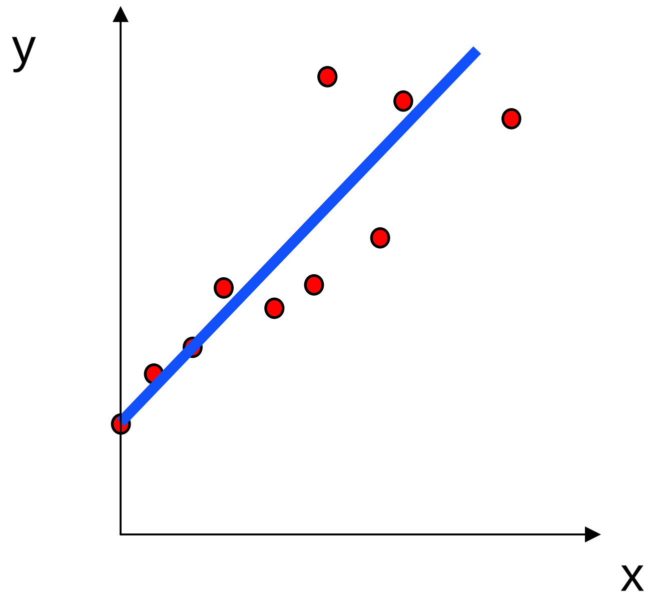
  $|Y-f(X)|$

- These two error functions imply that errors in both directions are considered the same way.

# Univariate Linear Model

$$\hat{y}_1 = 1\beta_0 + x_1\beta_1$$
$$\hat{y}_2 = 1\beta_0 + x_2\beta_1$$

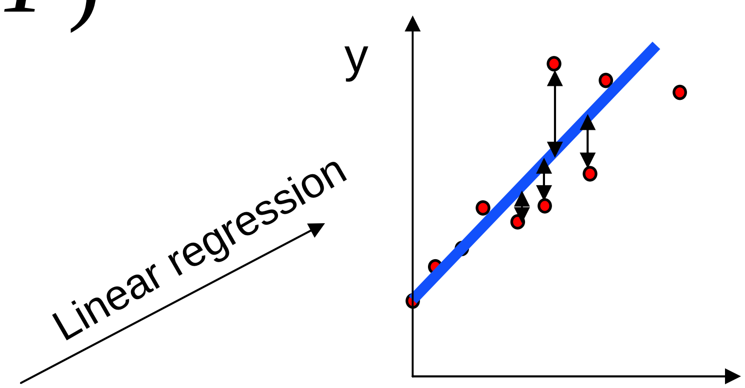
$$y_1 = 1\beta_0 + x_1\beta_1 + error$$
$$y_2 = 1\beta_0 + x_2\beta_1 + error$$



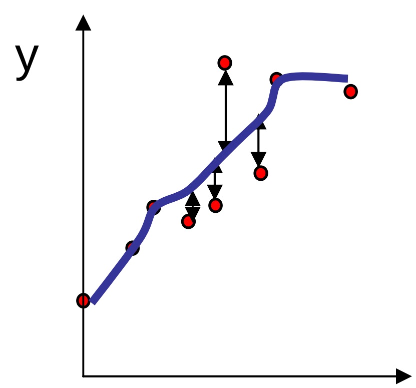
# Squared Errors

$$SSE = \sum_{i=1}^{n} (Y - \hat{Y})^2$$

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$




$$\hat{Y} = f(X)$$



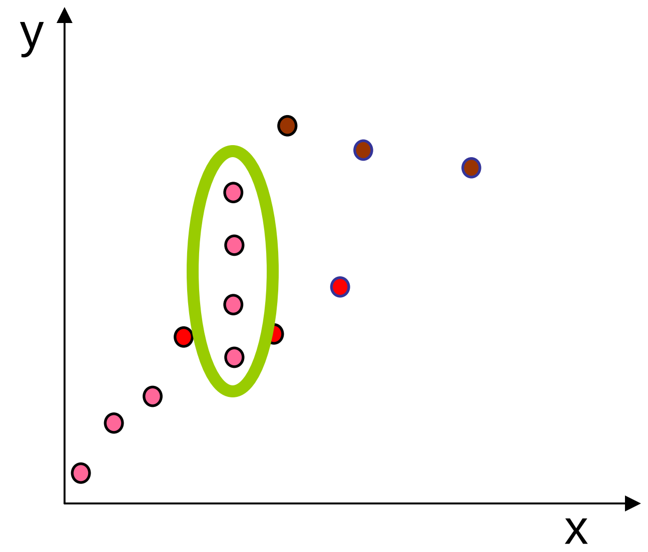
# Conditioning on x

- x is a certain value

$$\hat{y}(x) = 1/k \sum_{x_i = x)} y_i$$

$$f(x) = Ave(y_i \mid x_i = x)$$

- Expectation is approximated by average
- Conditioning is on x



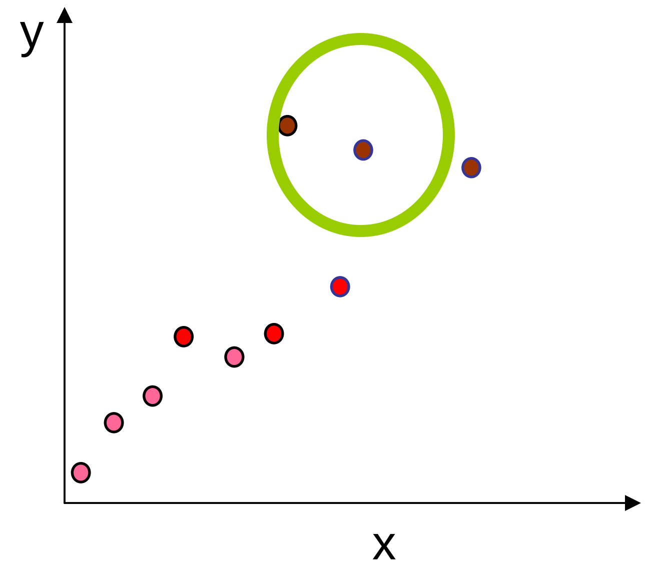
# k -Nearest Neighbors

- N is neighborhood

$$\hat{y}(x) = 1/k \sum_{x_i \in N_k(x)} y_i$$

$$f(x) = Ave(y_i \mid x_i \in N_k(x))$$

- Expectation is approximated by average
- Conditioning is on neighborhood

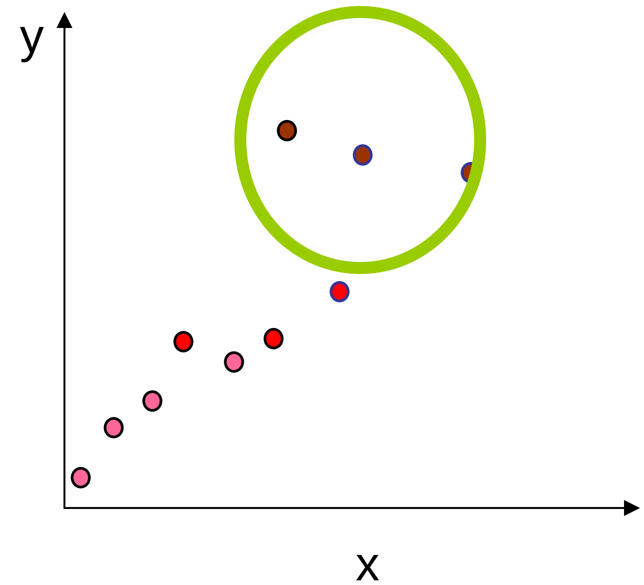


# Nearest Neighbors

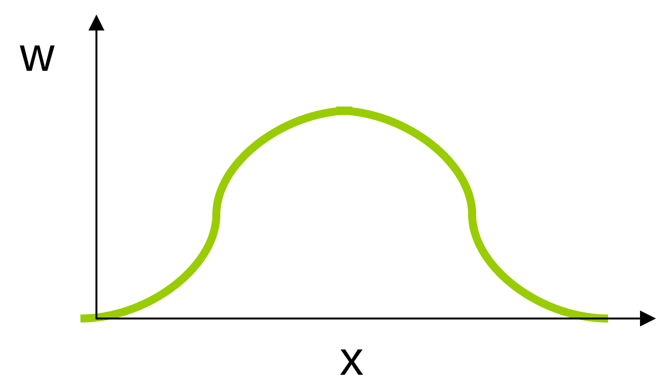
- N is continuous neighborhood

$$\hat{Y}(x) = 1/n \sum y_i w_i$$

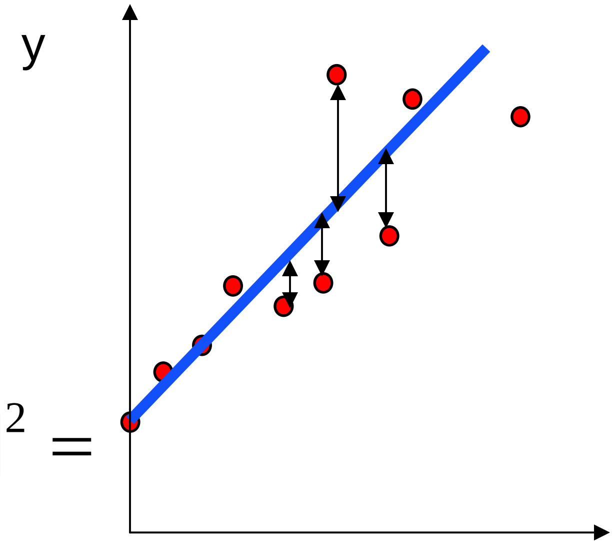
$$f(x) = WeightedAve(y_i \mid x)$$




- Expectation is approximated by weighted average



# Minimize Sum of Squared Errors




$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

$$SSE = \sum_{i=1}^{n} (y - \hat{y})^2 = \sum_{i=1}^{n} [y - (\beta_0 + \beta_1 x)]^2 =$$

$$= \sum_{i=1}^{n} (y^2 - 2y(\beta_0 + \beta_1 x) + (\beta_0 + \beta_1 x)^2) =$$

$$= \sum_{i=1}^{n} (y^2 - 2y\beta_0 - 2y\beta_1 x + \beta_0{}^2 + 2\beta_0\beta_1 x + \beta_1{}^2 x^2)$$

# (derivative wrt $\beta_0$) = 0

$$\sum_{i=1}^{n}(y^2 - 2y\beta_0 - 2y\beta_1 x + \beta_0{}^2 + 2\beta_0\beta_1 x + \beta_1{}^2 x^2)$$

$$\frac{\partial SE}{\partial \beta_0} = 2\sum_{i=1}^{n}(-y + \beta_0 + \beta_1 x) = 0$$

$$\beta_0 n + \beta_1 \sum x = \sum y$$

Normal equation 1

# (derivative wrt $\beta_1$) = 0

$$\sum_{i=1}^{n}(y^2 - 2y\beta_0 - 2y\beta_1 x + \beta_0{}^2 + 2\beta_0\beta_1 x + \beta_1{}^2 x^2)$$

$$\frac{\partial SE}{\partial \beta_1} = 2\sum_{i=1}^{n}(-yx + \beta_0 x + \beta_1 x^2) = 0$$

$$\beta_0 \sum x + \beta_1 \sum x^2 = \sum yx$$ Normal equation 2

# Solve system of normal equations

$$\beta_0 n + \beta_1 \sum x = \sum y$$

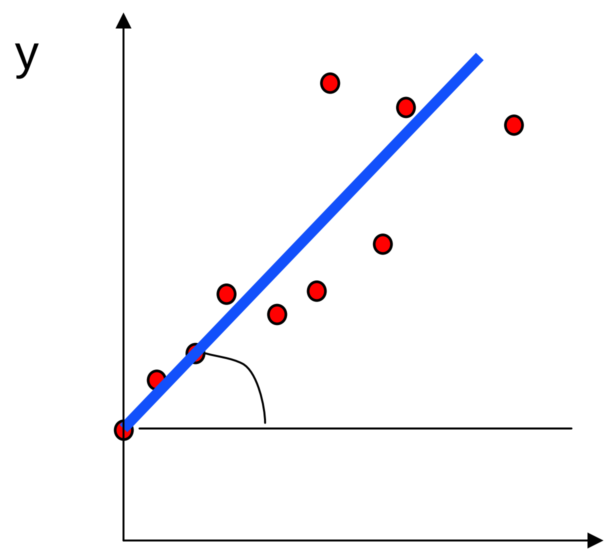
Normal equation 1

$$\beta_0 \sum x + \beta_1 \sum x^2 = \sum yx$$

Normal equation 2

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$



# Limitations of linear regression

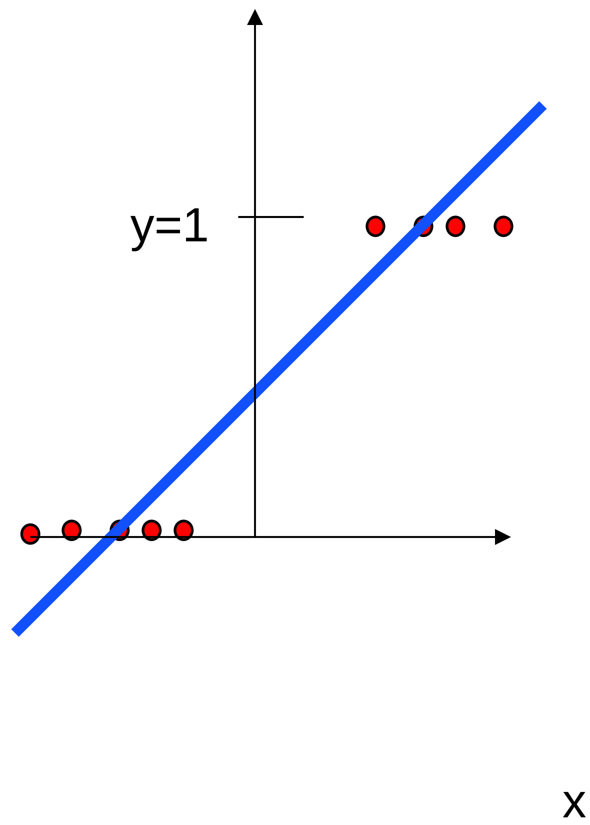
- Assumes conditional probability p(Y|X) is normal
- Assumes equal variance in every X
- It's linear ☺

(but we can always use interaction or transformed terms)

# Linear Regression for Classification?

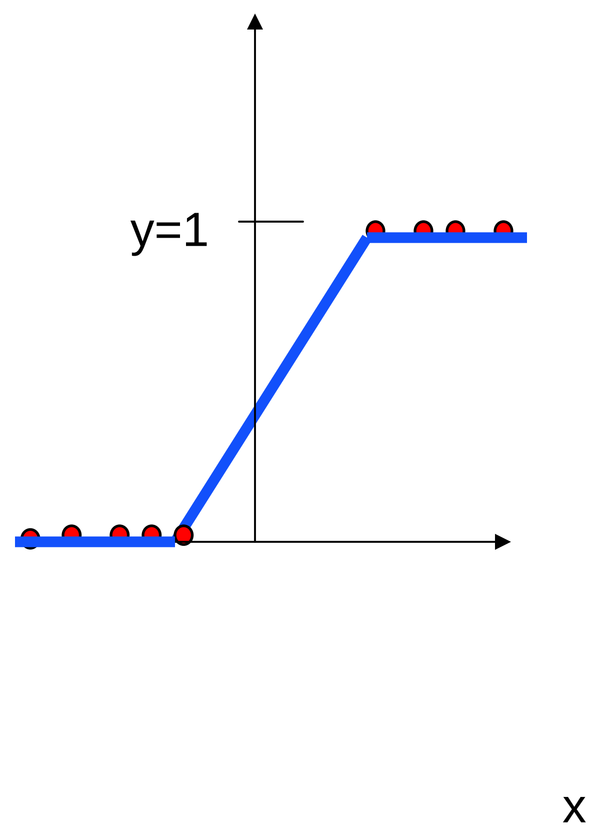
$$y = p$$

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$



# Linear Probability Model

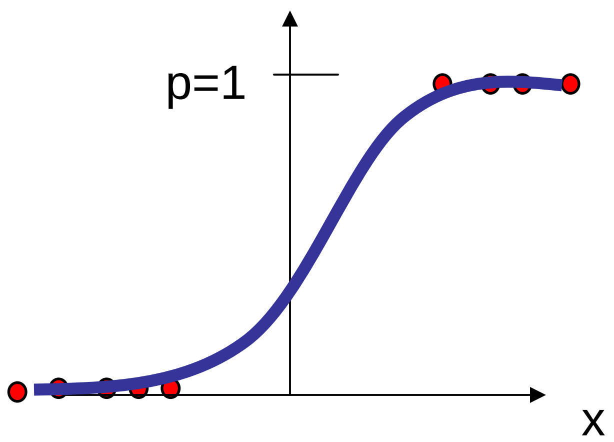
$$\hat{y}_i = \begin{cases} 0 & \text{for } 0 >= \beta_0 + \beta_1 x_i \\ \beta_0 + \beta_1 x_i & \text{for } 0 <= \beta_0 + \beta_1 x_i <= 1 \\ 1 & \text{for } \beta_0 + \beta_1 x_i >= 1 \end{cases}$$



# Logit Model

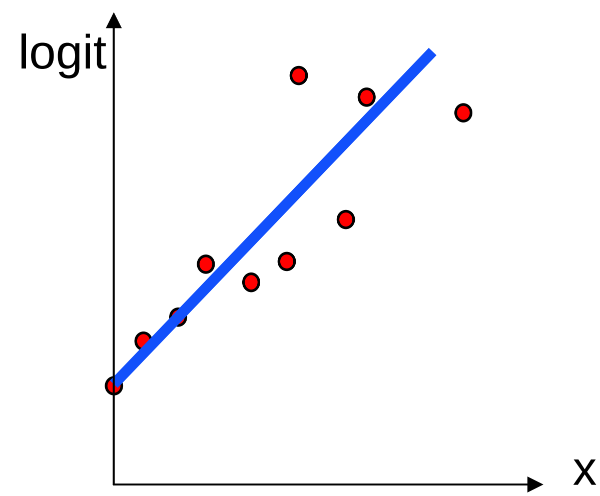
$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

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




$$\log\left[\frac{p_i}{1 - p_i}\right] = \beta_0 + \beta_1 x_i$$

$$\log\left[\frac{p_i}{1 - p_i}\right] = \sum_i \beta x_i$$



# Two dimensions