

**Base initiale (DATASET)**

Variables prédictives (X1, ..., Xp)	Cible (Y)	statut
		train
		test

**Base d'apprentissage (TRAIN SET)**

Variables prédictives (X1, ..., Xp)	Cible (Y)
(1)	

statut = train

Variables prédictives (X1, ..., Xp)	Cible (Y)
(3)	

statut = test

**Base de test (TEST SET)**

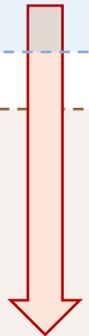
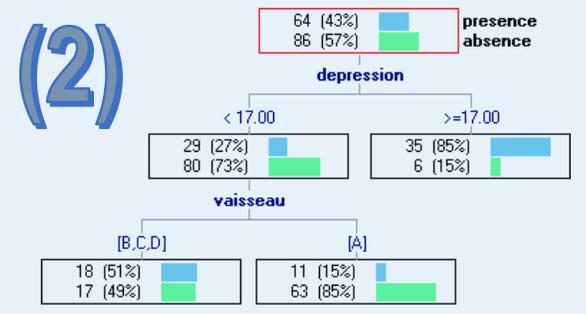
Subdivision des données (0) en échantillons d'apprentissage (1) et de test (3).



**Modélisation (2)**

**Modélisation – Arbre de décision:**

- (a) Construction de l'arbre
- (b) Extraction des règles d'affectation



**Prédiction**

Application des règles prédictives sur l'échantillon test. Construction de la prédiction  $Y^{\wedge}$

**(4)**

Cible (Y)	$Y^{\wedge}$

**Mesure des performances**

Confrontation entre valeurs observées de Y et prédictions

**(5)**

**Matrice de confusion**

		$Y^{\wedge}$	
		$+^{\wedge}$	$-^{\wedge}$
Y	+	a	b
	-	c	d

**Indicateurs**

- Taux d'erreur
- Rappel (sensibilité)
- Précision

(Voir Support de cours, page 10)

Evaluation du modèle prédictif (4 et 5)