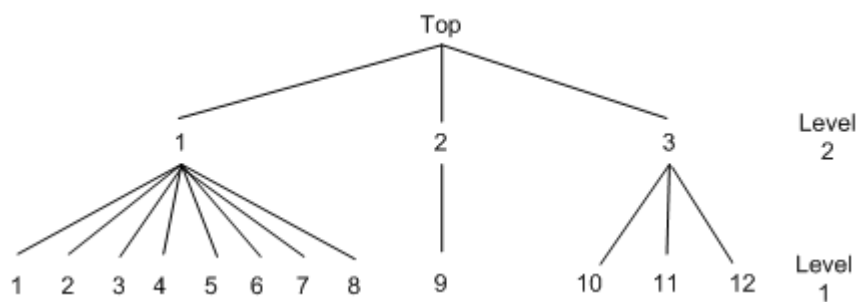
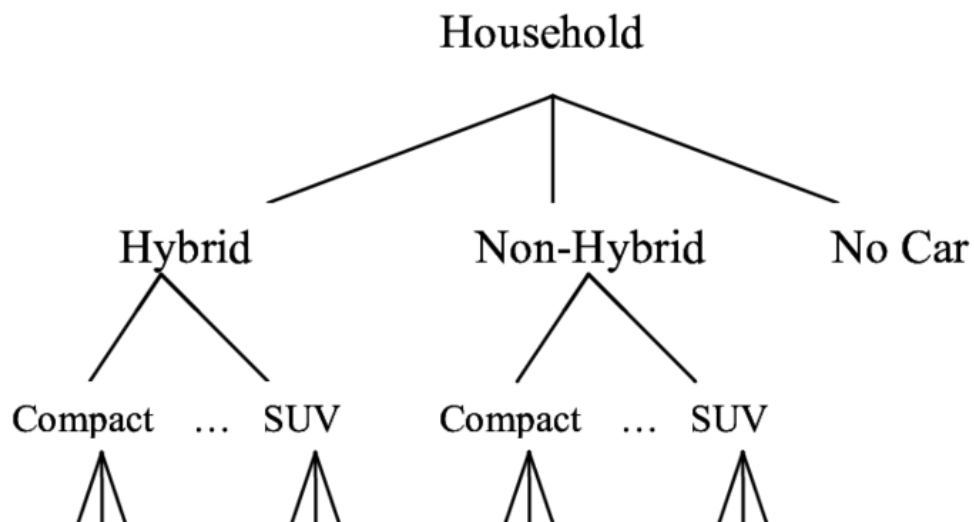
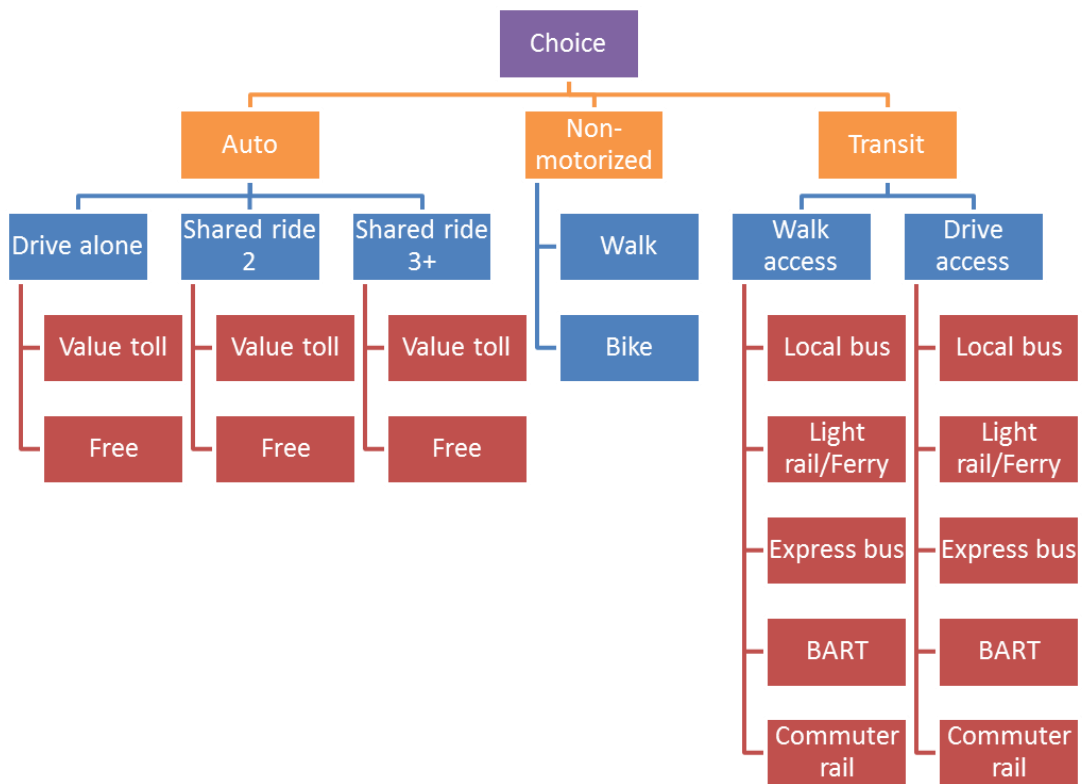
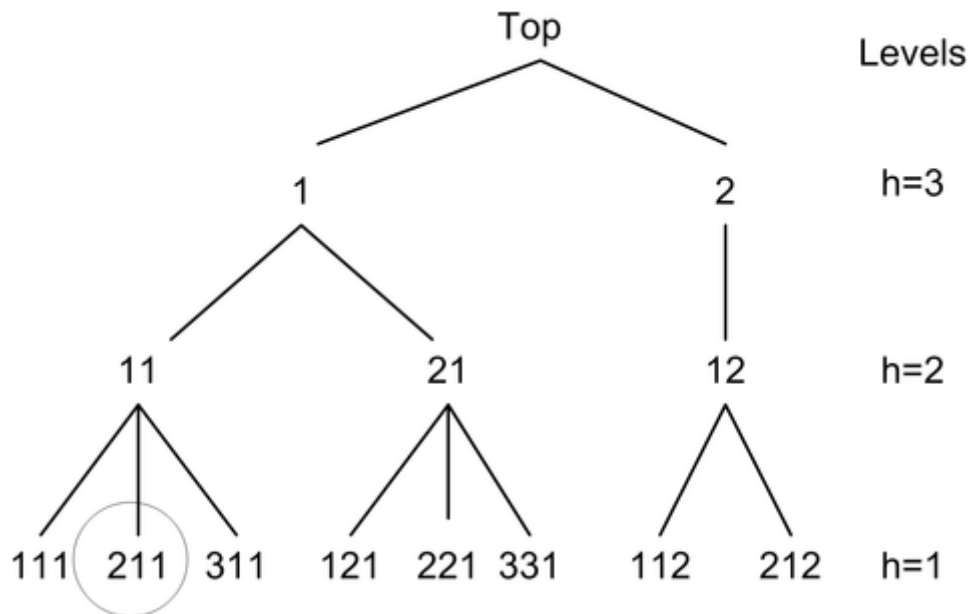


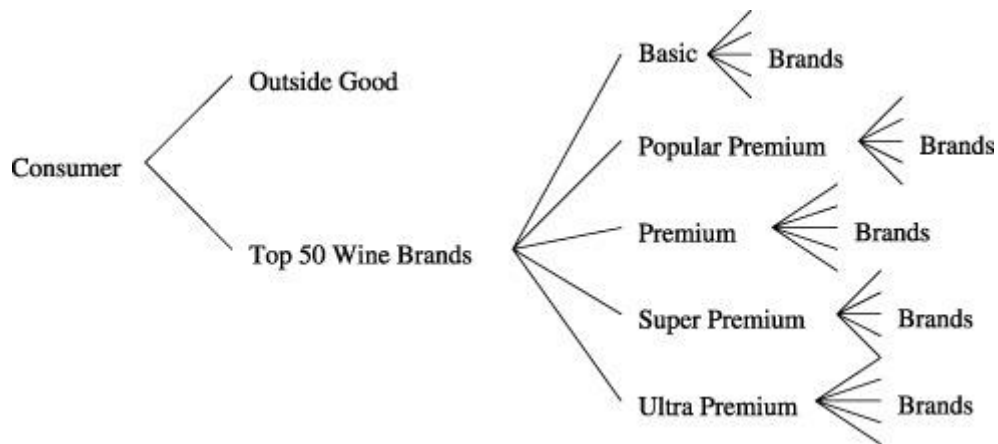
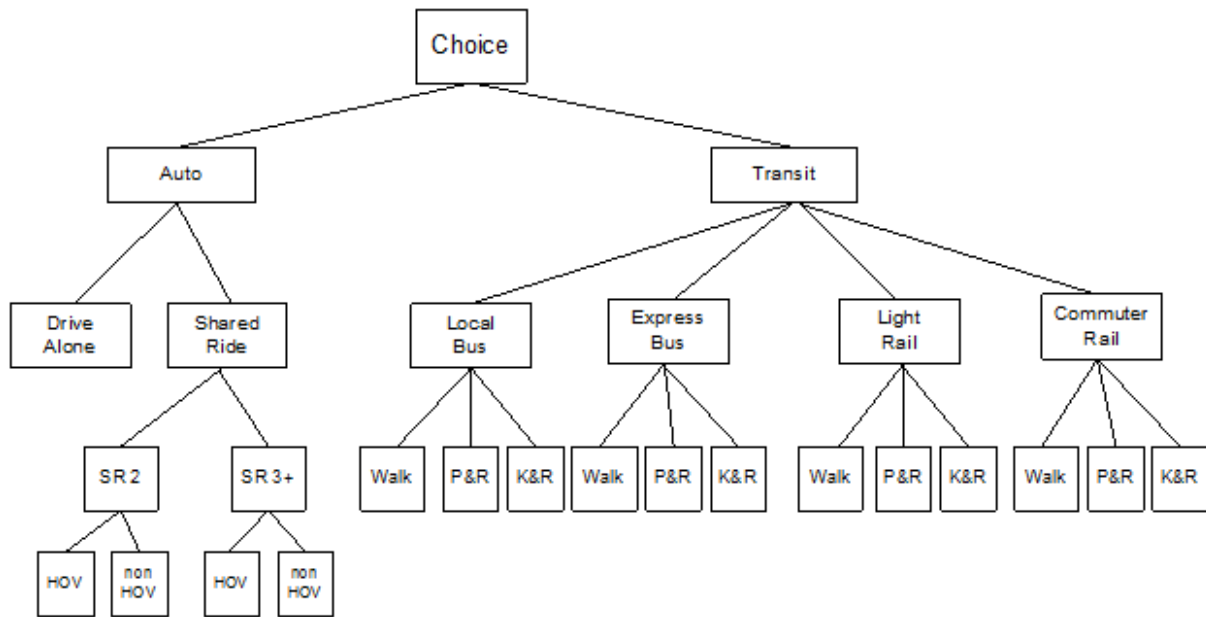
## **CHAPITRE 8 : ARBRES DE DECISION ET NESTED LOGIT**

**(imbriqué, hiérarchique...)**

### **SECTION 1 : Des exemples graphiques**





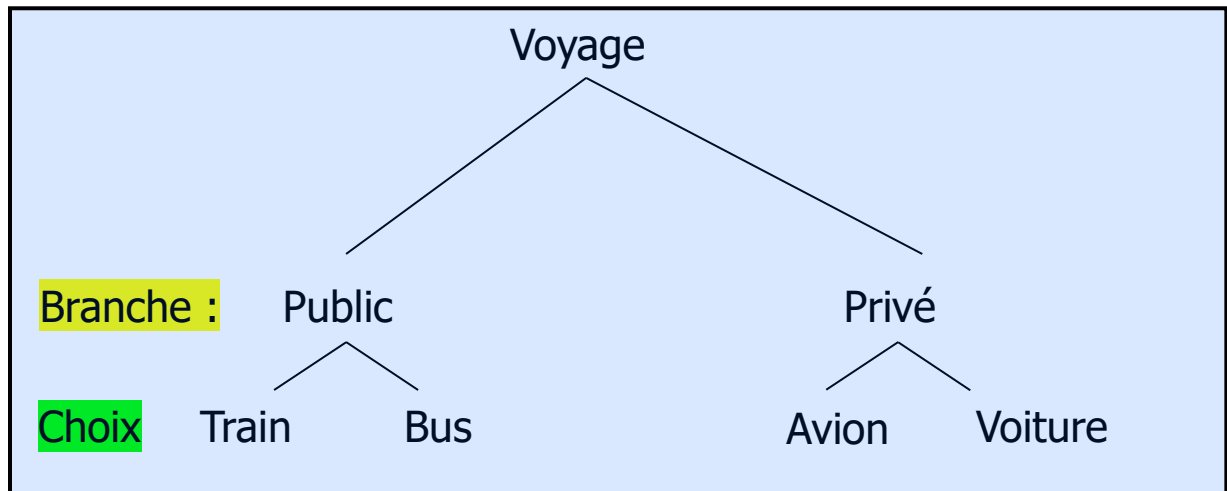


## SECTION 2 : La théorie

Nous avons **j choix** (appelé aussi alternatives).

Ces choix sont regroupés en **branches** ( $B_s$ ). Ces branches sont imbriquées (nested). On suppose des corrélations dans la branche mais pas entre les branches (indépendance).

Exemple :



Il y a 4 choix : avion-voiture-train-bus  
 Il y a 2 branches : privé-public

Avec pour hypothèses :

- (avion-voiture) corrélés,
- (train-bus) corrélés

mais (avion-train) pas corrélés.

Les individus (n) choisissent les options avec la plus haute utilité :

$$U_{nj} = X_{nj}\beta + Y_s\alpha + \varepsilon_{nj}$$

Avec X des exogènes

Y des caractéristiques de la branche

Nous pouvons décomposer l'utilité de la manière suivante :

$$U(\text{Choix}) = U(\text{Choix/Branche}) + U(\text{Branche})$$

Avec :

$U(\text{Branche})$  = fonction de certaines variables (Y) caractéristiques de la branche

$U(\text{Choix/Branche})$  = fonction de certaines variables (X) qui varient selon les choix.

On en déduit :

$$\text{Prob}[\text{choix}, \text{branche}] = \text{Prob}(\text{branche}) * \text{Prob}[\text{Choix/Branche}]$$

Avec :

$$P(\text{choix} = k/j) = \frac{\exp(bX_{k/j})}{\sum_{m=1}^{k/j} \exp(bX_{m/j})}$$

Les valeurs inclusives pour la branche sont :

$$IV(j) = \log \left[ \sum_{m=1}^{k/j} \exp(bX_{m/j}) \right]$$

Les probabilités de branches sont :

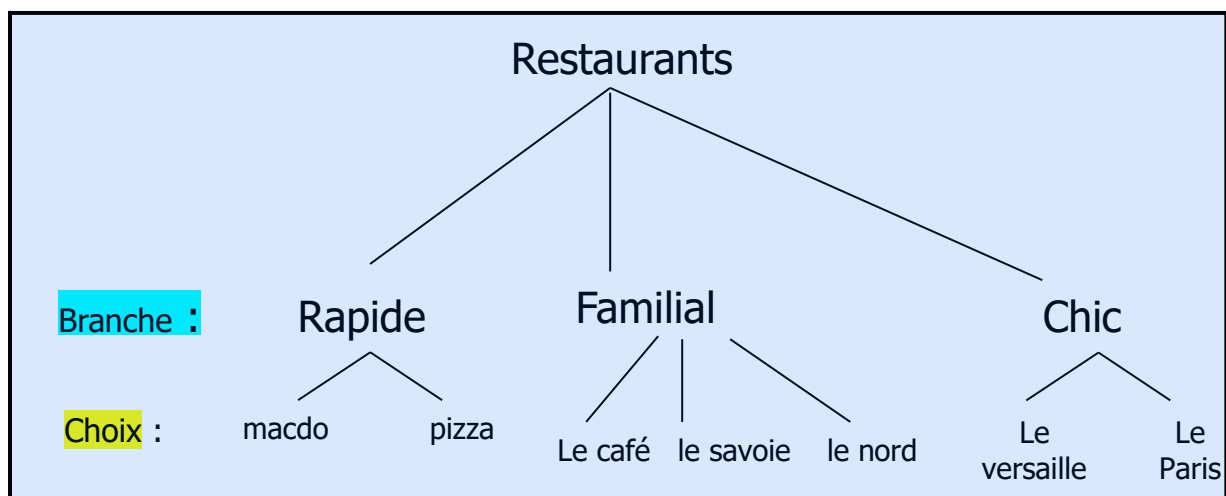
$$\text{Prob}(\text{Branch} = j) = \frac{\exp \left[ \lambda_j (\mathbf{y}' \mathbf{y}_j + IV(j)) \right]}{\sum_{b=1}^B \exp \left[ \lambda_b (\mathbf{y}' \mathbf{y}_b + IV(b)) \right]}$$

Si  $\lambda_j = 1$  nous obtenons le modèle logit multinomial.

### SECTION 3 : Exemple n°1 : choix d'un restaurant

Base de données : RESTAURANTS.XLS

N=2100 observations avec 3 choix au niveau de la **branche** et 7 choix au niveau des **alternatives** :



Variable endogène : CHOIX (un seul restaurant parmi les 7)

Restaurant						
rapide		familial			chic	
macdo	pizza	lecafé	lesavoie	lenord	le versaille	le paris

ID	restaurant	altresto	REVENU	COUT	ENFANTS	EVALUATION	distance	CHOIX	type	atype
1	MACdo	1	39	5,444695	1	0	1,245553	1	rapide	1
1	PIZZA	2	39	6,19446	1	1	2,82493	0	rapide	1
1	LECAFE	3	39	8,182085	1	2	4,21293	0	familial	2
1	LESAVOIE	4	39	9,861741	1	3	4,167634	0	familial	2
1	LENORD	5	39	9,667909	1	2	6,330531	0	familial	2
1	LEVERSAILLE	6	39	25,95777	1	4	10,19829	0	chic	3
1	LEPARIS	7	39	28,99846	1	5	5,601388	0	chic	3
2	MACdo	1	58	5,26874	3	0	4,162657	0	rapide	1
2	PIZZA	2	58	5,728618	3	1	2,865081	0	rapide	1
2	LECAFE	3	58	7,054855	3	2	5,337799	0	familial	2
2	LESAVOIE	4	58	10,78514	3	3	4,282864	1	familial	2
2	LENORD	5	58	8,313948	3	2	8,133914	0	familial	2
2	LEVERSAILLE	6	58	21,2801	3	4	8,664631	0	chic	3
2	LEPARIS	7	58	25,87567	3	5	9,119597	0	chic	3
3	MACdo	1	30	4,616315	3	0	2,112586	1	rapide	1
3	PIZZA	2	30	5,992166	3	1	2,215329	0	rapide	1
3	LECAFE	3	30	7,980528	3	2	6,978715	0	familial	2
3	LESAVOIE	4	30	10,0605	3	3	5,117877	0	familial	2
3	LENORD	5	30	8,76644	3	2	5,312941	0	familial	2
3	LEVERSAILLE	6	30	23,64499	3	4	9,551273	0	chic	3
3	LEPARIS	7	30	24,72128	3	5	5,539806	0	chic	3
4	MACdo	1	24	4,974085	5	0	2,401803	0	rapide	1
4	PIZZA	2	24	6,244085	5	1	2,059263	0	rapide	1
4	LECAFE	3	24	8,451072	5	2	4,944536	0	familial	2
4	LESAVOIE	4	24	10,14855	5	3	4,783116	1	familial	2
4	LENORD	5	24	9,105867	5	2	5,897751	0	familial	2
4	LEVERSAILLE	6	24	22,22009	5	4	7,547688	0	chic	3
4	LEPARIS	7	24	26,61075	5	5	7,36597	0	chic	3
5	MACdo	1	35	4,681487	0	0	2,669857	1	rapide	1
5	PIZZA	2	35	6,325174	0	1	4,312208	0	rapide	1
5	LECAFE	3	35	7,587141	0	2	4,816533	0	familial	2
5	LESAVOIE	4	35	10,04325	0	3	4,730175	0	familial	2
5	LENORD	5	35	9,46889	0	2	5,296418	0	familial	2
5	LEVERSAILLE	6	35	25,24608	0	4	6,08509	0	chic	3
5	LEPARIS	7	35	23,20879	0	5	5,77845	0	chic	3

## Variables exogènes :

**REVENU** : de la personne en \$

**COUT** : montant en \$ des repas

**ENFANTS** : nb d'enfants

**EVALUATION** : de 1 à 5 d'un guide spécialisé

**DISTANCE** : en kms de la maison



Nous obtenons :

Mesures du critère qualificatif de lissage		
Mesure	Valeur	Formule
Likelihood Ratio (R)	143.53	$2 * (\text{LogL} - \text{LogL0})$
Upper Bound of R (U)	1167.5	$-2 * \text{LogL0}$
Aldrich-Nelson	0.3236	$R / (R+N)$
Cragg-Uhler 1	0.3803	$1 - \exp(-R/N)$
Cragg-Uhler 2	0.3882	$(1 - \exp(-R/N)) / (1 - \exp(-U/N))$
Estrella	0.3998	$1 - (1 - R/U)^{(U/N)}$
Adjusted Estrella	0.386	$1 - ((\text{LogL} - K) / \text{LogL0})^{(-2/N * \text{LogL0})}$
McFadden's LRI	0.1229	$R / U$
Veall-Zimmermann	0.4068	$(R * (U+N)) / (U * (R+N))$
N = # d'observations, K = # de régresseurs		

### nested logit

The MDC Procedure

Nested Logit Estimates

Résultats estimés des paramètres					
Paramètre	DDL	Valeur estimée	Erreur type	Valeur du test t	Approx. de Pr >  t
COUT_L1	1	-0.0174	0.0168	-1.04	0.2993
distance_L1	1	-0.0700	0.0423	-1.65	0.0979
EVALUATION_L1	1	0.1981	0.1150	1.72	0.0852
INC_L2G1C1	1	5.1690	4.8674	1.06	0.2883
INC_L2G1C2	1	4.3982	2.2347	1.97	0.0491
INC_L2G1C3	1	4.7140	3.2245	1.46	0.1438

### Prévisions nested modèle

#### Procédure MEANS

Variable d'analyse : probs	
altresto	Moyenne
1	0.0420670
2	0.0505237
3	0.2350890
4	0.2747976
5	0.2292322
6	0.0765284
7	0.0917621

### Prévisions nested modèle

#### Procédure MEANS

Variable d'analyse : probs	
atype	Moyenne
1	0.0462954
2	0.2463729
3	0.0841453

Ce qui donne en termes de probabilités :

Probabilités des 7 alternatives :

Restaurant						
rapide		familial			chic	
0.0420	0.050	0.235	0.274	0.229	0.076	0.091
1	2	3	4	5	6	7

### Probabilités des 3 branches :

Restaurant		
<b>0.092</b>	<b>0.738</b>	<b>.167</b>
1	2	3

Remarque : pour obtenir les probabilités des 3 branches :

-soit on additionne les probabilités alternatives par alternatives :

$$.0420 + .050 = .092$$

$$.235 + .274 + .229 = .738$$

$$.076 + .091 = .167$$

-soit on multiplie par 2 puis 3 puis 2 le résultat de 'atype' :

$$.0462954 * 2 = .09259$$

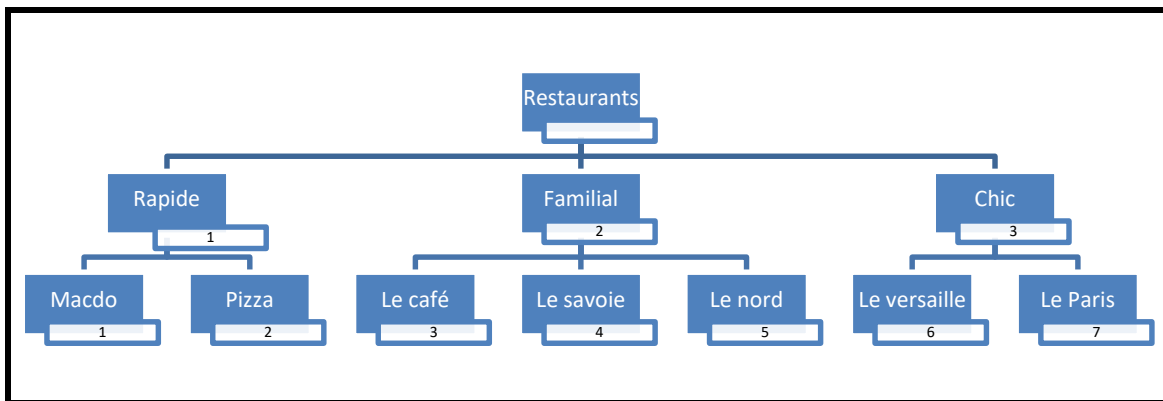
$$.24637729 * 3 = .739$$

$$.0841453 * 2 = .16829$$

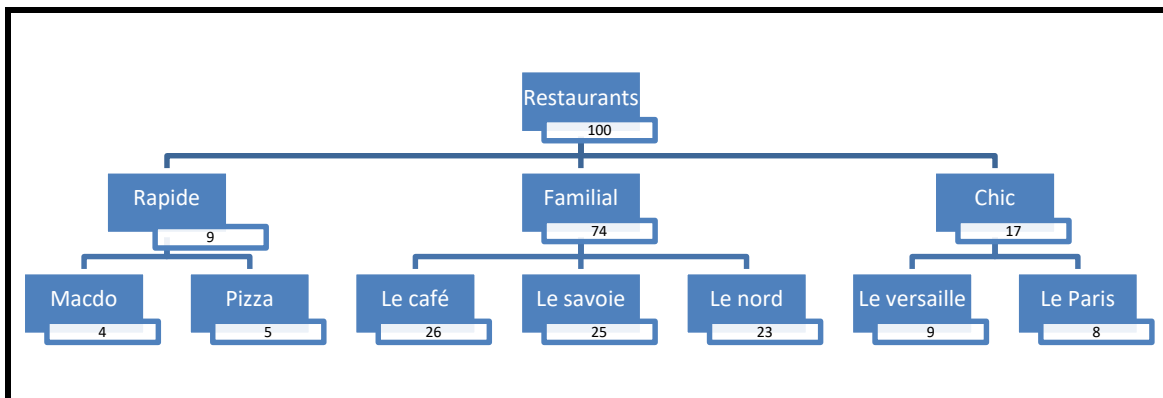
## ETUDE DETAILLEE

### Partie I : Présentation de l'organigramme

2 niveaux :    niveau 1 : altresto :    1-2-3-4-5-6-7  
                  Niveau 2 : atype :        1-2-3



### Partie II: Probabilités réelles de chaque choix (en %)

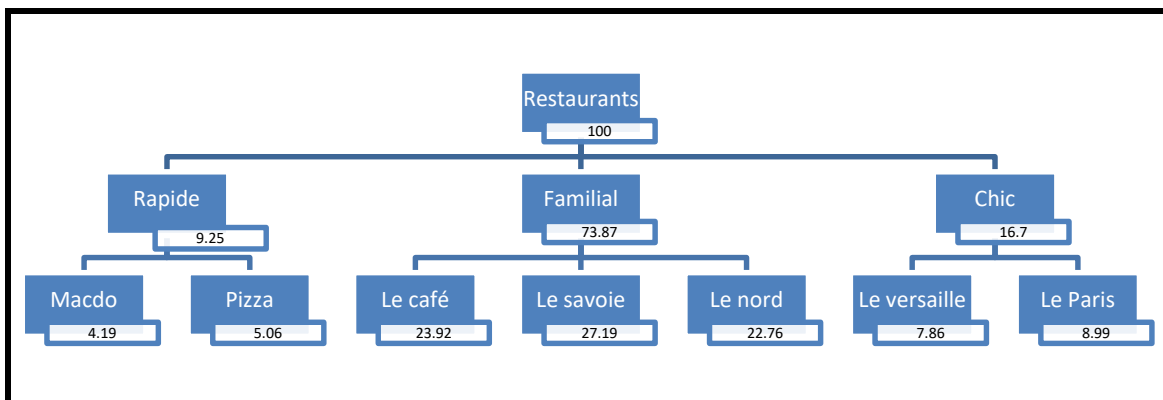


### Partie III: Probabilités calculées de chaque choix (en %)

Il y a 5 variables exogènes disponibles : revenu-coût-enfants-évaluation-distance

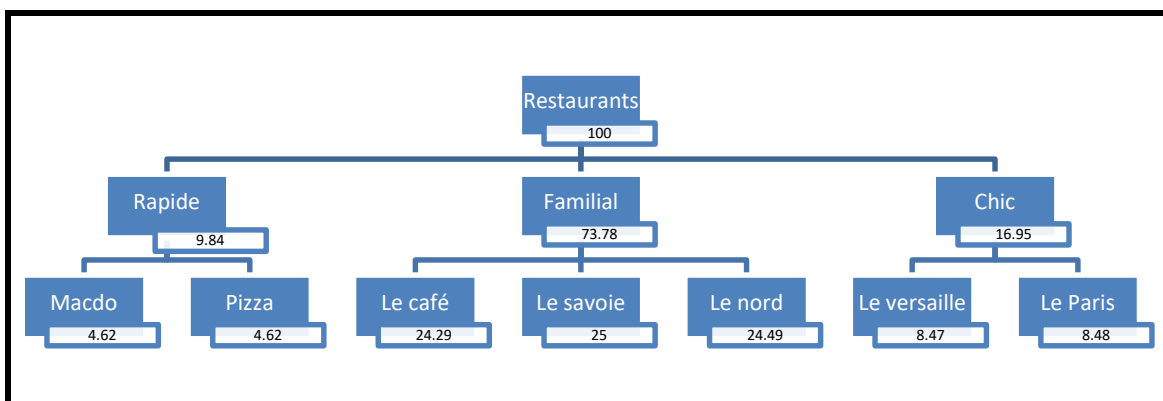
#### Version 1 : même choix de variables exogènes pour le niveau 1

revenu-coût-enfants-évaluation-distance

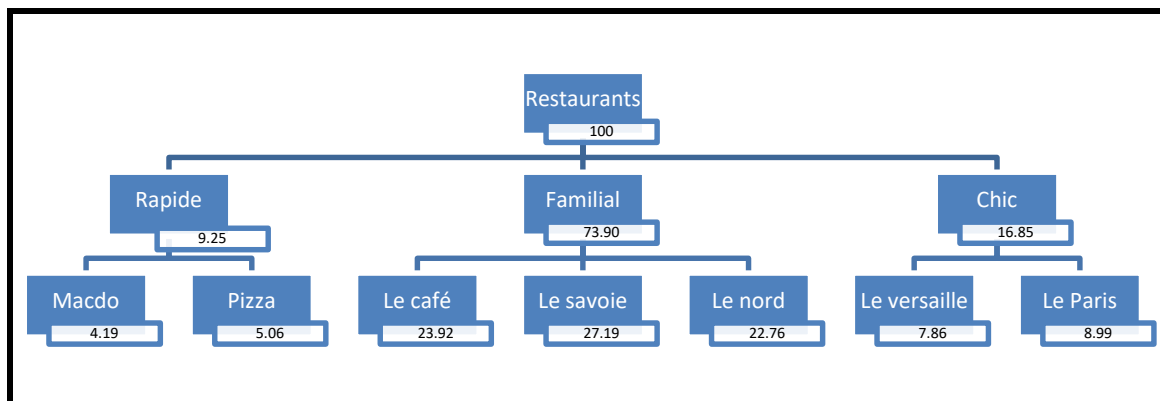


#### Version 2 : différents choix de variables exogènes pour le niveau 1 et le niveau 2

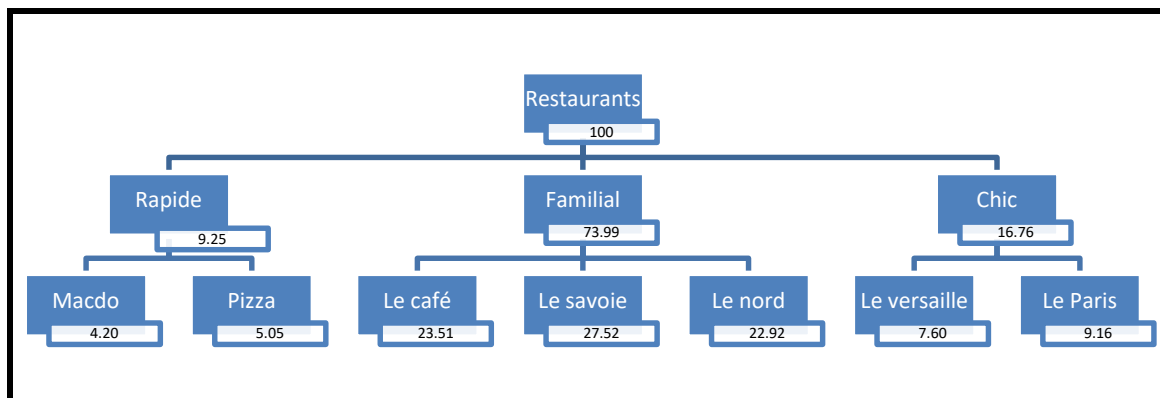
utility  $u(1,6,7 @3) = \text{distance}$  ,  
 $u(1,3,4,5 @2) = \text{cout distance}$  ,  
 $u(1,1,2 @1) = \text{revenu}$  ,  
 $u(2,1,2,3) = \text{cout distance}$  ;



### Version 3 : IV identiques pour 2-3 niveau 2



### Version 4 : IV identiques pour 1-2-3 niveau 2



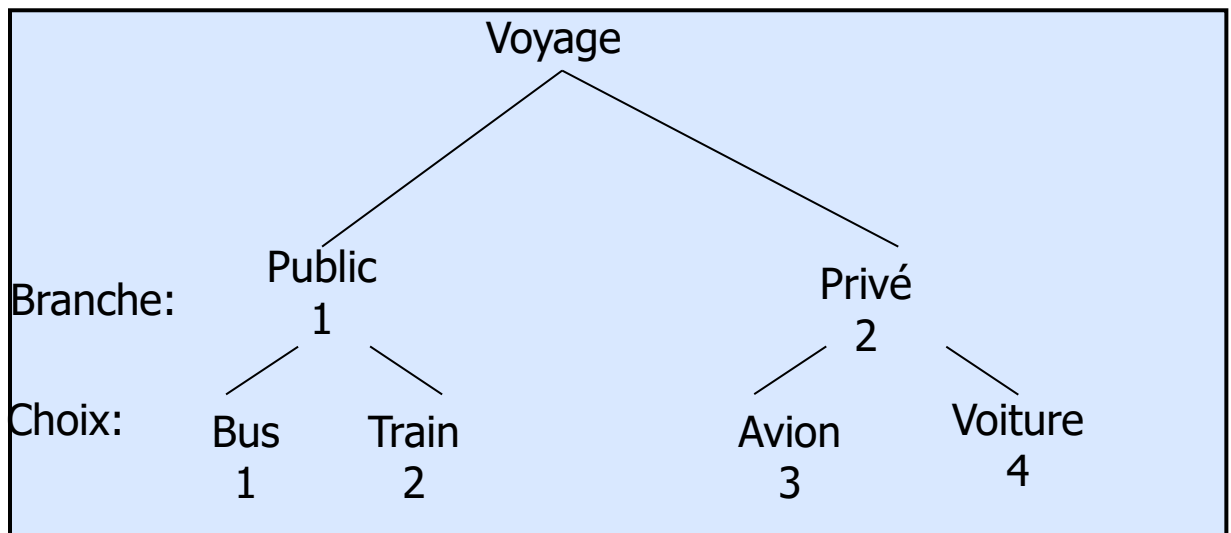
### Résumé

Choix	Réalité	V1	V2	V3	V4
Macdo	4	4.19	4.62	4.19	4.20
Pizza	5	5.06	4.62	5.06	5.05
Le café	26	23.92	24.29	23.92	23.51
Le savoie	25	27.19	25	27.19	27.52
Le nord	23	22.76	24.49	22.76	22.92
Le Versailles	9	7.86	8.47	7.86	7.60
Le Paris	8	8.99	8.48	8.99	9.16

## SECTION 4 : Exemple n°2 : choix d'un mode de transport

Base de données : CHOIXTRANSPORT.XLSX

N=840 observations avec 2 choix au niveau de la **branche** et 4 choix au niveau des **alternatives** sur un ensemble de 210 individus :



Variable endogène : CHOIX (un seul mode de transport parmi les 4)

CHOIX	TRANSPORT	ALTTRANSPORT	MODE	ALTMODE	TEMPS ATTENTE	COUT DU TRAJET	TEMPS DE VOYAGE	COUT TOTAL	REVENUMENAGE
0	BUS	1	PUBLIC	1	69	59	100	70	35
0	TRAIN	2	PUBLIC	1	34	31	372	71	35
0	AVION	3	PRIVE	2	35	25	417	70	35
1	VOITURE	4	PRIVE	2	0	10	180	30	35
0	BUS	1	PUBLIC	1	64	58	68	68	30
0	TRAIN	2	PUBLIC	1	44	31	354	84	30
0	AVION	3	PRIVE	2	53	25	399	85	30
1	VOITURE	4	PRIVE	2	0	11	255	50	30
0	BUS	1	PUBLIC	1	69	115	125	129	40
0	TRAIN	2	PUBLIC	1	34	98	892	195	40
0	AVION	3	PRIVE	2	35	53	882	149	40
1	VOITURE	4	PRIVE	2	0	23	720	101	40
0	BUS	1	PUBLIC	1	64	49	68	59	70
0	TRAIN	2	PUBLIC	1	44	26	354	79	70
0	AVION	3	PRIVE	2	53	21	399	81	70
1	VOITURE	4	PRIVE	2	0	5	180	32	70
0	BUS	1	PUBLIC	1	64	60	144	82	45
0	TRAIN	2	PUBLIC	1	44	32	404	93	45
0	AVION	3	PRIVE	2	53	26	449	94	45
1	VOITURE	4	PRIVE	2	0	8	600	99	45
0	BUS	1	PUBLIC	1	69	59	100	70	20
1	TRAIN	2	PUBLIC	1	40	20	345	57	20
0	AVION	3	PRIVE	2	35	13	417	58	20
0	VOITURE	4	PRIVE	2	0	12	284	43	20
1	BUS	1	PUBLIC	1	45	148	115	160	45
0	TRAIN	2	PUBLIC	1	34	111	945	213	45
0	AVION	3	PRIVE	2	35	66	935	167	45
0	VOITURE	4	PRIVE	2	0	36	821	125	45
0	BUS	1	PUBLIC	1	69	121	152	137	12
0	TRAIN	2	PUBLIC	1	34	52	889	149	12
0	AVION	3	PRIVE	2	35	50	879	146	12
1	VOITURE	4	PRIVE	2	0	50	780	135	12
0	BUS	1	PUBLIC	1	69	59	100	70	40
0	TRAIN	2	PUBLIC	1	34	31	372	71	40
0	AVION	3	PRIVE	2	35	25	417	70	40

Variables exogènes :

**TEMPS ATTENTE** : temps d'attente (0 pour la voiture)

**COUT DU TRAJET** : coût du trajet seul en \$

**TEMPS DE VOYAGE** : en minutes

**COUT TOTAL** : coût total du trajet en \$

**REVENUMENAGE** : revenu du ménage en M\$

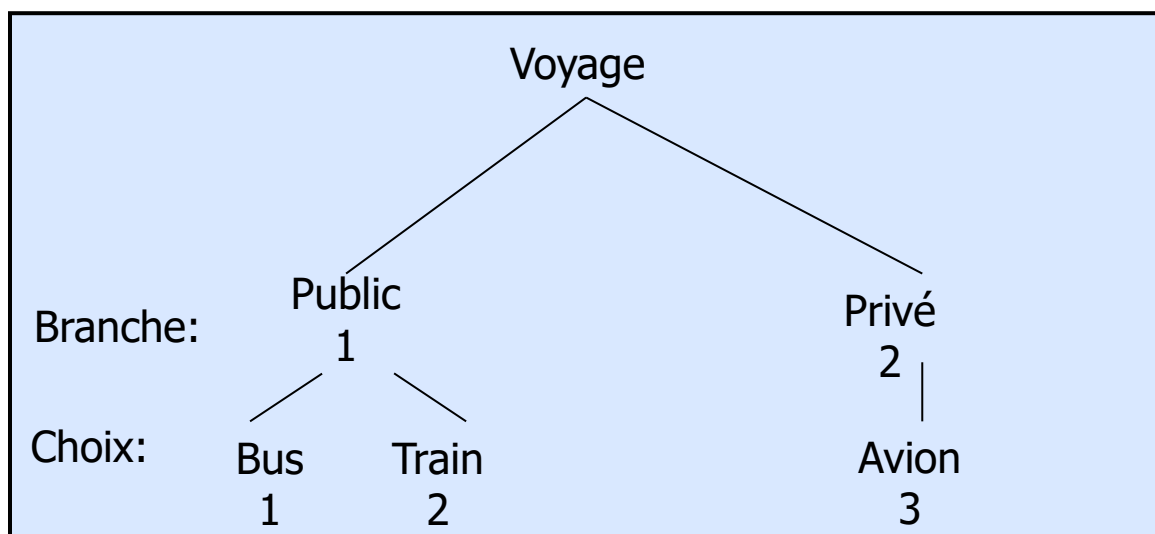


## SECTION 5 : Exemple n°2 bis : choix d'un mode de transport

Base de données : ~ CHOIXTRANSPORT.XLSX

N = ?

2 choix au niveau de la **branche** et 3 choix au niveau des **alternatives** sur un ensemble de ? individus :



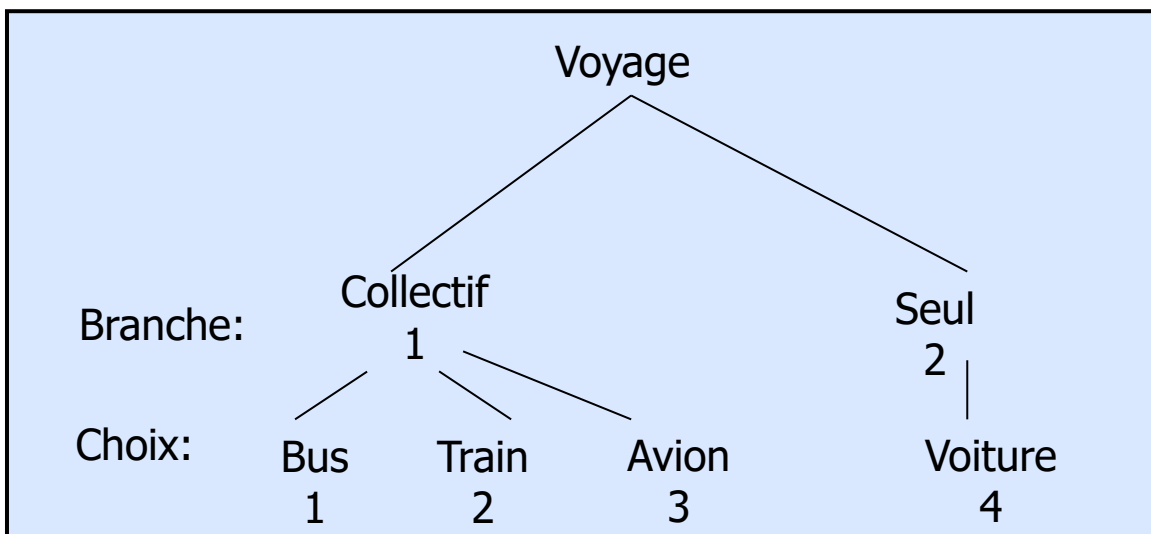
Variable endogène : CHOIX (un seul mode de transport parmi les 3)

## SECTION 6 : Exemple n°2 ter : choix d'un mode de transport (ie dégénérescence des arbres)

Base de données : ~ CHOIXTRANSPORT.XLSX

N = ?

2 choix au niveau de la **branche** et 4 choix au niveau des **alternatives** sur un ensemble de ? individus :



Variable endogène : CHOIX (un seul mode de transport parmi les 4)

---