



UFR SES/SET
Master Science des
Données et
Applications

Statistiques et Econométrie

TRAITEMENT DES VARIABLES QUALITATIVES

Ousmane Dia

Table des matières

1-Calcul les statistiques descriptives associées aux variables du modèle.	3
2-Détermination du niveau de corrélation entre les variables exogènes.	4
3-Estimation du modèle logit multinomial permettant de prévoir la classe de vente d'un article à partir des facteurs explicatifs proposés.	5
4-Test de la significativité des variables explicatives :	5
5-la significativité globale du modèle	6
6-Interprétation économiquement les paramètres estimés.....	6
7-Interprétation de la valeur du R^2 de McFadden.....	6

Remarque sur la variable manquante : Il est très fréquent dans une étude statistique d'être confronté à des problèmes de données manquantes. Pour pallier ce problème, plusieurs méthodes peuvent être retenue pour imputer des valeurs aux cellules manquantes ; on peut par exemple choisir la valeur la plus fréquente de l'échantillon ou la moyenne... Parmi les nombreuses méthodes qui existent il est recommandé de supprimer les lignes des données manquantes si elles sont très négligeables ou tout simplement de ne rien faire. Dans le cas de notre travail nous avons une seule donnée manquante donc nous laissons la base comme telle.

	Obs	VENTES	WE	EXPO	REDUC
78	78	0	1	.5	0
79	79	0	0	3.5	0
80	80	0	1	7	0
81	81	0	1	7	0
82	82	0	1	3.5	0
83	83	.	1	7	0

S

1-Calcul les statistiques descriptives associées aux variables du modèle.

. sum

Variable	Obs	Mean	Std. Dev.	Min	Max
Obs	83	42	24.10394	1	83
VENTES	82	.597561	.7176738	0	2
WE	83	.5421687	.5012473	0	1
EXPO	83	5.439759	2.054735	.5	9
REDUC	83	.0879518	.1086476	0	.3
DIRECT	83	.0361446	.1877845	0	1

2-Détermination du niveau de corrélation entre les variables exogènes.

```
. corr WE EXPO REDUC DIRECT
(obs=83)
```

	WE	EXPO	REDUC	DIRECT
WE	1.0000			
EXPO	-0.1396	1.0000		
REDUC	0.0094	-0.1918	1.0000	
DIRECT	0.1780	-0.1839	0.2607	1.0000

```
. corr VENTES WE EXPO REDUC DIRECT
(obs=82)
```

	VENTES	WE	EXPO	REDUC	DIRECT
VENTES	1.0000				
WE	-0.3186	1.0000			
EXPO	0.4583	-0.1495	1.0000		
REDUC	0.1798	0.0187	-0.1856	1.0000	
DIRECT	-0.0722	0.1811	-0.1828	0.2599	1.0000

3-Estimation du modèle logit multinomial permettant de prévoir la classe de vente d'un article à partir des facteurs explicatifs proposés.

. ologit VENTES WE EXPO REDUC DIRECT						
Iteration 0: log likelihood = -79.482191						
Iteration 1: log likelihood = -62.514701						
Iteration 2: log likelihood = -61.628095						
Iteration 3: log likelihood = -61.623502						
Iteration 4: log likelihood = -61.623501						
Ordered logistic regression						
			Number of obs	=	82	
			LR chi2(4)	=	35.72	
			Prob > chi2	=	0.0000	
Log likelihood = -61.623501			Pseudo R2	=	0.2247	
VENTES	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
WE	-1.395403	.5091821	-2.74	0.006	-2.393382	-.3974245
EXPO	.6220682	.1460136	4.26	0.000	.3358868	.9082496
REDUC	7.434029	2.448079	3.04	0.002	2.635881	12.23218
DIRECT	.2246324	1.333186	0.17	0.866	-2.388364	2.837628
/cut1	3.500077	.9680215			1.60279	5.397365
/cut2	5.960463	1.144376			3.717527	8.203399

4-Test de la significativité des variables explicatives :

Règle générale de teste de significativité des paramètres :

Pour tester la significativité des variables, il suffit de regarder les « z-stat » ou bien les « P-Value » ($P > z$). Pour faire simple ; par exemple une variables est significative avec un intervalle de confiance 95 % si son z-tat est supérieur à un seuil de 1.96 en valeur absolue ou si sa P value est inférieure à 0.05. La distribution des rapports du coefficient et de l'écart type suit une loi normale ; Le z-statistique s'interprète de manière classique à partir des probabilités critique et permet de tester la significativité des paramètres du model :

Hypothèses : $H_0 : B : \text{« Non Significatif »}$, $H_1 : B : \text{« Significatif »}$

Règle de décision : si $(P > z) \leq (1, 5 \text{ et } 10) \%$ On rejette H_0 L'effet est significatif

WE (Type de jour) : la probabilité est de 0.6 % inférieur au seuils 5 % donc on peut dire que l'effet est largement significatif

EXPO (Temps d'exposition) : la probabilité est inférieure aux seuils l'effet est significatif on rejette H0

REDUCT (Réduction sur le prix) : la probabilité est inférieure au seuil l'effet est significatif

DIRECT (variable indicatrice d'émission enregistrée) : La probabilité est de 86.6 %, très largement supérieure au seuil ; on accepte l'hypothèse H0. La variable n'est pas significative

5-la significativité globale du modèle

H0 : « l'ensemble des coefficients est nul »

H1 : « Au moins un coefficient du modèle est non nul »

Le test de significativité global du modèle montre que nous avons trois variables explicatives qui sont significatives donc le modèle a un intérêt d'être gardé

6-Interprétation économiquement les paramètres estimés.

WE : l'indication du jour de diffusion de l'émission a un effet négatif sur le niveau des ventes par article pour chaque émission. Autrement dit le niveau des ventes par article diminue par rapport au jour diffusion de l'émission

EXPO : le temps d'exposition du produit en minute a un effet positif sur le niveau des ventes par article pour chaque émission. Plus l'émission dure plus on enregistre des ventes

REDUCT : La réduction sur le prix des produits a un effet positif sur le niveau des ventes par article pour chaque émission. Plus le prix baisse plus les ventes augmentent

DIRECT : L'indication d'émission enregistrée a un effet positif mais non significatif sur le niveau des ventes par article pour chaque émission

7-Interprétation de la valeur du R^2 de McFadden.

Le « R -carré » mesure la proportion de la variance de Y (variable dépendante) qui est expliquée par la variation de toutes les variables explicatives. Le R-carré est par construction compris entre 0 et 1 ; plus on se rapproche de 1, plus le modèle

est précis. Dans notre exemple, 22.47% de la variation de **VENTE** peut être expliquée par les variations de **WE**, **EXPO**, **REDUC**, **DIRECTE**. Le pourcentage de significativité des variables explicatives est trop faible pour expliquer le modèle

8-D'effectuer une prévision pour un article présenté lors d'une émission en différé diffusée en semaine dont le temps d'exposition est de 7 minutes et sans réduction.

« predict P » : **Pour créer une nouvelle variable**

« predict VENTEprev » : **Pour prédire variable endogène VENTES**

« list Obs WE EXPO REDUC VENTESprev if WE=1 & EXPO==7 & REDUC==0 » : **Pour afficher les prévisions en tenant compte des conditions**

```
. list Obs WE EXPO REDUC VENTESprev if WE==1 & EXPO==7 & REDUC==0
```

	Obs	WE	EXPO	REDUC	VENTES~v
6.	6	1	7	0	.6320457
24.	24	1	7	0	.6320457
36.	36	1	7	0	.6320457
40.	40	1	7	0	.6320457
42.	42	1	7	0	.6320457
46.	46	1	7	0	.6320457
48.	48	1	7	0	.6320457
61.	61	1	7	0	.6320457
67.	67	1	7	0	.6320457
68.	68	1	7	0	.6320457
70.	70	1	7	0	.6320457
71.	71	1	7	0	.6320457
74.	74	1	7	0	.6320457
80.	80	1	7	0	.6320457
81.	81	1	7	0	.6320457
83.	83	1	7	0	.6320457