

Exercice3: Régression[4pts]

1. La régression linéaire permet d'étudier les individus atypiques. Décrire les différents cas possibles d'anormalité. [1pt]  
Atypique aberrant: valeur inhabituelle sur une variable ou combinaison de variables  
Influent: pèse de manière exagérée dans la régression  
Atypique régression: cible très mal expliquée, erreur très élevée
2. Quelles sont les méthodes pas à pas pour la construction de modèles réduits? Donner brièvement l'idée de chacune. [1pt]  
méthode descendante ou élimination en arrière lorsqu'on élimine des variables  
-méthode ascendante ou sélection en avant lorsque on ajoute des variables  
-méthode mixteest une combinaison de ces deux méthodes
3. Expliciter une méthode pour l'obtention des coefficients estimateurs dans la régression linéaire.[1pt]

Exercice2 : On propose de construire un modèle permettant de prédire les ventes semestrielles d'un produit agroalimentaire via la régression linéaire multiple.  
Les variables explicatives sont: MT: Besoin du marché,RG: Remises aux grossistes,PRIX: Prix du produit, BR: Budget de Recherche, INV: Investissement, PUB: Publicité, FV = Frais de ventes, TPUB = Total budget publicité de la branche.  
Les résultats du modèle sont affichés ci-dessus.

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	3129.231	641.355	4.879	.000
	MT	4.423	1.588	2.785	.009
	RG	1.676	3.291	.509	.614
	PRIX	-13.526	8.305	-1.629	.114
	BR	-3.410	6.569	-.519	.608
	INV	1.924	.778	2.474	.019
	PUB	8.547	1.826	4.679	.000
	FV	1.497	2.771	.540	.593
	TPUB	-2.15E-02	.401	-.054	.958

a. Dependent Variable: VENTES

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.898 <sup>a</sup>	.806	.752	256.29

1. En étudiant individuellement les coefficients liés à chaque variable explicative, déterminer les variables significativement pertinentes dans l'augmentation des ventes semestrielles du produit. [1pt]  
PUB, MT ont les plus faibles valeurs significatives donc elles sont les plus pertinentes.
2. En se basant sur les résultats des coefficients de détermination, conclure sur la qualité du modèle construit. [1pt]  
R² = 0.806 donc il s'agit d'un bon modèle.
3. On propose de sélectionner les variables pertinentes du modèle en utilisant la méthode de sélection pas à pas descendante (Backward). Expliquer brièvement les étapes de cette méthode. [2pts]  
AIC, le critère de choix du modèle le plus performant : itération par itération on élimine la variable la moins pertinente.
4. Donner l'équation du modèle permettant de prédire les ventes semestrielles en fonction de toutes les variables explicatives. [2pts]

VENTES = 3129,231 + 4,423 MT + 1,676 RG - 13,526 PRIX – 3,410 BR +1,924 INV + 8,328 PUB + 1,497 FV – 0,00215 TPUB

Exercice 3: Régression Linéaire Multiple[5pts]

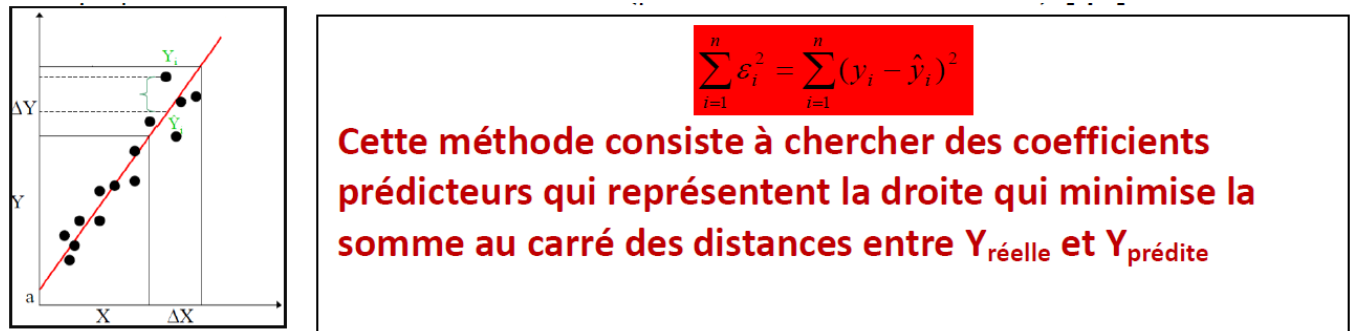
On propose de construire le meilleur modèle permettant de prédire une variable Y via la régression linéaire multiple.  
Les variables explicatives sont:X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub> et X<sub>4</sub>.

Les résultats des modèles global et réduit sont affichés ci-dessous.

Modèle Global	Coefficients
Constante	10
X <sub>1</sub>	1
X <sub>2</sub>	1.5
X <sub>3</sub>	2.5
X <sub>4</sub>	5

Modèle Réduit	Coefficients
Constante	15
X <sub>3</sub>	3
X <sub>4</sub>	1

1. Expliquer la méthode des **moindres carrées** (possibilité d'utiliser un schéma)



2. Décrire l'**utilité** de la régression linéaire. [2pts]

**La régression linéaire permet :**

- de prédire une variable décisionnelle quantitative via un modèle global ou un modèle avec sélection de variables pertinentes (modèle réduit via le critère AIC)
- d'avoir une vision sur l'ordre de pertinence des variables prédictives à l'aide des valeurs des coefficients prédictors
- de détecter les anomalies, les individus atypiques ou aberrants

3. Donner la **formule** du modèle global et celle du modèle réduit. [1pt]

$$Y_{global} = 10 + X_1 + 1.5 X_2 + 2.5 X_3 + 5 X_4$$

$$Y_{réduit} = 15 + 3 X_3 + X_4$$

4. Calculer les **prédictions** d'un individu I ( $X_1=1, X_2=1, X_3=1, X_4=2$ ).

4. Sachant que la valeur réelle de la variable décisionnelle de l'individu I est  $Y_{réelle} = 20$ , Comparer les deux prédictions obtenues et **Conclure**. [0.5pt]

$$Y_{réduit} = Y_{réelle}$$

Donc on peut constater que prédire Y à partir des valeurs de  $X_3$  et  $X_4$  est mieux que prédire à l'aide de toutes les variables prédictives // le modèle réduit mieux que le modèle global