

Directives pédagogiques :

- Trois pages de documents sont autorisées.
- Il sera tenu compte de la clarté des raisonnements.

Toute tentative de fraude sera sanctionnée par la note zéro.

Lisez attentivement les recommandations données à la fin de l'épreuve.

Nom :

Prénom :

Partie I : Echantillonnage (prière de répondre dans une feuille séparée des autres parties)

Un agent d'assurance voudrait estimer la proportion p d'employés, dans une certaine entreprise, qui possèdent au moins un véhicule. Pour cela, l'agent d'assurance décide de recourir à l'étude d'un échantillon. Se basant sur les fiches de salaires, il a décidé de répartir les 7500 employés de l'entreprise en trois strates:

- Strate 1 : Constituée des bas salaires,
- Strate 2 : Constituée des salaires moyens, et
- Strate 3 : Constituée des haut salaires.

Le tableau ci-après contient les résultats de l'enquête menée par l'agent d'assurance :

	$h=1$	$h=2$	$h=3$
N_h	3500	2000	2000
n_h	500	300	200
p_h	0.13	0.45	0.50

Où, pour $h=1, 2$ ou 3

N_h = la taille de la Strate h .

n_h = la taille de l'échantillon tiré de la Strate h (échantillonnage aléatoire simple),

p_h = la proportion d'employés, dans l'échantillon tiré de la Strate h , possédants au moins un véhicule.

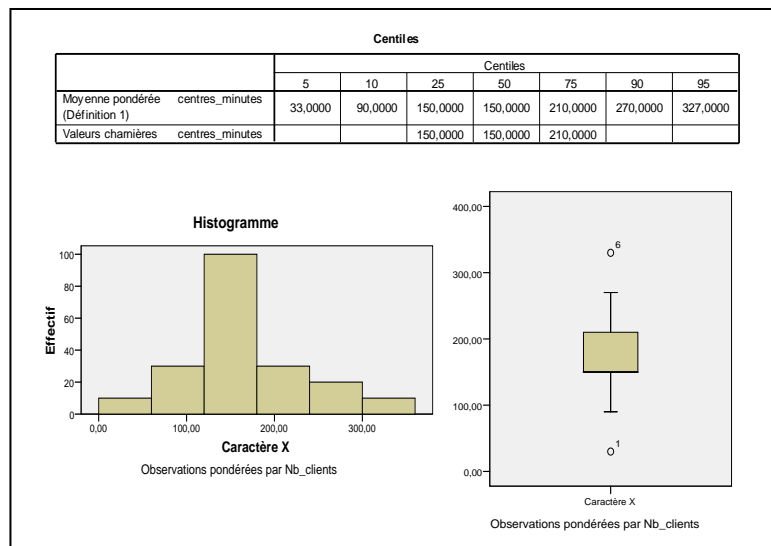
1. Citez deux raisons justifiant le recours de l'agent d'assurance à l'étude d'un échantillon.
2. Pensez-vous qu'un échantillonnage stratifié est approprié dans cette situation ? Justifiez votre réponse.
3. En utilisant les résultats ci-dessus, donner un estimateur sans biais pour p .
4. Calculer la variance de cet estimateur.
5. Donner un intervalle de confiance de niveau 95% pour p .

Partie II : Statistique Descriptive

La compagnie INTER-WEB compile le nombre de minutes de connexion de chacun de ses 200 clients au cours du son premier mois d'existence. On regroupe les observations de la façon suivante :

Nombre de minutes	Nombre de clients
$[0, 60[$	10
$[60, 120[$	30
$[120, 180[$	100
$[180, 240[$	30
$[240, 300[$	20
$[300, 360]$	10

- a) Identifier la population d'analyse :
.....
- b) Quel est le caractère étudié ici?Notons par X ce caractère.
- c) De quel type de caractère s'agit-il?
.....
- d) Peut-on calculer la moyenne de ce caractère ?
Si Oui, donner la formule : ainsi que la valeur
.....
- e) Donnez la valeur du mode de ce caractère :
- f) Donnez la valeur de la médiane de ce caractère : ainsi que la l'étendu :
.....
- g) Quel est le nb de clients dont le nb de connexion est au moins 270 minutes ?
.....
- h) Quelle devrait être la tarification (en Dh/**heure** de connexion) par client si la compagnie estime le revenu moyen des ventes de ce premier mois à 30 000 Dh ?
.....
.....
- i) Quelle serait la nouvelle estimation du revenu moyen si la compagnie introduit des frais de maintenance de 4Dh par mois pour chaque client ?



Partie 3 : Régression linéaire (cf. sorties correspondantes dans l'annexe) **les retourner avec vos feuilles de réponses commentées et avec une légende.**

3.1 Question de cours : décrire les différentes étapes de mise en œuvre d'une régression linéaire.

Application : On s'intéresse à la consommation d'essence pour 31 véhicules (en Suisse). On souhaite déterminer le critère qui a un impact important sur l'augmentation de la consommation parmi le prix, le poids, la puissance et cylindrée.

Soit Y : la variable à expliquer : Consommation (L/100 Km), et X_1 : Prix, X_2 : Cylindrée, X_3 : Puissance et X_4 : Poids les variables prédictives.

3.2 Existe-t-il une relation entre la consommation et chacune des variables ? OUI ☐ NON ☐
Justifiez :

3.3 Est-ce que les variables X_1, \dots, X_4 sont indépendantes ? OUI ☐ NON ☐
Justifiez :

Que concluez vous ?

3.4 Quelle que soit la réponse précédente, on décide d'effectuer une régression multiple entre Y et X_i ($i=1, \dots, 4$).
Quelle méthode a été utilisée ?

3.5 Est-ce que le modèle est de bonne qualité ? justifiez
.....

3.6 Donnez l'expression de la liaison qui existe entre Y et X_i
.....
.....
.....

3.7 Est-ce que toutes les variables sont importantes dans l'explication de la consommation ? OUI ☐ NON ☐
Justifiez en considérant la significativité à 5% :
.....
.....

3.8 D'après vous, qu'est ce qui influence le plus la consommation d'essence d'une voiture ?
.....
.....

3.9 En analysant les résidus, pensez vous que le modèle ajuste bien les données ?
.....
.....

3.10 Existe -il de l'information non expliquée ?
.....
.....

3.11 En appliquant la méthode pas à pas, quel modèle retenez vous ? justifiez
.....
.....
.....

3.12 Donnez son expression en justifiant la significativité des variables à 5%.

.....

.....

.....

.....

Partie 4 : Analyse factorielle

(cf. sorties correspondantes dans l'annexe) les retourner avec vos feuilles de réponses commentées et avec une légende.

Les questions suivantes concernent une analyse factorielle effectuée sur les mêmes données que celles de la partie 3. Elle concerne l'élaboration d'une typologie de voitures selon les caractéristiques le prix, le poids, la puissance et cylindrée.

Question 4.1 L'analyse factorielle est-elle pertinente ? justifier votre réponse en utilisant tous les critères possibles.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 4.2 : Quelles sont les variables qui vont le plus contribuer à la construction des facteurs ? justifier

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 4.3 L'extraction de facteurs a été faite à l'aide de l'Analyse en composantes principales

a) Quels nombre de facteurs retenez vous ? justifiez le

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

b) Est-il nécessaire de faire une rotation varimax ?

Justifiez :

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 4.4 Donnez une interprétation des facteurs si l'on décide de ne pas faire la rotation et donnez l'expression mathématique de chaque facteur.

.....

```

.....
.....
.....
.....

```

.....

.....

Les facteurs ainsi construits ont la propriété intéressante d'indépendance et de la normalité. On peut donc expliquer la consommation par ces facteurs. D'après vous, la consommation dépend de quel facteur ?

.....

.....

.....

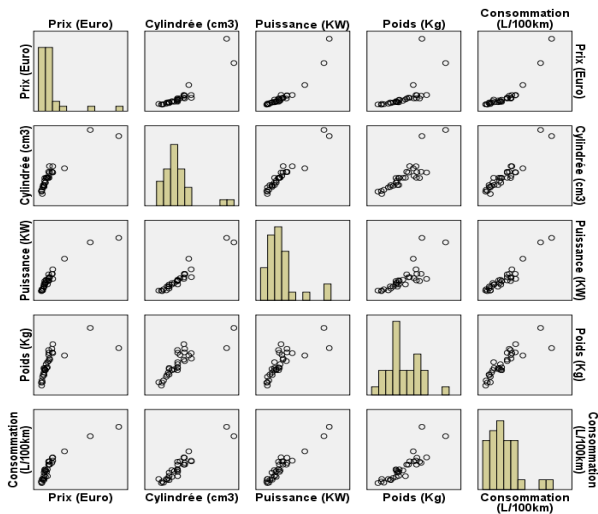
.....

.....

.....

.....

Sorties de la partie 3 : régression multiple



		Corrélations				
		Prix (Euro)	Cylindrée (cm3)	Puissance (KW)	Poids (Kg)	Consommation (L/100km)
Prix (Euro)	Corrélation de Pearson	1	,898 ^{**}	,935 ^{**}	,635 ^{**}	,891 ^{**}
	Sig. (bilatérale)		,000	,000	,000	,000
	N	31	31	31	31	31
Cylindrée (cm3)	Corrélation de Pearson	,898 ^{**}	1	,963 ^{**}	,838 ^{**}	,941 ^{**}
	Sig. (bilatérale)	,000		,000	,000	,000
	N	31	31	31	31	31
Puissance (KW)	Corrélation de Pearson	,935 ^{**}	,963 ^{**}	1	,780 ^{**}	,953 ^{**}
	Sig. (bilatérale)	,000	,000		,000	,000
	N	31	31	31	31	31
Poids (Kg)	Corrélation de Pearson	,635 ^{**}	,838 ^{**}	,780 ^{**}	1	,864 ^{**}
	Sig. (bilatérale)	,000	,000	,000		,000
	N	31	31	31	31	31
Consommation (L/100km)	Corrélation de Pearson	,891 ^{**}	,941 ^{**}	,953 ^{**}	,864 ^{**}	1
	Sig. (bilatérale)	,000	,000	,000	,000	
	N	31	31	31	31	31

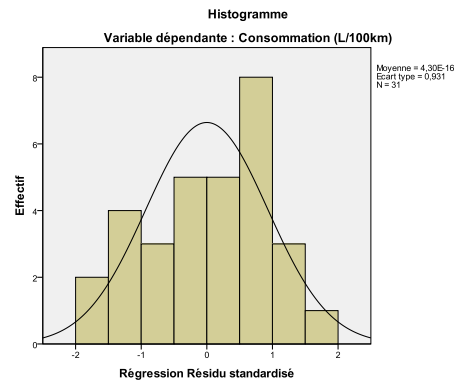
** La corrélation est significative au niveau 0.01 (bilatéral).

Variables introduites/supprimées^b

Modèle	Variables introduites	Variables supprimées	Méthode
1	Poids (Kg), Prix (Euro), Cylindrée (cm3), Puissance (KW) ^a	.	Entrée

a. Toutes variables requises saisies.

b. Variable dépendante : Consommation (L/100km)



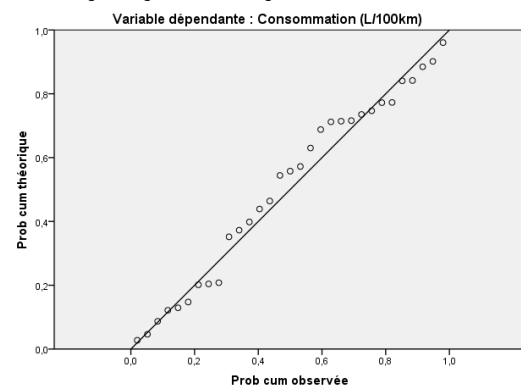
Récapitulatif des modèles^b

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Durbin-Watson
1	,977 ^a	,955	,948	,8172	2,180

a. Valeurs prédites : (constantes), Poids (Kg), Prix (Euro), Cylindrée (cm3), Puissance (KW)

b. Variable dépendante : Consommation (L/100km)

Diagramme gaussien P-P de régression de Résidu standardisé



Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Sig.
		A	Erreur standard	Bêta		
1	(Constante)	2,456	,627		3,919	,001
	Prix (Euro)	2,042E-5	,000	,315	2,339	,027
	Cylindrée (cm3)	-,001	,001	-,162	-,871	,392
	Puissance (KW)	,025	,010	,486	2,501	,019
	Poids (Kg)	,004	,001	,421	4,734	,000

a. Variable dépendante : Consommation (L/100km)

Récapitulatif des modèles^d

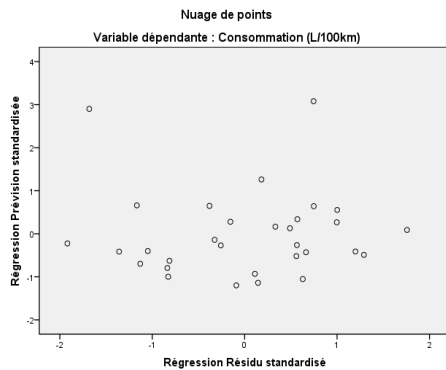
Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Durbin-Watson
1	,953 ^a	,907	,904	1,1041	
2	,972 ^b	,945	,941	,8676	
3	,976 ^c	,953	,948	,8136	2,262

a. Valeurs prédites : (constantes), Puissance (KW)

b. Valeurs prédites : (constantes), Puissance (KW), Poids (Kg)

c. Valeurs prédites : (constantes), Puissance (KW), Poids (Kg), Prix (Euro)

d. Variable dépendante : Consommation (L/100km)



Coefficients ^a					
Modèle	Coefficients non standardisés		Coefficients standardisés	t	Sig.
	A	Erreur standard	Bêta		
1 (Constante)	5,194	,345		15,057	,000
Puissance (KW)	,049	,003	,953	16,867	,000
2 (Constante)	2,561	,663		3,864	,001
Puissance (KW)	,037	,004	,712	10,041	,000
Poids (Kg)	,003	,001	,309	4,355	,000
3 (Constante)	2,499	,622		4,017	,000
Puissance (KW)	,020	,008	,391	2,441	,021
Poids (Kg)	,004	,001	,377	5,139	,000
Prix(Euro)	1,852E-5	,000	,286	2,201	,037

a. Variable dépendante : Consommation (L/100km)

Sortie de la partie 4 : analyse factorielle

Indice KMO et test de Bartlett		
Mesure de précision de l'échantillonnage de Kaiser-Meyer-Olkin.		,770
Test de sphéricité de Bartlett	Khi-deux approximé	172,378
	ddl	6
	Signification de Bartlett	,000

Statistiques descriptives

	Moyenne	Ecart-type	n analyse
Prix (Euro)	43755,65	55065,258	31
Cylindrée (cm3)	2093,68	1152,357	31
Puissance (KW)	97,10	69,341	31
Poids (Kg)	1256,13	360,753	31

Matrices anti-images

		Prix (Euro)	Cylindrée (cm3)	Puissance (KW)	Poids (Kg)
Covariance anti-images	Prix (Euro)	,096	-,017	-,039	,070
	Cylindrée (cm3)	-,017	,051	-,027	-,059
	Puissance (KW)	-,039	-,027	,046	-,016
	Poids (Kg)	,070	-,059	-,016	,222
Corrélation anti-images	Prix (Euro)	,764 ^a	-,250	-,590	,481
	Cylindrée (cm3)	-,250	,780 ^a	-,559	-,557
	Puissance (KW)	-,590	-,559	,778 ^a	-,157
	Poids (Kg)	,481	-,557	-,157	,752 ^a

a. Mesure de précision de l'échantillonnage

Variance totale expliquée

Composante	Valeurs propres initiales			Extraction Sommes des carrés des facteurs retenus			Somme des carrés des facteurs retenus pour la rotation		
	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	3,534	88,346	88,346	3,534	88,346	88,346	2,235	55,870	55,870
2	,386	9,661	98,006	,386	9,661	98,006	1,555	38,866	94,736
3	,050	1,239	99,245	,050	1,239	99,245	,142	3,544	98,280
4	,030	,755	100,000	,030	,755	100,000	,069	1,720	100,000

Méthode d'extraction : Analyse en composantes principales.

Qualité de représentation à 2 facteurs

	Initial	Extraction
Prix (Euro)	1,000	,978
Cylindrée (cm3)	1,000	,972
Puissance (KW)	1,000	,977
Poids (Kg)	1,000	,993

Méthode d'extraction : Analyse en composantes principales.

Corrélations reproduites

		Prix (Euro)	Cylindrée (cm3)	Puissance (KW)	Poids (Kg)
Corrélation reproduite	Prix (Euro)	,978 ^a	,913	,950	,624
	Cylindrée (cm3)	,913	,972 ^a	,967	,849
	Puissance (KW)	,950	,967	,977 ^a	,785
	Poids (Kg)	,624	,849	,785	,993 ^a
Résidu ^b	Prix (Euro)		-,015	-,015	,011
	Cylindrée (cm3)	-,015		-,004	-,011
	Puissance (KW)	-,015	-,004		-,005
	Poids (Kg)	,011	-,011	-,005	

Méthode d'extraction : Analyse en composantes principales.

a. Qualité de représentation reproduite

b. Les résidus sont calculés entre la covariance observée et la covariance reproduite. Il y a 0 (0,0%) résidus non redondants avec des valeurs absolues supérieures à 0,05.

Matrice des composantes^a

	Composante	
	1	2
Cylindrée (cm3)	,986	
Puissance (KW)	,981	-,118
Prix (Euro)	,927	-,345
Poids (Kg)	,860	,503

Méthode d'extraction : Analyse en composantes principales.

a. 2 composantes extraites.

Matrice des composantes après rotation^a

	Composante	
	1	2
Prix (Euro)	,942	,300
Puissance (KW)	,845	,513
Cylindrée (cm3)	,774	,810
Poids (Kg)	,366	,927

Méthode d'extraction : Analyse en composantes principales.

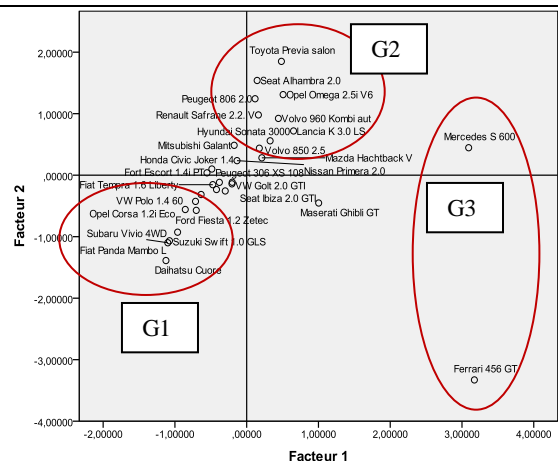
Méthode de rotation : Varimax avec normalisation de Kaiser.

a. La rotation a convergé en 3 itérations.

Matrice des coefficients des coordonnées des composantes

	Composante	
	1	2
Prix (Euro)	,262	-,893
Cylindrée (cm3)	,279	,007
Puissance (KW)	,278	-,305
Poids (Kg)	,243	1,303

Méthode d'extraction : Analyse en composantes principales.
Scores composante.



Corrélations

		Facteur 1	Facteur 2	Consommation (L/100km)
Facteur 1	Corrélation de Pearson	1	,000	,971 ^{**}
	Sig. (bilatérale)		1,000	,000
	N	31	31	31
Facteur 2	Corrélation de Pearson	,000	1	,045
	Sig. (bilatérale)	1,000		,809
	N	31	31	31
Consommation (L/100km)	Corrélation de Pearson	,971 ^{**}	,045	1
	Sig. (bilatérale)	,000	,809	
	N	31	31	31

** . La corrélation est significative au niveau 0.01 (bilatéral).

Bonne Chance