

Séminaire SPSS

Responsable : S. ZKIM

TP 1

Objectifs : être en mesure de **créer une base de données** sur SPSS. Découvrir l'éditeur des données ; affichage des variables et des données ; ...

La direction centrale d'une société de télécommunication veut analyser l'activité de ses 15 agences. Elle dispose pour cela de plusieurs indicateurs (qualitatifs et quantitatifs) concernant un mois d'activité:

Descriptif des indicateurs :

code (code l'agence); **Effectif** (nombre d'employés de l'agence); **ville** (ville d'implantation, Rabat ou casa); **Horcont** (horaire continue : oui (=1) ou non (=0)); **emplace** (lieu d'emplacement : quartier(=1), place commerciale(=2), grande avenue (=3)); **CA** (chiffre d'affaire du dernier mois, en milliers de DH), **Abonn** (nombre d'abonnements); **CartePP** (carte prépayées vendus); **visit** (nombre de visiteurs enregistrés); **sav** (service après vente assuré : oui (=1), non (=0))

Les données enregistrées sont les suivantes :

code	effectif	ville	horcont	emplace	abonn	cartePP	sav	CA
A1	3	casa	1	1	40	210	1	117,54
A2	1	casa	0	1	25	170	2	74
A3	5	rabat	0	2	35	195	1	121
A4	5	casa	0	3	24	245	1	79,65
A5	2	rabat	1	2	19	187	9	75
A6	1	rabat	0	1	16	125	0	58
A7	9	casa	1	2	44	366	1	135
A8	4	casa	1	3	36	288	0	115
A9	5	rabat	1	3	33	999	1	101
A10	6	casa	1	3	29	255	1	86,1
A11	2	rabat	0	2	22	209	0	67
A12	2	casa	0	2	18	188	0	68
A13	1	casa	0	1	09	133	0	60,88
A14	8	casa	1	2	55	233	1	155
A15	3	rabat	1	2	35	170	0	81,32

La valeur 9 désigne le code des valeurs manquantes pour la variable sav.

La valeur 999 désigne le code des valeurs manquantes pour la variable cartePP

1. Démarrer SPSS et créer un nouveau fichier contenant les données. On précisera toutes les déclarations nécessaires dans la feuille de l'affichage des variables.
2. La variable 'visit' a été oubliée et devrait être saisie entre les variables sav et CA.
3. Procéder à la correction du chiffre d'affaire de l'agence A11 en remplaçant la valeur 67 par 69,55.
4. Enregistrer le fichier obtenu sous le nom 'TP1.sav'.
5. Afficher dans une fenêtre résultat toutes les informations sur ce fichier. Enregistrer le fichier résultat sous le nom 'TP1RES.spo'.

Séminaire SPSS

Responsable :S.ZKIM

TP 2

Objectifs : Préparation des données, réaliser des sélections, construire des échantillons aléatoires, Tris composés d'observations, **premiers traitements**, etc

1. Ouvrir le fichier 'Employes.sav' et lister les informations sur les variables (étiquettes, étiquettes des valeurs, valeurs manquantes...).
2. En utilisant un tri suivant la variable 'salaire courant', déterminer le plus grand salaire et le plus petit salaire des employés.
3. En utilisant un tri suivant les variables sexe et salaire courant, déterminer le plus grand salaire et le plus petit salaire des hommes et des femmes.
4. Déterminer le plus grand salaire des femmes managers.
5. Sélectionner les employés de sexe féminin. Donner leur nombre.
6. Sélectionner les hommes managers. Puis, à l'aide d'un tri, déterminer le plus grand salaire dans cette catégorie.
7. Combien y a-t-il de femmes dont le salaire ne dépasse pas 25 000 ? (utiliser une sélection).
8. Choisir un échantillon aléatoire de 30 employés.
9. Reprendre l'ensemble des employés, choisir à nouveau un échantillon aléatoire contenant 20% des employés. Compter le nombre d'individus choisis.
10. Déterminer, sans utiliser le menu Statistiques, le pourcentage des femmes managers et qui ont un salaire d'embauche supérieur à 20 000.

Séminaire SPSS

Responsable : S. ZKIM

TP 3

Objectifs : Sélection et échantillonnage, calcul de nouvelles variables, recodage de variables , Discrétisation, recodage automatique, etc .

1/Ouvrir le fichier '**Employes.sav**' et lister les informations sur les variables (étiquettes, étiquettes des valeurs, valeurs manquantes...).

2/Calculer une nouvelle variable dénommée 'Progression' , égale à la différence entre le salaire courant et le salaire d'embauche.

3/Calculer une nouvelle variable dénommée 'Rendement' , égale au rapport entre le rendement et l'expérience passée. (attention à la division par 0).

4/Calculer la variable 'age des employés' à partir des dates de naissances. (utiliser la fonction Xdate.Year disponible dans la liste des fonctions de SPSS).

5/Recoder la variable 'salaire courant' en 3 classes :

- a. classe des petits salaires, codée par 1, contenant les salaires < 20 000
- b. classe des salaires moyens, codée par 2, contenant les salaires compris entre 20 000 et 40 000.
- c. Classe des grands salaires, codée par 3, contenant les salaires dépassant 40 000 .

La variable recodée prendra le nom 'Csalaire'. (ON n'oubliera pas de préciser les étiquettes des valeurs de Csalaire).

6/Refaire la question 5 pour la variable Age . On vous laisse le soin de choisir quatre classes raisonnables...

7/Utiliser un recodage automatique pour coder la variable nominale 'Sexe' en variable numérique

8/Utiliser une discrétisation pour obtenir un recodage rapide en 4 classes de la variable age. Comparer les classes obtenues avec celles de la question 6.

9/Calculer une variable 'prime' égale au cinquième du salaire courant pour les managers, au sixième du salaire courant pour les employés de bureau et au huitième du salaire courant pour les agents de sécurité. (utiliser la condition Si).

Séminaire SPSS

Responsable : S. ZKIM

TRAVAIL PRATIQUE

DIRECTIVES IMPORTANTES

D1) Ce travail peut être fait seul ou en équipe de deux au maximum.

D2) Date de remise : Semaine prochaine.

D3) **À remettre** : En plus de remettre vos documents sous la forme d'une copie papier, vous allez également nous rendre sur " disquette propre"; les fichiers «NIV5G3.SAV», «NIV5G3.SPO», «ecole casa.Sav» et «TP3B.spo».

D4) Évitez de produire des résultats inutiles ou superflus. Par exemple, lorsqu'on demande uniquement la moyenne, vous ne devez pas calculer toutes les autres statistiques ni produire la table de fréquences ni produire de graphiques.

D5) Supposez toujours que les résultats pourraient être facilement lus par un non initié au logiciel. Pour ce faire, des étiquettes (labels) doivent être associées aux noms et aux valeurs des variables lorsque nécessaire. N'oubliez pas d'associer également des étiquettes aux valeurs manquantes.

Dans la partie A, tous ces textes explicatifs doivent être ajoutés via la fenêtre «Affichage des variables»

D6) Assurez-vous de bien «aérer» le programme généré dans la partie B en y insérant des lignes blanches et des commentaires appropriés.

PARTIE A

Cette première partie du TP consiste à utiliser SPSS afin de créer une banque de données provenant d'une enquête réalisée auprès d'une classe d'élèves du niveau 5-ième année primaire d'une école à casablanca. Vous trouverez à la page suivante un imprimé de la banque de données ainsi qu'une brève description de son contenu. Malheureusement, les données n'ont pas encore été saisies à l'ordinateur, vous devrez donc le faire vous-même directement dans SPSS.

[Voici les étapes à suivre pour réaliser la partie A:](#)

1. Démarrez SPSS et définissez la structure de la banque de données (nom des variables ainsi que leurs types, textes explicatifs et valeurs manquantes) telle qu'elle est décrite à la page suivante. N'oubliez pas d'enregistrer ensuite cette structure dans le fichier «NIV5G3.SAV» . Le nom «NIV5G3» est justifié puisqu'il s'agit du groupe 3 du niveau 5.
2. Entrez ensuite les données pour les 15 élèves puis enregistrez de nouveau la banque de données.
3. Corrigez l'âge de l'élève #12: Remplacez 16 par 10.
4. Une erreur s'est glissée dans la liste qui vous a été fournie car la situation parentale de l'élève #5 est invalide. Remplacez le 7 par 1.
5. Nous avons également oublié d'inclure une variable nommée «TRANSPOR» qui doit être insérée entre les variables «Numéro de l'élève» et «Sexe». Insérez cette variable avec les paramètres suivants:
(«Moyen de transport», 1=Pied, 2=Parent, 3=Autobus) et entrez les données 1,2,1,3,3,1,3,3,1,1,2,1,3,1,1 aux élèves #1 à #15 respectivement.
6. Enregistrez une dernière fois la banque de données puisque vous venez d'y apporter des corrections.

7. Avant d'imprimer le contenu de la fenêtre des données, assurez-vous que les largeurs des colonnes sont suffisantes pour que le nom de chaque variable soit bien visible. Imprimez ensuite le contenu de la banque de données.

8. Pour fin de correction, nous aimerions obtenir une description détaillée de la structure de la banque de données «NIV5G3.SAV» que vous venez de définir. Pour chacune des variables, nous voulons obtenir son nom, son type, ses valeurs manquantes ainsi que les textes explicatifs associés au nom de la variable et aux valeurs qu'elle peut prendre. Pour ce faire, utilisez la séquence : **Outils / Informations sur le fichier.**

9. Imprimez finalement cette description .

10. Quittez SPSS sans oublier d'enregistrer votre fenêtre de résultats dans le fichier «NIV5G3.SPO».

NIVEAU SCOLAIRE	GROUPE	NUMÉRO DE L'ÉLÈVE	SEXE	AGE	SITUATION PARENTALE	NOMBRE D'ENFANTS	DINER	SERVICE DE GARDE
5,00	3,00	1,00	F	9,0	1,00	6,00	1,00	,00
5,00	3,00	2,00	F	0	1,00	3,00	3,00	,00
5,00	3,00	3,00	F	10,	2,00	4,00	2,00	1,00
5,00	3,00	4,00	M	0	1,00	2,00	1,00	,00
5,00	3,00	5,00	M	9,0	7,00	1,00	1,00	1,00
5,00	3,00	6,00	F	0	3,00	1,00	3,00	,00
5,00	3,00	7,00	M	9,0	2,00	3,00	3,00	,00
5,00	3,00	8,00	F	0	3,00	3,00	2,00	1,00
5,00	3,00	9,00	F	10,	1,00	4,00	2,00	,00
5,00	3,00	10,00	M	0	.	2,00	1,00	,00
5,00	3,00	11,00	M	11,	2,00	1,00	1,00	,00
5,00	3,00	12,00	M	0	1,00	6,00	1,00	,00
5,00	3,00	13,00	F	10,	2,00	1,00	2,00	,00
5,00	3,00	14,00	M	0	1,00	,00	1,00	,00
5,00	3,00	15,00	F	,00	3,00	1,00	2,00	1,00
				9,0				
				0				
				9,0				
				0				
				9,0				

				0				
				16,				
				0				
				11,				
				0				
				12,				
				0				
				10,				
				0				

DESCRIPTION DE LA CODIFICATION DU CONTENU DE LA BANQUE DE DONNÉES

NIVEAU SCOLAIRE 0=Maternelle, 1=1^{ère} année, 2=2^{ème} année, ..., 6=6^{ème} année

GROUPE Numéro du groupe dans ce niveau scolaire, Valeurs entre 1 et 4.

NUMÉRO DE L'ÉLÈVE anonyme de l'élève dans sa classe

AGE Âge de l'élève au moment de la rentrée scolaire en septembre

SITUATION PARENTALE 1=Ensemble, 2=Séparé, 3=Divorcé.

NOMBRE D'ENFANTS : Nombre d'enfants dans la famille incluant le répondant

DINER Situation habituelle pour le dîner : 1=Mon lunch à l'école, 2=Repas chaud servi à l'école,

3=Je dîne à la maison

SERVICE DE GARDE : Vas-tu au service de garde de l'école à 15:30 après la classe?

0=Non, 1=Oui

N.B.: Une valeur de 0 pour les variables «AGE» et «NOMBRE D'ENFANTS» indique un refus de

répondre par l'élève et doit être considérée comme une valeur manquante.

PARTIE B

La deuxième partie du TP consiste à utiliser SPSS pour analyser une banque de données dont la structure est identique à celle que vous avez créée dans la partie

A. Cependant, cette banque de données contient beaucoup plus d'observations
(un groupe par niveau)

Voici les étapes à suivre pour réaliser la partie B:

1. Avant de commencer, Explorez la structure de cette banque de données en parcourant chaque variable dans la fenêtre d'information sur les variables.
2. Démarrez ensuite SPSS et ouvrez la banque de données : **ecole casa**
3. Utilisez SPSS pour répondre aux 10 questions suivantes.

N'oubliez pas d'enregistrer la fenêtre des résultats
(ecolecasa.SPO) lorsque vous quitterez le logiciel SPSS

!

Q1. Illustrez la distribution de fréquences de l'âge des élèves à l'aide d'un histogramme puis.

Q2. Combien y a-t-il d'élèves dont les parents sont divorcés?

Q3. Quel est le 80^{ième} percentile de l'âge et quelle en est la médiane?

Q4. Produisez une distribution de fréquences du niveau scolaire . Quel est le niveau scolaire le moins représenté dans cette école?

Q5. Déterminez l'âge du plus jeune élève, du plus vieux ainsi que l'âge moyen.

Q6. Illustrez la distribution de fréquences du sexe des élèves par un diagramme en bâtons .

Q7. Combien y a-t-il d'élèves dans la classe la plus peuplée? Calculez cette valeur sans produire la table détaillée des fréquences et sans consulter l'ensemble de la fenêtre des données. Vous devez donc faire un choix judicieux de la bonne variable et surtout de la bonne statistique.

Q8. Déterminez le moyen de transport le plus utilisé à cette école mais sans produire la table

détaillée des fréquences. (Indice: faites encore un choix judicieux de la bonne statistique)

Q9. Utilisez la procédure statistique «Descriptives» afin de calculer le pourcentage des élèves qui utilisent le service de garde après l'école? (Indice: le codage 0-1 de cette variable est très utile)

Q10. Utilisez la commande **Transformer/ Recoder / Création de variables...** afin de créer une nouvelle variable qui indiquera le cycle d'étude de l'élève. Le 1^{er} cycle s'étend de la maternelle à la 3^{ème} année tandis que le 2^{ème} cycle comprend les 4^{ème}, 5^{ème} et 6^{ème} années. N'oubliez pas d'ajouter des étiquettes à cette nouvelle variable. Produisez une distribution de fréquences de cette variable et déterminez combien d'élèves sont inscrits au 2^{ème} cycle?

--- Bon travail à tous ---

Séminaire SPSS

Responsable :S.ZKIM

TP 4

Objectifs : Sélection et échantillonnage, calcul de nouvelles variables, recodage de variables
Discretisation, recodage automatique, etc .

1/Ouvrir le fichier '**Lemonde95.sav**' et lister les informations sur les variables (étiquettes, étiquettes des valeurs, valeurs manquantes...).

2/Calculer une nouvelle variable dénommée 'espvie' ; espérance de vie, égale à la moyenne entre les variables 'espvieh' et 'espvief'.

3/Calculer une nouvelle variable dénommée 'Logpop' , égale au logarithme de 'nbpop'.

4/Recoder la variable 'nbpop' en 4 classes :

- d. classe des petits pays, codée par 1, contenant les pays dont la population est inférieure à 2 000000
- e. classe des moyensinf pays, codée par 2, contenant les pays dont la population est comprise entre 2 000000 et 10000000
- f. Classe des moyensup pays, codée par 3, contenant les pays dont la population est comprise entre 10 000000 et 100 000000
- g. Classe des grands pays, codée par 4, contenant les pays dont la population dépasse 100 000 000 .

5/Refaire la question 4 pour la variable 'espvief' . On vous laisse le soin de choisir quatre classes raisonnables...

6/Utiliser un recodage automatique pour recoder la variable nominale 'religion' en variable numérique.

7/Utiliser une discrétisation pour obtenir un recodage rapide en 4 classes de la variable espvief. Comparer les classes obtenues avec celles de la question 5.

Séminaire SPSS

Responsable : . S. ZKIM

TP 5

Objectifs : Premiers traitements statistiques, **Tri à plat**, graphiques simples, lecture et gestion des résultats, premières interprétations .

1/Ouvrir le fichier 'employes.sav' et lister les informations sur les variables (étiquettes, étiquettes des valeurs, valeurs manquantes...). Enregistrer la fenêtre résultats sous le nom 'RESTP5 .spo'.

2/En utilisant le menu *Analyses-Stat.descriptives-Frequences*, dresser un **tableau de fréquences** des variables sexe et statut. Illustrer ces deux variables par deux diagrammes en secteurs.

3/Produire pour chacune des deux variables précédentes un diagramme en **bâtons** en pourcentages. (Utiliser deux manières différentes)

4/Produire pour chacune des deux variables précédentes un diagramme en **secteurs** avec des pourcentages. (Utiliser deux manières différentes)

5/Personnaliser les graphiques précédents (couleurs, tailles et titres des diagrammes, pourcentages).

6/Déterminer **les indicateurs** classiques de description pour la variable Salaire (utiliser le menu 'fréquences'). Produire aussi un histogramme pour cette variable.

7/Produire une **boîte à moustaches** pour la variable salaire (menu Graphes). Interpréter la. Masquer les valeurs extrêmes de la boîte .

8/Refaire les questions 5 et 6 pour la variable expérience passée. Interpréter les résultats.

9/Utiliser le menu 'graphes' pour produire **l' histogramme** du salaire. Comparer avec la question 5.

10/Utiliser le menu 'ANALYSES-stat descrip-**caractéristiques**' pour retrouver les indicateurs statistiques du salaire de la question 5.

11. Utiliser le menu ANALYSES-stat descrip-**explorer**' pour explorer d'un seul coup la variable salaire (histogramme, boîte à moustaches, statistiques) .

12. Décrire la variable exp-passée (statistiques et graphiques).

13. Nettoyer votre fichier des résultats, donner des titres aux graphiques et aux tableaux, enregistrer votre **fichier résultat**.

Séminaire SPSS

Responsable : S.ZKIM

TP 6

Objectifs : Premiers traitements statistiques, **Tri à plat**, graphiques simples, lecture et gestion des résultats, premières interprétations .

1/Ouvrir le fichier 'Lemonde95' et lister les informations sur les variables (étiquettes, étiquettes des valeurs, valeurs manquantes...). Enregistrer la fenêtre résultats sous le nom 'RESTP6 .spo'.

2/Dresser un tableau de fréquences des variables région et religion. Illustrer ces deux variables par deux diagrammes. Commenter.

3/Produire pour chacune des deux variables précédentes un diagramme en secteurs avec des pourcentages. Personnaliser les graphiques précédents (couleurs, tailles et titres des diagrammes, pourcentages).

4/Déterminer **les indicateurs** classiques de description pour la variable Nbpop (utiliser le menu 'fréquences'). Produire aussi un histogramme pour cette variable. Commenter.

5/Produire une **boîte à moustaches** pour la variable espvief (menu Graphes). Interpréter la.

7/Refaire les questions 4 et 5 pour la variable espviah. Interpréter les résultats.

8/Utiliser le menu 'graphes' pour produire **l'histogramme** de espviah et espvief.

9/Explorer la variable literat (pourcentage de lettrés) (histogramme, boîte à moustaches, statistiques). Interpréter les résultats obtenus.

10/Produire un histogramme de la variable Pdb (produit domestique brut par tête). Commenter le. Donner ensuite les statistiques de base de cette variable. Interpréter.

11/Donner les indicateurs de dispersion de la variable Nbpop. Donner un intervalle contenant la taille de population de 90% des pays de l'échantillon.

12/Comparer sur un même graphique les variables %hommes lecteurs et %femmes lectrices. Peut-on dire qu'elles sont identiquement distribuées ?

13/Déterminer les indicateurs de symétrie de la variable fertilité. Quel indicateur est le meilleur pour résumer cette variable ?

TP 7

On a fait aux vingt sept élèves d'une classe de collège un test de lecture rapide , la note **R** évalue la rapidité de lecture et la note **comp** indique la compréhension du texte mesurée par un pourcentage de bonnes réponses à un ensemble de questions posées après lecture . Les enseignants ont en outre noté les trois variables :sexe , niveau(A=très bien, B=bien ,C=MOYEN , D=médiocre ,E=faible) , Age (0 c'est l'age normal pour la classe , 1 c'est un an de retard , -1 c'est un an d'avance) et 2 ans de retard. Les données observées sont classées dans le tableau ci-dessous :

- 1) précisez s'il s'agit d'un échantillon ou d'une population . Donner le type de chacune des variables enregistrées

Tri à plat

- 2) illustrer graphiquement les différentes variables qualitatives du problème
- 3) construire des boites à moustaches pour les variables quantitatives
- 4) présenter dans un tableau récapitulatif les différentes statistiques pertinentes pour les variables du problème
- 5) à l'aide de ce qui précède , étudier la dispersion et la symétrie des variables quantitatives

Tri croisé

- 6) donner un graphique qui illustre la variable niveau par sexe
- 7) peut on dire que le niveau dépend de la variable sexe , justifiez votre réponse
- 8) produire un graphique qui illustre la liaison entre la compréhension et la rapidité et commentez le

9) quel type de graphique peut on utiliser pour l'étude de liaison entre la variable compréhension et la variable niveau ? entre la variable âge et la variable rapidité

Elève	Rapidité	Compréh	Sexe	Niveau	Age
1	19	70	F	C	0
2	47	82	F	D	0
3	18	70	F	A	0
4	17	76	F	A	0
5	32	45	F	B	1
6	33	60	G	C	-1
7	27	45	G	E	2
8	47	72	F	B	0
9	45	76	G	A	0
10	41	89	F	A	-1
11	28	26	G	C	-1
12	45	76	F	A	0
13	19	66	F	B	0
14	23	35	G	D	2
15	22	46	G	D	2
16	23	23	G	C	0
17	24	50	G	D	0
18	25	60	G	B	0
19	41	82	F	B	0
20	21	50	G	E	0
21	21	92	G	E	2
22	24	70	F	C	0
23	21	75	G	C	1
24	15	45	G	D	0
25	29	81	F	A	0
26	20	43	G	C	0
27	26	72	G	B	0

Séminaire SPSS

Responsable :S.ZKIM

TP 8

Objectifs : Etude des **liaisons**, **Tri croisé**, graphiques croisés, lecture et interprétation des indicateurs de **corrélations**.

- 1/Ouvrir le fichier 'Lemonde95' et lister les informations sur les variables (étiquettes, étiquettes des valeurs, valeurs manquantes).
- 2/Produire un nuage de points (*diagramme de dispersion*) entre les variables *espvieh* et *espvief*. Commenter.
- 3/A partir du menu *Analyses-Corrélation*, calculer la corrélation entre les variables *espvieh* et *espvief*. Cette corrélation est-elle significative à 95% de confiance ?
- 4/Refaire les questions 2 et 3 pour les variables *Nb d'individus/km2* et *Cas de sida*.
- 5/Produire, en utilisant le menu *Graphes-boite à moustaches*, la répartition de la variable *lettrés* par rapport à la *religion*. Peut-on dire que ces deux variables sont corrélées ?
- 6/même question entre les variables *région* et *Nb d'individus/km2*
- 7/En utilisant le menu *Analyses-Statistiques-descrip-tabcroisé*, produire un tableau d'effectifs entre les variables *région* et *religion*.
- 9/Produire les pourcentages des religions selon les régions. En déduire la part des pays musulmans dans les pays d'Afrique.
- 10/Produire les pourcentages des régions selon les religions. En déduire la part des pays d'Afrique dans les pays musulmans.
- 11/Utiliser un diagramme interactif pour représenter les religions selon les régions. Peut-on dire que ces deux variables sont liées ?
- 12/Répondre à la question précédente en utilisant un test du χ^2 (*tab.croisé – statistiques*). La liaison est-elle significative ?
- 13/Étudier la liaison entre les variables *climat* et *religion*. Utiliser des graphiques et le test du χ^2 pour répondre. Interpréter le sens de la liaison.

Séminaire SPSS

Responsable :S.ZKIM

TP 9

Une enquête a été réalisée sur un échantillon de 16 PME de différents secteurs , l'objectif de l'enquête est de décrire et de détecter les liaisons entre les indicateurs classiques sur lesquels porte l'étude (chiffre d'affaire, secteur d'activité etc...)

Voici la liste des déclarations des variables retenues pour l'analyse.

Descriptif des valeurs

Code (code de l'entreprise), secteur(secteur d'activité :commerce=1, industrie=2, service=3), ville(ville d'activité : rabat=1,casa=2,autres=3) ,effectif(nombre d'employés),CA(chiffre d'affaires), réseau (réseau intranet : oui=1,non=0)

Les données enregistrées sont regroupées au tableau ci-dessous

La valeur 9 désigne une valeur manquante pour la variable réseau

La valeur 999 désigne une valeur manquante pour la variable chiffre d'affaires

Code	Secteur	Ville	Effectif	CA	Réseaux
1	1	1	10	315	1
2	1	3	6	230	0
3	3	3	18	450	0
4	2	1	8	250	0
5	1	2	10	190	1
6	3	3	12	430	1
7	2	1	12	999	9
8	2	2	9	445	0
9	3	1	14	285	0
10	1	2	8	520	0
11	2	2	22	500	1
12	1	1	11	410	1
13	3	2	14	275	0
14	2	1	7	350	0
15	3	2	8	290	9
16	1	3	15	320	1

Répondre aux questions suivantes en justifiant chaque fois la réponse par le maximum d'arguments :

- 1) a : donner le pourcentage valide des entreprises disposant d'un réseau local
b : quel est le CA moyen des entreprises de cet échantillon ?
- 2) en utilisant les annexes fournies répondre aux questions suivantes
a : peut on dire que la répartition des effectifs est symétrique
b : la variable CA est elle uniformément dispersée ou y a t-il des zones où la concentration est plus grande
c : quelle valeur proposeriez vous pour résumer la distribution du CA
d : construire un histogramme de la variable CA en 4 classes dont les extrémités sont définies par les quartiles
- 3) construire un tableau croisant les modalités des variables secteur et réseau en déduire

a : le pourcentage d'entreprises disposant d'un réseau local parmi celles opérant dans le secteur commercial

b : déterminer le pourcentage d'entreprises du secteur service et disposant d'un réseau

C : représentez graphiquement la répartition des réseaux selon le secteur

4) existe il une liaison entre la variable 'secteur' et 'réseau', justifiez (on pourra utiliser aussi les annexes)

5) calculez le CA moyen pour chaque ville . Que peut on déduire à propos de la liaison de ces deux variables ? justifier

6) Produire un graphique qui illustre la liaison entre 'effectif' et 'CA' y a il liaison significative ? justifier (on pourra utiliser les annexes).

7) Produire un graphique qui illustre la liaison entre ' le chiffre d'affaire' et ' Réseau ' y a il liaison significative ?

Annexes :

Statistiques	Effectif	Chiffre d'affaire
Effectif valide	16	15
Manquantes	0	1
Minimum	6	190
Maximum	22	520
Somme	184	5260
Moyenne	11,5	350,6
Ecart-type	4,29	102,11
Variance	18,4	10428,1
Quartile (25%)	8	277,5
Quartile (50%)	10,5	320
Quartile (75%)	14	441,25
Asymétrie	0,95	0,19
Applatissement	0,35	-1,17

Test de corrélation CA et Effectif :

Nombre d'observation valide :	15
Observations manquantes :	1
Covariance	175,7
Indice de corrélation (R)	0,387
Indice de détermination	0,15
Significativité (sig)	0,006

Test de Ki-2 entre CA et Secteur :

Nombre d'observation	14
Observation manquante	2
Statistique du Ki-2	452,12
Significativité (Sig)	0,09

Séminaire SPSS

Responsable : S.ZKIM

TP 10

Objectifs : tableau statistiques généraux, tableaux statistiques de base, tableau de bord, tableaux pivotants
Test de l'ANOVA.

- 1/Ouvrir le fichier 'Employes' et lister les informations sur les variables (étiquettes, étiquettes des valeurs, valeurs manquantes).
- 2/En utilisant un tableau de statistiques de base (menu Tabuler) donner les statistiques sur le salaire courant selon la variable sexe.
- 3/Peut-on dire que la répartition des salaires (courants) dépend du sexe des employés ? justifier en se basant aussi sur des graphiques.
- 4/Faire un test de l'ANOVA pour répondre à la question précédente.
- 5/Illustrer graphiquement la corrélation entre la variable salaire et ancienneté. Cette corrélation est-elle significative ?
- 6/Calculer le pourcentage des femmes managers dans l'échantillon.
- 7/Calculer le pourcentage des femmes parmi les managers, des managers parmi les femmes.
- 8/Illustrer graphiquement la répartition de la variable sexe selon le statut professionnels. Commenter.
- 9/Utiliser un test du χ^2 pour savoir si la corrélation entre les deux variables précédentes est significative ou non.
- 10/Visualiser la répartition de l'ancienneté par statut professionnel. Y a-t-il une liaison ?
- 11/Produire un graphique (interactif) illustrant la liaison entre le salaire et l'ancienneté par statut professionnel (un nuage par modalité de statut). Commenter et interpréter.
- 12/Même question mais en changeant la variable statut par la variable Sexe.
- 13/Produire un diagramme sectoriel interactif du statut professionnel selon le sexe des employés.
14. Etudier la liaison entre le salaire et le sexe pour chaque catégorie de la variable statut (utiliser un diagramme interactif). Cette liaison est-elle la même pour les trois catégories du statut professionnel ?

Objectifs : COMPARAISON DE MOYENNES et ANALYS DE VARIANCE**1) Chronométrage de l'assemblage d'un montage transistorisé : fichier spss: t test**

Le responsable du département d'Organisation et Méthodes 'une entreprise mentionne qu'en moyenne, le temps requis pour effectuer l'assemblage d'un montage transistorisé est de 10 minutes. Un chronométrage de cette opération sur 25 unités de même type donne les temps ci-contre.

Peut-on conclure, au seuil $\alpha = 0,05$, que le temps moyen pour effectuer l'assemblage de ce montage transistorisé est supérieur à la norme spécifiée ?

(fichier spss : t test)

Temps observés (minutes)				
9,1	11,3	11,9	10,9	11,9
11,6	10,3	11,8	11,0	10,9
12,3	9,6	11,9	10,8	12,9
10,3	10,1	12,2	11,1	10,3
10,0	10,5	9,8	11,8	12,0

2) EFFET D'UN PROGRAMME D'APPRENTISSAGE : fichier spss : t test apparié)

La directrice des ressources humaines d'une veut suggérer à la direction de l'entreprise de mettre en œuvre un programme spécial d'apprentissage pour les employés affectés au département d'assemblage. Pour évaluer l'efficacité de ce programme d'apprentissage d'une durée de 3 semaines, on a choisi au hasard 15 employés et on a observé le nombre de pièces assemblées durant une certaine période de temps. Par la suite, on a administré à ces mêmes employés le programme d'apprentissage et on a observé à nouveau le nombre de pièces

assemblées durant la même période de temps. Les résultats sont présentés dans le tableau ci-dessous :

Individus	Après le programme	Avant le programme
1	17	15
2	16	13
3	10	8
4	9	9
5	9	7
6	13	12
7	14	11
8	15	12
9	14	11
10	11	9
11	14	10
12	11	12
13	13	11
14	10	7
15	13	12

Ce programme est-il efficace au seuil de signification 1% ?

3) Evaluation de caractéristiques managériale : (fichier spss : comparaison moyennes et ANOVA)

On fait subir à des cadres de trois grandes entreprises oeuvrant dans trois secteurs différents (transport, fabrication de produits électroniques et le secteur des multimédias) un test d'appréciation et d'évaluation de diverses caractéristiques managériales et ceci à l'aide de diverse simulation de cas d'entreprises présentant différent niveaux de difficulté . Les résultats obtenus au test d'appréciation globale pour chaque groupe sont présentés dans le tableau ci-après .

Secteur d'équipements de transport	Secteur de produits électriques	Secteur de produits multimédias
200	210	202
198	208	256
177	183	225
210	223	202
214	227	246
189	197	168
168	172	204
168	172	181
178	184	164
191	199	168
213	225	170
195	204	205
195	204	234
184	191	189
201	211	202
170	174	187
206	217	208
151	162	198
188	196	233
204	215	222
183	190	239
129	155	165
221	235	172
178	184	172
206	218	163
147	147	210
212	225	204
178	183	200
182	189	196
150	150	203
192	201	

174	179	
	199	
	186	

- 1) L'appréciation globale moyenne est elle identique pour les cadres des secteurs transport et produits électriques ?
- 2) L'appréciation globale moyenne est elle identique pour les cadres des trois secteurs

4) Analyse de la variance à deux facteurs avec effet d'interaction : (fichier spss : MANOVA)

Les données du tableaux ci-après représentent les résultats obtenus par 36 employés à une évaluation , par le psychologue industriel de l'entreprise , concernant l'aspect névrotique de ces employés selon le niveau de stress de leur tâche et du nombre d'années associés à la tâche que l'employés effectue .

	Très stressant	stressant	Peu stressant
Moins de 5 ans	26	19	18
	29	24	25
	23	20	20
	24	25	21
Entre 5 et 10 ans	29	17	19
	33	25	23
	31	20	20
	29	22	20
Plus de 10 ans	26	15	10
	36	17	12
	29	18	9
	33	20	13

- 1) Existe-il un effet du facteur nombre d'années sur le résultat moyen d'évaluation ?
- 2) Existe-il un effet du facteur niveau de stress sur le résultat moyen de l'évaluation ?
- 3) Existe-il un effet d'interaction des deux facteurs sur le résultat moyen d'évaluation ?

Séminaire SPSS

Responsable : S.ZKIM

TP 12

Objectif :

- maîtrise de la procédure régression dans SPSS
- Interprétation des résultats d'une régression .

Les données suivantes proviennent de l'évaluation faite par le département des ressources humaines d'une entreprise , spécialisée dans la vente des systèmes informatisés , Cette entreprise possède déjà un nombre important de représentants mais doit accroître son personnel de vente pour assurer un service meilleur à sa clientèle , la plupart des représentants actuels avaient des formations universitaires provenant des divers secteurs (ingénierie, informatique, gestion) et l'engagement se faisant simplement à partir d'entrevue entre les divers chefs de service de l'entreprise , cette façon de faire crée des fois des situations conflictuelles en ce qui a trait à l'engagement de nouveaux personnels de vente , il a donc été proposé lors de la réunion mensuelle du groupe d'amélioration continue de l'entreprise de soumettre des vendeurs actuels à certains tests reconnus pour l'évaluation de certaines habilités et de déterminer si les résultats de cette évaluation peuvent être liés à l'indice de performance du vendeur , cet indice correspond au ratio des ventes du vendeur au cours de la dernière année et de l'objectif fixé par le responsable des ventes en début d'année , un indice supérieur à cent indique une performance qui dépasse l'objectif fixé .

Les données relatives à cette évaluation sont celles du tableau suivant : (fichier SPSS régression . SAV).

Performance	Intelligence	Estime	Personnalité	Expérience	Sexe
95	99	58	74	44	1
121	85	89	85	70	1
103	114	55	78	67	0
92	85	62	72	40	1
93	85	57	67	52	0
96	92	63	78	64	0
92	86	59	89	46	0
88	89	55	58	40	1
112	94	72	86	74	1
115	98	85	88	73	0
101	104	72	69	54	0
89	86	58	64	46	1
115	102	78	84	65	1
100	87	71	77	42	0
106	108	67	70	55	0
97	98	69	80	44	1
95	81	56	74	40	1
103	96	60	86	50	0
105	100	76	72	48	0
122	130	86	92	79	1
84	80	72	64	46	1
108	90	80	80	68	0
94	91	50	66	43	1
105	98	73	68	62	0
88	87	70	67	65	0
102	84	72	70	71	1
111	96	64	81	67	0
103	89	73	84	52	0
99	92	81	76	50	1
113	94	70	72	60	1
114	103	69	76	58	0
110	114	76	72	59	0

Dans quelle mesure Peut-on expliquer la variable indice de performance par les autres variables ?

Séminaire SPSS

Responsable : S.ZKIM

TP 13

Statistiques descriptives

	Moyenne	Ecart-type	n analyse
CYLINDRE	1906,1250	527,9087	24
PUISSANC	113,6667	38,7844	24
VITESSE	183,0833	25,2154	24
POIDS	1123,3333	248,4328	24
LONGUEUR	421,5833	41,3405	24
LARGEUR	168,8333	7,6537	24

Matrice de corrélation

		CYLINDRE	PUISSANC	VITESSE	POIDS	LONGUEUR	LARGEUR
Corrélation	CYLINDRE	1,000	,861	,693	,897	,864	,709
	PUISSANC	,861	1,000	,894	,769	,689	,552
	VITESSE	,693	,894	1,000	,507	,532	,363
	POIDS	,897	,769	,507	1,000	,863	,700
	LONGUEUR	,864	,689	,532	,863	1,000	,864
	LARGEUR	,709	,552	,363	,700	,864	1,000

Matrices des anti-images

		CYLINDRE	PUISSANC	VITESSE	POIDS	LONGUEUR	LARGEUR
Covariance des anti-images	CYLINDRE	9,806E-02	-1,493E-02	-5,627E-03	-2,932E-02	-2,318E-02	6,234E-04
	PUISSANC	-1,493E-02	4,721E-02	-5,606E-02	-4,068E-02	3,221E-02	-3,935E-02
	VITESSE	-5,627E-03	-5,606E-02	8,245E-02	5,763E-02	-4,190E-02	5,807E-02
	POIDS	-2,932E-02	-4,068E-02	5,763E-02	8,070E-02	-4,689E-02	4,963E-02
	LONGUEUR	-2,318E-02	3,221E-02	-4,190E-02	-4,689E-02	7,743E-02	-9,299E-02
	LARGEUR	6,234E-04	-3,935E-02	5,807E-02	4,963E-02	-9,299E-02	,193
Corrélation des anti-images	CYLINDRE	,934 ^a	-,219	-6,258E-02	-,330	-,266	4,531E-03
	PUISSANC	-,219	,625 ^a	-,898	-,659	,533	-,412
	VITESSE	-6,258E-02	-,898	,521 ^a	,706	-,524	,460
	POIDS	-,330	-,659	,706	,651 ^a	-,593	,398
	LONGUEUR	-,266	,533	-,524	-,593	,658 ^a	-,761
	LARGEUR	4,531E-03	-,412	,460	,398	-,761	,661 ^a

a. Mesure de précision de l'échantillonnage

Indice KMO et test de Bartlett

Mesure de précision de l'échantillonnage de Kaiser-Meyer-Olkin.		,669
Test de sphéricité de Bartlett	Khi-deux approché	171,188
	ddl	15
	Signification	,000

.....

.....

.....

.....

Qualité de représentation

	Initial	Extraction
CYLINDRE	1,000	,924
PUISSANC	1,000	,965
VITESSE	1,000	,946
POIDS	1,000	,859
LONGUEUR	1,000	,938
LARGEUR	1,000	,864

Méthode d'extraction : Analyse des principaux composants.

.....

.....

.....

.....

Variance expliquée totale

Composante	Valeurs propres initiales			Sommes des carrés chargées		
	Total	% de la variance ==	% cumulés	Total	% de la variance ==	% cumulés
1	4,617	76,956	76,956	4,617	76,956	76,956
2	,879	14,647	91,603	,879	14,647	91,603
3	,304	5,058	96,661			
4	,106	1,759	98,420			
5	7,316E-02	1,219	99,639			
6	2,165E-02	,361	100,000			

Méthode d'extraction : Analyse des principaux composants.

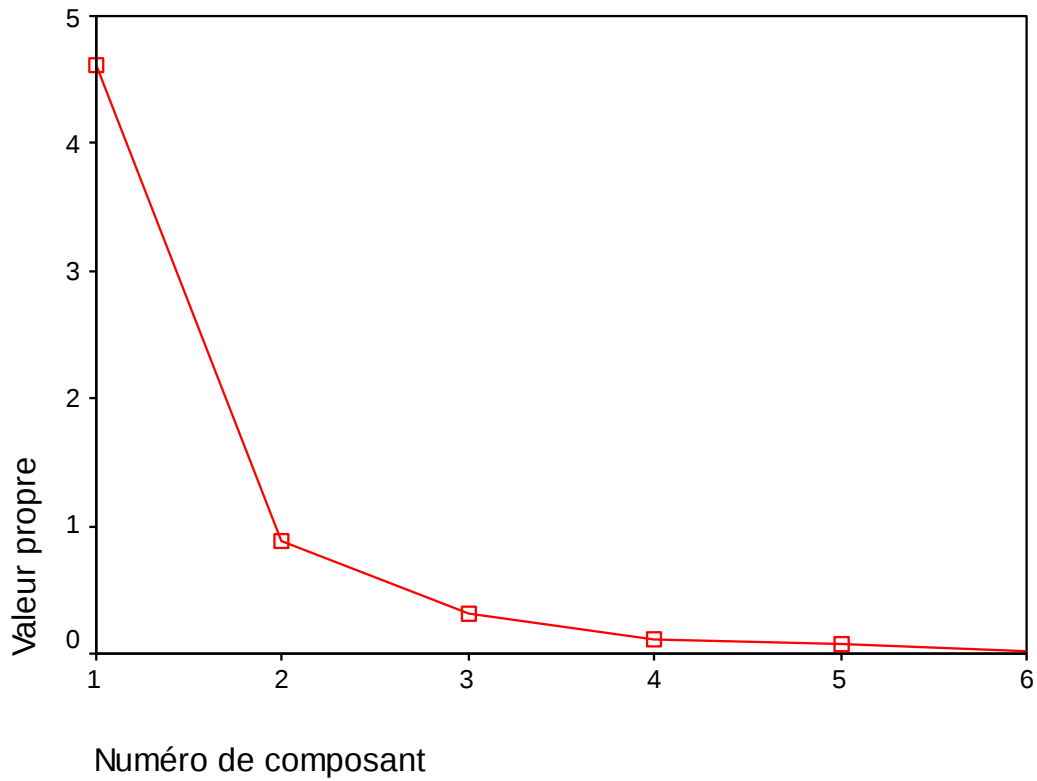
.....

.....

.....

.....

Graphique des valeurs propres



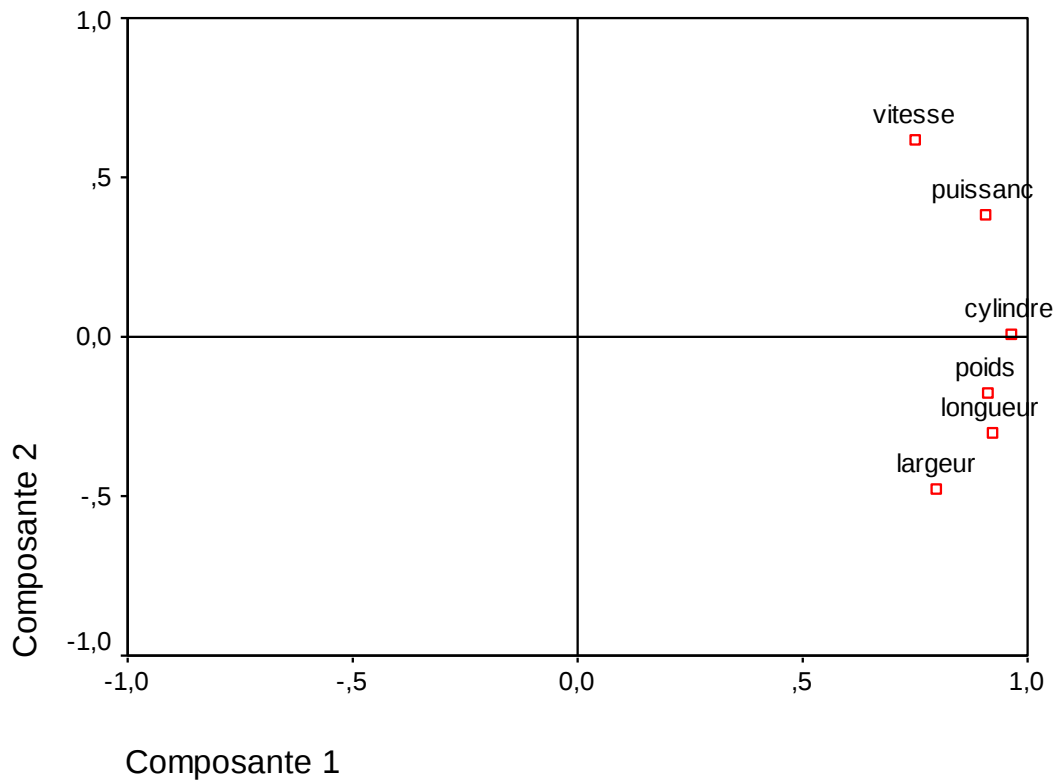
Matrice des composantes

	Composante	
	1	2
CYLINDRE	,961	1,095E-02
PUISSANC	,905	,382
VITESSE	,751	,618
POIDS	,910	-,176
LONGUEUR	,920	-,303
LARGEUR	,798	-,477

Méthode d'extraction : Analyse en composantes principales.

a. 2 composantes extraites.

Diagramme de composantes



Matrice des coefficients des composantes

	Composante	
	1	2
CYLINDRE	,208	,012
PUISSANC	,196	,435
VITESSE	,163	,703
POIDS	,197	-,201
LONGUEUR	,199	-,345
LARGEUR	,173	-,543

Méthode d'extraction : Analyse en composantes principales.

Scores composante.

Diagramme de composantes dans l'espace ap

