

Projet SAS Master 1 Mathématiques et Finance 2016 / 2017

Le projet est à réaliser en binôme. Il doit être envoyé par mail (thibaut.deloriot@outlook.com) avant le 06/01/2017 dans un dossier zippé portant le nom des membres du binôme et contenant :

- Un fichier .sas avec le code et les réponses aux questions sous forme de commentaires.
- Un fichier Excel pour l'exercice 3.

Exercice 1 : SNCF.

L'exercice doit être réalisé sans utiliser la procédure PROC SQL.

1. Exécuter le programme suivant. Commenter chacune des instructions. Analyser le résultat obtenu.

Indication : l'option @ de l'instruction INPUT permet d'interrompre momentanément la lecture d'un enregistrement sans passer au suivant.

```
DATA IMPORT;
INFILE CARDS4 DLM=",";
INPUT TEST $ @;
IF TEST=59 THEN DO;
    INPUT CODE $ LIBELLE :$17.;
    OUTPUT;
END;
DROP TEST;
CARDS4;
62;62041;ARRAS
59;59009;VILLENEUVE-D'ASCQ
59;59350;LILLE
62;62498;LENS
59;59360;LOOS
62;62193;CALAIS
;;;;
RUN;
```

2. Le fichier « Régularité mensuelle TGV.txt » contient, pour chaque mois, le nombre de TGV programmés, le nombre de TGV annulés et le nombre de TGV en retard à l'arrivée entre 2012 et 2015. Ouvrir le fichier avec le Bloc-notes. Comment est-il organisé ? Quelles sont les lignes qui peuvent poser problème lors d'une étape d'importation ?
3. En vous inspirant du programme de la première question, importer le fichier « Régularité mensuelle TGV.txt » dans la Table **TGV** stockée dans votre Bibliothèque personnelle. Les totaux annuels ne doivent pas être importés.
4. Afficher les 20 premières observations de cette Table.
5. Analyser la structure de cette Table. Combien y a-t-il d'observations ? de variables ? Identifier les attributs de chacune des variables. La Table est-elle triée ? Quelles sont les variables quantitatives ? qualitatives ?
6. Calculer la somme, la moyenne, le minimum, le maximum et la médiane des variables quantitatives. Y a-t-il des observations avec des valeurs manquantes pour ces variables ? Si oui, afficher ces observations.

7. Reprendre la question précédente et réaliser les calculs en fonction de l'année. Vérifier que le résultat obtenu corresponde aux totaux annuels contenus dans le fichier source.
8. Réaliser un tableau de cardinalités sur les variables qualitatives. Combien de modalités différentes peut prendre la variable ID_TRAJET ?
9. Produire un tableau de fréquences pour chaque variable qualitative.
Indication : l'instruction WEIGHT permet de préciser la variable quantitative à utiliser pour pondérer les observations (ici Nb de TGV programmés).
10. Afficher les 20 premières observations de la Table **STRUCTURE_SNCF**.
11. Analyser la structure de la Table **STRUCTURE_SNCF**. Combien y a-t-il d'observations ? de variables ? Identifier les attributs des différentes variables ? La Table est-elle triée ? Si oui, selon quelle(s) variable(s) ?
12. Réaliser un tableau de cardinalités. Combien de modalités différentes peut prendre la variable ID_TRAJET ?
13. Afficher les différentes modalités de la variable AXE.
14. Quelle est la variable qui doit être utilisée comme clé de jointure pour fusionner les Tables **TGV** et **STRUCTURE_SNCF** ? Quels sont les points à vérifier avant d'effectuer une jointure ? Réaliser la jointure entre ces deux Tables. Stocker le résultat dans une Table intermédiaire située dans la Bibliothèque temporaire.
15. A partir de la Table créée à l'étape précédente, répondre aux questions suivantes.
 - ▶ Quel est l'axe qui comptabilise le plus grand nombre de TGV programmés en 2015 ? Quel pourcentage cela représente-t-il ?
 - ▶ Calculer le nombre moyen de TGV programmés mensuellement par trajet en 2015 pour chaque axe. Quel est l'axe avec le nombre moyen le plus élevé ? le plus faible ?
 - ▶ Afficher les gares d'arrivées par ordre décroissant du nombre de TGV programmés pour chaque axe. Quelle ville est la mieux desservie ?
 - ▶ En quelle année le nombre de TGV annulés a t'il été le plus élevé ? Quel mois en particulier ? Indiquer le poids que cela représente.
16. Consolider les données en fonction de l'année et du croisement axe - année. Stocker le résultat dans la Table **RESULTAT** située dans la Bibliothèque temporaire.
Indication : Utiliser l'instruction TYPES pour définir les croisements de variables souhaités.
17. A partir de la Table **RESULTAT**, créer les indicateurs suivants.
 - ▶ Le pourcentage de trains annulés (PCT_ANNULATION).
 - ▶ Le pourcentage de trains à l'heure parmi ceux ayant circulés (PCT_REGULARITE).
 - ▶ Le nombre de trains à l'heure pour un train en retard à l'arrivée (RAPPORT).

18. Produire le rapport de synthèse affiché ci-dessous. Ce rapport est enregistré au format PDF dans le dossier associé à votre Bibliothèque personnelle. Il est composé d'un tableau global et d'un tableau pour chaque axe. Les différents tableaux sont affichés sur la même page.

Indications :

- Une étape de tri préalable est nécessaire. Identifier bien la ou les variables à utiliser.
- L'instruction *TITLE* permet d'insérer un titre dans un rapport.
- Le *STYLE=PLATEAU* est utilisé ici.
- Le format numérique avec séparateur de milliers : *NLNUM10*.
- Le format pourcentage : *PERCENT8.1*.

Régularité des TGV par Axe en 2015, 2014, 2013 et 2012

ANNEE	NB_TGV_PROGRAMMES	NB_TGV_ANNULES	PCT_ANNULATION	NB_TGV_EN_RETARD	PCT_REGULARITE	RAPPORT
2015	368 599	927	0.3%	39 606	89.2%	8,3
2014	354 634	3998	1.1%	33 546	90.4%	9,5
2013	354 923	1128	0.3%	41 402	88.3%	7,5
2012	365 252	526	0.1%	38 732	89.4%	8,4

AXE=ATLANTIQUE

ANNEE	NB_TGV_PROGRAMMES	NB_TGV_ANNULES	PCT_ANNULATION	NB_TGV_EN_RETARD	PCT_REGULARITE	RAPPORT
2015	132 356	220	0.2%	12 251	90.7%	9,8
2014	125 627	1608	1.3%	9 776	92.1%	11,7
2013	124 792	433	0.3%	12 012	90.3%	9,4
2012	128 521	136	0.1%	11 863	90.8%	9,8

AXE=EST

ANNEE	NB_TGV_PROGRAMMES	NB_TGV_ANNULES	PCT_ANNULATION	NB_TGV_EN_RETARD	PCT_REGULARITE	RAPPORT
2015	29 975	53	0.2%	2 357	92.1%	11,7
2014	30 813	268	0.9%	1 896	93.8%	15,1
2013	31 223	64	0.2%	2 810	91.0%	10,1
2012	32 541	41	0.1%	2 786	91.4%	10,7

AXE=NORD

ANNEE	NB_TGV_PROGRAMMES	NB_TGV_ANNULES	PCT_ANNULATION	NB_TGV_EN_RETARD	PCT_REGULARITE	RAPPORT
2015	43 864	81	0.2%	5 430	87.6%	7,1
2014	38 241	544	1.4%	4 218	88.8%	7,9
2013	38 482	236	0.6%	5 963	84.4%	5,4
2012	41 207	105	0.3%	4 962	87.9%	7,3

AXE=SUD-EST

ANNEE	NB_TGV_PROGRAMMES	NB_TGV_ANNULES	PCT_ANNULATION	NB_TGV_EN_RETARD	PCT_REGULARITE	RAPPORT
2015	162 404	573	0.4%	19 568	87.9%	7,3
2014	159 953	1578	1.0%	17 656	88.9%	8,0
2013	160 426	395	0.2%	20 617	87.1%	6,8
2012	162 983	244	0.1%	19 121	88.3%	7,5

Exercice 2 : CAC40

L'exercice doit être réalisé sans utiliser la procédure PROC SQL.

La Table **CAC40_2016** contient l'historique des cotations de l'indice CAC40 (FR0003500008) en 2016.

1. Analyser la structure de la Table (nombre d'observations, nombre de variables, attributs des variables). La Table est-elle triée ? si oui, selon quelle(s) variable(s) ?
2. Calculer, pour le dernier jour de cotations de chaque mois, les indicateurs suivants.
 - Le nombre de jours de cotations dans le mois.
 - Le cours de clôture du dernier jour de cotation du mois.
 - La variation mensuelle du cours de clôture en nombre et en pourcentage.
 - La variation depuis le 31/12 du cours de clôture en nombre et en pourcentage.

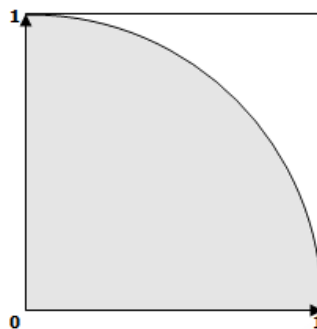
Indication : à partir de la variable DATE_COTATION créer deux nouvelles variables ANNEE et MOIS pour définir les blocs d'observations.

Obs	DATE_COTATION	NB_JOURS_M	COURS_CLT	VAR_M	VAR_M_PCT	VAR_A	VAR_A_PCT
1	29/01/2016	20	4 417,0	-220,0	-4.7%	-220,0	-4.7%
2	29/02/2016	21	4 353,6	-63,5	-1.4%	-283,5	-6.1%
3	31/03/2016	21	4 385,1	31,5	0.7%	-252,0	-5.4%
4	29/04/2016	21	4 429,0	43,9	1.0%	-208,1	-4.5%
5	31/05/2016	22	4 505,6	76,7	1.7%	-131,4	-2.8%
6	30/06/2016	22	4 237,5	-268,1	-6.0%	-399,6	-8.6%
7	29/07/2016	21	4 439,8	202,3	4.8%	-197,3	-4.3%
8	31/08/2016	23	4 438,2	-1,6	-0.0%	-198,8	-4.3%
9	30/09/2016	22	4 448,3	10,0	0.2%	-188,8	-4.1%
10	31/10/2016	21	4 509,3	61,0	1.4%	-127,8	-2.8%
11	30/11/2016	22	4 578,3	69,1	1.5%	-58,7	-1.3%

Exercice 3 : Approximation de π

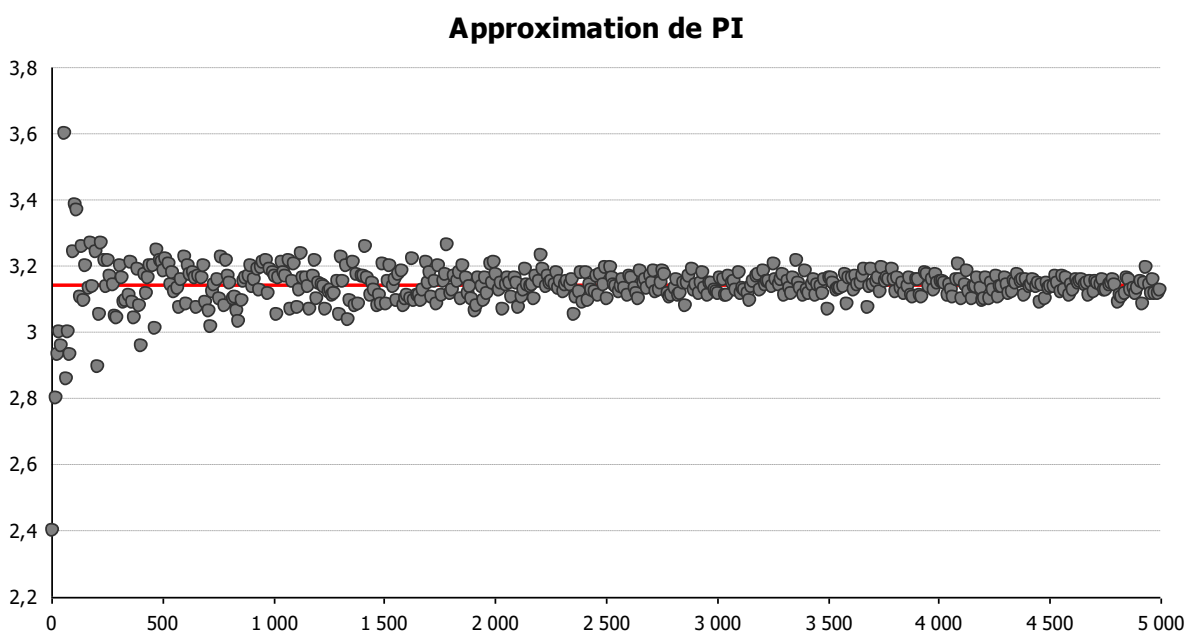
Cet exercice a pour but de déterminer une approximation de π par la méthode de Monte-Carlo.

1. Soit C un quart de cercle de centre $(0, 0)$ et de rayon 1 inscrit dans un carré de coté 1. Soit P un point de coordonnées (x, y) avec x et y pris au hasard dans l'intervalle $[0; 1]$. Quelle est la probabilité que le point P appartienne au quart de cercle C ?



2. Quelle condition doivent vérifier les coordonnées x et y du point P pour qu'il appartienne au quart de cercle C ?
3. Soit n le nombre de points tirés au hasard et k le nombre de points appartenant au quart de cercle C . Vers quelle valeur tend le ratio k/n ? En déduire une approximation de π .
4. Écrire un macro-programme **%Pi(n)** prenant comme paramètre le nombre de points à tirer au hasard et qui retourne, dans le Journal, une valeur approchée de π en utilisant la méthode décrite ci-dessus. Tester le macro-programme avec différentes valeurs de n .
Indication : la fonction RANUNI(0) génère un nombre aléatoire selon une loi uniforme sur $[0; 1]$.
5. Modifier le macro-programme précédent de manière à ce que le résultat ne soit plus affiché dans le Journal mais stocké dans une macro-variable exploitable en dehors de ce macro-programme.
6. Écrire un macro-programme **%Pi2(deb, fin, pas)** qui exécute le macro-programme créé précédemment un certain nombre de fois en faisant varier la valeur du paramètre n (à l'aide des paramètres *deb*, *fin* et *pas*). Pour chaque itération, conserver la valeur prise par le paramètre n et l'approximation de π ainsi obtenu dans une Table **PI** stockée dans la Bibliothèque temporaire. Exécuter le macro-programme avec les paramètres suivants : *deb*=100, *fin*=5 000 et *pas*=99.
7. Exporter le contenu de la Table **PI** dans un fichier texte avec séparateur de champs « ; ». Ouvrir le fichier avec Excel et représenter graphiquement la série de données. Commenter le résultat obtenu.

Exemple de représentation graphique avec $deb=100$, $fin=5\ 000$ et $pas=99$.



Exercice 4 : Décès 2015

La Table **BASE_DECES_2015** contient le détail des décès des personnes domiciliées en France (hors Mayotte) en 2015. Les informations sont issues des bulletins de décès émis par les communes. Il s'agit de la commune du domicile du défunt. Les données sont anonymes et communiquées mensuellement.

1. Afficher les 20 premières observations de cette Table.
2. Analyser la structure de cette Table. Combien y a-t-il d'observations ? de variables ? Identifier les attributs de chacune des variables. La Table est-elle triée ? Quelles sont les variables quantitatives ? qualitatives ?
3. Réaliser un tableau de cardinalités sur les variables qualitatives. Combien de modalités différentes peut prendre la variable COD_COMMUNE ?
4. Produire un tableau de fréquences pour chaque variable qualitative. Analyser et commenter les résultats obtenus.
5. Réaliser un tableau de fréquences pour les croisements de variables qui vous semblent pertinents. Analyser et commenter les résultats obtenus.
6. Calculer la moyenne, le minimum, le maximum et la médiane de la variable AGE. Y a-t-il des observations avec des valeurs manquantes pour cette variable ? Si oui, afficher ces observations.
7. Reprendre la question précédente et réaliser les calculs en fonction de la variable SEXE. Quel est l'âge moyen de décès pour un homme ? pour une femme ?
8. Indiquer la valeur de l'écart interquartile de la variable AGE en fonction du SEXE. Interpréter le résultat.
9. Représenter la distribution de la variable AGE en fonction du SEXE à l'aide d'un graphique de type boîte à moustaches (boxplot).
10. A partir de la variable AGE, créer une variable TRANCHE_AGE contenant les classes suivantes :
 - ▶ 0 – 17 ans.
 - ▶ 18 – 44 ans.
 - ▶ 45 – 59 ans.
 - ▶ 60 – 74 ans.
 - ▶ 75 – 84 ans.
 - ▶ 85 – 94 ans.
 - ▶ 95 ans et plus.
11. Calculer la fréquence de chacune des classes de la variable TRANCHE_AGE.
12. Réaliser un tableau de fréquences pour le croisement des variables TRANCHE_AGE et SEXE. Analyser et commenter le résultat obtenu. Les variables sont-elles indépendantes ?
13. Calculer pour chaque commune, le nombre de décès, le nombre d'hommes décédés, le nombre de femmes décédées et le nombre de décès pour chacune des classes de la variable TRANCHE_AGE. Stocker le résultat dans la Table **CONSO_COMMUNE** située dans la Bibliothèque temporaire.

	COD_COMMUNE	NB_DECES	NB_DECES_H	NB_DECES_F	NB_DECES_0_17	NB_DECES_18_44	NB_DECES_45_59	NB_DECES_60_74	NB_DECES_75_85	NB_DECES_85_94	NB_DECES_SUP_95
1	01001	5	3	2	0	0	0	1	2	2	0
2	01002	1	0	1	0	0	0	0	0	1	0
3	01004	121	60	61	1	2	7	16	39	45	11
4	01005	7	6	1	0	0	1	3	2	1	0
5	01006	2	1	1	0	0	0	0	1	1	0

La Table **STRUCTURE_TERRITOIRE** contient la liste des régions (niveau=1), des départements (niveau=2) et des communes (niveau=3) de la France métropolitaine et des départements d'outre-mer (hors Mayotte).

14. Analyser la structure de la Table (nombre d'observations, nombre de variables, attributs des variables). La Table est-elle triée ? Si oui, selon quelle(s) variable(s) ?
15. Ouvrir la Table **STRUCTURE_TERRITOIRE**. Comment est-elle organisée ? Comment peut-elle être exploitée ?
16. En utilisant des jointures réflexives, créer la Table **STRUCTURE** organisée de la façon suivante. Elle sera stockée dans la Bibliothèque temporaire.

	COD_REGION	LIB_REGION	COD_DEPARTEMENT	LIB_DEPARTEMENT	COD_COMMUNE	LIB_COMMUNE
1	01	GUADELOUPE	971	GUADELOUPE	97101	LES ABYMES
2	01	GUADELOUPE	971	GUADELOUPE	97102	ANSE-BERTRAND
3	01	GUADELOUPE	971	GUADELOUPE	97103	BAIE-MAHAULT
4	01	GUADELOUPE	971	GUADELOUPE	97104	BAILLIF
5	01	GUADELOUPE	971	GUADELOUPE	97105	BASSE-TERRE

Indication : l'utilisation d'alias de Table est nécessaire lors de la réalisation de jointures réflexives.

17. Combien y a-t'il de régions ? de départements ? de communes ?
18. Afficher pour chaque région, le nombre de départements et de communes qui la compose.
19. Afficher les communes des départements du Nord et du Pas-de-Calais qui ont le même nom.
20. Quel est le nom de commune le plus répandu en France ? Combien de communes portent ce nom ?
19. Quelle est la variable qui doit être utilisée comme clé de jointure pour fusionner les Tables **STRUCTURE** et **CONSO_COMMUNE** ? Quel type de jointure allez-vous utiliser ? Justifier ? Réaliser la jointure entre ces deux Tables. Stocker le résultat dans la Table **DECES_2015** située dans la Bibliothèque temporaire.
20. Combien y a-t-il de communes qui n'ont pas eu de décès en 2015 ?
21. Afficher, pour chaque région, le nombre de décès, le nombre et le pourcentage de décès en fonction du sexe.
22. Créer une variable **POIDS_DEP** qui indique le poids de la commune dans son département en fonction du nombre de décès.
23. Exporter les données de la région Nord Pas-de-Calais Picardie dans un fichier Excel composé d'un onglet par département. Les communes seront triées par ordre décroissant du nombre de décès enregistrés.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	COD_DEPARTEMENT=02												
2	COD_COMMUNE	LIB_COMMUNE	NB_DECES	POIDS_DEP	NB_DECES_H	NB_DECES_F	NB_DECES_0_17	NB_DECES_18_44	NB_DECES_45_59	NB_DECES_60_74	NB_DECES_75_85	NB_DECES_85_94	NB_DECES_SUP_95
3	2691	SAINT-QUENTIN	639	11,15%	298	341	3	18	64	138	143	230	43
4	2722	SOISSONS	300	5,24%	156	144	2	13	33	60	72	104	16
5	2408	LAON	268	4,68%	124	144	1	9	27	56	50	109	16
6	2168	CHATEAU-THIERRY	173	3,02%	81	92	3	7	18	27	32	63	23
7	2738	TERGNIER	170	2,97%	86	84	3	7	17	32	40	58	13
8	2173	CHAUNY	164	2,86%	88	76	1	1	17	36	35	65	9
9	2381	HIRSON	132	2,30%	65	67	0	2	9	28	31	55	7
10	2064	BELLEU	101	1,76%	31	70	1	1	3	12	27	53	4
11	2361	GUISE	101	1,76%	48	53	0	2	11	19	18	44	7
12	2095	BOHAIN-EN-VERMANDOIS	93	1,62%	45	48	0	0	4	17	32	38	2
13	2810	VILLERS-COTTERETS	74	1,29%	46	28	0	4	10	15	23	18	4
14	2680	SAINT-GOBAIN	72	1,26%	34	38	0	4	3	8	20	28	9
15	2558	LE NOUVION-EN-THIERACHE	58	1,01%	29	29	0	0	1	7	15	31	4
16	2304	LA FERRE	54	0,94%	21	33	0	3	3	6	9	27	6
17	2305	FERRE-EN-TARDENOIS	53	0,92%	27	26	0	0	6	8	12	24	3
18	2340	GAUCHY	44	0,77%	25	19	1	1	3	10	13	14	2
19	2468	MARLE	40	0,70%	19	21	1	1	2	6	10	15	5
20	2114	BRASLES	40	0,70%	14	26	0	0	1	2	12	20	5

24. Transformer le programme de la question précédente en un macro-programme *%Export* prenant comme paramètre le code de la région à exporter. Le nom du fichier créé contiendra le code de la région. Tester le macro-programme avec différentes valeurs.
25. Reprendre le macro-programme précédent et apporter les modifications nécessaires pour pouvoir saisir une liste de codes région dans le paramètre. Les différents codes région seront séparés par le caractère « - ». Un fichier Excel sera alors créé pour chaque région. Tester le macro-programme avec différentes valeurs.
26. Modifier le macro-programme de manière à traiter l'ensemble des régions si l'utilisateur ne précise pas de valeur dans le paramètre au moment de l'appel du macro-programme. Quel type de paramétrage devez-vous utiliser ? Tester le macro-programme.