votrenom.zip. Ce r´epertoire contiendra le fichier votrenom.sascorrespondant au code ´ecrit en r´eponse au

contrˆole, ainsi qu’un r´epertoirebib copie du r´epertoire vers lequel pointent vos biblioth`eques. Les deux parties

sontind´ependantes. Les donn´ees sont t´el´echargeables `a l’adresse http://filez.univ-amu.fr/7dr4vxryo7

sous la forme d’une archive nomm´eedata CC SAS M2MASSPOP1617.zip

PARTIE 1

**Question 1.1**

La variable **PV1MATH** correspond au score synthétique de l’élève aux évaluations de mathématique. Le tableau suivant en synthétise la distribution (non-pondérée) :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis Variable : PV1MATH** | | | | | |
| **N** | **Mean** | **Median** | **Variance** | **Minimum** | **Maximum** |
| **1368** | **498.1789377** | **498.6836** | **9232.63** | **230.807** | **781.4379** |

**a.** Quelle **procédure** a été utilisée pour produire ces résultats ?

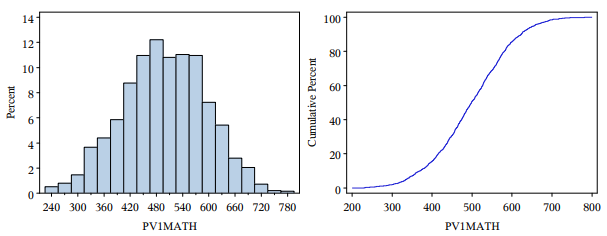
**b.** Interprétez la valeur de la moyenne. En analysant les autres statistiques présentées, que pensez-

vous de l’influence des valeurs extrêmes ?

**c.** Calculez l’écart-type et le coefficient de variation de la variable **PV1MATH**.

**Question 1.2**

Les deux graphiques suivants représentent la distribution de la variable **PV1MATH.**

**a.** Quel est le nom de ces deux graphiques ? En quoi leur analyse confirme-t-elle les résultats de la

question précédente quant à l’influence des valeurs extrêmes ?

**b.** Utilisez ces graphiques pour déterminer (approximativement) la valeur des quartiles de la variable

PV1MATH ainsi que celle des premier et neuvième déciles.

**c.** Quelle **procédure** auriez-vous pu utiliser pour obtenir directement les quantiles de la variable PV1MATH ?

**d.**Pourriez-vous-expliquer l’output de la procédure suivante :

procboxplot data=sashelp.cars;

plotlength\*make;

inset min mean max std;

labellength = 'The length'

make='The cars';

run;

**e.** Comment peut afficher la distribution de la variable length (en utilisant un boxplot) sur toute la population en se basant sur la procédure suivante ?

**Question 1.3**

1. On donne deux tables **dads** et **kids** de la façon suivante :

|  |  |
| --- | --- |
| **Pour dads**  famid name inc  2 Art 22000  1 Bill 30000  3 Paul 25000  **;** | **Pour kids**  famidkidname birth age wt sex  1 Beth 1 9 60 f  1 Bob 2 6 40 m  1 Barb 3 3 20 f  2 Andy 1 8 80 m  2 Al 2 6 50 m  2 Ann 3 2 20 f  3 Pete 1 6 60 m  3 Pam 2 4 40 f  3 Phil 3 2 20 m  **;** |

**Faites le match fusion des deux tables**.

1. Comment obtenons-nous les résultats des fusions suivantes à partir de la table suivante :

Table

Description automatically generated

**Résultat 1 :**

Table

Description automatically generated

**Résultat 2 :**

Table

Description automatically generated

**Résultat 3 :**

Table

Description automatically generated

Table

Description automatically generated

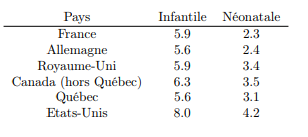
1. Quelle est la différence entre ces trois résultats ?

**Question 1.4**

1. Enregister la table CARS dans la librarie Work en utilisant une macro-variable data pour la table sashelp.cars
2. Afficher les dix premières lignes de la table sashelp.class en utilisant une macro-variable qui s’appelle DATA pour sashelp.class

**Question 1.5** (Mortalité infantile)

En 1994, les indicateurs démographiques donnaient les taux de mortalité infantile et de mortalité néonatale précoce (pour 1000 naissances vivantes) suivants :



1. Créer une table SAS contenant ces données(avec proc sql).
2. Etudier la corrélation entre les deux variables.
3. Tracer le nuage de points correspondant
4. Ecrire une requête permettant d’afficher les indicateurs de la France et le Canada.
5. Quel pays a le taux de mortalité infantile le plus élevé.
6. Afficher les statistiques de distribution des deux variables