UNIVERSITE DE CARTHAGE

**ECOLE SUPERIEURE DE LA STATISTIQUE**

**ET DE L’ANALYSE DE L’INFORMATION**

**Matière** : TP SAS **Devoir surveillé :** Octobre 2021 **Filière :** 1ère année 2021

**Exercice 1**

Nous disposons d’un fichier **gains.xlsx** **(tableau 1 page 3)** contenant, pour un numéro de joueurs de tennis donnée, le **nom du tournois**, **l’année**, le montant de la **prime** et le **sponsor**.

1. Créez la bibliothèque “examen”, importez le fichier « gains.xlsx » et sauvegardez le dans cette bibliothèque.

2. Calculez les fréquences pour la variable « lieutournoi » (proc freq).

3. Filtrez les valeurs où le lieu de tournoi est Roland Garros et sponsor est Peugeot.

4. Calculez la moyenne de la variable « prime » par l’année (proc sort et proc means).

5. Créez une nouvelle variable « country » selon le critère de « lieutournoi ».

**Les valeurs de la variable « country » sont France, UK, et US (if else ou encore select when). Utilisez proc freq pour calculer les fréquences de la variable «country ».**

**Annexe 1 :pour les réponses**

/\* La bibliothèque \*/

--------------------- TD2 "/home/u43532959/TD/----------";

run;

**/\* Importation des données 'gains.xlsx'\*/**

proc ---------- datafile="/home/u43532959/TD/gains.xlsx"

dbms=----------------

out=-------------------

replace;

run;

**/\* 1.Calculez les fréquences pour la variable lieutournoi \*/**

proc ------ data=TD2.gains;

-------------- lieutournoi; run;

**/\* 2. Filtrez les valeurs ou le lieu de tournoi est Roland Garros et sponsor est Peugeot.\*/**

data gains\_filtrage;

set --------------;

---------------- like ---------- and ------------------- like -----------------; run;

/\* **3.Calculez la moyenne de variable prime par l'année (proc sort et proc means) \*/**

proc --------- data=------- out=gains\_sort;

by --------------------------;

--------------------;

proc ------------------- data=------------ noprint;

by ------------------;

---------------- prime;

output out=gains\_moyenne(drop= \_type\_ \_freq\_)

------------=moyenne\_prime;------------------- ;

\* 4. Créez une nouvelle variable country selon le critère de lieutournoi.

Les valeurs de la variable country sont France, UK, et US

**(if else ou encore select when)**

data gains\_country;

set ----------;

if ----------='Roland Garros' then ------------='France';

--------------- lieutournoi= -------------- then

country='UK';

else if ---------------------**='**-----------------------' then

----------------------**=**-----------------------;

---------;

proc --------- data=gains\_country;

table -------------------; run;

Tableau 1 gains.xls

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| nujoueur | lieutournoi | annee | prime | sponsor |
| 14 | Roland Garros | 1992 | 200000 | Peugeot |
| 14 | Roland Garros | 1994 | 1800000 | Reebok |
| 14 | Wimbledon | 1992 | 700000 | Peugeot |
| 14 | Wimbledon | 1993 | 1400000 | Peugeot |
| 13 | Roland Garros | 1990 | 1100000 | Kennex |
| 13 | Roland Garros | 1992 | 1500000 | Kennex |
| 13 | Wimbledon | 1989 | 350000 | Donnay |
| 13 | Wimbledon | 1992 | 400000 | Kennex |
| 12 | Roland Garros | 1990 | 400000 | Dunlop |
| 12 | Roland Garros | 1992 | 200000 | Dunlop |
| 12 | Wimbledon | 1989 | 600000 | Dunlop |
| 12 | Flushing Meadow | 1989 | 1600000 | Dunlop |
| 12 | Flushing Meadow | 1991 | 1800000 | Lacoste |
| 11 | Roland Garros | 1990 | 700000 | Kennex |
| 11 | Roland Garros | 1992 | 500000 | Kennex |
| 11 | Wimbledon | 1989 | 1000000 | Dunlop |
| 10 | Roland Garros | 1994 | 600000 | Peugeot |
| 8 | Roland Garros | 1992 | 500000 | Lacoste |
| 8 | Roland Garros | 1994 | 1000000 | Reebok |
| 8 | Wimbledon | 1989 | 350000 | Peugeot |
| 7 | Roland Garros | 1992 | 200000 | Donnay |
| 7 | Wimbledon | 1993 | 800000 | Reebok |
| 7 | Flushing Meadow | 1991 | 1000000 | Donnay |
| 6 | Roland Garros | 1992 | 900000 | Dunlop |
| 6 | Wimbledon | 1992 | 1200000 | Dunlop |
| 5 | Wimbledon | 1992 | 300000 | Dunlop |
| 5 | Wimbledon | 1993 | 350000 | Reebok |
| 4 | Roland Garros | 1994 | 400000 | Lacoste |
| 4 | Wimbledon | 1992 | 300000 | Lacoste |
| 4 | Wimbledon | 1993 | 350000 | Lacoste |
| 9 | Roland Garros | 1990 | 400000 | Peugeot |
| 9 | Roland Garros | 1992 | 200000 | Peugeot |
| 9 | Roland Garros | 1994 | 600000 | Reebok |
| 9 | Wimbledon | 1992 | 400000 | Peugeot |
| 9 | Wimbledon | 1993 | 500000 | Reebok |
| 9 | Flushing Meadow | 1989 | 900000 | Lacoste |
| 3 | Roland Garros | 1990 | 500000 | Donnay |
| 3 | Roland Garros | 1992 | 550000 | Donnay |
| 3 | Roland Garros | 1994 | 400000 | Reebok |
| 3 | Wimbledon | 1992 | 850000 | Donnay |

**EXERCICE 2**

**Une équipe cycliste suit la progression de la VO2max de ses coureurs au fil des années.**

**Pour 10 coureurs cyclistes, les VO2max suivantes ont été notées en 2015 et en 2016.**

|  |  |  |
| --- | --- | --- |
| **Coureur** | **2015** | **2016** |
| **C1** | **68** | **69** |
| **C2** | **60** | **65** |
| **C3** | **74** | **70** |
| **C4** | **68** | **69** |
| **C5** | **67** | **73** |
| **C6** | **74** | **72** |
| **C7** | **70** | **69** |
| **C8** | **63** | **72** |
| **C9** | **74** | **65** |
| **C10** | **71** | **74** |

**1. Rentrez ce jeu de données via une étape data.**

**2. Créez une nouvelle variable qui définit la différence de VO2max entre 2016 et 2015.**

**3. Créer un variable sex pour chaque coureur de telle sorte que les cinq premiers sont « homme » et les cinq derniers sont « femme ». Utilisez le proc format pour le label de cette variable.**

**4. Créez une nouvelle table dans laquelle la différence de VO2max entre 2016 et**

**2015 soit positive**