# Exercices pour la séance 2

## Exercice 1

Cet exercice est tiré d’un ancien devoir (rapport 1\_Automne 2016). C’est la même mise en situation que l’exercice 1 de la séance 1.

La base de donnée avec laquelle vous allez faire une partie de vos analyses est intitulée ***« Données\_énergie\_2009\_A16.xlsx »***. Le fichier comporte deux onglets : un onglet de donnée ***« Données\_cas »*** et un onglet explicatif des variables retenues ***« Liste\_variables\_Description ».***

## **Question 1**

Le gouvernement américain veut lancer un programme incitatif, dans le cadre d’un projet pilote, visant la réduction du coût total de l’énergie dans les bâtiments de type « Maison Mobile, Maison attenante et Maison unifamiliale ».

Vous êtes donc appelés à évaluer la répartition du **coût total par pieds carré** ($) en 2009 des maisons construites avant l’an 2000 et la comparer aux maisons construites après cette date. Les variables « Coût\_total\_énergie » et « Pied\_carré » sont les deux variables importantes à utiliser.

Comparez les statistiques descriptives pour les deux groupes et commentez leurs Boxplots.

**NB : *la variable Coût total par pieds carré n’est pas dans la base de données. Vous êtes appelés à la créer avant de répondre à la question 2.***

## **Question 2**

Vous avez certainement remarqué un comportement atypique en termes de **coût total par pied carré** chez un ménage dont la propriété est **récente.**

1. Quel est le numéro d’identification de ce ménage?
2. Énumérez une dizaine de caractéristiques qui vous paraissent intéressantes parmi les 80 variables disponibles pour ce ménage.

## **Question 3**

Vous n’êtes peut être pas un spécialiste en énergie mais vous êtes capable, sur la base de l’information disponible sur les **maisons rurales similaires**, **construites en 2008**, de déduire si ce comportement relève d’une ou des caractéristiques spécifiques de ce ménage ou s’il s’agit d’une erreur de transcription des données.

***NB : Les maisons comparables doivent répondre aux critères suivants : même région de recensement, même état, même type de maison, même année de construction et même climat.***

1. Combien de maisons comparables avez-vous trouvés dans la base de données?
2. Commentez le résultat et formulez au moins 3 hypothèses pour expliquer ce coût exagérément élevé.
3. Votre supérieur immédiat vous demande d’éliminer cette observation et de refaire l’analyse comparative de la question 2. Commentez le résultat.

## Exercice 2

***Note : Dans les trois questions qui suivent, on s’intéresse à la valeur des variables par match joué. Il faut donc diviser au préalable les variables par le nombre de matchs joués.***

**Question 1 :** À partir du fichier 1-605-Hockey2014.xlsx, calculer les statistiques descriptives (moyenne, médiane, coefficient d’asymétrie, étendue, écart-type, IQR) et tracer le boxplot du temps de jeu **par match** par position. Que remarquez-vous?

**Question 2 :** À partir du fichier 1-605-Hockey2014.xlsx, calculer les statistiques descriptives (moyenne, médiane, coefficient d’asymétrie, étendue, écart-type, IQR) et tracer le boxplot du nombre de tirs **par match** par position. Que remarquez-vous?

**Question 3 :** À partir du fichier 1-605-Hockey2014.xlsx, calculer les statistiques descriptives (moyenne, médiane, coefficient d’asymétrie, étendue, écart-type, IQR) et tracer le boxplot du nombre de minutes de pénalité **par match** par position. Que remarquez-vous?

## Exercice 3

*Cet exercice est tiré de l’examen Intra (Automne 2016).*

La chaîne de supermarchés « ***Finebouche***», chef de file de l'alimentation au détail, s’engage à offrir à ses clients une expérience unique de magasinage.

Les hauts dirigeants de la chaîne ne s’entendent pas sur la politique de gestion des plaintes des clients. Certains prétendent qu’une compensation monétaire sous forme d’un coupon de 20 $ devrait être envoyée par la poste aux clients non satisfaits et d’autres pensent qu’un message de courtoisie personnalisé serait suffisant pour rétablir la confiance des clients insatisfaits.

Les analystes de la chaîne ont décidé d’analyser l’effet de ces deux politiques afin de parvenir à une recommandation.

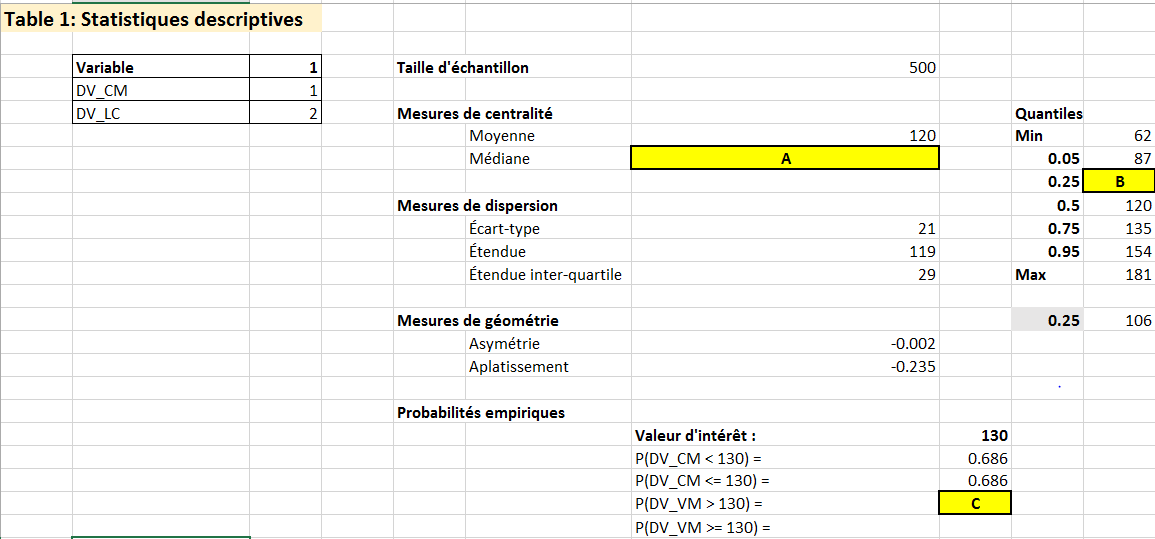
Deux échantillons aléatoires simples de 500 clients chacun ont été tirés. Une lettre de courtoisie a été envoyée au premier échantillon et une compensation monétaire a été accordée au deuxième échantillon.

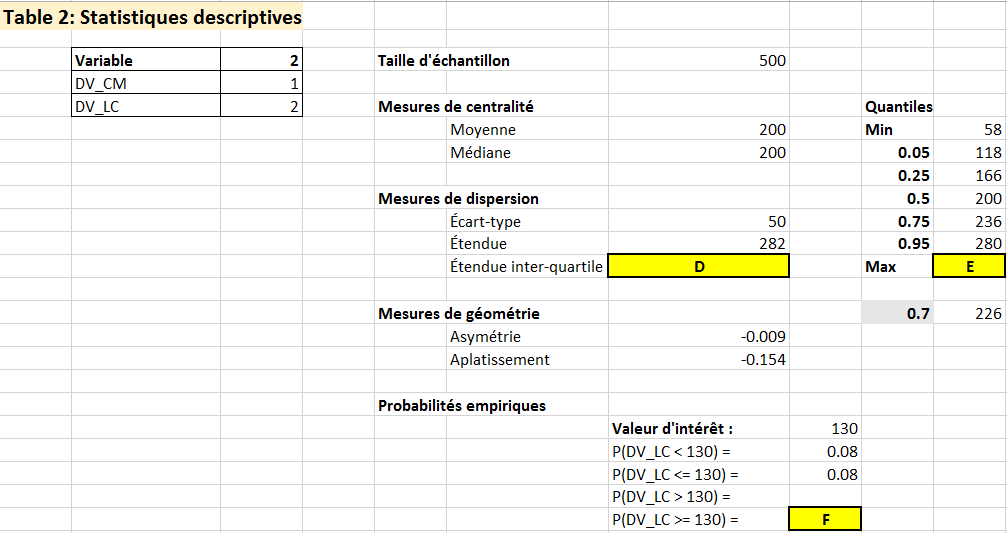
Vous trouverez à **la Figure 1 (Table 1 & Table 2)** les statistiques descriptives des dépenses par visite observées un mois après la mise en œuvre de chaque politique.

***Note :***

Veuillez noter que l’on désignera par **DV\_CM** les dépenses par visite observées un mois après l’envoi de la compensation monétaire et par **DV\_LC** les dépenses par visite observées un mois après l’envoi de la lettre de courtoisie.

**Figure 1 : Statistiques descriptives des dépenses par visite observées un mois après la mise en œuvre de chaque politique.**



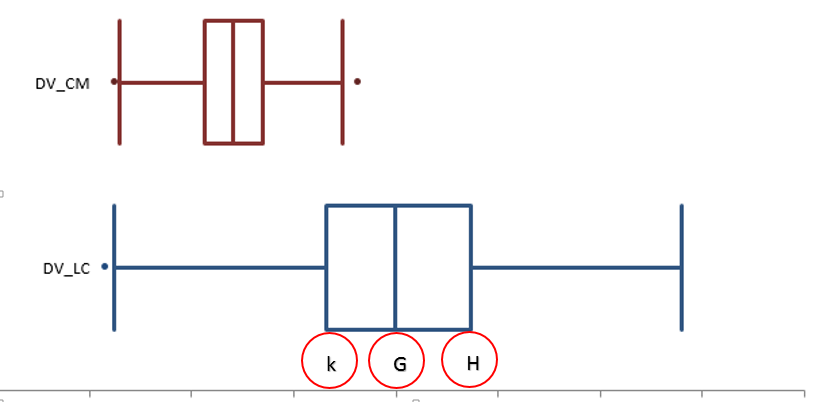


**Question 1** : Identifiez les variables correspondantes aux deux tables de la **Figure 1**.

**Question 2** : Donnez les valeurs manquantes (**A, B et C**) de la **Table 1** et (**D, E et F**) de la **Table 2** de la

**Figure 1**.

**Question 3** : Donnez les valeurs correspondantes aux barres verticales identifiées par les lettres **k, G et H** dans la **Figure 2** ci-dessous.

**Figure 2 :** **Répartition des dépenses par visite observées un mois après la mise en œuvre de chaque politique**.

**Réponses**

## Exercice 1

Les réponses de cet exercice se trouvent dans le fichier : « Données\_Énergie\_2009\_solution\_Exercice1\_Questions1&3.xlsx »

## Exercice 2

**Question 1 : Temps de jeu par match**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **RW** | **LW** | **C** | **D** |
| **Taille d'échantillon** | | 162 | 163 | 258 | 303 |
|  | **Moyenne** | 12.83 | 13.10 | 13.81 | 18.13 |
|  | **Médiane** | 13.07 | 13.07 | 13.87 | 17.97 |
|  | **Coef. d’asymétrie** | -0.13 | -0.16 | -0.12 | -0.07 |
|  | **Étendue** | 16.94 | 17.06 | 19.98 | 22.91 |
|  | **Écart-type** | 4.16 | 4.24 | 4.12 | 4.03 |
|  | **IQR** | 6.36 | 6.00 | 6.46 | 5.57 |



On constate :

1. En moyenne, le temps de jeu par match des défenseurs est plus élevé que celui des autres alors qu’il est similaire pour les trois types d’attaquants. La boîte du boxplot des défenseurs est également plus à droite que celle des attaquants.
2. Au niveau de la dispersion, l’écart-type est similaire pour toutes les positions malgré le fait que l’IQR soit plus faible chez les défenseurs. On constate des valeurs plus extrêmes chez les défenseurs (étendue plus grande). Chez les attaquants, les centres ont un IQR plus faible mais une étendue plus grande que les ailiers.
3. Les coefficients sont tous légèrement négatifs (mais près de 0). Les distributions sont relativement symétriques. Ce qui est confirmé par les boxplots et surtout les histogrammes disponibles dans le fichier Excel ainsi que le fait que les moyennes sont près des médianes.

**Question 2 : Nombre de tirs par match**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **RW** | **LW** | **C** | **D** |
| **Taille d'échantillon** | | 162 | 163 | 258 | 303 |
|  | **Moyenne** | 1.57 | 1.61 | 1.50 | 1.22 |
|  | **Médiane** | 1.49 | 1.50 | 1.40 | 1.16 |
|  | **Coef. d’asymétrie** | 0.20 | 0.94 | 0.48 | 0.55 |
|  | **Étendue** | 3.72 | 4.95 | 3.68 | 3.28 |
|  | **Écart-type** | 0.86 | 0.92 | 0.79 | 0.62 |
|  | **IQR** | 1.20 | 1.08 | 1.01 | 0.71 |



On constate que :

1. En moyenne, les défenseurs ont moins de tirs que les attaquants et les ailiers en ont légèrement plus que les centres.
2. Les moyennes sont toujours un peu plus élevées que les médianes, ce qui est laisse croire à une asymétrie à gauche. Les coefficients positifs le confirment et montrent que cette asymétrie est plus prononcée chez les ailiers gauches (l’histogramme disponible sur le fichier Excel le montre aussi).
3. Chez les attaquants, on constate des boxplots dont la boîte mesure environ 1 unité (IQR d’environ 1). On constate que 50% des attaquants font entre 1 et 2 tirs par match environ (voir les quartiles pour des valeurs plus précises), environ 25% des attaquants font moins d’un tir par match alors qu’environ 25% en font plus de 2. Chez les défenseurs, les boîtes sont encore plus étroites (IQR d’environ 0.7), 50% des défenseurs font entre 0.84 et 1.56 tir par match (voir les quartiles dans l’onglet Statistiques).
4. La dispersion est plus grande chez les attaquants que chez les défenseurs (voir les écart-types, étendues et IQR et même les boxplots).

**Question 3 : Minutes de pénalité par match**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **RW** | **LW** | **C** | **D** |
| **Taille d'échantillon** | | 162 | 163 | 258 | 303 |
|  | **Moyenne** | 0.62 | 0.75 | 0.46 | 0.58 |
|  | **Médiane** | 0.42 | 0.50 | 0.36 | 0.46 |
|  | **Coef. d’asymétrie** | 2.32 | 1.89 | 4.28 | 2.39 |
|  | **Étendue** | 3.86 | 4.25 | 5 | 3.69 |
|  | **Écart-type** | 0.72 | 0.74 | 0.47 | 0.50 |
|  | **IQR** | 0.52 | 0.83 | 0.41 | 0.48 |



On constate :

1. Une asymétrie à droite pour les 4 positions qui peut être constaté par la valeur des coefficients d’asymétrie qui sont tous positifs mais aussi en regardant les boxplots et les histogrammes (voir le fichier Excel).
2. Les moyennes sont sensiblement plus élevées que les médianes, ce qui est cohérent avec l’asymétrie à droite.
3. Les valeurs extrêmes à droite (joueurs très punis) font que les étendues sont très élevées par rapport aux IQR. Ces valeurs viennent tirer les moyennes vers le haut.
4. Les boîtes des défenseurs, ailiers droits et centres sont toutes situées à droite de 1. Ainsi, plus de 75% des joueurs obtiennent moins d’une minute de pénalité par match.
5. Les ailiers gauches sont les plus punis (moyennes et médianes plus élevées) et aussi boîte un peu plus à droite (3e quartile dépassant légèrement 1).

## Exercice 3

**Question 1**

|  |  |  |
| --- | --- | --- |
|  | **Table 1** | **Table 2** |
| **Nom de la Variable** | DV\_CM | DV\_LC |

**Question 2**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** | **E** | **F** |
| **Table 1** | **120** | **106** | **0.314** |  |  |  |
| **Table 2** |  |  |  | **70** | **340** | **0.92** |

**Question 3**

|  |  |  |
| --- | --- | --- |
| **K** | **G** | **H** |
| **166** | **200** | **236** |