



DUT STID, 1^{ème} année
Statistique descriptive I
Devoir du mercredi 6 novembre 2013

Nom : _____/34

Consignes

- Les réponses sont à donner directement sur le sujet. N'oubliez pas de noter votre nom.
- Toute réponse doit être précisément justifiée. Les réponses insuffisamment justifiées ne donneront droit à aucun point.
- *Matériel autorisé* (à l'exclusion de toute autre chose) : crayons, calculatrices (pas d'ordinateur, pas de téléphone portable), cerveau (pour ceux qui en possèdent un). **Les téléphones portables sont formellement interdits sur les tables, sur vos genoux, dans vos poches : ils doivent être déposés, avec vos sacs, à côté de mon bureau.**
- Les deux exercices sont indépendants ainsi que la plupart des questions à l'intérieur des exercices.
- Il est formellement interdit de parler (même en langage des signes et même pour demander une gomme, un crayon, etc à son voisin).

Exercice 1 /17

Cet exercice utilise les données **precip** extraites du livre : McNeil, D.R. (1977) *Interactive Data Analysis*. New York : Wiley. Elles contiennent les précipitations annuelles, en pouces, de 70 villes des États-Unis :

Mobile : 67,0 ; Juneau : 54,7 ; Phoenix : 7,0 ; Little Rock : 48,5...

(pour simplifier, seules les premières observations sont reproduites dans l'énoncé). On donne :

- le total des précipitations sur l'ensemble des 70 villes est 2 442 pouces ;
- la somme des carrés des précipitations par ville est 98 154,1 pouces².

La plupart des questions suivantes sont indépendantes :

1. Quelle est la population étudiée, quelle est sa taille ?

2. Quelle est la variable étudiée, quel est son type ?

3. Quelle est la moyenne des précipitations sur ces 70 villes ?

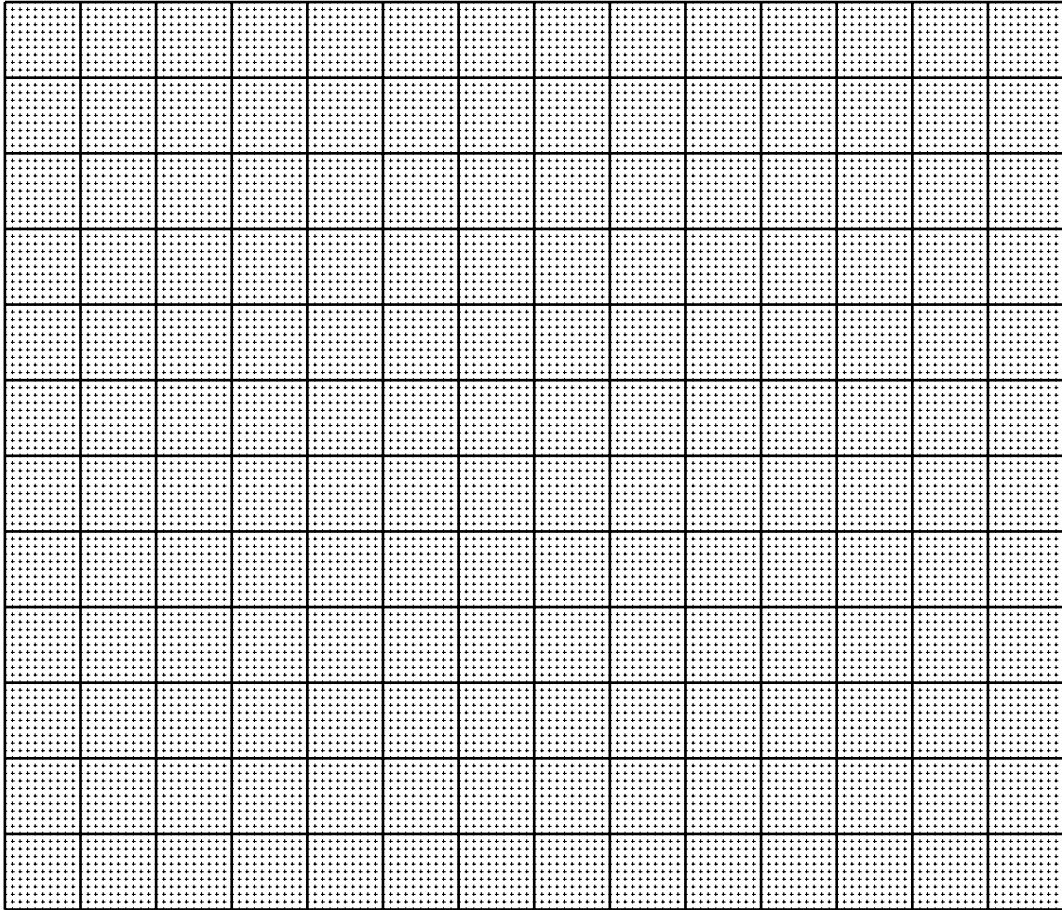
4. Quel est l'écart type des précipitations sur ces 70 villes ?

Dans la suite de l'exercice, on utilisera le regroupement en classes suivant des données :

précipitation	[0,20[[20,30[[30,35[[35,40[[40,45[[45,70[
effectifs	13	5	11	14	13	14

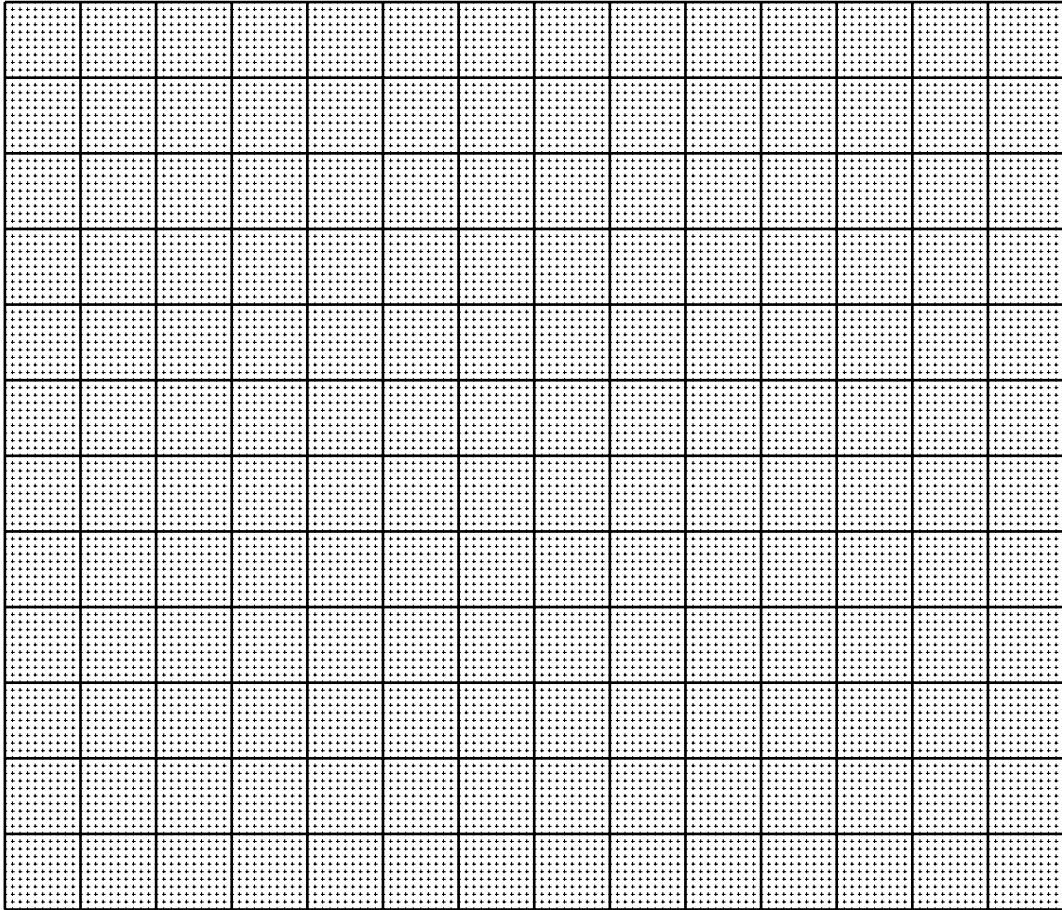
5. À partir du regroupement en classes, recalculer la moyenne des précipitations sur les 70 villes. Comparer à la valeur trouvée dans la question 3 : que peut-on en conclure ?

6. Sur le quadrillage suivant, construire l'histogramme des données selon le regroupement en classes données précédemment. *On n'oubliera pas de donner les détails de la construction.*

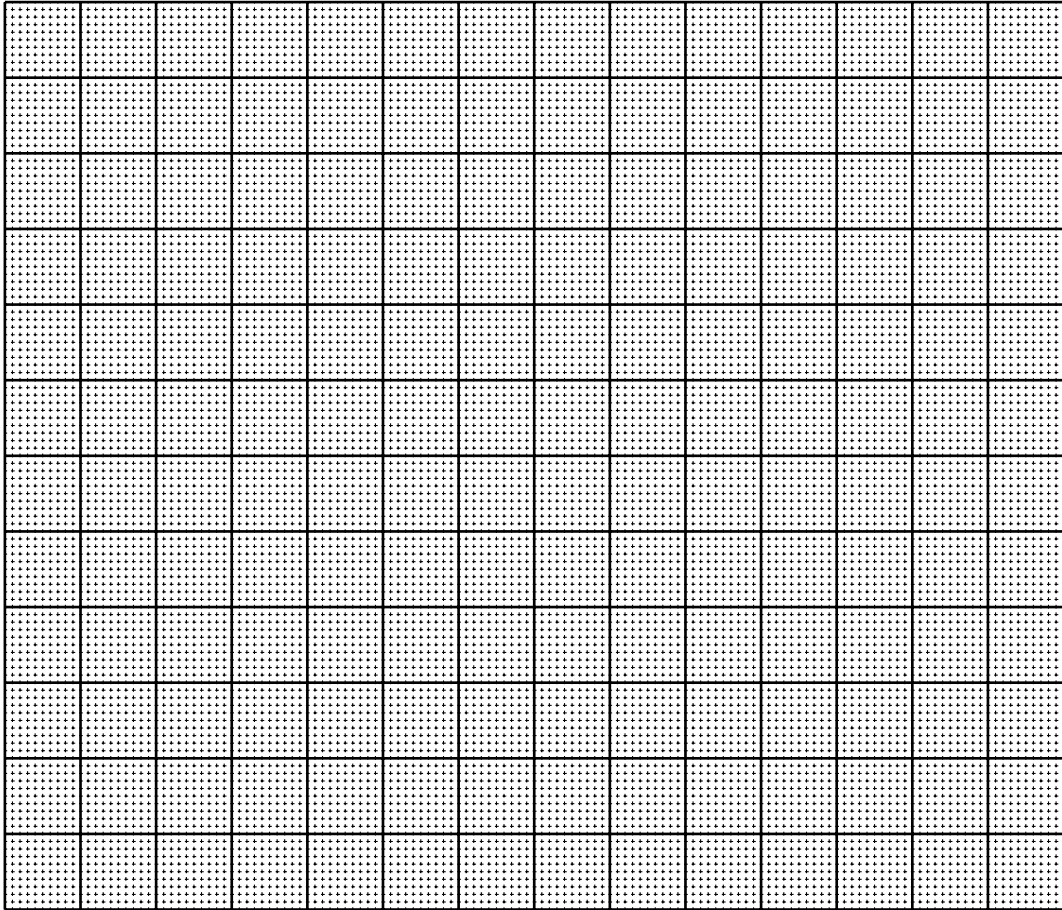


7. À partir du regroupement en classes, calculer la médiane. Comparer la valeur de celle-ci à la moyenne et commenter.

8. Sur le quadrillage suivant, construire le polygone cumulatif et retrouver graphiquement la valeur de la médiane. *On fera clairement apparaître les traits de construction et où la valeur de la médiane est lue.*



9. Déterminer la courbe de Lorenz de cette distribution statistique sur le quadrillage ci-dessous. Commenter son allure. *On n'oubliera pas de donner les détails de la construction.*



10. Compléter la phrase suivante : 1/5 des villes américaines les plus pluvieuses totalisent _____ des précipitations des 70 villes étudiées.

Espace supplémentaire (au besoin)

Exercice 2/17

La plupart des questions de cet exercices sont indépendantes.

1. Entourer la commande permettant d'importer dans R le fichier de données au format texte partiellement reproduit ci-dessous :

```
"Nom ", "Gout ", " Amertume ", "Soif ", "TxAlcool ", " Ferment  
""109""", 3.0,0.0,1.5,6.0, "Haute ", "Ambrée", "France", "Non", "S:  
"1356 Jean le Bon ", 2.5,0.5,1.0,5.6, "Haute ", "Blonde", "France'  
"1664 Gold ", 1.0,1.0,2.0,6.1, "Basse ", "Blonde", "France", "Non".  
""1845""", 3.0,1.5,0.5,6.3, "Haute ", "Autre", "Iles britanniques:  
"732 Charles Martel ", 2.5,0.5,1.0,5.6, "Haute ", "Blonde", "Fran  
"A.K. Damm ", 2.0,0.5,2.0,4.8, "Basse ", "Blonde", "Espagne", "Non'  
"Abbaye d'Aulnes Blonde 6 ", 1.5,1.0,1.5,6.0, "Haute ", "Blonde".  
"Abbaye d'Aulnes Blonde 8 ", 2.0,1.0,0.5,8.0, "Haute ", "Blonde".  
"Abbaye d'Aulnes Brune 6 ", 3.0,1.0,0.5,6.0, "Haute ", "Brune", "f  
"Abbaye d'Aulnes Brune 8 ", 3.5,0.5,0.5,8.0, "Haute ", "Brune", "f  
"Abbaye d'Aulnes Val de Sambre ", 2.5,1.0,0.5,7.0, "Haute ", "Ami  
"Abbaye de Gembloux ", 1.5,0.5,0.5,8.0, "Haute ", "Blonde", "Belg:  
"Abbaye de St Landelin Spéciale ", 2.0,0.0,0.5,6.8, "Haute ", "Ar  
"Abbaye des Roccs Brune ", 4.0,0.5,0.0,9.0, "Haute ", "Brune", "Be  
"Abbaye des Roccs Grand Cru ", 4.0,1.0,0.5,10.0, "Haute ", "Brune'  
"Abbaye des Roccs Spéciale ", 4.0,1.0,0.0,9.0, "Haute ", "Brune", '  
"Abbaye du Val Dieu Grand Cru (l') ", 4.0,1.0,0.0,10.5, "Haute '  
"Abbaye du Val Dieu (l') ", 3.5,0.0,2.0,7.0, "Haute ", "Blonde", '  
"Abdis Brune ", 2.0,0.5,0.0,6.5, "Haute ", "Brune", "Belgique", "D
```

```
donnees <- read.table("dataset.txt", header=FALSE, sep=",", dec=".")  
donnees <- read.table("dataset.txt", header=TRUE, sep=",", dec=".")  
donnees <- read.table("dataset.txt", header=FALSE, sep=".", dec=",")  
donnees <- read.table("dataset.txt", header=TRUE, sep=".", dec=",")
```

La suite des questions porte sur le fichier de données C02 qui est disponible dans R en tapant `data(C02)`. Il n'est pas nécessaire de savoir ce que contient ce fichier de données pour répondre aux questions. La commande `summary(C02)` donne

```

      Plant           Type      Treatment      conc      uptake
Qn1   : 7   Quebec      :42   nonchilled:42   Min.    : 95   Min.    : 7.70
Qn2   : 7   Mississippi:42   chilled  :42   1st Qu.: 175  1st Qu.:17.90
Qn3   : 7                                     Median : 350  Median :28.30
Qc1   : 7                                     Mean   : 435  Mean   :27.21
Qc3   : 7                                     3rd Qu.: 675  3rd Qu.:37.12
Qc2   : 7                                     Max.   :1000  Max.   :45.50
(Other):42

```

2. Quel est le type de chacune des variables du fichier de données (selon la typologie des variables vue en cours). *Remarque : lorsqu'une ambiguïté sur le sous-type de la variable existe, on ne donnera pas ce sous-type.*

3. Quelle est la taille de la population de ce fichier de données ?

4. Quelle valeur renvoie la commande `quantile(C02$conc,0.75)` ?

5. Quelle est la moyenne de la variable `conc` ? Quelle est sa médiane ? Commenter la différence.

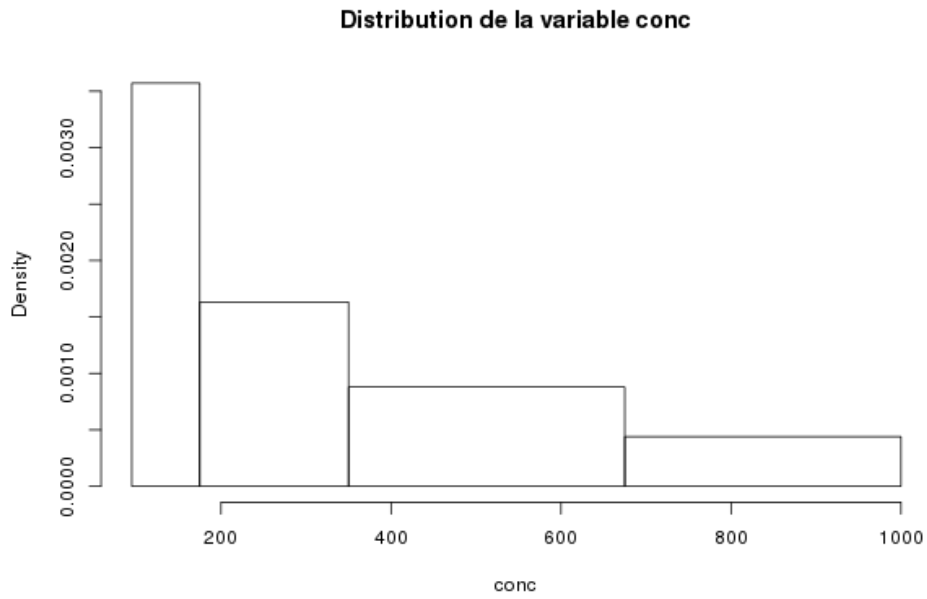
6. Quelle commande permet d'obtenir l'écart type de la variable `uptake` ?

7. Quelle commande permet d'obtenir le tableau d'effectifs complet de la variable `Plant` ?

8. Quelles sont les commandes qui permettent de stocker dans une nouvelle variable du jeu de données, que l'on nommera `uptakeC`, le découpage en classes de la variable `uptake` délimitées par les bornes 7, 20, 30, 40 et 50 puis d'obtenir le tableau d'effectifs de ces classes ?

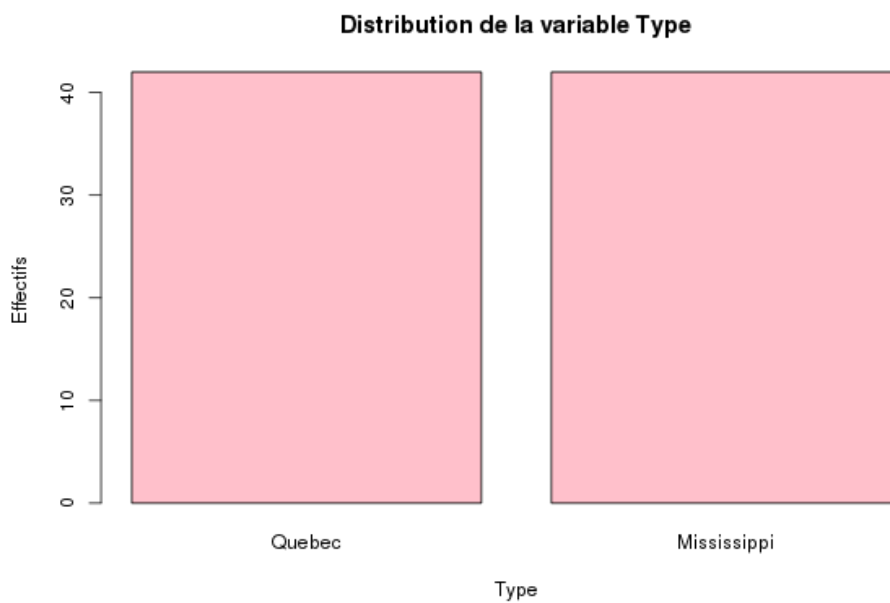
9. Quelle commande permet de stocker dans une nouvelle variable du jeu de donnée, que l'on nommera `uptakeC2`, le découpage en 5 classes de même amplitudes de la variable `uptake` ?

10. L'histogramme ci-dessous est celui de la variable `conc`.



Quelle commande permet de l'obtenir? (Les bornes des classes sont 95, 175, 350, 675 et 1 000.)
 Commenter la forme de la distribution.

11. Le diagramme en tuyau d'orgues ci-dessous est celui de la variable Type.



Quelle commande permet de l'obtenir ? (La couleur des tuyaux d'orgue est le rose.)