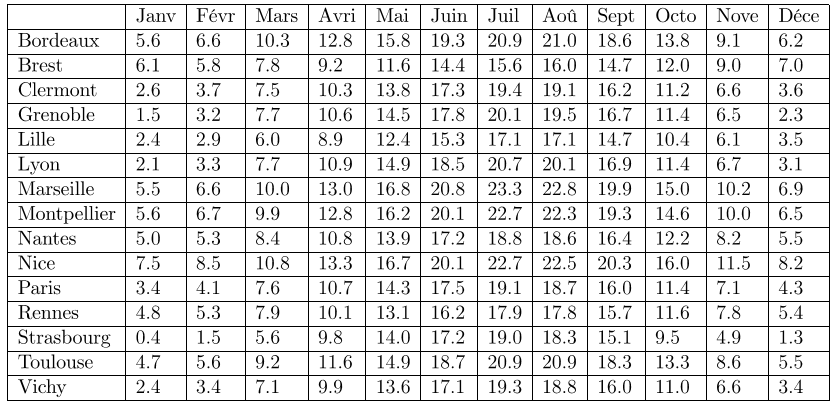
**TP : Analyse en composantes principales**

Moyennes mensuelles des températures pour 15 villes françaises.

**1 Description des données**

On dispose, pour 15 villes de France, des moyennes mensuelles de températures calculées sur 30 ans (de 1931 `a 1960). Ces données sont rassemblées dans la Table ci-jointe qui croise les 15 villes (lignes) et les 12 mois de l’année (colonnes).



Tab. Températures mensuelles de 15 villes de France.

**2 Travail à effectuer**

Le but de l’´etude est de comparer les températures mensuelles des différentes villes. L’analyse en composantes principales (ACP) sur ce tableau devra préciser les points suivants. Il s’agit entre autre de :

— Réaliser une typologie des villes, c’est-`a-dire celles qui se ressemblent et celles qui diﬀèrent du point de vue des températures mensuelles ;

— Proposer un bilan des liaisons entre les variables et, dans la mesure du possible, résumer approximativement l’ensemble des variables par un petit nombre de variables synthétiques, c’est-`a-dire non pas extraites du tableau mais les combinant ;

— Etudier si les ressemblances ou les dissemblances correspondent à des proximités ou des éloignements géographiques.

**2.1 Question de cours**

1. Comment mesure-t-on la dissemblance entre les individus ?

2. Le nuage est toujours centré. Pourquoi ? Quel est l’eﬀet du centrage sur l’analyse du nuage des individus ? Quel type de liaisons évalue-t-on dans le nuage des variables ?

3. Montrer que si on ne réduit pas les variables, on accorde aux variables un poids égal à leur écart-type. La liste des écart-types permet-elle, a priori, de prévoir une diﬀérence signiﬁcative entre une ACP normée et une ACP non-normée ?

**2.2 Analyse des données**

1. Tracer l’´evolution de température pour les diﬀérentes villes (graphes supperposés). Que remarquez vous ?

2. Centrer et réduire les données et tracer une fois que plus l’´evolution des températures. peut-on reconnaître des groupes de villes ?

**2.3 Analyse en composantes principales**

Tous les calculs suivant se feront sur des données centrée et réduites.

1. Estimer l’impact de chaque variable si les données n’avaient pas été réduites.

2. Analyser les statistiques descriptives de chacune des 16 variables.

3. Analyser la matrice des corrélations. Quelles sont les variables les plus corrélées ?

4. Analyser les valeurs propres. Quel est le pourcentage d’inertie expliquée par chaque axe ? Combien d’axes peut-on conserver ?

5. Quelle est la première composante principale ? Etudier la représentation des individus sur le premier axe. Quels sont les points dont la contribution au premier axe est la plus grande ? Quelles sont leurs coordonnées ? Comparer les résultats avec l’interprétation de la première composante principale obtenue.

6. Reprendre le travail de la question précédente avec la seconde composante principale.

7. Visualiser les villes sur le plan 2d contenant les deux axes principaux.

8. Donner une interprétation globale du premier cercle des corrélations (prendre en compte l’apport des variables illustratives que sont la latitude et la longitude des villes). Proposer une typologie des 15 villes `a partir de ces résultats.