

TD 3 - Etude bivari ee

But du TD : Coefficient de corr elation, variance inter-classe, variance intra-classe,  equation d'analyse de la variance.

1 Etude des donn ees women

- Charger le dataframe `women` (`W<-women`) dans l'espace de travail et d ecrire les variables en jeu.
- Repr esenter sur un graphique l' evolution conjointe des deux variables, donner un nom au graphique et aux axes (commandes `plot`, `names`, `main`). Commenter le graphique.
- Calculer le coefficient de corr elation entre les deux variables ( a l'aide de `cor`). Commenter le r esultat et comparer avec l'intuition apport ee par le graphique.

2 Etude des donn ees iris

Charger le dataframe `iris` dans l'espace de travail (`D<-iris`) et d ecrire les variables en jeu.

2.1 Analyse de variance

Le terme *analyse de variance* se r ef ere  a l' etude de l'effet d'un caract ere qualitatif (ici l'esp ece d'iris) sur un caract ere quantitatif. On  etudie ici la relation entre la vari et e et la variable `Sepal.Length`. Nommer `espece` la colonne de iris correspondant  a l'esp ece et `sepale` la colonne correspondant  a la colonne `Sepal.Length`.

- Calculer la variance totale de `sepale` : $VT = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$. Utiliser la fonction `var` (avec pr ecaution).
- Calculer la variance inter-classe sur l'ensemble des k classes (ici, $k = 3$) : $V_{inter} = \sum_{j=1}^k \frac{n_j}{n} (\bar{x}^j - \bar{x})^2$. On utilisera la fonction `tapply`.
- Calculer la variance sur chacune des k classes : $V_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (x_i^j - \bar{x}^j)^2$ (en utilisant `tapply`). Commenter.
- En d eduire la variance intra-classe : $V_{intra} = \sum_{j=1}^k \frac{n_j}{n} V_j$.
- V erifier et retrouver num eriquement l' equation d'analyse de la variance : $VT = V_{inter} + V_{intra}$.

2.2 Analyse de corr elations

On s'int eresse maintenant  a l'analyse de corr elations entre les variables `Sepal.Length` et `Petal.Length` (on nommera la colonne correspondante `petale`).

- Comme pr eliminaire, r ealiser le graphique suivant :

```
x=D[,1:4]
y=D[,5]
i=1;j=3;
sepale<-x[,i]; petale<-x[,j];
plot(sepale,petale,main='iris study', xlab=names(x)[i],ylab=names(x)[j]);
ind=(y==levels(y)[2]);
points(sepale[ind],petale[ind], col='green');
ind=(y==levels(y)[3]);
points(sepale[ind],petale[ind], col='blue');
```

Comprendre le code (en particulier, la fonction `levels`). Expliquer ce que repr esente le graphique obtenu.

- Calculer la corr elation totale entre `sepale` et `petale`, puis la corr elation  a l'int erieur de chaque classe (esp ece). Quelles conclusions en tirer ?
- Calculer la corr elation inter-classe. Conclure.