

Fiche 1 - R - M1 Maths-fi

Tran Viet Chi, `chi.tran@math.univ-lille1.fr`, bureau 316 (bâtiment M3).

Exercice 1 (Loi normale bivariée)

On se donne, par simulation, une suite de couples i.i.d. $(X_i, Y_i)_{i \in \{1, \dots, n\}}$ tels que :

$$(X_i, Y_i) \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

avec $\rho \in [-1, 1]$ et on souhaite représenter graphiquement les points correspondants.

1. Simuler deux vecteurs indépendants X et Z de $n = 1000$ v.a. $\mathcal{N}(0, 1)$ indépendantes en utilisant `rnorm`.
2. Montrer (sur papier) que $Y = \rho X + \sqrt{1 - \rho^2} Z$ suit une loi $\mathcal{N}(0, 1)$ et que $\text{Cov}(X, Y) = \rho$.
3. Pour $\rho = 0.75$, créer Y avec R. Calculer $\text{Var}(Y)$ et $\text{Cov}(X, Y)$.
3. Représenter graphiquement les couples (X_i, Y_i) pour différentes valeurs de ρ entre -1 et 1 .

Exercice 2 (Des hommes et des femmes)

Les données du fichier `salaires.xls` fournissent le salaire, le sexe, la catégorie socio-professionnelle (CSP) et le nombre de jours d'absence pour chaque salarié d'une entreprise.

1. Télécharger le fichier `salaires.csv` depuis la page :

`http://labomath.univ-lille1.fr/~tran/enseignements.html`

Puis, importer ce fichier sous R à l'aide des commandes :

```
chemin<-"G:/Enseignements/TISD/TD sas/Donnees/"
salaires<-read.table(paste(chemin,"salaires.csv",sep=""),header = TRUE, sep = ";")
```

où il faut remplacer `chemin` par l'emplacement dans lequel vous avez enregistré votre fichier.

2. Nous nous intéressons au nombre de jours d'absence. Obtenir les moyennes, écarts-type et variances pour les hommes, pour les femmes et pour l'ensemble. Combien y a-t-il d'hommes, de femmes ?
3. A partir des données précédentes, calculer la variance inter et la variance intra. Conclusion ?
4. On s'intéresse maintenant aux salaires. Créer un découpage des salaires comme suit :
 - on crée une variable auxiliaire *groupe* qui vaut 1 si l'individu gagne moins que 11450, 2 s'il gagne entre 11 450 et 15 650, 3 s'il gagne entre 15 650 et 39 150 et 4 sinon.
 - calculer l'effectif, la moyenne et la variance des salaires dans chacun des groupes.
5. On suppose qu'on ne dispose maintenant plus que des données résumées par classe. Retrouver la moyenne et la variance.
6. Supposons qu'on ne dispose plus que du nombre d'employés par classe. Donner une approximation de la moyenne et de la variance.

Exercice 3 (Passagers du Titanic)

1. Les données sur les passagers du Titanic, fournies par **R**, peuvent-être chargées par la commande `data(Titanic)`. Demander la description de cette table par

```
help("Titanic")
```

2. Explorer la table (en particulier sa structure, demander `dim(Titanic)`) et décrire les différentes variables. Calculer les fréquences marginales et tracer leurs diagrammes en secteur.
3. Obtenir la table de contingence du croisement de la variable `Survived` avec la variable `Class`. Calculer les indicateurs de liaison χ^2 , Φ^2 et coefficient de Cramer avec les commandes `chisq.test`.
4. Faire de même avec le croisement de `Survived` et `Sex` et `Survived` et `Age`. Avec quelle variable `Survived` est-elle le plus liée ?
5. Recommencer pour le croisement de `Survived` avec `Class` par sexe. La liaison de ces deux variables est-elle plus forte pour les hommes ou pour les femmes ?