DM - Master Pro

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

Données : Les fichiers `cac40.sas` et `salaires.xls` sont disponibles à l'adresse `http://math.univ-lille1.fr/~tran/enseignements.html`.

Rque : j'attends des commentaires pertinents et pas une simple obtention des chiffres demandés. Toute remarque bienvenue sera appréciée, même si elle n'était pas exigée dans la question !

Exercice 1 (Etude d'une série financière)
Le CAC40, qui prend son nom du système de *Cotation Assistée en Continu*, est le principal indice boursier de la place de Paris. Créé avec 1 000 points de base au 31 décembre 1987 par la Compagnie des agents de change, il est déterminé à partir des cours de quarante actions cotées en continu parmi les cent sociétés dont les échanges sont les plus abondants sur Euronext Paris. Ces actions, représentatives des différentes branches d'activités, reflètent en principe la tendance globale de l'économie des grandes entreprises françaises et leur liste est revue régulièrement pour maintenir cette représentativité. Le CAC40 est publié du lundi au vendredi de 9 h 00 à 17 h 30 et mis à jour toutes les 15 secondes. On utilise ici la table `cac40.sas7bdat`.

Partie A : Description générale des données

1. Tracer graphiquement, et sur un même graphique les évolutions des cours en ouverture, en cloture, les maxima et minima. On pourra préalablement créer une variable t indiquant le numéro de l'observation.

2. A l'aide d'une `PROC UNIVARIATE`, sortir des statistiques résumant la distribution de l'indice. La réponse sera présentée sous forme d'un paragraphe de quelques lignes où seules les informations pertinentes seront mises en valeur.

3. Tracer l'histogramme de la variable `ouverture`.

4. Peut-on supposer que cette variable suit une loi log-normale ? Détaillez le test et expliquez comment votre conclusion est tirée.

Partie B : Etude des rendements

Pour un cours $(x_t; t \in \{1, \ldots, T\})$, on définit les rendements par :

$$r_t = \log\big(\frac{x_t}{x_{t-1}}\big), \qquad \text{rendements logarithmiques}$$
$$\rho_t = \frac{x_t - x_{t-1}}{x_{t-1}}, \qquad \text{rendements arithmétiques.}$$

5. Exprimer pour chacun de ces rendements le cours x_t au temps t en fonction de x_0 et des valeurs des rendements aux temps intermédiaires $(r_s; s \in \{0, \ldots, t\})$ et $(\rho_s; s \in \{0, \ldots, t\})$.

6. Calculer les rendements logarithmiques et arithmétiques pour le cours `ouverture`. On pourra créer au préalable une variable auxiliaire $y_t = x_{t-1}$ en utilisant `y=lag(ouverture)`.

7. Tracer l'évolution de ces rendements au cours du temps. Que pensez-vous de ces séries en terme de stationnarité ?

8. Décrire les distributions de ces rendements. Encore une fois, la réponse sera sous forme d'un paragraphe de quelques lignes où seules les informations pertinentes seront mises en valeur.

9. Les rendements sont-ils gaussiens ? Ceci était-il prévisible, en particulier pour le rendement logarithmique ?

10. Comparez les deux séries de rendements.

Partie C : Autres indicateurs financiers

11. Pour un cours $(x_t; t \in \{1, \ldots, T\})$, on définit les indicateurs financiers suivants qu'on vous demande de calculer et d'analyser (en quelques brèves lignes) :

1. le rendement annualisé est obtenu à partir des rendements arithmétiques mensuels ρ_t, en utilisant le caractère géométrique obtenu à la question **5.** et $\bar{\rho}$ le rendement mensuel moyen :
$$\rho_{an} = (1 + \bar{\rho})^{12} - 1.$$
2. le *drawdown* (perte maximale historique) donne une indication sur la perte depuis le dernier pic : au temps t, il est donné par
$$D_t = \frac{\max_{0 \leq s \leq t} x_s}{x_t}.$$
3. Soit $\alpha \in [0, 1]$. La *Value at Risk* (VaR) est le quantile $1-\alpha$ de la distribution des rendements arithmétiques, c'est-à-dire la valeur telle que : $\alpha = \mathbb{P}(r > \text{VaR}(\alpha))$.

Exercice 2 (Discrimination salariale hommes-femmes)
Les données du fichier `salaires.xls` contiennent les salaire, sexe, catégorie socio-professionnelle (CSP) et nombre de jours d'absence pour chaque salarié d'une entreprise. On s'intéresse plus particulièrement au salaire X et au sexe S.

1. Importer ce fichier sous SAS.

2. En utilisant la `PROC UNIVARIATE`, sortir des statistiques résumant les distributions de salaires par sexe. Tracer les histogrammes du salaires. La réponse sera présentée sous forme d'un paragraphe de quelques lignes où seules les informations pertinentes seront mises en valeur.

3. On étudie toujours la variable `salaire`. Obtenir en utilisant la `PROC MEANS` avec l'option `by sexe` : les effectifs, moyennes, minima, maxima, variances, pour les hommes, pour les femmes et pour l'ensemble. Récupérer ces statistiques dans une table de sortie avec l'option `output`.

4. Comparer les moyennes par sexe. Que cela vous inspire-t-il ?

On cherche à savoir quelle est la part de la variable `sexe` dans la différence observée. On note `1` les hommes et `2` les femmes, n_g l'effectif du groupe $g \in \{1, 2\}$

$$\bar{X}_g = \frac{1}{n_g} \sum_{i \text{ t.q. } S_i = g} X_i, \qquad s_g^2 = \frac{1}{n_g} \sum_{i \text{ t.q. } S_i = g} (X_i - \bar{X}_g)^2, \qquad i \in \{1, 2\}$$

les moyennes et variances du salaire dans le groupe des hommes et des femmes, et $\bar{X}$ et s^2 les moyenne et variance sur tous les individus (hommes et femmes). Notre but est de montrer que :

$$\begin{aligned} s^2 &= s^2_{\text{intra}} + s^2_{\text{inter}} \qquad (1) \\ s^2_{\text{intra}} &= \frac{n_1 s_1^2 + n_2 s_2^2}{n} \\ s^2_{\text{inter}} &= \frac{n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2}{n}. \end{aligned}$$

(1) montre que l'on peut décomposer la variance en deux termes. Le terme s^2_{intra} est la moyenne des variances de chaque groupe, qui représentent la variation du salaire indépendamment du sexe puisque tous les individus d'un même groupe ont même sexe. Le terme s^2_{inter} est la variance que l'on aurait obtenue si le salaire de tous les hommes était $\bar{X}_1$ et le salaire de toutes les femmes $\bar{X}_2$. C'est donc au contraire la variabilité du salaire qui ne dépend que du sexe. La part de la variabilité du salaire due au sexe est donc :

$$\rho = \frac{s^2_{\text{inter}}}{s^2} = \frac{s^2_{\text{inter}}}{s^2_{\text{intra}} + s^2_{\text{inter}}}.$$

5. Etablissons (1) :
5.1. En partant de la définition de s^2 séparer les sommes correspondant à $S_i = 1$ et $S_i = 2$.
5.2. Utiliser que $X_i - \bar{X} = (X_i - \bar{X}_g) + (\bar{X}_g - \bar{X})$ pour obtenir (1).

6. Sur les données, calculer la variance inter, la variance intra, ρ. On utilisera des `PROC MEANS` et les tables créées en question **3**. Conclusion ?