
PROJET 5 - TASK PLANNING

OBJECTIF

- Cibler plus particulièrement certains pays, dans le but d'approfondir ensuite l'étude de marché.
- Idéalement, produire des "groupes" de pays, plus ou moins gros, dont on connaît les caractéristiques.

La stratégie est plutôt d'exporter les produits plutôt que de produire sur place, c'est-à-dire dans le(s) nouveau(x) pays ciblé(s).

Préambule : Choix des variables (4 imposées + 3/4 personnelles)

1. Récupération données FAO (nom des zones & critères) → 1df

2. Récupération données WBD (nom des zones & critères) → 2df

3. Création Table de Correspondance Pays (Codes ISO-3166, FAO, WBD) → 1 df

4. Import Data Notebook & Construction Dataframe 1 - 4 variables du cours

5. Ajout 3/4 variables :

- Zone calc Fao → Qté Volaille (Kg / Hab) = (Qté Prod + Qt_Import – Qt Export) / population_pays
- Trade_value_france_pays_prospect_\$ (wbd)
- PIB/h ou PIB tot 2013 (wbd - cnucead)
- Goods_Import (% PIB) (wbd – data eco)

6. Nettoyage des données résiduelles

Explications rejets certains pays (pas de datas pour 2010/2013, pas de ratio, etc...), traitement des NA & valeurs aberrantes

7. Réalisation Dendrogramme - sur n pays (174 a priori)

Export Image png

8. Découpage en « n » clusters/groupes

Export fichier groupes.csv

9. Calcul des Centroïdes par Groupe/Clusters

Export fichier centroides.csv

10. Critique/Analyse/Comparatif des Valeurs centroides

Exemple

#c1	PIB/h très faible			
#c2	PIB/h élevé	/ croissance demo assez forte	/ 3eme Var (Qt.Volailles) forte	/ 4eme Trade_Value\$?
#c3	PIB/h élevé	/ croissance demo assez faible	/ 3eme Var (Qt.Volailles) faible	/ 4eme Trade_Value\$?
#c4	PIB/h moyen	/ croissance demo assez forte	/ 3eme Var (Qt.Volailles) très forte	/ 4eme Trade_Value\$?
#c5	PIB/h moyen	/ croissance demo correcte	/ 3eme Var (Qt.Volailles) faible	/ 4eme Trade_Value\$?

Choix des groupes #c3 & #c5 par exemple

11. Affinage sur une 10aine de pays dans les #C3/#C5

Affichage par tri sur les critères qui nous intéressent
(forte Trade value, faible Qt.volaille dispo, fort PIB, forte croissance demo)

12. Visualisez vos partitions dans le premier plan factoriel obtenu par ACP

Utilisation du module "functions.py" issu du cours contenant les fonctions suivantes

- Centrage/réduction données (`preprocessing.fit.transform`)
- Projections des individus sur 1^{er} plan (`display_factorial_planes`)
- Cercle des corrélations (`display_circles`)
- Eboulis des valeurs propres – scree plot

13. Test d'adéquation à la Loi Normale - Avec Shapiro-Wilk

Variables des Calories par exemple (`qt_disp_alim_(kcal_p_j)`)
`stats.shapiro(df["var"])`

14. Test Comparaison de 2 variables (cas gaussien)

D'abord trouver une variable qui suit une loi normale dans 2 groupes différents (#c1, #c2)

- On fait deux fois shapiro `stats.shapiro(#c1["var"])` `stats.shapiro(#c2["var"])`

H0 = La variable suit une loi normale

H1 = La variable ne suit donc pas une loi normale

Fixons le risque standard à 5% (0.05) voir 1% (0.01)

Si la p-value > alpha => on ne rejette pas H0, donc la variable suit une loi normale

On les compare

- Avec Bartlett `stats.bartlett(#c1["var"], #c2["var"])`
Hypothèse nulle (H0) => Egalité des variances
Si la p-value est élevée => on ne rejette pas H0 donc égalité des variances
- Avec T Test (Student) `stats.ttest_ind(#c1["var"], #c2["var"], equal_var=True)`
Hypothèse nulle (H0) => Egalité des moyennes
Si la p-value est < risk => Rejet de l'hypothèse H0 de l'égalité des moyennes.

Exemples de tests shapiro :

data: `rnorm(100, mean = 5, sd = 3)`
W = 0.9895, **p-value = 0.6211**

L'exemple ci-dessus renvoie une p-value non significative. L'échantillon suit donc une loi normale.

data: `runif(100, min = 2, max = 4)`
W = 0.9337, p-value = 8.077e-05

Dans l'exemple 2, ci-dessus, la p-value est significative. L'échantillon ne suit donc pas une loi normale.