

TP LANGAGE R

Le fichier ventes.csv contient des données indiquant les chiffres d'affaires réalisés par 440 antennes d'une chaîne de distribution, pour divers types de produits :

« ProduitsFrais », « Lait », « Epicerie », « Surgele », « Detergents » et « Traiteur ».

Importer ce fichier dans R et vérifier son bon chargement avec la fonction `str`.

Partie 1 :

Supprimer la colonne « Canal », puis ajouter successivement à la droite du tableau :

- une colonne « Total » qui indique la somme des chiffres d'affaires des colonnes « ProduitsFrais », « Lait », « Epicerie », « Surgele », « Detergents » et « Traiteur » ;
- une colonne « % ProduitsFrais » qui indique les pourcentage (arrondis à l'entier près) de chiffre d'affaires des produits « ProduitsFrais » par rapport au « Total », et ainsi de suite pour tous les produits.

Le début de celui-ci devrait être :

	Region	ProduitsFrais	Lait	Epicerie	Surgele	Detergents	Traiteur	Total	% ProduitsFrais	% Lait	% Epicerie	% Surgele	% Detergents	% Traiteur
1	3	12669	9656	7561	214	2674	1338	34112	37	28	22	1	8	4
2	3	7057	9810	9568	1762	3293	1776	33266	21	29	29	5	10	5
3	3	6353	8808	7684	2405	3516	7844	36610	17	24	21	7	10	21
4	3	13265	1196	4221	6404	507	1788	27381	48	4	15	23	2	7
5	3	22615	5410	7198	3915	1777	5185	46100	49	12	16	8	4	11
6	3	9413	8259	5126	666	1795	1451	26710	35	31	19	2	7	5

Exporter ce tableau avec l'instruction :

```
write.csv2(x, file = "TPR.csv")
```

où `x` désigne le nom que vous avez donné à votre tableau final dans R.

Enfin, afin de comparer la performance de chaque région, afficher dans la console R, pour chacune des régions, le chiffre d'affaire « Total » moyen :

CA Total moyen Region 1 : 30997.57 euros

etc.

Partie 2 :

Dans cette deuxième partie, on travaillera sur le tableau original (celui qui contient les données du fichier ventes.csv).

On demande :

- de déterminer la paire de variables qui a le meilleur coefficient de corrélation linéaire, parmi les variables produits (« ProduitsFrais », « Lait », « Epicerie », « Surgele », « Detergents » et « Traiteur ») ;
- d'effectuer les deux régressions linéaires sur cette paire de variables (Y par rapport à X et X par rapport à Y) ;
- d'effectuer une prévision, avec chacune de ces régressions, lorsque la valeur d'une des deux variables est égale à sa valeur moyenne (autrement dit, prévoir Y pour la valeur moyenne de X et prévoir X pour la valeur moyenne de Y) ;
- d'afficher le nuage de points et la droite de régression correspondants à l'une des deux régressions effectuées.

Bien sûr, le script devra fonctionner d'un bout à l'autre sans intervention manuelle sur la console R.

Livrable : votre script au format .r (fonctionnel, efficace et agréable à lire) à envoyer à jc.duchain@wanadoo.fr