PLANS ET ANALYSES D'EXPÉRIENCES
STT 3410

Devoir 2 : à remettre le mercredi 10 novembre à 8h30.

1. Analyse des infections nosocomiales en fonction des régions aux États-Unis :

Une étude sur l'efficacité du contrôle des infections nosocomiales (infections causées par l'hospitalisation) a été effectuée en 1975-76 dans 113 hôpitaux américains. Les données sont disponibles à la page web
`http://www.dms.umontreal.ca/~leger/STT3410/senic.html`. Il y a 12 variables.

Variable	Nom	Description
1	Code d'identification	1–113
2	V2	
3	V3	
4	Risque d'infection	Probabilité estimée d'attraper une infection à l'hôpital (en pourcent)
5	V5	
6	V6	
7	V7	
8	V8	
9	Région	1 = Nord-est, 2 = Centre-nord, 3 = Sud, 4 = Ouest
10	V10	
11	V11	
12	V12	

Analysez le risque d'infection en fonction de la région. Les hypothèses de base du modèle semblent-elles satisfaites ? Y a-t-il des différences entre les groupes ? Si oui, entre quels groupes ? Donnez-moi les résultats de l'analyse sous forme d'un rapport à un client qui n'est pas nécessairement un statisticien avec les sorties pour supporter vos conclusions soigneusement organisées dans une annexe.

2. Attention aux logiciels statistiques... :

Considérez les données de Kenton vues en classe dont les valeurs pour le troisième groupe ont été légèrement modifiées (la même constante a été ajoutée à chaque valeur). Elles sont disponibles à la page web
`http://www.dms.umontreal.ca/~leger/STT3410/kenton-mod.html`. La première colonne contient le nombre de caisses vendues, la deuxième identifie le design alors que la troisième identifie le magasin (cette colonne ne sera pas utilisée). (Oui, je sais, le nombre de caisses n'est plus un entier pour le troisième groupe).

a) À l'aide de la procédure `anova` de SAS, obtenez des intervalles de confiance de niveau 95% pour la différence de chacune des paires de moyennes selon chacune des méthodes suivantes : Scheffé, Tukey et Bonferroni. Quels sont les groupes significativement différents ? Y a-t-il des différences de conclusion entre les différentes méthodes ?
b) En conservant un maximum de chiffres significatifs durant les calculs, vérifiez l'intervalle de confiance selon Scheffé pour la différence entre les moyennes des groupes 3 et 2. Y a-t-il une différence significative entre les moyennes des groupes 3 et 2 ?
c) L'option `lines` de l'énoncé `means` de la procédure `anova` permet de présenter les résultats des comparaisons multiples différemment. Voici une description de l'option telle que trouvée dans *SAS/STAT User's Guide* :

> presents results of the BON, DUNCAN, GABRIEL, REGWF, REGWQ, SCHEFFE, SIDAK, SMM, GT2, SNK, T, LSD, TUKEY, and WALLER options by listing the means in descending order and indicating nonsignificant subsets by line segments beside the corresponding means. The LINES option is appropriate for equal cell sizes, for which it is the default. LINES is also the default if DUNCAN, REGWF, REGWQ, SNK, or WALLER is specified, or if there are only two cells of unequal size. If the cell sizes are unequal, the harmonic mean of the cell sizes is used, which may lead to somewhat liberal tests if the cell sizes are highly disparate. LINES cannot be used in combination with the DUNNETT, DUNNETTL, or DUNNETTU options. In addition, the procedure has a restriction that no more than 24 overlapping groups of means can exist. If a mean belongs to more than 24 groups, the procedure issues an error message. You can either reduce the number of levels of the variable, or use a multiple comparison test that allows the CLDIFF option rather than the LINES option.

Utilisez l'option `lines` afin de trouver les groupes statistiquement différents au niveau 5% selon les méthodes Scheffé, Tukey et Bonferroni. Y a-t-il des conclusions différentes de celles obtenues en a) ?

d) La moyenne harmonique d'un vecteur de nombres $(x_1, \ldots, x_n)$ est $\left(n^{-1} \sum_{i=1}^{n} (1/x_i)\right)^{-1}$. Calculez la moyenne harmonique des tailles échantillonnales des quatre groupes. Reprenez le calcul de l'intervalle en b) en remplaçant les deux tailles échantillonnales par la moyenne harmonique des quatre tailles échantillonnales. Que peut-on maintenant dire de l'hypothèse de l'égalité des moyennes des groupes 3 et 2 ?