

Master M1 Finance – logiciels

TP4: Tests du χ^2

Emeline Schmisser, `emeline.schmisser@math.univ-lille1.fr`, bureau 314 (bâtiment M3).

Exercice 1 : Loi d'une variable χ^2

1. Tracer sur le même graphique la densité d'une loi du χ^2 à 1 degré de liberté et la densité d'une loi du χ^2 à 2 degrés de liberté entre 0 et 5. Utiliser la fonction `dchisq` de R.
2. Tracer sur le même graphique les densités des loi du χ^2 de 2 à 5 degrés de liberté entre 0 et 10.
3. Tracer sur un même graphique les fonctions de répartition du χ^2 à 1, 4 et 10 degrés de liberté.
4. Est-ce que ces graphiques sont cohérents avec les résultats théoriques ?

Exercice 2 : Erreurs dans un texte tapé à l'ordinateur

On admet généralement que le nombre d'erreurs dans un texte tapé suit une loi de Poisson de paramètre 4 pour 1000 mots. On observe 100 échantillons aléatoires de 1000 mots chacun et on compte le nombre d'erreurs.

| | | | | | | |
|--------------|----|----|----|----|----|----|
| nb erreurs | 0 | 1 | 2 | 3 | 4 | 5 |
| observations | 10 | 16 | 20 | 28 | 12 | 14 |

1. Est-ce que les observations suivent bien une loi $\mathcal{P}(4)$? (Faire le test à un niveau $\alpha = 10\%$).
2. Est-ce que ces données suivent une loi de Poisson ? (Faire le test à un niveau $\alpha = 10\%$).

Exercice 3 : Titanic

Les données sur les passagers du Titanic, fournies par R, peuvent être chargées grâce à la commande `data(Titanic)`. On peut afficher ces données en tapant simplement `Titanic` et on obtient de l'aide par la commande `help("Titanic")`.

1. Quelle est la dimension de la table de données Titanic ? On utilisera la commande `dim`. À quoi correspondent chacune de ces dimensions ?
2. Utiliser la commande `mosaicplot` pour représenter ces données.
3. On voudrait la table de contingence de la survie en fonction de la classe. On utilisera la fonction `apply`. Représenter ces données sur un graphique. Est-ce que on a l'impression que la survie dépend de la classe ? Comment serait le graphique si les données ne dépendaient pas de la classe ?
4. Calculer pour chaque classe, le pourcentage de gens qui ont survécu ou non.
5. Effectuer un test du χ^2 pour cette table de contingence. Pour cela, on utilise la commande `chisq.test`. On stocke le résultat de ce test dans une variable `t`. Afficher `t`. Que peut-on en conclure ? Le résultat est-il cohérent avec la première visualisation ?
6. La commande `summary(t)` donne la liste des différentes composantes de `t`. À quoi correspondent toutes ces composantes ? (On peut les afficher en faisant `t$nomvariable`. Que nous apprennent les résidus ?

7. Effectuer deux autres tests du χ^2 pour tester si la survie dépend du sexe et de l'âge.

Exercice 4 : un peu de programmation

1. Créer une fonction `t.chi.ad` qui étant donné un vecteur x et un vecteur de probabilités p :
 - vérifie que p est un vecteur de probabilité et renvoie un message d'erreur sinon
 - vérifie les coordonnées du vecteur x sont plus grandes que 5 et renvoie un message d'erreur sinon.
 - Si les conditions sont vérifiées, fait le test du χ^2 d'adéquation et donc calcule la statistique de test T , le nombre de degrés de liberté et la p -valeur stocke ces trois variables dans une liste.
2. Tester cette fonction sur les données d'un exercice précédent.