Université de Nice — L1MASS, année 2014-2015
Département de Mathématiques — Statistique
NOM : Date : .
PRENOM : Salle et machine : .

**Feuille-Réponse 9**
**Test du $\chi^2$ (chi2)**

Les exercices suivants sont à faire au moyen de R. On chargera donc ce logiciel et on veillera à diviser son écran en deux parties : la console pour les réponses et messages d'erreurs et l'éditeur (script) dans lequel sera saisi le code, qu'on veillera à sauvegarder en fin de séance.

# 1 Un test du $\chi^2$ d'indépendance

1. Le tableau suivant donne la répartition en pauvres et non-pauvres d'un échantillon de 1000 familles dans quatre régions des USA, en 1979 (pauvre signifiait revenu inférieur à 7.400 USD pour une famille de 4 personnes) :

| • | Northeast | Midwest | South | West |
|---|---|---|---|---|
| pauvres | 112 | 105 | 154 | 113 |
| non-pauvres | 888 | 895 | 846 | 887 |

   On veut tester au test du chi2 l'hypothèse nulle $H_0$ que le taux de pauvreté est indépendant de la région, c'est à dire statistiquement le même partout. Quel est ce taux $p$ ? On admet que le nombre de famille est sensiblement le même dans toutes les régions. Quel est le nombre $d$ de degrés de liberté du test envisagé ?

2. Au seuil $\alpha = 10\%$ quelle est la valeur de la statistique du chi2 au-delà de laquelle on rejettera $H_0$ ? On rappelle que le quantile $1-\alpha$ de la loi du $\chi^2(k)$ est donné par la commande `qchisq(1-alpha,k)`.

3. Au seuil $\alpha = 1\%$ quelle est la valeur de la statistique du chi2 au-delà de laquelle on rejettera $H_0$ ?

4. Calculer la statistique [1] $S$ pour cet échantillon.

5. Quelle est votre conclusion quant-au rejet ou non de $H_0$ ?

6. Quelle est, sous $H_0$, la probabilité (*p-value*) d'avoir trouvé une valeur aussi grande que celle que vous avez trouvée pour $S$.

1. Il y a ici double indexation des cas : $i = 1..2$ pour "pauvre" et "non-pauvre", et $j = 1..4$ pour les quatre régions. La statistique utilisée est alors $S = \sum_{j=1}^{4} \sum_{i=1}^{2} \frac{(O_{ij}-T_{ij})^2}{T_{ij}}$

## 2 Simulation d'un dé non pipé

Nous envisageons dans cette section et la suivante l'hypothèse nulle $H_0$ que les six faces d'un dé sont équiprobables. Nous utilisons le générateur de nombres aléatoires de `R` pour simuler un dé non-pipé (puis un dé pipé, dans la section suivante) et expérimentons la méthode en calculant une statistique du $\chi^2$ de l'échantillon obtenu.

1. En vous inspirant du code du cours simulez un échantillon de taille 840 de tirages d'un dé non-pipé. Quels effectifs obtenez-vous ? Quitte à refaire suffisamment de tirages aléatoires, veillez à ne pas avoir les même effectifs que vos voisins (n'omettez pas d'indiquer le numéro de votre machine à la place prévue à cet effet en haut de votre copie).

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $O_i$ | | | | | | |
| $T_i$ | | | | | | |

2. Donnez un intervalle de confiance au seuil $\alpha = 5\%$ pour un échantillon de cette taille pour les proportions des classes, sous l'hypothèse que le dé n'est pas pipé ? Puis en déduire l'intervalle de confiance pour les effectifs. Les effectifs de votre échantillon sont-ils dans cet intervalle ?

3. On fait l'hypothèse nulle $H_0$ que le dé n'est pas pipé. Quel statistique $S$ du chi2 obtenez-vous pour votre échantillon ? Détaillez ici la valeur qui doit être calculée et donnez le résultat trouvé par `R`.

4. Rejetez-vous l'hypothèse $H_0$ ?(Faites une phrase précisant votre seuil)

# 3 Simulation d'un dé pipé

1. En vous inspirant du code du cours simulez un échantillon de taille 840 de tirages d'un dé pipé avec `pipage=0.02`. Quels effectifs obtenez-vous ? Quitte à refaire suffisamment de tirages aléatoires, veillez à ne pas avoir les même effectifs que vos voisins. Voyez si les intervalles de confiances trouvés à la question 2.2 contiennent les valeurs que vous obtenez et commentez.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $O_i$ | | | | | | |
| $T_i$ | | | | | | |

2. On fait l'hypothèse nulle $H_0$ que le dé n'est pas pipé. Quel statistique $S$ du chi2 obtenez-vous pour votre échantillon ? Détaillez ici la valeur qui doit être calculée et donnez le résultat trouvé par `R`.

3. Rejetez-vous l'hypothèse $H_0$, au seuil $\alpha = 10\%$, $\alpha = 5\%$, et $\alpha = 1\%$ ?

# 4 Loi du $\chi^2$

1. Quelles est la valeur en 0 de la densité de la loi du $\chi^2$ à 1, 2, 3, 4, et 5 degrés de liberté.

2. Représenter avec soin le graphe de ces densités pour $x \in [0, 20]$. Indication : éviter de commencer par $\chi^2(1)$ ; commencer par exemple par

```
xx=seq(0,20,by=0.05);
plot(xx,dchisq(xx,2),type="l",col="green");
curve(dchisq(x,1),add=TRUE,col="red");
```