Université de Bourgogne Année 2017-2018

L3 de mathématiques Sciences, Technologies, Santé - Statistique

---

Examen de l'U.E. STATISTIQUE - jeudi 17 mai 2018

Questions sur R (4 pts)

Soit la fonction suivante codée en R :

```
f=function(N)
{
  N1=10000
  X=rep(0,N1)
  for(i in seq(1,N1))
  {
    a=rbinom(N,1,0.2)
    X[i]=mean(a)
  }
  hist(X,probability=TRUE)
  curve(dnorm(x,0.2,sqrt(0.2*0.8/N)),add=TRUE,col='red')
}
```

1. Que fait l'instruction *X=rep(0,N1)* ?
2. Que fait l'instruction *a=rbinom(N,1,0.2)* ?
3. Que peut-on attendre pour les valeurs de X après la boucle *for* ? (Justifier avec un théorème)
4. A quoi sert l'option *probability=TRUE* dans l'instruction *hist(X,probability=TRUE)* ?
5. Que fait l'instruction *curve(dnorm(x,0.2,sqrt(0.2*0.8/N)),add=TRUE,col='red')* ?
6. Que constate-t-on quand on appelle successivement *f(5) ; f(30) ; f(100) ; f(1000)* ? (Justifier avec un théorème)

Exercice 1 : (8 pts)

Des relevés nationaux concernant le travail intérimaire indiquent que la durée moyenne d'un contrat est de 9 jours. Dans une certaine agence, on prélève deux échantillons, l'un (de taille 260) constitué de contrats effectués en 2015, l'autre (de taille 310) constitué de contrats effectués en 2016.

Durée du contrat (en jours)	$[0, 2[$	$[2-6[$	$[6-16[$	$[16-30[$	$[30-100]$	total
2015	152	65	28	12	3	260
2016	184	88	22	11	5	310

1. Tester, avec un risque d'erreur de 5%, l'hypothèse que la durée moyenne d'un contrat en 2016 dans cette agence est différente de la moyenne nationale.
2. Tester, avec un risque d'erreur de 5%, l'hypothèse que la durée moyenne d'un contrat en 2015 est plus faible qu'en 2016.
3. Durant le mois de janvier 2017, 20 offres d'emploi sur 250 n'ont pas trouvé preneur. Tester, avec un risque d'erreur de 5%, l'hypothèse que cette proportion est supérieure à la proportion usuelle de 5%.
4. A partir des données de **2015**, tester, avec un risque d'erreur de 5%, l'hypothèse que la durée d'un contrat en 2015 suit une loi exponentielle dont la fonction de répartition est $F_0(x) = 1 - \exp(-\frac{x}{4,6})$ pour $x \geq 0$.

   Indications : pour faire les calculs, on pourra utiliser le tableau suivant où F_n désigne la fonction de répartition empirique (rappelons que lorsque l'échantillon est découpé en classes, on se contente d'étudier l'écart entre les fonctions de répartition aux points formant les bornes des classes).

	2	6	16	30	100
F_n					
F_0					

Exercice 2 : (8 pts)

On rappelle que la densité de la loi normale centrée réduite est

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

1. Soit Z une variable aléatoire suivant une loi normale centrée réduite.
   a) Montrer (en intégrant) que $E(Z^2) = 1$.
   b) Montrer (en intégrant) que $E(Z^4) = 3$.
   c) En déduire la valeur de $V(Z^2)$.
2. Soit X une variable aléatoire suivant une loi normale centrée (d'espérance 0) et de variance σ^2.
   a) Déduire de ce qui précède $E(X^2)$.
   b) Déduire de ce qui précède $V(X^2)$.

3. Soit $(X_1, ..., X_n)$ n variables aléatoires indépendantes et identiquement distribuées de loi normale centrée (d'espérance 0) et de variance σ^2. On considère l'estimateur suivant pour σ^2

$$T = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

   a) L'estimateur T a-t-il un biais ?

   b) Calculer la variance et l'erreur quadratique de T

   c) Quel résultat donne la loi faible des grands nombres appliquée à T ?

   d) Quel résultat donne le théorème central limite appliqué à T ?

Exercice 3 : (4 pts)

On souhaite estimer la proportion π d'adultes illettrés parmi les personnes inscrites à Pôle Emploi dans une certaine agence suivant 1000 chômeurs. On sait par ailleurs que dans les agences semblables environ 15% des personnes sont illettrées. On se propose de sélectionner un échantillon au moyen d'un sondage aléatoire simple.

1. Quelle taille d'échantillon faut-il sélectionner pour que la demi-amplitude d'un intervalle de confiance avec un niveau de confiance 0,95 soit inférieure à 0,03 dans le cas :

   a) avec remise

   b) sans remise

2. De façon littérale maintenant : notons n_0 la taille de l'échantillon nécessaire pour avoir un intervalle de confiance avec remise ayant une demi-amplitude inférieure à h.

   a) Exprimer n_0 en fonction de h, f et z_α.

   b) Exprimer la taille de l'échantillon nécessaire pour avoir un intervalle de confiance **sans remise** ayant une demi-amplitude inférieure à h en fonction de n_0 et de N uniquement. (où N désigne la taille de la population)