Université de Strasbourg
UFR de Mathématiques et d'Informatique
Han-Ping LI

Année 2018/2019
Statistique
Etude de cas L3

TD 1 Notions statistiques

Exercices sur l'utilisation de R

Voir "*Introduction au logiciel R*" par S. Déjean

http ://www.math.univ-toulouse.fr/~sdejean/PDF/semin-R_juin_2014.pdf

Trois commandes très utiles pour les exercices :

1) La boucle "for" dont le syntaxe est le suivant :
for (compteur **in** vecteur) **{**
liste d'instructions à exécuter
}

Exemple : pour calculer la somme et le produit de 10 premiers entier :
p=1 ;s=0
for (i in 1 :10){s=s+i ; p=p*i}
on obtient donc s=0+...10 et p=10 !

2) La boucle "while" dont le syntaxe est le suivant :

while (condition) **{**
liste d'instructions à exécuter
}

Exemple : pour trouver une approximation du nombre pi avec le dénominateur 113 :

```
k=1
while (k/113 <pi) { k=k+1}

print (k/113)
```

3) Echantillonnage "**sample**" dont le syntaxe est le suivant :

sample(vecteur, n, replace = TRUE) # population : 1 :N <-> N
fournit n valeurs prises au hasard avec remise dans le vecteur donné, représentant une réalisation $(x_1 \ldots, x_n)$ d'un échantillon de taille n $(X_1 \ldots, X_n)$.

Exemple 1 : On prélève d'un échantillon de taille 8 (avec remise) $(X_1 ... X_8)$ dans la population constituant de $\{1, 2, 3\}$ On peut générer 15 réalisations de l'échantillon $(X_1 ... X_8)$ plus 15 réalisations de la moyenne d'échantillon Xbar=sum(Xi)/n qui est un variable aléatoire :

```
Population=1 :3
N=12 ; n=8
Echantillon=matrix(NA, nrow=N,ncol=n) # N réalisations
Moyenne.echantillon=rep(NA,N) ; # N réalisations

for (j in 1 :N) {
Echantillon[j,]=sample(Population, n, replace=T)
Moyenne.echantillon[j]=mean(Echantillon[j,]) }

Echantillon
Moyenne.echantillon
```

Exercice 1 :

a) Dans un bassin où vivent $N = 10$ poissons $P_1, \ldots, P_N$. Si on en pêche (simultanément) 3, quels sont les possibilités ? Que signifie **choisir** (simultanément) $k = 3$ poisson **au hasard** ?

L=1 :10 ; combn(L, 3)

b) On suppose que les poissons sont tous identiques et qu'il y a $K = 4$ poissons marqués parmi les 10. Quelle est la probabilité quand on en pêche (simultanément) $k = 3$ d'en trouver i=2 qui sont marqués ? Quelle est la loi de probabilité ?

On Simule $M = 500$ réalisations d'une variable aléatoire suivant une loi hypergéométrique $H(10, 4, 3)$ dans la suite.

hyper(K, N-K, k)

c) Déterminer les fréquences d'obtenir $i = 0, \ldots, 3$ poissons marqués parmi 4, comparer avec les effectifs théoriques.

d) Tracer un diagramme à bâtons à barres accolées des effectifs observés et des effectifs théoriques.

e) Comparer la moyenne théorique avec la moyenne observée de ces M réalisations.

f) Mêmes questions pour $N = 40, K = 10$ et $k = 9$.

Exercice 2 :

a) On lance un dé équilibré et on définit une v.a. $X = 0$ si le résultat du dé est inférieur ou égal à 4, $X = 1$ sinon.

On note S le nombre de 1 après $n = 5$ lancers. Quelle est la loi de S ?

b) Simuler $M = 1000$ réalisations de la v.a. S, déterminer les fréquences d'avoir $S = 0, \ldots, 5$. Comparer avec les effectifs théoriques.

c) Tracer un diagramme à bâtons à barres accolées des effectifs observés et des effectifs théoriques.

d) Comparer la moyenne théorique avec la moyenne observée de ces M réalisations.

e) Quelle est la différence entre l'exercice actuel et le précédent ?

Exercice 3 : On s'intéresse à la largeur des anneaux de croissance des arbres. On prend le fichier "treering" dans la librairie "datasets" avec 7980 comme étant le cardinal de la population. On note μ la moyenne (théorique) de largeurs et on souhaite avoir une idée sur la grandeur du paramètre μ. Pour ce faire, on va construire $M = 500$ réalisations, basé sur un échantillon de taille $n = 65$.

a) Sauvegarder ces 7980 valeurs dans un vecteur nommé **Population**, étudier ses caractéristiques en utilisant la commande "summary".

b) Générer $M = 500$ réalisations de l'échantillon $(X_1, ..., X_n)$ de taille $n = 65$, sauvegarder 500 réalisations de l'échantillon ainsi que sa moyenne.

c) Comparer la moyenne théorique, la moyenne d'échantillon ainsi que la moyenne des moyennes observées. Quelle est la différence entre la nature des trois moyennes ?

Exercice 4 :

On continue à lancer un dé équilibré. On définit une v.a. Y le nombre de lancers nécessaire pour obtenir un résultat supérieur ou égal à 5 pour la première fois.

a) Déterminer la loi de Y.

b) Simuler $M = 1000$ réalisations de la v.a. Y, déterminer les fréquences d'avoir $Y = 1, \ldots, 14$, puis $Y \geq 15$. Comparer avec les effectifs théoriques.

c) Tracer un diagramme à bâtons à barres accolées des effectifs observés et des effectifs théoriques.

d) Comparer la moyenne théorique avec la moyenne observée de ces M réalisations.

Exercice 5 :

Si U est une v.a. de loi uniforme $U(0,1)$, alors $Y = -\frac{\log(U)}{\lambda}$ suit une loi d'exponentielle de paramètre λ. On pose $\lambda = 1/2$ dans la suite. On simule $M = 5000$ réalisations de la v.a. U de loi uniforme $U(0,1)$, par la transformation on en obtient ainsi 1000 réalisations d'une v.a. d'une loi d'exponentielle de paramètre λ. on note le vecteur obtenu par x.

a) Tracer l'histogramme de ces M réalisations en utilisant

(m=floor(min(x))) ; (M=ceiling(max(x))) ; brk=c(seq(m,M,0.3)) ;

hist (x, probability = TRUE, col = "light blue",breaks=brk)

b) On superpose le graphique de la densité théorique exp(1/2), indépendant des données, sur le graphique précédent
u=seq(0,15,length=300) ;v=dexp(t,rate=1/2)

lines(u,v, type="l", ,col="red",lwd = 2,add=T)

c) On peut calculer la proportion des données comprises entre 8 et 12 :
(length(x[(x $\geq$ 8)&(x$\leq$12)]))/M

d) aussi la probabilité P(8<=X<=1 2)=F(12)-F(8)
où la fonction de répartition de la loi est donnée par "pexp(t, rate=1/2)"
en utlisant la fonction de répartition
pexp(12,rate=1/2)-pexp(8,rate=1/2)

f=function(x)exp(-x/2)/2 ; integrate(f, 8, 12) ;

e) Son graphique
x=seq(0.5,1.4,length=300) ; y=dexp(x,rate=1/2)

```
plot(x,y,type="l", lwd=2, col="red")

x=seq(8,12,length=200) ; y=dexp(x,rate=1/2)

polygon(c(8,x,12),c(0,y,0),col="gray")
```

Exercice 6 :

a) On simule $M = 5000$ réalisations d'une v.a. de loi $N(-10, 1.5^2)$, noté xn.

b) On affiche l'ensemble de xn tels que $-12 < xn <= -10$.

c) On compare la proportion de xn tels que $-12 < xn <= -10$ avec leur probabilité.

d) On superpose la fonction de densité de X avec l'histogramme.

Exercice 7 : Triangle aléatoire

On fixe trois points $A_1 = (1, 1), A_2 = (50, 50 * sqrt(3))$ et $A_3 = (100, 1)$.

a) On choisit au hasard un point $P_1 = (x_1, y_1)$ dont les deux composantes sont prises au hasard dans le carré de sommets (1,1), (100,1), (100,100) et (1,100).

b) On sélectionne un des 3 points A_i avec probabilité 1/3. Si A_i est choisi, le prochain point P_{i+1} se trouvera au mi-chemin entre le point actuel et le point P_i, ainsi de suite, $i = 2, \ldots, N = 5000$.

c) Tracer le graphique de ces 50000 points ainsi obtenus. Que constatez -vous ?