

Université Claude Bernard - Lyon 1

M1 - IGAPA
Statistique

Année 2007-2008
Session 2

Durée : 2 heures

A.B. Dufour & J. Bastien

LES ETUDIANTS DOIVENT CHOISIR DE PREFERENCE LES EXERCICES PROPOSES
PAR LEUR ENSEIGNANT DE TD. DANS LE CAS CONTRAIRE, CHOISIR QUATRE
EXERCICES PARMIS CEUX PROPOSES.

NOTER LE NOM DE L'ENSEIGNANT EN TETE DE COPIE

Tous les documents sont autorisés.

Exercice 1 - AB Dufour

Probabilités liées à une loi de Poisson

Une loi de Poisson est une variable aléatoire désignant le temps d'attente entre deux événements, dont la caractéristique principale est l'égalité de la moyenne et de la variance. X varie de 0 à l'infini. Soit X définie par son unique paramètre $\lambda = 1$. Nous donnons les valeurs de 0 à 10.

`dpois(0:10, 1)`

```
[1] 3.679e-01 3.679e-01 1.839e-01 6.131e-02 1.533e-02 3.066e-03 5.109e-04 7.299e-05  
[9] 9.124e-06 1.014e-06 1.014e-07
```

`ppois(0:10, 1)`

```
[1] 0.3679 0.7358 0.9197 0.9810 0.9963 0.9994 0.9999 1.0000 1.0000 1.0000 1.0000
```

A l'aide des informations ci-dessus, calculer les probabilités :

1. $P(X = 2)$
2. $P(3 \leq X \leq 4)$
3. $P(X \leq 6)$
4. $P(X \geq 5)$

Probabilités liées à une loi normale centrée réduite

On donne quelques valeurs liées à cette loi.

```
axex <- seq(-3, 3, by = 0.3)
axex

[1] -3.0 -2.7 -2.4 -2.1 -1.8 -1.5 -1.2 -0.9 -0.6 -0.3  0.0  0.3  0.6  0.9  1.2  1.5
[17]  1.8  2.1  2.4  2.7  3.0

pnorm(axex)

[1] 0.001350 0.003467 0.008198 0.017864 0.035930 0.066807 0.115070 0.184060 0.274253
[10] 0.382089 0.500000 0.617911 0.725747 0.815940 0.884930 0.933193 0.964070 0.982136
[19] 0.991802 0.996533 0.998650
```

A l'aide des informations ci-dessus, calculer les probabilités :

1. $P(X \leq -2.4)$
2. $P(X > 2.8)$
3. $P(0.2 \leq X \leq 1.2)$
4. On note z une valeur particulière de la distribution normale centrée réduite (z est un nombre positif). Donner la relation liant $P(X > z)$ et une probabilité (à définir) de $-z$.

Exercice 2 - AB Dufour

Dans un sondage réalisé pour l'université de Montréal, Sylvie Beauchamp et Pierre Bouchard se sont interrogés sur les habitudes de lecture d'une population adulte. Répondre, en vous aidant des informations à votre disposition, à la question : avez-vous lu au moins un livre au cours des trois mois précédant l'interview ?

Les données :

	langue maternelle			
	français		anglais	
	hommes	femmes	hommes	femmes
oui	265	406	40	65
non	319	264	20	22

Comparaisons de proportions :

Comparaison	groupe 1 vs groupe 2	f_1	f_2	χ^2	p -value
français	hommes vs femmes	0.4538	0.6060	28.4477	9.62e-8
anglais	hommes vs femmes	0.6667	0.7471	0.7667	0.3812
hommes	français vs anglais	0.4538	0.6667	9.0568	0.002617
femmes	français vs anglais	0.6060	0.7471	5.9402	0.0148

Exercice 3 - AB Dufour

On a mesuré la quantité d'alcool (g/l) contenue dans 10 cidres doux d'un marché de Bretagne.

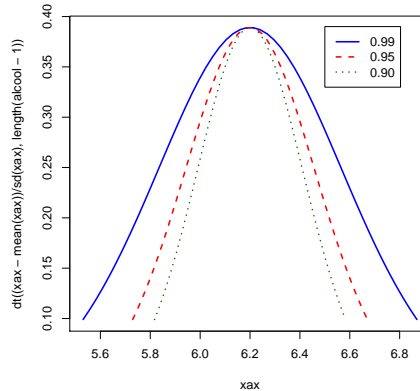
```
alcool <- c(5.42, 5.55, 5.61, 5.91, 5.93, 6.15, 6.2, 6.79, 7.07,
7.37)
sum(alcool)
```

```
[1] 62
```

```
sum((alcool - mean(alcool))^2)
```

[1] 4.012

1. Donner la moyenne et la variance descriptives de la quantité d'alcool de cette série statistique.
2. On représente les intervalles de confiance de la moyenne de l'échantillon pour les niveaux de confiance 0.90, 0.95 et 0.99. Commenter.



3. La norme officielle moyenne de quantité d'alcool dans du cidre doux est $\mu = 6.2$ g/l. La série observée est-elle, en moyenne, conforme à cette norme officielle?
 - a) Donner le nom du test approprié et écrire la fonction de **R** qui permet d'obtenir le résultat.
 - b) Répondre à la question à l'aide des informations ci-dessous.

One Sample t-test

```
data: alcool
t = 0, df = 9, p-value = 1
alternative hypothesis: true mean is not equal to 6.2
95 percent confidence interval:
 5.722 6.678
sample estimates:
mean of x
 6.2
```

- c) Le test était-il utile? Pourquoi? Généralisez votre réponse.

Exercice 4 - AB Dufour

Un test du Chi-Deux dit sur les "outliers" permet de détecter si la valeur d'une distribution peut être considérée comme extrême ou non. Prenons par exemple la série suivante :

```
echantillon <- c(48, 52, 56, 71, 45, 50, 40, 48, 44, 62)
```

Elle contient les résultats à la détente verticale de 10 hommes belges âgés de 18 ans. La question que l'on se pose est sur le sujet ayant réalisé un saut de 71cm. Fait-il partie de la même population ou d'une autre population? Il peut tout simplement pratiquer un sport alors que les autres non. La question peut se formuler autrement : le sujet sautant 71cm est-il un outlier?

La valeur de la statistique du test est $\chi^2 = \frac{(valext - mean(echantillon))^2}{varpop}$ où *valext* représente la valeur à tester, *mean(echantillon)* la moyenne de l'échantillon étudié et *varpop* la variance de la population connue par ailleurs.

- 1) Donner les deux hypothèses H_0 et H_1 .
- 2) Dans notre cas, la valeur de la statistique du test vaut $\chi^2 = 5.925$. Parmi les deux chi-deux ci-dessous, choisir celui qui définit la probabilité associée c'est-à-dire la p-value.

```
pchisq(5.925, 1)
```

```
[1] 0.985
```

```
1 - pchisq(5.925, 1)
```

```
[1] 0.01493
```

3) Donner une conclusion à cette étude.

Exercice 5 - J Bastien

- Des quatre fonctions graphiques de \mathbb{R} : `barplot`, `dotchart`, `pie`, `hist`, quelle est celle qui ne s'adresse pas à une variable qualitative ?
- De ces trois instructions, quelle est celle qui donne les pourcentages liés à une variable qualitative A ?

```
summary(A)
summary(A)/length(A)
100*(summary(A)/length(A))
```

Exercice 6 - J Bastien

Dans cet exercice, on étudie la taille et la masse de sportifs de haut niveau (fichier disponible sur internet : <http://pbil.univ-lyon1.fr/R/donnees/morphosport.txt>, selon le sport pratiqué. Le nombre de pratiquants par sport est donné dans le tableau :

sport	nombre
athlétisme	20
basket	2
foot	33
hand	44
judo	13
natation	24
volley	19

TAB. 1 – Nombre de pratiquants par sport.

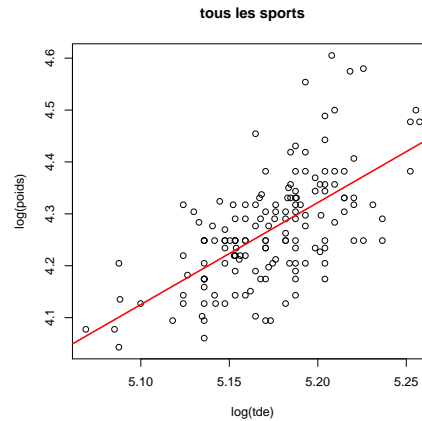
- 1) a) Sous \mathbb{R} , que permet de faire la suite d'instructions suivante,

```
donnees <- read.table("morphosport.txt", header = TRUE)
attach(donnees)
```

L'instruction `head(donnees)` donne le résultat suivant :

```
 sport dbi tde tas lms poids
1  athl 33.1 174 77.8 78.0    70
2  athl 31.5 179 85.6 81.8    80
3  athl 35.5 179 80.0 80.0    63
4  athl 34.5 170 79.3 79.4    70
5  athl 37.3 182 79.5 86.0    67
6  athl 32.7 170 74.4 69.0    58
```

- b) On s'intéresse aux variables `tde` (taille debout) et `poids`. De quel type sont ces variables ?
- a) On souhaite tracer le logarithme du poids en fonction du logarithme de la taille pour l'ensemble des sportifs. Quelles sont les instructions sous \mathbb{R} pour obtenir le graphique ci-dessous ?



En tapant `cor(log(tde),log(poids))`, on obtient :

[1] 0.6362

Que cela signifie-t-il ?

- b) On souhaite maintenant tracer le logarithme du poids en fonction du logarithme de la taille selon les différents sports pratiqués (à l'exception du basket). Quelles sont les instructions sous `R` pour obtenir la figure donnée en annexe (à la fin du sujet) ?

Comment peut-on obtenir les renseignements fournis dans le tableau ci-dessous ? Commenter le.

sport	corélation r	pente de la droite p
Athlétisme	0.675851	2.186065
foot	0.6172698	1.658025
hand	0.4916253	1.426944
judo	0.8487247	3.869508
natation	0.7026234	2.081003
volley	0.7330231	2.145380

TAB. 2 – Corrélations linéaires et pentes par sport.

Exercice 7 - J Bastien

Dans cet exercice, on étudie un groupe d'étudiant en L3 APA (fichier disponible sur internet : <http://pbil.univ-lyon1.fr/R/donnees/L3APA06.txt>).

L'instruction `head(read.table("L3APA06.txt",header=T))` donne le résultat suivant :

```

groupe identifiant  sexe poids  taille rythmcard  age baccaalaureat mention
1      A            1 masculin  70   170      60 27/01/1985      S      P
2      A            2 masculin  80   185      65 13/01/1984      S      P
3      A            3 féminin  54   159      62 18/04/1984      S      P
4      A            4 féminin  60   170      58 20/05/1985      ES     P
5      A            5 féminin  60   170      56 30/04/1985      S      AB
6      A            6 féminin  59   165      56 23/02/1985      L      P
hmental hmoteur  hsensoriel pblesocial pratique sport  niveau mecriture
1      0      0      0      1      non gymnastique loisir droite
2      1      1      1      0      oui handball compétition droite
3      0      1      0      1      non <NA> <NA> droite
4      0      1      0      0      oui tennis compétition gauche
5      1      0      0      0      oui danse compétition droite
6      1      1      0      0      oui gymnastique compétition gauche
mfourchette pballon  oeil rotation pappui
1 gauche droit droit gauche droit
2 gauche droit droit gauche gauche
3 droite droit droit gauche droit
4 gauche droit gauche droit droit
5 droite droit gauche droit droit
6 gauche droit droit gauche droit

```

Choisir une variable qualitative et une variable quantitative et expliquer comment on peut vérifier sous `R` si ces deux variables sont liées.

Annexe

