

Questions

1. Quel est l'équivalent en statistique descriptive de la probabilité d'un évènement ?
2. Une population est étudiée via un caractère (ou variable) quantitatif X. On calcule la variance de X et on trouve 0 ; que peut on dire de X ?
3. Une population est étudiée via deux caractères (ou variables) quantitatives X et Y . On calcule le quotient de la covariance de X et Y par le produit des écarts type de X et de Y et on trouve 1 ; que peut on dire de X et Y ?

TP

1. Le dessin ci-dessous est une saisie d'écran de LibreOffice Calc. La cellule F5 a pour contenu la formule =SOMME.SI(A1:D5,"=2") ; quelle sera la valeur ? Quelle valeur obtient on si on copie la cellule F5 et qu'on la colle en G6 ? Comment doit on modifier la formule pour que le résultat soit insensible au copie-coller ?

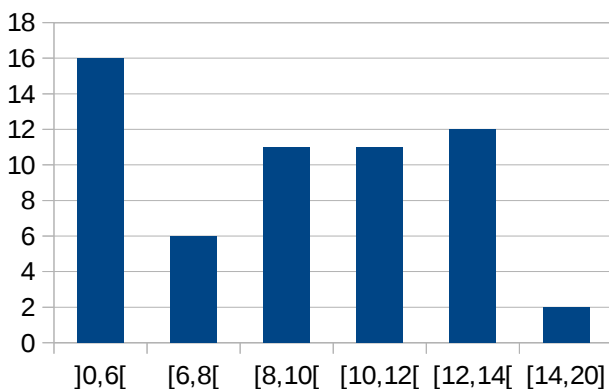
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	0	1	2	1									
2	0	1	1	0									
3	1	2	1	2									
4	1	2	0	2									
5	1	1	0	1		=SOMME.SI(A1:D5,"=2")							
6													

2. Que font chacune des lignes du script R suivant ?

```
library(MASS);data(survey)
surveyc=survey[complete.cases(survey),]
x=surveyc$Wr.Hnd;y=surveyc$NW.Hnd
a=cov(x,y)/var(x);b=mean(y-a*x)
cat(a,b)
plot(x,y)
abline(b,a)
```

Exercices

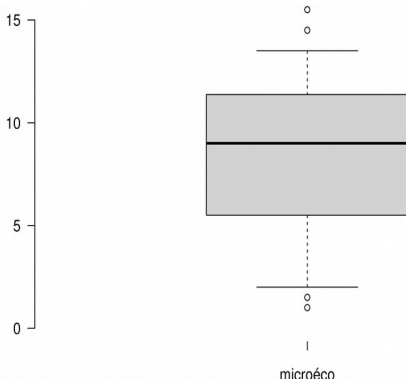
1. Le dessin suivant représente l'histogramme des notes de microéconomie en Licence MASS.



Quelle est la taille de la population ?

On rencontre un étudiant au hasard qui dit avoir validé la matière microéconomie. Quelle est la probabilité qu'il ait eu 12 ou plus ?

2. Le dessin ci-dessous représente la boîte à moustaches des notes en microéconomie



Que vaut approximativement le troisième quartile ?

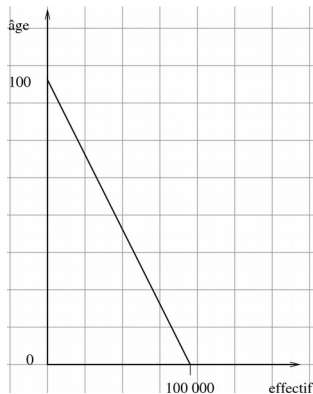
La boîte à moustache est elle cohérente avec l'histogramme de la question 1 ? Expliquer.

3. Le dessin ci-dessous représente la pyramide des âges des hommes d'une population fictive.

Quelle est la taille de la population ?

Quelle est l'âge médian ?

Quelle est la probabilité qu'un individu choisi au hasard ait plus de 50 ans ?



4. Une population de 1000 individus est étudiée à travers deux caractères quantitatifs X et Y . On représente ci-dessous le nuage des points $(X(i), Y(i))$ et la droite de régression $y = ax + b$. Pour chaque individu i on note $R(i)$ la différence entre la valeur de Y et la valeur prédite par X et la régression linéaire : $R(i) = Y(i) - aX(i) - b$.

a. D'après le graphique, que valent approximativement les coefficients a et b de la droite de régression ?

b. Sachant que la moyenne de X vaut $\frac{1}{2}$, que vaut approximativement la moyenne de Y ?

c. Que vaut la moyenne de Y conditionnée à l'évènement " X est proche de 0" ? Que vaut la moyenne de R conditionnée à " X est proche de 0" ?

d. Que vaut les moyennes de Y et de R conditionnées à " X est proche de 1" ?

e. La régression linéaire est elle satisfaisante ? Quel changement de variable sur X peut on essayer pour améliorer la régression ?

