

Probabilités et Statistiques

T.P. 1

- Les **statistiques sous Maple** s'utilisent à l'aide du "package" *stats* :

```
> with(stats);
```

qui contient des "sub-packages" : *describe*, *importdata*, *statevalf*, *statplots*, *fit* etc... On les ouvre par :

```
> with(stats[describe]);
```

- Le subpackage "*describe*" contient les fonctions essentielles de statistiques descriptives : *mean* (moyenne), *standarddeviation* (écart-type), *median* (médiane) etc...

- Le subpackage "*statplots*" permet de tracer toutes les figures associées (histogrammes,...)

- Le subpackage "*fit*" permet d'ajuster un nuage de points par une droite ou autre courbe.

- Le mode d'emploi des commandes de ces subpackages s'obtient par l'aide. Exemple :

```
> ?histogram;
```

I. Prise en main.

1. Une *série statistique à une variable* est une liste (donc définie entre crochets) de nombres, ou d'intervalles.

(a) Définir la série X_1 considérée à l'exercice 3 du TD1 (liste de notes).

(b) Calculer sa moyenne, sa médiane, son écart-type et sa variance.

(c) Tracer son histogramme, en utilisant l'option *area=count* (elle permet d'avoir des rectangles de largeur égale et de hauteur proportionnelle à l'effectif).

(d) Sur cet histogramme, Maple choisit de lui-même un certain nombre de classes en abscisses (ici 6). On voudrait des classes plus petites (ne contenant qu'une note). Modifier ceci en utilisant l'option *numbars=k*. (Quel nombre k de colonnes faut-il utiliser si on veut des intervalles de longueurs $1/2$, couvrant $[2.5, 18.5]$? Attention : les colonnes de Maple sous *stats* sont de la forme $[\]$).

2. *Série statistique à deux variables*.

Définir les deux séries :

$$X := [0, 1, 2, 2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 9, 10]$$

$$Y := [4, 5, 5, 3, 5, 3, 4, 4, 4, 6, 6, 5, 9, 8, 8, 7].$$

(a) Calculer les moyennes et écart-types de ces distributions, ainsi que la covariance et le coefficient de corrélation linéaire r du couple (X, Y) (chercher les commandes correspondantes dans la liste du subpackage *describe*). r est-il proche de 1 ?

(b) Rechercher l'équation de la droite de régression de Y par rapport à X à l'aide des formules indiquées dans le cours.

(c) Rechercher cette droite à l'aide du subpackage *fit* de *stats* (to fit a curve : adapter une courbe à), et de la commande *leastsquare* (voir son aide). A l'aide de cette même commande trouver la courbe de degré deux $y = ax^2 + bx + c$ ajustant au mieux cette série.

- (d) Afficher sur un même dessin le nuage de points associé à (X, Y) , et la droite d'ajustement linéaire puis le nuage de points et la courbe d'ajustement quadratique (de degré deux). Indication : utiliser `scatterplot` pour le nuage de point, `plot` pour les fonctions linéaires ou quadratique. Pour les superposer, donner un nom à chaque représentation : $p := \text{scatterplot}(X, Y)$ $q := \text{plot}(a*x+b, x=0..10, y=0..10)$ et les superposer par `plots[display]([p, q])` ;

II. Traitement statistique d'un signal bruité.

On considère une suite de 100 points $(1, f(1)) \cdots (100, f(100))$, sur le graphe d'une fonction à priori inconnue f . La mesure de ces points n'est pas exacte mais perturbée par un bruit b_i : $Y := (i, f(i) + b_i)$. Les points mesurés ne sont donc plus sur cette courbe. On veut néanmoins retrouver cette courbe de manière approchée (par une droite ou une courbe de degré 2, 3, ...) à l'aide d'outils statistiques.

On commencera par réinitialiser (`restart`) puis charger le package `stats`, et les subpackages `fit` et `statplots`.

1. Créer la liste $X := [1, 2, \dots, 100]$ (utiliser `seq`, puis transformer en liste par des crochets).
2. Créer un bruit, c'est-à-dire une suite B de 100 nombres générés aléatoirement (selon une loi que nous étudierons plus tard, la loi normale) :

$$> B := [\text{random}[\text{normald}[0,1]](100)];$$
Explication : cette commande génère 100 nombres $B[i]$ de moyenne 0 et d'écart-type 1. On peut faire augmenter ou diminuer ces bruits en augmentant ou diminuant l'écart-type (ici 1).
3. Définir la fonction $f := x \mapsto 10e^{-x/30}$. Construire la liste Y des mesures bruitées : $Y := [f(1) + B[1], \dots, f(100) + B[100]]$.
4. Calculer l'équation de la droite d'ajustement linéaire de (X, Y) . Maple affiche une équation $y = ax + b$. Extraire le second membre de cette équation $ax + b$ afin d'avoir une expression que l'on pourra tracer : $C := \text{rhs}(\%)$ (rhs : right hand side). Placer sur un même dessin la courbe de f , le nuage de points (X, Y) , et la droite d'ajustement pour approcher la courbe.
5. Afin d'améliorer cette récupération, calculer l'équation de la courbe du second degré $y = cx^2 + dx + e$ approximant le nuage de points et comme précédemment, comparer avec le graphe de f sur un même dessin.
6. Recommencer avec une courbe du troisième degré $y = gx^3 + hx^2 + lx + m$.
7. Tester la qualité de cette récupération de la courbe de f en augmentant ou diminuant le bruit (écart-type de B).