

# Statistiques descriptives à 1 variable - Tri à plat

## A quoi vont-elles servir ?

⇒ **Décrire**

**Distribution**

**Position** : moyenne, mode, médiane, ...  
(ordre de grandeur)

⇒ **Résumer**

paramètres et graphes

→ **Forme** (symétrie, tendance...)

**Dispersion** : écart-type, variance, quantiles, ...  
(répartition autour de l'ordre de grandeur)

⇒ **Comparer**

**Echantillons – Populations**

# Statistiques descriptives à 1 variable - Tri à plat

## L'analyse Graphique

**Contexte** ⇒ contribue à **faire parler vos données**

- Rappels sur la manipulation des données sous R
- Premier graphe avec R
- Comparaison avec EXCEL
- Les fichiers : véhicules des données
- Instructions gérant les matrices permettant, entre autre, de réaliser des graphes
- Gestion des barres d'incertitude (erreur)
- Présentation *Powerpoint* de votre graphe
- La part du stochastique dans un graphe déterministe
- Quizz final
- De l'acquisition des données
- Exercice proposé (difficulté : 3 sur une échelle de 5 niveaux)

# Statistiques descriptives à 1 variable - Tri à plat

## L'analyse Graphique

### Avertissements :

- > ça commence doucement mais le rythme ne souffrira pas des manques d'attentions
- > on travaille en groupe
- > on signe un engagement à ne pas utiliser internet à des fins personnelles en classe

### Autre :

- > présentation du site web de l'enseignement

# Traitement des données (*data processing*)

- **Récupération et formatage des données brutes** : **entrées**  
(celles de l'expérience par ex, dans des fichiers)
- **Déclaration des acteurs des calculs** (variables, vecteurs, matrices, constantes, ...)
- **Phase de calcul**
- **Graphes + sorties** (fichiers résultats : images, tableaux, nombres)
- **Transmission des résultats** (serveur, web/FTP, mailing, clé usb, ...)

**nous allons nécessairement utiliser et manipuler différents fichiers**

# Statistiques descriptives à 1 variable - Tri à plat

## L'analyse Graphique

⇒ **Pour réaliser un graphe :**

**Les fichiers sont les véhicules des données**

(depuis l'acquisition jusqu'à l'analyse des données)

- Récupérer (clé usb, ordinateur, réseau) un fichier formaté (Texte ASCII)

- Lire le fichier (instructions réservées R)

- Transférer les données dans une matrice

(> instructions gérant les matrices)

- Utiliser les fonctions statistiques et graphiques appropriées

(dépend du type de contenu du fichier)

# Statistiques descriptives à 1 variable - Tri à plat

## L'analyse Graphique


### Processus déterministe

Processus dans lequel un antécédent produit toujours le même effet.

**cause**  **effet**

### Processus stochastique (aléatoire)

Processus qui, pour un antécédent donné, peut produire plusieurs effets, chacun ayant une probabilité déterminée.

	<u>résultat</u>	<u>Probabilité</u>
<b>cause</b> 	#1	P1
	#2	P2
	#3	P3
	#4	P4
	#5	P5

**Les processus stochastiques font l'objet de l'analyse statistique.**

## Statistiques descriptives à 1 variable - Tri à plat

### L'analyse Graphique

Exemple de processus déterministe :

Loi de Beer-Lamber  $DO^\lambda = \epsilon^\lambda . l . C$

On peut suivre une courbe de croissance (vers 620 nm) bactérienne à l'aide d'un spectrophotomètre. La cause *C augmente* provoque directement le même effet *DO augmente*.

Exemple de processus stochastiques :

- Résistance d'une souche bactérienne à un antibiotique donné  
(boîtes de Pétri)
- Naissance des alvins quelques jours après l'accouplement  
de 2 poissons

# Statistiques descriptives à 1 variable - Tri à plat

## L'analyse Graphique

temps	DO
0	0.005
5	0.006
10	0.007
15	0.008
20	0.011
30	0.017
40	0.023
45	0.033
60	0.080
75	0.085
90	0.120
105	0.250
120	0.330
135	0.600
150	1.020
165	1.950
170	2.320
175	2.660
180	3.340
200	4.050
210	4.250
220	4.170
230	4.280
240	4.250

← **identificateurs/labels** (non obligatoire mais conseillé)

← **données** : matrice [24,2]

← **variables aléatoires** : 2

← **Type de courbe** : déterministe

← **échantillon**: 1 très probablement = une série de mesures

← **unités**: reporter sur le graphe celles utilisées

← **précision des données**: il faut être raisonnable

← **autre** : nous verrons au cours de la réalisation du graphe



## QUIZZ

temps	DO
0	0.005
5	0.006
10	0.007
15	0.008
20	0.011
30	0.017
40	0.023
45	0.033
60	0.080
75	0.085
90	0.120
105	0.250
120	0.330
135	0.600
150	1.020
165	1.950
170	2.320
175	2.660
180	3.340
200	4.050
210	4.250
220	4.170
230	4.280
240	4.250

**Q1 : Peut-on utiliser ces données pour réaliser un test d'hypothèse ?**

**Q2 : Si oui quel type de test (expliquer en 1 ligne maxi) ?**

## Statistiques descriptives à 1 variable - Tri à plat

### L'analyse Graphique

⇒ Récupérer les données d'un **fichier** (logiciel R)

⇒ **Format ASCII** (texte universel)

**read.table**      récupérer un fichier formaté avec entêtes de colonnes

(ressemble à une feuille de calcul EXCEL)

**scan**      récupérer les données d'un fichier constitué uniquement de réels

(réels : au sens large)

**nous allons utiliser et comparer ces 2 instructions**

# bacteria\_data.txt

Col	1	2	3	4	5	6	7
Line	temps	DO1	DO2	DO3	DO4	DO5	DO6
1	0	0.005	0.005	0.003	0.005	0.005	0.005
2	5	0.004	0.008	0.006	0.007	0.005	0.006
3	10	0.008	0.007	0.007	0.006	0.006	0.007
4	15	0.007	0.005	0.005	0.005	0.013	0.012
5	20	0.005	0.013	0.010	0.014	0.010	0.016
6	30	0.012	0.022	0.016	0.021	0.018	0.017
7	40	0.024	0.036	0.031	0.018	0.018	0.012
8	45	0.042	0.021	0.028	0.038	0.038	0.033
9	60	0.082	0.080	0.067	0.090	0.091	0.073
10	75	0.086	0.084	0.067	0.096	0.099	0.080
11	90	0.097	0.132	0.139	0.112	0.118	0.121
12	105	0.258	0.217	0.251	0.214	0.254	0.303
13	120	0.371	0.327	0.273	0.276	0.326	0.380
14	135	0.583	0.666	0.580	0.624	0.499	0.657
15	150	1.000	0.920	1.104	1.072	1.135	1.014
16	165	1.917	1.892	1.825	2.137	1.894	1.937
17	170	2.410	2.267	2.250	2.245	2.286	2.361
18	175	2.599	2.748	2.585	2.662	2.593	2.708
19	180	3.331	3.275	3.305	3.334	3.316	3.267
20	200	3.998	4.085	4.075	4.047	4.123	4.147
21	230	4.360	4.194	4.320	4.318	4.268	4.206
22	240	4.291	4.241	4.365	4.261	4.272	4.284

← **identificateurs**  
(non utilisés)  
**skip**  
> sauter cette ligne  
du fichier

Tab[10,3]=0.084

nl = 22

ncol = 7

**scan** ⇒ récupérer les données d'un fichier constitué uniquement de réels

```

scan("./data/bacteria_data.txt",skip=2)
nl=length(mesures)/7
tab_val=matrix(mesures,ncol=7,nrow=nl,byrow=T)
temps=matrix(ncol=1,nrow=nl,byrow=T)
Abs_600=matrix(ncol=1,nrow=nl,byrow=T)
error=matrix(ncol=1,nrow=nl,byrow=T)
for (i in 1:nl)
{
  temps[i]=tab_val[i,1]
  Abs_600[i]=2*round(mean(tab_val[i,2:7]),3)
  error[i]=round(sd(tab_val[i,2:7]),3)
}

```

Graphe de Abs\_600 en fonction du temps : la moyenne des 6 mesures de DO effectuées à chaque temps

```

plot(x=temps,y=Abs_600)
arrows(temps, Abs_600-error, temps, Abs_600+error,
       length = .05, angle = 90, code = 3)

```