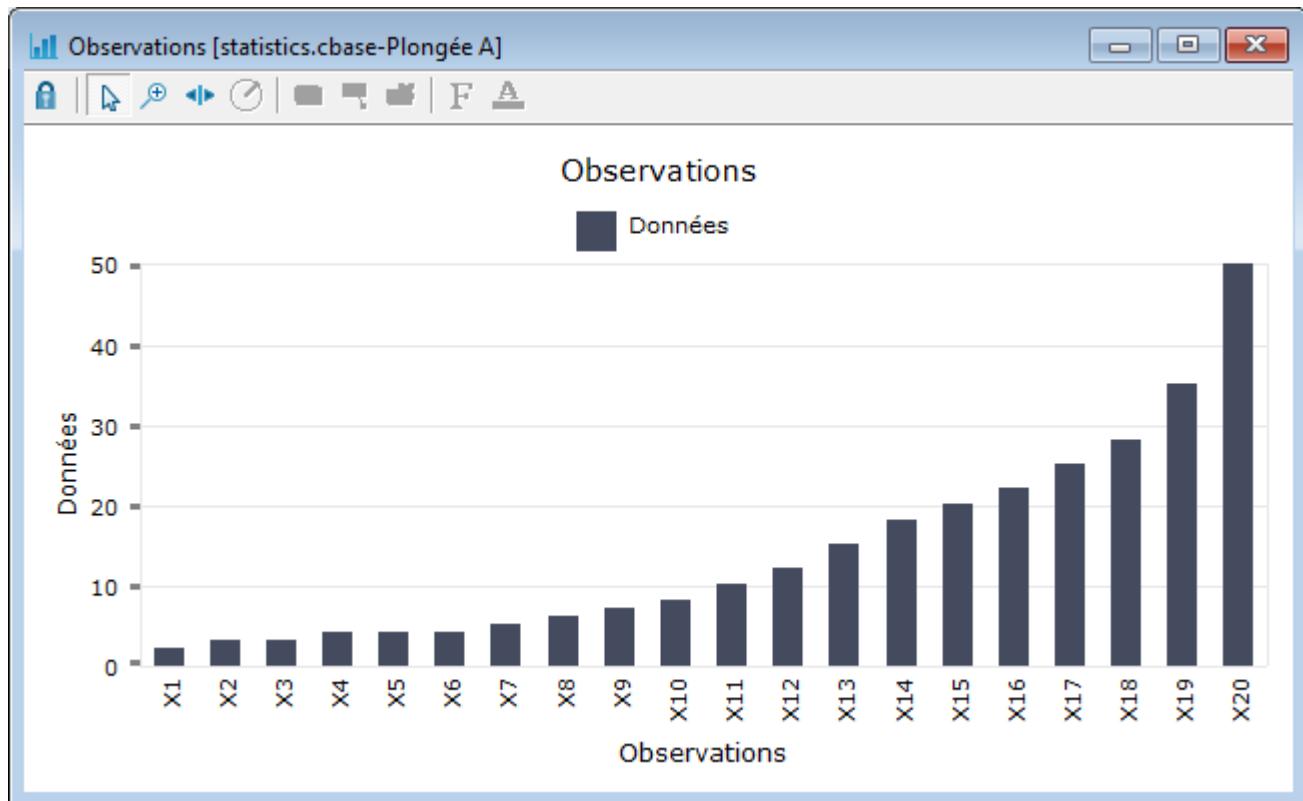


Utilisation de données statistiques avec l'aide de Spectre (1^{re} partie)

Description

Supposons que l'on ait un échantillon de données représentées sur le graphique suivant :



Nous souhaiterions connaître des données statistiques telles que la moyenne, la médiane, les quartiles.

Dans un script spectre **.build**, nous allons voir comment mettre en place les différents calculs permettant d'insérer ces données statistiques (et d'autres).

Le script **.build** à construire est constitué de plusieurs éléments :



Le fichier d'entrée de type texte contient deux colonnes **Observations** et **Données**. L'objet **calc** contient les différents calculs que nous allons détailler par la suite. Enfin un objet de sortie va permettre de créer le cBase.

Au niveau du fichier d'entrée, les deux colonnes prises en compte sont définies de la manière suivante :

Source :::::

Path (1)

Input Files	/data/data1.txt
-------------	-----------------

Columns (3)

Column	Type	Keep	Remove
Position	integer	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Observations	string	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Données	integer	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Source Test results

Il y a également une autre colonne **Position** correspondant au numéro de ligne.

Au niveau de l'objet **calc**, les calculs suivants ont été définis :

	Opération	Expression
1	Median	top(median(value("Données")))
1	Percentile 25	top(percentile(value("Données"),25))
1	Percentile 50	top(percentile(value("Données"),50))
1	Percentile 75	top(percentile(value("Données"),75))
1	Average	top(average(value("Données")))
1	Geomean	top(geomean(value("Données")))
1	Harmean	top(harmean(value("Données")))
1	Slope	top(slope(value("Données"), value("Position"))))
1	Intercept	top(intercept(value("Données"), value("Position"))))
1	Trend	value("Position")*value("Slope")+value("Intercept")

Au niveau du script, ces calculs sont codés de la façon suivante :

```
add "Median" `top(median(value("Données")))` name="Calculs"
add "Percentile 25" `top(percentile(value("Données"),25))` name="Calculs"
add "Percentile 50" `top(percentile(value("Données"),50))` name="Calculs"
add "Percentile 75" `top(percentile(value("Données"),75))` name="Calculs"
add "Average" `top(average(value("Données")))` name="Calculs"
add "Geomean" `top(geomean(value("Données")))` name="Calculs"
add "Harmean" `top(harmean(value("Données")))` name="Calculs"
add "Slope" `top(slope(value("Données"), value("Position"))))` name="Calculs"
add "Intercept" `top(intercept(value("Données"), value("Position"))))` name="Calculs"
add "Trend" `value("Position")*value("Slope")+value("Intercept")` name="Calculs"
```

La fonction top permet aux calculs d'afficher le résultat sur chaque ligne.

Nous avons également rajouté le calcul Trend qui permet de calculer la droite de régression linéaire de type : $y=ax+b$

Une fois le script .build exécuté, l'ouverture dans ProDiver du cBase résultant et une plongée sur la Dimension **Observations** donne ceci :

Observations	Données	Average	Geomean	Harmean	Intercept	Median	Percentile			Position	Slope
							25	50	75		
Totaux	281	281,00	189,44	130,02	-129,37	180,00	80,00	180,00	410,00	210	39,0
X1	2	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	1	1,9
X2	3	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	2	1,9
X3	3	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	3	1,9
X4	4	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	4	1,9
X5	4	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	5	1,9
X6	4	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	6	1,9
X7	5	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	7	1,9
X8	6	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	8	1,9
X9	7	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	9	1,9
X10	8	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	10	1,9
X11	10	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	11	1,9
X12	12	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	12	1,9
X13	15	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	13	1,9
X14	18	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	14	1,9
X15	20	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	15	1,9
X16	22	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	16	1,9
X17	25	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	17	1,9
X18	28	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	18	1,9
X19	35	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	19	1,9
X20	50	14,05	9,47	6,50	-6,47	9,00	4,00	9,00	20,50	20	1,9

Une fois les données calculées, le but maintenant est d'afficher dans DivePort les données sous forme de graphique à barres et d'afficher les lignes de régression linéaire, de la médiane et des différentes moyennes.

Pour cela, nous allons utiliser un Chart Portlet avec les propriétés suivantes :

Édition de Chart Portlet: i0001

Source :

Type :

Dimension :

Valeurs :

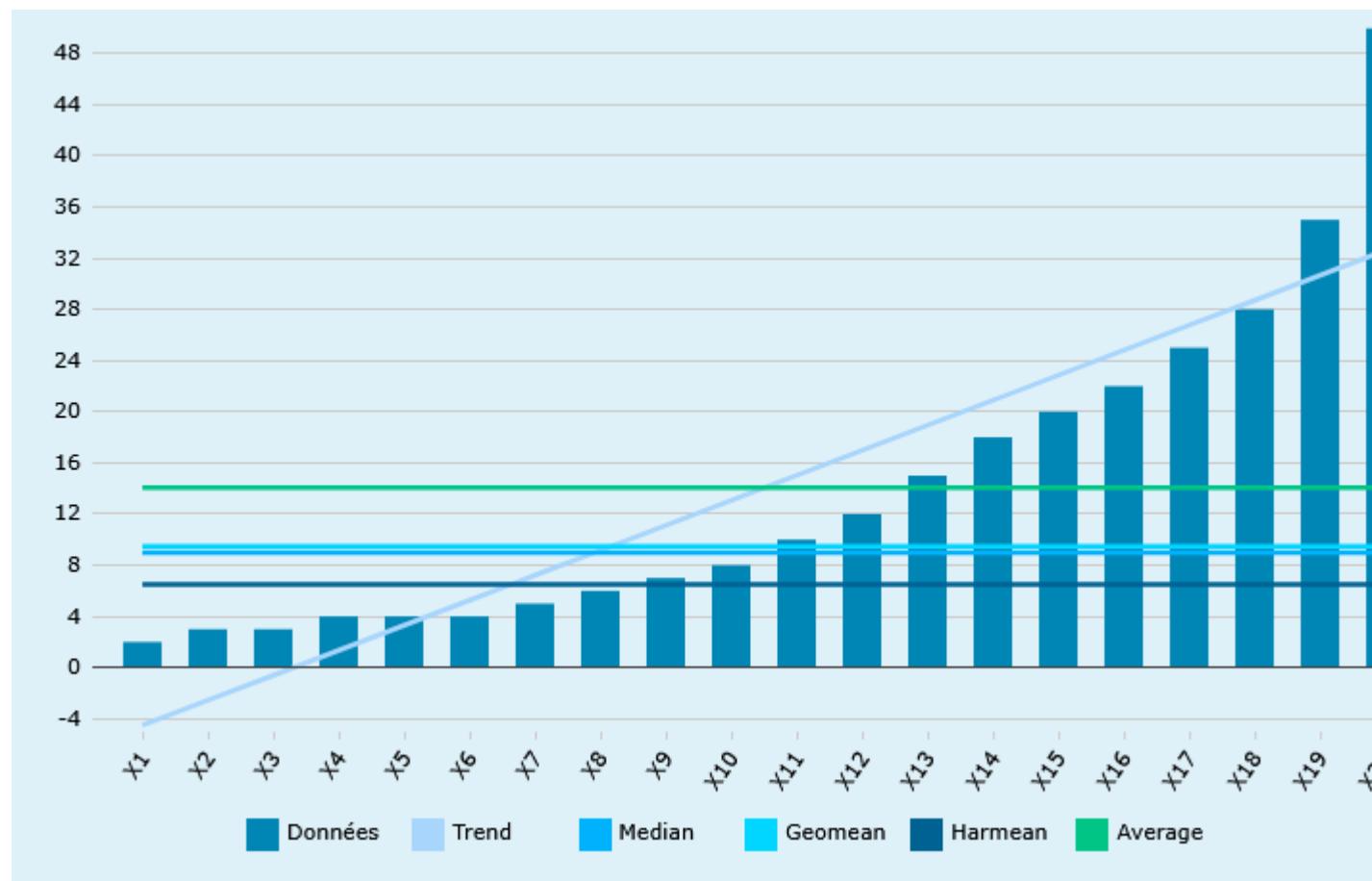
- Données
- Trend
- Median
- Geomean
- Harmean
- Average

[Ajouter une valeur...](#)

Paramètres supplémentaires :

[Axes](#) [Légende](#) [Texte](#) [Impression](#) [Affichage](#) [Avancé](#) [Pointage](#) [Actions sur clic](#)

Et voici le résultat :



Pour en faire plus, on peut également rajouter dans le titre les paramètres de la droite de régression linéaire.

Pour cela, aller dans **Texte > Paramètre du texte > Paramètre du titre** et ajouter dans **Texte du titre** l'information suivante :

`y= $(FIRST_VALUE_IN_COLUMN "Slope") * x + $(FIRST_VALUE_IN_COLUMN "Intercept")`

Voici le titre résultant :

y= 1,9541353 * x + -6,4684211

Rappels formules

Moyenne arithmétique (average) :

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Moyenne géométrique (geomean) :

$$\bar{x}_g = (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{\frac{1}{n}} = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}}$$

Pour des observations x1,x2,..., xn strictement positives

Moyenne harmonique (harmean) :

$$\bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

Pour des observations x1,x2,..., xn non nulles

Médiane (median) - également 50^e centile (percentile) :

Une fois les valeurs triées par ordre croissant, si le nombre de valeurs est **impair**, c'est la valeur du milieu :

$$\text{Médiane} = x_{\left(\frac{n+1}{2}\right)}$$

Une fois les valeurs triées par ordre croissant, si le nombre de valeurs est **pair**, c'est la moyenne des deux valeurs centrales :

$$\text{Médiane} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

Tags

1. calcul
2. cBase
3. DivePort
4. Portlet
5. script
6. Spectre
7. Workbench