Logiciel XLSTAT version 7.0

Contact : Addinsoft

40 rue Damrémont

75018 PARIS

2005 - 2006

Plan

- Présentation générale du logiciel
- Statistiques descriptives
- Histogramme – Discrétisation
- Tableau de contingence et Khi deux
- Régression linéaire
- Codage
- Quelques autres fonctionnalités
- Sources

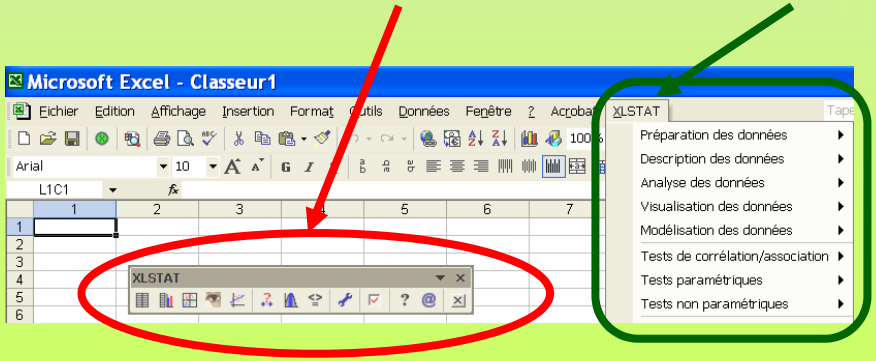
Présentation générale du logiciel

Caractéristiques du logiciel

XLSTAT permet d'utiliser les techniques

- **de statistique**
- **d'analyse de données**
- **de modélisation** *sans quitter Microsoft Excel*

- Sa particularité = être parfaitement intégré à Excel.
- On lance XLSTAT ⇨ barre d'outils et menu apparaissent



Caractéristiques du logiciel

- Une interface intuitive

- XLSTAT utilise Excel comme une interface

  de récupération des données

  et d'affichage des résultats.

- Tous les calculs sont réalisés en dehors d'Excel.

Quelques applications

- Marketing quantitatif
- Analyse d'enquêtes
- Analyse de portefeuille
- Modélisation de risques
- Analyse de données de survie
- Contrôle qualité
- Analyse sensorielle

Modules complémentaires à XLSTAT Pro

***XL*STAT-3DPlot** : pour la visualisation de données en 3 dimensions.

***XL*STAT-Time** : pour l'analyse des séries chronologiques

***XL*STAT-SPC** : pour le contrôle statistique des procédés

***XL*STAT-Life** : pour l'analyse de données de survie

***XL*STAT-Pivot** : pour créer des tableaux croisés intelligents

***XL*STAT-MX** : pour l'analyse sensorielle

***XL*STAT-Dose** : pour l'analyse des effets de dose

La barre d'outils principale de XLSTAT Pro

Préparation des données

Description des données

Analyse des données

Visualisation des données

Modélisation des données

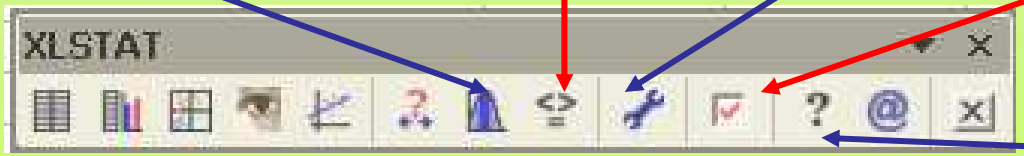
Tests de corrélation / association

Tests paramétriques

Tests non paramétriques

Utilitaires

Options



Aide

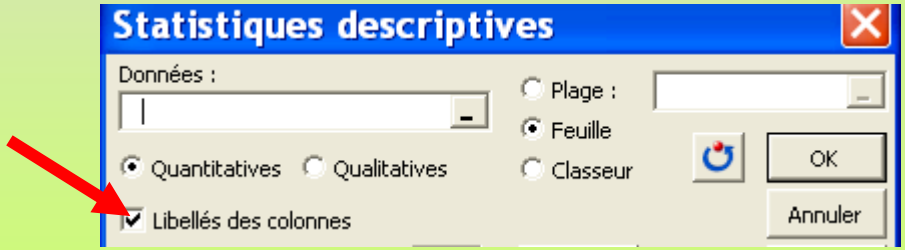
Sélection des données

Sélection rapide par colonnes :

Possible seulement si les données commence à la ligne 1.

Pour sélectionner une colonne, on clique sur la lettre de la colonne. *XL*STAT détecte automatiquement la fin du tableau de données par la présence d'une ligne vide.

L'option "Libellés des colonnes" permet de préciser à *XL*STAT si la 1re ligne contient des données ou des libellés.



Sélection multiple possible : on laisse la touche Ctrl enfoncée pendant la sélection des colonnes non adjacentes.

Sélection des données

Sélection par plages

- A utiliser quand le tableau de données ne commence pas dès 1re ligne de la feuille Excel.
- La plage de données est sélectionnée avec la souris.
- Sélection multiple possible : on laisse la touche Ctrl enfoncée pendant la sélection de plages de données non adjacentes.

Sélection des données

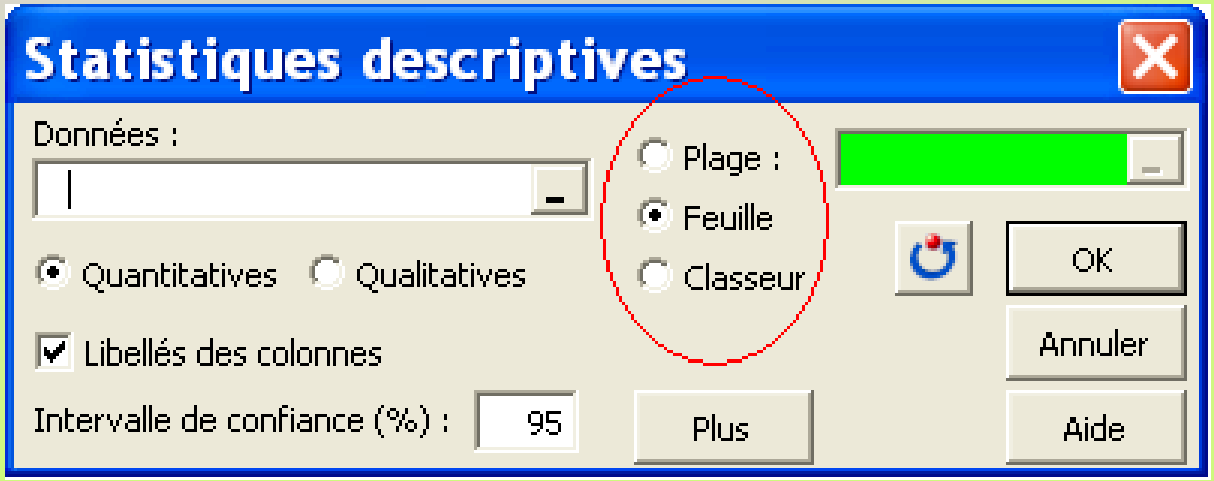
Raccourcis clavier :

- On peut sélectionner des données en utilisant les touches **Ctrl**, **Shift,** **déplacement** haut, bas, gauche, droite.

- On peut utiliser la touche **Tab** pour passer d'un élément au suivant dans la boîte de dialogue.

Affichage des résultats dans Excel

Dans chaque boîte de dialogue,
***XL*STAT propose 3 modes d'affichage des résultats :**



Affichage des résultats dans Excel

- **Plage :** les résultats sont affichés dans la feuille active, à partir d'une cellule **(à sélectionner)** qui sera le coin supérieur gauche de la plage

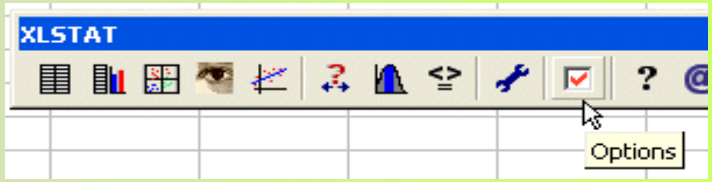
- **Feuille :** les résultats sont affichés dans une nouvelle feuille Excel créée par *XL*STAT

- **Classeur :** les résultats sont affichés dans un nouveau classeur Excel.

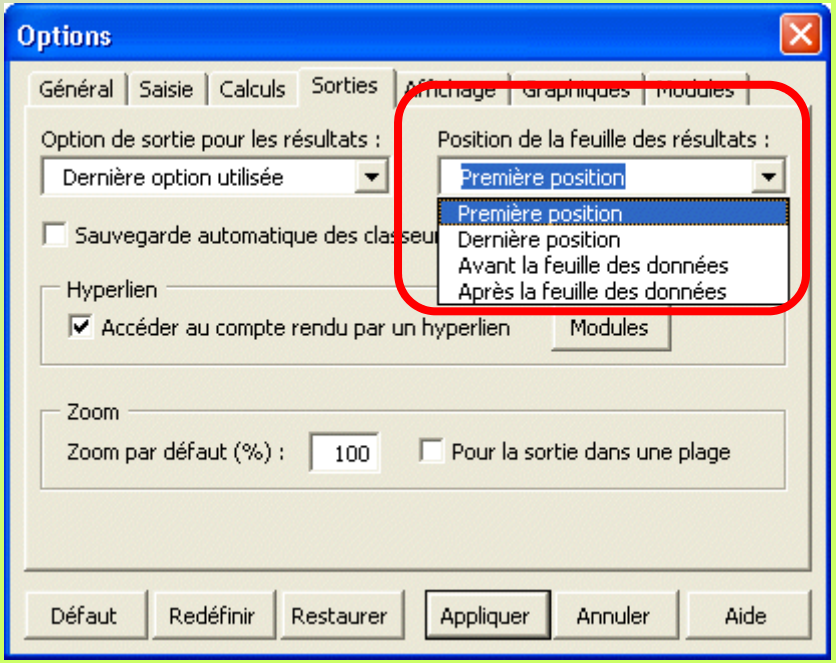
Affichage des résultats dans Excel

A propos de l'option "Feuille" : On peut choisir parmi différentes options d'affichage des résultats.

Cliquer sur le bouton Options



Puis sur l'onglet "Sortie"



Florence Nicolau

Affichage des résultats dans Excel

Remarques :

- Option "Feuille" :

  *XL*STAT donne à la feuille de résultats le nom de la méthode employée **(suivi si besoin est d'un numéro)**.

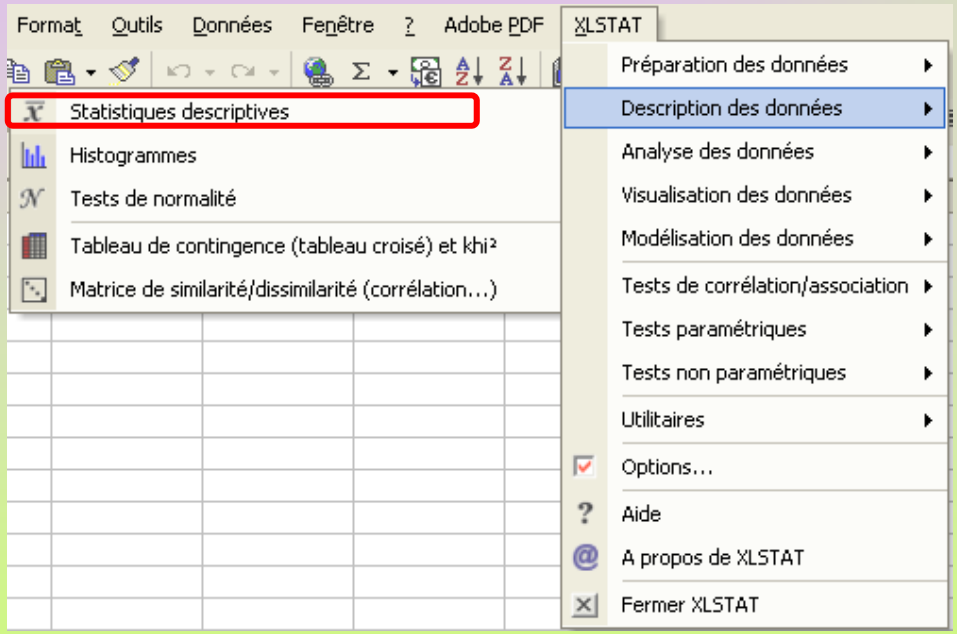
- Option "Classeur" :

  conseillée pour de gros volumes de données.

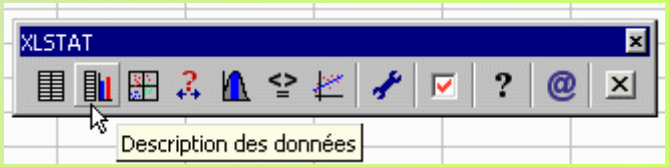
Statistiques descriptives

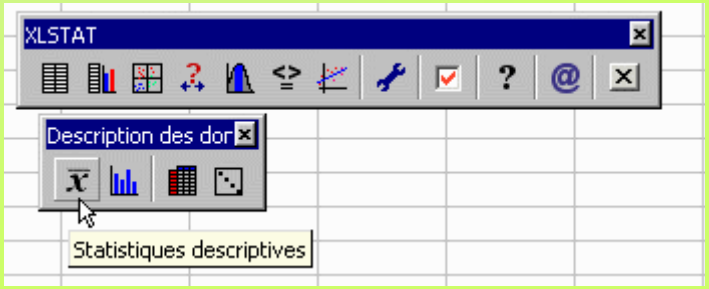
Statistiques descriptives

Ouverture par le menu XLSTAT :



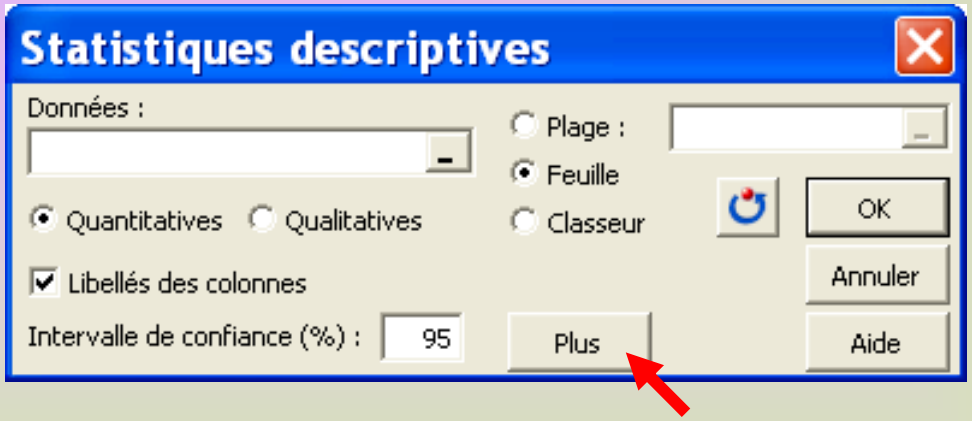
Ouverture par la barre d'outils XLSTAT :





Florence Nicolau

Statistiques descriptives – La fenêtre :



Après un clic sur le bouton Plus, la fenêtre complète s'affiche

Florence Nicolau

Statistiques descriptives – Ce qu'il faut compléter :

Données : sélectionner la plage des variables à décrire.

S'il y a des valeurs manquantes dans une colonne, XLSTAT propose de les ignorer. En cas de refus, le traitement est abandonné.

" quantitatives " / " qualitatives " :

choisir le type des variables

Libellés des colonnes :

à cocher si la 1re ligne contient le nom des variables

Statistiques descriptives – Données pondérées

Poids :

case à cocher si on veut pondérer les données

- puis sélectionner la colonne des poids

Les valeurs manquantes dans la colonne « Poids » sont remplacées par 0 et impliquent la non prise en compte de la ligne concernée.

- Cette fenêtre Statistiques Descriptives permet ainsi de faire des calculs que Excel ne sait pas faire.

Exemple : calcul de moyenne, écart type… en tenant compte du poids de chaque observation

Statistiques descriptives – Définition de groupes

- **Descripteur de groupe :**

  case à cocher si une variable doit définir des sous populations (groupes)

  puis sélectionner la colonne correspondante

- **Comparer :**

  case à cocher pour comparer les résultats obtenus pour chaque groupe avec ceux obtenus pour la totalité des valeurs.

Statistiques descriptives – **Liste des statistiques calculées dans le cas des données quantitatives**

Nbr de valeurs utilisées : nombre de valeurs effectivement utilisées dans les calculs (les valeurs manquantes ou poids nul sont exclues)

Nbr de valeurs ignorées : nombre de valeurs manquantes ou de poids nul

Nbr de val. min. : nombre de valeurs égales à la valeur minimale

% de val. min. : pourcentage du nombre de valeurs égales à la valeur minimale

- **Minimum** : valeur minimale

Statistiques descriptives – **Liste des statistiques calculées dans le cas des données quantitatives**

1er quartile

Médiane

3ème quartile

Maximum : valeur maximale

Étendue = maximum - minimum

Statistiques descriptives – **Liste des statistiques calculées dans le cas des données quantitatives**

Somme des poids des valeurs utilisées dans le cas de données pondérées

- **Total** : somme des valeurs, éventuellement pondérée

Moyenne : éventuellement pondérée

Moyenne géométrique : Elle n'est pas définie s'il y a des valeurs négatives ou nulles.

- **Moyenne harmonique** : Elle n'est pas définie s'il y a des valeurs nulles.

Statistiques descriptives – **Liste des statistiques calculées dans le cas des données quantitatives**

Aplatissement (Pearson) :

- valeur négative ⇨ distribution plus plate que la loi normale
- valeur positive ⇨ distribution plus pointue que la loi normale

Asymétrie (Pearson) : par rapport à sa moyenne.

- valeur négative ⇨ distribution étalée vers la gauche
- valeur positive ⇨ distribution étalée vers la droite

Aplatissement = le coefficient d'aplatissement calculé par Excel

Asymétrie = le coefficient d'asymétrie calculé par Excel

Statistiques descriptives – **Liste des statistiques calculées dans le cas des données quantitatives**

CV : coefficient de variation (écart-type/moyenne)

- **Variance d'échantillon** : variance des données

Variance estimée : estimation de la variance d'une population dont les données constituent un échantillon (dénominateur = n -1)

Écart type d'échantillon

Écart type estimé

- **Écart absolu moyen** = la moyenne des valeurs absolues des écarts de chaque valeur par rapport à la moyenne

- **Écart type de la moyenne**

Écart absolu médian = la médiane des écarts à la médiane

Statistiques descriptives – **Liste des statistiques calculées dans le cas des données qualitatives**

Nbr de modalités : nombre de modalités de la variable

- **Mode**

Effectif mode (*dans le cas de données non pondérées)*

Poids mode *(dans le cas de données pondérées)*

% mode : fréquence relative du mode exprimée en %

- **Fréq. rel. mode** : fréquence relative du mode.

Statistiques descriptives – **Liste des statistiques calculées dans le cas des données qualitatives**

Pour chaque modalité :

Effectif : dans le cas de données non pondérées

Poids : dans le cas de données pondérées

% : fréquence relative de la modalité exprimée en %

Fréq. rel. : fréquence relative de la modalité

Statistiques descriptives – **Liste des statistiques calculées dans le cas des données qualitatives**

Exemple de variable qualitative : variable « Parcours »

Synthèse :

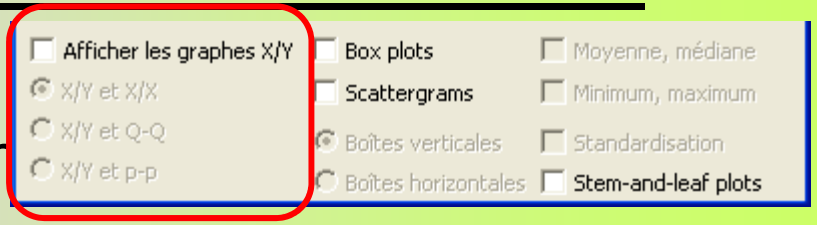
Variable	Nbr. De données utilisées	Nbr. de données ignorées	Nombre de modalités	Mode	Fréquence mode	% mode	Fréq. rel. mode
Parcours	64	0	6	SSS	33	51,56	0,516

Parcours / 6 Modalités :

Modalité	Effectif	%	Fréq. rel.
EEE	12	18,75	0,188
EES	4	6,25	0,063
ESS	13	20,31	0,203
SRR	1	1,56	0,016
SRS	1	1,56	0,016
SSS	33	51,56	0,516

Statistiques descriptives – **Graphiques dans le cas des variables quantitatives**

Afficher les graphes X/Y :



Tous les nuages bivariés obtenus en croisant 2 à 2 toutes les variables quantitatives

" X/Y et X/X " " X/Y et Q-Q " " X/Y et p-p " : pour choisir les nuages bivariés à afficher

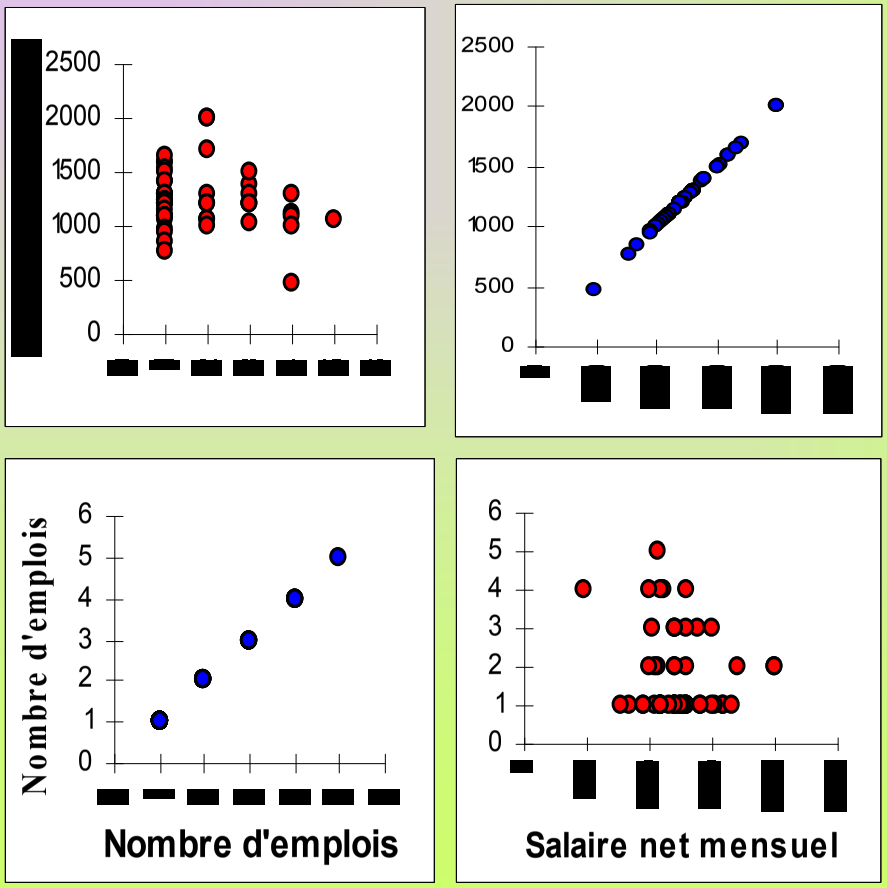
graphiques " Quantile-Quantile " ou Q-Q plots

graphiques " probabilité- probabilité " ou p-p plots

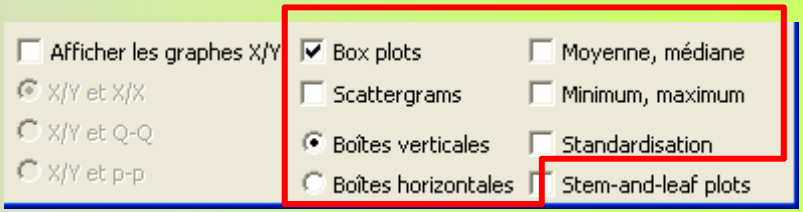
Ils ne peuvent pas être affichés s'il y a plus de 6 variables ou plus de 30 000 points.

Statistiques descriptives – **Graphiques dans le cas des variables quantitatives**

Exemple de graphe " X/Y et X/X " :



Statistiques descriptives – **Graphiques dans le cas des variables quantitatives**



- **Box plots** :

  graphiques " boîte à moustaches ".

  Non affichés s'il y a plus de 16 variables ou plus de 30 000 points.

- **Scattergrams** : nuages de points univariés.

  Non affichés s'il y a plus de 24 variables ou plus de 30 000 points.

Statistiques descriptives – **Graphiques dans le cas des variables quantitatives**

Les options pour les box plots et scattergrams

Moyenne, médiane : pour afficher ces valeurs sur le graphique

Minimum, maximum : pour afficher ces valeurs sur le graphique

" Boîtes verticales " / " Boîtes horizontales " :
pour choisir l'orientation des box plots et des scattergrams.

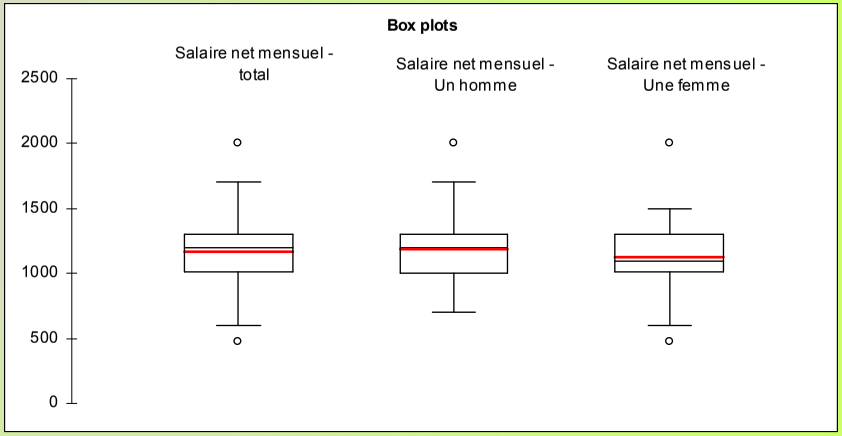
Standardisation : pour supprimer les différences d'ordre de grandeur entre les variables dans les box plots et scattergrams,
(On divise les valeurs de chaque variable par l'écart type)

Statistiques descriptives – **Graphiques dans le cas des variables quantitatives**

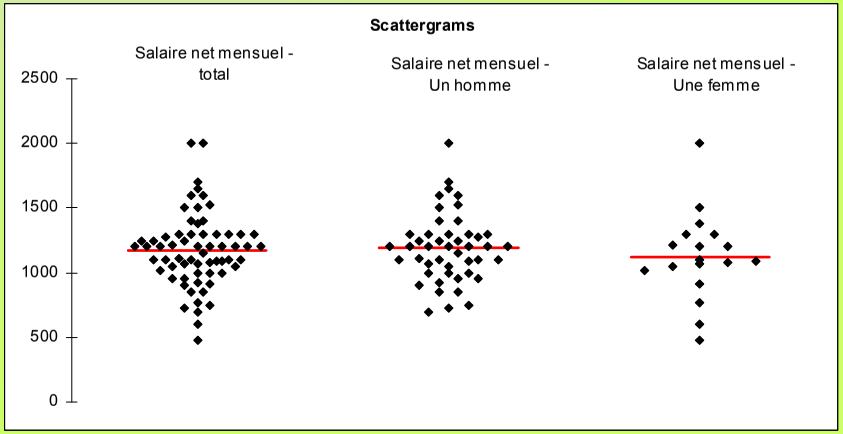
Exemple de variable quantitative : variable « Salaire net mensuel »

La variable « sexe » définit 2 groupes.

Box plot

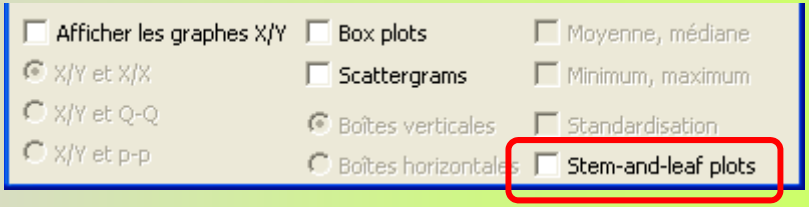


Scattergrams



Florence Nicolau

Statistiques descriptives – **Graphiques dans le cas des variables quantitatives**



- **Stem and leaf plots** :

  diagrammes " tige et feuille "

Remarque : Les graphiques générés par *XL*STAT peuvent être modifiés comme n'importe quel autre graphique Excel.

Statistiques descriptives – **Graphiques dans le cas des variables qualitatives**

Histogramme

pour le graphique en tuyaux d'orgue

Secteurs

pour le graphique en secteurs circulaires

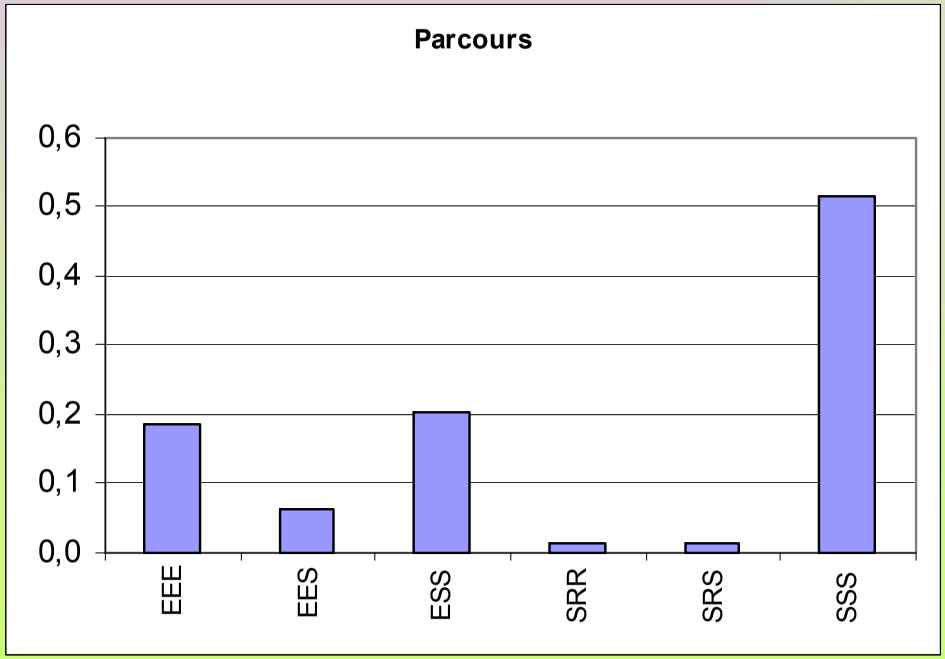
Regrouper graphes

pour afficher tous les graphiques sur une feuille à part

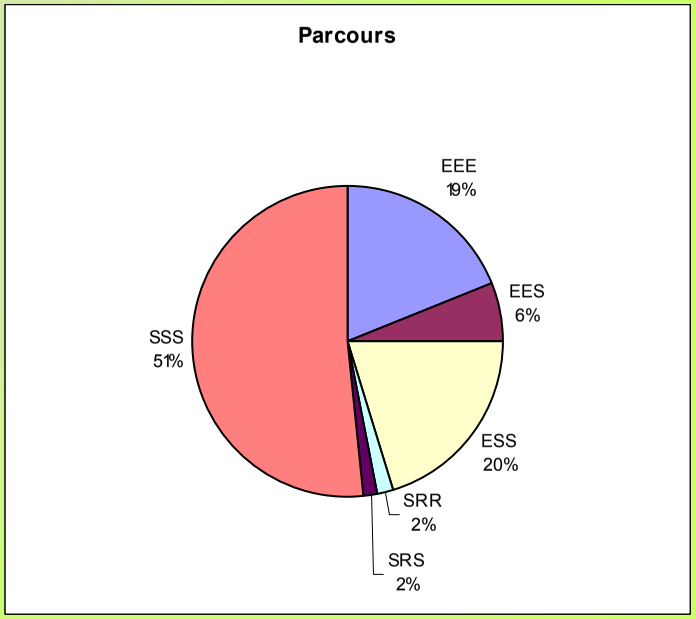
Statistiques descriptives – **Graphiques dans le cas des variables qualitatives**

Exemple de variable qualitative : variable « Parcours »

Tuyaux d'orgue :



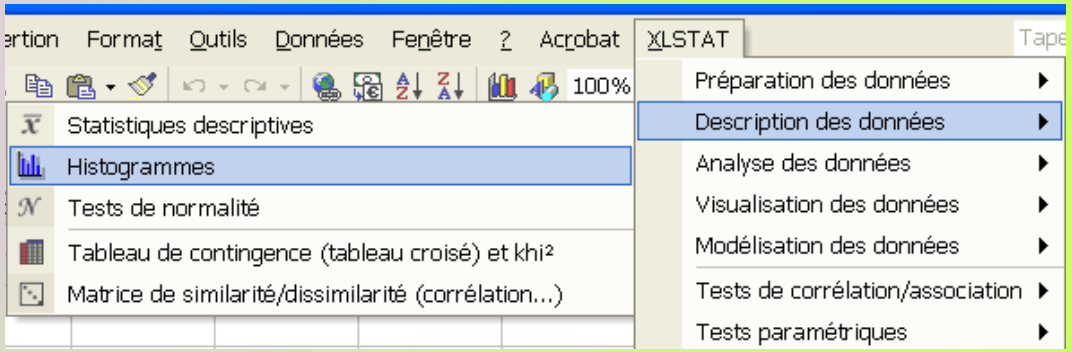
Secteurs circulaires :



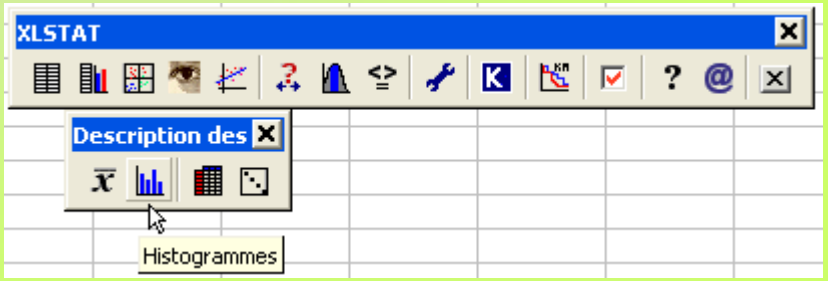
Histogramme – Discrétisation

Histogramme

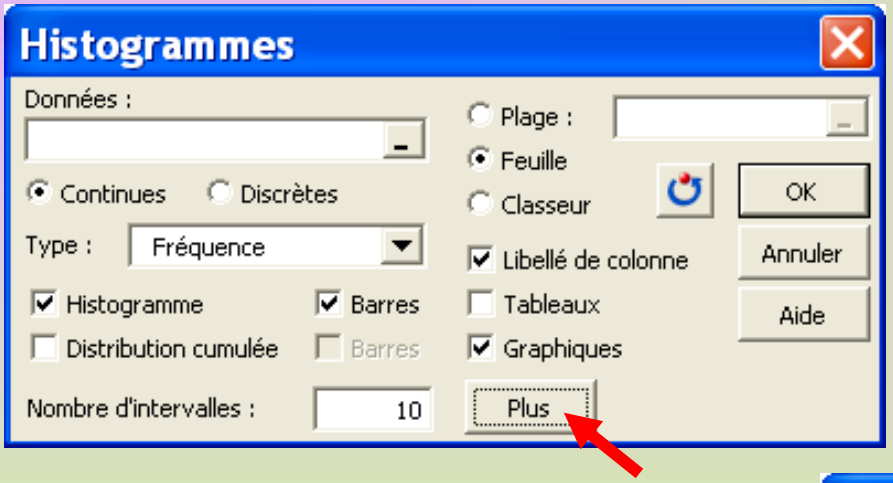
- Ouverture par le menu XLSTAT



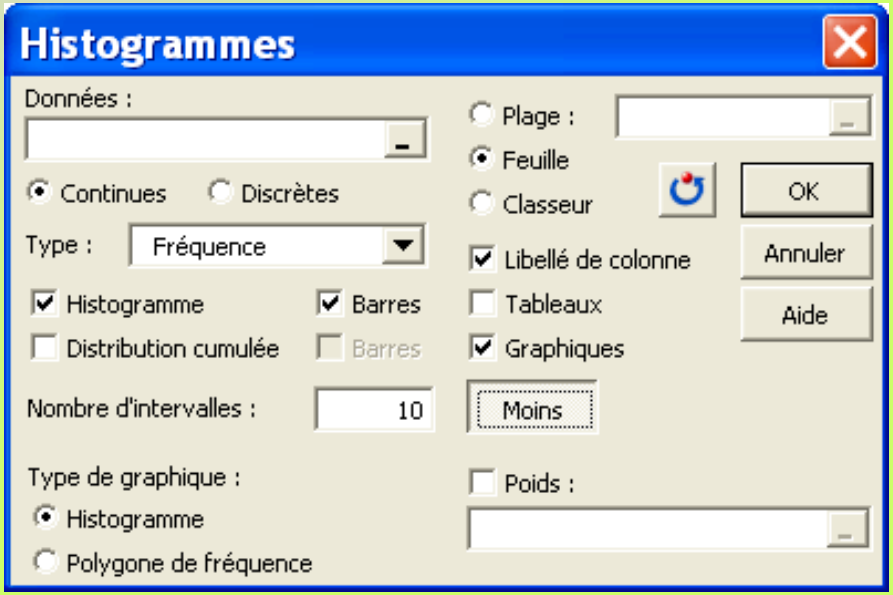
- Ouverture par la barre d'outils XLSTAT



Histogramme – **La fenêtre :**



Après un clic sur le bouton Plus, la fenêtre complète s'affiche



Florence Nicolau

Histogramme – **Ce qu'il faut compléter :**

Il faut sélectionner les données.

- Il faut choisir le type de la variable : discrètes ou continues

- Il faut choisir le type de graphique :

  histogramme

  polygones de fréquences

  distribution cumulée (courbe cumulative)

Histogramme – **Ce qu'il faut compléter :**

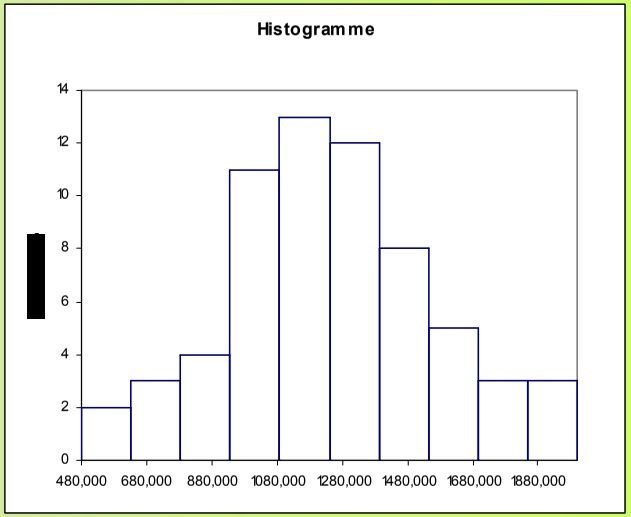
Et éventuellement, choisir

- Le nombre d'intervalles pour une variable continue
- Le poids pour des données pondérées
- Afficher le tableau avec les résultats

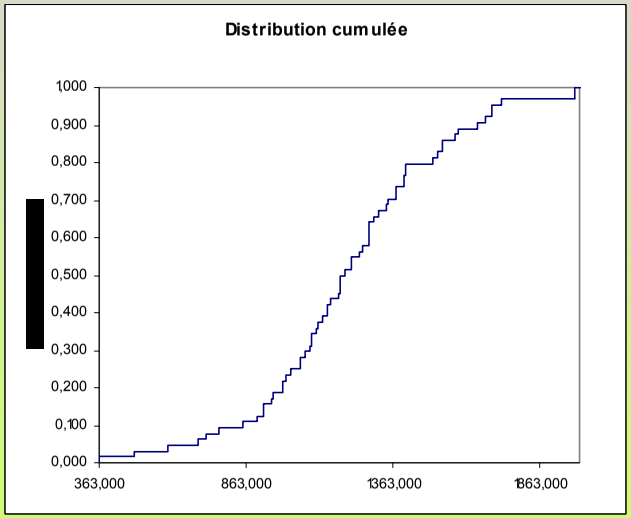
Histogramme – **Les différents graphiques :**

- Pour une variable quantitative continue :

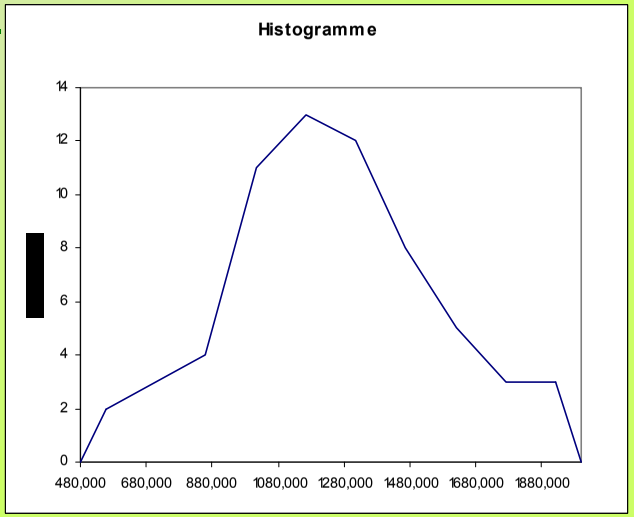
Histogramme



Courbe cumulative



Polygone des fréquences

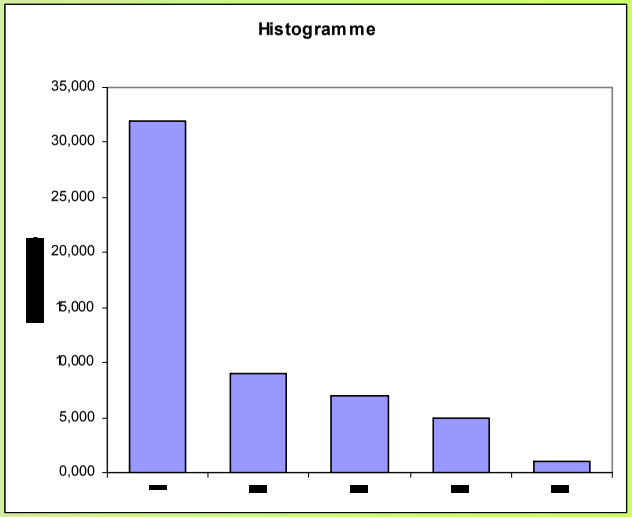


Variable « Revenu mensuel net »

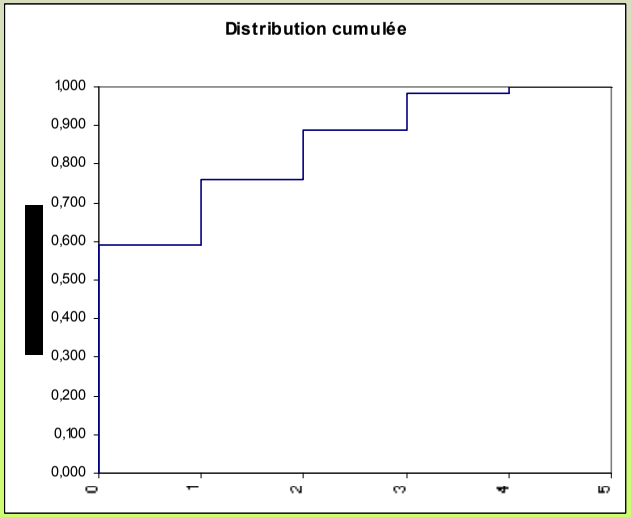
Histogramme – **Les différents graphiques :**

- Pour une variable discrète :

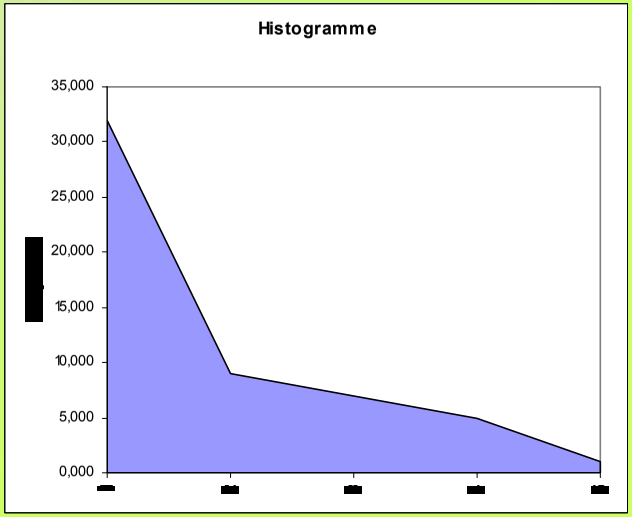
Diagramme en bâtons



Courbe cumulative



Polygone des fréquences

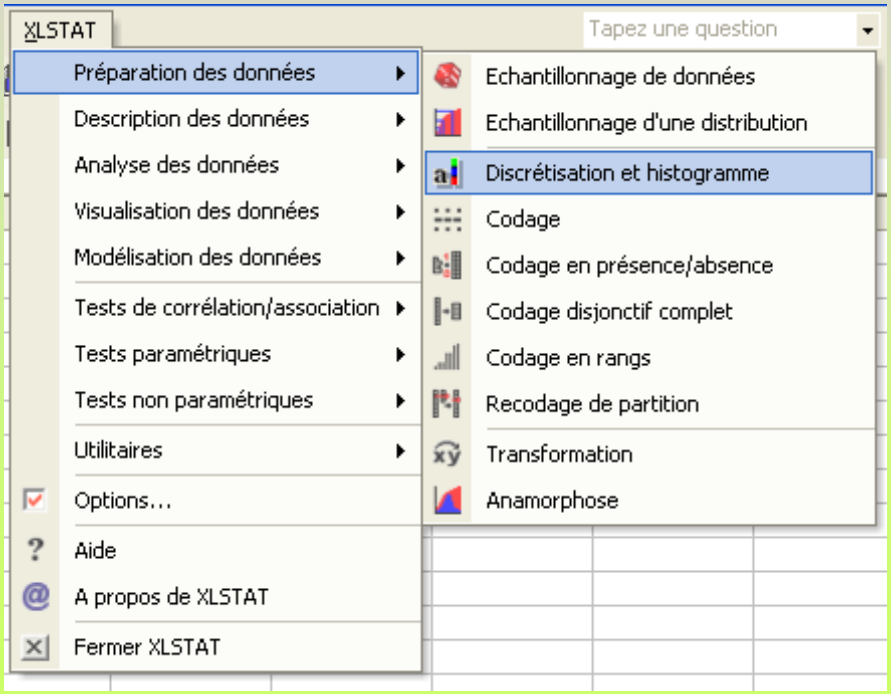


Variable « Nombre d'emplois »

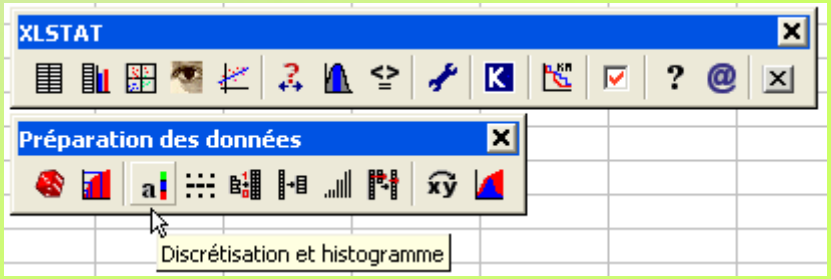
Discrétisation et histogramme

Cas d'une variable continue : L'outil "Discrétisation et histogramme" permet de créer librement des classes.

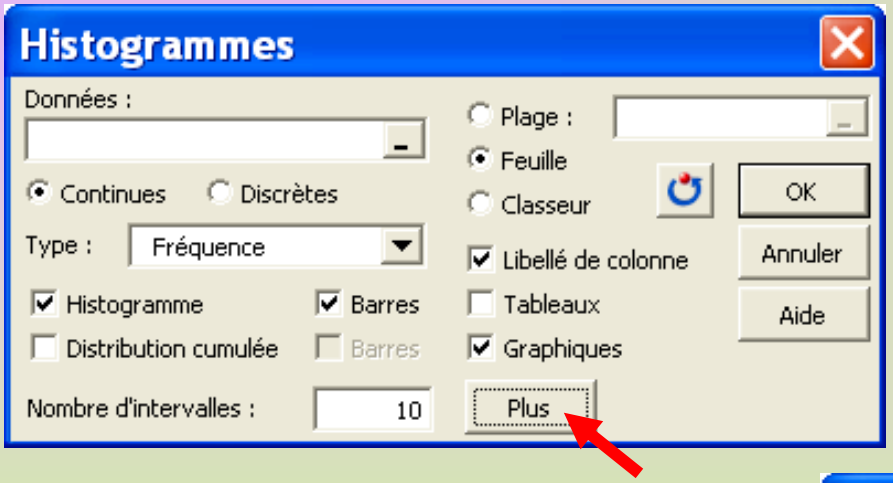
Ouverture par le menu XLSTAT



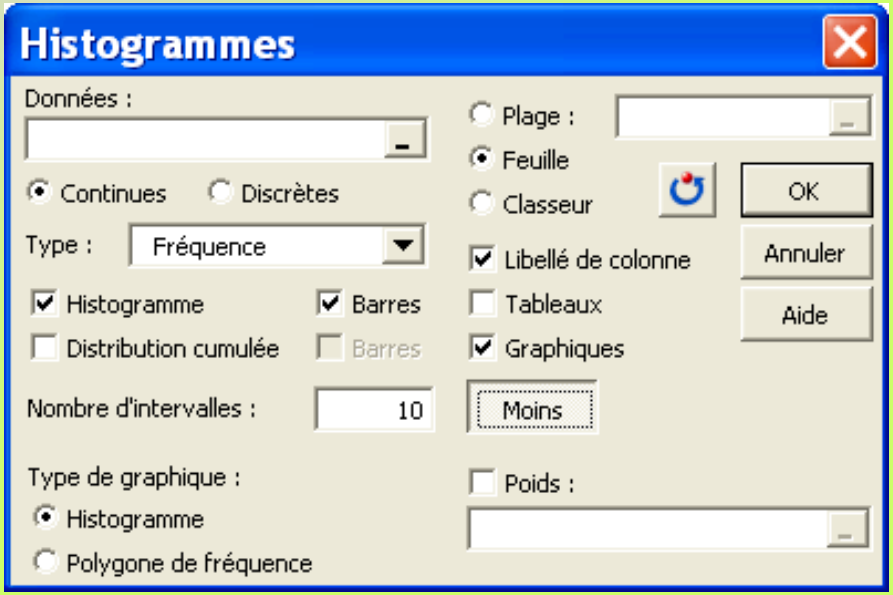
Ouverture par la barre d'outils XLSTAT



Histogramme – **La fenêtre :**



Après un clic sur le bouton Plus, la fenêtre complète s'affiche



Florence Nicolau

Discrétisation et histogramme – La fenêtre :

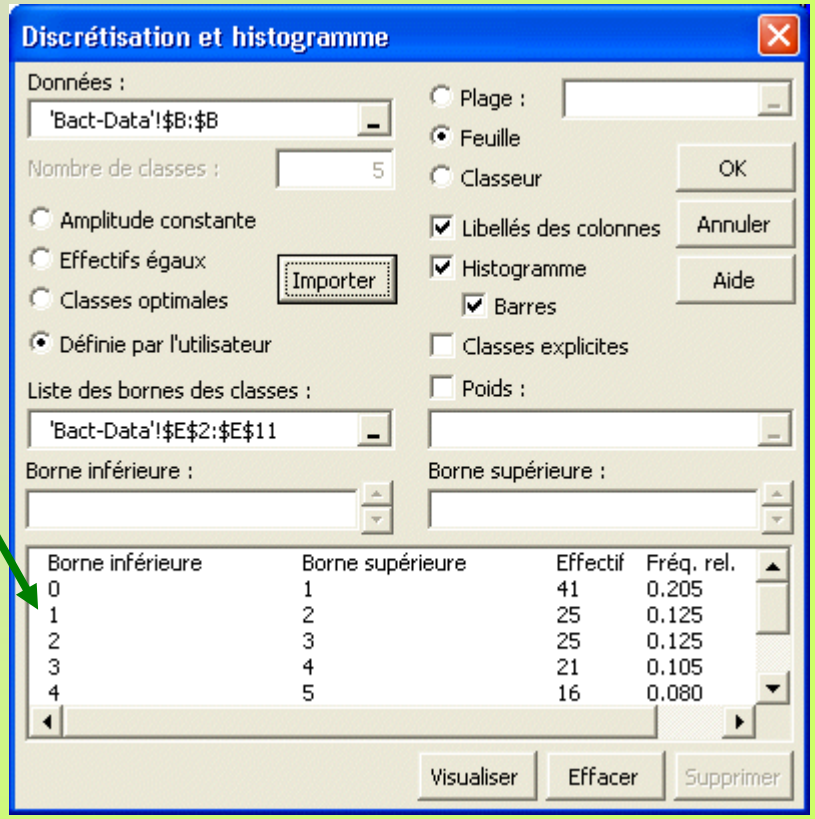
- On coche « Définie par l'utilisateur ».
- L'option « Liste des bornes » est activée.
- On sélectionne la liste de bornes.

  On clique « Importer ».

Discrétisation et histogramme – La fenêtre :

Les bornes et les fréquences apparaissent dans la partie inférieure de la boîte.

- On peut éventuellement modifier les bornes.



Discrétisation et histogramme – Graphique obtenu :

- On clique sur Ok et on obtient un histogramme selon les classes choisies.

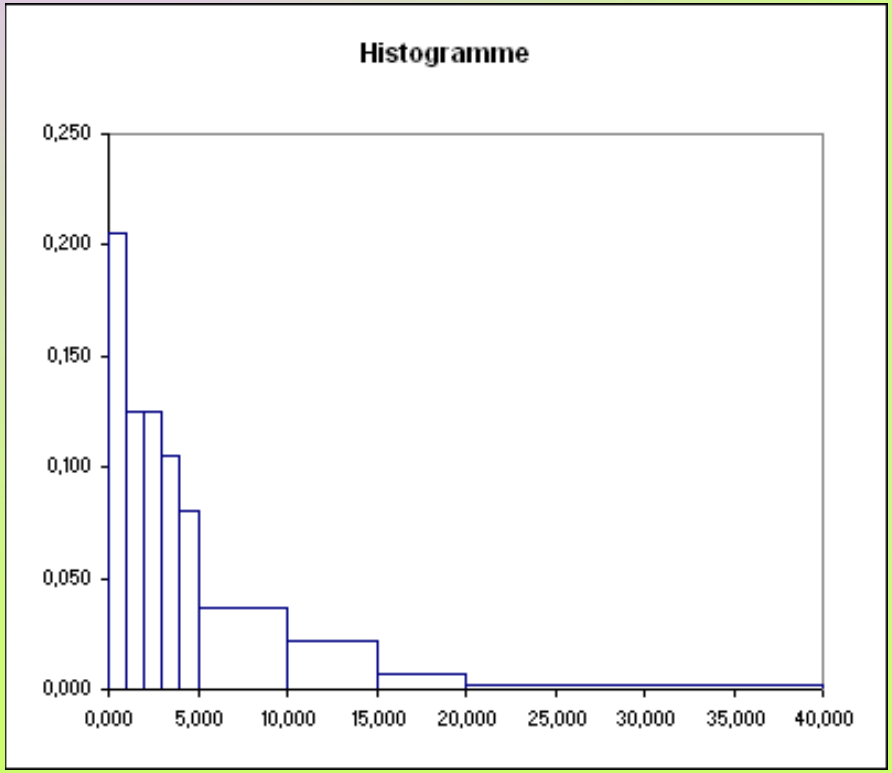
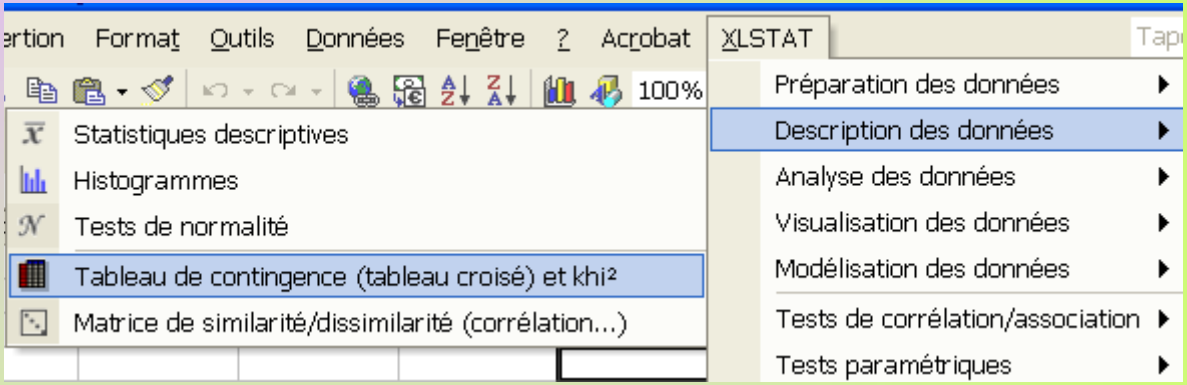


Tableau de contingence et Khi deux

Tableau de contingence (Tableau croisé) et Khi deux

- Ouverture par le menu XLSTAT



- Ouverture par la barre d'outils XLSTAT

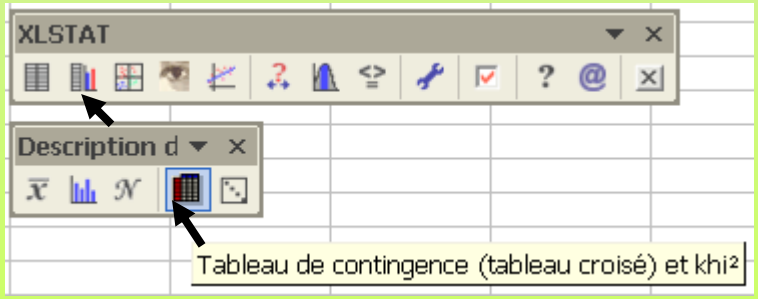
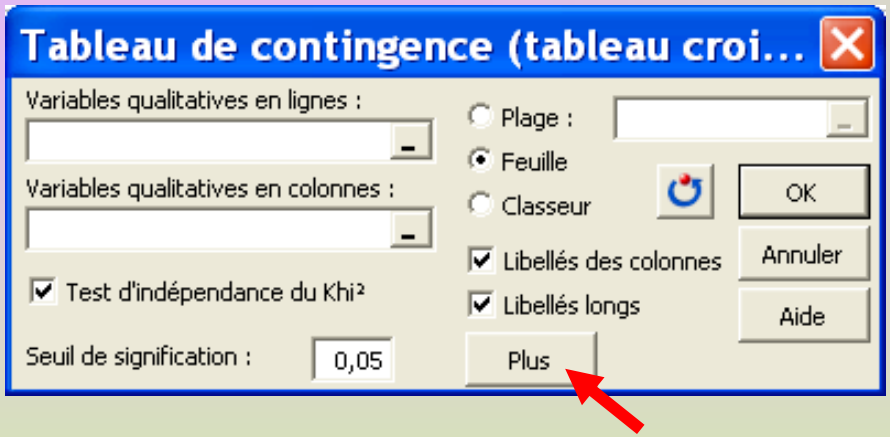
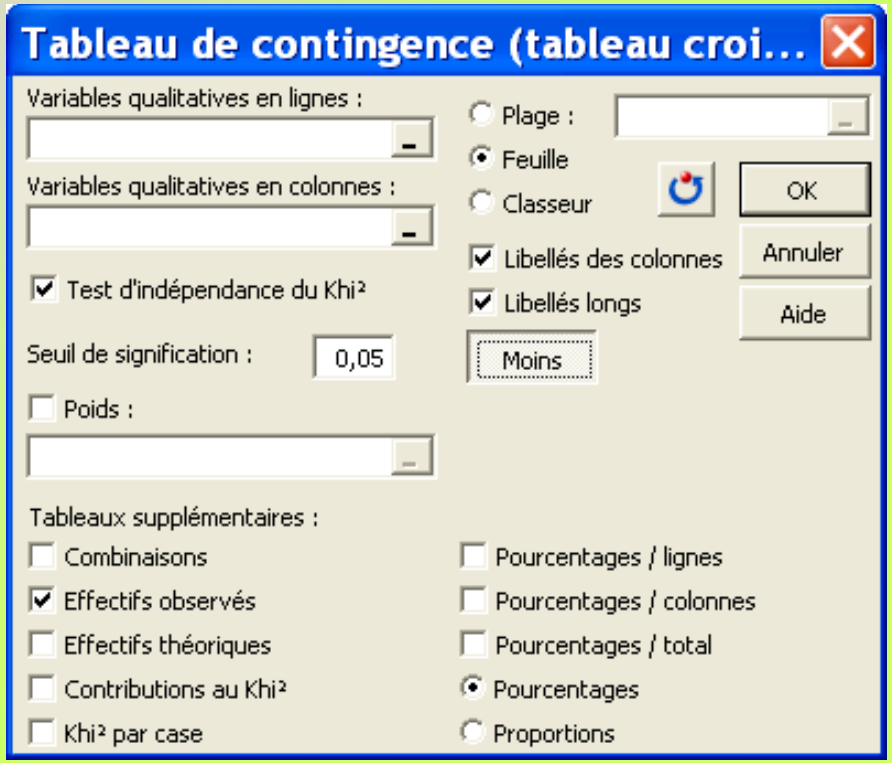


Tableau de contingence… – **La fenêtre :**



Après un clic sur le bouton Plus, la fenêtre complète s'affiche



Florence Nicolau

Tableau de contingence… – **La fenêtre :**

- On sélectionne la variable à mettre en ligne et celle à mettre en colonne.

On sélectionne éventuellement le poids des données

On choisit les effectifs, les fréquences totales ou conditionnelles

On peut calculer un test du Khi² et afficher les effectifs théoriques, les contributions au Khi².

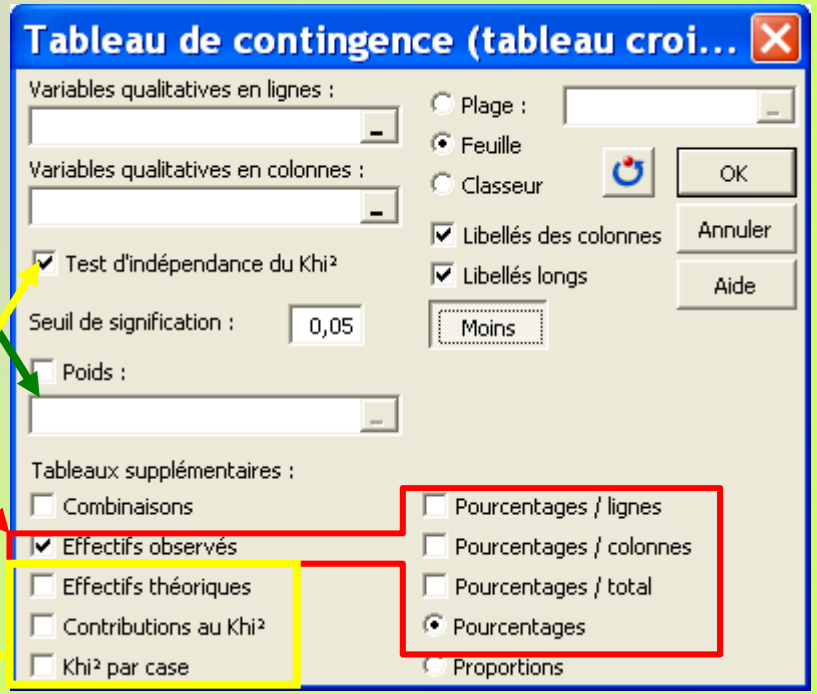


Tableau de contingence… – **Les résultats :**

- Tableau des n_{ij} et $f_{j/i}$:

Tableau des effectifs observés :

	Activité entreprise - Autres ser	té entreprise -	prise - Banqu	mmerce, tran	entreprise - Inds (administr	Total
Sexe - Un homme	6	3	3	13	12	9	46
Sexe - Une femme	7	0	2	4	0	3	16
Total	13	3	5	17	12	12	62

Tableau des pourcentages / lignes :

	Activité entreprise - Autres ser	té entreprise -	prise - Banqu	mmerce, tran	entreprise - Inds (administr	Total
Sexe - Un homme	13,04	6,52	6,52	28,26	26,09	19,57	100
Sexe - Une femme	43,75	0,00	12,50	25,00	0,00	18,75	100
Total	20,97	4,84	8,06	27,42	19,35	19,35	100

- Khi deux :

Test du Khi² d'indépendance :

Khi² (valeur observée)	11,132
Khi² (valeur critique)	11,070
ddl	5
p-value unilatérale	0,049
Alpha	0,05

Valeur critique :
valeur de la statistique sous l'hypothèse nulle pour la probabilité 1-alpha (test unilatéral à droite).

On rejette l'hypothèse nulle lorsque la valeur observée est supérieure à la valeur critique.

Conclusion :

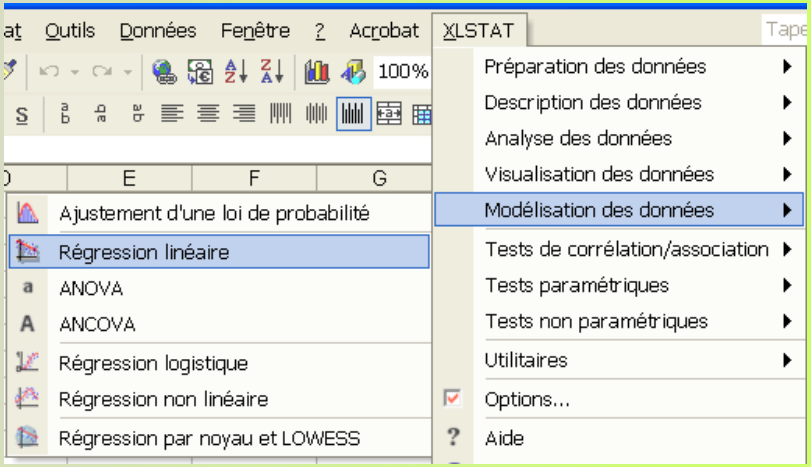
Au seuil de signification alpha=0,050 on peut rejeter l'hypothèse nulle d'indépendance entre les lignes et les colonnes.
Autrement dit, la dépendance entre les lignes et les colonnes est significative.

Florence Nicolau

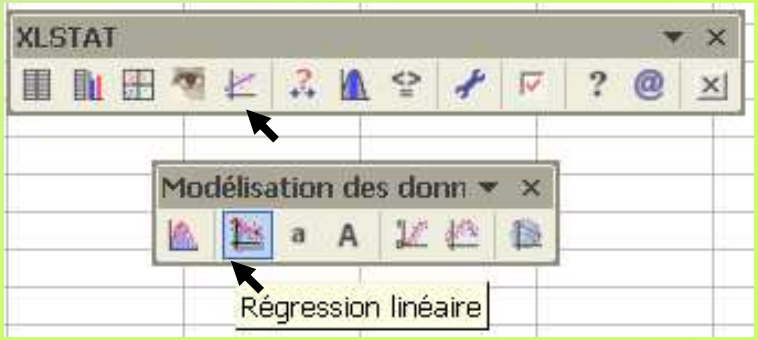
Régression linéaire

Régression linéaire

- Ouverture par le menu XLSTAT



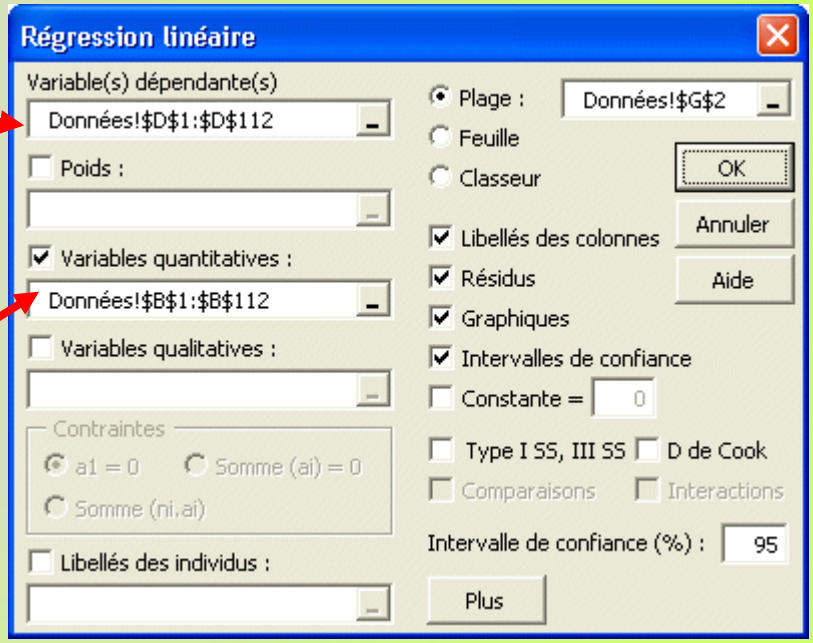
- Ouverture par la barre d'outils XLSTAT



Régression linéaire – **La fenêtre :**

La « Variable dépendante » est la variable expliquée (variable à modéliser)

La « Variable quantitative » est la variable explicative.



Exemple : On veut expliquer « le poids » par « la taille ».

La variable dépendante est « le poids ».

La variable quantitative est « la taille ».

Régression linéaire – **Les résultats :**

Résumé pour la variable dépendante :

Variable	Nbr. de valeurs total	Nbr. de valeurs utilisées	Nbr. de valeurs ignorées	Somme des poids	Moyenne	Ecart-type
Poids	111	111	0	111	98,878	18,616

Résumé pour les variables quantitatives :

Variable	Moyenne	Ecart-type
Taille	60,526	3,358

Coefficients d'ajustement :

R (coefficient de corrélation)	0,751
R^2 (coefficient de détermination)	0,564
R^2aj. (coefficient de détermination ajusté)	0,560
SCR	16614,585

Evaluation de la valeur de l'information apportée par les variables (H0 = Y=Moy(Y)) :

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Pr > F
Modèle	1	21506,523	21506,523	141,094	< 0.0001
Résidus	109	16614,585	152,427		
Total	110	38121,108			

Paramètres du modèle :

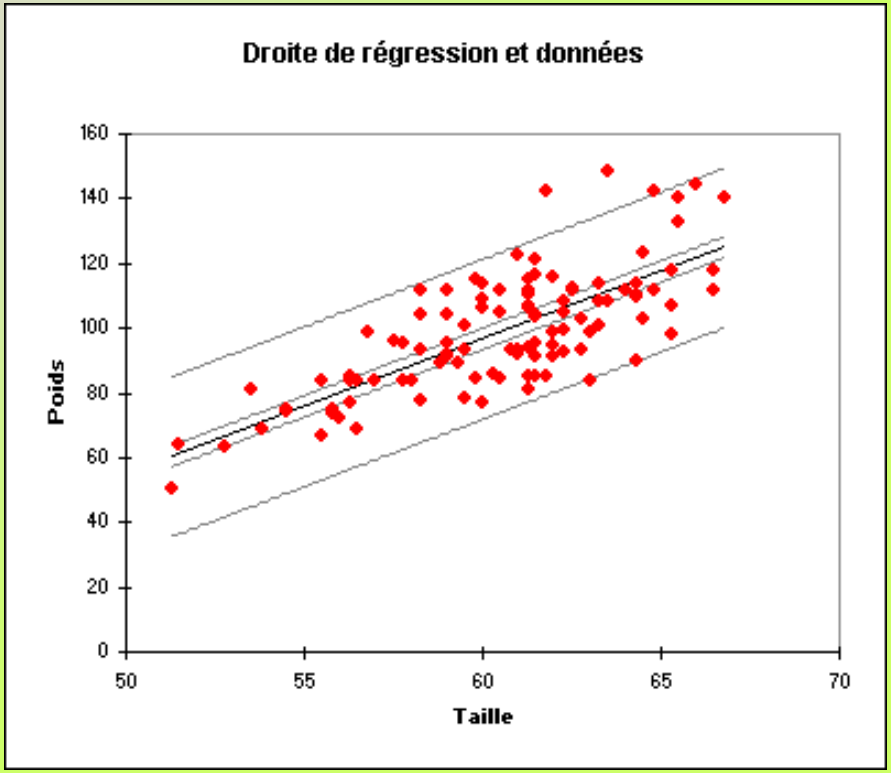
Paramètre	Valeur	Ecart-type	t de Student	Pr > t	ıe inférieure	Ə supérieure
Constante	-153,129	21,248	-7,207	< 0.0001	-195,242	-111,016
Taille	4,164	0,351	11,878	< 0.0001	3,469	4,858

*L'équation du modèle s'écrit : Poids = -153.128910179342 + 4.16361172748079*Taille*

Régression linéaire – **Le 1er graphique :**

Ce graphique permet de visualiser :

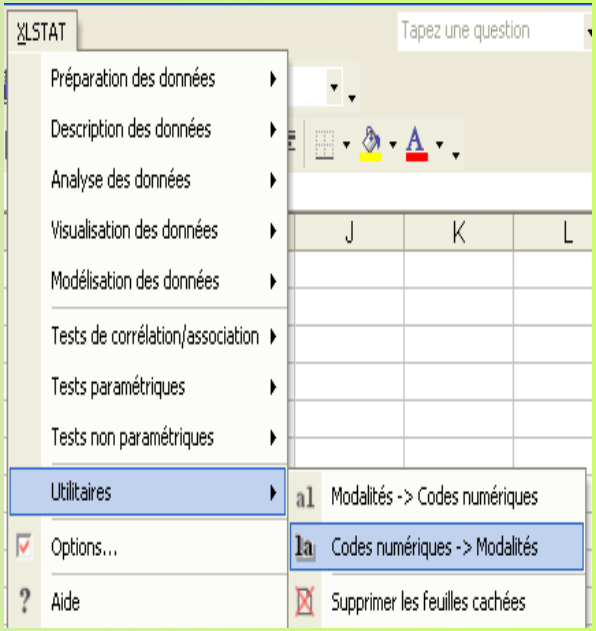
- les données,
- la droite de régression,
- les deux intervalles de confiance.



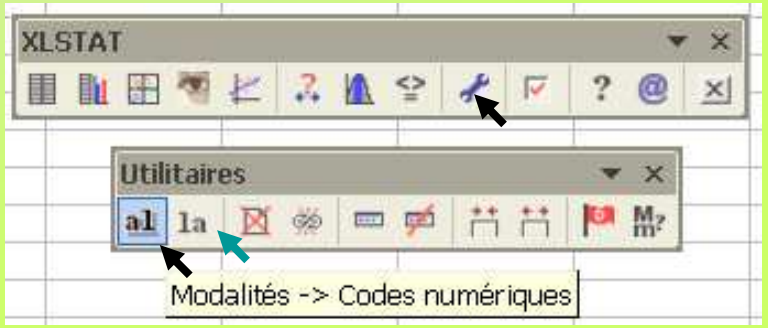
Codage

Codage

- Ouverture par le menu XLSTAT



- Ouverture par la barre d'outils XLSTAT



Modalités -> Codes numériques

Cela permet de remplacer les modalités d'une variable qualitative par des codes numériques.

- Apparaît une zone de commentaire sur la cellule contenant le libellé de la variable :
  - La 1re ligne contient le libellé de la variable tel qu'il figure dans la 1re cellule de la colonne sélectionnée
  - La 2e ligne contient par défaut le libellé de la variable : on peut le remplacer par un libellé détaillé
  - Les autres lignes correspondent aux différentes modalités de la variable, codées à partir de 1.
- Cette procédure peut être exécutée sur plusieurs colonnes à la fois, placées n'importe où dans la feuille.

Codes numériques -> Modalités

Cela permet de remplacer les codes numériques d'une variable qualitative par les libellés des modalités figurant dans la zone de commentaire de la cellule contenant le libellé de la variable.

Typiquement,

la 1re ligne du commentaire contient le libellé abrégé de la variable (ex. " Q9 "),

la 2e contient le libellé détaillé de la variable (ex. " Satisfaction globale "),

- les lignes suivantes correspondent aux différentes modalités de la variable, codées à partir de 1.

Codes numériques -> Modalités

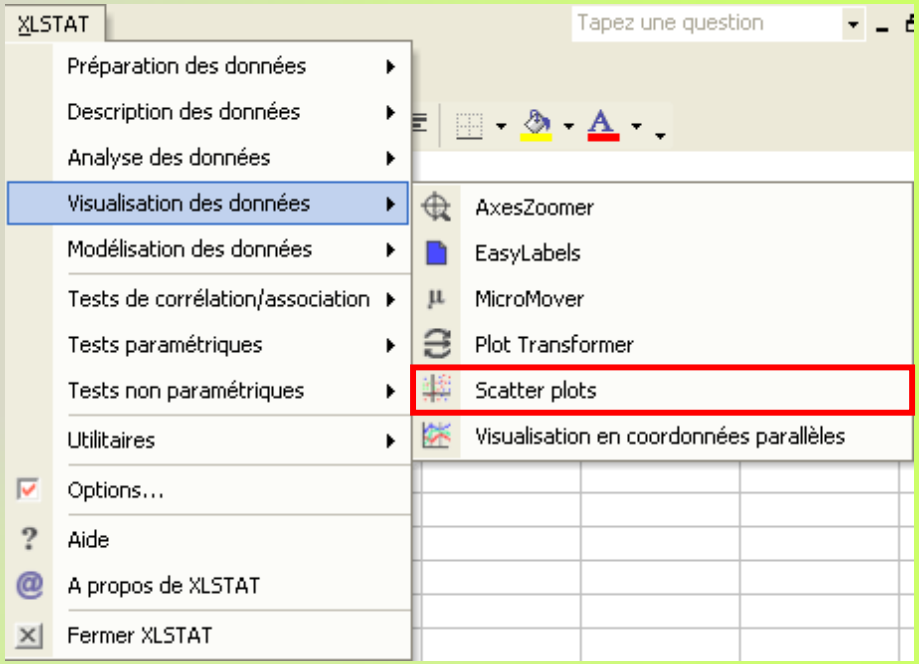
Remarques :

- Le 0 est considéré comme la valeur codant une valeur manquante.
- A l'issue de cette procédure, la zone de commentaire est supprimée.
- Si une seule ligne précède la liste des codes, alors le contenu de la 1re ligne est considéré comme le libellé détaillé de la variable.
- Cette procédure peut être exécutée sur plusieurs colonnes à la fois.

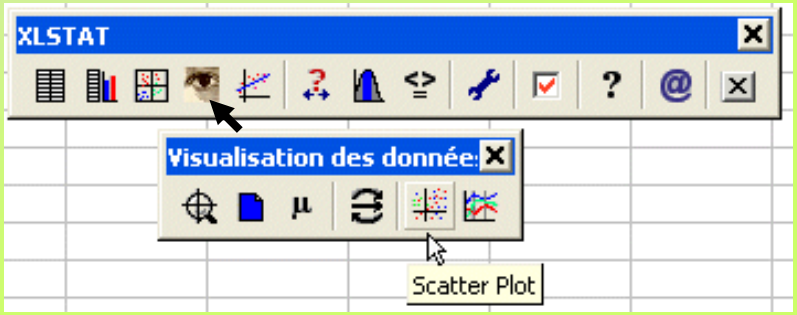
Quelques autres fonctionnalités

Scatter plots

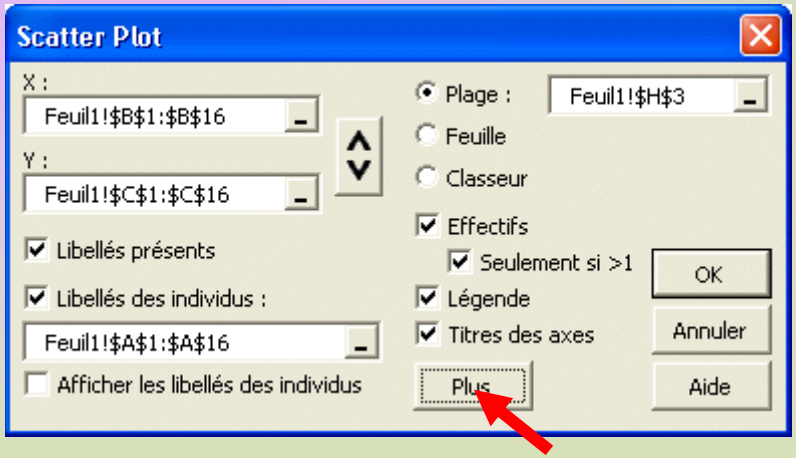
- Ouverture par le menu XLSTAT



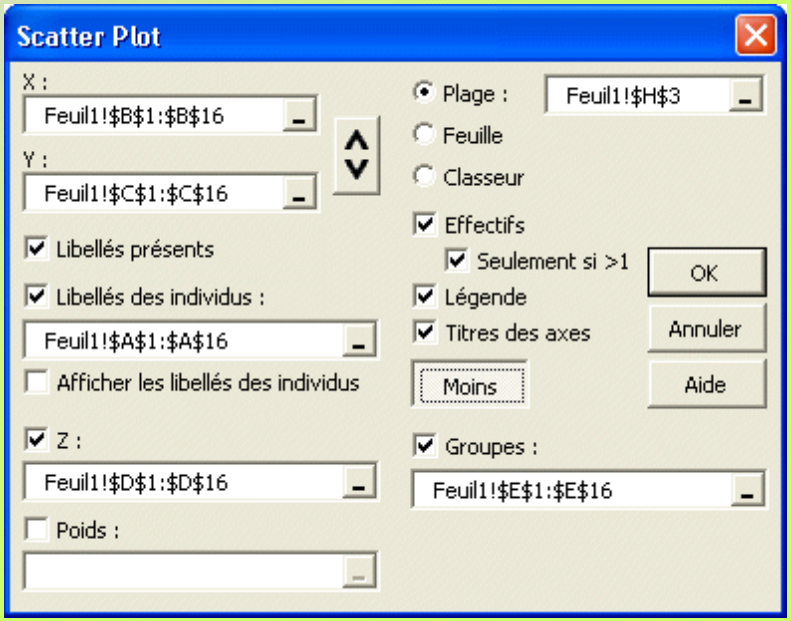
- Ouverture par la barre d'outils XLSTAT



Scatter plots – **La fenêtre :**



Après un clic sur le bouton Plus, la fenêtre complète s'affiche



Florence Nicolau

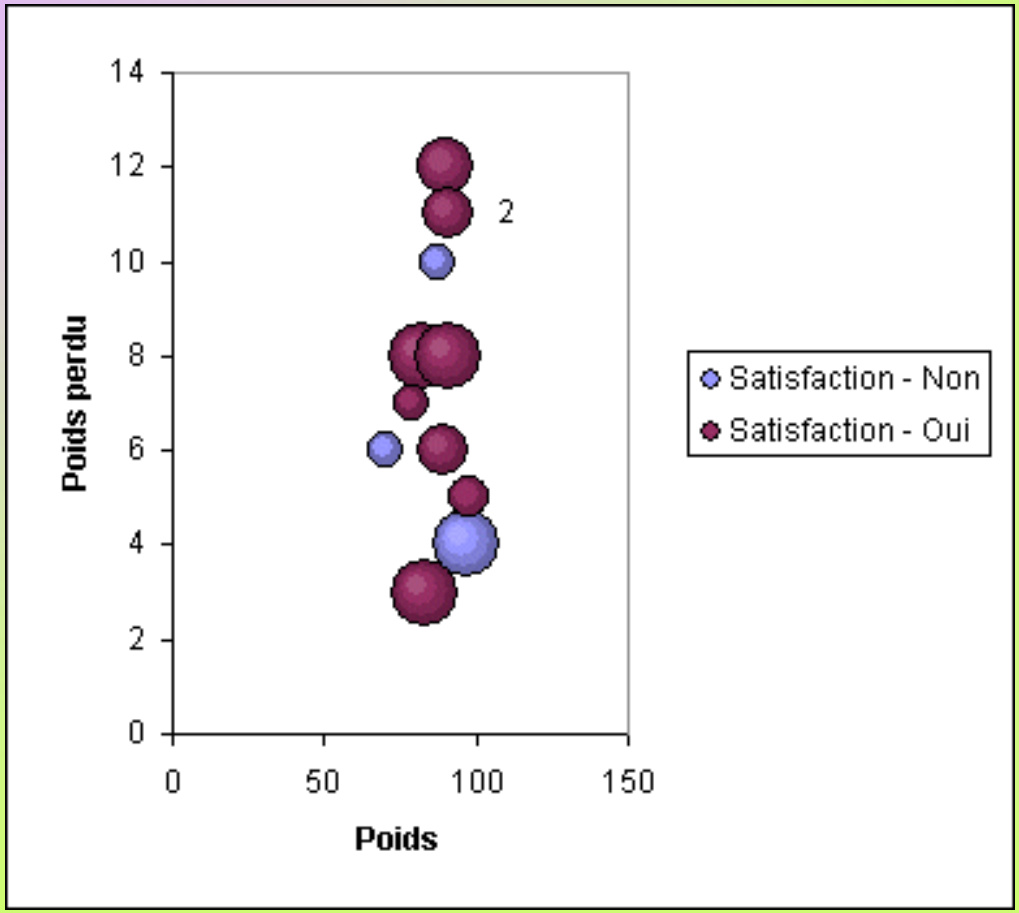
Scatter plots – **Un exemple :**

Exemple : pour un groupe de patients qui ont suivi un régime, on a relevé

- leur poids avant le régime (kg),
- le poids qu'ils ont perdu,
- leur satisfaction quant aux résultats du régime
- leur âge.

Libellés des individus	En X	En Y	Taille des cercles	Groupes
	Poids	Poids perdu	Age	Satisfaction
Patient1	86	3	39	Non
Patient2	91	11	42	Oui
Patient3	91	8	57	Oui

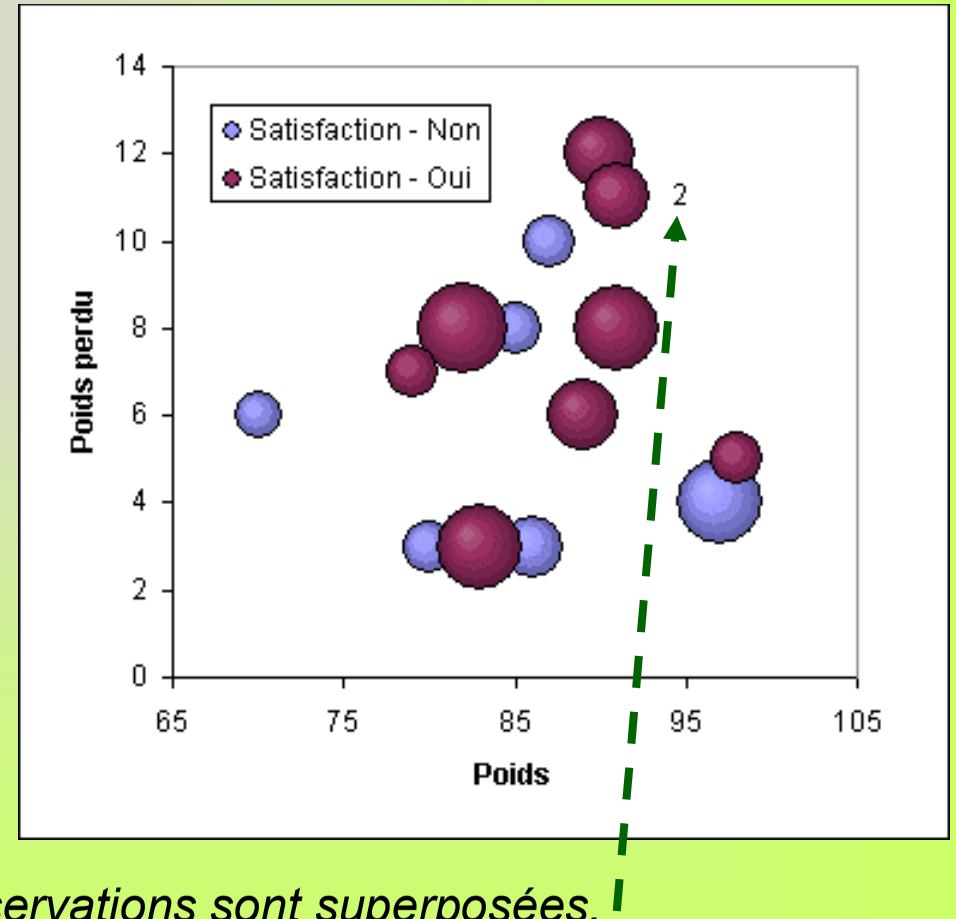
Scatter plots – **Le graphique :**



Scatter plots – **Le graphique :**

Pour améliorer la lisibilité du graphique, modifier l'échelle de l'axe des abscisses :

Visualisation des données
AxesZoomer



L'étiquette "2" signifie que deux observations sont superposées.

Analyse de données

Analyse en Composantes Principales (ACP),

Analyse Factorielle Discriminante (AFD),

- Analyse Factorielle des Correspondances (AFC),

Analyse des Correspondances Multiples (ACM ou AFCM),

Classification Ascendante Hiérarchique (CAH),

- Nuées dynamiques (k-means),

- …

Modélisation

Ajustement d'une loi de probabilité,

Régression,

- ANOVA / ANCOVA (tests de comparaisons multiples),

Régression logistique,

Régression non-linéaire,

- …

Tests d'hypothèses

- Tests sur les tableaux de contingence,
- Tests de corrélation,
- Tests paramétriques pour la comparaison de deux échantillons
- Tests non-paramétriques

Sources

- Brochure Xlstat
- Tutoriels de Xlstat
- Rubriques d'aide de Xlstat
- Site web : http://www.xlstat.com