

# Atelier de formation à la Gestion-Analyse des données et interprétation des résultats des ESS



## Analyse des données d'enquête (avec le TCD)



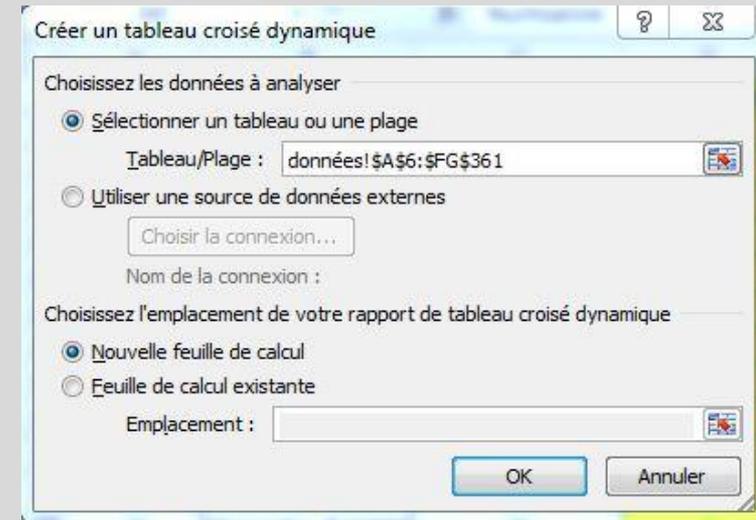
# Objectifs

1. **Comprendre les fonctionnalités** de l'outil d'analyse tableau croisé dynamique (TCD) dans Excel.
2. Avoir une meilleure **compréhension des statistiques descriptives** afin d'exécuter une analyse statistique
3. Être **capable de lancer des analyses statistiques** en utilisant le TCD



# Introduction

- Avant de lancer l'analyse de données, il est important de savoir le type d'information et d'analyse souhaité pour la rédaction du rapport.
- **Note:** Il existe un large éventail d'informations dans la base de données mais à importantes variées.
- La fonctionnalité TCD est un outils facile d'utilisation, pertinent pour les analyses dans le cadre des ESS



# Types d'analyses

- Les statistiques descriptives sont utilisées pour résumer les données quantitatives sous une forme gérable. Cela peut toutefois dénaturer le sens des données.

## Types d'analyse

- Analyse univariée: cela implique l'examen d'une variable à la fois. Il y a 3 caractéristiques majeurs relatives à une seule variable sur lesquelles nous allons nous pencher: a) la distribution, b) **la tendance centrale** et c) la dispersion
  - **Ex: Analyse du niveau d'accès au crédit par type de ménage**
- Analyse multivariées: L'analyse simultanée de 2 ou plusieurs variables (dépendantes) (n'a pas été largement utilisé dans les analyses ESS)
  - **Ex: Analyse du niveau d'accès au crédit par type de ménage selon la zone agro-écologique**

# Types de questions et analyses

1. Q?. Réponse simple avec 2 options: *oui* ou *non*, *homme* ou *femme* – codé comme 0 et 1
  - ✓ Analyse: nombre (Fréquence) et/ou pourcentage
2. Q?. Réponse simple avec plusieurs options prédéfinies (Résident, Réfugié, Déplacé)
  - ✓ Analyse: nombre (Fréquence) et/ou pourcentage
3. Q?. Donnée quantitative, ex: Superficie, Quantité, volume, distance, durée, etc.
  - ✓ Analyse: somme, moyenne, médiane, mode, maximum, minimum, amplitude, Déviation standard (écart type)
4. Q?. Donnée qualitative - chevauchement entre les définitions qualitatives et quantitatives
  - ✓ Classes /groupes - nombre (Fréquence) et/ou pourcentage
5. Questions ouvertes (qualitative)
  - ✓ Classes/regroupement - nombre (Fréquence) et/ou pourcentage

# Analyse Tableau croisé dynamique

**Les tableaux croisés dynamiques sont l'une des fonctions les plus robustes d'Excel.** Un tableau croisé dynamique vous permet d'extraire la substance d'un grand ensemble de données détaillées.

Fichier Accueil Insertion Mise en page Formules Données Révision

TblCroiséDynamique Tableau Image Images clipart Formes SmartArt Capture Illustrations

Ab Numsaie

6 Numsaie enquêteur Nom repondar Sexreponda Age

17 64

24 64

25 44

26 70

27 49

28 45

29 68

30 55

31 67

34 70

37 59

38 63

39 47

41 48

42 62

Créer un tableau croisé dynamique

Choisissez les données à analyser

Sélectionner un tableau existant

Tableau/Plage : données!\$A\$6:\$FG\$361

Utiliser une source de données externes

Choisir la connexion...

Nom de la connexion :

Choisissez l'emplacement de votre rapport de tableau croisé dynamique

Nouvelle feuille de calcul

Feuille de calcul existante

Emplacement :

OK Annuler

TCD

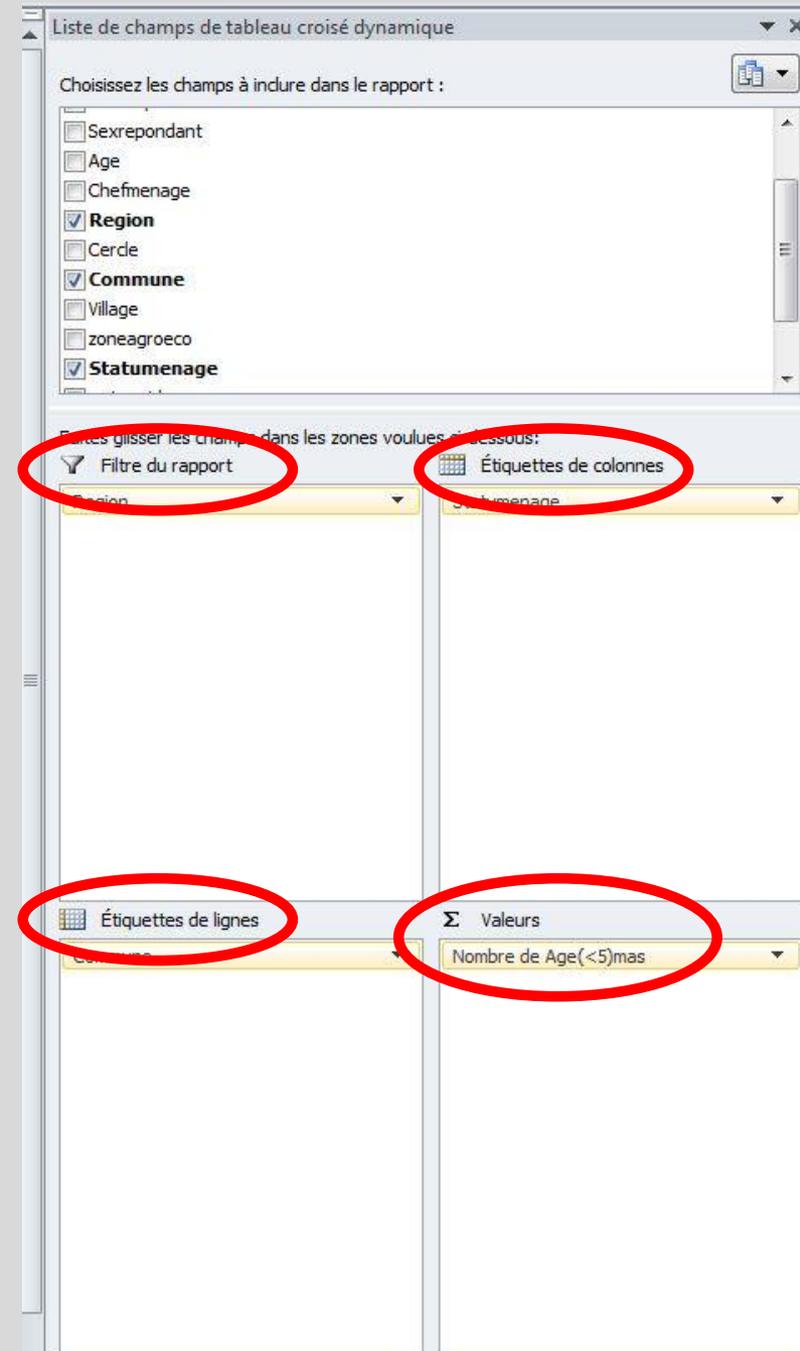
Plage de données

# Production d'un TCD

1. **SELECTIONNER** les variables à analyser (y compris les variables indépendantes: localisation, genre, zone agro-éco, status ménages... ), y compris les données relatives à ces variables (contenus des cellules).
2. Allez sur INSERTION et cliquez sur tableau croisé dynamique. Une boîte de dialogue "**Créer un tableau croisé dynamique**" apparaît.
3. Dans **la boîte de dialogue**, la plage de données apparaît dans la 1er fenêtre. En dessous, sélectionnez "**Nouvelle feuille de calcul**" puis validez par **OK**. Une nouvelle feuille de calcul apparaît à la gauche de votre base de données avec **une liste de champs** présentant toutes les variables sélectionnées.

# TCD – liste de champs

- Liste de champ de TCD:
  - L'étiquette de ligne et l'étiquette de colonnes sont utilisées pour créer la structure du tableau.
  - $\Sigma$  Valeurs est utilisée pour lancer les analyses.
  - Filtre du rapport est utilisé pour filtrer le champ d'intérêt lors de l'analyse



# TCD– Génération des tableaux

## 4. Génération d'un tableau

- d) Tirer une ou plusieurs variables (ex: Region, cercles, zone agro-éco) dans l'étiquette des lignes.
- e) Tirer la **variable à analyser** (ex: taille ménage, Age<5ans (mas)) dans la case de  $\Sigma$  des valeurs et lancer l'analyse.
- f) Pour les variables qualitatives, elles doivent d'abord être introduites dans l'étiquette des colonnes puis dans le champs de  $\Sigma$  valeurs pour le comptage (fréquence).

Liste de champs de tableau croisé dynamique

Choisissez les champs à inclure dans le rapport :

- Numsaisie
- enquêteur
- Nom repondant
- Sexrepondant
- Age
- Chetmenage
- Region
- Cercle
- Commune
- Village
- zoneagroeco
- Statumenage
- moisresid
- Age(<5)mas
- Age(<5)fem

Faites glisser les champs dans les zones voulues ci-dessous:

Filter du rapport: Cercle

Étiquettes de colonnes: (empty)

Étiquettes de lignes: Sexrepondant

Σ Valeurs: Moyenne de Age

Différer la mise à jour de la disposition

Mettre à jour

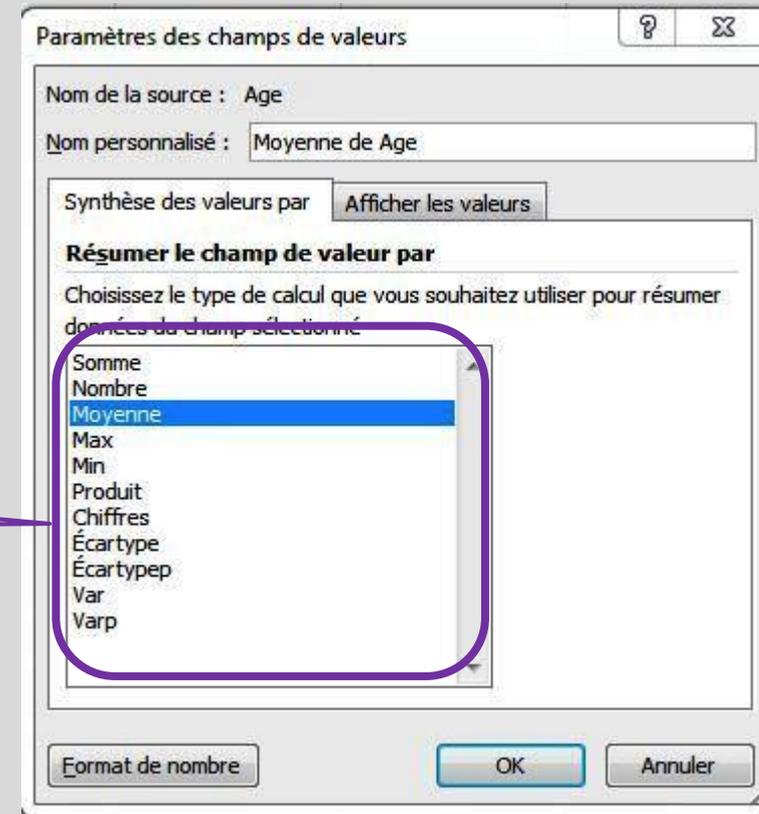
# TCD– Calculs statistiques

## 5. Lancer l'analyse statistique:

- Cliquez sur la variable dans le champs  $\Sigma$  **valeurs**, puis, sur la bande déroulante affichée, cliquez sur **Paramètre des champs de valeurs**.
- Une boîte d'analyse "**paramètre des champs de valeur**" s'affiche avec une plage de fonctions (somme, nombre, moyenne, Max, .....).

Champs de calcul

- Choisir le type de calcul voulu et cliquer sur **OK** (ex: ici c'est la **moyenne** qui est choisie)



# TCD– Calculs statistiques

## 5. Lancer l'analyse statistique:

- a) Générez la table – Puis celui-ci apparaît à la gauche de l'écran. Vérifiez que le tableau reflète l'analyse souhaitée.

Type de calcul

- b) Le type de calcul est présenté dans la case supérieure droite de la table (ex: calcul du nombre moyen de garçons <5ans par ménage dans les 9 différentes cercles)

Étiquettes de lignes ▼	Moyenne de Age(<5)mas
1	2,524590164
2	1,764705882
3	3,4
4	1,125
5	1,175
6	1,6
7	1,8
8	1,275
9	1,45
<b>Total général</b>	<b>1,743661972</b>



# Tableaux croisés dynamiques à plusieurs niveaux (Démonstration)

## Multiples champs de lignes

A3	fx Sum of Amount			
	A	B	C	D
1	Drop Report Filter Fields Here			
2				
3	Sum of Amount			
4	Category	Country	Total	
5	<input checked="" type="checkbox"/> Fruit	Australia	91221	
6		Canada	82338	
7		France	125931	
8		Germany	66430	
9		New Zealand	62392	
10		United Kingdom	87786	
11		United States	176971	
12	Fruit Total		693069	
13	<input checked="" type="checkbox"/> Vegetables	Australia	40492	
14		Canada	12407	
15		France	15125	
16		Germany	88738	
17		New Zealand	4390	
18		United Kingdom	85351	
19		United States	90162	
20	Vegetables Total		336665	
21	Grand Total		1029734	
22				
23				

## Multiples variables dans le champs $\Sigma$ valeurs

A3	fx			
	A	B	C	D
1	Drop Report Filter Fields Here			
2				
3		Data		
4	Country	Sum of Amount	Sum of Amount2	
5	Australia	131713	131713	
6	Canada	94745	94745	
7	France	141056	141056	
8	Germany	155168	155168	
9	New Zealand	66782	66782	
10	United Kingdom	173137	173137	
11	United States	267133	267133	
12	Grand Total	1029734	1029734	
13				
14				

# Tableaux croisés dynamiques à plusieurs niveaux (Démonstration)

## Multiples champs de filtres

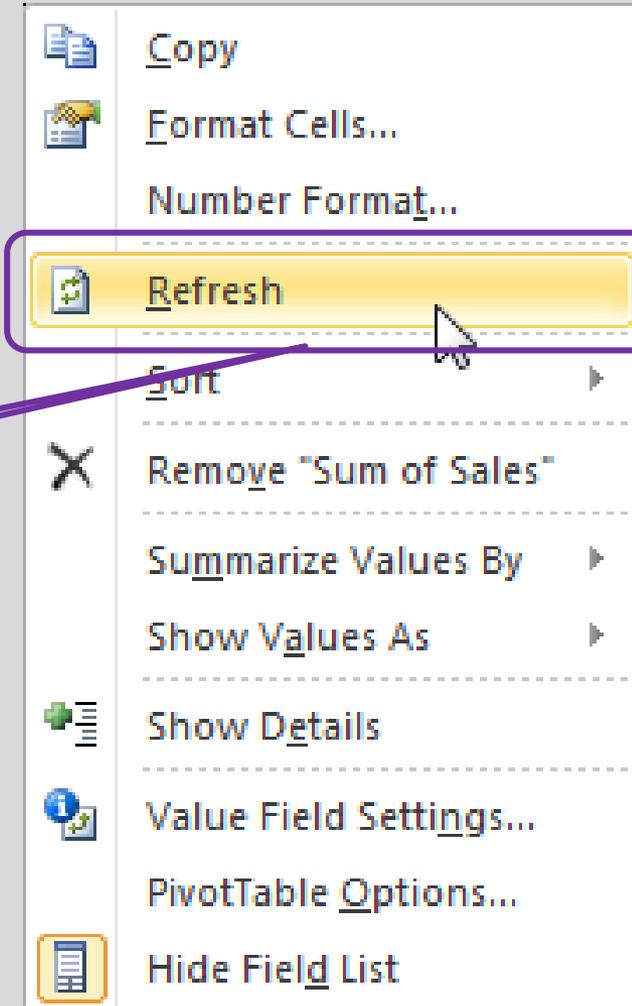
	A	B	C	D
1	Country	United Kingdom		
2	Product	Broccoli		
3				
4	Sum of Amount			
5	Order ID	Total		
6	2	8239		
7	19	3595		
8	86	2054		
9	92	2011		
10	112	7231		
11	121	6343		
12	126	3027		
13	163	5936		
14	Grand Total	38436		
15				
16				

# Mise à jour du TCD- actualiser

- **Note:** Tout changement opéré dans la base de données n'est pas automatiquement pris en compte par le TCD. **Actualiser** le TCD ou **changer la source de données** pour la mise à jour du TCD avec la prise en compte des modifications.

**actualiser**

- **Actualiser:** Si vous changez tout nombre ou texte dans la base de données, il faut actualiser le TCD.
1. Cliquer sur une cellule dans le TCD.
  2. Puis clic droit et actualiser.



# Mise à jour du TCD– Changement de la source de données

- **Changement de la source de données:** Si vous changez la plage de données en ajoutant ou retirant une ligne/colonne, vous devez faire la mise à jour de la source de données pour le TCD.

1. Cliquer sur n'importe quelle cellule dans le TCD (**E 1**)
2. Dans la barre d'outils, cliquez sur option (**E 2**) puis sur changer sources de données (**E3**).

The screenshot displays the 'Outils de tableau croisé dynamique' (PivotTable Tools) ribbon in Microsoft Excel. The 'Options' tab (E 2) is active, showing options like 'Grouper la sélection', 'Dissocier', and 'Grouper les champs'. The 'Création' tab (E 3) is also visible, showing options like 'Trier', 'Insérer un segment', 'Actualiser', and 'Changer la source de données'. Below the ribbon, a PivotTable is shown with the following data:

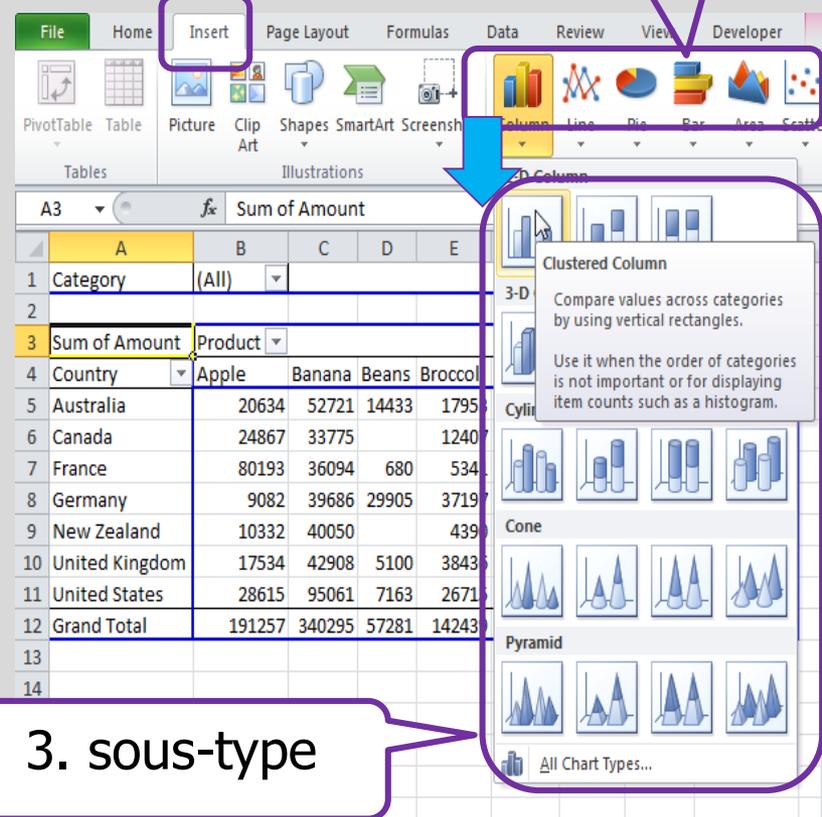
	A	B	C	D	E	F	G
<b>Nombre de Statumenage</b>							
<b>Étiquettes de lignes</b>			0	1	<b>Total général</b>		
1			40	155	195		
2			40	120	160		
<b>Total général</b>			80	275	355		

# Graphiques croisés dynamiques

- Un **Graphique CD** est une visualisation du TCD dans Excel. Les graphiques CD et les TCD sont liés.
- **Création d'un graphique CD:** Allez dans **insertion** (dans la barre d'outil).
  1. Cliquez sur une cellule dans le TCD.
  2. Dans la barre Insertion, sélectionner le type de graphique (ex: **colonnes**) et
  3. Sélectionner un **sous-type de colonnes** et double clic

1. Insertion

2. Type de graphique



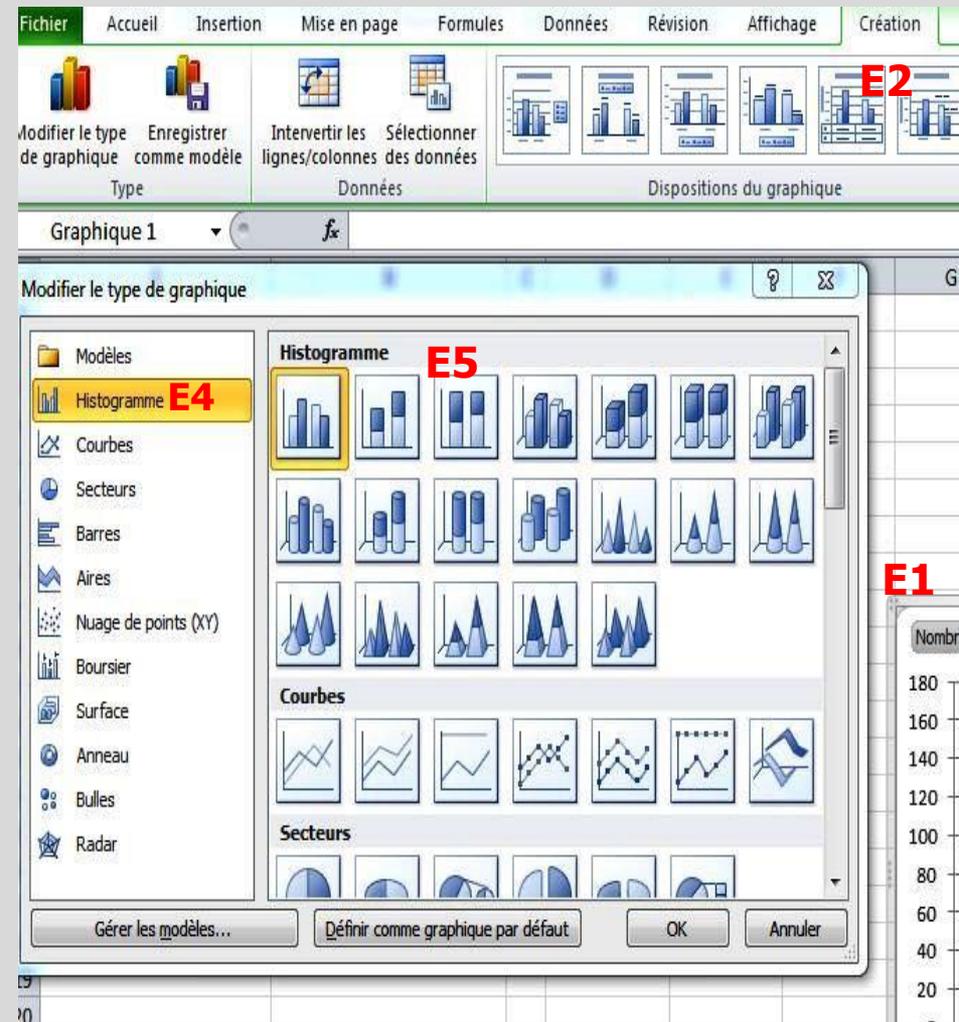
3. sous-type

	A	B	C	D	E
1	Category	(All)			
2					
3	Sum of Amount	Product			
4	Country	Apple	Banana	Beans	Broccoli
5	Australia	20634	52721	14433	1795
6	Canada	24867	33775		1240
7	France	80193	36094	680	534
8	Germany	9082	39686	29905	3719
9	New Zealand	10332	40050		439
10	United Kingdom	17534	42908	5100	3843
11	United States	28615	95061	7163	2671
12	Grand Total	191257	340295	57281	14243
13					
14					

# Graphiques croisés dynamiques

- **Changer le type de graphique CD:** Vous pouvez changer le type de graphiques à tout moment. **E3**

1. **Selectionner le graphique.** La barre d'outil graphique s'active (**E1**).
2. Cliquez sur **création** (**E2**) puis sur modifier type de graphique (**E3**).
3. Choisir un type de graphique (**E4**) et un sous-type (**E5**)
4. Click **OK**

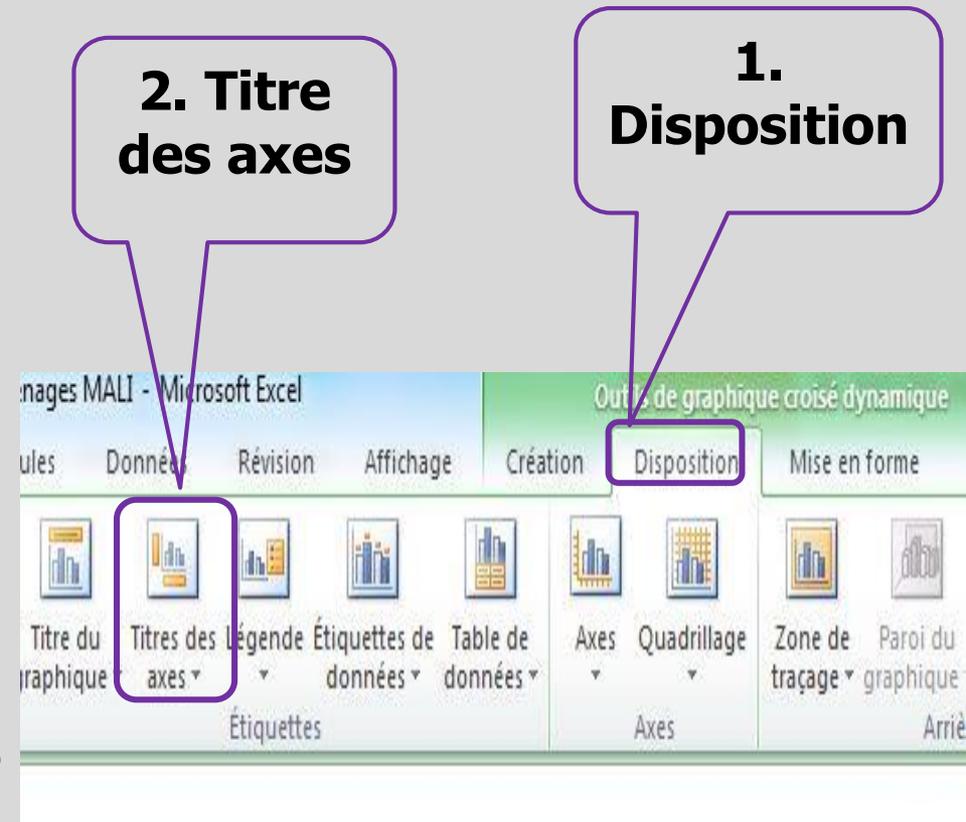


# Graphiques croisés dynamiques

- **Axes** : La majorité des graphiques ont 2 **axes**: Les abscisses et les ordonnées.

- **Titre des axes** : Pour ajouter un titre en abscisse ou en ordonnée, exécutez les étapes suivantes.

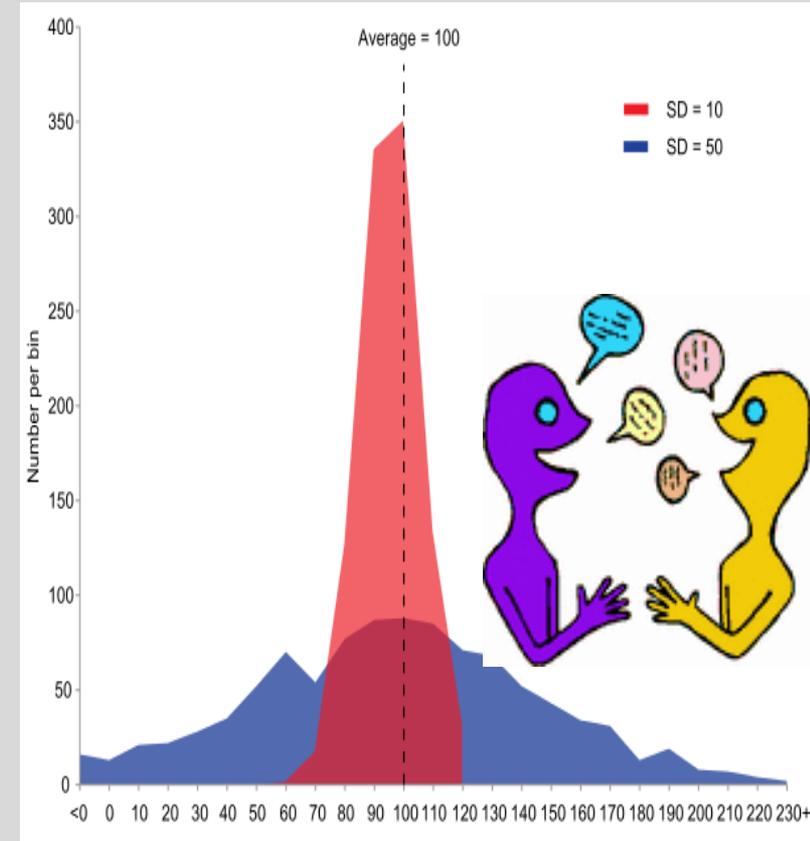
1. Sélectionner le graph. La barre d'outil graph CD s'active.
2. Dans **Disposition**, cliquez sur **Titre des axes**.
3. Vous avez des options pour le titre de l'axe vertical/horizontal. Choisir et valider



# Déviatiion standard/Ecart type ( $\sigma$ or Std. Dev.)

- **Ecart-type** ( $\sigma$  ou Std. Dev.) – *racine carrée de la variance*  
C'est une quantité qui exprime de combien les membres d'un groupe diffèrent de la valeur moyenne du groupe (pour une variable donnée)

- **Un faible écart type** indique que les points de données ont tendance à être très proche de la moyenne-**données assez homogènes**
- Alors que **l'écart-type élevé** indique que les données sont réparties sur une large gamme de valeurs. **Données assez hétérogènes**



Deux ensembles de données avec la **même moyenne** mais des **écarts-types différents?**

# Déviatiion standard/Ecart type ( $\sigma$ or Std. Dev.)

- Il existe deux types d'écart types : l'écart type de la population et celui de l'échantillon
  - Écart type de l'échantillon est utilisé quand nous voulons tirer des conclusions sur la population à partir de l'échantillon.
  - Si une valeur de l'échantillon se situe en dehors de ( $M \pm$  l'écart type) de l'échantillon alors nous pouvons dire avec un niveau de confiance élevé que la valeur est significativement différente de **la moyenne de la population.**
- **L'écart type est une analyse robuste qui nous aide à savoir si les différences entre certains groupes sont représentatives des différences au sein de la population générale à partir de laquelle l'échantillon est tiré.**

# Déviatión standard et barre d'erreur

- Calcul de l'écart-type : Utilisation de la formule excel:
  - Dans G2, taper "=stdev (B2:E2)" , et...entrez
  - Etendre la formule à G3 et G4

The screenshot shows the Microsoft Excel interface. The formula bar at the top displays the formula `=ECARTYPE(B2:E2)` for cell G2. The spreadsheet contains the following data:

	A	B	C	D	E	F	G
1	<b>Statut</b>	<b>Mali</b>	<b>BF</b>	<b>Niger</b>	<b>Tchad</b>	<b>Moy</b>	<b>ecart-type</b>
2	<b>Pop hôte</b>	20	17	33	43	28,25	12,04
3	<b>Déplacés</b>	27	26	28	32	28,25	2,63
4	<b>Retounés</b>	2	0	8	15	6,25	6,75
5	<b>Réfugiés</b>	1	14	3	24	10,5	10,66
6							
7							

# Déviatiun standard et barre d'erreur

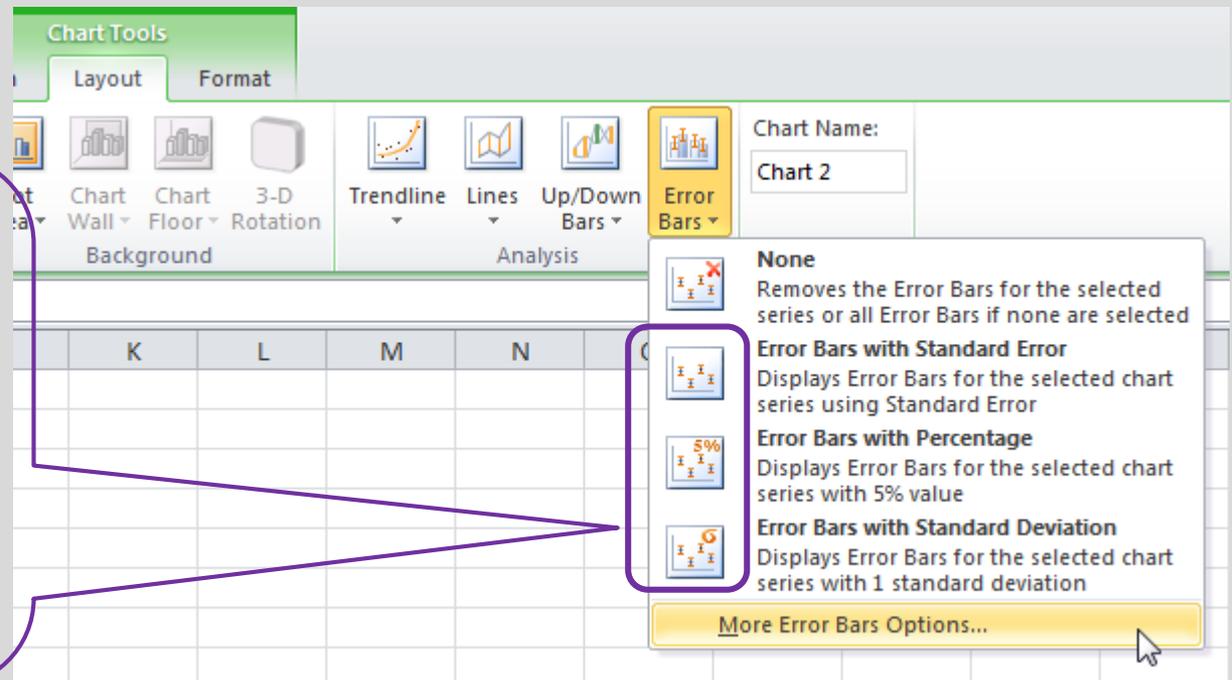
- Les barres d'erreur sont des représentations graphiques de la variabilité des données et sont utilisées sur les graphes pour indiquer l'erreur, ou l'incertitude dans les mesures présentées
- Les barres d'erreur peuvent être utilisées pour comparer visuellement deux quantités lorsque d'autres conditions sont identiques. Ceci peut aider à déterminer si des différences sont statistiquement significatives.

# Déviatation standard et barre d'erreur

## • Comment ajouter une barre d'erreur ?

1. Sélectionner le graphique. La barre d'outil graph croisé dynamique s'active.
2. Dans **disposition** , cliquez sur barre d'erreur, puis sur **autre option de barre d'erreur...**

Remarquez les raccourcis pour afficher rapidement les barres d'erreur en utilisant l'erreur standard, d'une valeur de 5% ou de 1 écart-type.



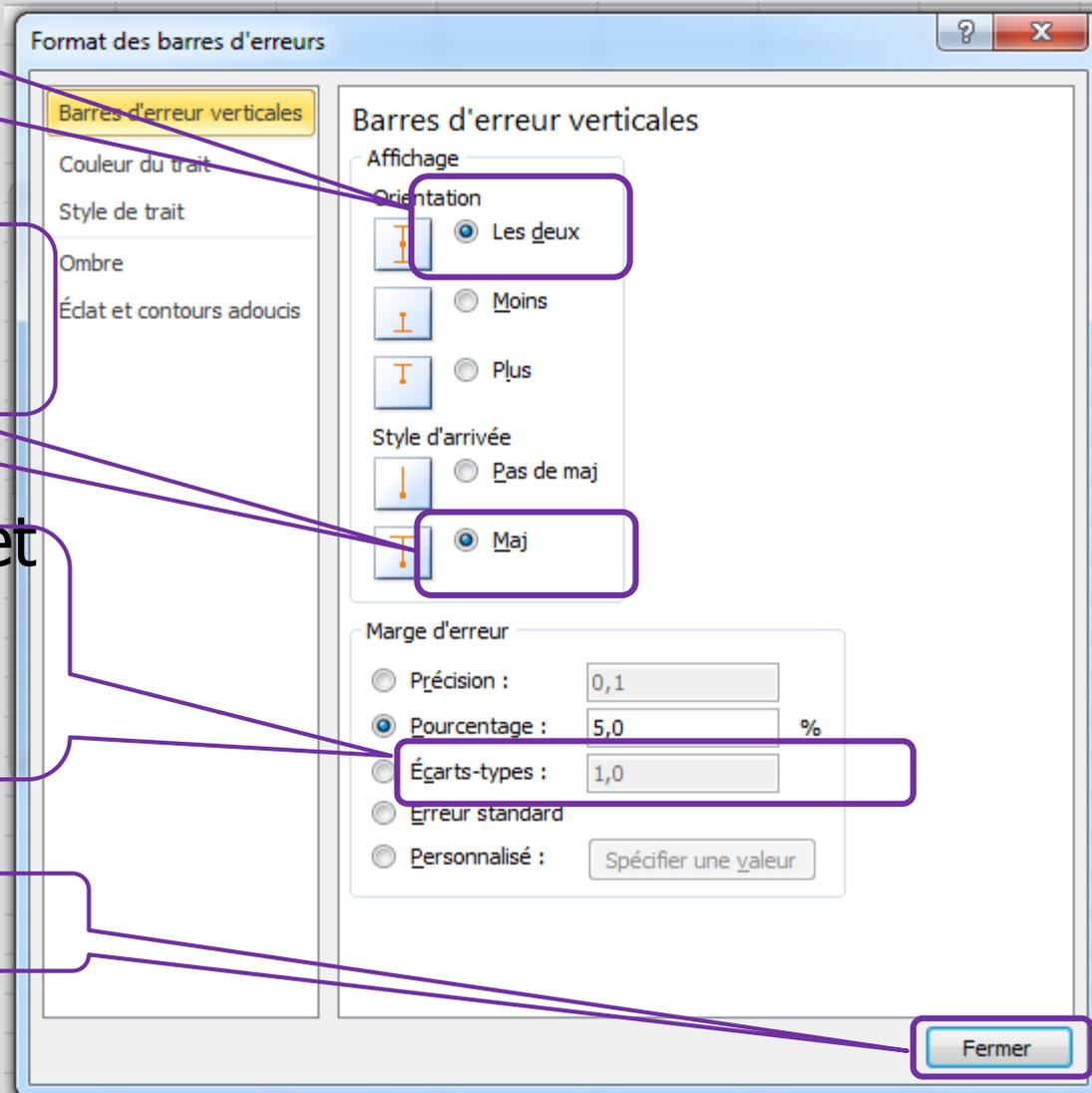
# Déviation standard et barre d'erreur

3. Choisir Orientation.  
Clique sur les deux.

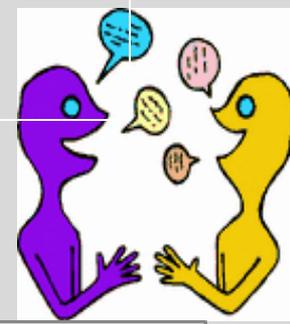
4. Choisir le style d'arrivée.

5. Choisir l'écart-type et entrer la valeur 10.

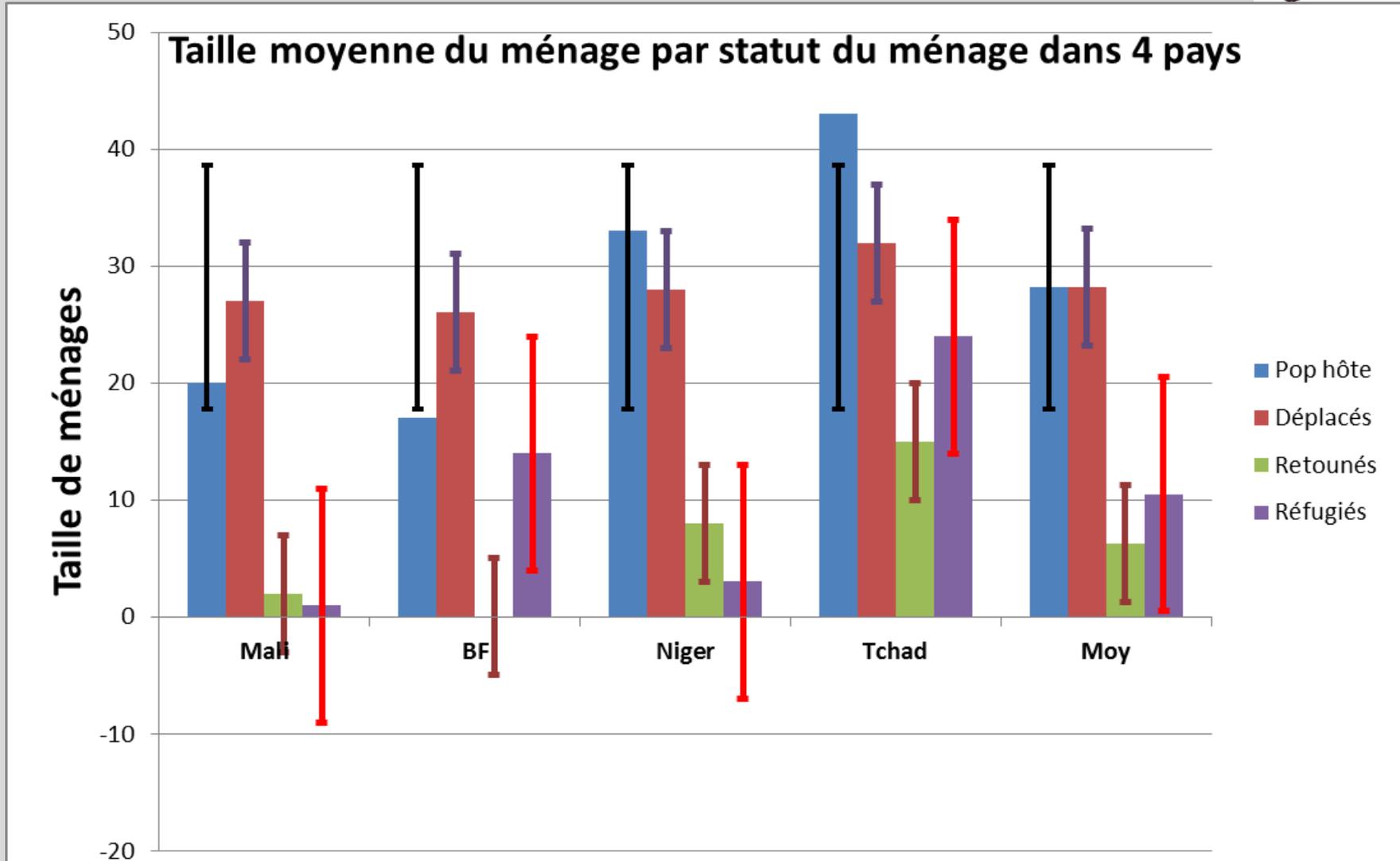
6. Cliquer sur fermer



# Déviation standard et barre d'erreur



- Résultat à discuter.



# Exercice 4.1: Création de TCD

- Vous avez reçu la base de données de l'ESS du Mali (Gao et Tombouctou). Examiner les données et lancer l'analyse TCD.
  1. Quelle est la quantité moyenne de semences pour les différentes cultures plantées en 2014 par les ménages (**ref: cultprin A**)?
  2. Quelle a été la somme totale de semences plantées par les paysans échantillonnés?
  3. Quelle est la pratique culturelle la plus utilisée? Quel est le pourcentage de ménages concernés par cette pratique
  4. Ressortir aussi cette analyse par type de vulnérabilité et par zone agro-écologique?
  5. Le type de pratique cultural est-il fonction du niveau de vulnérabilité?

# Exercice 4.2: Graphiques et barres d'erreurs

En utilisant les TCD créés en 4.1 (Q 1 et Q 3)

1. Ressortir un graphique (utilisant les colonnes) relatif à la quantité moyenne de semence par culture (cultprin A) en 2014 ? Nommer les axes
2. Fait un graphique de fromage avec les différents variétés de arachides qui sont utilisées dans le Nord de Mali. Inclue la fréquence et les pourcentages dans le graph.
3. Fait un graphique a barres pour les cinq cultures principales en montrant les proportion des variétés améliorées et locales.

## **Sources de semences**

4. Tableaux et graphiques le plus appropriés qui montrent les sources des semences pour les différentes cultures.

## **Disponibilité des semences**

5. En utilisant les graphique le plus appropriés, présenter les perception des paysan les paramètres de la disponibilité des semences, par statu des ménages.

## **Acces aux semences**

7. En utilisant les graphique le plus appropriés, présenter comment les paysans ont obtenue leurs semences en 2014 et 2015.

**SAVOIR PLUS**

# Types d'analyses

- **La Distribution.** La distribution est un résumé de la **fréquence** des valeurs individuelles ou les plages de valeurs pour une variable.

- **Fréquence** est le nombre d'apparition d'un événement dans une population donnée.

- Il est possible de grouper les valeurs ci-contre de la manière suivante: 1-2 (7), 3-4 (6) and 5-6 (1) .

Scores: 1, 1,2, 2,2, 2,2, 3,3,3,3,4,4,5	
Score	Fréquences
1	2
2	5
3	4
4	2
5	1

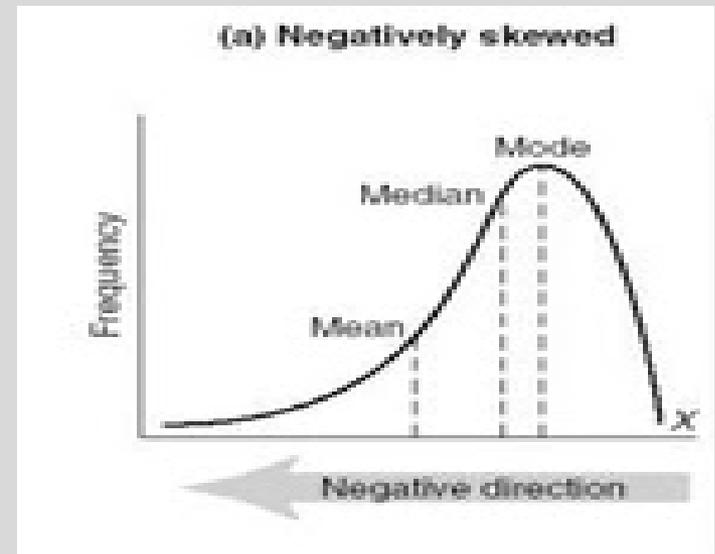
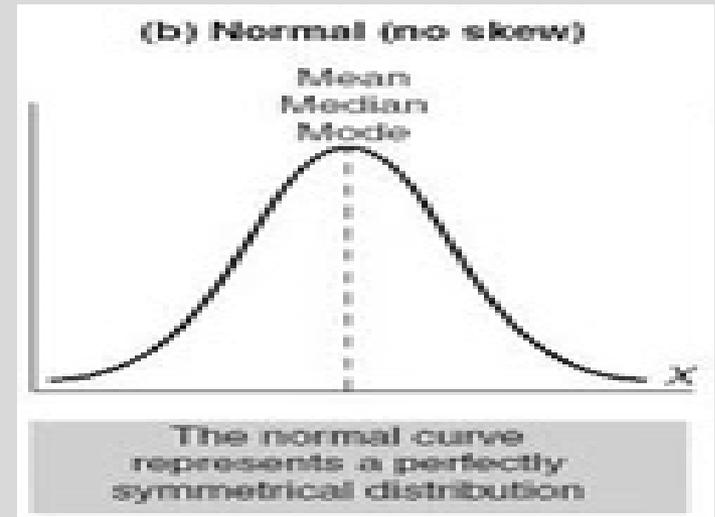
# Types d'analyses

• Un paramètre statistique est dit de **tendance centrale** s'il s'agit d'un nombre clé autour duquel les observations sont réparties : **mode**, **médiane**, **moyenne** sont des paramètres de tendance centrale

• La **Tendance centrale** d'une distribution est une estimation du "centre" de la distribution de valeurs.

• Les trois principaux types d'estimations de la tendance centrale sont

- La moyenne,
- La médiane (valeur médiane).
- Le mode (la valeur dominante)



# Types d'analyses

a) La moyenne,

**La moyenne arithmétique est la somme des valeurs de la variable divisée par le nombre d'individus**

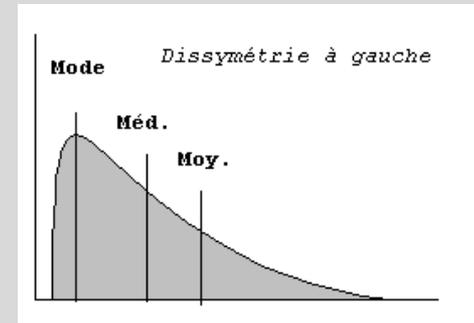
a) La médiane (valeur médiane)

**La médiane est la valeur centrale qui partage l'échantillon en 2 groupes de même effectif : 50 % au-dessus et 50 % en dessous. Si l'effectif est un nombre pair, on prendra comme valeur médiane la moyenne arithmétique des deux valeurs centrales**

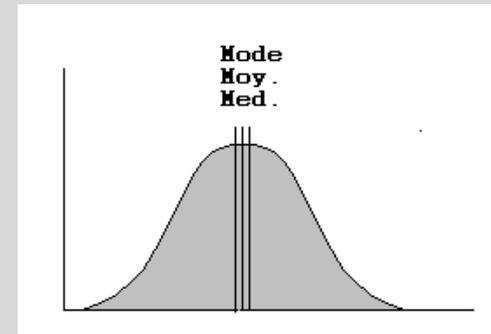
a) Le mode (la valeur dominante)

**Le mode correspond à la réalisation la plus fréquente.**

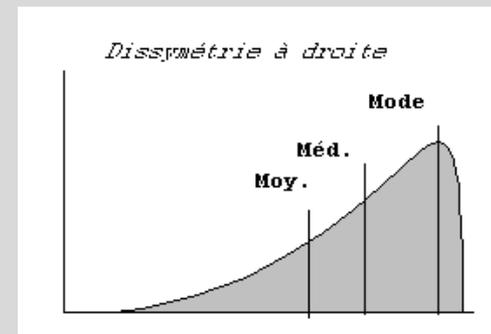
Dissymétriques (asymétriques) à gauche



symétriques (graphique du milieu)



symétriques (graphique du milieu)



# Types d'analyses

- **Dispersion**: La répartition des valeurs autour de la tendance centrale. Deux mesures communes de la dispersion sont utilisées:

- L'Amplitude est la différence entre la plus grande valeur et la plus petite valeur.

- **Note**: Une valeur aberrante peut affecter l'amplitude.

- La **Déviat**ion standard (écart type): permet d'évaluer la dispersion des mesures autour de la valeur moyenne.

