

APPLI-COURS CORRIGE :

ANALYSE STATISTIQUE DES VARIABLES QUALITATIVES



- 2 EXEMPLES -

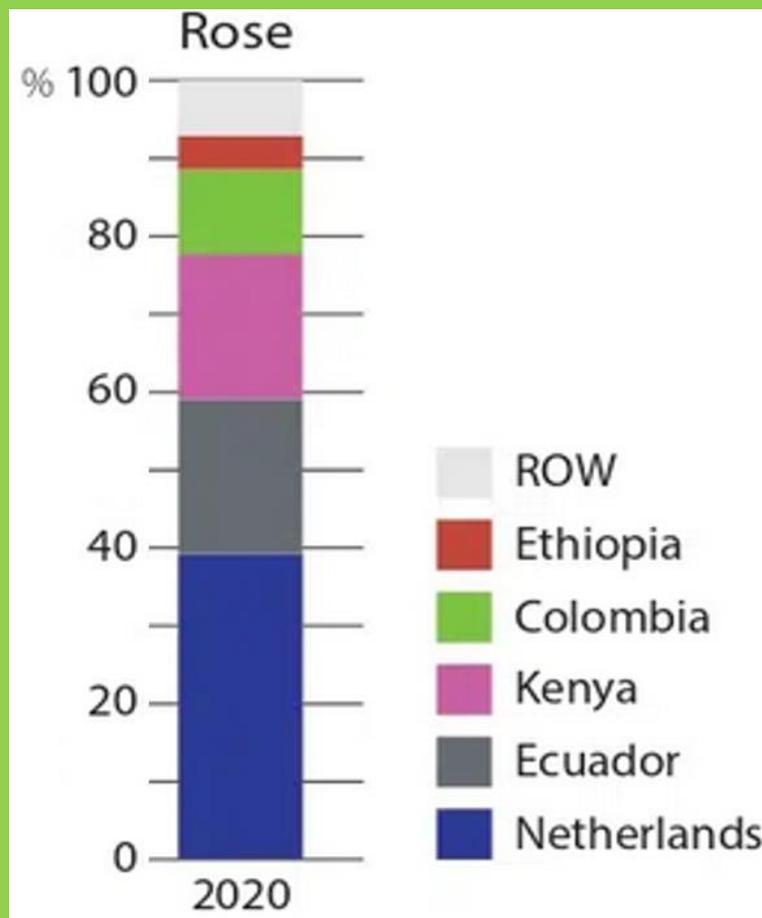
EXEMPLE 1

Après lecture du diagramme ci-dessous, veuillez répondre aux questions sous le diagramme (page suivante).

Global cut flower exports by flower type, 2020

Exportations mondiales de fleurs coupées par type, en 2020 : **cas des roses**

Source : <https://www.jardinerie-animalerie-fleuriste.fr/actualites/etude-rabobank-royal-floraholland-la-situation-de-la-floriculture-mondiale-en-2021/764795>



NB : RWO = *Rest of the world* = reste du monde

QUESTIONS :

- 1) De quelle nature est la variable étudiée ? *dites pourquoi.*

La variable est une **nomenclature alphabétique** de 6 pays exportateurs de roses. Elle est donc **qualitative**, car ses modalités (les pays) ne sont pas numériques.

- 2) Comment dénomme t'on ce diagramme ?

Le diagramme est un **diagramme empilé 100%**, construit au moyen des fréquences cumulées $F_i\%$ en ordonnées.

- 3)) Veuillez déduire du diagramme les **fréquences simples ($f_i\%$)** pour les reporter dans un **tableau de distribution** que vous aurez construit. Vos fréquences peuvent être approximatives. Les décimales ne sont pas exigées sauf si vous le souhaitez.

Une fréquence simple $f_i\%$ est la différence entre deux fréquences cumulées $F_i\%$, successives

$$\text{Ce que l'on écrit : } F_i\% = F_{i-1}\% + f_i\% \text{ donc } f_i\% = F_i\% - F_{i-1}\%$$

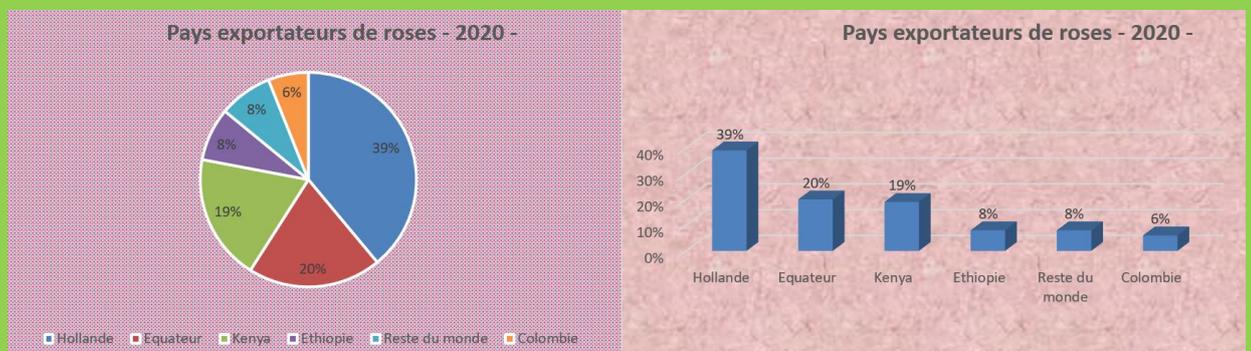
Une fréquence simple mesure **la part** des exportations de chaque pays dans l'ensemble.

La lecture du diagramme permet d'estimer les $f_i\%$ en les déduisant des $F_i\%$.

Pays x_i	Freq Cum $F_i\%$	freq simples $f_i\%$
Hollande	39%	39%
Equateur	59%	20%
Kenya	78%	19%
Colombie	84%	6%
Ethiopie	92%	8%
Reste du monde	100%	8%
Ensemble		100%

- 4) A l'aide du tableau de distribution, représenter un **diagramme différentiel** de votre choix.

Il est possible de représenter les $f_i\%$ dans divers types de *diagrammes différentiels*. On présente ci-dessous 2 modèles : **un diagramme en secteur**, et **un diagramme en tuyaux d'orgue**.



5) Commentez en une phrase, de deux manières :

a. En vous adressant à un novice en statistique

Par exemple : *Il y a dans le monde trois principaux pays exportateurs de roses : La Hollande, l'Equateur, et le Kenya.*

b. En vous adressant à un étudiant en statistique

Par exemple : *L'exportation de roses dans le monde est une activité concentrée dans 3 pays, qui réalisent plus des ¾ des exportations (78%) : La Hollande, l'Equateur, le Kenya. La Hollande est le principal pays exportateur avec près de 40% des exportations.*

Le reste des exportations est disséminé dans le monde avec néanmoins une part significative pour l'Ethiopie (8%) et la Colombie (8%).

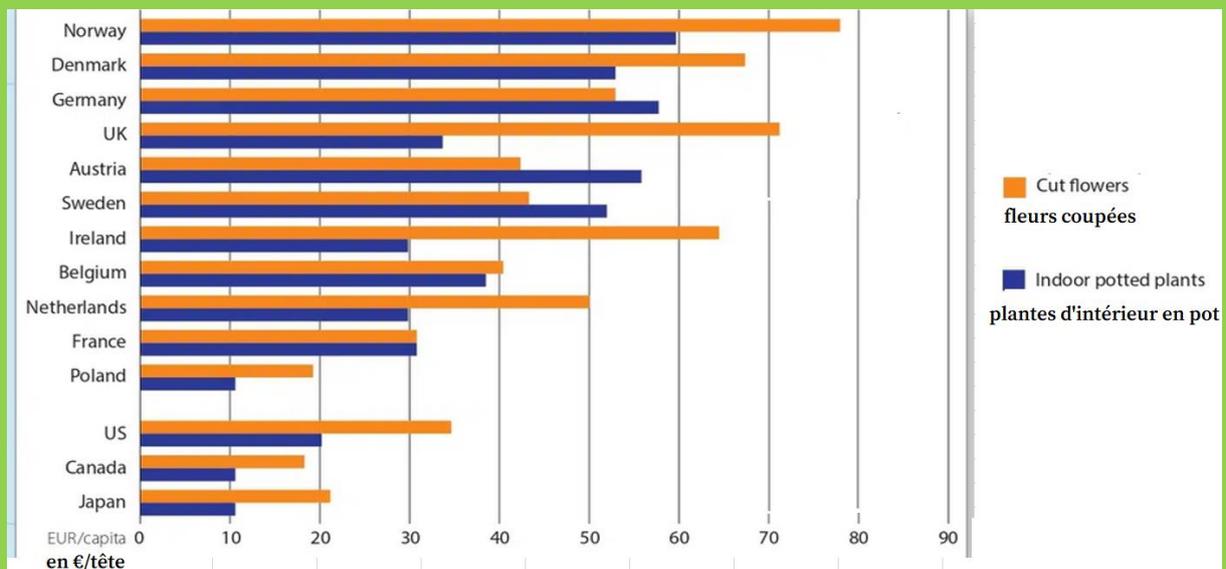


EXEMPLE 2

On a extrait de la même source de données (voir plus haut) la distribution suivante :

Floriculture expenditure per capita in selected countries, 2020

Dépenses en floriculture par tête dans une sélection de pays, en 2020



QUESTIONS :

1) La variable est-elle de même nature que celle de l'exemple 1 ? *dites pourquoi.*

La variable est ici le PAYS consommateur de fleurs et de plantes d'intérieur. Les 14 pays forment une **nomenclature alphabétique**. Il s'agit donc d'une variable **qualitative**, tout comme celle de l'exemple 1.

2) Quelle sont les différences entre les deux diagrammes (exemple 1 et exemple 2).

Il y a trois différences entre les deux diagrammes :

- Le diagramme 1 est un diagramme **intégral**. Le diagramme 2 est un diagramme **différentiel** appelé **diagramme en barres**.
- Le diagramme 1 est celui d'une **distribution unique** (les pays sont distingués par leurs seules exportations de roses). Le diagramme 2 comporte **deux distributions**, différenciées par leur couleur. En orange, les 14 pays sont distribués selon leur dépense en fleurs coupées. En bleu, selon leur dépense en plantes d'intérieur.
- Les données du diagramme 1 sont en **fréquences** (Cumulées $F_i\%$), tandis que celles du diagramme 2 sont en **effectifs n_i** (l'abscisse est en valeur absolue ou Euro/tête).

3) Prouvez la différence en construisant le **tableau de distribution** de cet exemple

L'exemple 2 est susceptible de deux présentations du tableau de distribution :

- soit un **tableau par distribution** : dépenses en fleurs coupées, dépenses en plantes d'intérieur
- soit un **tableau unique** en distinguant pour chaque pays : sa dépense en fleurs coupées (**effectifs n_{i1}**), et sa dépense en plantes d'intérieur (**effectifs n_{i2}**). Ce choix est préférable, car il permet le calcul de la dépense totale = $n_{i1} + n_{i2} = n_{i3}$. Ce qui enrichit le commentaire.

xi pays	n1 effectifs fleurs	n2 effectifs plantes	f1% freq. Fleurs	f2% freq. Plantes	n3 = n1+n2 ensemble	f3% ensemble
Norvège	79	59	12,5%	11,9%	138	12,2%
Danemark	68	53	10,7%	10,7%	121	10,7%
Allemagne	52	58	8,2%	11,7%	110	9,8%
Royaume Uni	71	34	11,2%	6,9%	105	9,3%
Autriche	42	56	6,6%	11,3%	98	8,7%
Suède	43	52	6,8%	10,5%	95	8,4%
Irlande	64	30	10,1%	6,1%	94	8,3%
Belgique	40	38	6,3%	7,7%	78	6,9%
Hollande	50	30	7,9%	6,1%	80	7,1%
France	31	31	4,9%	6,3%	62	5,5%
Pologne	19	11	3,0%	2,2%	30	2,7%
Etats Unis	35	20	5,5%	4,0%	55	4,9%
Canada	18	11	2,8%	2,2%	29	2,6%
Japon	21	11	3,3%	2,2%	32	2,8%
Total	633	494	100,0%	100,0%	1127	100,0%

4) Quel(s) moyen(s) peut on employer pour classer les pays en « **groupes de pays** » ? Réaliser ce(s) classement(s) : d'une part dans le tableau, d'autre part à l'aide d'un (ou plusieurs) diagramme (s) différentiel(s).

Dans la mesure où l'on dispose des montants absolus des dépenses (ou effectifs n_{i1} , n_{i2} , n_{i3}), il est possible de regrouper les pays autour d'un indicateur central, **la moyenne des dépenses (en fleurs coupées, en plantes d'intérieur, et pour l'ensemble)**.

Ressortiront alors les pays dont la dépense par tête est **supérieure** à la moyenne, ou **inférieure**. En triant ces groupes dans l'ordre décroissant de la dépense, on obtiendra de plus, dans chaque groupe, une **hiérarchie des pays**.

Les colonnes du tableau permettant le calcul des trois dépenses moyennes, sont données ci-dessous.

La **dépense arithmétique moyenne** (ou \bar{D}), pour les 14 pays i , est donnée par :

$$\bar{D} = \frac{\sum_{i=1}^{14} ni}{14}$$

Fleurs coupées		Plantes d'intérieur		Ensemble	
xi	$ni1$	xi	$ni2$	xi	$ni3 = ni1+ni2$
Norvège	79	Norvège	59	Norvège	138
Royaume Uni	71	Allemagne	58	Danemark	121
Danemark	68	Autriche	56	Allemagne	110
Irlande	64	Danemark	53	Royaume Uni	105
Allemagne	52	Suède	52	Autriche	98
Hollande	50	Belgique	38	Suède	95
Suède	43	Royaume Uni	34	Irlande	94
Autriche	42	France	31	Hollande	80
Belgique	40	Irlande	30	Belgique	78
Etats Unis	35	Hollande	30	France	62
France	31	Etats Unis	20	Etats Unis	55
Japon	21	Pologne	11	Japon	32
Pologne	19	Canada	11	Pologne	30
Canada	18	Japon	11	Canada	29
Total/tête	633	Total/tête	494	Total/tête	1127
Moyenne (D)	45	Moyenne (D)	35	Moyenne (D)	81

On lit que les pays représentatifs de la dépense moyenne (ou Pays moyen) sont :

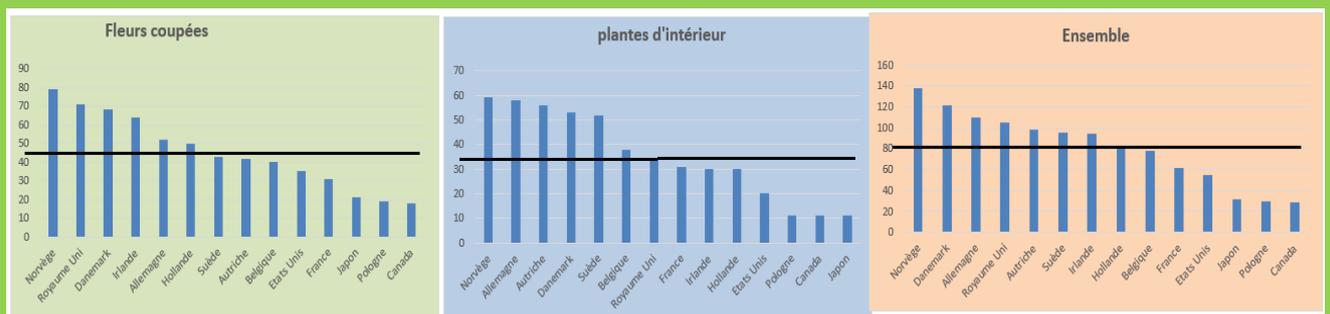
Fleurs coupées : Suède

Plantes d'intérieur : Royaume uni

Ensemble (Fleurs et plantes) : Hollande

Au-delà de chaque moyenne, se situe le groupe des pays *sur consommateurs* (en couleur), et en deçà, celui des pays *sous-consommateurs* (en blanc).

Le diagramme différentiel peut être « **en tuyaux d'orgue** », ici en ordonnées ni , si l'on veut éviter le calcul des fréquences ($fi\%$). Il permet une lecture plus rapide de l'enseignement du tableau, l'*horizontale représentant la dépense moyenne (ou le pays moyen)* :

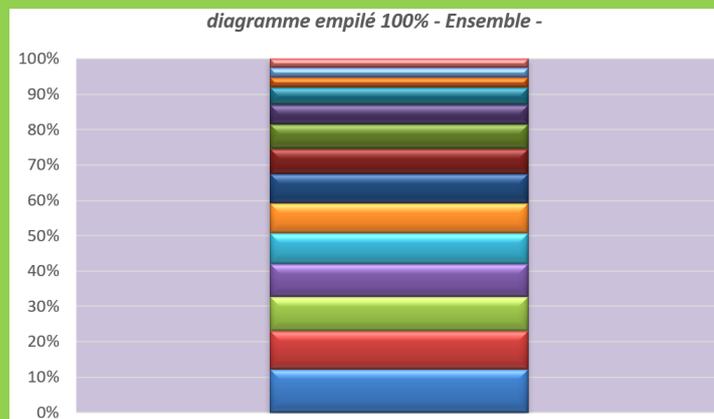


5) Pour l'ensemble des dépenses (fleurs et plantes) représenter le diagramme intégral.

Le diagramme intégral est **le diagramme empilé 100%**. Il suppose le calcul des fréquences simples ($fi3\%$), que l'on cumule en Fréquences Cumulées ($Fi3\%$). Des colonnes de calcul additionnelles sont donc nécessaires, soit :

$$(fi_3)\% = (ni_3 / \sum_{i=1}^{14} ni_3) \times 100\% \quad \text{et} \quad (Fi_3)\% = ((fi_3)\% - 1) + (fi_3)\%$$

Ensemble			
x_i	$ni_3 = ni_1 + ni_2$	$(fi_3)\%$	$(Fi_3)\%$
Norvège	138	12%	12%
Danemark	121	11%	23%
Allemagne	110	10%	33%
Royaume Uni	105	9%	42%
Autriche	98	9%	51%
Suède	95	8%	59%
Irlande	94	8%	68%
Hollande	80	7%	75%
Belgique	78	7%	82%
France	62	6%	87%
Etats Unis	55	5%	92%
Japon	32	3%	95%
Pologne	30	3%	97%
Canada	29	3%	100%
Total/tête	1127	100%	
Moyenne (D)	81		



6) Commentez en plusieurs phrases, de deux manières :

a. En vous adressant à un novice en statistique et en floriculture

Par exemple : « C'est en Norvège que l'on dépense le plus pour acheter des fleurs et des plantes ».

ou : « Le Canada est le pays où les dépenses en fleurs sont les plus faibles ».

b. En vous adressant à un étudiant en statistique qui n'ignore pas les fleurs.

Par exemple : La distribution des dépenses en fleurs et plantes d'intérieur, dans le monde, montre de fortes disparités. Les $\frac{3}{4}$ des dépenses sont réalisées par 8 pays. Les pays du Nord sont les plus « florifères » : La Norvège qui se distingue, le Danemark et l'Allemagne réalisent à eux seuls le $\frac{1}{3}$ des dépenses. Toutefois, les dépenses au Royaume Uni et en Irlande sont au-delà de la moyenne.

A l'opposé, les pays les moins florifères sont : Canada, Pologne, Japon, Etats Unis, et France.

Une interrogation sur les causes permettrait d'approfondir ce constat. Mais il ne semble pas que les régimes politiques aient une influence notable !!.

La dépense en floriculture ne montre une supériorité des dépenses en plantes d'intérieur par rapport aux fleurs coupées que pour quelques pays : Allemagne, Autriche et Suède. La France se distingue par une égalité des deux types de dépenses. Cependant, la supériorité de dépenses en fleurs coupées, lorsqu'elle existe, est le plus souvent importante : du tiers au double. Ressortent donc des préférences marquées.

*Enfin, s'il est admis que la Hollande est **le pays des fleurs**, il n'apparaît pas qu'il est celui où l'on dépense le plus pour les fleurs. L'expression a donc plutôt trait non à la consommation, mais à **la production et à l'exportation de fleurs et plantes**.*

