CHAPITRE 5 - RESUMER UNE DISTRIBUTION : LES PRINCIPAUX INDICES STATISTIQUES.

1 Résumé

Pour résumer une distribution, on calcule des indices de position ou de tendance centrale et des indices de dispersion. Le choix des indices dépend de ce qu'on souhaite résumer dans la distribution et de l'échelle de mesure de la variable (se reporter au tableau suivant). Les questions qu'on peut se poser sont :

- Sur quelle(s) modalité(s) se concentre(nt) les observations (concentration) ?
- Comment les observations se répartissent-elles dans la distribution (répartition) ?
- Quel est le centre de gravité de la distribution et la variation moyenne autour de ce centre (centre et variation) ?

		Echelles de mesure		
Questions	**Indices**	**nominale**	**Ordinale**	**Numérique**
Concentration	de position	mode Mode secondaire	mode Mode secondaire	mode Mode secondaire
Répartition	de tendance centrale de dispersion		Médiane Quartile 1 et 3	Médiane Quartile 1 et 3
Centre et variation	de tendance centrale de dispersion			Moyenne Ecart-type

Définitions :

Le *mode* est la modalité la plus fréquente.

Le *mode secondaire* est la modalité la plus fréquente après le mode.

La *médiane* (ou quartile 2) est la modalité dont l'effectif cumulé correspond à n/2.

Le *quartile 1* est la modalité dont l'effectif cumulé à gauche correspond à n/4.

Le *quartile 3* est la modalité dont l'effectif cumulé à gauche correspond à n*3/4.

La *moyenne* est la somme des observations divisée par le nombre d'observations. On la note m. Elle représente le centre de gravité de la distribution : $m=\Sigma x/n$

La *variance* est la moyenne des carrés des écarts à la moyenne. On la note s^2 :

$$s^2 = \frac{\sum (x_i - m)^2}{n} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}$$

L'écart-type est la racine carrée de la variance. On le note s. Il représente la variation moyenne des observations autour de la moyenne du protocole. On l'utilise comme indice de dispersion pour résumer un protocole.

La *variance corrigée* est la somme des carrés des écarts à la moyenne pondérée par n-1. On la note $s^2_{corr} = \text{var}_{corr}$:

$$s^2_{corr} = \frac{\sum (x_i - m)^2}{n-1} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{s^2 * n}{n-1}$$

L'écart-type corrigé est la racine carrée de la variance corrigée. Il représente d'un point de vue descriptif une mesure de la variabilité intersujet, et d'un point de vue inférentiel, une estimation de l'écart-type de la population parente. On l'utilise lorsqu 'on veut essayer de généraliser les observations faites sur l'échantillon à l'ensemble de la population dont il est issu.

2 Les erreurs fréquentes à éviter

Déclarer un effectif cumulé au lieu d'une modalité comme médiane ou quartile. C'est une erreur assez fréquente qui consiste à annoncer que la médiane est égale à n/2 au lieu d'annoncer la modalité correspondante. Pour l'éviter, il faut garder à l'esprit que médiane et quartiles sont des coupures de l'échelle de mesure (et non de la distribution cumulée). Ce qu'on désire savoir avec ces indices c'est sur quelles modalités se répartissent les observations et non la valeur de n/2.

Confondre la somme des observations avec la somme des modalités. Cette erreur très fréquente résulte d'une confusion entre le tableau de protocole et le tableau de distribution. Il est donc tout à fait essentiel de bien les différencier. Le moyen le plus efficace pour faire la différence est de vous demander à quoi correspondent les lignes du tableau. Si les lignes correspondent aux individus statistiques alors vous travaillez sur le tableau du protocole. Si les lignes correspondent aux modalités de la variable alors vous travaillez sur le tableau de distribution. Pour vous y retrouver, il est impératif d'avoir au préalable identifié vos unités statistiques et vos variables.

Confondre le nombre d'observations et le nombre de modalités. Cette erreur tout aussi fréquente que la précédente relève des mêmes difficultés et sera évitée de la même façon.

Confondre le carré de la somme des observations et la somme des carrés des observations. Cette erreur est liée entre autres à une difficulté à intégrer les écritures formelles et à un non-respect de l'ordre de priorité des opérations. Le carré de la somme des observations $(\Sigma x)^2$ suppose qu'on fasse d'abord la somme des observations (souvenez-vous qu'il faut d'abord traiter l'intérieur des parenthèses). La somme des carrés des observations Σx^2 requiert de faire d'abord l'élévation au carré, puis la somme (l'ordre de priorité des opérations n'a pas changé depuis que vous avez quitté le collège, on commence d'abord par les puissances, on fait ensuite les multiplications ou les divisions et on termine par les sommes et les différences).

Annoncer une variance ou un écart-type négatif. La variance est une moyenne de carrés. Un carré étant forcément positif (la multiplication de deux nombres relatifs de même signe donne un nombre positif), leur moyenne ne saurait être négative. Elle représente une distance et une distance négative n'a pas de sens.

Annoncer une moyenne qui sort de l'intervalle de variation de la variable. Cette erreur, pour le moins étonnante, consiste a annoncer une moyenne qui est située au delà de la plus petite ou de la plus grande des observations. La moyenne est le centre de la distribution, elle est donc forcément située entre les deux extrêmes observés.

Annoncer un écart-type plus grand que la moyenne. Cette erreur résulte d'une mauvaise représentation de ce que sont un écart-type et une moyenne. L'écart-type représente la distance moyenne qui sépare les observations de la moyenne du protocole, la moyenne étant le centre de la distribution de ce protocole. La plus grande distance qu'on puisse trouver entre la moyenne et les deux extrémités de la distribution, c'est lorsque la moyenne est au milieu de l'intervalle de variation. L'écart moyen entre les observations et cette moyenne est alors au maximum de la moitié de l'intervalle de variation. L'écart-type est donc forcément inférieur ou égal à la moyenne (sauf si votre échelle de mesure comprend des nombres négatifs). Le cas d'égalité entre la moyenne et l'écart-type est un cas extrême et rarissime. Il suppose que la moyenne soit située à mi-chemin entre la plus grande et la plus petite valeur observée et que toutes les valeurs observées soient situées uniquement sur les extrémités de la distribution. Dans ce cas, on peut douter du caractère numérique de la variable (quel est le sens de la notion d'intervalle si seulement deux modalités sont observées ?) et donc de la pertinence du calcul de la moyenne et de l'écart-type.

Confondre la moyenne du protocole avec la moyenne scolaire. Cette erreur consiste à prendre comme moyenne 10 au lieu de la moyenne du protocole, erreur surtout observée lorsque la variable va de 0 à 20. Le langage courant, sous l'effet de notre passage à tous dans le milieu scolaire, désigne sous le nom de moyenne la note 10. C'est effectivement une moyenne, au sens arithmétique du terme. C'est la moyenne des modalités de la variable (pour une note sur 20). Il ne faut, bien entendu, pas la confondre avec la moyenne des observations dont nous nous occupons en statistiques.

3 Exercices

Exercice 10. Situez le mode de la distribution en classes sur la note au test de l'exercice 9 que nous reproduisons ci-dessous.

Intervalle	Valeur centrale	Limites inférieures	Effectifs
6-10	8	5,5	3
11-15	13	10,5	3
16-20	18	15,5	9
21-25	23	20,5	13
26-30	28	25,5	14
31-35	33	30,5	17
36-40	38	35,5	23
41-45	43	40,5	24
46-50	48	45,5	7
		Total	113

Exercice 11. Situez la médiane et les quartiles de la distribution donnée à l'exercice précédent.

Exercice 12. Calculer la moyenne et l'écart-type des notes au test, à partir du protocole, puis à partir de la distribution. Pour plus de commodités, nous rappelons l'énonce de l'exercice 6 ci-dessous.

Voici un protocole correspondant à la passation, par un ensemble de 113 sujets d'un test (Faverge, 1966). Les individus statistiques sont les sujets. La variable correspond au nombre de réponses correctes sur un ensemble de 50 items. Nous avons une seule variable, l'échelle de mesure est une échelle d'intervalle. Nous avons donc ici un protocole univarié non structuré.

I	**Note**	**I**	**Note**	**I**	**Note**	**I**	**Note**	**I**	**Note**	**I**	**Note**	**I**	**Note**	**I**	**Note**
S1	43	S16	37	S31	49	S46	36	S61	35	S76	37	S91	50	S106	33
S2	31	S17	40	S32	24	S47	30	S62	30	S77	45	S92	42	S107	25
S3	38	S18	32	S33	18	S48	12	S63	25	S78	43	S93	44	S108	18
S4	40	S19	13	S34	49	S49	25	S64	20	S79	44	S94	20	S109	44
S5	40	S20	24	S35	41	S50	16	S65	39	S80	19	S95	44	S110	37
S6	41	S21	28	S36	25	S51	36	S66	37	S81	34	S96	21	S111	48
S7	28	S22	33	S37	27	S52	25	S67	39	S82	39	S97	26	S112	29
S8	41	S23	18	S38	43	S53	32	S68	36	S83	24	S98	32	S113	35
S9	43	S24	36	S39	11	S54	41	S69	10	S84	34	S99	36		
S10	41	S25	47	S40	33	S55	31	S70	26	S85	34	S100	44		
S11	41	S26	29	S41	33	S56	36	S71	23	S86	21	S101	37		
S12	41	S27	38	S42	35	S57	29	S72	27	S87	38	S102	8		
S13	35	S28	33	S43	45	S58	22	S73	26	S88	19	S103	29		
S14	40	S29	44	S44	43	S59	36	S74	20	S89	42	S104	42		
S15	48	S30	44	S45	9	S60	37	S75	24	S90	46	S105	27		

4 Solutions

Solution de l'exercice 10

Le mode correspond à la modalité dont l'effectif est le plus élevé. Dans notre exemple Il s'agit de la classe 41-45. L'effectif de la classe 36-40 est très proche. Cette classe constitue le mode secondaire.

Solution de l'exercice 11

Détermination de la médiane. Pour déterminer la médiane et les quartiles, il faut une distribution cumulée à gauche ou à droite ce qui implique que la variable soit ordinale ou numérique. Nous reprendrons la même distribution en classes pour exemplifier la détermination de la médiane.

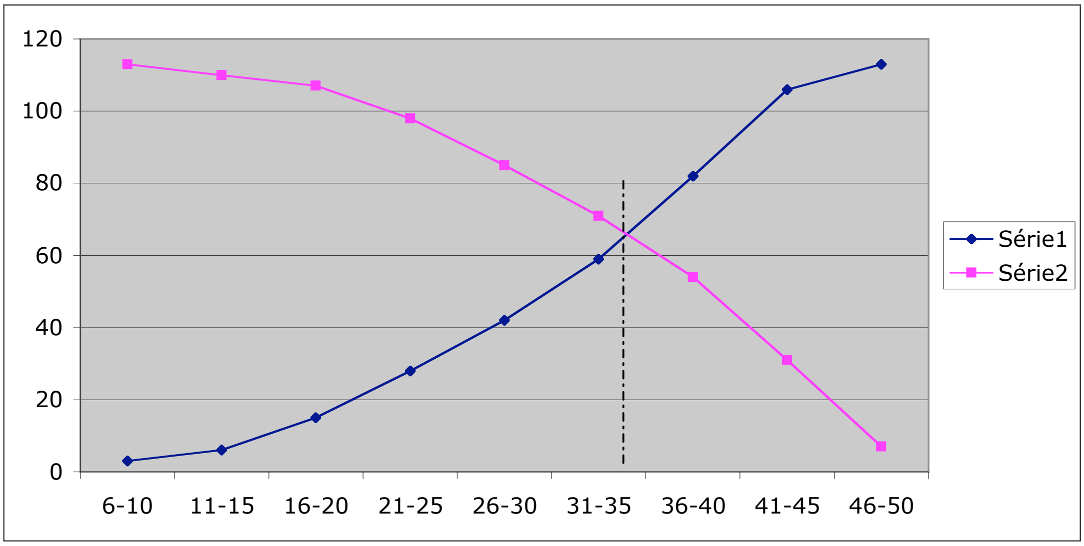
Valeur mini	Valeur maxi	Valeur centrale	Effectif	Effectifs cumulés à gauche	Effectifs cumulés à droite
6	10	8	3	3	113
11	15	13	3	6	110
16	20	18	9	15	107
21	25	23	13	28	98
26	30	28	14	42	85
31	35	33	17	59	71
36	40	38	23	82	54
41	45	43	24	106	31
46	50	48	7	113	7

Le détermination de la médiane se fait en deux temps :

1) Calculer l'effectif cumulé correspondant à la moitié des observations soit n/2. Dans notre exemple n vaut 113 et n/2=113/2=66,5

2) Chercher dans les effectifs cumulés n/2. Dans notre cas, on cherchera 66,5. Si une modalité correspond à cet effectif cumulé, on la prend comme médiane. Mais, dans notre exemple, et c'est souvent le cas, cet effectif cumulé ne correspond pas à une modalité. On prendra donc l'effectif cumulé le plus proche. Il s'agit de la classe 31-35. Dans le cas où l'effectif cumulé correspondant à la médiane est à égale distance entre deux modalités, on s'abstiendra de choisir et on parlera de coupure quasi-médiane entre les deux modalités en question. Si plusieurs modalités correspondent à l'effectif cumulé (cas où l'on a des effectifs nuls dans des classes successives), on prend la première modalité.

Détermination graphique de la médiane. On peut déterminer graphiquement la médiane en faisant sur un même graphique la courbe des effectifs cumulés à gauche (série 1) et à droite (série 2). Pour trouver la médiane, on abaisse la perpendiculaire à l'axe des abscisses qui passe par l'intersection des deux courbes.



Détermination des quartiles. La procédure est similaire à celle de la détermination de la médiane, mais, dans ce cas, on cherche pour le premier quartile la modalité dont l'effectif cumulé correspond à un quart des observations soit n/4. Pour le troisième quartile, on cherche la modalité dont l'effectif cumulé est égal aux trois quarts des observations soit n*3/4 (le deuxième quartile correspond à la médiane). Dans notre exemple on cherche donc les modalités dont l'effectif cumulé correspond à :

- Pour le premier quartile 113/4 soit 28,25,
- Pour le troisième quartile 113*3/4=84,75.

On peut voir que la modalité dont l'effectif cumulé est le plus proche de 28,25 est la classe 21-25, elle constituera donc notre premier quartile (Q1). La modalité dont l'effectif cumulé est le plus proche de 84,75 est la classe 36-40. Ce sera notre troisième quartile (Q3).

Interprétation des quartiles et de la médiane. La médiane nous indique quelle modalité coupe en deux la distribution. Elle montre donc que la moitié des sujets est située en dessous de cet échelon et la moitié au-dessus. Dans notre exemple, nous pouvons donc dire qu'un sujet sur deux appartient aux classes inférieures (ou supérieures ce qui est équivalent) à la classe 31-35. Un commentaire similaire peut-être fait pour Q1 et Q3, mais, dans ce cas les classes supérieures et inférieures ne sont plus symétriques. On aura ainsi

- Avec Q1 : Un quart des sujets appartient à une classe inférieure à la classe 21-25 et par complémentarité, les trois quarts des sujets, appartiennent à une classe supérieure à la classe 21-25.
- Avec Q3 : Les trois quarts des sujets appartiennent à une classe inférieure à la classe 36-40 et par complémentarité, un quart des sujets appartient à une classe supérieure à la classe 36-40.

Mais, plutôt que de commenter chacun de ces indices séparément, il est plus informatif et plus synthétique de fonder son commentaire sur la combinaison des trois indices. On pourra ainsi dire que la moitié des sujets a obtenu une note comprise entre 23 (valeur centrale de Q1) et 38 (valeur centrale de Q3) avec une médiane à 33 (valeur centrale de la classe médiane). Puisqu'on sait, avec la distribution, que les notes observées vont de 8 à 50, on situe tout de suite la répartition des observations du côté des valeurs élevées, mais aussi que la répartition des notes autour de la médiane n'est pas symétrique.

Solution de l'exercice 12

Commençons par la moyenne. On sait calculer une moyenne depuis l'école primaire, Il s'agit de la somme des observations divisée par le nombre d'observations. Mais lorsqu'on a un ensemble important de données, cela peut créer quelques difficultés. Nous noterons la somme des observations Σx, le total des observations n et la moyenne m. On a donc m = Σx/n.

En pratique, pour calculer une moyenne, on peut partir du protocole ou de la distribution. Il est tout à fait important de bien différencier les deux, car la procédure n'est pas la même. Il est rare que le protocole se présente sous la forme d'une série de données en vrac. On les présente en général sous la forme d'un tableau et c'est ce qui induit en erreur bon nombre d'étudiants parce qu'ils ne savent pas différencier le tableau d'un protocole de celui d'une distribution. Pour bien les différencier, il convient de se demander à quoi correspondent chacune des lignes. Dans un tableau de protocole, les lignes correspondent aux individus statistiques. Dans une distribution, les lignes correspondent aux modalités de la variable. Bien rangé, le tableau de protocole correspondant à notre exemple aurait l'allure suivante :

Individus (i)	**Observation (x)**
1	43
2	31
3	38
...	...
110	37
111	48
112	29
113	35

Il fait correspondre à un ensemble d'individus (notés i) un ensemble d'observations (notées x). Chaque observation relative à un sujet est notée x_i (le i renvoyant à l'individu statistique). La somme des observations se notera donc Σx_i. La moyenne sera $m=\Sigma x_i/n$. Concrètement cela veut dire qu'on fait la somme de la colonne des observations et qu'on la divise par le nombre des observations. Dans notre exemple, on aura : (43+31+38+...+37+48+29+35)/113=32,95

Calcul de la moyenne à partir de la distribution. Si on part de la distribution, la procédure est un peu différente. Le tableau de distribution nous donne le nombre de fois où chaque modalité a été observée (effectif). Il convient alors, pour faire le total des observations, de multiplier les modalités par leur effectif. Pour bien comprendre cela, imaginez que dans votre bulletin scolaire vous ayez obtenu 3 fois la note 9 et 2 fois la note 11 (oui, je sais, ce n'est pas cher payé), vous comprenez aisément qu'il est équivalent pour calculer la moyenne de faire : (9+9+9+11+11)/5=9,8 (ce qu'on ferait en partant du protocole) ou (9*3+11*2)/5=9,8 (ce qu'on fait en partant de la distribution). Concrètement, nous venons de résumer le protocole (votre bulletin scolaire) sous la forme d'une distribution qu'on pourrait représenter sous la forme du petit tableau suivant :

Notes (ou modalités de la variable)	**Effectifs (ou nombre de notes)**
9	3
11	2

Si vous partez de la distribution, il vous faudra

1) Multiplier chaque modalité par son effectif,
2) Faire la somme de ces produits,
3) Diviser par le total des effectifs (nombre d'observations).

Un peu de formalisme. Appelons maintenant u^k les modalités de la variable et n^k les effectifs (pour faire simple, nous interpréterons k comme un renvoi à une ligne du tableau de distribution sans tenir compte de sa position en indice ou en exposant). La somme des observations est alors $\Sigma n_k u^k$, ce qui se lit somme des produits de n_k par u^k, et la moyenne est alors égale à $\Sigma n^k u_k$ /n. Concrètement sur l'exemple des notes au test, on construira le tableau suivant :

u^k	n_k	$n_k u^k$
6	0	0
7	0	0
8	1	8
...	...	...
...	...	...
47	1	47
48	2	96
49	2	98
50	1	50
Total	113	3723

La moyenne est donc de 3723/113=32,95

Calcul de la variance et de l'écart-type. En pratique, on ne peut pas calculer directement l'écart-type. Il faut pour cela d'abord calculer la variance et ensuite extraire sa racine carrée. Comme pour la moyenne (mais après tout la variance est aussi une moyenne), on peut partir du protocole ou de la distribution. Dans ces deux cas, on a deux formules permettant de faire ce calcul: la formule de définition et la formule dite "de calcul" parce qu'elle est plus souvent utilisée car elle permet d'éviter les erreurs liées aux arrondis. Rappelons ici la définition de la variance (il va falloir s'en imprégner) : la variance est la moyenne des carrés des écarts à la moyenne.On note la variance : s^2 et l'écart-type : s. On trouvera parfois la notation suivante pour la variance : var ou var(x). Concrètement qu'est-ce que cela signifie ? Les écarts à la moyenne s'écrivent (x_i-m) ce qui signifie qu'il faut soustraire la moyenne du protocole à chaque observation. Le carré des écarts à la moyenne s'écrira donc $(x_i-m)^2$, on élève donc chacun de ces écarts au carré. La moyenne de ces carrés des écarts à la moyenne s'écriera donc $s^2 =\Sigma(x_i-m)^2/n$. Ce qui signifie qu'on va faire la somme des carrés des écarts à la moyenne et les diviser par le nombre d'observations.

$$s^2 = \frac{\sum (x_i - m)^2}{n}$$

Utilisation de la formule de définition à partir du protocole. Cette petite explication détaillée de la formule de définition vous donne la procédure pour calculer la variance à partir du protocole. Concrètement voici comment cela on procède :

- On reprend le tableau du protocole, puis on ajoute deux colonnes, l'une pour calculer les écarts à la moyenne, l'autre pour calculer les carrés de ces écarts.
- On calcule la moyenne. Dans notre exemple, m=32,95
- Pour chaque individu statistique, on calcule son écart à la moyenne soit (x_i-m)
    - Pour i_1, on a : 43-32,95=10,05
    - Pour i_2, on a : 31-32,95=-1,95
    - Pour i_3, on a : 38-32,95=5,05
- Pour chaque individu statistique, on élève son écart à la moyenne au carré soit $(x_i-m)^2$
    - Pour i_1, on a : 10,052=101,06
    - Pour i_2, on a : -1,952=3,79
    - Pour i_3, on a : 5,052=25,53
- On fait ensuite la somme de ces carrés qu'on divise par le nombre d'individus statistiques soit $\Sigma(x_i-m)^2$ /n. Nous avons donc s^2 = (101,06+3,79+25,53+...+15,58+4,22) / 113 = 97,06

i	**x_i**	**x_i-m**	**$(x_i-m)^2$**
1	43	10,05	101,06
2	31	-1,95	3,79
3	38	5,05	25,53
...	...		
110	37	4,05	16,43
111	48	15,05	226,6
112	29	-3,95	15,58
113	35	2,05	4,22
Total	3723	0	10967,68

L'écart-type est la racine carrée de la variance soit : $s = \sqrt{\text{var}} = \sqrt{97,06} = 9,85$

Utilisation de la formule de définition à partir de la distribution. La difficulté, ici, est similaire à celle que nous avons rencontré lors du calcul de la moyenne à partir de la distribution. On peut bien calculer l'écart à la moyenne de chaque modalité, mais il ne faut pas oublier de multiplier cet écart par le nombre de fois où il a été observé, c'est-à-dire par l'effectif. Concrètement, on commence par préparer le tableau de distribution en lui ajoutant trois colonnes:

- Une pour le calcul de l'écart à la moyenne.
- Une autre pour l'élévation de cet écart au carré
- Et une troisième pour le produit de ce carré par l'effectif de la modalité.

On calcule l'écart de chaque modalité avec la moyenne du protocole soit u^k – m.

- Pour u^1, on a : 32,95-6 = -26,95
- Pour u^2, on a : 32,95-7 = -25,95
- Pour u^3, on a : 32,95-8 = -24,95 etc.

Pour chaque modalité, on calcule le carré de cet écart, soit $(u^k - m)^2$.

- Pour u^1, on a : $-26,95^2$ = 726,14
- Pour u^2, on a : $-25,95^2$ = 673,24
- Pour u^3, on a : $-24,95^2$ = 622,35 etc.

Pour chaque modalité, on calcule le produit de cet écart et de son effectif soit: $(u^k - m)^2 * n_k$

- Pour u^1, on a :726,14 * 0 = 0
- Pour u^2, on a : 673,24* 0 = 0
- Pour u^3, on a : 622,35 * 1 = 622,35 etc.

On fait ensuite la somme de ces produits et on la divise par n soit $s^2=\Sigma((u^k - m)^2 * n_k)/n$. Dans notre exemple, nous avons donc $s^2=(0+0+622{,}35+...+515{,}40+290{,}81)/113=10967{,}68/113= 97{,}06$

u^k	n_k	u^k-m	$(u^k$-m)2	$n(u^k-m)^2$
6	0	-26,95	726,14	0
7	0	-25,95	673,24	0
8	1	-24,95	622,35	622,35
46	1	13,05	170,38	170,38
47	1	14,05	197,49	197,49
48	2	15,05	226,6	453,19
49	2	16,05	257,7	515,4
50	1	17,05	290,81	290,81
Total	113		Total	10967,68

L'écart-type se calcule de la même façon que précédemment variance soit $s = \sqrt{\text{var}} = \sqrt{97{,}06} = 9{,}85$

Utilisation de la formule de calcul. On peut montrer que la formule de définition est équivalente à la seconde formule:

$$s^2 = \frac{\sum (x_i - m)^2}{n} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}$$

Cette deuxième formule signifie qu'il faut faire la différence entre la somme des carrés des observations (Σx^2) et le carré de la somme des observations pondérée par n soit $(\Sigma x)^2/n$) et diviser ensuite cette différence par n. Pour ceux que cela intéresse, vous pouvez essayer de retrouver la suite de transformations algébriques qui permettent de passer de l'une à l'autre. Pour les autres, passons à sa mise en application. Encore une fois, nous devons différencier l'utilisation du protocole ou de la distribution comme point de départ des calculs.

Utilisation de la formule de calcul à partir du protocole. Concrètement, il va falloir calculer la somme des carrés et la somme des observations pour pouvoir appliquer la formule. On prépare donc un tableau correspondant à son protocole et on lui ajoute une colonne pour calculer les carrés des observations. On voit tout de suite l'avantage de cette procédure, Puisque la somme des observations a déjà été calculée pour la moyenne, il n'y a plus qu'une colonne de calculs à faire. Pour chaque observation, on calcule son carré.

- Pour i_1, on a : 43^2=1849
- Pour i_2, on a : 31^2=961
- Pour i_3, on a : 38^2=1444

On fait ensuite la somme des carrés. Dans notre exemple, nous avons Σx^2= 133629. Si ce n'est déjà fait pour la moyenne, on calcule ensuite la somme des observations. Dans notre exemple, cette somme est de Σx = 3723

i	$\mathbf{x_i}$	$\mathbf{x_i^2}$
1	43	1849
2	31	961
3	38	1444
4	40	1600
...	...	....
110	32	1024
111	36	1296
112	44	1936
113	37	1369
Total	3723	133629

Dans la mesure où l'application de la formule est la même que l'on parte d'un protocole ou d'une distribution, nous développerons son application à la fin.

Utilisation de la formule de calcul à partir de la distribution. Le problème est toujours le même dans le passage du protocole à la distribution pour le calcul de ces indices, il ne faut pas oublier de multiplier par l'effectif. Concrètement on prépare un tableau de distribution en lui ajoutant deux colonnes. La première servira à calculer le carré des modalités soit u^{k2}, la seconde servira à multiplier ce carré par l'effectif de la modalité soit $n_k(u^{k2})$. Si la somme des observations n'a pas été calculée pour la moyenne, on aura besoin d'une troisième colonne pour calculer le produit de chaque modalité par son effectif soit $n_k u^k$.

On calcule alors la somme des carrés des observations (Σx^2) Pour chaque modalité, on commence par calculer son carré. Par exemple,

- Pour u^1, on a 6^2=36
- Pour u^2, on a 7^2=49
- Pour u^3, on a 8^2=64 etc.

Ensuite, pour chaque modalité, on multiplie le carré par l'effectif de la modalité.

- Pour u^1, on a 36*0=0
- Pour u^2, on a 49*0=0
- Pour u^3, on a 64*1=64 etc.

<u>Attention,</u> il s'agit bien ici de multiplier le carré de la modalité par son effectif et non de faire le carré du produit de l'effectif par la modalité. Ainsi, pour la note 48, on a 48^2*2=2401*2=4802 et à ne pas confondre avec le carré du produit de l'effectif par la modalité : $(48*2)^2$=9216

On fait ensuite la somme de ces produits. Dans notre exemple, Σx^2=133629

Si ce n'est déjà fait lors du calcul de la moyenne, on calcule la somme des observations (Σx).On multiplie chaque modalité par son effectif, par exemple

- Pour u^1, on a 6*0=0
- Pour u^2, on a 7*0=0
- Pour u^3, on a 8*1=8 etc.

On fait la somme de ces produits. Dans notre exemple, $\Sigma x=3723$

u^k	n_k	u^{k2}	$u^{k2}*n$	$u^k n_k$
6	0	36	0	0
7	0	49	0	0
8	1	64	64	8
9	1	81	81	9
...	...	...	...	...
47	1	2209	2209	47
48	2	2304	4608	96
49	2	2401	4802	98
50	1	2500	2500	50
Total	113		133629	3723

On applique ensuite la formule en calculant dans l'ordre

- Le carré de la somme des observations $(\Sigma x)^2 = 3723^2=13860729$
- On divise ensuite ce carré par n, soit $(\Sigma x)^2/n$ =13860729/113=122661,319
- On fait ensuite la différence avec la somme des carrés, soit :
- $\Sigma x^2-((\Sigma x)^2/n$=133629-122661,319=10967,68
- Enfin on divise le tout par n soit : $s^2=(\Sigma x^2-((\Sigma x)^2/n)/n$ = 10967,68/113=97,06

L'écart-type se calcule de la même façon que précédemment. Il est égal à la racine carrée de la variance, soit s=9,85. Le résultat est bien sûr identique à celui que nous avions trouvé précédemment. En pratique, ce n'est pas toujours le cas, notamment si on doit effectuer le calcul à la main. Cela tient au fait que, dans l'utilisation de la formule de définition, on est conduit à faire beaucoup plus d'arrondis (dans les différences avec la moyenne) ce qui entraîne une perte de précision. Il est donc préférable d'utiliser la formule de calcul.