

## Laboratoire 2 : Gestion des données et statistiques descriptives

Pour ce laboratoire, nous utiliserons les données du Programme pancanadien d'évaluation (PPCE) 2013 élaboré par le Conseil des ministres de l'Éducation (Canada). Cette évaluation est menée auprès d'élèves de 8<sup>e</sup> année tous les trois ans pour recueillir des données sur leurs résultats scolaires en mathématiques, en lecture et en sciences partout au Canada. Le premier PPCE a été réalisé en 2007. Un objectif du PPCE est de déterminer si les élèves aux quatre coins du pays atteignent des niveaux de rendement similaires à environ la même période de leur parcours scolaire, malgré les différences dans les programmes scolaires d'un bout à l'autre du pays.

Nous vous avons fourni les données du PPCE 2013 pour les élèves et le personnel enseignant de l'Alberta. Dans ce laboratoire, nous nous concentrerons sur les réponses des élèves. Veuillez suivre ce laboratoire pour bien comprendre l'ensemble de données. Pour ce faire, vous devrez peut-être consulter votre **manuel** ou **Internet**. Ce laboratoire a été conçu pour faire progresser votre apprentissage, ce qui signifie que les premières questions seront accompagnées d'instructions complètes, alors que les questions subséquentes vous donneront moins d'instructions. Ainsi, vous pourrez perfectionner vos compétences au fur et à mesure que les instructions deviennent moins détaillées au cours du laboratoire (p. ex., vous verrez à quoi ressemble exactement le Tableau 1, mais ce sera à vous de créer le Tableau 4 au complet.)

### 1. Statistiques descriptives : examen des valeurs aberrantes et des erreurs de saisie

Les erreurs de saisie sont difficiles à repérer dans un tableur, mais peuvent avoir une incidence **MAJEURE** sur les analyses subséquentes. Par conséquent, il est important de vous assurer que tous vos ensembles de données sont corrects avant de commencer toute analyse. Les analyses, si bonnes soient-elles, n'ont aucune utilité si elles sont basées sur des données incorrectes!!! Une façon de repérer les valeurs aberrantes et les erreurs de saisie est de réaliser des statistiques descriptives pour chaque variable. Veuillez procéder comme suit : Analyse → Statistiques descriptives → Descriptives → sélectionnez toutes les variables → Options → veillez à ce que les options « Minimum », « Maximum » et « Plage » soient cochées → Poursuivre → OK.

#### Question :

1. En regardant les résultats de vos examens des valeurs aberrantes, avez-vous trouvé des valeurs aberrantes? Si oui, veuillez énumérer les valeurs aberrantes. Qu'est-ce

qui aurait pu causer ces valeurs aberrantes dans votre ensemble de données? En regardant les résultats, comment avez-vous constaté qu'il y avait des valeurs aberrantes, s'il y a lieu?

## 2. Compréhension de vos données

Pour mieux comprendre vos données, veuillez fournir quelques statistiques descriptives de base concernant chacune de vos variables. Veuillez calculer la moyenne, l'aplatissement (kurtosis), l'asymétrie (skewness), le minimum, le maximum, la plage, la variance et l'écart type pour les scores en sciences et chacun des items concernant l'intimidation, en procédant comme suit : Analyse → Statistiques descriptives → Descriptives → sélectionnez les variables appropriées et déplacez-les vers « Variable(s) » → Options → sélectionnez les analyses appropriées → Poursuivre → OK.

### Questions :

2. Veuillez répondre à ces questions pour les variables de l'échelle :
  - a) Combien de personnes ont répondu à votre sondage? \_\_\_\_\_
  - b) Minimum \_\_\_\_\_ Maximum \_\_\_\_\_
  - c) Moyenne \_\_\_\_\_ Plage \_\_\_\_\_ Écart type \_\_\_\_\_ Variance \_\_\_\_\_
  - d) Skewness (asymétrie) \_\_\_\_\_ Kurtosis (aplatissement) \_\_\_\_\_ Qu'indiquent les valeurs « Skewness » (asymétrie) et « Kurtosis » (aplatissement)? (Vous aurez peut-être besoin d'utiliser Google.)
  
3. Quelle est la différence entre la variance et l'écart type?

Pour présenter les données de la question 2, il faut utiliser un tableau. Veuillez remplir le tableau ci-dessous avec les réponses pour les variables de l'échelle :

Tableau 1

*Titre (veuillez créer un titre pour ce tableau)*

	n	Moyenne	Écart type	Skewness (asymétrie)	Kurtosis (aplatissement)
Score en sciences	??	??	??	??	??
SQ13A	??	??	??	??	??
SQ13B	??	??	??	??	??
SQ13C	??	??	??	??	??
<b>Etc.</b>		<b>Etc.</b>	<b>Etc.</b>	<b>Etc.</b>	<b>Etc.</b>

**\*\*\*Veuillez à ce que le tableau soit en format APA\*\*\***

Maintenant, il faut réaliser des statistiques descriptives sur des variables nominales. Veuillez calculer des comptes d'effectifs pour ces variables en procédant comme suit : Analyse → Statistiques descriptives → Effectifs → sélectionnez les variables appropriées et déplacez-les vers « Variable(s) » → OK.

### Questions :

4. Veuillez répondre à cette question pour CHACUNE de vos variables nominales.
  - a) Quel était le nombre de personnes dans chaque catégorie?
  
5. Pourquoi fallait-il réaliser deux différents types d'analyse (c.-à-d., statistiques descriptives et effectifs) pour mieux comprendre les données? N.B. : toutes les analyses de type « compréhension de vos données » sont appelées *statistiques descriptives* même si SPSS leur donne des noms différents.
  
6. Pourquoi serait-il inapproprié de calculer une moyenne pour chaque variable?

Maintenant que vous avez travaillé avec votre propre ensemble de données (qui était un échantillon de l'ensemble de données de l'Alberta), vous allez devoir apprendre à travailler avec de plus vastes ensembles de données. Veuillez télécharger le fichier « Données du PPCE 2013\_Alberta » pour toutes les questions subséquentes de ce laboratoire. Cet ensemble de données comprend tous les élèves de l'Alberta qui ont participé à l'évaluation du PPCE 2013. L'ensemble de données de l'Alberta sera utilisé pour les activités du laboratoire liées à SPSS, alors que l'ensemble de données pour l'ensemble du Canada sera utilisé pour les tâches d'apprentissage n<sup>os</sup> 3 et 4.

### 3. Compréhension des variables dans l'ensemble de données

Avant de commencer toute analyse, il est important de comprendre votre ensemble de données. Veuillez télécharger et ouvrir l'ensemble de données intitulé « Données du PPCE 2013\_Alberta », puis sélectionnez l'onglet « Affichage des variables » en bas, à gauche.



Nom	Type	Largeur	Décimales	Etiquette
1 STUDENT_I	Numérique	5	0	STUDENT_ID...
2 TEACHER_I	Numérique	5	0	TEACHER_ID...
3 SCHOOL_I	Numérique	4	0	SCHOOL_ID_E...
4 PROVINCE	Chaîne	6	0	PROVINCE
5 POPULATION	Chaîne	9	0	POPULATION
6 LANGUAG...	Numérique	3	0	LANGUAGE_L...
7 GENDER_S	Chaîne	3	0	GENDER_SEXE
8 BOOKLET_I	Chaîne	3	0	BOOKLET_ID...
9 STUDENT...	Numérique	6	2	STUDENT_WEI...
10 SCHOOL...	Numérique	12	0	SCHOOL_WEI...
11 SCIENCES	Numérique	6	2	SCIENCES
12 D_LS	Numérique	6	2	D_LS
13 D_PS	Numérique	6	2	D_PS
14 D_ES	Numérique	6	2	D_ES
15 D_NS	Numérique	6	2	D_NS
16 C_SI	Numérique	6	2	C_SI
17 C_PS	Numérique	6	2	C_PS
18 C_SR	Numérique	6	2	C_SR
19 MATH	Numérique	6	2	MATH
20 READING...	Numérique	6	2	READING_LEC...
21 PROFICIEN...	Chaîne	3	0	PROFICIENCY...
22 SQ02A	Numérique	2	0	SQ02A
23 SQ02B	Numérique	4	0	SQ02B
24 SQ03	Numérique	1	0	SQ03
25 SQ04	Numérique	1	0	SQ04
26 SQ05	Numérique	1	0	SQ05
27 SQ06A	Numérique	1	0	SQ06A
28 SQ06B	Numérique	1	0	SQ06B
29 SQ06C	Numérique	1	0	SQ06C
30 SQ06D	Numérique	1	0	SQ06D
31 SQ06E	Numérique	1	0	SQ06E
32 SQ06F	Numérique	1	0	SQ06F
33 SQ07	Numérique	1	0	SQ07
34 SQ08	Numérique	1	0	SQ08
35 SQ09	Numérique	1	0	SQ09
36 SQ10	Numérique	1	0	SQ10
37 SQ11	Numérique	1	0	SQ11
38 SQ12A	Numérique	1	0	SQ12A
39 SQ12B	Numérique	1	0	SQ12B
40 SQ12C	Numérique	1	0	SQ12C
41 SQ13	Numérique	1	0	SQ13
42 SQ14	Numérique	1	0	SQ14
43 SQ15	Numérique	1	0	SQ15
44 SQ16	Numérique	1	0	SQ16
45 SQ17	Numérique	1	0	SQ17
46 SQ18	Numérique	1	0	SQ18
47 SQ19	Numérique	1	0	SQ19
48 SQ20	Numérique	1	0	SQ20
49 SQ21	Numérique	1	0	SQ21
50 SQ22	Numérique	1	0	SQ22
51 SQ23	Numérique	1	0	SQ23
52 SQ24	Numérique	1	0	SQ24
53 SQ25	Numérique	1	0	SQ25
54 SQ26	Numérique	1	0	SQ26
55 SQ27	Numérique	1	0	SQ27
56 SQ28	Numérique	1	0	SQ28
57 SQ29	Numérique	1	0	SQ29
58 SQ30	Numérique	1	0	SQ30
59 SQ31	Numérique	1	0	SQ31
60 SQ32	Numérique	1	0	SQ32
61 SQ33	Numérique	1	0	SQ33
62 SQ34	Numérique	1	0	SQ34
63 SQ35	Numérique	1	0	SQ35
64 SQ36	Numérique	1	0	SQ36
65 SQ37	Numérique	1	0	SQ37
66 SQ38	Numérique	1	0	SQ38
67 SQ39	Numérique	1	0	SQ39
68 SQ40	Numérique	1	0	SQ40
69 SQ41	Numérique	1	0	SQ41
70 SQ42	Numérique	1	0	SQ42
71 SQ43	Numérique	1	0	SQ43
72 SQ44	Numérique	1	0	SQ44
73 SQ45	Numérique	1	0	SQ45
74 SQ46	Numérique	1	0	SQ46
75 SQ47	Numérique	1	0	SQ47
76 SQ48	Numérique	1	0	SQ48
77 SQ49	Numérique	1	0	SQ49
78 SQ50	Numérique	1	0	SQ50
79 SQ51	Numérique	1	0	SQ51
80 SQ52	Numérique	1	0	SQ52
81 SQ53	Numérique	1	0	SQ53
82 SQ54	Numérique	1	0	SQ54
83 SQ55	Numérique	1	0	SQ55
84 SQ56	Numérique	1	0	SQ56
85 SQ57	Numérique	1	0	SQ57
86 SQ58	Numérique	1	0	SQ58
87 SQ59	Numérique	1	0	SQ59
88 SQ60	Numérique	1	0	SQ60
89 SQ61	Numérique	1	0	SQ61
90 SQ62	Numérique	1	0	SQ62
91 SQ63	Numérique	1	0	SQ63
92 SQ64	Numérique	1	0	SQ64
93 SQ65	Numérique	1	0	SQ65
94 SQ66	Numérique	1	0	SQ66
95 SQ67	Numérique	1	0	SQ67
96 SQ68	Numérique	1	0	SQ68
97 SQ69	Numérique	1	0	SQ69
98 SQ70	Numérique	1	0	SQ70
99 SQ71	Numérique	1	0	SQ71
100 SQ72	Numérique	1	0	SQ72
101 SQ73	Numérique	1	0	SQ73
102 SQ74	Numérique	1	0	SQ74
103 SQ75	Numérique	1	0	SQ75
104 SQ76	Numérique	1	0	SQ76
105 SQ77	Numérique	1	0	SQ77
106 SQ78	Numérique	1	0	SQ78
107 SQ79	Numérique	1	0	SQ79
108 SQ80	Numérique	1	0	SQ80
109 SQ81	Numérique	1	0	SQ81
110 SQ82	Numérique	1	0	SQ82
111 SQ83	Numérique	1	0	SQ83
112 SQ84	Numérique	1	0	SQ84
113 SQ85	Numérique	1	0	SQ85
114 SQ86	Numérique	1	0	SQ86
115 SQ87	Numérique	1	0	SQ87
116 SQ88	Numérique	1	0	SQ88
117 SQ89	Numérique	1	0	SQ89
118 SQ90	Numérique	1	0	SQ90
119 SQ91	Numérique	1	0	SQ91
120 SQ92	Numérique	1	0	SQ92
121 SQ93	Numérique	1	0	SQ93
122 SQ94	Numérique	1	0	SQ94
123 SQ95	Numérique	1	0	SQ95
124 SQ96	Numérique	1	0	SQ96
125 SQ97	Numérique	1	0	SQ97
126 SQ98	Numérique	1	0	SQ98
127 SQ99	Numérique	1	0	SQ99
128 SQ100	Numérique	1	0	SQ100

Regarder le *Nom* des variables n'est souvent pas très utile, car les noms sont normalement abrégés. C'est pourquoi nous utilisons souvent un manuel de codage pour déchiffrer toutes les différentes variables dans un ensemble de données. Veuillez consulter le fichier « Manuel de codage fusionné du PPCE 2013 » pour mieux comprendre les variables recueillies et leur nom dans votre fichier SPSS.

### Questions :

7. Veuillez indiquer ce que chacun de ces noms de variables SPSS signifie :
  - a) STUDENT\_WEIGHT\_PONDERATION\_ELEVE
  - b) SCIENCES
  - c) C\_SI
  - d) SQ13A
  - e) SQ26B
  - f) SQ31B
8. Pour un grand nombre des items du questionnaire (p. ex., SQ03 et SQ04), les réponses des élèves ont été saisies comme « -9 » ou « -8 » dans SPSS. Que signifient ces valeurs?
9. Pourquoi est-il important d'utiliser une valeur pour représenter une réponse manquante au lieu de laisser la cellule vide?

### 4. Pondération des personnes participantes

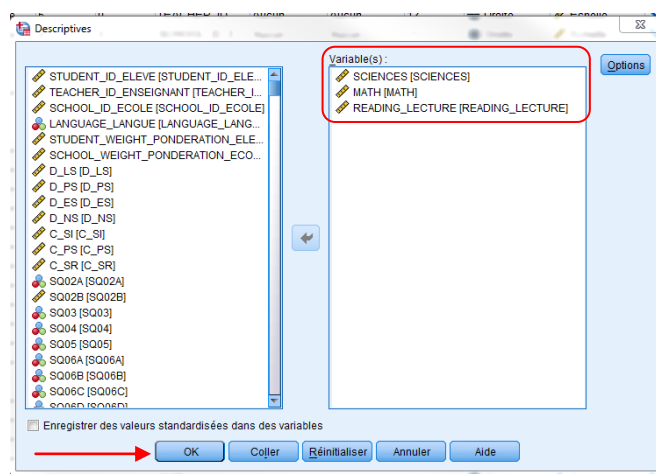
Une variable *Pondération des élèves* a été créée pour les données du PPCE 2013. Il est important d'utiliser cette variable tout au long du laboratoire.

**Questions :**

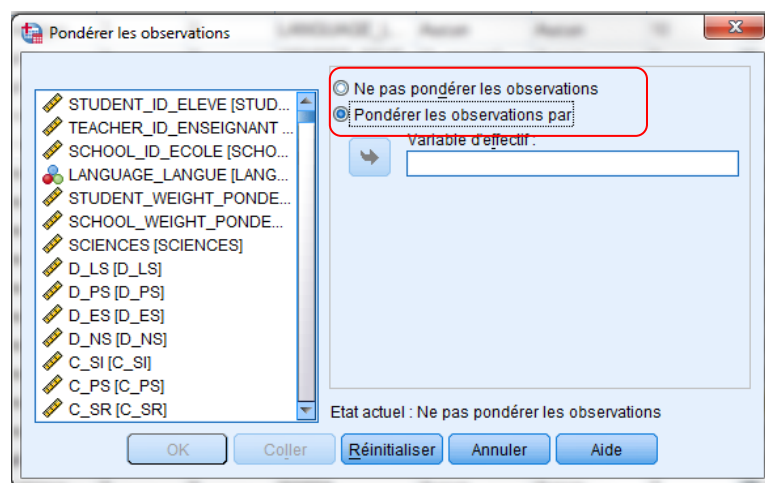
**10.** Quelle est l'utilité d'une variable de pondération?

**11.** Veuillez calculer les scores moyens des élèves en sciences, en mathématiques et en lecture :

- a) en n'utilisant aucune pondération des élèves. Procédez comme suit : Analyse → Statistiques descriptives → Descriptives → sélectionnez SCIENCES, MATH et READING\_LECTURE et déplacez-les vers la boîte « Variable(s) » → sélectionnez OK.



- b) en utilisant les pondérations des élèves. Premièrement, vous devez incorporer la pondération des élèves à votre ensemble de données. Procédez comme suit : Données → Pondérer les observations → sélectionnez Pondérer les observations par → sélectionnez STUDENT\_WEIGHT\_PONDERATION\_ELEVE et déplacez-la vers la boîte « Variable d'effectif » → OK.



Maintenant, réalisez la même analyse descriptive qu'à la section

« **3. Compréhension des variables dans l'ensemble de données** ».

- c) Veuillez remplir le tableau suivant (c.-à-d., remplir chaque champ « ?? ») avec les statistiques descriptives calculées ci-dessus :

Tableau 1

*Comparaison des scores de rendement pour le PPCE en utilisant les pondérations des élèves*

	Aucune pondération des élèves utilisée ( $n=??$ )		Pondérations des élèves utilisées ( $n=??$ )	
	Moyenne	Écart type	Moyenne	Écart type
Sciences	??	??	??	??
Mathématiques	??	??	??	??
Lecture	??	??	??	??

- d) Quelles sont les principales différences entre l'utilisation ou non des pondérations des élèves?

**12.** Pourquoi est-il important de s'assurer que cette variable de pondération des élèves est utilisée lors de toutes nos analyses des données du PPCE?

**13.** Veuillez comparer la moyenne des scores en sciences calculée à partir de l'ensemble de données saisi (c.-à-d., laboratoire 1 pour SPSS) et l'ensemble de données de l'Alberta (en utilisant les pondérations). Quelle est la différence? Pourquoi existe-t-il une différence entre ces deux valeurs?

Vous avez maintenant terminé le laboratoire.