

TP 1 : Correction

Manipulations de base dans SPSS

Ceci est un corrigé type. D'autres procédures peuvent être utilisées pour obtenir des résultats semblables.

1. Faire une analyse descriptive (résumés numériques et graphiques) des variables *Poids, Taille, Sexe et Fumeur*.

La fonction FREQUENCIES (Analyse -> Statistiques_descriptives -> Effectifs) permet d'obtenir à la fois des résumés numériques, des graphiques et des tableaux de fréquences. On l'utilise deux fois, tout d'abord pour les deux variables numériques (Poids, Taille) en demandant des résumés numériques, des histogrammes, mais pas de tableaux de fréquences (trop de modalités), puis pour les variables catégorielles (Sexe, Fumeur) en demandant les tableaux de fréquence, des graphiques en bâtons et seulement le mode comme résumé numérique.

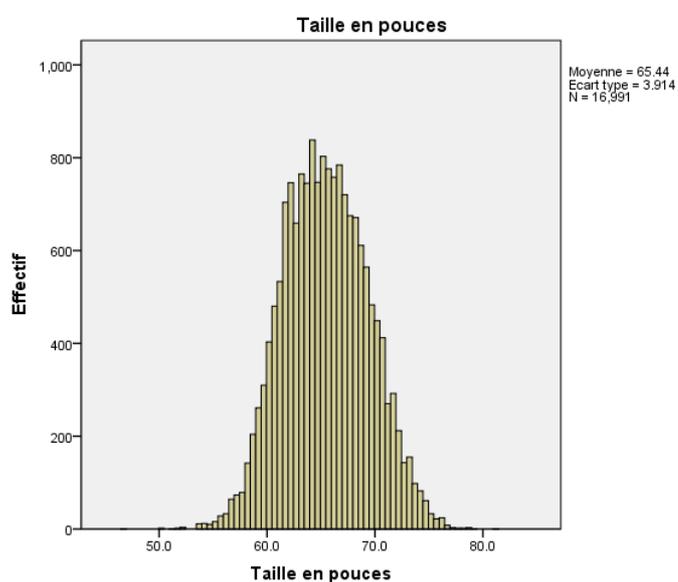
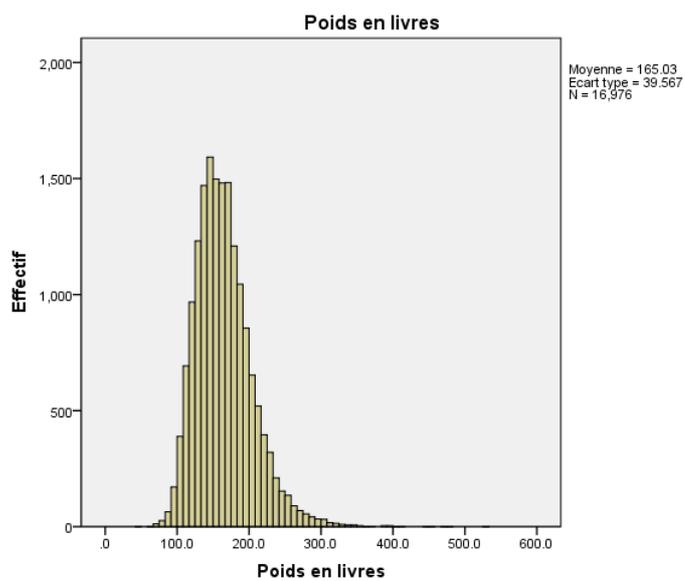
Poids et Taille :

		Statistiques	
		Poids en livres	Taille en pouces
N	Valide	16976	16991
	Manquante	54	39
Moyenne		165.030	65.438
Médiane		160.450	65.300
Ecart-type		39.5675	3.9139
Variance		1565.586	15.318
Minimum		48.0	46.7
Maximum		532.0	81.3
Centiles	25	137.400	62.600
	50	160.450	65.300
	75	186.500	68.200

Il y a quelques données manquantes pour chaque variable (54 et 39).

Le poids est distribué de manière légèrement asymétrique avec un étalement vers les grandes valeurs, alors que la taille est beaucoup plus symétrique. On peut voir cela soit sur les histogrammes, soit en comparant dans chaque cas la moyenne avec la médiane (moyenne = médiane -> symétrie, ...).

Les deux variables n'étant pas dans la même unité, il n'est pas possible de comparer directement leurs résumés numériques.



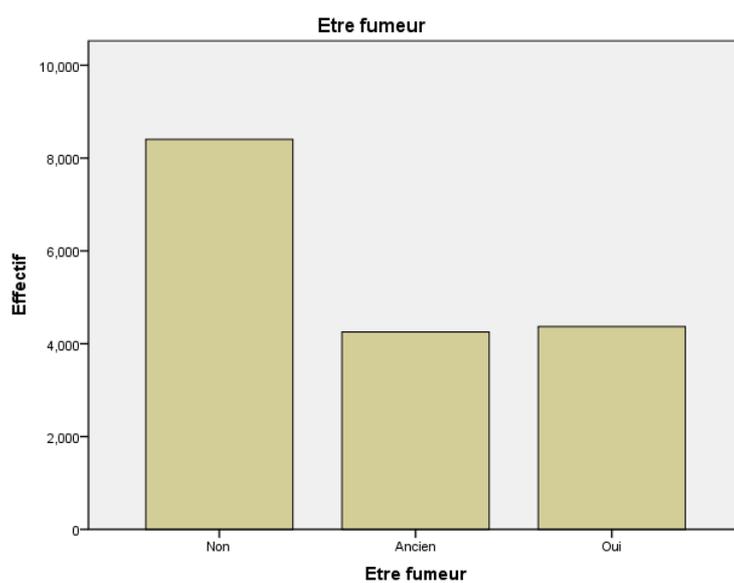
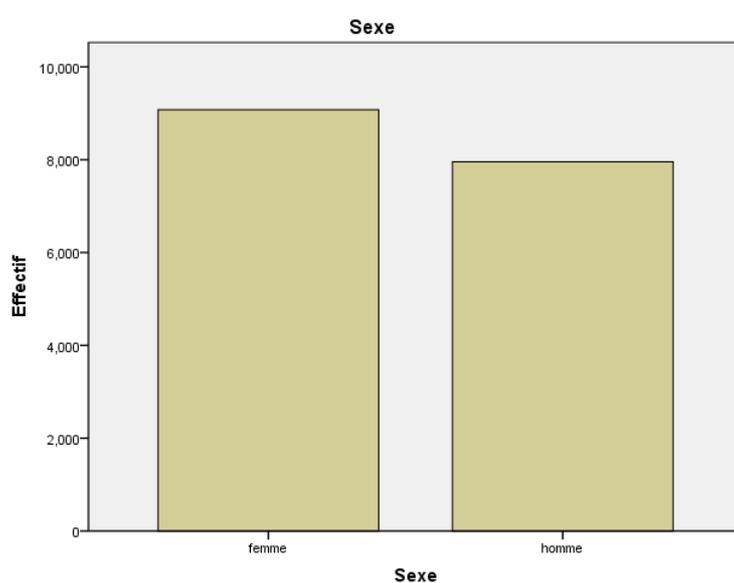
Sexe et Fumeur :

		Sexe			
		Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	femme	9077	53.3	53.3	53.3
	homme	7953	46.7	46.7	100.0
Total		17030	100.0	100.0	

Il y a un peu plus de femmes que d'hommes (53.3% contre 46.7%).

		Etre fumeur			
		Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Non	8403	49.3	49.3	49.3
	Ancien	4255	25.0	25.0	74.3
	Oui	4371	25.7	25.7	100.0
	Total	17029	100.0	100.0	
Manquante	-99	1	.0		
Total		17030	100.0		

Il y a une donnée manquante. La modalité la plus fréquente est « non-fumeur » (49.3%), suivie de « fumeur » (25.7%) et « ancien fumeur » (25.0%).



2. Refaire l'analyse des variables *Poids, Taille et Fumeur* en effectuant les calculs séparément pour les femmes et les hommes et en comparant les résultats obtenus.

On utilise la fonction EXAMINE (Analyse -> Statistiques_descriptives -> Explorer) en utilisant la variable Sexe comme facteur et en sélectionnant à la fois des résumés numériques descriptifs et un graphique de type boxplot (boîte à moustaches).

Récapitulatif du traitement des observations

		Observations					
		Valide		Manquante		Total	
		N	Pourcent	N	Pourcent	N	Pourcent
Poids en livres	femme	9036	99.5%	41	.5%	9077	100.0%
	homme	7932	99.7%	21	.3%	7953	100.0%
Taille en pouces	femme	9036	99.5%	41	.5%	9077	100.0%
	homme	7932	99.7%	21	.3%	7953	100.0%
Etre fumeur	femme	9036	99.5%	41	.5%	9077	100.0%
	homme	7932	99.7%	21	.3%	7953	100.0%

Descriptives

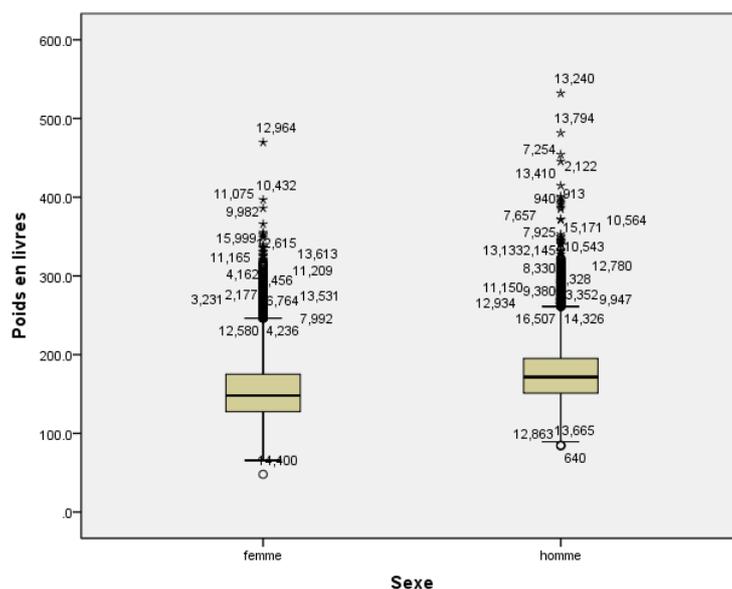
Sexe			Statistique	Erreur standard		
Poids en livres	femme	Moyenne	155.220	.4137		
		Intervalle de confiance à 95% pour la moyenne	Borne inférieure		154.409	
			Borne supérieure		156.031	
		Moyenne tronquée à 5%			152.626	
		Médiane			148.000	
		Variance			1546.747	
		Ecart-type			39.3287	
		Minimum			48.0	
		Maximum			469.7	
		Intervalle			421.7	
		Intervalle interquartile			47.5	
		Asymétrie			1.171	.026
		Aplatissement			2.401	.052
		Poids en livres	homme		Moyenne	176.208
Intervalle de confiance à 95% pour la moyenne	Borne inférieure			175.399		
	Borne supérieure			177.017		
Moyenne tronquée à 5%				174.082		
Médiane				171.600		
Variance				1351.325		
Ecart-type				36.7604		

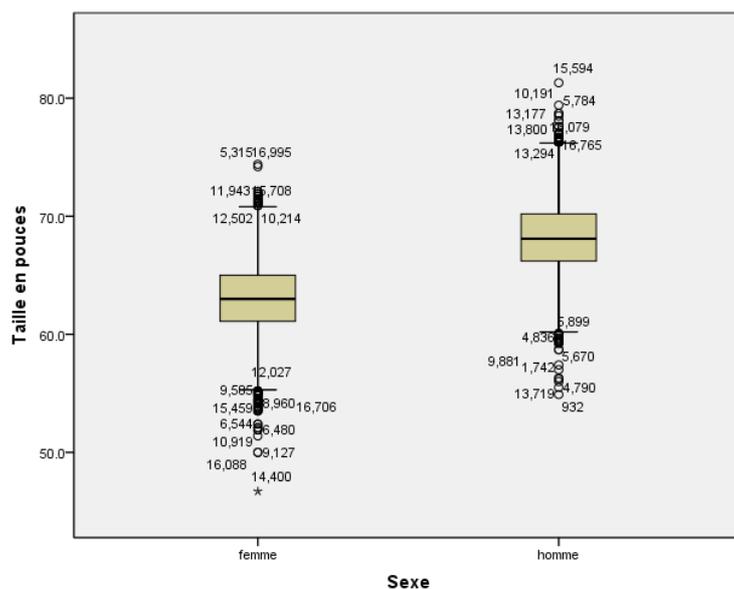
		Minimum		84.3	
		Maximum		532.0	
		Intervalle		447.7	
		Intervalle interquartile		44.0	
		Asymétrie		1.310	.027
		Aplatissement		4.928	.055
Taille en pouces	femme	Moyenne		63.009	.0301
		Intervalle de confiance à 95% pour la moyenne	Borne inférieure	62.950	
			Borne supérieure	63.068	
		Moyenne tronquée à 5%		63.023	
		Médiane		63.000	
		Variance		8.185	
		Ecart-type		2.8610	
		Minimum		46.7	
		Maximum		74.4	
		Intervalle		27.7	
		Intervalle interquartile		3.9	
		Asymétrie		-.089	.026
		Aplatissement		.315	.052
			homme	Moyenne	
Intervalle de confiance à 95% pour la moyenne	Borne inférieure			68.151	
	Borne supérieure			68.283	
Moyenne tronquée à 5%				68.210	
Médiane				68.100	
Variance				8.949	
Ecart-type				2.9915	
Minimum				54.9	
Maximum				81.3	
Intervalle				26.4	
Intervalle interquartile				4.0	
Asymétrie				.029	.027
Aplatissement				.130	.055
Etre fumeur	femme			Moyenne	
		Intervalle de confiance à 95% pour la moyenne	Borne inférieure	1.58	
			Borne supérieure	1.62	
		Moyenne tronquée à 5%		1.55	
		Médiane		1.00	
		Variance		.667	
		Ecart-type		.816	
		Minimum		1	

	Maximum		3	
	Intervalle		2	
	Intervalle interquartile		1	
	Asymétrie		.856	.026
	Aplatissement		-.963	.052
homme	Moyenne		1.95	.009
	Intervalle de confiance à 95% pour la moyenne	Borne inférieure	1.93	
		Borne supérieure	1.97	
	Moyenne tronquée à 5%		1.95	
	Médiane		2.00	
	Variance		.660	
	Ecart-type		.812	
	Minimum		1	
	Maximum		3	
	Intervalle		2	
	Intervalle interquartile		2	
	Asymétrie		.088	.027
	Aplatissement		-1.480	.055

Les hommes ont tendance à être un peu plus lourds et un peu plus grands que les femmes. En revanche, il y a moins de différences au niveau de la dispersion pour ces deux variables. Sur les boxplots, les données représentées par des points sont des données extrêmes (beaucoup plus grandes ou petites que les autres observations).

Attention, la variable « Fumeur » étant ordinale et non-numérique, la plupart des résultats donnés par SPSS ne font pas de sens et ne doivent pas être considérés ! Il faudrait refaire des analyses séparées par sexe en utilisant la fonction FREQUENCIES pour obtenir quelque chose d'utile !





3. Créer une nouvelle variable *Poids2* contenant le poids en kilos et une nouvelle variable *Taille2* contenant la taille en mètres.

On utilise dans les deux cas la fonction COMPUTE (Transformer -> Calculer la variable). Pour transformer le poids de livres en kilos, il faut multiplier le poids en livres par 0.45359 :

```
COMPUTE Poids2=0.45359 * Poids.
VARIABLE LABELS Poids2 'Poids en kilos'.
EXECUTE.
```

Pour transformer la taille de pouces en mètres, il faut multiplier la taille en pouces par 0.0254 :

```
COMPUTE Taille2=0.0254 * Taille.
VARIABLE LABELS Taille2 'Taille en mètres'.
EXECUTE.
```

4. Calculer l'indice de masse corporelle à partir de *Poids2* et *Taille2*, puis analyser la distribution du résultat.

On utilise à nouveau la fonction COMPUTE pour calculer l'IMC :

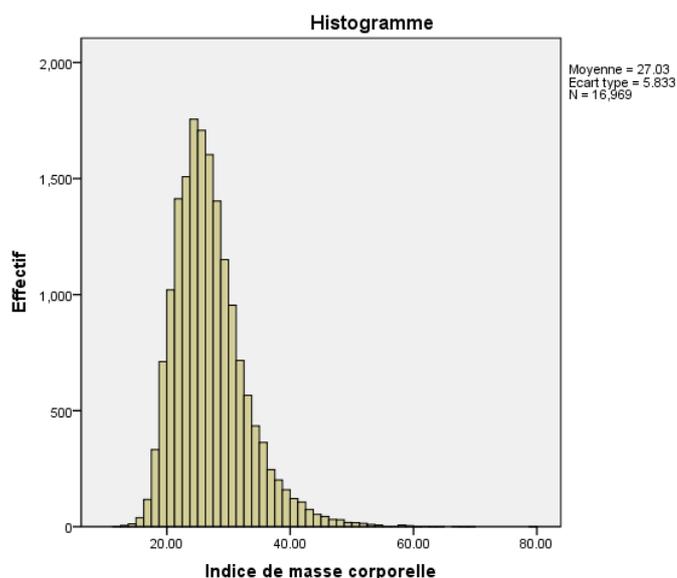
```
COMPUTE IMC=Poids2/(Taille2 ** 2).
VARIABLE LABELS IMC 'Indice de masse corporelle'.
EXECUTE.
```

On utilise ensuite la fonction FREQUENCIES pour obtenir une analyse descriptive de l'IMC, à nouveau sans table de fréquences en raison du grand nombre de valeurs différentes possibles :

Statistiques

Indice de masse corporelle

N	Valide	16969
	Manquante	61
Moyenne		27.0330
Médiane		26.1377
Ecart-type		5.83345
Variance		34.029
Minimum		11.68
Maximum		79.38
Centiles	25	22.9974
	50	26.1377
	75	29.9419



L'IMC est asymétrique avec un étalement sur la droite. Il vaut 27.03 en moyenne pour un intervalle s'étendant de 11.68 à 79.38 (valeur extrêmement élevée).