Indices de poids et taille dans les études morphologiques

Nabila MIMOUNI[1] ; Jacques PONTIER[2]

1. i.n.f.s./ s.t.s. Dely- brahim ; Cité Les Vergers, Bt C2, n° 57, Birkhadem, Alger.

2. Université Claude Bernard Lyon 1 ; 40 rue Edouard Millaud F 69230 Saint Genis Laval

Résumé

Poids et taille, caractères facilement mesurables, sont à la base de beaucoup d'études morphologiques humaines. Leur coefficient de corrélation non négligeable exprime leur redondance naturelle. Celle-ci perturbe les analyses statistiques utilisant ces deux caractères. Divers indices (Livi, Kaup, Schreider, etc.) relativisant le poids par rapport à la taille essaient d'éviter ce phénomène.

Constatant les liaisons entre ces caractères et indices, nous voulons détecter ceux qui restituent l'information totale véhiculée par le couple poids et taille, tout en éliminant la redondance. Notre démarche s'appuie sur des méthodes statistiques appropriées classiques (analyse en composantes principales, régression linéaire multiple).

Observant de jeunes adultes algériens, sportifs, nous montrons que taille et indice de Kaup, pris ensemble, répondent à la question : ils sont sans corrélation entre eux, et contiennent la totalité de l'information initiale. De plus, taille et indice de Kaup permettent le découpage du poids en somme de deux composantes indépendantes : l'une, dépendant uniquement de la stature, est le "poids statural", l'autre, liée à la notion de corpulence, est l'"écart pondéral" par rapport au poids statural, écart qui peut être positif (traduisant une surcharge pondérale) ou négatif (déficit pondéral).

Au passage, notre étude confirme que poids et taille sont insuffisants pour discriminer entre eux des sportifs adultes d'activités sportives différentes.

Toutes nos conclusions qualitatives sont valables pour les deux sexes. Par contre il existe un dimorphisme sexuel des valeurs numériques des coefficients, dans les formules.

Introduction

Le poids et la taille sont les deux caractères "de base", ceux qui sont mesurés en priorité sur une personne dont on veut obtenir une image quantitative de la morphologie globale. Leur mesure ne nécessite qu'un matériel réduit : un pèse personne, précis à l'hectogramme, et un ruban à mesurer, précis au millimètre, suffisent en général. Certes, la connaissance des valeurs numériques de ces deux caractères seuls ne donne qu'une idée bien incomplète de cette morphologie, et toute étude plus fouillée se doit de les compléter par d'autres mensurations, dont le choix dépend des objectifs de l'étude.

Précisément, dans ces études plus complètes, où poids et taille sont inclus parmi d'autres mensurations, les analyses statistiques que l'on souhaite réaliser sont perturbées par le fait que ces deux caractères sont liés entre eux de manière assez importante. Un coefficient de corrélation de l'ordre de 0,6 à 0,7, indique que près de 50 % de la variabilité inter–individuelle du poids est expliquée par la seule variabilité inter–individuelle de la taille. C'est là une manifestation de "l'effet taille", qui veut que, biologiquement, les individus de grande taille ont tendance à être plus pesants que les individus de petite taille : simple question de volume corporel, lui-même fonction directe de la taille. Mais cet effet taille n'explique pas tout : à taille égale, il existe une variabilité importante entre individus, lesquels sont plus ou moins gros ou maigres (par rapport à une silhouette "moyenne"), variabilité qui introduit la notion de "corpulence". Divers indices sont utilisés dans le but de mesurer cette corpulence, chacun ayant ses interprétations et ses mérites. Ces indices sont des rapports d'une fonction du poids sur une fonction de la taille : généralement une puissance du poids sur une puissance de la taille :

$$indice = k \infty \, P^{\alpha} / T^{\beta}$$

(k est un coefficient dépendant des unités choisies, α et β sont des exposants de valeurs numériques assez simples : voir plus loin).

Dans une étude multivariée de la morphologie humaine (par exemple, dans le domaine sportif, un objectif d'une telle étude peut être la recherche d'une morphologie différentielle entre pratiquants de disciplines différentes), se posent dès le départ les questions suivantes :

- le poids et la taille étant simultanément pris en compte dans l'étude, comment dissocier, pour le poids, la part due à l'effet taille et la part due à la corpulence ?
- parmi les indices proposés pour réduire cet effet taille, le(s) quel(s) faut-il choisir de préférence ?
- la part due à l'effet taille et la part due à la corpulence sont-elles indépendantes ?

Dans cet article nous nous proposons d'apporter des réponses précises à ces questions.

Matériel et méthodes

Les sujets sur lesquels nous avons travaillé sont des sportifs algériens, hommes et femmes, jeunes adultes (âges échelonnés de 20 à 30 ans chez les femmes, 18 à 35 ans chez les hommes). Tous, membres d'équipes nationales, pratiquent la compétition de haut niveau (compétitions nationales et internationales). Au total, huit disciplines sportives sont représentées, dont quatre dans des sports collectifs (basket-ball, football, handball, volley-ball), et quatre dans des sports individuels (athlétisme, gymnastique, judo, natation).

La répartition entre les disciplines, des 57 femmes et des 327 hommes examinés, est la suivante :

	Athl	Gymn	Judo	Nata	Bask	Foot	Hand	Voll	Totaux
♀	8	8	6	10	6	0	14	5	57
♂	40	7	27	47	15	66	87	38	327
Tot.	48	15	33	57	21	66	101	43	384

La collecte des données s'est étalée sur une période de onze années (1983-1993), dans le cadre de l'enseignement du cours de morphologie sportive de l'Institut Supérieur des Sciences et de la Technologie du Sport à Alger. Les mesures ont toujours été réalisées dans la matinée, dans les locaux du laboratoire de morphologie de l'Institut, la température étant toujours maintenue entre 20° et 22°C.

Pour chaque individu examiné, nous avons noté les valeurs de nombreux caractères : âge, poids, taille, hauteurs de 9 points anthropométriques, 10 diamètres ou largeurs, 11 périmètres ou circonférences, 9 plis cutanés. Ces mesures ont été partiellement exploitées dans une thèse de Doctorat soutenue par l'un de nous (Mimouni 1996), et le présent travail est un complément de cette étude antérieure. Comme expliqué plus haut en Introduction, nous nous préoccupons ici exclusivement de quelques questions relatives à l'utilisation conjointe du poids et de la taille, et d'indices rapportant le poids à la taille.
Sur le plan des méthodes statistiques que nous utiliserons, nous distinguerons deux catégories de variables :

- les "variables actives", celles sur lesquelles seront réalisés les calculs, et qui sont au nombre de sept :
  - le poids P, en kilogrammes, avec une éventuelle décimale (hectogrammes)
  - la taille T, en centimètres, avec une éventuelle décimale (millimètres)
  - cinq indices de poids et taille choisis parmi les plus connus : Livi, Quételet, Kaup, Röhrer, Schreider (voir plus loin quelques détails)
- les "variables supplémentaires", au nombre de neuf : l'âge des sujets (variable 1), et les indicatrices des huit disciplines sportives (variables 9 à 16). Ces variables sont susceptibles d'apporter des informations supplémentaires, par exemple pour aider à l'interprétation des résultats obtenus à partir des variables actives.

Revenons sur les cinq indices choisis, dans les variables actives. Les définitions que nous avons utilisées sont les suivantes (compte tenu du poids P en kilogrammes, de la taille T en centimètres) :

indice pondéral de Livi : $L = 10^3 \times \frac{\sqrt[3]{P}}{T}$

indice de Quételet inversé : $Q = 10^2 \times \frac{P}{T}$

indice de corpulence de Kaup : $K = 10^4 \times \frac{P}{T^2}$

indice de Röhrer : $R = 10^6 \times \frac{P}{T^3}$

indice de Schreider : $S = 139{,}19822 \times \frac{P^{0{,}575}}{T^{0{,}725}}$

Dans ces diverses formules, les puissances de 10 utilisées comme coefficients sont destinées à homogénéiser les ordres de grandeurs de ces divers indices. Cette opération est sans influence sur les résultats de l'A.C.P. normée (qui analyse les coefficients de corrélation linéaire), tout en évitant des désagréments dans la réalisation des calculs proprement dits.

Nous avons choisi d'inverser l'indice de Quételet (qui est le rapport de la taille sur le poids), pour le rendre homogène par rapport aux autres indices (qui sont tous des rapports d'une fonction directe du poids sur une fonction directe de la taille).

Quant à l'indice de Schreider, qui est simplement le rapport du poids sur la surface du corps, nous l'avons exprimé en fonction du poids et de la taille, utilisant dans ce but l'expression de la surface du corps (dont la mesure directe est difficile) par la formule des Du Bois :

$$\text{Surface} = 0{,}007184 \infty P^{0{,}425} \infty T^{0{,}725}.$$

Nous avons traité séparément l'échantillon féminin et l'échantillon masculin. Dans le premier cas, l'ensemble des 57 femmes a été pris comme "échantillon actif", c'est-à-dire que la totalité des sujets a servi aux calculs. Par contre, dans le cas masculin, compte tenu de l'effectif relativement important (327 sujets), nous avons préféré scinder l'ensemble du groupe en deux sous-ensembles d'effectifs pratiquement égaux. L'un (d'effectif 164) nous a servi d'échantillon actif, c'est sur lui qu'a été réalisée l'analyse proprement dite, selon la même démarche que pour l'échantillon féminin. L'autre (d'effectif 163), échantillon supplémentaire, nous a servi à contrôler les résultats obtenus à partir de l'échantillon actif.

La démarche utilisée pour analyser le comportement des variables actives, tant dans l'échantillon actif féminin que dans l'échantillon actif masculin, consistait :

1) à explorer l'ensemble des données, par l'analyse en composantes principales normée (A.C.P.N.), méthode destinée à mettre en évidence d'une part les éventuelles structures de l'ensemble des variables actives, d'autre part les éventuelles structures de l'ensemble des individus actifs.

2) à exploiter les résultats apportés par cette A.C.P.N., pour répondre aux questions que nous nous posons, et notamment, comme on le verra, à utiliser la méthode de régression linéaire multiple pour parvenir à dissocier, dans la valeur numérique du poids, la part due à l'effet taille et la part supplémentaire, positive ou négative.

Nous donnerons ci-après le maximum de détails sur le développement de la méthode, dans l'exposé du cas féminin. La démarche étant identique dans le cas masculin, nous nous contenterons alors d'en indiquer seulement les principaux résultats.

Résultats

1. Cas féminin

L'exploration des données (57 sujets), par l'analyse en composantes principales normée, indique que, sur les sept composantes, la première rend compte de 68,35 % de la variabilité totale, et la deuxième de 31,59 %, soit un total de 99,94 % pour ces seules deux premières composantes. La part de variabilité attribuable aux autres composantes, soit 0,06 %, est donc négligeable. Ceci signifie que les 7 variables actives sont liées entre elles par des relations linéaires de rang 2 : l'une quelconque d'entre elles peut toujours être exprimée comme fonction linéaire d'au plus deux prises parmi les six autres.

Ceci se traduit, dans le cercle des corrélations relatif aux deux premières composantes, par le fait que les sept points représentant les variables actives sont tous disposées sur la circonférence (fig. 1). Chacune des variables actives peut donc être exprimée comme fonction linéaire de ces deux composantes, mais aussi, plus généralement, comme fonction linéaire de deux autres variables actives, *a priori* choisies arbitrairement (sauf le couple L et R, ces deux variables étant fonctions linéaires l'une de l'autre : les deux points sont confondus). Le choix le plus indiqué est celui de la taille T, et de l'indice de Kaup K. En effet, K est pratiquement une fonction linéaire de la première composante principale, T est pratiquement une fonction linéaire de la deuxième (voir fig. 1, les positions des points K et T proches d'extrémités des deux axes). T et K ont entre elles un coefficient de corrélation linéaire pratiquement nul (-0,04442 dans cet échantillon) (voir fig. 1, les rayons vecteurs K et T sont presque perpendiculaires).

Quant aux variables supplémentaires (variable 1 = âge exactement au centre, variables 9 à 16 = les disciplines sportives dans l'environnement immédiat du centre), leur position proche du centre du cercle signifie que ces variables sont pratiquement sans corrélation avec les variables actives. A ce stade, nous pouvons prévoir que les résultats que nous obtiendrons dans la suite de l'exploitation des variables actives, seront sans lien avec l'âge des sujets ou avec la discipline sportive pratiquée.

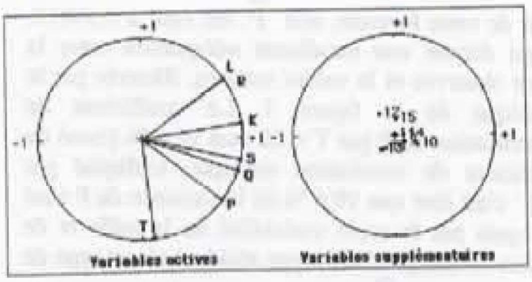

Figure 1. Cas féminin. Cercle des corrélations relatives aux deux premières composantes principales (première composante dans le sens horizontal, deuxième composante dans le sens vertical). Commentaires dans le texte.

La figure 2 montre, au moyen d'une carte factorielle, la répartition des 57 sportives dans le plan des deux premières composantes principales. La position des individus "marginaux" (c'est-à-dire les plus éloignés du centre du nuage de points) illustre bien le fait que la première composante principale coïncide avec le gradient "indice de Kaup" (les plus fortes valeurs à droite), et la deuxième composante avec le gradient "taille" (les plus fortes valeurs en bas).

Afin de permettre au lecteur de repérer ces individus, on a rassemblé dans le tableau 1 d'une part les indices de position de ces deux variables parmi les 57 sportives, d'autre part les valeurs de ces variables pour quelques uns des individus marginaux. Par exemple, l'individu n° 1, qui est le plus grand en taille (185 cm), est représenté par le point le plus vers le bas. Son indice de Kaup (21,9) est au-dessus de la moyenne, mais n'est pas le plus élevé du groupe, qui est celui du n° 45 (25,5) (représenté par le point le plus à droite). Ce dernier, sans être le plus petit, est tout de même de petite taille, ce qui se traduit graphiquement par sa position vers le haut de la carte.

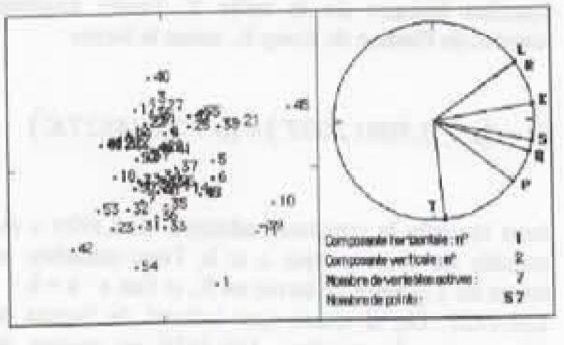

Figure 2. A gauche, représentation des 57 sportives sur la carte factorielle relative aux deux premières composantes principales. A droite, représentation des variables actives dans le cercle des corrélations (identique à celui de la fig. 1). Commentaires dans le texte.

Tableau 1. Quelques valeurs particulières du poids (P), de la taille (T) et de l'indice de Kaup (K). Les valeurs soulignées sont les valeurs extrêmes, pour chacune des trois variables. Pour les individus seulement, K est égal à $10^4 \infty P / T^2$

	Poids (kg)	Taille (cm)	Indice de Kaup
maximum	75,0	185	25,5
moyenne	58,658	167,7	20,8
minimum	50,0	155	17,3
n° 1	75,0	185	21,9
n° 10	74,0	174	24,4
n° 11	75,0	178	23,7
n° 39	75,0	177	23,9
n° 40	50,0	155	20,8
n° 42	55,0	178	17,3
n° 45	67,0	162	25,5
n° 54	64,0	182	19,3

L'A.C.P.N. a donc mis en évidence quelques propriétés importantes :

- les sept variables actives sont liées entre elles par des relations linéaires de rang 2 : chacune d'entre elles, qui est une fonction linéaire des deux premières composantes, peut également être exprimée comme fonction linéaire d'au plus deux autres variables actives ;
- les deux premières composantes principales peuvent pratiquement s'interpréter chacune comme une fonction linéaire, la première, de l'indice de Kaup K, la seconde, de la taille T ;
- ainsi, chacune des sept variables actives peut être exprimée comme fonction linéaire de T et de K, qui sont elles-mêmes sans corrélation linéaire entre elles (donc linéairement indépendantes entre elles ; cette dernière propriété entraîne, compte tenu de la quasi normalité de ces variables, qu'elles sont indépendantes entre elles).
-

Pour l'une quelconque de ces variables, soit X, son expression comme fonction linéaire de T et de K, c'est-à-dire sous la forme $X = \alpha T + \beta K + \gamma$, peut être obtenue facilement par un calcul de régression linéaire multiple de X sur les deux variables T et K. Les résultats numériques de ces régressions, pour les cinq variables actives autres que T et K, sont rassemblés dans le tableau 2.

Tableau 2. Valeurs des coefficients des deux régresseurs T et K, et de la constante de décentrage (coefficient constant), dans l'expression de chacune des variables P, L, Q, R, S comme fonction linéaire de T et de K. Valeur du coefficient de corrélation multiple correspondant (soit le coefficient de corrélation linéaire entre la vraie valeur de la variable, et sa valeur estimée par régression linéaire multiple). Valeur du coefficient de détermination (soit le carré du coefficient de corrélation multiple, multiplié par 100, qui exprime le pourcentage de la variabilité de la variable attribuable aux seuls deux régresseurs T et K). Voir dans le texte l'exemple développé de la variable P (le poids).

variables	coeff. de T	coeff. de K	coeff. constant	corr. mult.	déterm. (%)
P	0,7081209	2,884827	-120,1929	0,9981896	99,64
L	-0,04634711	0,3642660	23,33449	0,9995056	99,90
Q	0,2092228	1,695984	-35,48634	0,9993409	99,87
R	-0,07332139	0,5913302	12,41851	0,9991981	99,84
S	0,0878672	0,9685437	0,3217374	0,9994928	99,90

Ce qui nous intéresse ici, c'est l'expression du poids P en fonction de la taille T et de l'indice de corpulence K, car, ces deux derniers caractères étant indépendants entre eux, il nous paraît tout à fait intéressant de pouvoir décomposer le poids total d'une personne en deux composantes indépendantes:

- l'une, liée à la taille de manière "naturelle", le poids dépendant forcément du volume, lequel peut être "résumé", d'une certaine manière, par la taille ;

- l'autre, ne dépendant pas directement de la taille, mais étant un "plus" ou un "moins" particulier à l'individu lui-même, à son mode de vie, qui fait que l'individu est plus ou moins "maigre", plus ou moins "gros" par rapport à la moyenne.

La régression linéaire multiple du poids P sur la taille T et l'indice de corpulence K, dans le cas des 57 jeunes femmes, nous donne la relation (voir tab. 2) :

$$P \approx \hat{P} = 0{,}7081209T + 2{,}884829K - 120{,}1929$$

Le coefficient de corrélation multiple, c'est-à-dire le coefficient de corrélation linéaire entre la valeur observée du poids, soit P, et la valeur estimée à partir de cette formule, soit $\widehat{P}$, est égal à 0,99819, ce qui dénote une excellente adéquation entre la valeur observée et la valeur estimée, illustrée par le graphique de la figure 3. Le coefficient de détermination de P par T et K vaut 99,6 % (carré du coefficient de corrélation multiple, multiplié par 100) : c'est dire que 99,6 % de la variance de P sont expliqués par la seule variabilité de la taille et de l'indice de Kaup. L'écart type résiduel (écart type de la différence P - $\widehat{P}$) donne une idée de la qualité de l'approximation réalisée en substituant la valeur estimée $\widehat{P}$ à la vraie valeur P. Cet écart type résiduel vaut ici 0,370649, soit environ 0,4 kg [rappel : la moyenne de P - $\widehat{P}$ est nulle].

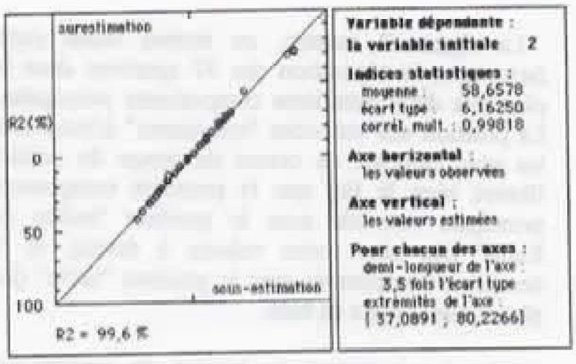

Figure 3. Représentation graphique du poids $\widehat{P}$ estimé par la formule de régression linéaire en fonction de la taille T et de l'indice de Kaup K (axe vertical), en fonction du poids P observé (axe horizontal). Le coefficient de corrélation linéaire étant très élevé (0,99818), il est normal que les 57 points individuels se trouvent pratiquement tous sur la diagonale du carré, qui figure l'égalité entre une variable et son estimée.

Cette formule linéaire permet de présenter l'estimation du poids comme somme de deux composantes indépendantes entre elles, l'une fonction linéaire de la taille T, l'autre fonction linéaire de l'indice de Kaup K, selon la forme :

$$\hat{P} = (a + 0{,}70812097T) + (b + 2{,}884827K)$$

dans laquelle la constante additive -120,1929 a été scindée en deux parties a et b, l'une rattachée au terme en T, l'autre au terme en K, et l'on a : a + b = -120,1929. Or, il existe une infinité de façons de décomposer le nombre -120,1929 en somme de deux nombres a + b. C'est dire qu'un certain arbitraire règne sur le choix de a et de b. Un choix que nous pouvons défendre ici — mais il n'est pas le seul — s'appuie sur notre objectif de décomposition du poids en une part attribuable à l'effet taille seul, et une part attribuable à la "corpulence" seule. Pour faciliter l'utilisation de cette dernière part, nous souhaiterions qu'elle s'exprime par une valeur positive si le sujet est plus "gros" que la moyenne, et par une valeur négative si le sujet est plus "maigre" que la moyenne. Rappelons que Davenport (cité par Vandervael 1980) propose un barème situant les sujet normaux dans une fourchette de valeurs allant de 21,5 à 25,6 (valeurs transcrites selon nos unités) — valeurs à comparer avec, par exemple, la moyenne de notre échantillon, soit 20,83, qui montre que nos sportives sont en moyenne plutôt maigres. La valeur 0 de la part attribuable à la corpulence, pourrait donc correspondre à une valeur intermédiaire dans la fourchette de Davenport. Choisissons (assez arbitrairement !) le milieu de cette fourchette, soit la valeur 23,5. Dans ces conditions, pour tout individu ayant cette valeur de K, l'expression b + 2,884827 K est nulle, d'où b = -67,793435, et par conséquent a = -120,1929 - b = -52,399466. Soit donc :

$$\hat{P} = (0,7081209T - 52,399466) + (2,884827K - 67,793435)$$

Cette expression de l'estimation du poids est donc la somme de deux termes :

- l'un, fonction linéaire de la taille T, est la valeur du poids résultant de l'effet taille ; on peut pour cette raison l'appeler le "poids statural" (toujours positif chez des sujets adultes normaux) ;

- l'autre, fonction linéaire de l'indice de Kaup K, est la valeur de l'écart présenté, par rapport à ce poids statural, par l'estimation du poids du sujet ; on peut l'appeler l'"écart pondéral" ; il peut être positif (on peut alors parler de "surcharge pondérale") ou négatif (alors "déficit pondéral").

Rappelons que le poids réel P du sujet ne coïncide pas forcément avec son poids estimé $\hat{P}$. La différence entre les deux est le résidu, à caractère aléatoire, dont les propriétés statistiques dans l'échantillon actif sont connues : sa moyenne est nulle, son écart type est de l'ordre de 0,4 kg. Le poids réel du sujet peut donc être écrit sous la forme:

$$P = (0,7081209T - 52,399466) + (2,884827K - 67,793435) + \text{résidu}$$

Notons que ces trois termes du second membre, sont exprimés dans l'unité de poids choisie (ici, en kilogrammes). L'écart pondéral traduit en kilogrammes l'indice de Kaup, et en particulier le barème de Davenport peut être exprimé ainsi :

	Davenport		Ecart pondéral (kg)	
Très maigre	de 14,0	à 18,0	de -27,4	à -15,7
Maigre	de 18,1	à 21,4	de -15,7	à -5,9
Moyen	de 21,5	à 25,6	de -5,9	A + 6,2
Corpulent	de 25,7	à 30,5	de +6,2	à +20,2
Obèse	plus de 30,5		plus de 20,2	

A titre d'exemple numérique, reprenons quelques-uns des sujets vus au tableau 1. Nous avons choisi les quatre sujets les plus extrêmes selon la taille ou selon l'indice de Kaup [rappelons que, dans les estimations par régression linéaire, les résidus ont tendance à augmenter avec l'éloignement par rapport à la moyenne ; notre choix nous met donc dans des conditions plutôt défavorables]. Les valeurs correspondantes de P, $\hat{P}$, poids statural, écart pondéral et résidu figurent au tableau 3. Selon le barème de Davenport, traduit en écart pondéral, nous voyons par exemple que les sujets n° 1 et 45 sont "moyens", que le n° 40 est "maigre", que le n° 42 est "très maigre". Il manquerait environ 17 à 18 kg à ce dernier pour être "moyen".

Tableau 3. Exemples. P et T sont les valeurs mesurées du poids et de la taille. K est la valeur (calculée) de l'indice de Kaup. PS et EP sont le poids statural et l'écart pondéral calculés à partir de la formule. $\hat{P}$, poids estimé par régression linéaire multiple de P sur T et K, est la somme de PS et de EP. Le résidu est la différence P - $\hat{P}$.

Sujet	P	T	K	PS	EP	$\hat{P}$	Résidu
n° 1	75,0	185	21,9	78,603	-4,576	74,027	0,973
n° 40	50,0	155	20,8	57,359	-7,755	49,604	0,396
n° 42	55,0	178	17,3	73,788	-17,828	55,959	-0,959
n° 45	67,0	162	25,5	62,316	5,855	68,171	-1,171

Pour l'utilisation pratique de la formule donnant l'estimation $\hat{P}$ de P, il n'est pas nécessaire que les coefficients conservent toutes les décimales obtenues dans nos calculs (précision numérique qui n'a de sens que dans notre échantillon). La formule à trois décimales, soit :

$$\widehat{P}_{(3)} = (0{,}708\ T - 52{,}399) + (2{,}885\ K - 67{,}793)$$

donne une moyenne résiduelle de 0,016 kg (alors que l'utilisation des six décimales donne une moyenne résiduelle nulle, conforme à la théorie ; il y a donc introduction d'un léger biais de 16 grammes en trop), avec un écart type résiduel inchangé de 0,371 kg. Par contre, avec la formule à deux décimales, soit :

$$\widehat{P}_{(2)} = (0{,}71\ T - 52{,}40) + (2{,}89\ K - 67{,}79)$$

La moyenne résiduelle est assez considérablement dégradée, passant à -0,218 kg, alors que l'écart type résiduel reste égal à 0,371. Il est donc loisible, dans la pratique, de retenir la formule à trois décimales, soit :

$$P = (0{,}708\ T - 52{,}399) + (2{,}885\ K - 67{,}793) + \text{résidu}$$

2) Cas masculin

L'analyse des données de l'échantillon masculin actif (164 sujets) aboutit à des résultats qualitativement identiques à ceux obtenus avec l'échantillon féminin. Par contre, comme on pouvait s'y attendre, les résultats quantitatifs (valeurs des coefficients dans les formules) sont un peu différents, ce qui est naturel dans une étude morphologique.

Les résultats de l'A.C.P.N. des sept variables actives sont remarquablement similaires à ceux obtenus dans le cas féminin : les seules deux premières composantes principales restituent 99,94 % de la variabilité globale des sept variables actives, ce qui indique que celles-ci présentent entre elles des liaisons linéaires de rang 2. L'indice de Kaup K et la taille T sont pratiquement des fonctions linéaires, respectivement de la première et de la deuxième composante principale, indépendante entre elles. Pour cette raison, nous choisissons T et K comme variables fondamentales, à partir desquelles nous pouvons exprimer l'une des quelconques cinq autres variables actives au moyen d'une fonction linéaire de T et de K. L'âge et les huit disciplines sportives, dont les points représentatifs sont proches du centre du cercle de corrélation, sont pratiquement sans lien avec ces deux composantes principales, donc finalement avec la taille et l'indice de Kaup.

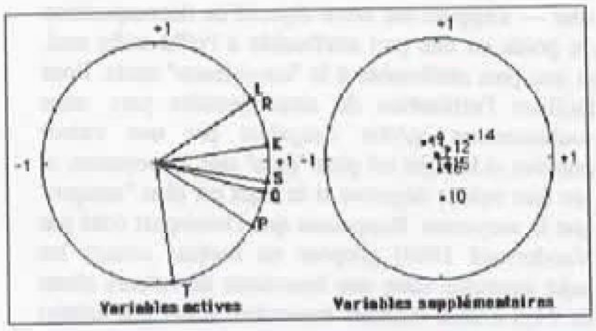

Figure 4. Cas masculin. Cercle des corrélations relatives aux deux premières composantes principales (première composante dans le sens horizontal, deuxième composante dans le sens vertical). Commentaires dans le texte.

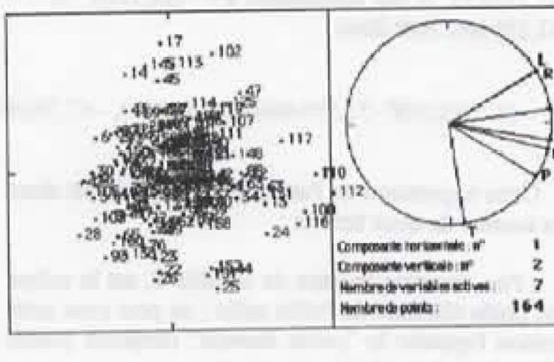

Figure 5. Cas masculin. La représentation graphique des 164 sujets dans le plan des deux premières composantes principales ne révèle aucune structure particulière liée aux variables supplémentaires (âge, disciplines sportives). Les sujets "marginaux" sont les plus marqués pour la taille (les plus grands en bas, les plus petits en haut) et/ou pour l'indice de Kaup (fortes valeurs à droite, faibles valeurs à gauche).

L'analyse ci-dessus ayant donc mis en évidence le fait que le poids P est pratiquement une fonction linéaire de la taille T et de l'indice de Kaup K, nous explicitons cette fonction linéaire grâce à une régression linéaire multiple, dont voici les résultats :

$$P \sim \widehat{P} = 0{,}8191975\ T + 3{,}213975\ K - 146{,}768$$

Le coefficient de corrélation linéaire, entre la valeur observée du poids et la valeur estimée à partir de cette formule, est égal à 0,9987379 : les deux valeurs sont pratiquement égales (voir fig.6). La moyenne résiduelle est nulle, l'écart type résiduel, égal à 0,41372, est donc de l'ordre de 0,4 kg.

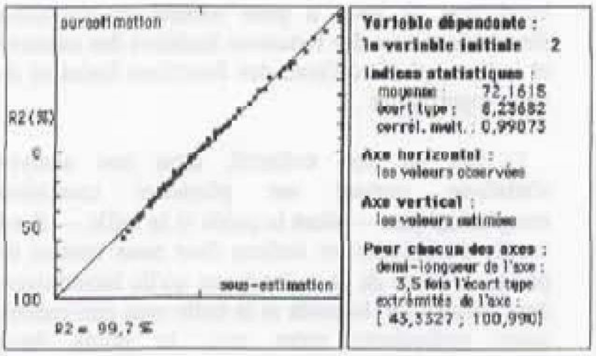

Figure 6. Cas masculin. Représentation graphique du poids $\widehat{P}$ estimé par la formule de régression linéaire sur la taille T et l'indice de Kaup K (axe vertical), en fonction du poids P observé (axe horizontal).

Nous ventilons la constante de décentrage, soit -146,768, entre le terme dépendant de la taille et celui dépendant de K, de façon à obtenir la forme :

$$\widehat{P} = (a + 0{,}8191975\ T) + (b + 3{,}213975\ K)$$

Où :

a + b = - 146,768, et en imposant la contrainte que, pour un individu dont l'indice de Kaup est moyen (valeur 23,5 selon le barème de Davenport), le terme en K soit nul. Donc :

$$b = -3{,}213975 \propto 23{,}5 = -75{,}528413 \quad \text{et}$$
$$a = -146{,}768 + b = -71{,}239588$$

D'où la formule de décomposition de l'estimation $\widehat{P}$ du poids :

$$\widehat{P} = (0{,}8191975\ T - 71{,}239588) + (3{,}213975\ K - 75{,}528413)$$

Le terme en T s'interprète comme le "poids statural", poids résultant du seul effet taille, et le terme en K s'interprète comme l'"écart pondéral" (positif ou négatif) par rapport à ce poids statural.

Les calculs réalisés à partir de formules modifiées par conservation de 3 décimales, puis de 2 décimales, ont abouti à modifier légèrement les moyennes résiduelles, soit 0,034 kg avec 3 décimales, -0,048 kg avec 2 décimales, alors que l'écart type résiduel est resté inchangé à 0,414 kg. Il est donc loisible, dans la pratique, d'utiliser la formule à 3 décimales, comme dans le cas féminin, soit :

$$P = (0{,}819\ T - 71{,}240) + (3{,}214\ K - 75{,}528) + \text{résidu}$$

3) Contrôle dans le cas masculin

Dans quelle mesure les formules d'estimation de P (à 6 décimales, à 3 décimales, à 2 décimales), obtenues ci-dessus par l'analyse de l'échantillon masculin actif, peuvent-elle être appliquées avec succès à des individus n'appartenant pas à cet échantillon ? Pour apporter un élément de réponse à cette question, nous appliquons ces formules aux 163 individus supplémentaires, dont aucun n'est intervenu dans les calculs qui précèdent. Les indices statistiques des résidus (différences entre le poids réel et l'une ou l'autre des estimations obtenues par ces formules) sont les suivants, sur l'ensemble des 163 individus supplémentaires :

	6 décimales	3 décimales	2 décimales
moyenne	- 0,024	+ 0,011	- 0,071
écart type	0,471	0,471	0,468

Aucune des moyennes n'est nulle, ce qui est normal pour un échantillon supplémentaire. L'écart type résiduel est à peu près constant, de l'ordre de 470 grammes, légèrement plus élevé que dans le cas de l'échantillon actif (de l'ordre de 420 grammes). Avec les formules à 6 ou 3 décimales, le résidu moyen reste de l'ordre de celui obtenu avec la formule à 3 décimales dans l'échantillon actif. Ces observations nous amènent à la conclusion que les formules à 6 ou à 3 décimales sont pratiquement aussi efficaces dans le cas des individus supplémentaires que dans le cas des individus actifs. Dans ces conditions, nous pouvons privilégier la formule à trois décimales, c'est-à-dire :

$$P = (0{,}819\ T - 71{,}240) + (3{,}214\ K - 75{,}528) + \text{résidu}$$

plus simple à manipuler que la formule à six décimales. La figure 7 est la représentation graphique du poids estimé par cette formule à 3 décimales, en fonction du poids réel observé. Les points représentant ces 163 sujets supplémentaires sont presque disposés en ligne droite. En fait, le coefficient de corrélation linéaire entre le poids estimé et le poids observé est ici égal à 0,99801, valeur qui traduit bien l'efficacité de la formule lorsqu'on l'applique à des individus supplémentaires.

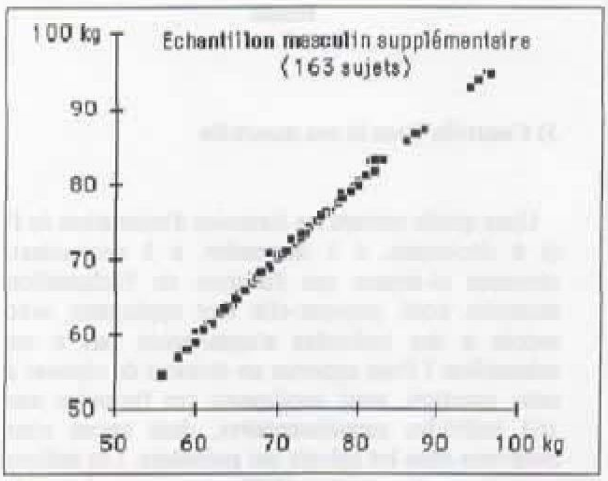

Figure 7. Cas masculin (échantillon supplémentaire).

Représentation graphique du poids $\widehat{P}$ estimé par la formule à 3 décimales de régression linéaire sur la taille T et l'indice de Kaup K (axe vertical), en fonction du poids P observé (axe horizontal).

Discussions et conclusions

Tant dans le cas féminin que dans le cas masculin, notre analyse met en évidence les faits suivants, valables au moins chez de jeunes adultes sportifs :

1) Le poids, la taille, et les cinq indices choisis pour relativiser le poids par rapport à la taille (Livi, Quételet inversé, Kaup, Röhrer, Schreider) sont très redondants entre eux : chacun peut être exprimé comme fonction linéaire d'au plus deux des six autres. A première vue, le caractère linéaire de ces fonctions peut surprendre, car les indices dont il est question sont définis par des formules non linéaires. Cependant, ces formules sont des rapports de puissances, respectivement du poids P et de la taille T. Il en découle que les logarithmes de ces indices sont des fonctions linéaires des logarithmes de P et de T. Or, dans l'intervalle de variation ordinaire, tant du poids que de la taille, la fonction logarithme est très proche d'une simple fonction linéaire, et il en est de même pour ces indices. Par acquit de conscience, nous avons vérifié ce phénomène, et réalisé aussi les calculs précédents sur les logarithmes de toutes les variables actives : les résultats, évidemment différents sur le plan numérique (coefficients dans les formules), ont été identiques sur le plan qualitatif. Dans ces conditions, il nous a paru naturel de travailler directement avec des fonctions linéaires des mesures et indices, plutôt qu'avec des fonctions linéaires de leurs logarithmes.

2) Sur un plan collectif, dans une analyse statistique portant sur plusieurs caractères morphologiques — dont le poids et la taille — il est inutile d'adjoindre les indices dont nous venons de parler, à cause de la redondance qu'ils introduisent dans l'étude. Or, le poids et la taille sont eux-mêmes assez redondants entre eux, le poids étant partiellement dépendant de la taille. Par contre, nous avons vérifié que la taille T et l'indice de Kaup K sont pratiquement indépendants, et constituent de ce fait un excellent choix pour, dans une analyse statistique, être les "représentants" (par des fonctions linéaires) non seulement du poids P, mais aussi au moins des autres indices rapportant le poids à la taille.

3) Sur le plan individuel, le poids P du sujet peut être décomposé en trois parts. L'une est le "poids statural", découlant naturellement de la taille, dont il est une fonction linéaire. Une autre part est l'"écart pondéral", lié à la corpulence du sujet, fonction linéaire de l'indice de Kaup ; cet écart est négatif si le sujet est plus maigre que la moyenne (déficit pondéral), positif si le sujet est plus gros (surcharge pondérale). Enfin, une dernière part est le "poids résiduel", différence entre le poids réel observé, et la somme des deux parts précédentes ; ce poids résiduel (positif ou négatif) est une quantité aléatoire, de faible valeur par rapport au poids total (de l'ordre de 1 %). Dans la pratique, le calcul de ces trois parts peut être réalisé à partir de formules, différentes selon le sexe du sujet. Partant du poids P (en kilogrammes) et de la taille T (en centimètres), on calcule d'abord $K = 10000 \infty P / T^2$, puis :

$$\widehat{P} = (0{,}708\ T - 52{,}399) + (2{,}885\ K - 67{,}793)$$ dans le cas féminin,

$$\widehat{P} = (0{,}819\ T - 71{,}240) + (3{,}214\ K - 75{,}528)$$ dans le cas masculin.

Dans chacune de ces deux formules, le premier terme du second membre est le poids statural, le second terme est l'écart pondéral. La somme des deux, soit $\widehat{P}$, est le "poids estimé". Enfin, on calcule le poids résiduel, soit $P - \widehat{P}$ (ceci dans le but de vérifier qu'il constitue une part très faible du poids total).

Les conclusions exprimées ci-dessus proviennent d'échantillons particuliers, ce qui est inévitable : il s'agit de jeunes adultes, algériens, hommes et femmes, sportifs, pratiquant à un haut niveau leurs disciplines sportives. Nous avons pu observer que les phénomènes observés concernant P et K sont indépendants des disciplines sportives (malgré la diversité de ces dernières), et de l'âge des sujets (malgré un étalement sur plus d'une dizaine d'années).

Si sur le plan qualitatif les phénomènes observés sont identiques chez les hommes et chez les femmes, les valeurs numériques des coefficients dans les formules sont différentes d'un sexe à l'autre. Ceci n'est évidemment pas une surprise. Mais, pour chacun des sexes pris séparément, nous pouvons légitimement nous poser la question de l'"universalité" de la formule correspondante. Dans l'état actuel de notre recherche, nous n'avons vérifié, ci-dessus, que la validité de la formule masculine, appliquée à des sujets provenant du même groupe humain que ceux de l'échantillon actif (jeunes adultes, algériens, sportifs, pratiquant à un haut niveau leurs disciplines sportives). La vérification de leur validité — ou leur éventuelle adaptation — dans d'autres groupes humains reste à faire.

Remarque technique. Dans l'établissement des formules de décomposition du poids P en une part fonction linéaire de la taille T, et une part fonction linéaire de l'indice de Kaup K, nous avons utilisé la technique de régression linéaire multiple (de P sur T et K), puis nous avons réparti le terme constant de cette régression, entre les deux autres termes. Par exemple, dans le cas féminin, la régression linéaire s'exprime sous la forme :

$$\widehat{P} = (a + 0{,}7081209\ T) + (b + 2{,}884827\ K) \text{ avec } a + b = -120{,}1929$$

Nous avons attiré l'attention du lecteur sur le fait que cette répartition du terme constant pouvait se faire d'une infinité de façons, et nous avons à ce moment-là fait un choix, utilisant le barème de Davenport. Au cours de notre recherche, nous avons étudié d'autres solutions, parmi lesquelles celle consistant à travailler sur "l'individu moyen", partant de l'idée que, pour un individu "moyen" (c'est-à-dire de poids $\overline{P}$, de taille $\overline{T}$, et d'indice de Kaup $\overline{K}$), le poids réel $P = \overline{P}$ devrait coïncider avec le poids statural $a + 0{,}7081209\ \overline{T}$, et l'écart pondéral devrait être nul, soit $b + 2{,}884827\ \overline{K} = 0$. Cette façon de procéder aboutit à une solution ($a = -60{,}11386$; $b = -60{,}07909$). La formule obtenue est non biaisée (résidu moyen nul), mais sur le plan de l'interprétation concrète elle se heurte à une impossibilité "physique" : l'individu "moyen" tel que défini ci-dessus ne peut exister ($\overline{K}$, moyenne de $10^4\ P/T^2$ n'est pas égale à $10^4\ \overline{P}/\overline{T}^2$, bien qu'elle en soit proche numériquement : la moyenne d'un rapport n'est pas égale au rapport des moyennes).

Références

Mimouni N. (1996). *Contribution de méthodes biométriques à l'analyse de la morphotypologie des sportifs.* Thèse de Doctorat, Univ. Claude Bernard Lyon I.

Vandervael F. (1980). *Biométrie humaine.* Éd. Masson, Paris.