

sion simple $a_{k1}x_0^{Jk1}$ de y_0^J sur x_0^{Jk1} . La variable $a_{k1}x_0^{Jk1}$, projection orthogonale de y_0^J sur x_0^{Jk1} (régression simple), est aussi projection orthogonale de \tilde{y}_0^J sur x_0^{Jk1} (théorème des 3 perpendiculaires), alors que la variable $b_{k1}x_0^{Jk1}$ est la projection de \tilde{y}_0^J sur x_0^{Jk1} parallèlement à x_0^{Jk2} (régression partielle).

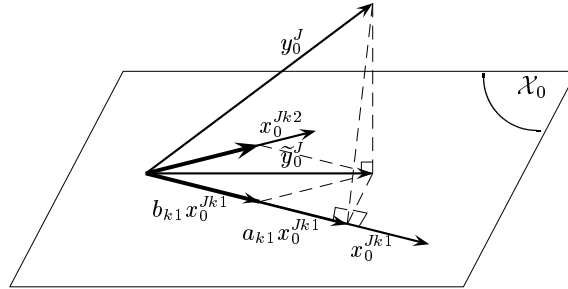


FIG. IV.10. Régression multiple et régression simple.

Exemple Psychométrie. Comme on l'a vu au §IV-2-b (p. 82), le coefficient de régression partielle $b_{k1} = 0.106$ diffère du coefficient de régression simple $a_{k1} = 0.118$. Le rapport de structure b_{k1}/a_{k1} vaut $0.106/0.118 = 0.90$ est peu inférieur à 1 : il y a légère atténuation (voir p. 96 l'étude de l'effet de structure pour deux variables).

Variables réduites. Posons $\text{Ety } x^{Jk} = \sigma_k$ et $\text{Ety } y^J = \sigma$, d'où les variables réduites $\frac{1}{\sigma_k}x_0^{Jk}$ et $\frac{1}{\sigma}y_0^J$ et rappelons que $\text{Corr}(y^J|x^{Jk}) = r_k$.

La régression simple de $\frac{1}{\sigma}y_0^J$ sur $\frac{1}{\sigma_k}x_0^{Jk}$ est $\frac{r_k}{\sigma_k}x_0^{Jk}$, d'où $a_k = r_k \frac{\sigma}{\sigma_k}$.

La régression multiple de $\frac{1}{\sigma}y_0^J$ sur $(\frac{1}{\sigma_k}x_0^{Jk})_{k \in K}$ est⁷ $\sum_k \frac{\beta_k}{\sigma_k}x_0^{Jk}$; d'où

$b_k = \beta_k \frac{\sigma}{\sigma_k}$, et la propriété : $\forall k \in K : \frac{b_k}{a_k} = \frac{\beta_k}{r_k}$.

3-e Cas particuliers et extensions

Cas singulier (colinéarité). Le cas singulier de la régression peut être caractérisé de plusieurs manières :

- dans l'espace d'observables, le nuage projeté parallèlement à l'axe des y est dans un sous-espace de dimension inférieure à K ;
- dans l'espace des variables, le sous-espace \mathcal{X} est de dimension inférieure à $K + 1$; ou encore, entre les variables $(x^{Jk})_{k \in K}$ et 1^J , on a des relations linéaires ; d'où le terme de *colinéarité* ;
- la matrice des variances-covariances des K variables prédictrices

⁷La notation β_k ("beta-weights") pour désigner les coefficients de régression partielle réduits est usuelle dans les logiciels.