

Exercise

Consider the following dataset (Royston & Sauerbrei, 2008):

A.2.3 Educational Body Fat

To exemplify some issues in multiple regression analysis, Johnson (1996) explored a dataset comprising a response variable (the estimated percentage of body fat) and 13 continuous covariates (age, weight, height and 10 body circumference measurements) in 252 men (see Table A.4). The mixture of imperial and metric measurement units is how the data were provided. The aim was to predict percentage body fat from the covariates.

Table A.4 Educational body-fat data.

Variable	Name	Details	Variable	Name	Details
x_1	age	Age (years)	x_8	thigh	Circumference (cm)
x_2	weight	Weight (lb)	x_9	knee	Circumference (cm)
x_3	height	Height (in)	x_{10}	ankle	Circumference (cm)
x_4	neck	Circumference (cm)	x_{11}	biceps	Circumference (cm)
x_5	chest	Circumference (cm)	x_{12}	forearm	Circumference (cm)
x_6	ab	Circumference (cm)	x_{13}	wrist	Circumference (cm)
x_7	hip	Circumference (cm)	y	pcfat	Body fat ^a (%)

^a Response variable.

Download the data by using the code in the course web-page (link: Exercises) and perform model selection with:

- backward elimination;
- forward selection;
- stepwise selection;
- stepback selection;
- best subset selection

Report the regression coefficient estimates (with the corresponding standard errors) and comment the results.

References

ROYSTON, P. & SAUERBREI, W. (2008). *Multivariable Model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables*. Wiley, Chichester.