

The data frame HSWRESTLER, (from package PASWR2) contains information on nine variables for a group of 78 high school wrestlers that was collected by the human performance lab at Appalachian State University. The variables are

- age (in years)  $\textcircled{0}$
- ht (height in inches)  $\textcircled{0}$
- wt (weight in pounds)  $\textcircled{b}$
- abs (abdominal skinfold-measure)  $\textcircled{D}$
- triceps (tricep skinfold measure)  $\textcircled{Q}$
- subscap (subscapular skinfold measure)  $\textcircled{Q}$
- hwfat (hydrostatic determination of fat)  $\checkmark$
- tanfat (Tanita determination of fat)  $\texttimes$
- skfat (skinfold determination of fat)  $\texttimes$

It is of interest to predict wrestler's hydrostatic fat (hwfat) using predictors age, ht, wt, abs, triceps and subscap. Remove observations 22, 27, 32, 35, and 60, which may have been poorly measured. Use 5-fold cross validation to find the best regression model.

To find the best set of predictors with the highest predictive performance, we will

- Select best set of predictors using `regsubsets()`
- Select the number of predictors using cross validation
- Build the model and find coefficient estimates using the full data set