

## Angewandte Regression — Serie 7

1. Für ein Modell der Lawinenprognose ist es wichtig, die Dichte der Schneedecke zu kennen. Da man dies im Allgemeinen nicht lokal messen kann, wird versucht, ein Modell zu entwickeln, mit dem man die Dichte aus meteorologischen Variablen bestimmen kann.

Der Datensatz `snowdens1` enthält die folgenden Variablen:

Date	date
Density	density of new snow (Zielgrösse)
TA (AirTemp)	air temperature
TSS (SurfTemp)	surface temperature
RH (RelHumid)	relative humidity
VW (WindVeloc)	wind velocity
HH (Elev)	elevation of the measurement site: 1560 for Davos, 2290 for Weissfluhjoch

Der Datensatz `snowdens2` liefert die gleiche Information (ausser Elev) für eine zweite Messkampagne, die wie in Davos gemessen wurde (Elev = 1560).

Entwickeln Sie ein Modell, dass die Dichte gut vorhersagt.

**Hinweise:**

- Setzen Sie die beiden Daten zusammen.
- Machen Sie einen Scatterplot und betrachten Sie besonders `plot(~Density+VW)`. Was fällt auf?  
Hilfe dazu: Im File `r-swnodens_help.R` (siehe <ftp://stat.ethz.ch/WBL/Source-WBL-2/R/>) wird veranschaulicht, wie und weshalb die erklärende Variable mit `pmin()` und `pmax()` gesplittet wird. Die damit erzeugten Abbildungen sind auf der Rückseite abgedruckt.
- Optimales  $\lambda$  in Lasso: die Graphik der Cross-Validation kann mit `cv.lasso(...)` oder `plot(..., type="criteria", cv=TRUE)` erstellt werden. Das optimale  $\lambda$  ist das Minimum der roten Kurve.

