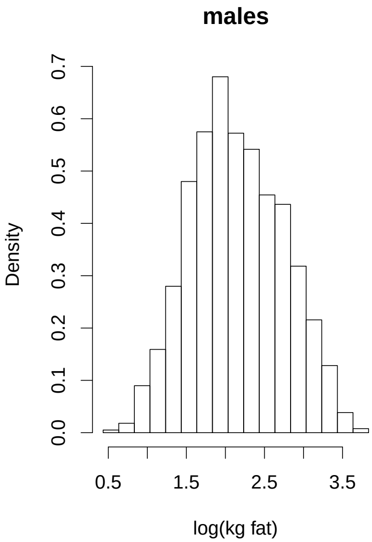
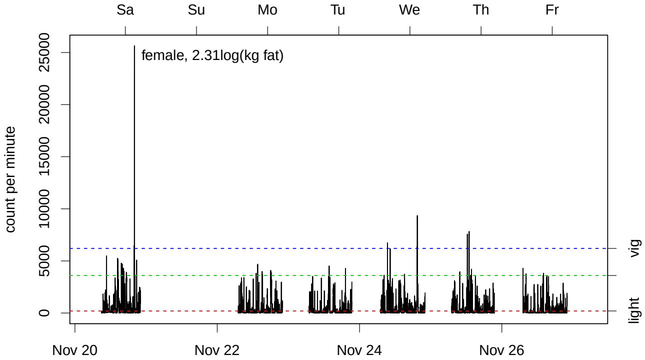
Modelling Obesity as a function of weekly physical activity profiles measured by Actigraph accelerometers

Nicole Augustin[1]
[1]**Department of Mathematical Sciences, University of Bath, UK**
In collaboration with Callum Mattocks, Chris Riddoch, Andy Ness and Julian Faraway.

Health outcome: fat mass at age 12, n = 4120



Predictor: physical activity profile at age 12

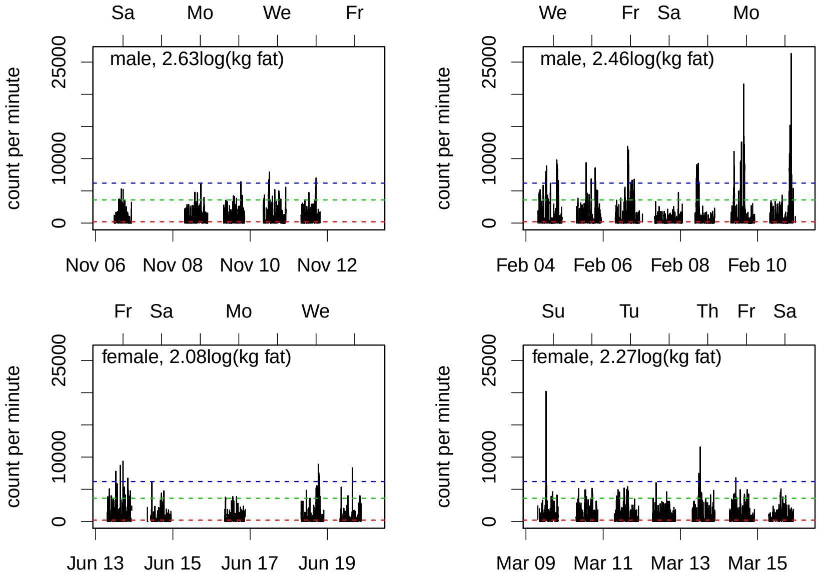


- time series of measurements by minute over 7 days of 4120 individuals at age 12;
- after some pre-processing;
- cutpoints for light/moderate/vigorous activity estimated from a calibration study (Mattocks et al, 2007);
- mean hours worn per day: 11 hours (SD 4.9 hours).

Protocol for pre-processing of activity profiles

1. replace any sequence with more than 10 zeros by missing values;
2. exclude days if:
   - mean count < 150;
   - mean count > 3 SD above overall mean (prior to exclusions);
   - monitor was worn < 10h;
3. Exclude weekly profiles if < 3 valid days were observed.

random sample of 4 profiles



Objectives

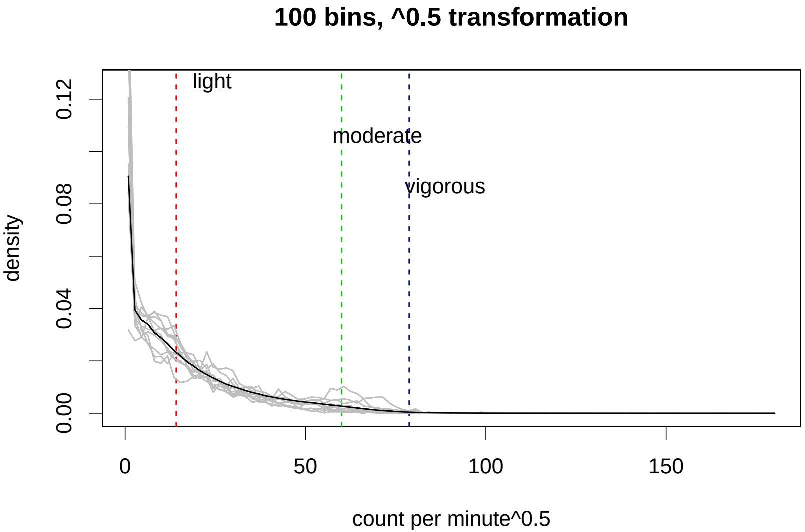
- Scientific question: What is the relationship between physical activity profile and fat mass?
- So far only single summary statistics of physical activity profiles are used, e.g. moderate to vigorous physical activity (MVPA).
  - MVPA neglects a large part of the activity profile where activity is light or below;
  - relies on setting cut-points for moderate and vigorous activity;
  - cut-points may change with age;
  - the pattern of physical activity may be relevant.
- Aim: develop a statistical tool to explore the relationship between physical activity and fat mass.

Requirement: a functional summary of the profiles

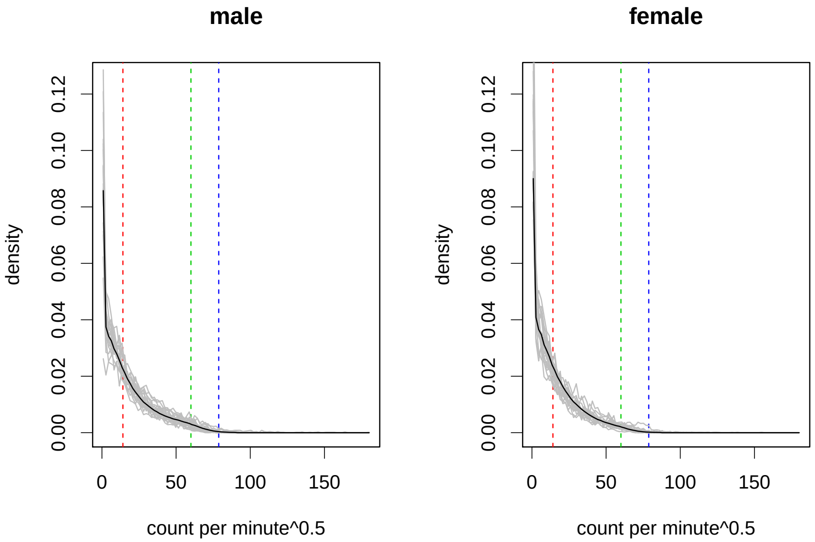
- cannot compare individuals using profiles directly;
- need to reduce the dimension of data;
- the function should be simple and easy to interpret;
- different possibilities: spectrum, quantiles, cumulative density function, histogram, ...

Here we use the histogram as a function.

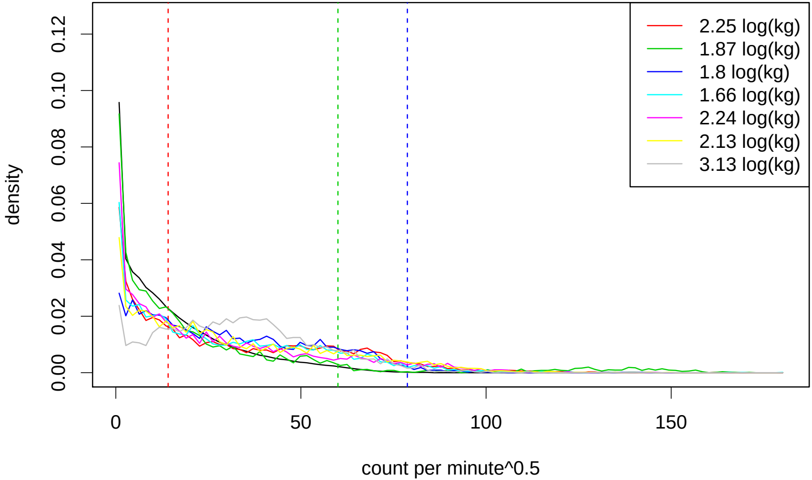
Mean histogram



Mean histogram



Some extreme histograms



Model fat mass as a function of the accelerometer profile

- response y_{ik} total fat mass for individual i at age k;
- vector x_{ik} is the accelerometer profile, with 10080 entries;
- $z_{ik}(x)$ is the histogram with some given number of mid-points x_j;

Start with a linear model:

$$log(y_{ik}) = \alpha + \sum_j \beta_j z_{ik}(x_j) + \sum_l \gamma_l \texttt{confounder}_{lik} + \epsilon_{ik}.$$

with $\epsilon_{ik} \sim N(0, \sigma^2)$ and confounders `sex, height, height`^{2}.

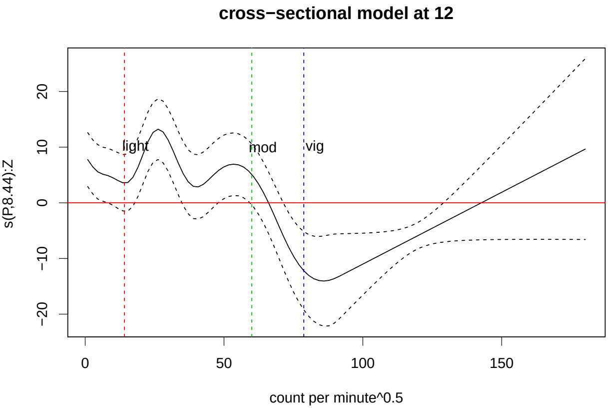
A generalised regression of scalars on functions (Ramsay and Silverman, 2005)

Let the β_j vary smoothly, where $\beta_j = f(x_j)$:

$$log(y_{ik}) = \alpha + \sum_j f(x_j) z_{ik}(x_j) + \sum_l \gamma_l \texttt{confounder}_{lik} + \epsilon_{ik}.$$

- $f(x)$ is an unknown 'coefficient' function to be estimated;
- $f(x)$ is represented using an adaptive smooth with a P-spline basis;
- use penalised iteratively re-weighted least squares for parameter estimation;
- Wood (2010), implemented in `R mgcv package`.

Preliminary model results



- results robust to different bin widths;
- backs up the cut-point used for MVPA
- moderate and vigorous activity has a negative effect on fat mass;

Further work

- Should the very low levels of activity with counts close to zero be included?
- Which transformation of counts should be used?
- classification of activity profiles;
- investigate other types of summary functions;
- look at different ages;
- how sensitive are results to different ways of pre-processing.

References

Mattocks, C., Ness, A., Leary, S., Tilling, K., Blair, S., Shield, J., Deere, K., Saunders, J., Kirkby, J., Smith, G. D., Wells, J., Wareham, N., Reilly, J., and Riddoch, C. (2008). Use of accelerometers in a large field based study of children: Protocols, design issues, and effects on precision. *Journal or Physical Activity and Health*, 5:S94–S107.

Ramsay, J.O. and Silverman, B.W. 2005. Functional Data Analysis. Springer Verlag.

Wood, S. (2010). Fast stable REML and ML estimation of semiparametric GLMs. *Journal of the Royal Statistical Society, Series B*, in press.