SSA and TRF Analysis of 2003-2004 NHANES Actigraphy Data by Age and Employment

We analyzed the difference between ages in actigraphy data from the CDC's NHANES in 2003-2004. Since physical activity is a component of health, we wanted to understand physical activity patterns by finding differences in patterns. Different age groups should not have had different patterns of activity based solely on the number of years alive, but confounders such as health status and employment, which come with age, might have affected physical activity. An actigraph device recorded the physical movement for 7167 participants every minute for an entire week. Each participant had 10,080 measurements (60*24*7) of time series data that would need to be simplified for meaningful interpretation.

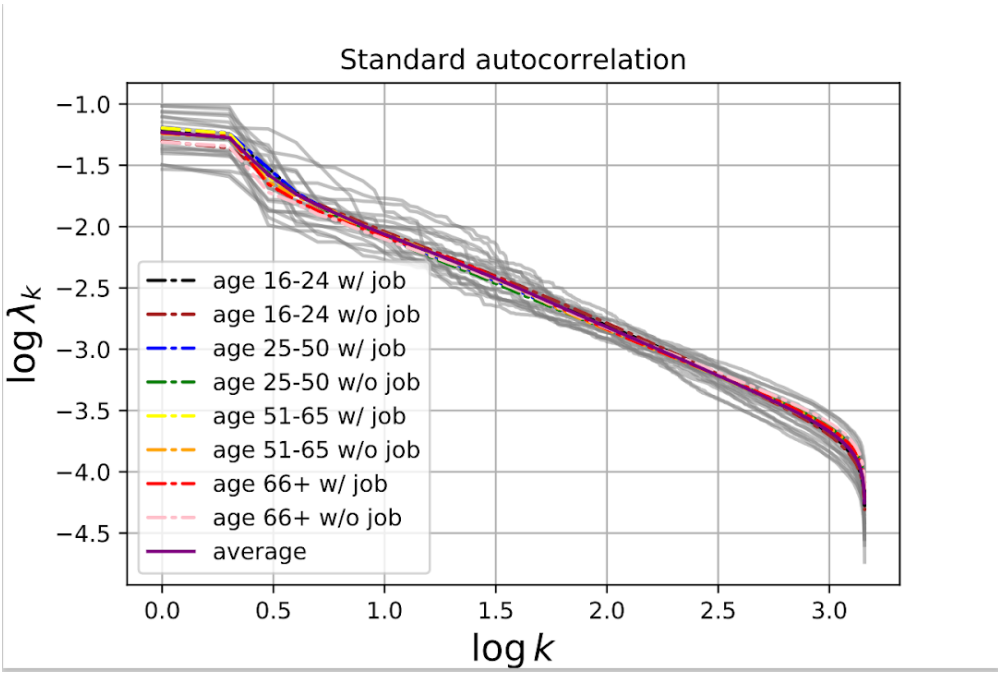
Singular spectrum analysis (SSA) reduced dimensionality by converting the time series data into eigenvalues that we could compare. We converted the original data into a lag-shifted matrix, the Hankel matrix, and performed singular value decomposition (SVD) for a row of eigenvalues. The standard autocorrelation plot of logged and ordered eigenvalues versus their logged positions was similar to a scree plot for PCA in that we can tell which components can describe most of the patterns in the original actigraphy data. We expected that the relationship between the ordered eigenvalues and their positions would follow the power law, so the logged relationship should and was linear for a range of positions. The eigenvalues before this linear relationship explained most of the variance in the NHANES data so those first few eigenvalues described the dominant patterns of physical activity for each participant. While these first eigenvalues themselves did not have much meaning, the aggregate of these first eigenvalues could visually represent which eigenvalues would be important in describing physical activity patterns.

We first did SSA on age for a general idea of how age affected physical activity. We created age categories (16-24, 25-50, 51-65, 66+) based on popular life events such as education and retirement for interpretability. The 66+ group was distinctly below the overall average while the 16-24 group was slightly below the overall average, suggesting that these two groups were visually distinct in physical activity compared to the 25-50 and 51-65 groups. We included other factors such as gender, cholesterol, and blood pressure in hopes that physical attributes contributed to this difference between age groups, but none of those factors revealed any visible differences. Including whether the participants were employed, simplified to the 'job factor', finally showed a meaningful interaction. The job factor did not show a difference in physical activity pattern based on whether the participant was working during the week of

observation but did show a difference based on whether the participant was contemporarily employed. Combined with jobs, age stratified further in the standard autocorrelation plot.

We used individual time frequency representations (TFR) to qualify the validity of each physical activity pattern. We sampled 5 participants in each age-job category and visually analyzed their TFR to gain insight into each group. The 16-24 groups, both with a job and without a job, had various TFRs in which some barely showed viable patterns and others had a consistent 7 day pattern. The same was true for the 66+ groups, both with and without a job. The other groups also had a variety of patterns, but they had more consistent and viable 7 day patterns with one or two days of non-activity. These analyses suggested that, again, the youngest groups and the oldest groups were the most different in physical activity patterns but also that the youngest and oldest groups were more likely to have unviable outliers that are unrepresentative of the groups.

We concluded that employment was more important in explaining the variance in physical activity patterns than gender, cholesterol, or blood pressure. The TRFs suggested that the measured patterns for the youngest and oldest groups were either not representative or had greater variance than that for the 4 middle aged groups. Assuming that all the measurements were representative of the participants' general physical activities, the standard schedule of a job seemed to have a great influence on participants' activity patterns. The 66+ without a job and 16-24 without a job groups were the most different from the average while the 16-24 with a job and 25-50 with a job groups were also different from the average to a lesser degree but the opposite direction. The 25-50 without a job and 51-65 without a job groups were still similar to the 51-65 with a job and 66+ with a job groups, all of which were similar to the average pattern. Many 25-65 year old participants without a job might either have been between jobs and maintained the activity pattern ingrained by their previous jobs or might have chosen domicile work, which NHANES did not measure, that mirrors non-domicile work in physical activity patterns. Simply not having the pressure of a consistent job, which would be more frequent for 16-24 year olds who focused on education and 66+ year olds who focused on retirement, increased the variance in physical activity. Changes in physical activity pattern, then, would be related to experience and expectation of employment.


Standard autocorrelation
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
−4.0
−4.5
log λ_k
0.0
0.5
1.0
1.5
2.0
2.5
3.0
log k
age 16-24 w/ job
age 16-24 w/o job
age 25-50 w/ job
age 25-50 w/o job
age 51-65 w/ job
age 51-65 w/o job
age 66+ w/ job
age 66+ w/o job
average