Statistical Data Analysis
TKO_7093-3004

Group 160

Ayana Kotuwegoda Guruge	2406865
Sheheryar Wahidi	2413773
Yagya Yadav	2409273

UNIVERSITY
OF TURKU

Statistical Analysis of Daily Activity Patterns in Finland

An examination of the Finnish time-use survey (habits.data) to understand how individuals allocate their time across daily activities, explore differences between population groups, and identify relationships between activity patterns.

745

Person-Day Observations

Total data points captured

378

Unique Individuals

Independent participants surveyed

378

Households

All single-person households

RESEARCH FRAMEWORK

Research Objectives

01

Characterise Population

Demographics and household composition analysis

02

Estimate Time Allocation

Average minutes per activity with 95% confidence intervals

03

Compare Groups

Weekday vs. weekend patterns and living environment differences

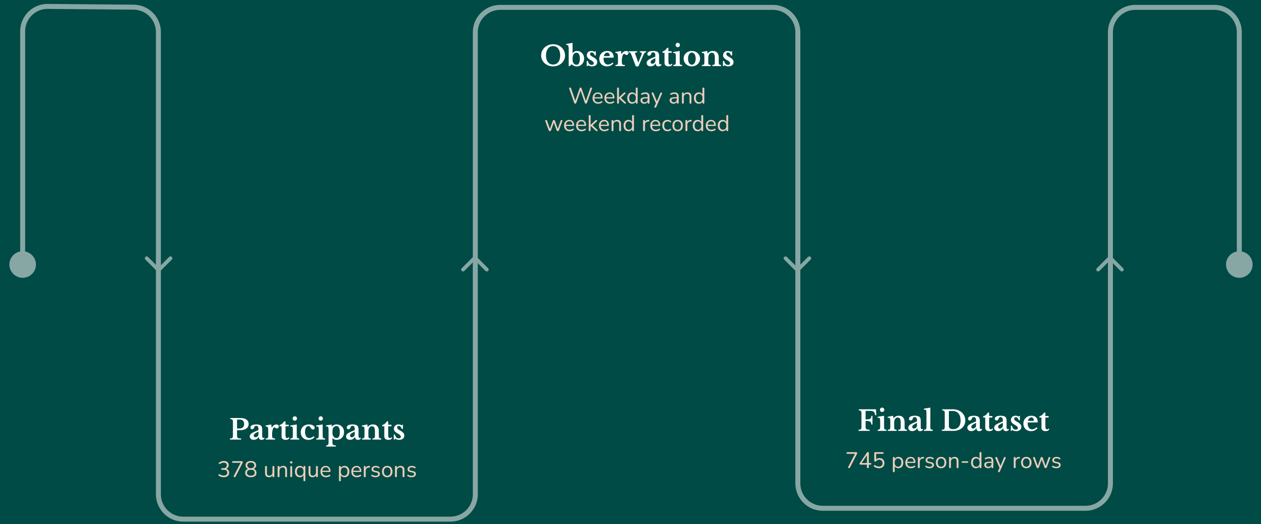
04

Identify Associations

Correlation analysis and principal component patterns

Data Structure & Analytical Approach

The dataset comprises person-day observations, capturing both weekday and weekend activities for each participant. This design choice creates duplicate rows per individual, resulting in 745 total observations from 378 unique participants.



Participants

378 unique persons

Observations

Weekday and weekend recorded

Final Dataset

745 person-day rows

This longitudinal structure enables within-person comparisons across day types while maintaining individual-level variation.

Key Design Decision

Analysis uses household-level aggregation for statistical inference. Given that all 378 households consist of single individuals, household-level aggregates are mathematically equivalent to individual-level measures in this dataset.

Activity Variables & Measurement Scales

The survey captures five distinct activity categories, each measured using specific scales appropriate to the nature of the activity and data distribution.

1

Activities A1–A4

Continuous measurement: Minutes spent per day on each activity. Treated as numeric variables after data cleaning and validation processes.

These variables exhibit right-skewed distributions with substantial zero values, reflecting varied participation patterns across the population.

2

Activity A5

Categorical measurement: Participation indicator derived from mixed original values (0, 1, 2, 60, 120, 420 minutes).

A new binary variable A5_binary was created to enable valid categorical analysis while preserving the original A5 data for reference.

Creating A5_binary: A Methodological Choice

Rather than replacing the original A5 variable, researchers created a separate derived variable to maintain data integrity and analytical flexibility.

1

Raw A5 Data

Mixed values: 0, 1, 2, 60, 120, 420

Not interpretable as pure yes/no

2

A5_binary Created

Valid categorical variable

Enables chi-square testing

3

Original Preserved

Raw data unchanged

Maintains audit trail

This approach follows best practices in data management: preserve original data while creating analysis-ready variables with clear documentation.

Population Characteristics

The sample represents a diverse cross-section of Finnish adults, with balanced gender distribution and broad age representation across living environments.

Gender Distribution

Female: 196 (51.9%)

Male: 182 (48.1%)

Age Range Representation

20–24: 12 | **25–34:** 57

35–44: 64 | **45–54:** 81

55–64: 93 | **65–74:** 44

75+: 27

Living Environment

City: 240 (63.5%)

Rural: 80 (21.2%)

Municipality: 58 (15.3%)

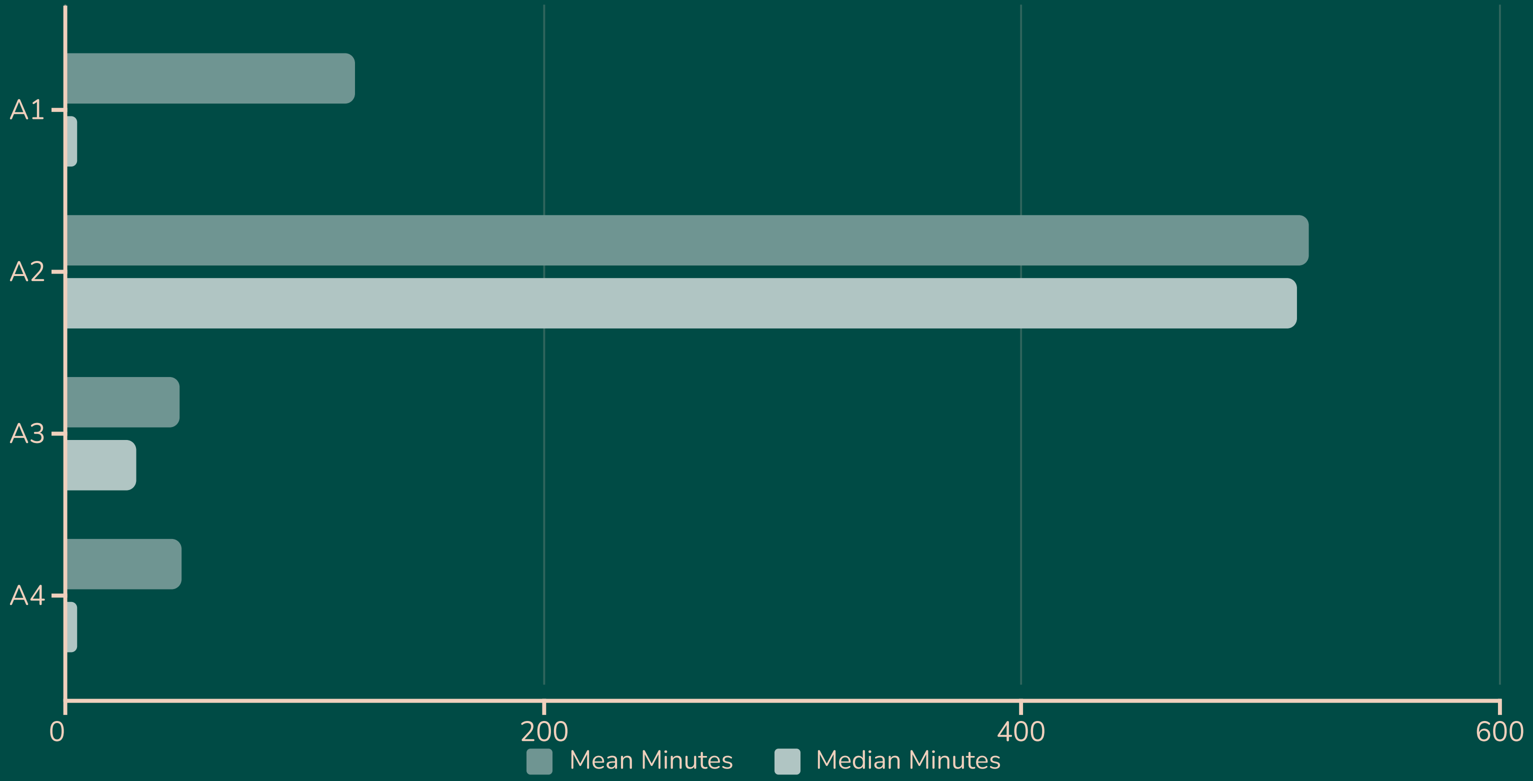
Household Composition

All 378 households consist of single individuals (mean household size = 1.0). This uniform household structure simplifies interpretation: individual-level and household-level analyses are mathematically equivalent.

The majority of participants reside in urban areas, reflecting Finland's demographic distribution with approximately two-thirds of the population living in cities.

Average Daily Time Allocation Patterns

Descriptive statistics reveal substantial variation in activity participation, with some activities showing universal engagement while others demonstrate selective participation patterns.



The stark contrast between mean and median values for A1 and A4 (median = 0) indicates highly right-skewed distributions with many non-participants. A2 shows the most consistent engagement with near-equal mean and median. Analysis employed household-level averages, though these equal individual-level measures given uniform single-person households.

STATISTICAL PRECISION

Confidence Intervals for Time Estimates

95% confidence intervals quantify uncertainty in population mean estimates, revealing which activities show more stable versus variable time allocations across the sample.

Activity A1

Mean: 121.6 minutes

95% CI: [106.6, 136.7]

Range: ±30.1 minutes

Activity A2

Mean: 520.2 minutes

95% CI: [512.6, 527.8]

Range: ±15.2 minutes

Activity A3

Mean: 48.6 minutes

95% CI: [42.5, 54.7]

Range: ±12.1 minutes

Activity A4

Mean: 49.1 minutes

95% CI: [44.7, 53.5]

Range: ±8.8 minutes

Interpretation

Activity A2 demonstrates the most stable estimate with the narrowest confidence interval relative to its mean, indicating consistent time allocation across participants.

Activity A1 shows the widest absolute interval, reflecting greater individual variation in participation and duration for this activity type.

Weekday vs. Weekend Activity Patterns

Welch independent-samples t-tests compared time allocation between weekday and weekend observations, accounting for potentially unequal variances between groups.

Activity A1

p = 0.746

No significant difference

Activity A2

p = 0.404

No significant difference

Activity A3

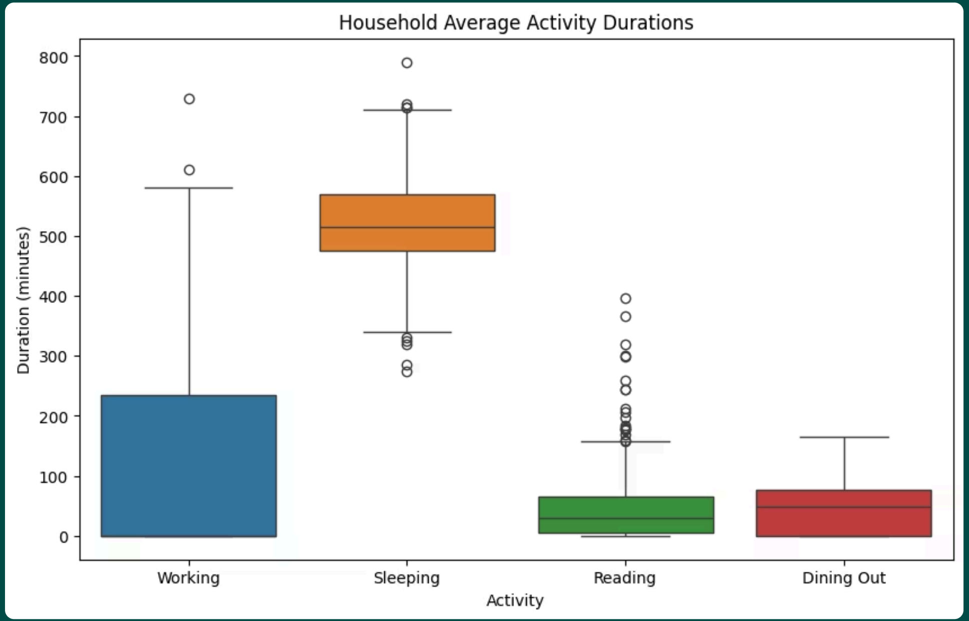
p = 0.098

No significant difference

Activity A4

p = 0.006 **

Significantly higher on weekends



Key Finding: Only Activity A4 (dining out) shows statistically significant temporal variation, with substantially more time allocated on weekends compared to weekdays.

This pattern aligns with expected social behavior, as individuals typically have more leisure time and flexibility for restaurant dining during weekend periods.

Living Environment Comparisons

Kruskal-Wallis tests examined whether time allocation differs across living environments (city, municipality, rural), using a non-parametric approach appropriate for skewed distributions with zero values.

Activity A1

p = 0.052

Marginally non-significant (approaching threshold)

Activity A2

p = 0.261

No significant difference across environments

Activity A3

p = 0.012 *

Significant environmental effect

Activity A4

p = 0.737

No significant difference across environments

Interpretation: Activity A3 time allocation varies significantly by living environment, suggesting geographic factors influence participation in this activity. Follow-up Dunn post-hoc tests would identify which specific environment pairs differ significantly.

Post-hoc Testing: Activity A3 by Living Environment

Following the significant Kruskal-Wallis result for Activity A3, Dunn's post-hoc tests with Bonferroni adjustment were applied to identify which specific living environment pairs exhibited significant differences in time allocation.

City vs. Municipality

Adjusted $p = 0.026893$

Significantly Different

City vs. Rural

Adjusted $p = 1.0000$

No significant difference

Municipality vs. Rural

Adjusted $p = 0.12023$

No significant difference

- A statistically significant difference was found between city and municipality households (adjusted p = 0.027).
- No significant differences were observed between city and rural or municipality and rural households.
- This suggests that differences in Reading time across living environments are mainly driven by the contrast between city and municipality residents.

Binary Outcome Analysis (A5)

Chi-square tests were conducted to assess the association between Activity A5 participation (a binary outcome: participated/did not participate) and two key demographic factors: living environment and weekday/weekend observations.

Participation by Living Environment

City	311	128	439
Municipality	79	35	114
Rural	87	63	150

Chi-square = 8.584, df = 2, p = 0.0137.

Activity A5 participation significantly differs across living environments, suggesting that geographical location plays a role in whether individuals engage in this activity.

Participation by Weekday/Weekend

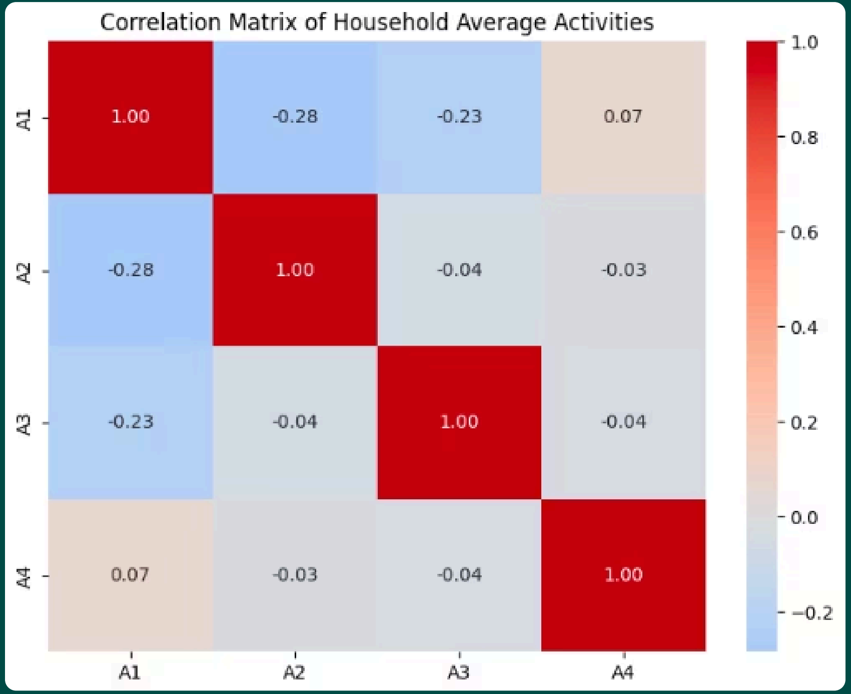
Weekday	239	116	355
Weekend	238	110	348

Chi-square = 0.049, p = 0.8243.

There is no significant difference in Activity A5 participation between weekdays and weekends, indicating that temporal factors do not influence engagement in this specific activity.

Activity Associations

Understanding the relationships between different daily activities helps identify potential dependencies and behavioral patterns. Pearson correlation analysis was used to quantify the linear relationships between the time spent on each activity.



Key Insights from Correlation Analysis

The analysis of activity associations revealed that most correlations between daily activities are relatively weak, suggesting a degree of independence in how individuals allocate their time across these different categories.

- **A1 vs. A2:** Corr = -0.284 (Weak negative)
- **A1 vs. A3:** Corr = -0.233 (Weak negative)
- **A1 vs. A4:** Corr = 0.068 (Very weak positive)
- **A2 vs. A3:** Corr = -0.040 (Very weak negative)
- **A2 vs. A4:** Corr = -0.027 (Very weak negative)
- **A3 vs. A4:** Corr = -0.036 (Very weak negative)

Negative correlations suggest that as time spent on one activity increases, time on the other tends to decrease, though these effects are minor.

Why PCA (and Why Not Clustering)

Principal Component Analysis (PCA) was employed to reduce the dimensionality of activity data, identify underlying patterns, and avoid issues with highly correlated variables. Clustering was deemed unsuitable due to the continuous nature of the data distribution.

Methodological Approach

- **Variable Standardization:** Activities were standardized to prevent dominant activities (e.g., sleep) from disproportionately influencing the analysis.
- **Missing Data Handling:** Households with missing activity averages were removed, reducing the sample from 378 to 336 for PCA.

Principal Component Interpretation

No Clear Clusters

The PCA scatter plot revealed a continuous spread of data points rather than distinct groupings, indicating no natural clusters within the activity patterns, thus making clustering methods inappropriate.

PCA Explained Variance

PC1: 34.3%

PC2: 25.9% (Cumulative 60.2%)

PC3: 24.5%

PC4: 15.2%

PC1: Work–Rest Tradeoff

Strong positive loading for A1 (+0.71) and negative loadings for A2 (–0.53) and A3 (–0.41), suggesting a dynamic between productive time and rest/personal care.

PC2: Leisure vs. Sleep

High positive loading for A3 (+0.74) and negative for A2 (–0.64), indicating a contrast between leisure activities and sleep.

PC3: Occasional Activities

Dominated by A4 (+0.95), representing activities like dining out or special events that occur less frequently.

Final Conclusions

Our analysis provides several key insights into daily activity patterns in Finland:

Dominant Activities & Infrequent Engagements

Sleeping constitutes the largest portion of daily time, while activities such as reading and dining out are often reported as zero, indicating their occasional nature.

Weekend Impact on Dining Out (A4)

Day-type significantly influences only dining out (Activity A4), which shows a higher incidence on weekends, suggesting a leisure-oriented shift.

Living Environment Influences Activity A3

Time allocation for Activity A3 differs significantly by living environment (Kruskal-Wallis p=0.0117), primarily driven by disparities between city and municipality residents.

Activity A5 Participation Varies by Environment, Not Day-Type

Participation in Activity A5 is significantly influenced by living environment (p=0.0137) but shows no significant difference between weekdays and weekends (p=0.8243).

Weak Activity Correlations & PCA Insight

Pearson correlations between activities are generally weak, indicating that most daily activities are largely distinct. PCA further supports a continuous, gradual variation in activity patterns, rather than distinct clusters.