OCTAGON HEALTH DATA SCIENCE COMPETITION

NOVEMBER 5, 2021

Digital Biomarkers for Cognitive Performance

Amber, Gavincko, Kevin, and Zoila

Introduction

- Demand for **telehealth** increasing faster than ever
- Biomarker predicting cognitive task performance should be **personalized** [1]
- **Hexoskin & cognitive tests** can be used to train a personalized biomarker for cognition
- <u>**What statistical model and biometric data best predict cognitive performance?**</u>

[1] Patel, A. N., Howard, M. D., Roach, S. M., Jones, A. P., Bryant, N. B., Robinson, C., Clark, V. P., & Pilly, P. K. (2018). Mental State Assessment and Validation Using Personalized Physiological Biometrics. Frontiers in human neuroscience, 12, 221. https://doi.org/10.3389/fnhum.2018.00221

Study Design & Dataset

- Subject wore Hexoskin for **30 testing sessions**
  - One session is **48 hours** over **three days**
- Data collected
  - **Biometrics** (heart rate, breath rate, cadence, blood glucose, blood pressure)
  - **Cognitive test scores** (immediate & delayed recognition, memory)

Time of Day	Permutation 1	Permutation 2	Permutation 3
9 AM	TV	Reading	TV
3 PM	Aerobic Exercise	Aerobic Exercise	No Exercise
11 PM	Sugar	Sugar	No Sugar

Data Cleaning

HexoSkin biometric **datasets renamed** according to Session and Day; loaded dataframes **edited to reflect Day, Session, and Permutation**

Immediate actions defined by Permutation, Day, and Time; **assumption made that entries within 1 hour of the scheduled activity** timeslot reflected immediate activity.

Recent actions defined similarly; flexibility on **tagging entries between activity schedules** as opposed to within the expected activity period.

Datasets merged, then **incorporated with measures from cognition dataset** corresponding to entries with appropriate Day, Permutation, and Time.

Unfiltered extraction of entries was 4,576,160; final filtered yield was **4,268,132 entries**

Fitting 3 LASSO Prediction Models

Omit Observations with NA

4,268,132 observations & 14 variables

↓

Train & Validate LASSO model with 10-fold Cross Validation

Outcome: IR, DR or STROOP

Predictors: Systolic & Diastolic blood pressure, Heart rate, Breathing Rate, Cadence, Permutation, Immediate/Recent Action

Baseline Model: Numeric Predictors = 0, Permutation = 1, Immediate & Recent Action = Exercise

↓

Determine **Lambda** that Minimizes Cross Validation Mean Squared Error (MSE)

↓

Determine **Predictor Weight Estimates** in Optimal Model

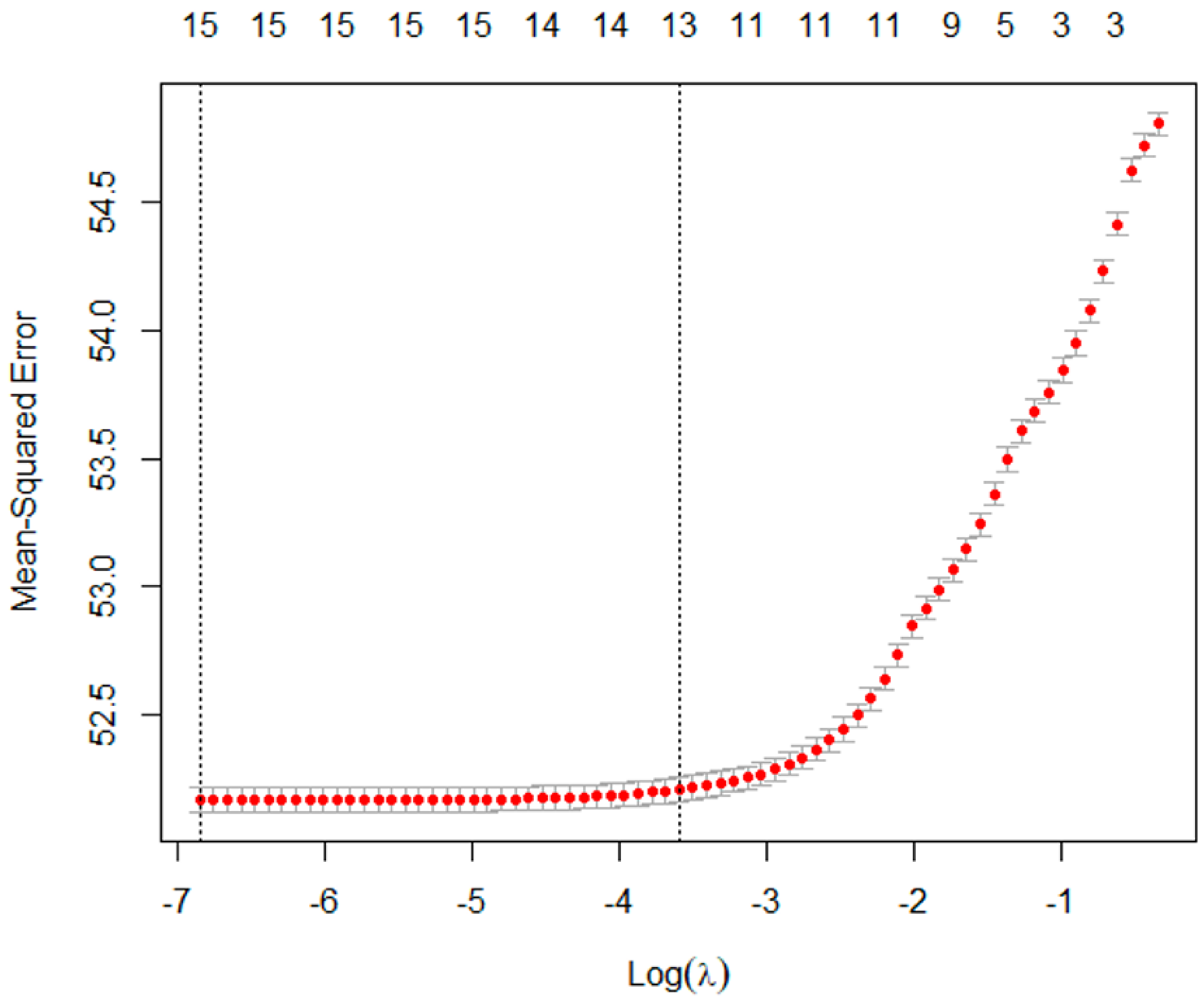

Figure 1. MSE for various Stroop model cross validation lambda values

IR = immediate recognition, DR = delayed recognition

Results from 3 Prediction Models

Predictor Number	Model 1: STROOP Prediction	Model 2: Immediate Recognition (IR) Prediction	Model 3: Delayed Recognition (DR) Prediction
Intercept	**STROOP score:** 125.48	**IR score:** 98.0480	**DR score:** 67.5994
Predictor 1	**Recent action: Reading** (W = 8.22) *vs. baseline: Exercise*	**Recent action: Reading** (W = 3.159) *vs. baseline: Exercise*	**Recent action: Reading** (W = -7.74) *vs. baseline: Exercise*
Predictor 2	**Immediate action: Sugar** (W = 5.75) *vs. baseline: Exercise*	**Immediate action: Sugar** (W = -4.77) *vs. baseline: Exercise*	**Recent Action: TV** (W = 5.52) *vs. baseline: Exercise*

Table 1. Model intercept and significant predictor weight estimates for STROOP, IR and DR prediction models

Intercept Interpretation: the model predicts a subject will on average have a **STROOP score of 125.48 points** when all numeric predictors (i.e heart rate) = 0 and all categorical variables (i.e recent action) are at baseline: exercise.

Predictor Interpretation: the model predicts that when a subject's recent action is reading (compared to exercising), their **STROOP score will on average be 8.22 points higher.**

Limitations & Future Steps

- **Braincheck Score units and scaling** - Need to understand how scores are calculated in order to develop a composite score
- **Subgroup analysis** - Statistical significance between various groups of observations. Are predictions significant? Are existing samples significantly different?
- **Lack of information on individual** - Adherence, external habits; each is their own.
- **Covariates** - Composite variables can be investigated; activities may influence biometrics.

Thank You!

Appendix

LASSO: Penalization

$$\sum_{i=1}^{N} \left(w_0 + \sum_{j=1}^{k} w_j x_j - y_i \right)^2 + \lambda \sum_{j=1}^{k} |w_j|$$

Stroop MSE Graph