Asthma Severity

Kadie Iverson & Matthew Donaldson

3/21/2022

Executive Summary

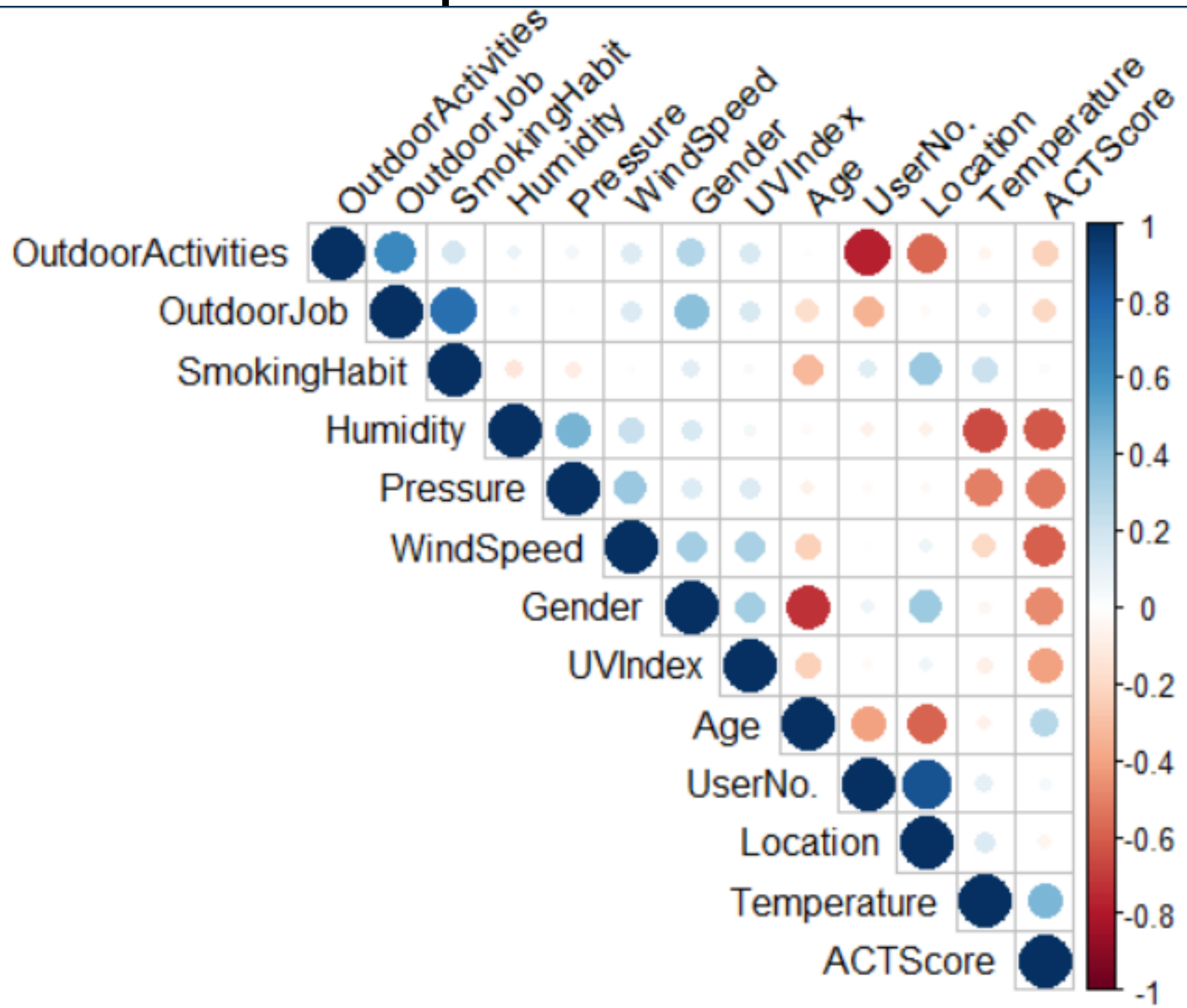
- Analytic Objective: To predict the severity of an asthma attack
- Decisions Impacted: Help people with asthma become more aware of having an asthma attack and if it could be severe
- Business Value: Could change the way inhalers are manufactured and create an app that accounts for the most significant predictors
- Data Assets: 11 features, from a study done by Radiah Haque

Data Asset Description

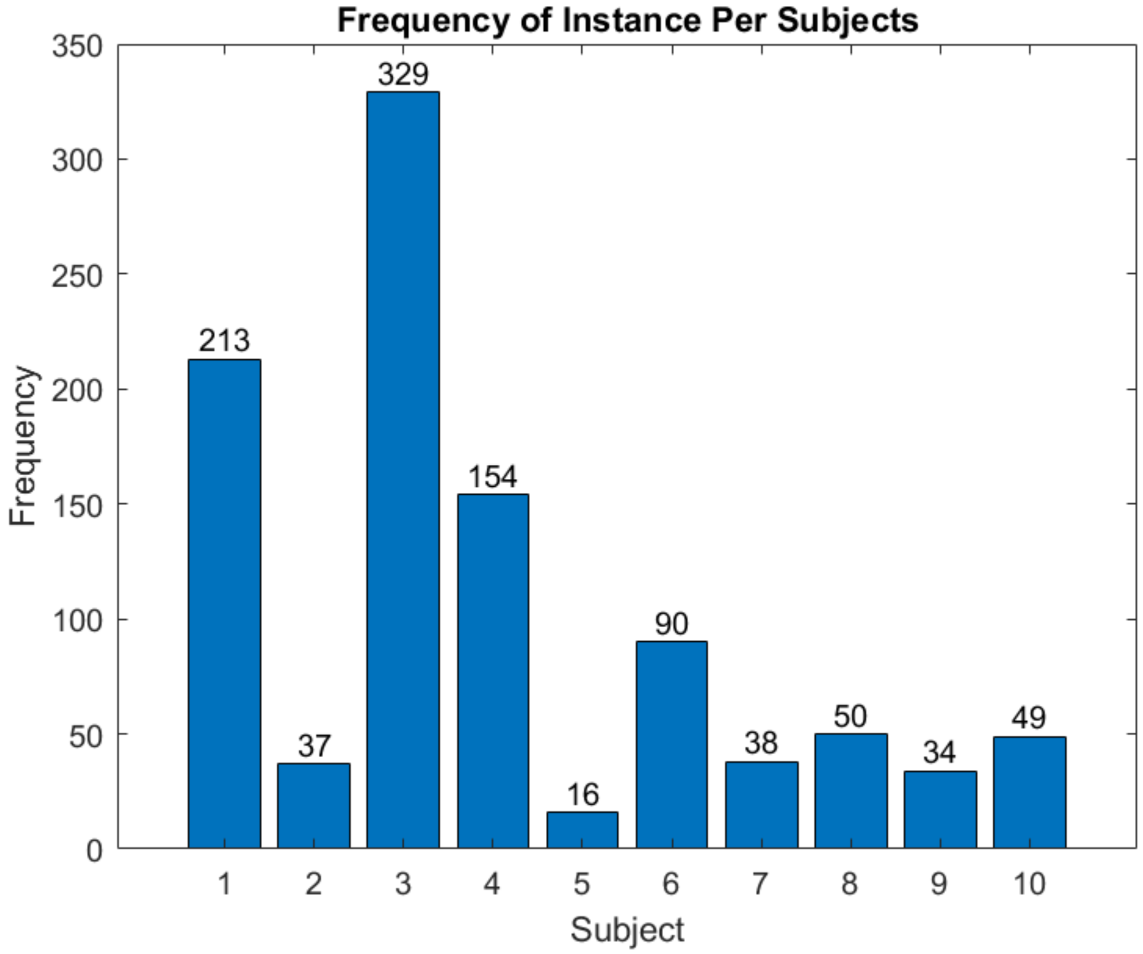
- Dimension: 1010 by 13

```
    UserNo.          Location              Age                Gender           OutdoorJob        OutdoorActivities  SmokingHabit
 Min.   : 1.000   Length:1010        Length:1010        Length:1010        Length:1010        Length:1010        Length:1010
 1st Qu.: 3.000   Class :character   Class :character   Class :character   Class :character   Class :character   Class :character
 Median : 3.000   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character
 Mean   : 3.933
 3rd Qu.: 6.000
 Max.   :10.000
    Humidity         Pressure      Temperature      UVIndex            WindSpeed        ACTScore
 Min.   : 40.00   Min.   :1003   Min.   :21.10   Length:1010        Min.   :0.000   Min.   : 8.00
 1st Qu.: 70.00   1st Qu.:1008   1st Qu.:25.20   Class :character   1st Qu.:1.000   1st Qu.:12.25
 Median : 87.00   Median :1009   Median :27.50   Mode  :character   Median :2.100   Median :17.00
 Mean   : 81.75   Mean   :1009   Mean   :27.61                      Mean   :2.373   Mean   :16.58
 3rd Qu.: 93.00   3rd Qu.:1011   3rd Qu.:30.10                      3rd Qu.:3.300   3rd Qu.:21.00
 Max.   :100.00   Max.   :1014   Max.   :34.50                      Max.   :6.700   Max.   :25.00
```

Data Asset Description



Data Asset Description



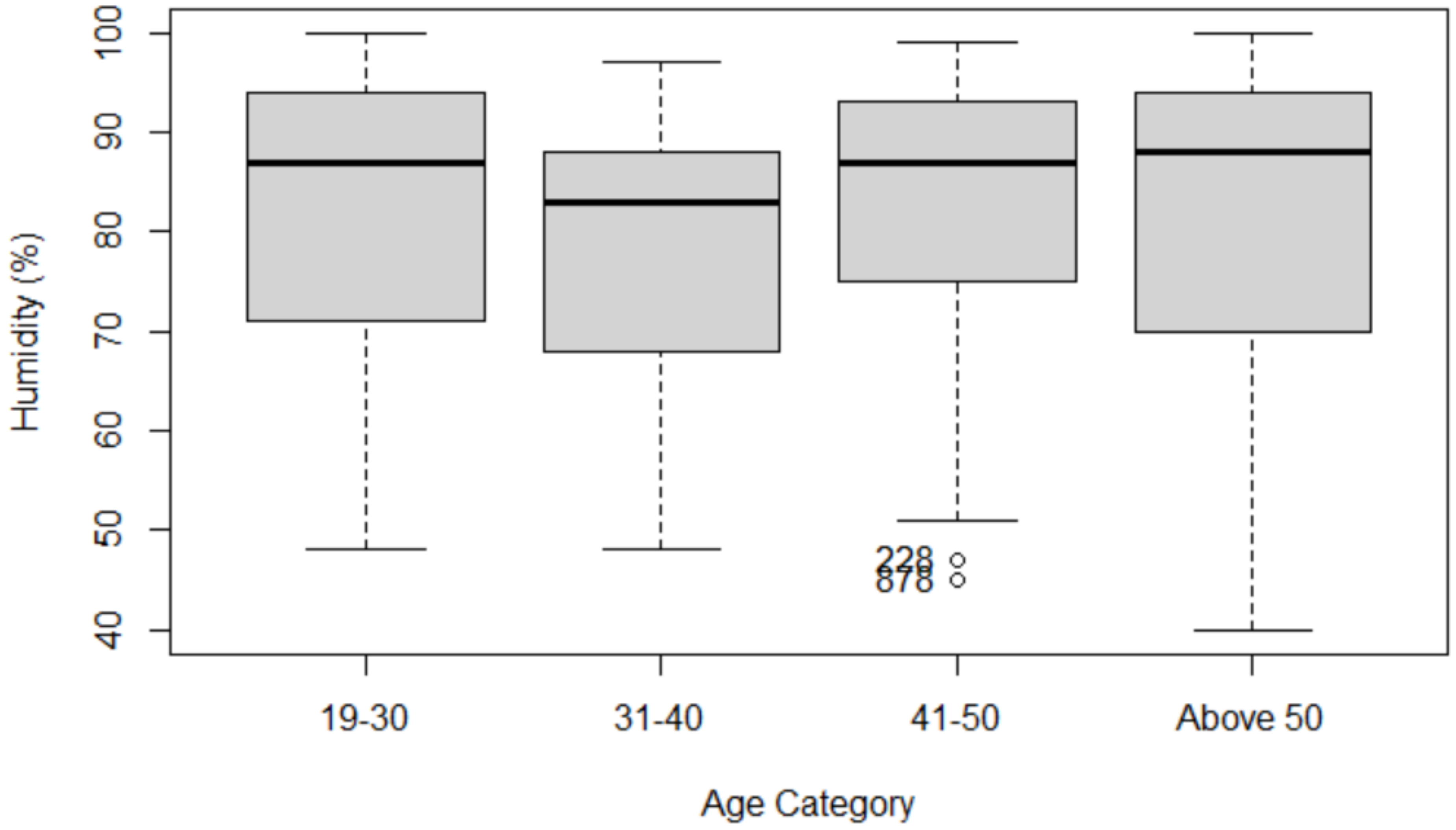
Data Preprocessing - Factoring

- Handling categorical data: lapply() function
- Converts the categories into numbers
- First category is considered 0 and each group afterward is a binary number
- Important for interpreting the intercept of the regression model
- Example: Outdoor Job

Category	Frequently	Occasionally	Rarely
Frequently	0	0	0
Occasionally	0	1	0
Rarely	0	0	1

Preprocessing – Outlier detection

- All categorical variables box plotted against continuous variables
- Any value outside of 1.5*IQR
- Values that showed up more than once deemed an outlier
  - 23 observations deemed as outliers



Humidity (%)
100
90
80
70
60
50
40
228
878
19-30
31-40
41-50
Above 50
Age Category

Preprocessing – Normalizing the data

- Standard Normal ~ N(0,1)
- Only for continuous variables
- Needed in order to be in form that is interpretable for a linear regression model

Model Update – Splitting the Data

- Kept 3 subjects out
- Less chance of bias when evaluating performance of test set
- Captures ruggedness
- More confidence that the model can accurately predict the severity of an attack for individuals that were not monitored for this study

Model Update – Mixed Effects Model

Model: $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$

- Easy to interpret
- Mixture Models: Incorporates random effects from each subject

- $\boldsymbol{Y}$ → Outcome
- $\boldsymbol{X}$ → Predictor variables
- $\boldsymbol{\beta}$ → Fixed effects regression coefficients
- $\boldsymbol{Z}$ → Random effects from each patient
- $\boldsymbol{u}$ → Random effects coefficients
- $\boldsymbol{\varepsilon}$ → noise

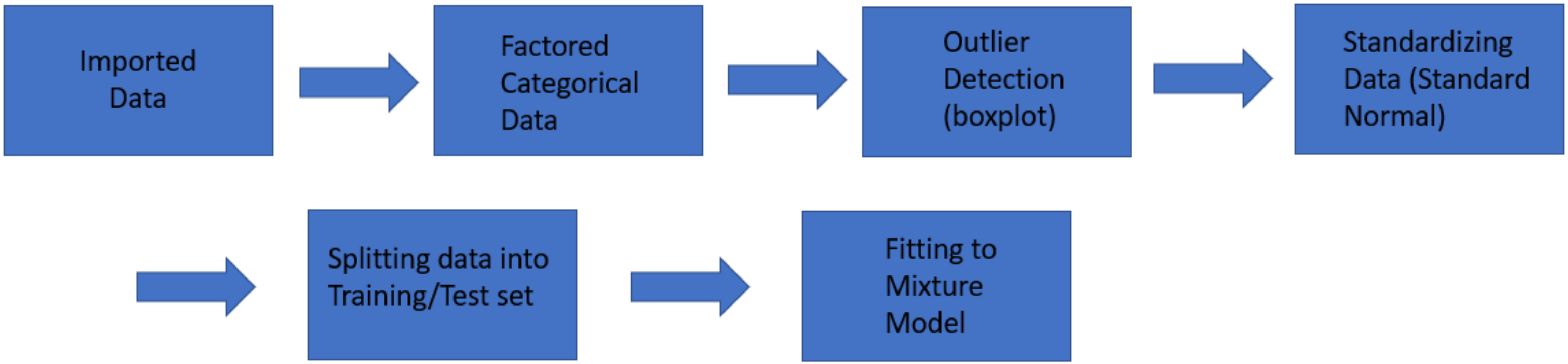
Model Update

```
fixed-effect model matrix is rank deficient so dropping 3 columns / coefficients
Linear mixed model fit by REML ['lmerMod']
Formula: ACT ~ wind + gender + age + ODJ + ODA + smoking + pressure +      temp + hum + (1 | subject)
   Data: train

REML criterion at convergence: 3827.2

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.75038 -0.56382  0.05585  0.66341  2.46544

Random effects:
 Groups   Name        Variance Std.Dev.
 subject  (Intercept) 1.895    1.377
 Residual             6.685    2.586
Number of obs: 811, groups:  subject, 7
```

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	14.18703	2.95972	4.793
wind	-1.49140	0.11404	-13.077
genderMale	1.77696	2.00193	0.888
age31-40	-2.51655	2.87676	-0.875
age41-50	0.18354	1.98708	0.092
ageAbove 50	0.80510	2.00593	0.401
ODJOccasionally	-0.31330	1.96376	-0.160
ODJRarely	0.92732	2.83601	0.327
pressure1004	1.21207	2.90687	0.417
pressure1005	3.97212	2.63786	1.506
pressure1006	3.10062	2.62912	1.179
pressure1007	2.62597	2.62804	0.999
pressure1008	0.77845	2.61577	0.298
pressure1009	1.80766	2.61074	0.692
pressure1010	2.61201	2.62584	0.995
pressure1011	0.89947	2.63178	0.342
pressure1012	-0.59421	2.64163	-0.225
pressure1013	-0.01395	2.65254	-0.005
pressure1014	0.98138	2.73770	0.358
temp	-0.01829	0.13625	-0.134
hum	-2.14252	0.13716	-15.620

Model Update – MLM



Next steps

- Determine best linear mixed-effects model through exhaustive search to predict asthma attacks
  - Based on R^2 and Akaike Information Criterion (AIC)

- Compare best linear mixed-effects model to a model that uses lasso regression with the linear mixed model

- Look into K-fold cross validation

- Build Rshiny feature

- Research previous studies done using weather features