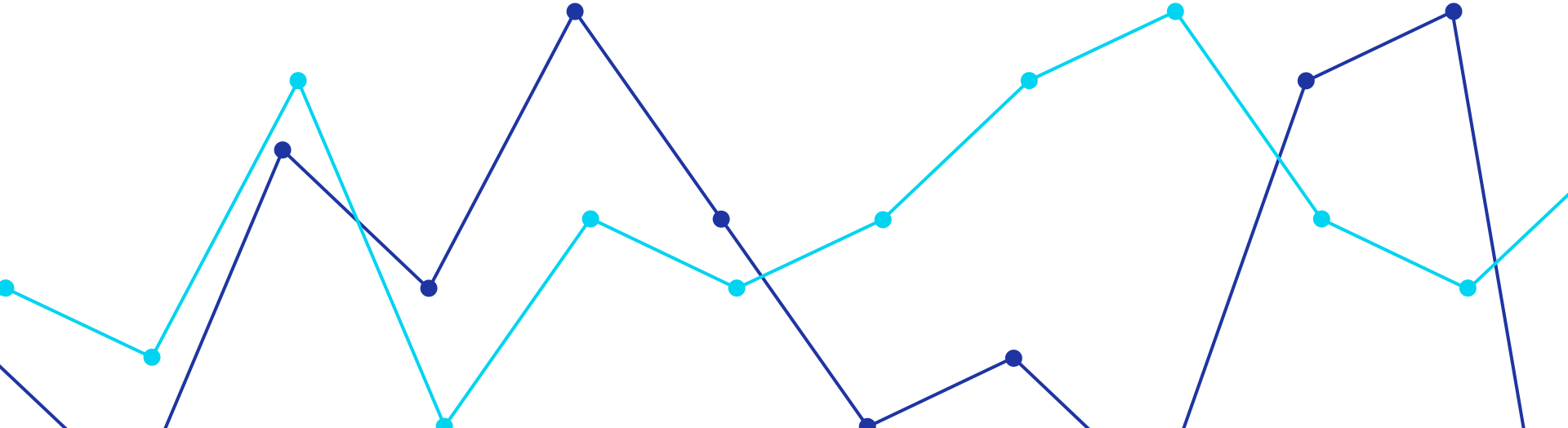


Evidation Health

By: Priscilla, Sarah, Ali, and Eric

Sponsor: Eric J. Daza

Mentor: Professor Franks

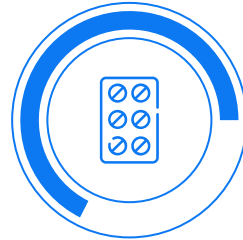


Our Motivation



Question

Which variables are most indicative of stress in our population?



Tier 1

Explore wearable data, and participant surveys (Adverse Childhood Experience, PTSD, Self Assessment Mannequin)



Tier 2

Create visualizations of data and run a logistic regression model to determine significant variables

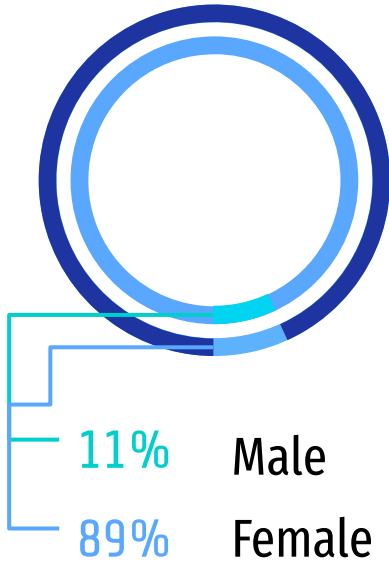


Tier 3

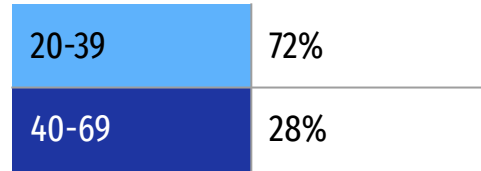
Create a Mixed Effects Model that can predict stress of participants

Our Population

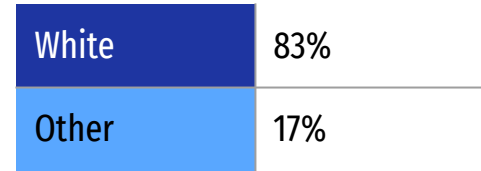
GENDER



AGE



Race



365

COVID-19 Healthcare
Workers

Our Data

	HR-AVG	HR-low	Breath-AVG	awake	light	rem	deep	Effcy	restless	S-Total	S-REM	S-Deep	S-Effcy	S-latency	S-Disturbances	S-Alignment	Temp-Delta	Shifts	Covid-Shift	ACE	PTSD	Gender	Age	Stressed
0	57.35	50.0	15.75	0.79	3.82	1.12	2.50	0.90	0.29	0.79	0.57	1.00	0.96	0.64	0.66	1.00	-0.17	0.0	0	0	32	0.0	29	0.0
1	62.21	55.0	16.00	0.62	4.16	1.17	2.52	0.93	0.29	0.85	0.58	1.00	0.98	0.72	0.76	1.00	0.22	1.0	1	0	32	0.0	29	1.0
2	61.81	55.0	15.50	0.82	4.02	1.74	2.85	0.91	0.26	0.96	0.88	1.00	0.96	0.94	0.80	0.88	0.26	0.0	0	0	32	0.0	29	0.0
3	67.13	58.0	15.62	1.54	3.88	1.37	1.60	0.82	0.27	0.71	0.69	0.95	0.76	0.01	0.80	1.00	0.15	0.0	0	0	32	0.0	29	0.0
4	62.80	56.0	15.75	1.51	5.60	1.43	1.69	0.85	0.35	0.97	0.72	0.96	0.83	0.19	0.64	0.98	-0.04	1.0	0	0	32	0.0	29	0.0

Data Units are the same as Raw Units

Data was in seconds, divided by 3600 to convert to hours

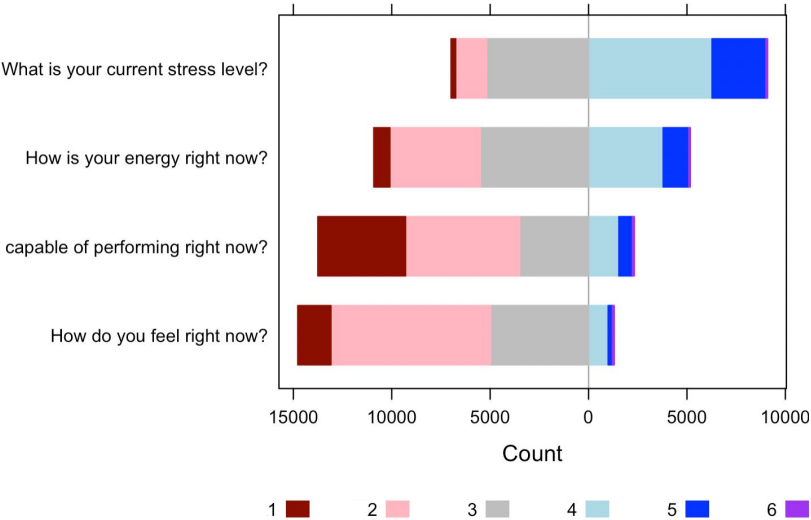
Data was a score between 0-100, scaled by 100 to be between 0-1

Originally, Female=1.0 and Male=2.0, encoded Female=0.0 and Male=1.0

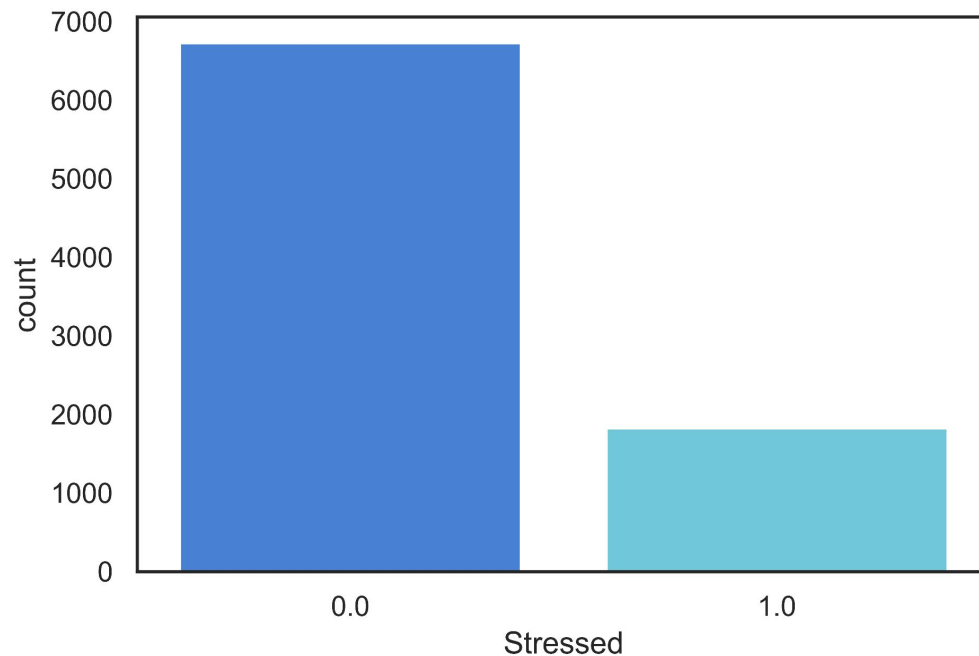
Data Limitations

Question	Variable Name	Variable Details
What is your current stress level?	sam8	🌀 Very Stressed (1); 😫 Stressed (2); 😐 Neutral (3); 😌 Calm (4); 😊 Very Calm (5); 🙄 No Answer (6)
How is your energy right now?	sam4	⚡⚡ Very Energetic (1); ⚡ Energetic (2); 🤔 Neutral (3); 😞 Low Energy (4); 😫 Very Low Energy (5); 🙄 No Answer (6)
What is your brain capable of performing right now?	sam6	All Complex Tasks 🧠💡💡💡💡 (1); Some Complex Tasks 🧠💡💡 (2); Moderate Tasks 🧠💡 (3); All Simple Tasks 🧠💡 (4); Some Simple Tasks 🧠💡 (5); 🙄 No Answer (6)
How do you feel right now?	sam2	😄 Very Positive (1); 😊 Positive (2); 😐 Neutral (3); 😞 Negative (4); 😫 Very Negative (5); 🙄 No Answer (6)

Likert Scale of SAM Data



Data Limitations



- Large Class Imbalance in Data
- Solutions? SMOTE, Undersampling, Oversampling, Logistic Regression

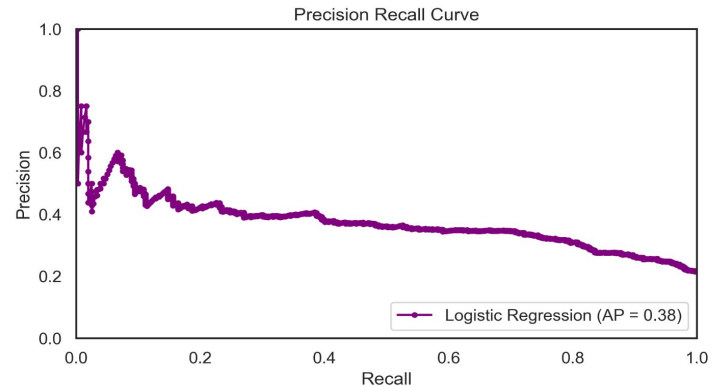
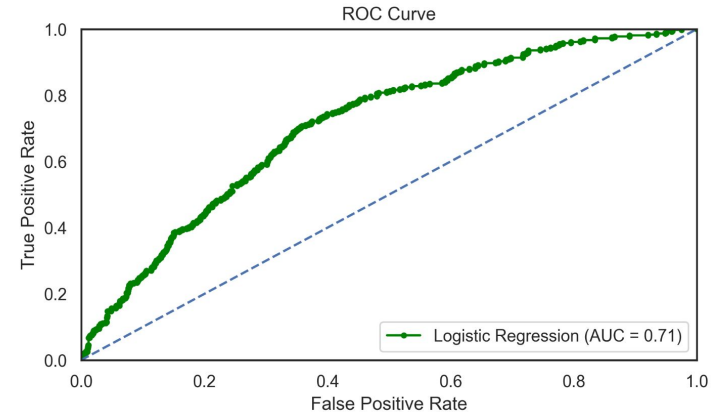
Our Logistic Regression Model

Results: Logit

```
=====
Model:          Logit          Pseudo R-squared: 0.092
Dependent Variable: daily_stressed AIC:          6392.7481
Date:           2022-05-22 08:33 BIC:          6481.4575
No. Observations: 6794          Log-Likelihood: -3183.4
Df Model:        12             LL-Null:       -3506.7
Df Residuals:    6781          LLR p-value:    1.1585e-130
Converged:       1.0000         Scale:        1.0000
No. Iterations:  6.0000
=====
```

	Coef.	Std.Err.	z	P> z	[0.025	0.975]
const	-3.3797	0.4299	-7.8616	0.0000	-4.2223	-2.5371
hr_average	0.0482	0.0117	4.1105	0.0000	0.0252	0.0711
hr_lowest	-0.0354	0.0126	-2.7986	0.0051	-0.0601	-0.0106
rem	-0.2609	0.0526	-4.9608	0.0000	-0.3640	-0.1578
restless	1.5677	0.3333	4.7035	0.0000	0.9144	2.2210
score_deep	-0.1497	0.1593	-0.9396	0.3474	-0.4618	0.1625
score_latency	-0.4726	0.2164	-2.1842	0.0289	-0.8968	-0.0485
score_alignment	-0.3603	0.1179	-3.0563	0.0022	-0.5913	-0.1292
daily_shifts	1.1581	0.0538	21.5436	0.0000	1.0527	1.2635
ace_total	0.0411	0.0152	2.7075	0.0068	0.0113	0.0709
ptsd_total	0.0134	0.0034	3.9439	0.0001	0.0067	0.0201
Gender	-0.8042	0.1677	-4.7963	0.0000	-1.1328	-0.4756
dem_age	0.0135	0.0032	4.1900	0.0000	0.0072	0.0199

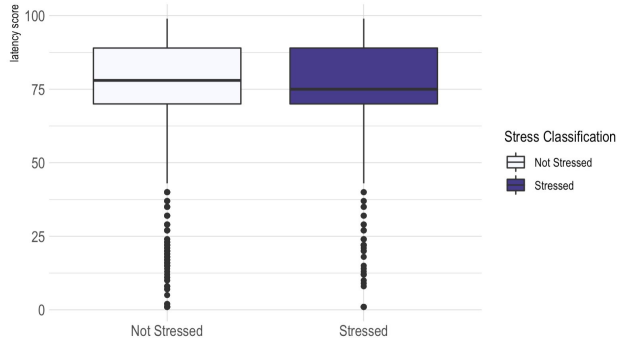
```
=====
```



Population Level Indicative Variables

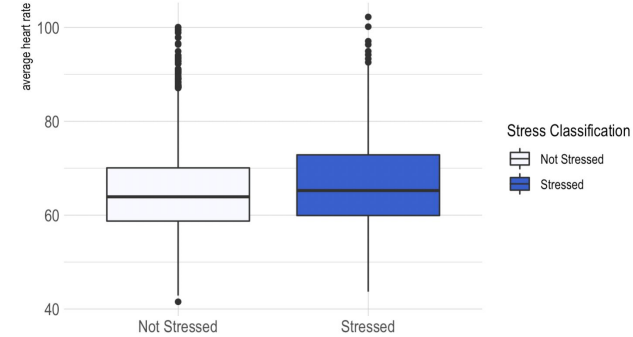
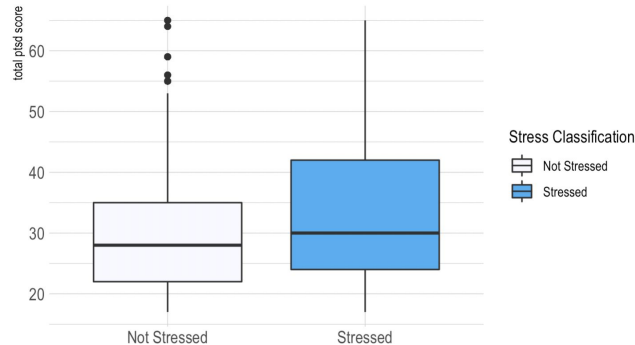
Latency Score

A score based on how long it takes for a participant to fall asleep. A time of 15 minutes gives the best score.



PTSD Survey

percentage of time a user was moving during their total sleep period

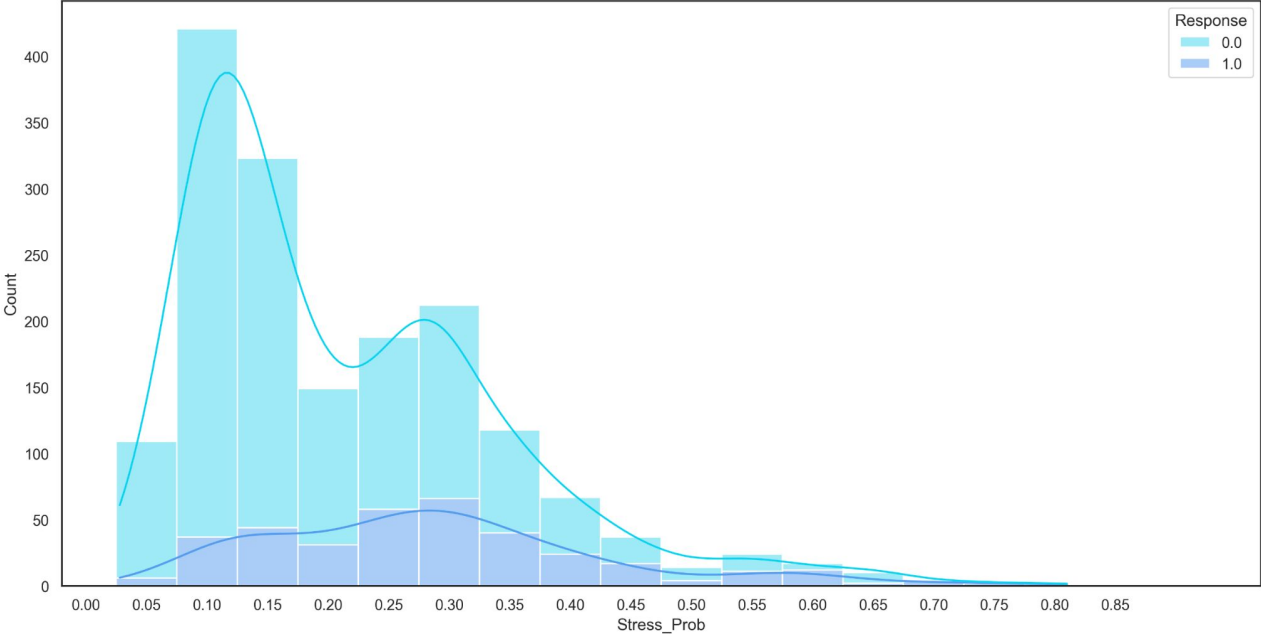


Average Heart Rate

participant's average heart rate captured during their sleep period

Chance of Being Stressed

	Constant	HR-AVG	HR-LOW	REM	Restless	Deep-S	Latency-S	Align-S	daily_shifts	ACE-S	PTSD-S	Gender	Age	Stress_Prob	Chance-Stress
5492	1.0	81.94	70.0	1.925000	0.33	0.96	0.67	0.97	1.0	8	49	0.0	31	0.464906	46.5%
7508	1.0	65.29	55.0	1.075000	0.38	0.78	0.89	1.00	0.0	2	27	0.0	57	0.175510	17.6%
8396	1.0	67.83	59.0	2.050000	0.21	0.92	0.75	0.89	0.0	0	34	0.0	32	0.088732	8.9%
6635	1.0	69.07	63.0	2.216667	0.22	0.96	0.89	1.00	1.0	6	24	0.0	39	0.234143	23.4%
380	1.0	77.79	67.0	2.541667	0.44	0.97	0.62	0.09	0.0	1	45	0.0	49	0.242452	24.2%



But what if each participant has a different baseline for stress?

Mixed Effects Models

What is a mixed effects model?

In general, a statistical model that combines random effects and fixed effects.

Random effects: Used when we think/know that there is variation within groups. Very useful in biostatistics! Can think of them as “nuisance variables” since we’re only interested in the fixed effects. Can be used to vary intercepts or slopes.

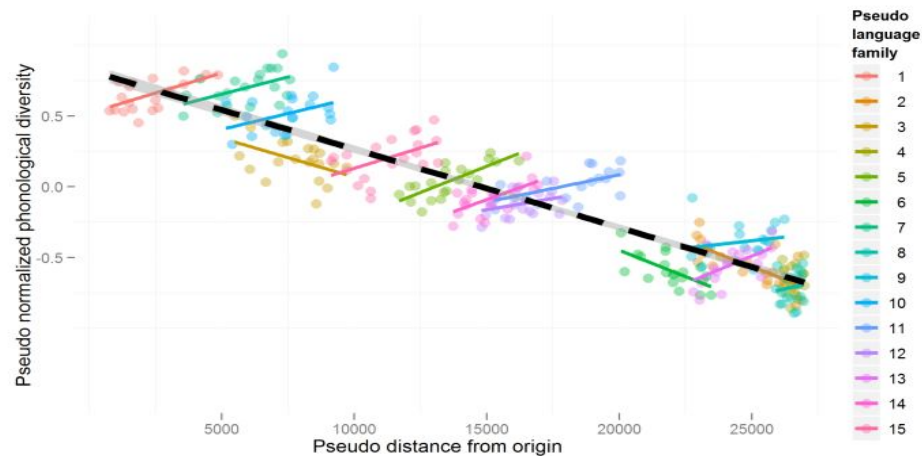
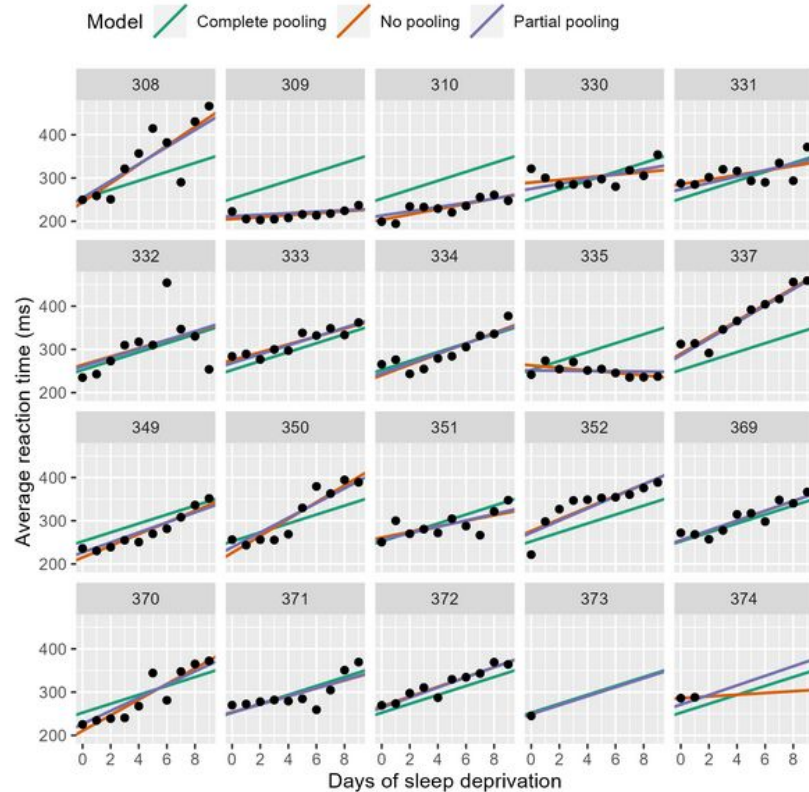
Fixed effects: Variables that stay constant within each group. These are effects that we are interested in analyzing. Ex. Predictors in a linear regression or logistic regression.

We’ll be dealing with generalized linear mixed models (GLMMs).

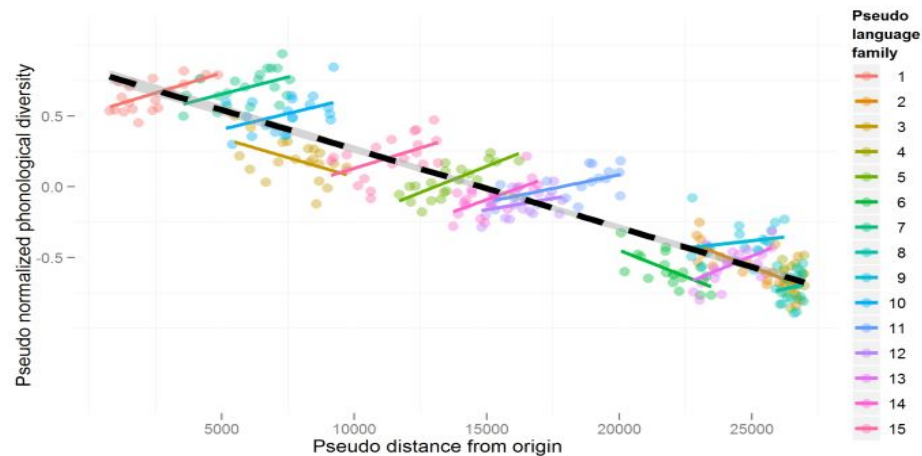
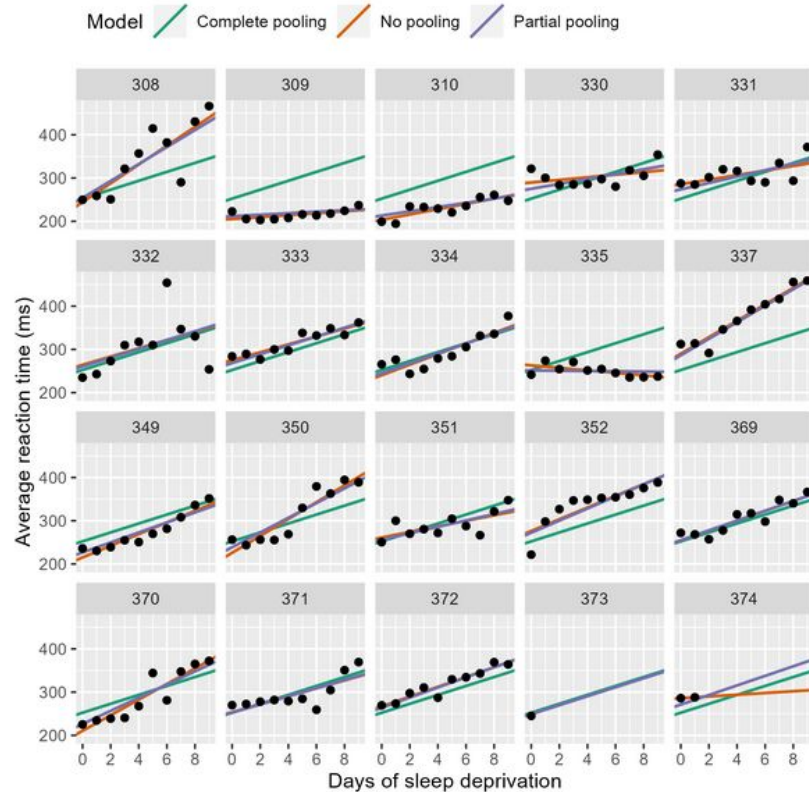
Brief Overview of GLMMs

- GLMMs are interested in capturing the variation within each group
 - If we specify a random intercept based on participant id. their deviations from the mean are assumed to be normally distributed, $N(0, \sigma_b^2)$
 - Soft restriction
 - Makes computation a lot easier
 - Extreme values will be shrunk towards 0 (partial pooling)
- Very useful for our data because participants have a different baseline for how stressed they are, among other factors
 - In order to have a complete study of our population, we need to account for these variations
 - Logistic regression cannot account for this directly, since it is a generalized model (complete pooling)
 - We tried adding in participant id as a predictor, but coefficients can blow up rapidly for low sample participants (no pooling)

GLMMs are Great!

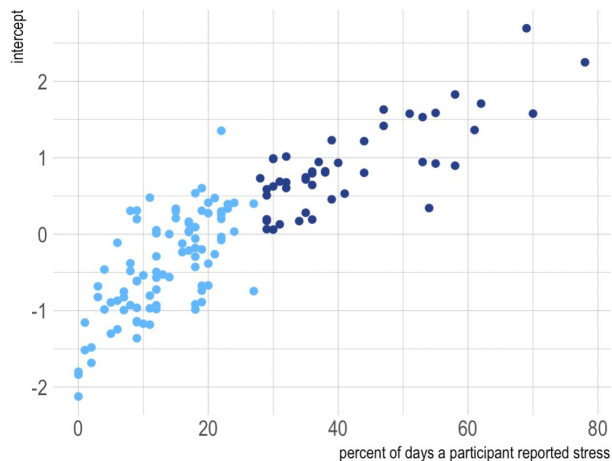


GLMMs are Great!



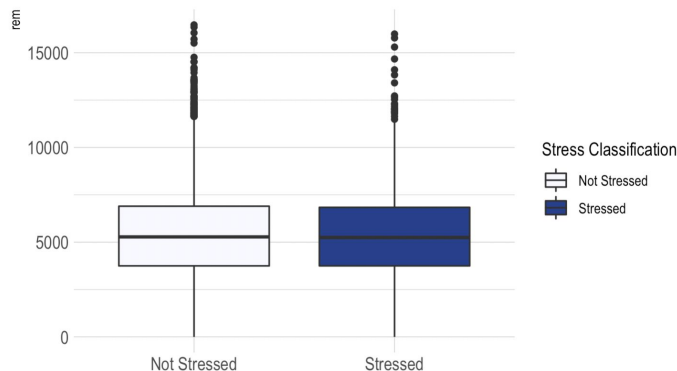
Our Mixed Effects Model

Random Intercept by participant ID vs. percent days stressed



Model Coefficients

Variables that are not indicative of stress at the population level become significant once a participants baseline is adjusted for ex. REM



Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.36374	0.46156	-2.955	0.00313	**
hr_average	0.13059	0.05519	2.366	0.01796	*
restless	0.03360	0.05001	0.672	0.50175	
daily_shifts	1.24630	0.05391	23.116	< 2e-16	***
dem_gender	-1.03670	0.41425	-2.503	0.01233	*
rem	-0.10512	0.03718	-2.828	0.00469	**
total_ace	0.10263	0.04202	2.443	0.01458	*

Questions?