

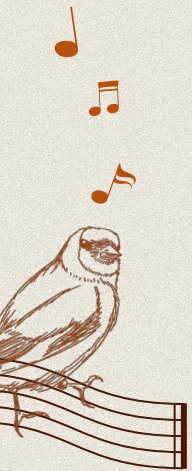
# *Introduction to Targeted Maximum Likelihood Estimation*

Xiqian Yuan  
Mentor: Nolan Cole



# How can we do inference with machine learning?

- Targeted Maximum Likelihood Estimation (TMLE) is a solution
- Pairing TMLE with causal inference allows for causal discovery



# Outline

1

*Algorithm*



2

*Application*



3

*Conclusion*



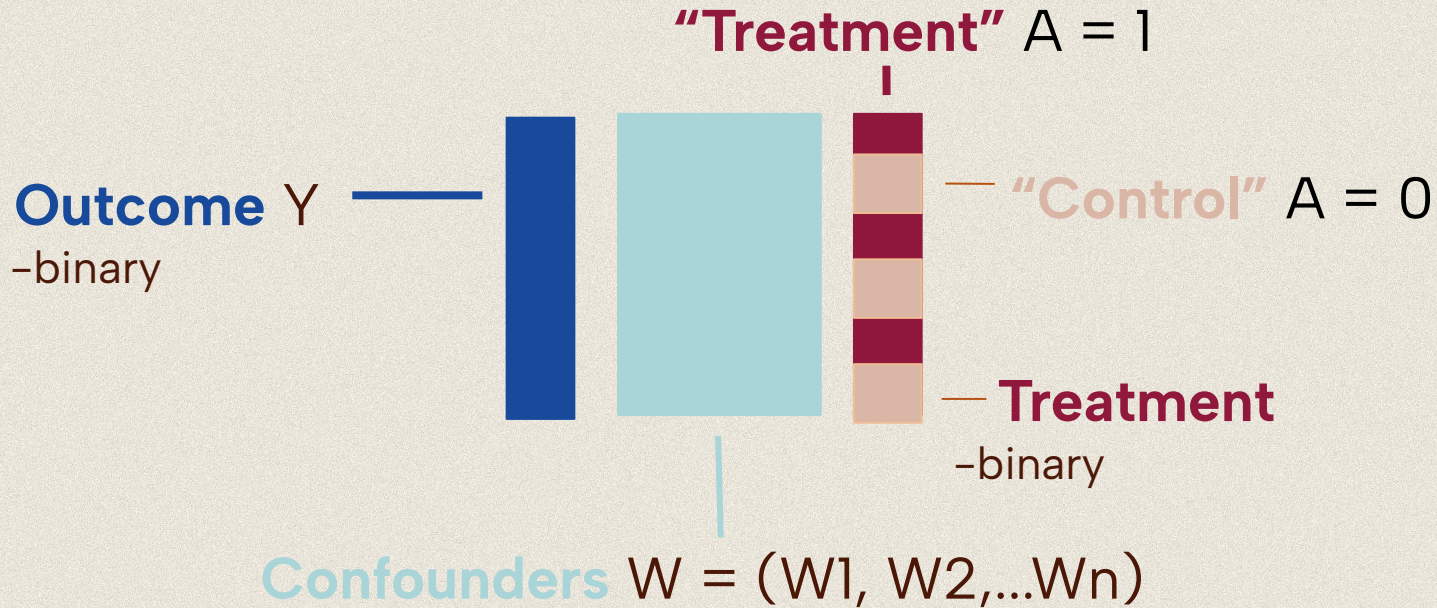


01

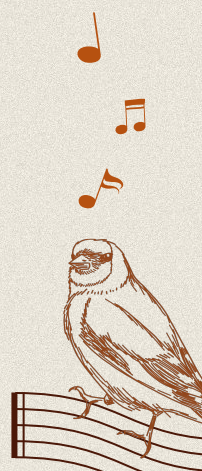
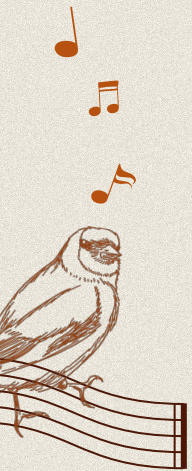
# Algorithm

How do we do TMLE?

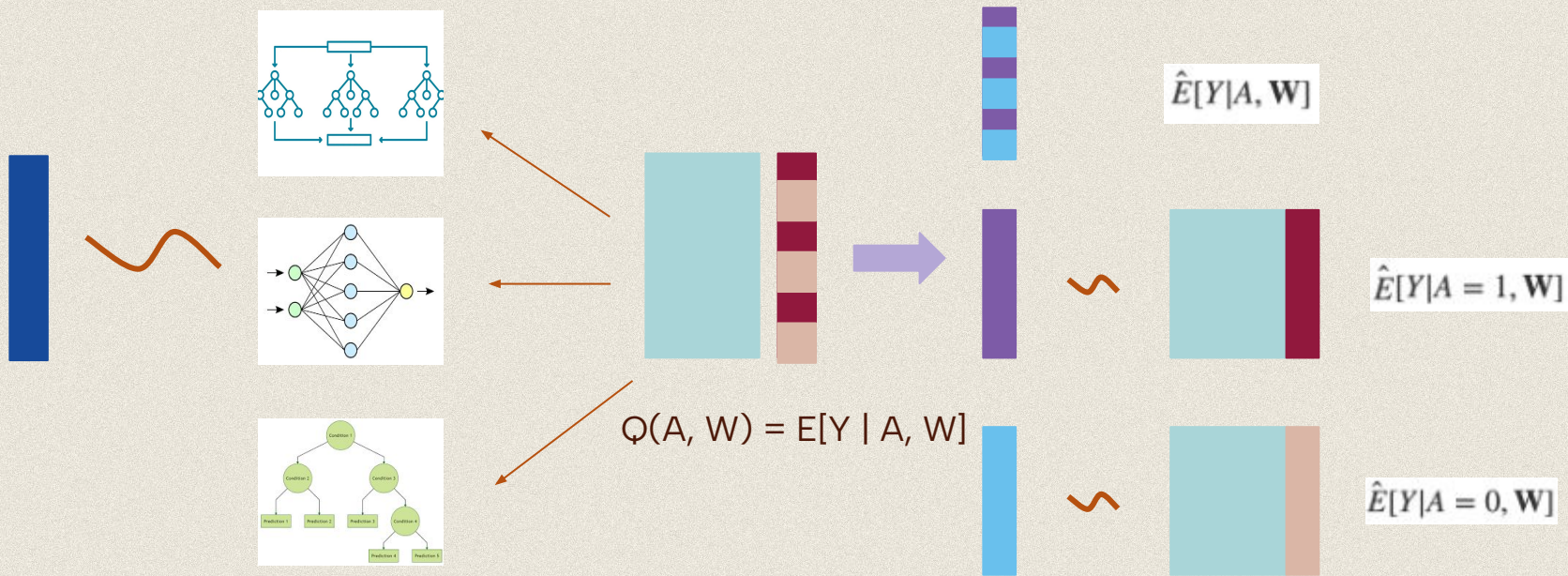
# Data



$$\text{ESTIMAND: ATE} = E[E[Y \mid A = 1, W]] - E[E[Y \mid A = 0, W]]$$

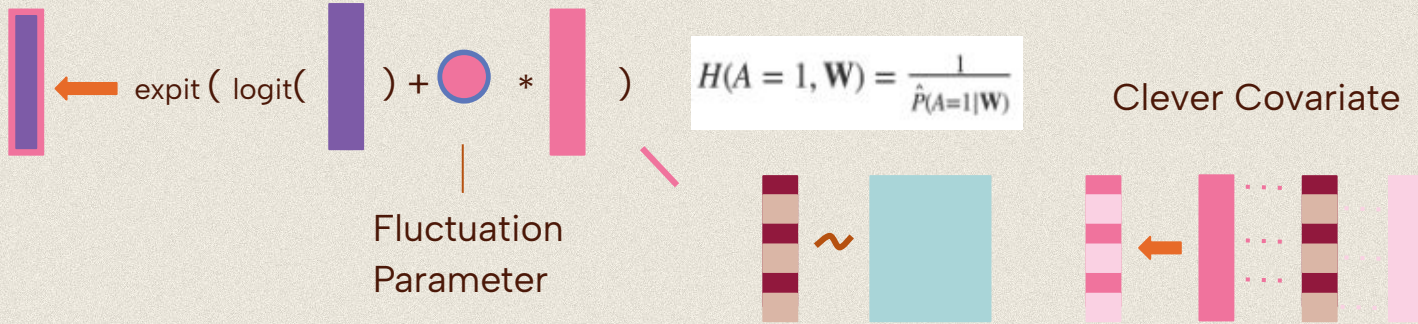


# 1. Get Initial Estimate



# 2. Update Initial Outcomes

$$\hat{E}^*[Y|A = 1, \mathbf{W}] = \text{expit}(\text{logit}(\hat{E}[Y|A = 1, \mathbf{W}]) + \hat{e}H(1, \mathbf{W}))$$



$$g(\mathbf{W}) = P(A = 1 | \mathbf{W})$$

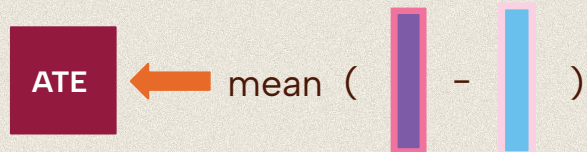


$$H(A = 0, \mathbf{W}) = -\frac{1}{\hat{P}(A=0|\mathbf{W})}$$

$$\hat{E}^*[Y|A = 0, \mathbf{W}] = \text{expit}(\text{logit}(\hat{E}[Y|A = 0, \mathbf{W}]) + \hat{e}H(0, \mathbf{W}))$$

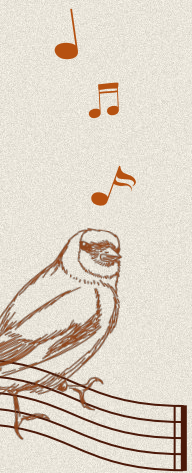


# 3. Inference



$$\hat{ATE} = \hat{E}[\hat{E}^*[Y|A = 1, \mathbf{W}] - \hat{E}^*[Y|A = 0, \mathbf{W}]]$$

- Semiparametric – combination of parametric and nonparametric model
- Doubly Robust!!
- Asymptotic Normality of Influence Function allows for construction of CI and test – inference !!







02

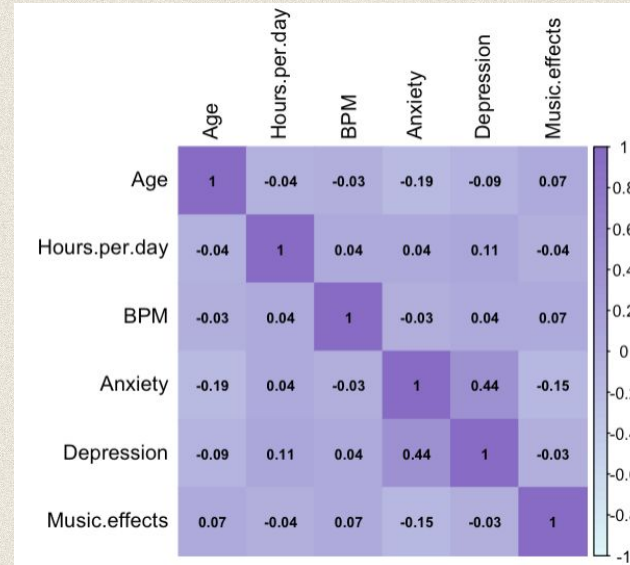
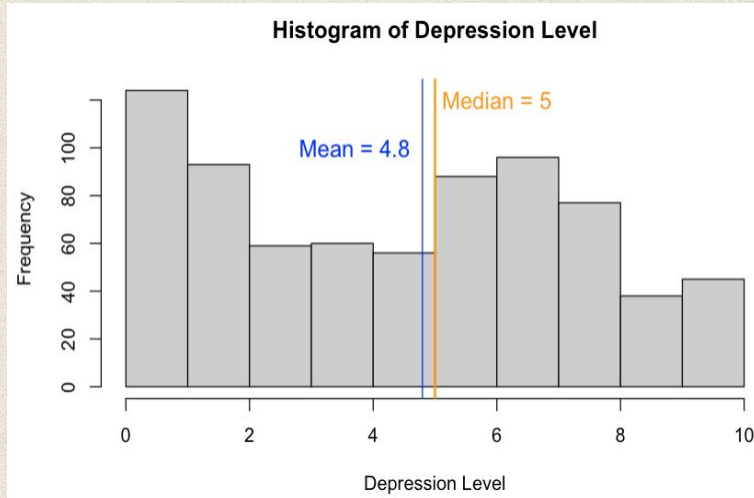
# Application

Music & Mental Health



# Music & Mental Health

- Music & Mental Health Survey from Kaggle
- 624 observations



# Data Structure

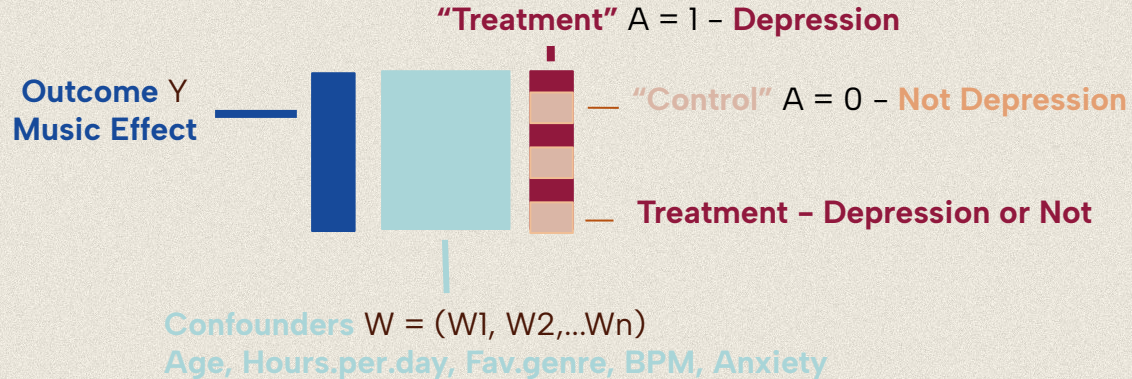


Define

Outcome Y - Music Effect (0 - Improve; 1 - Not Improve)

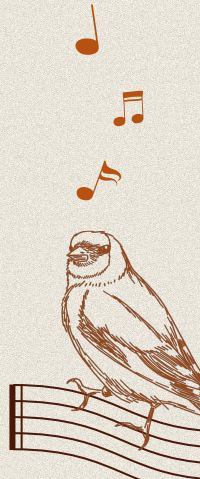
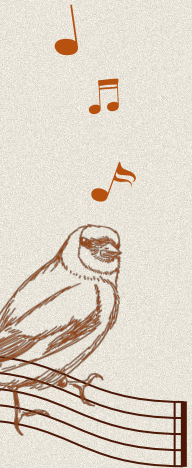
Exposure A - Depression (0 - Not Depressed; 1 - Depressed)

Confounders W - Age, Hours.per.day, Fav.genre, BPM, Anxiety



# Result – Differences in Means

Parameter Estimate	0.01546
Estimated Variance	0.0010993
p-value	0.64101
95% Conf Interval	(-0.049524, 0.080444)



03

# Conclusion



# Conclusion

- Inference with Machine Learning Model
- Doubly Robust Estimation
- Advancement in Causal Inference



# Reference



- Hoffman, K. (2020, December 11). *An illustrated guide to TMLE, part II: The Algorithm*. KHstats.  
<https://www.khstats.com/blog/tmle/tutorial-pt2>
- RASGAITIS, C. (n.d.). Music & Mental Health Survey Results. Kaggle.  
[https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results?select=mxmh\\_survey\\_results.csv](https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results?select=mxmh_survey_results.csv)
- M Van Der Laan, S Rose Targeted Learning: Causal Inference for Observational and Experimental Data (Springer, New York, 2011).