State-level Food Waste Policies In the U.S.: A Predictive Modelling

The Hotel School
Cornell
SC Johnson College of Business

GIZEM KILIC

CORNELL UNIVERSITY

SCHOOL OF HOTEL ADMINISTRATION

Purpose/Objective

The purpose of the initial stage of this research is to predict the relationship between rigidity of the food waste policies and demographic, political, economic, food safety, and environmental features at state level.

Research questions include:
- Which policy category (prevention, recycling, recovery, and/or all) has the best performance in predicting regulatory strictness in food waste policies across states?
- Which model has the best performance metrics in predicting regulatory strictness in food waste policies across states?
- Which variables have the highest importance in predicting regulatory strictness in food waste policies across states?

Introduction

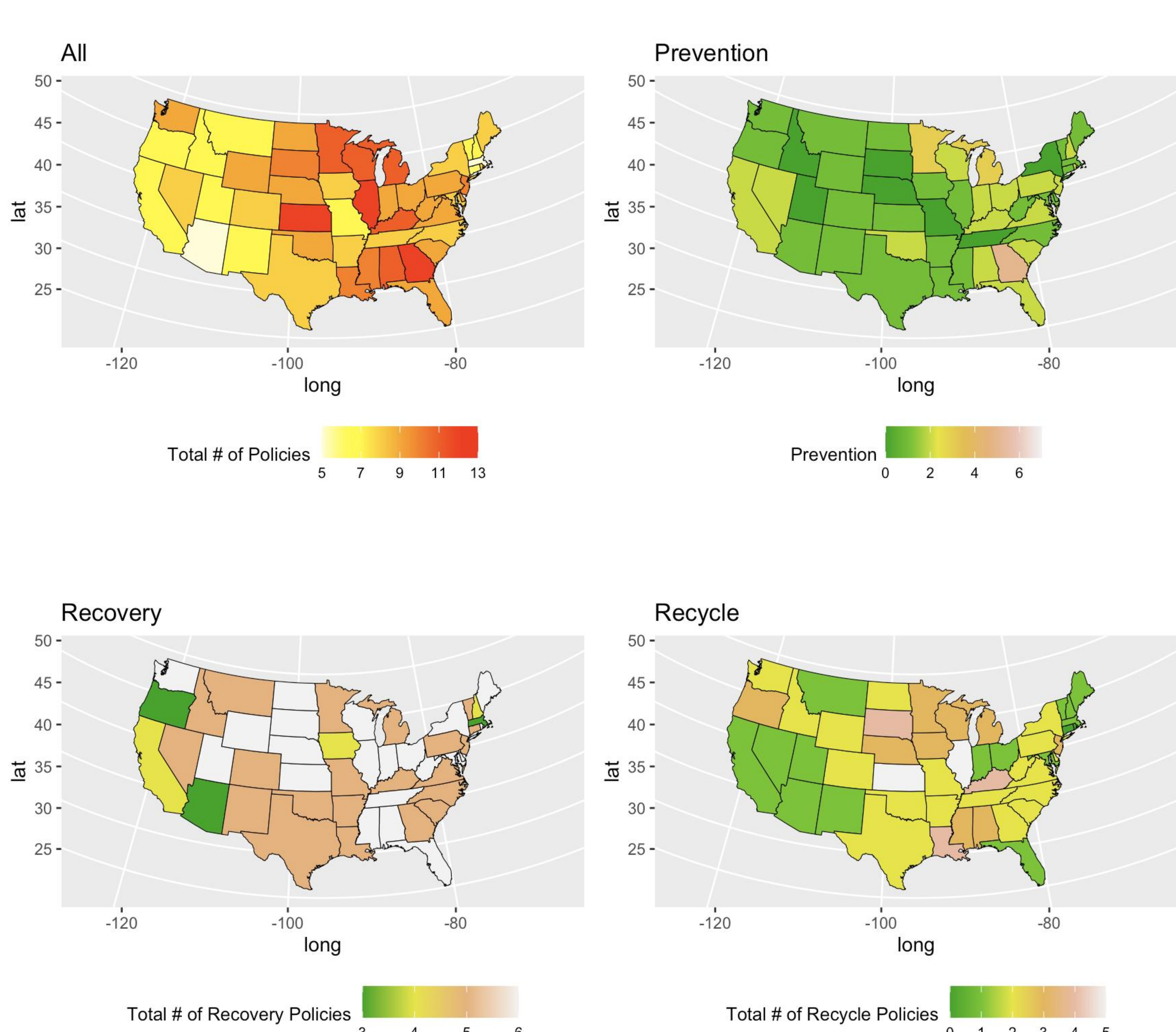
Prevention (Date Labelling):
- No uniform federal standards
- Label indicators
  - Safety vs. quality
- State restrictions

Recovery (Food donations):
- Liability protection
  - Federal civil and criminal
  - Additional state liability protection
- Tax deduction and credit
- Direct donations

Recycling (Animal Feed):
- Reusing food scraps
  - Heat treatment
  - Liability protection for qualified direct donors
- Restriction on animal protein
- Food safety controls
- Labelling and adulteration

Recycling (Waste Laws):
- No federal law
- State and municipality bans
  - Organic waste bans
  - Waste recycling laws

Figure #1. U.S. Heat Map of Level of Regulatory Strictness by Policy Category



Methodology

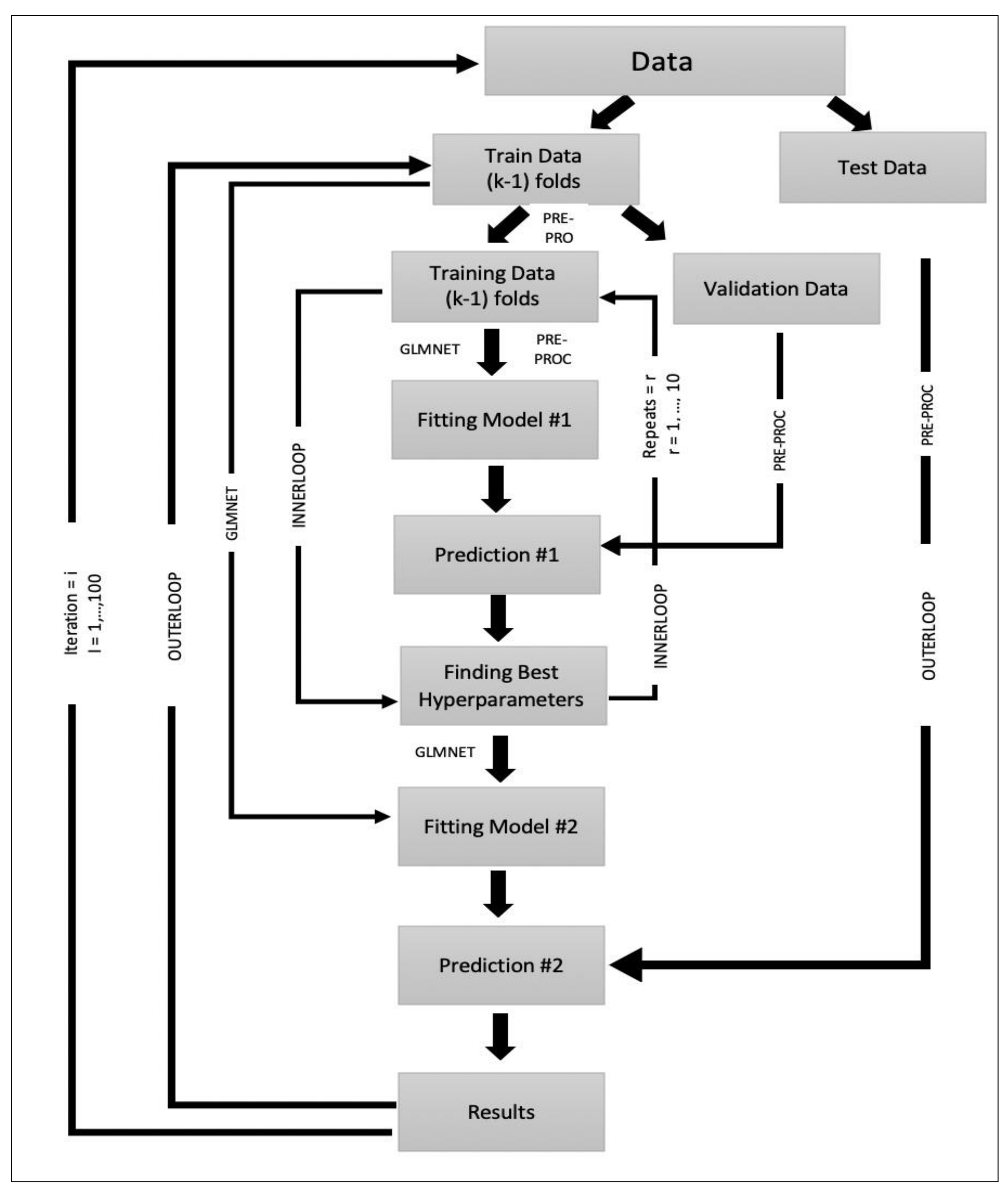
Generalized Linear Model Elastic Net (GLMNET) is a type of regularized linear regression model that incorporates tuning parameters for variable selection and shrinkage. The paper compares 3 levels of k-fold cross validation models to select the best model with optimal performance metrics. Too high penalty levels lead to simple model and underfitting[1,2,3,4,5].

$$\frac{\sum_{i=1}^{n}(y_i - x_i^J\hat{\beta})^2}{2n} + \lambda\left(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j=1}^{m}|\hat{\beta}_j|\right)$$

- Alpha ⟶ convexity ⟶ 0.0 – 1.0
- Lambda ⟶ degree of penalty ⟶ 0.0 – 1.0

Model Selection [6]
- Prevention, recovery, recycling, all categories
- 5, 10, LOOCV k-fold, repeated 10x

Figure #2. GLMNET Process Flow Chart



Performance metrics, iteration 100x [3,4,6]
- RMSE
- R-square

Wilcoxon Sum Rank Test [6,8,9,10]
- Non-parametric version of the two-sample t-test
- Ordinal level
- Normal distribution not required

Hypothesis 1:	Hypothesis 2:	Hypothesis 3:
$H_0: M_{i1} = M_{i3}$	$H_0: M_{i1} = M_{i2}$	$H_0: M_{i2} = M_{i3}$
$H_A: M_{i1} \neq M_{i3}$	$H_A: M_{i1} \neq M_{i2}$	$H_A: M_{i2} \neq M_{i3}$

Significant difference between performance of two models if P < 0.05 [32,36]. Mi1:10 k-fold, Mi2: LOOCV (50 k-fold), Mi3: 5 k-fold. Policy categories: i , for i = 1,2,3, and 4, which respectively represents sum of all categories, prevention, recovery, and recycling policy categories .

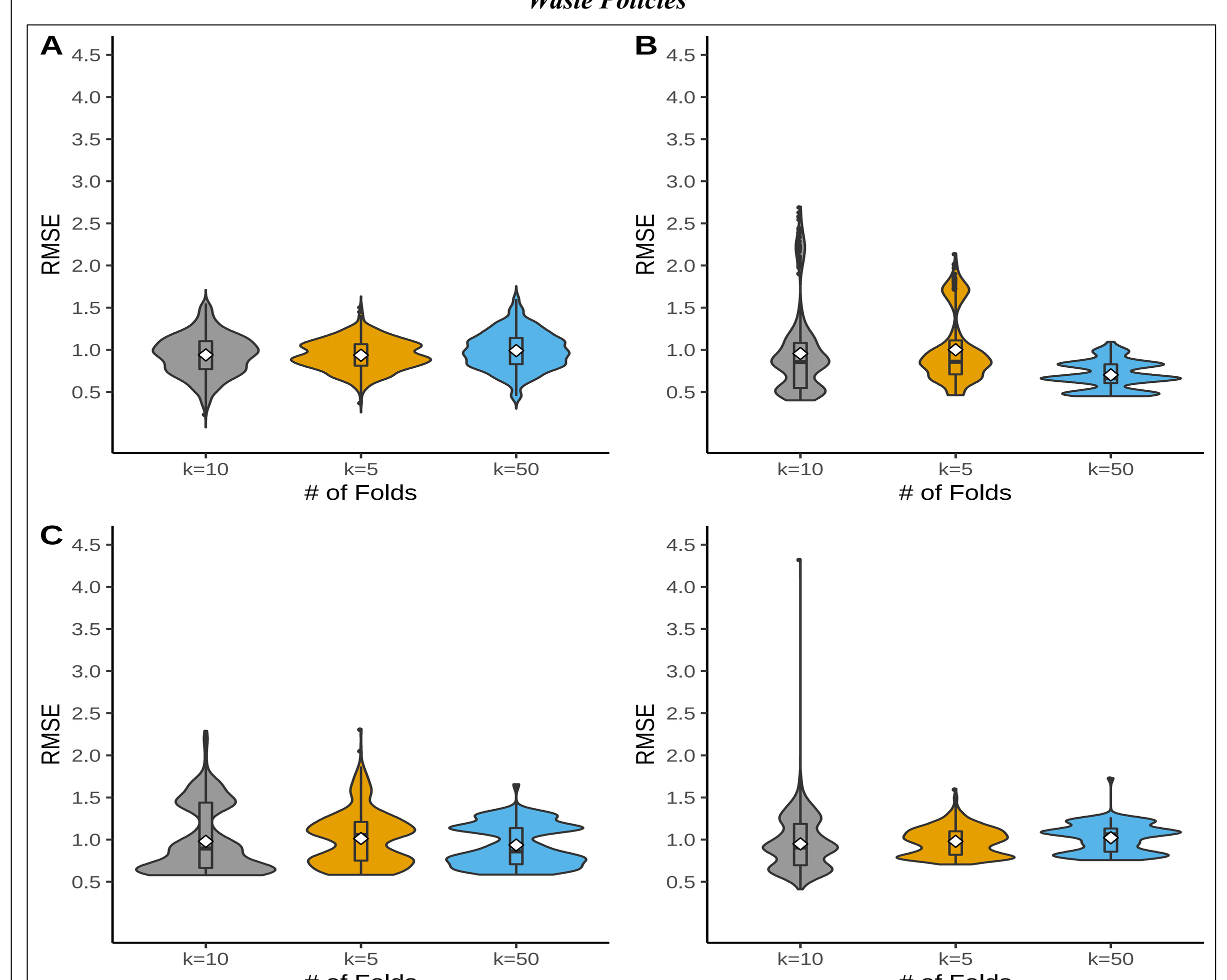
Findings

Figure #3. The R2 and RMSE Results of the K-Fold Methods in Predicting State Level Policies

	K-Fold	All	Prevention	Recovery	Recycling
R2	M_{i2}	0.1802 (0.0887, 0.2755)	0.0348 (0.0078, 0.1233)	0.0661 (0.0172, 0.1458)	0.1282 (0.053 0.2550)
	M_{i1}	**0.2105 (0.0647, 0.4337)**	0.0885 (0.0243, 0.2533)	**0.1995 (0.0634, 0.4386)**	**0.2387 (0.066 0.4672)**
	M_{i3}	0.1584 (0.0648, 0.3173)	**0.0756 (0.0244, 0.1273)**	0.0532 (0.0137, 0.1629)	0.1375 (0.048 0.3052)
RMSE	M_{i2}	**0.9220 (0.8111, 1.0651)**	0.8584 (0.7071, 1.1100)	1.0543 (0.7515, 1.2091)	0.9845 (0.820 1.0987)
	M_{i1}	0.9547 (0.7694, 1.1014)	**0.8527 (0.5451, 1.0827)**	0.8942 (0.6630, 1.4397)	**0.9061 (0.694 0.9489)**
	M_{i3}	0.9823 (0.8286, 1.1410)	0.6632 (0.6035, 0.8276)	**0.8591 (0.7079, 1.1368)**	1.0618 (0.856 1.1318)

- M_{i1}1:10 k-fold, M_{i2}: LOOCV (50 k-fold), M_{i3}3: 5 k-fold
- Policy categories are denoted as i , for i = 1,2,3, and 4, which respectively represents sum of all categories, prevention, recovery, and recycling policy categories

Figure #4. Distribution of RMSE Results of Regression Analysis in Predicting State Level Food Waste Policies



- (A) All, (B) Prevention, (C) Recovery, and (D) Recycling.
- Median, mean, first and third quartiles, mean values of the RMSE distribution
- Policy categories are calculated by 3 k-fold CV elastic net methods

5, 10, 50 k-fold

Figure #5. Wilcoxon Sum Rank Test by R2 and RMSE Results of the K-Fold Methods

	K-Fold Comparison	All	Prevention	Recovery	Recycling
R2	$\mathbf{M_{i1}} = \mathbf{M_{i3}}$	**0.0168***	**0.0432***	**0.0000****	**0.0003****
	$M_{i2} = \mathbf{M_{i3}}$	0.5905	0.1797	0.8940	0.7249
	$\mathbf{M_{i1}} = M_{i2}$	**0.0001****	**0.0000****	**0.0000****	**0.0000****
RMSE	$\mathbf{M_{i1}} = \mathbf{M_{i3}}$	0.0640	**0.0000****	0.4748	**0.0012****
	$M_{i2} = \mathbf{M_{i3}}$	**0.0303***	**0.0000****	0.1145	**0.0095***
	$\mathbf{M_{i1}} = M_{i2}$	0.6499	**0.0009****	**0.0000****	**0.0002****

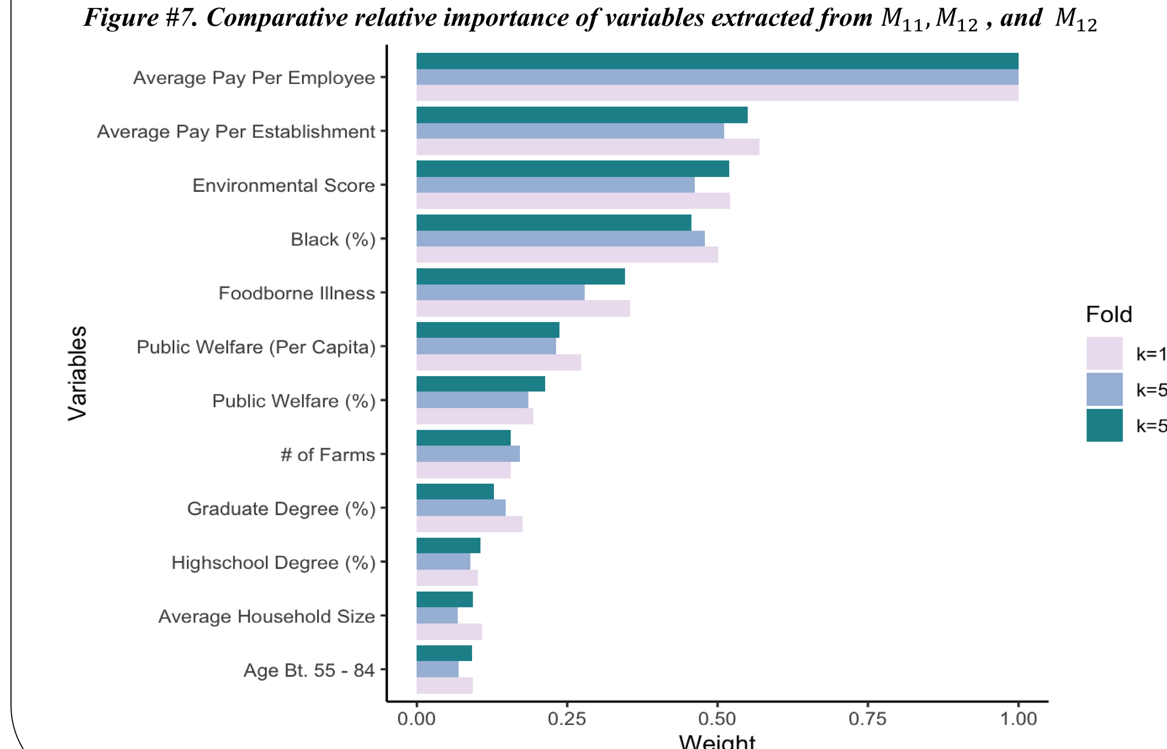
(* p value < 0.05 and ** p value < 0.005. $M_{1(1,2,3)}$:10 k-fold, $M_{2(1,2,3)}$:: LOOCV (50 k-fold), $M_{3(1,2,3)}$: 5 k-fold)

Findings

Figure #76 Coefficient Table of 12 Most Important Variables

	k=10	k=5	k=50
Average Pay Per Employee	-0.1661	-0.1610	-0.1820
Average Pay Per Establishment	-0.0946	-0.0823	-0.1001
Environmental Score	-0.0864	-0.0743	-0.0945
Black (%)	0.0832	0.0771	0.0832
Foodborne Illness	0.0590	0.0450	0.0630
Public Welfare (Per Capita)	-0.0454	-0.0372	-0.0432
Public Welfare (%)	-0.0321	-0.0299	-0.0388
Graduate Degree (%)	-0.0291	-0.0238	-0.0233
# of Farms	0.0260	0.0275	0.0285
Average Household Size	-0.0179	-0.0110	-0.0169
Highschool Degree (%)	0.0168	0.0142	0.0193
Age Bt. 55 - 84	-0.0156	-0.0112	-0.0167

Figure #7. Comparative relative importance of variables extracted from M_{11}, M_{12} , and M_{12}



Results

Figure #8. Policy Implications

Grocery stores impact on state decision
- Higher financial contribution to state finances
- Higher salary per employee + Higher total salary and larger size
- Opportunities for initiatives

Higher environmental score
- Less food waste
- Opportunity for lifting the restriction on the food labels and animal feed and adding extra incentive for food donation

Higher African American population
- Restrictions discourage donation & secondary markets
- Liability concerns
- Secondary food markets enable affordable food
- Food label restrictions hinder the growth of secondary food markets
- Increase awareness of the food insecurity cycle

Foodborne illness concerns
- Stricter food donation and food label regulations
- The perception of state government is parallel to the perception consumers
- Freshness concerns that drive consumers to throw away food
- Date labels not a good proxy for foodborne illness threats
- Businesses hesitant to donate food due to liability concerns

References:
[1] Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, *28*(2), 29–50. https://doi.org/10.1257/jep.28.2.29
[2] Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634
[3] Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
[4] Hastie, T., Tibshirani, R., & Friedman, J. (2020). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)* (2nd ed.). Springer.
[5] Buhlmann, P., & Geer, V. S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications (Springer Series in Statistics)* (2011th ed.). Springer.
[6] Tozlu, C., Edwards, D., Boes, A., Labar, D., Tsagaris, K. Z., Silverstein, J., Pepper Lane, H., Sabuncu, M. R., Liu, C., & Kuceyeski, A. (2020). Machine Learning Methods Predict Individual Upper-Limb Motor Impairment Following Therapy in Chronic Stroke. *Neurorehabilitation and Neural Repair*, *34*(5), 428–439. https://doi.org/10.1177/1545968320909796
[7] Snedecor, G. W., & Cochran, W. G. (1989). *Statistical Methods* (8th ed.). Iowa State University Press.
[8] McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. *The Corsini Encyclopedia of Psychology*. https://doi.org/10.1002/9780470479216.corpsy0524
[9] Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491
[10] Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, *1*(6), 80. https://doi.org/10.2307/3001968