

The background features a complex network diagram with numerous nodes of varying sizes (dark blue, light blue, and grey) connected by thin grey lines. Some nodes are highlighted with larger concentric circles. The overall aesthetic is modern and technical.

CAUSAL ANALYSIS: HOME NURSES

Jonah Winninghoff

OVERVIEW



STAR RATING SYSTEM ESTABLISHED IN 2009 TO HELP FAMILY MEMBERS TO DECIDE WHICH HOME NURSE THEIR SENIOR FAMILY MEMBERS LIVE IN.



THE HYPOTHESIS TESTING IS UNABLE TO CONFIRM THAT NUMBER OF COVID-19 DEATHS AT FIVE-STAR FACILITIES ARE SIGNIFICANTLY DIFFERENT FROM ONE-STAR FACILITIES.



THE U.S. CENTERS FOR MEDICARE AND MEDICAID SERVICES (CMS) MAY FACE A MASSIVE DECLINE IN PUBLIC TRUST.

WHAT TO DO?

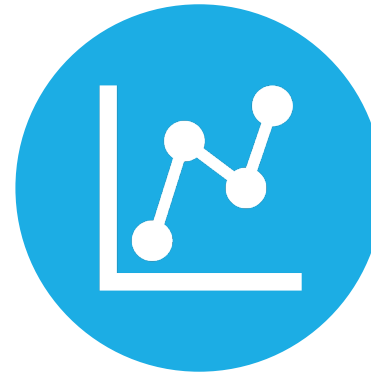


THE CAUSAL ANALYSIS IS A LINCHPIN IN ADDRESSING THIS PROBLEM.

TWO METHODS



DATA SCIENCE



ECONOMETRICS

TWO METHODS



DATA SCIENCE

Data Science:

Test the model against the data.

Econometrics:

Test the data against the model.

TWO DATASETS



MINIMUM DATA SET
QUALITY MEASURES



PROVIDER INFORMATION

TWO DATASETS



MINIMUM DATA SET
QUALITY MEASURES

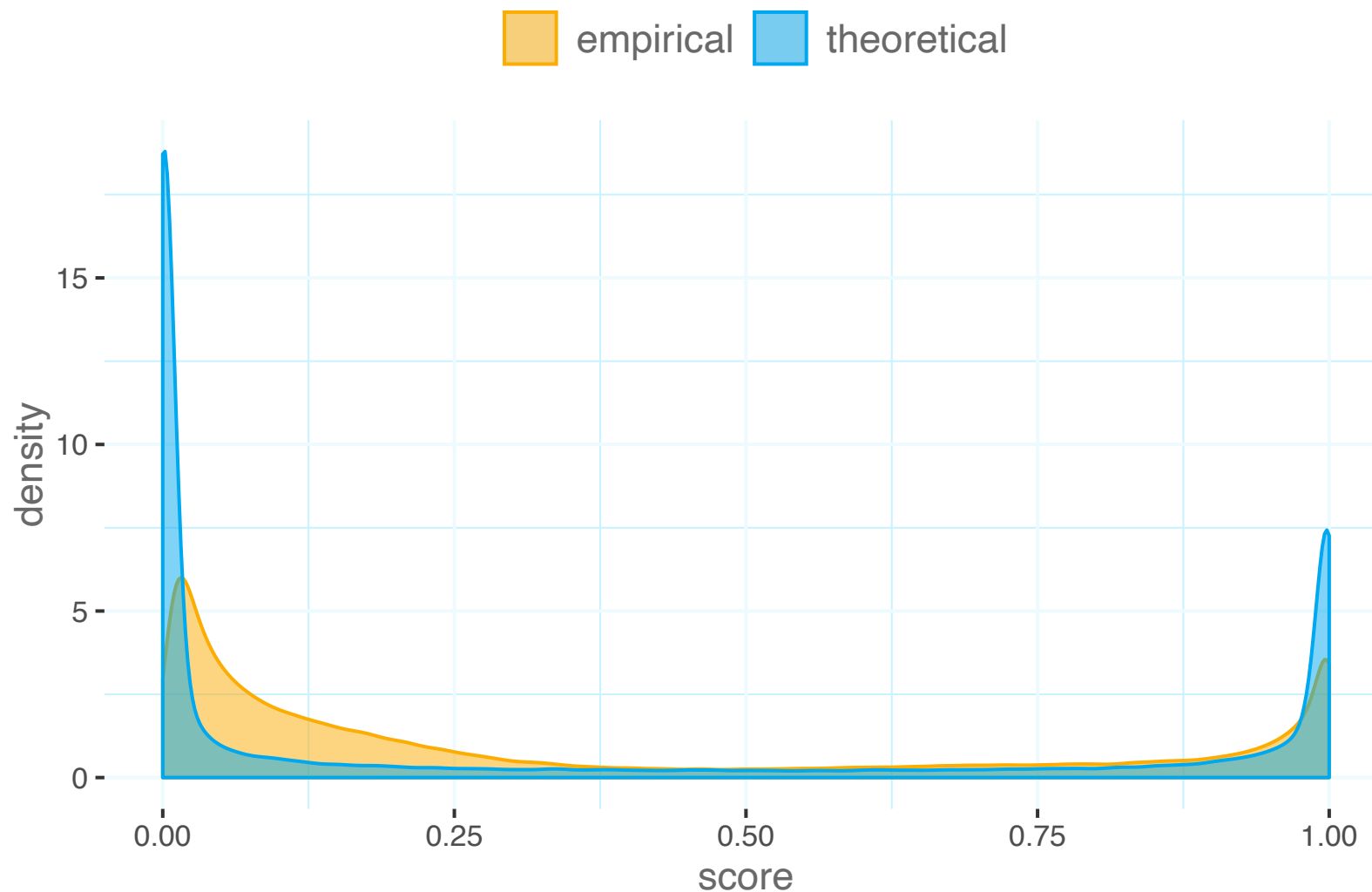
MDS Quality Measures dataset:

- Over 15,000 providers in US
- 200,000 entities
- No predictive power

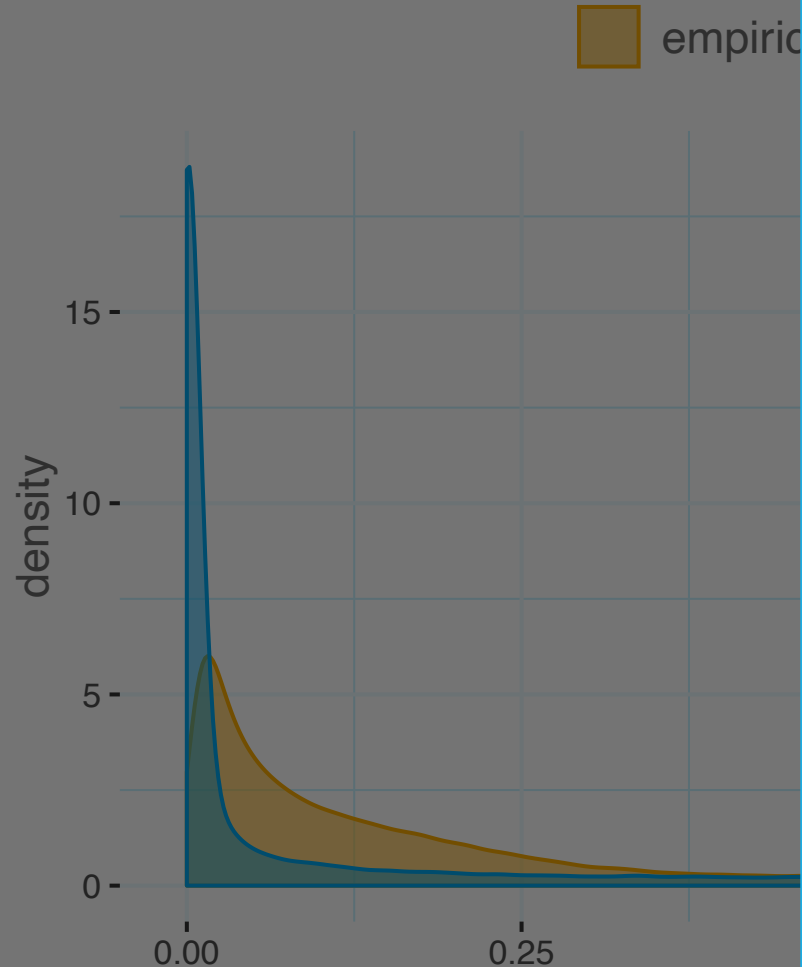
Provider Information dataset:

- Over 80 features
- 15,000 entities
- At least 70 have predictive power

MDS QUALITY MEASURE: STATISTICAL DESCRIPTIVE



MDS QUALITY MEASURE: S



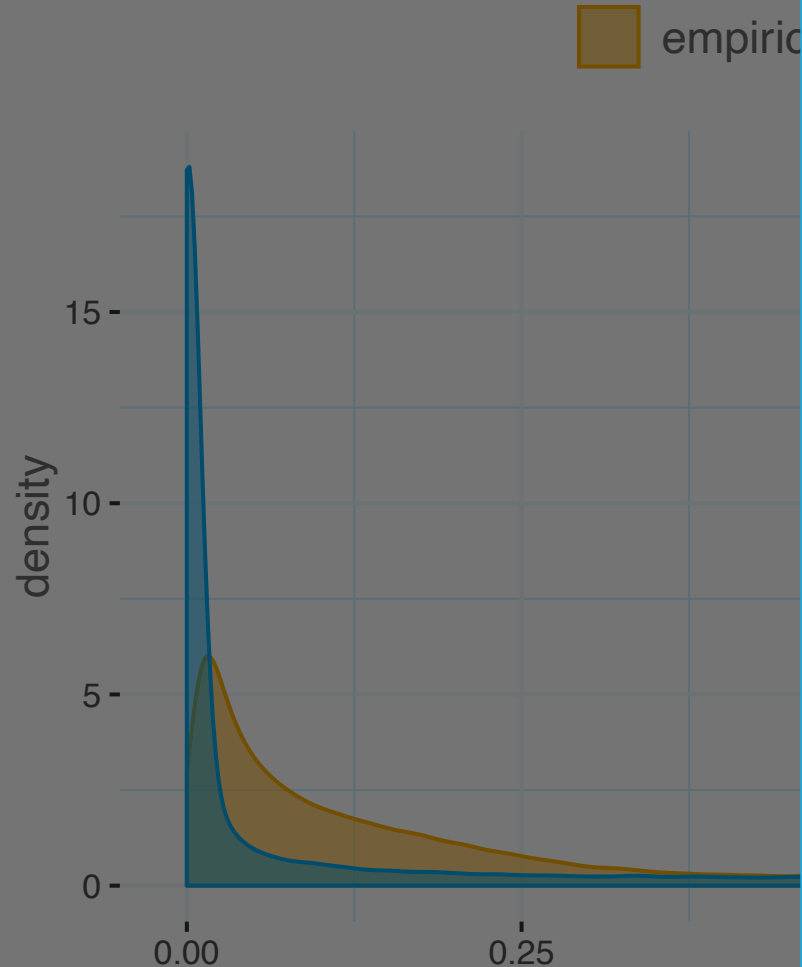
Purpose of this plot:

The empirical distribution is real shape of distribution that is based on MDS quality measure. The statistical descriptive shows the comparison between empirical and theoretical distribution. Both shapes should be similar. But this test shows that both shapes are not similar at all.

But why am I doing this plot?:

Normally, you want to test hypothesis to see if difference is statistically significant using Central Limit Theorem or any parametric tests. But this analysis shows that this method won't work.

MDS QUALITY MEASURE: S



Purpose of this plot:

The empirical distribution is real shape of distribution that is based on MDS quality measure. The statistical descriptive shows the comparison between empirical and theoretical distribution. Both shapes should be similar. But this test shows that both shapes are not similar at all.

But why am I doing this plot?:

Normally, you want to test hypothesis to see if difference is statistically significant using Central Limit Theorem or any parametric tests. But this analysis shows that this method won't work.

STATISTICAL DESCRIPTIVE

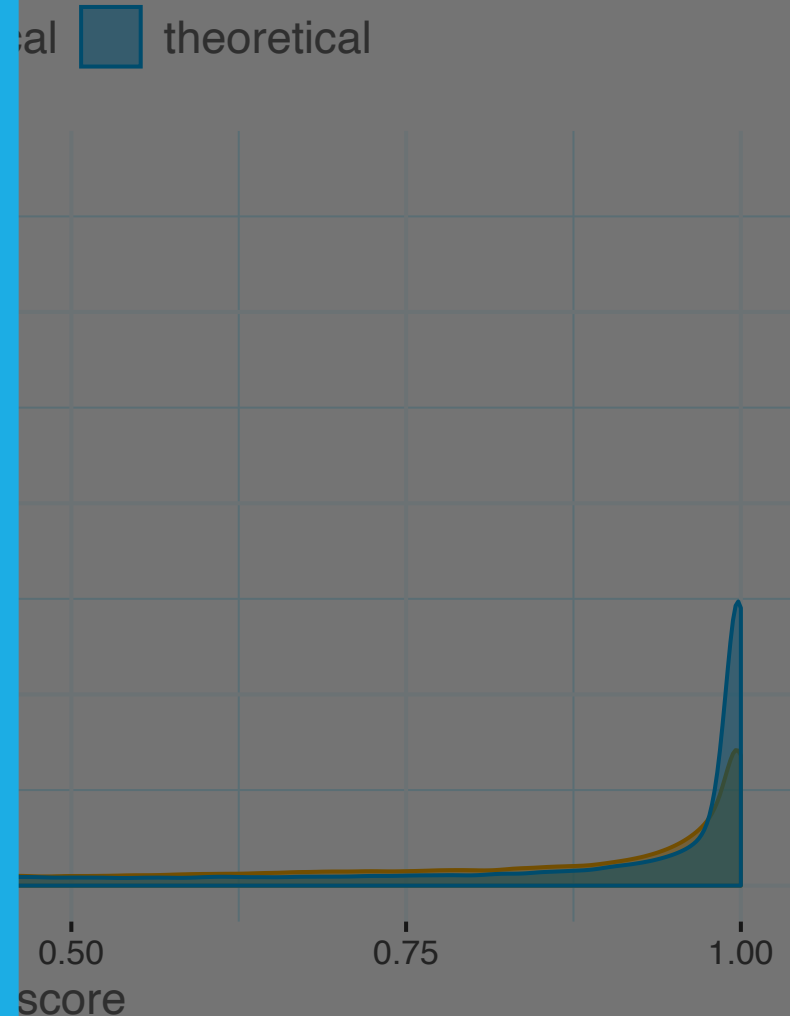
Theoretical Distribution:

The theoretical distribution is called *Beta* distribution that is in use. This distribution is a continuous version of binomial, which is suitable for this test.

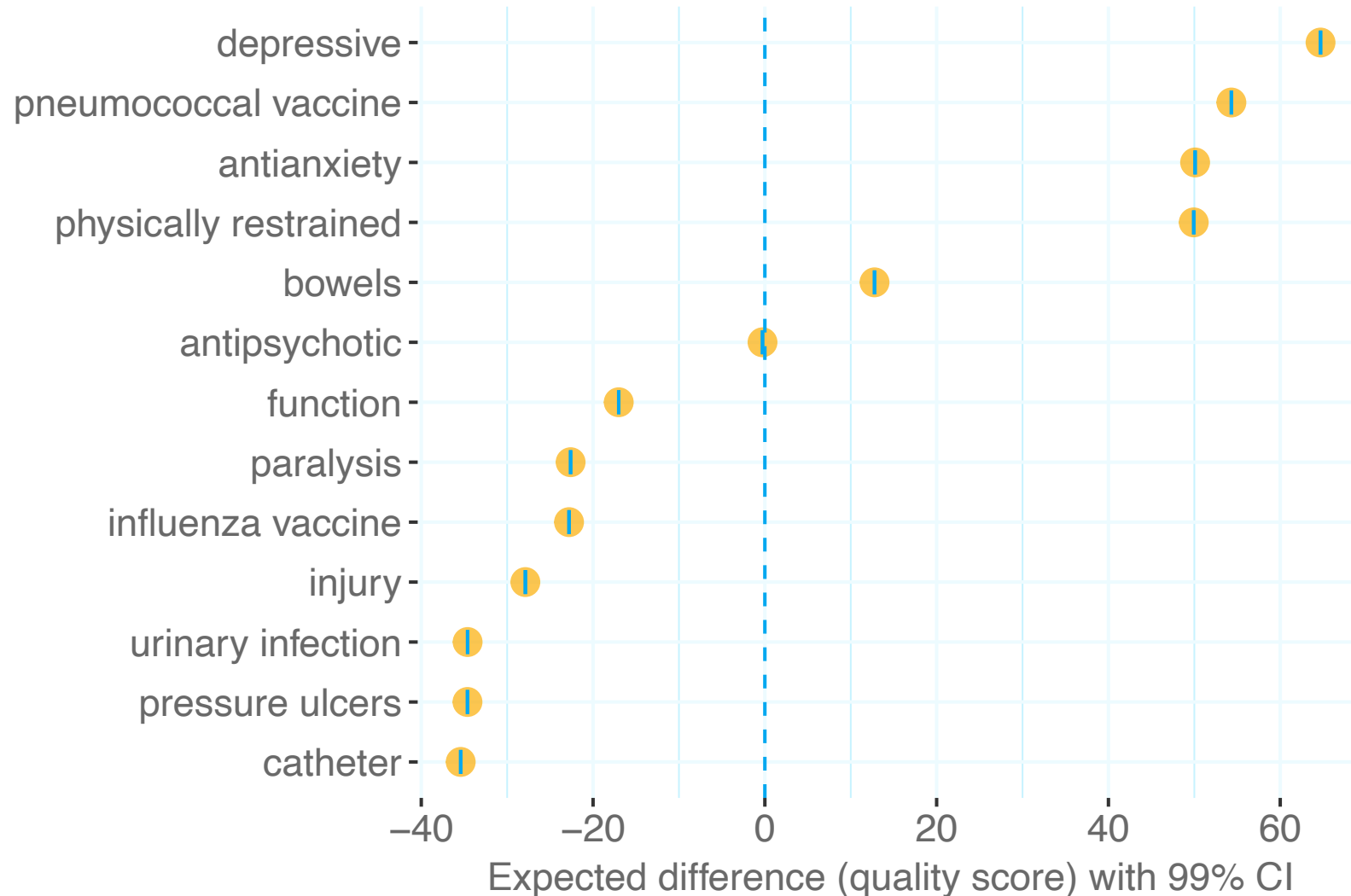
Empirical Distribution:

- Continuous
- $\varepsilon(0,100)$
- Cross-sectional
- Bimodally distributed

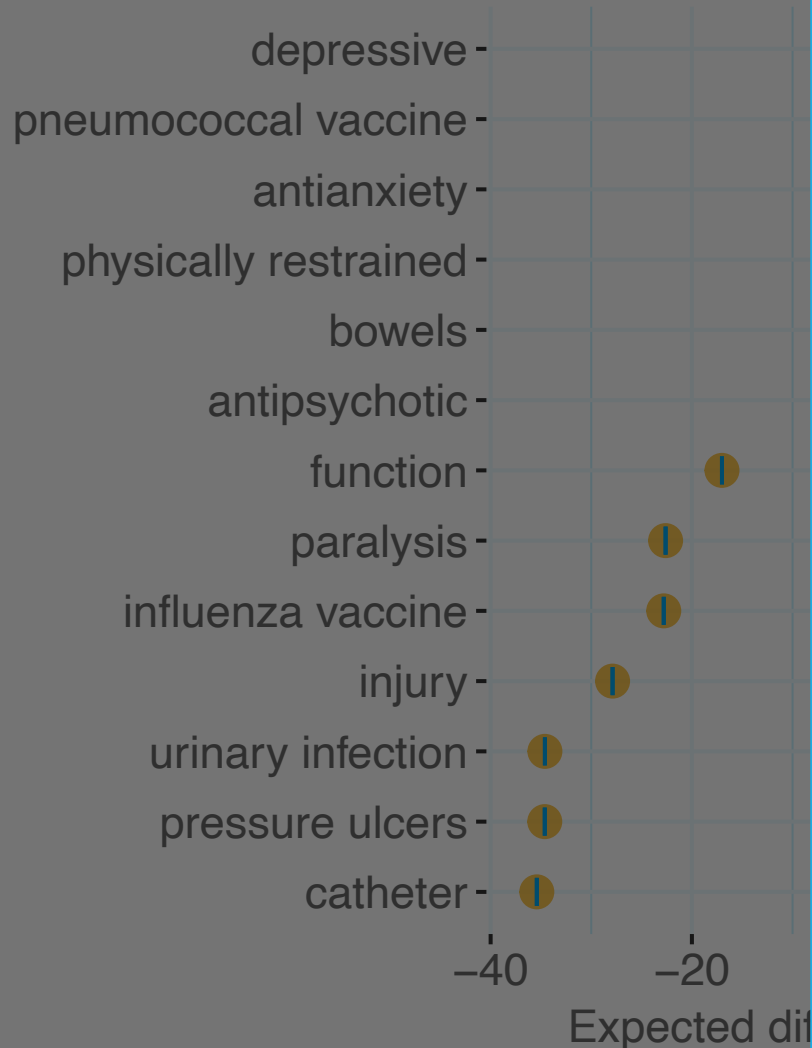
This empirical distribution has limited range from 0 to 100, and continuous. This distribution is cross-sectional.



MDS QUALITY MEASURE: STATISTICAL DESCRIPTIVE



MDS QUALITY MEASURE: S



Permutation Test:

H_0 : two groups are with the same distribution

H_a : two groups are NOT with the same distribution

The hypothesis testing is against 13 different measures codes, so the alpha level with Bonferroni correction is 0.38%. The purpose of this correction is to minimize the chance of Type I Error.

The 99% confidence interval is difficult to see due to small standard error.

The expected difference for catheter-related measure code is below -40 while the expected for difference for depressive-related measure is higher than 60. But this analysis is not established with a causal relationship due to self-report bias (CMS, 2021).

PROVIDER INFORMATION DATA: WRANGLING



THE PROVIDER INFORMATION DATASET HAS OVER 80 FEATURES BUT SOME ARE REDUNDANT (SOME ARE PERFECTLY CORRELATED). SOME HAVE LEAKAGES.



THE LEAKAGE REFERS TO WHICH THE FEATURES IN THE TRAINING PROCESS DO NOT EXIST WHEN INTEGRATING THE PRODUCTION, IN TURN, CAUSES THE PREDICTIVE SCORE TO OVERESTIMATE.

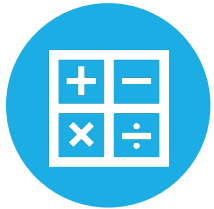


FOR EXAMPLE, THE OVERALL RATING PREDICTING THE HEALTH INSPECTION RATING IS A FORM OF LEAKAGE. 30 LEAKED FEATURES ARE FOUND AND ALL OF THEM ARE REMOVED.

FEATURE SELECTION



THIS DATASET IS SEPARATED INTO TRAINING, VALIDATION, AND TESTING SETS USING 60-20-20 RULE.



THERE ARE 36 FEATURES IN TOTAL BESIDES THE TARGET VARIABLE. THE TOTAL POSSIBLE SELECTIONS ARE 68,719,476,736. AN ATTEMPT TO OPTIMIZE THIS MODEL MANUALLY IS NOT POSSIBLE.



THIS IS WHY LASSO REGULARIZATION AND BAYES OPTIMAL FEATURE SELECTION ARE IN USE TO COMPARE BEST RESULT OF FEATURE SELECTION.

FEATURE SELECTION



THIS DATASET IS SEPARATED INTO TRAIN AND TEST SETS USING 60-20-20 RULE.



THERE ARE 36 FEATURES IN TOTAL. THE NUMBER OF POSSIBLE SELECTIONS ARE 68,719. A BRUTE FORCE MODEL MANUALLY IS NOT POSSIBLE.



THIS IS WHY LASSO REGULARIZATION AND CROSS-VALIDATION ARE IN USE TO COMPARE BEST RESULTS.

Lasso regularization:

This tool is popular in data science because this tool has the ability to identify unimportant predictors by setting predictors to zero.

Bayes optimal feature selection:

The Bayes optimal feature selection is an iterative process by balancing its needs of exploration and exploitation using three functions, as following:

- **Objective function:** the true shape cannot be seen and it can only reveal some data points that can otherwise be expensive to compute.
- **Surrogate function:** the probabilistic model is being built to exploit what is known and it is altered in the light of new information using objective function.
- **Acquisition function:** this function is to calculate a vector of hyperparameters that make this search smarter using surrogate function.

Lasso regularization:

- R-squared score is 32.19% for validation set.
- Number of features in total is 16.

Bayes optimal feature selection:

- R-squared score is 0.68% higher than lasso.
- Number of features in total is 21.

For this analysis, the lasso regularization is best because the simpler model is less likely to overfit.

TO TRAINING, VALIDATION, AND TESTING SETS

TOTAL BESIDES THE TARGET VARIABLE. THE TOTAL
19,476,736. AN ATTEMPT TO OPTIMIZE THIS
BLE.

TION AND BAYES OPTIMAL FEATURE SELECTION
SULT OF FEATURE SELECTION.

FEATURE SELECTION



THIS DATASET IS SEPARATED INTO TRAIN, VALIDATION, AND TESTING SETS USING 60-20-20 RULE.



THERE ARE 36 FEATURES IN TOTAL. THE TOTAL POSSIBLE SELECTIONS ARE 68,719. MANUALLY SELECTING THE BEST MODEL MANUALLY IS NOT POSSIBLE.



THIS IS WHY LASSO REGULARIZATION AND CROSS-VALIDATION ARE IN USE TO COMPARE BEST RESULTS.

Light Gradient Boosting Model:

- This model is optimized by Bayesian search theory.
- R-squared score is 43.47% for validation set.
- R-squared score is 41.92% for testing set.
- Root mean square error score is 0.97.

If this model predicts the health inspection rating equal to 3.5, the actual score may fall somewhere between 2.53 and 4.47. This shows how large the error is.

ECONOMETRIC METHOD



TO THE EXTENT THAT THE CAUSAL RELATIONSHIP FAILS TO ESTABLISH, THE PREVIOUS DISCUSSION DEMONSTRATES THE LIMITS OF DATA SCIENCE. FOR THAT REASON, THE ECONOMETRIC APPROACH IS ADOPTED.

- MINIMAL REQUIREMENT FOR THIS APPROACH IS THAT THE ADJUSTED R-SQUARED SCORE MUST BE POSITIVE.
- THE CAUSAL RELATIONSHIP CAN BE ESTABLISHED WHEN MODEL IS SPECIFIED.
- GAUSS MARKOV ASSUMPTIONS ARE FUNDAMENTAL OF THIS APPROACH, THOUGH THE CAUSAL MODEL MAY DEPART FROM SOME ASSUMPTIONS.

ECONOMETRIC METHOD



TO THE EXTENT THAT THE PREVIOUS DISCUSSION DEMONSTRATES, THE ECONOMETRIC METHOD IS A

- MINIMAL REQUIREMENT FOR THE OLS METHOD IS THAT THE SQUARED SCORE MUST BE
- THE CAUSAL RELATIONSHIP BETWEEN THE EXOGENOUS AND ENDOGENOUS VARIABLES
- GAUSS MARKOV ASSUMPTIONS ARE NOT NEEDED, HOWEVER, THE CAUSAL MODEL

Gauss Markov assumptions:

- A1:** linearity in parameters
- A2:** no perfect collinearity
- A3:** zero conditional mean error
- A4:** homoskedasticity and no serial correlation
- A5:** normality of the error

This causal model does not follow A1 and A4 assumptions because, firstly, the endogenous variable is Limited Dependent Variable (LDV). Secondly, the Breusch-Pagan test confirms that the error variances are all equal is rejected at the significance level. Finally, this model follows contemporaneous exogeneity; a weaker version of strict exogeneity.

Why not logarithmic transformation?

- The popular solution is logarithmic transformation without Monte Carlo simulation in order to preserve A1 assumption.
- This transformation is, unfortunately, incorrect. The model misspecification is called Duan's Smear where error term is equal to 1, not 0.
- Example: $e^{u_i} = e^0 = 1$.

Why Probit with Quasi-Maximum Likelihood Estimation (QMLE)?

- The Probit is more efficient to heteroskedasticity, which refers to the variance of residual term is not constant.
- QMLE is a weaker version of Conditional Maximum Likelihood Estimation but the condition for QMLE is consistent and asymptotically normal.
- QMLE condition may be little less efficient if the information loss is being minimized.

IF CAUSAL RELATIONSHIP FAILS TO ESTABLISH, THE
CONSTRATES THE LIMITS OF DATA SCIENCE. FOR THAT
APPROACH IS ADOPTED.

FOR THIS APPROACH IS THAT THE ADJUSTED R-
SQUARED IS POSITIVE.

CAUSAL RELATIONSHIP CAN BE ESTABLISHED WHEN MODEL IS SPECIFIED.
ASSUMPTIONS ARE FUNDAMENTAL OF THIS APPROACH,
MODEL MAY DEPART FROM SOME ASSUMPTIONS.

ECONOMETRIC METHOD



TO THE EXTENT THAT THE PREVIOUS DISCUSSION DEMONSTRATES, FOR ONE REASON, THE ECONOMETRIC METHOD IS USED FOR THE FOLLOWING REASONS:

- MINIMAL REQUIREMENT FOR THE ADJUSTED R-SQUARED SCORE MUST BE MET
- THE CAUSAL RELATIONSHIP BETWEEN VARIABLES IS NOT KNOWN
- GAUSS MARKOV ASSUMPTIONS ARE NOT MET, BUT THE CAUSAL MODEL IS USED

Causal Model:

$$ins_i = \phi(\beta_0 + \beta_1 bed_i + \beta_2 hr_i + cond_i \beta) + u_i$$

- ins_i is health inspection rating per inspection
- $\phi(.)$ is the Probit function.
- hr_i is amount of hours spending with residents per day
- bed_i is the number of certified beds
- $cond_i$ is a vector of conditions including the kind of quality care, competency, location area, environment, and residents' vulnerability.

Exogenous variables for $cond_i$:

- National Area Regional Code
- Number of Substantiated Complaints
- Number of Facility Incident Reports
- Number of Fines
- Amount of Fines in Total
- Family on the Council (dummy)
- Licensed Practical Nurses (dummy)
- Registered Nurses (dummy)
- Nurse Aides (dummy)
- Abuse Icon (dummy)
- Installed automatic sprinkler system (dummy)

ECONOMETRIC METHOD



TO THE EXTENT THAT THE PREVIOUS DISCUSSION DEMONSTRATES, FOR ONE REASON, THE ECONOMETRIC METHOD IS USED.

- MINIMAL REQUIREMENT FOR THE ADJUSTED R-SQUARED SCORE MUST BE MET
- THE CAUSAL RELATIONSHIP BETWEEN THE VARIABLES MUST BE ESTABLISHED
- GAUSS MARKOV ASSUMPTIONS MUST BE SATISFIED, EVEN THOUGH THE CAUSAL MODEL IS NOT KNOWN

Causal Model:

$$ins_i = \phi(\beta_0 + \beta_1 bed_i + \beta_2 hr_i + cond_i \beta) + u_i$$

- ins_i is health inspection rating per inspection
- $\phi(.)$ is the Probit function.
- hr_i is amount of hours spending with residents per day
- bed_i is the number of certified beds
- $cond_i$ is a vector of conditions including the kind of quality care, competency, location area, environment, and residents' vulnerability.

Exogenous variables for $cond_i$:

- National Area Regional Code
- Number of Substantiated Complaints
- Number of Facility Incident Reports
- Number of Fines
- Amount of Fines in Total
- Family on the Council (dummy)
- Licensed Practical Nurses (dummy)
- Registered Nurses (dummy)
- Nurse Aides (dummy)
- Abuse Icon (dummy)
- Installed automatic sprinkler system (dummy)

ECONOMETRIC METHOD



TO THE EXTENT THAT THE PREVIOUS DISCUSSION DEMONSTRATES, FOR THIS REASON, THE ECONOMETRIC METHOD IS USED.

- MINIMAL REQUIREMENT FOR THE ADJUSTED R-SQUARED SCORE MUST BE MET
- THE CAUSAL RELATIONSHIP MUST BE ESTABLISHED
- GAUSS MARKOV ASSUMPTIONS MUST BE MET, EVEN THOUGH THE CAUSAL MODEL IS NOT KNOWN

Wald Test:

The result shows that the Abuse Icon and Installed automatic sprinkler system are not significantly different from zero.

Exogenous variables for cond_i:

- National Area Regional Code
- Number of Substantiated Complaints
- Number of Facility Incident Reports
- Number of Fines
- Amount of Fines in Total
- Family on the Council (dummy)
- Licensed Practical Nurses (dummy)
- Registered Nurses (dummy)
- Nurse Aides (dummy)
- Abuse Icon (dummy)
- Installed automatic sprinkler system (dummy)

Wald Test of Exogeneity:

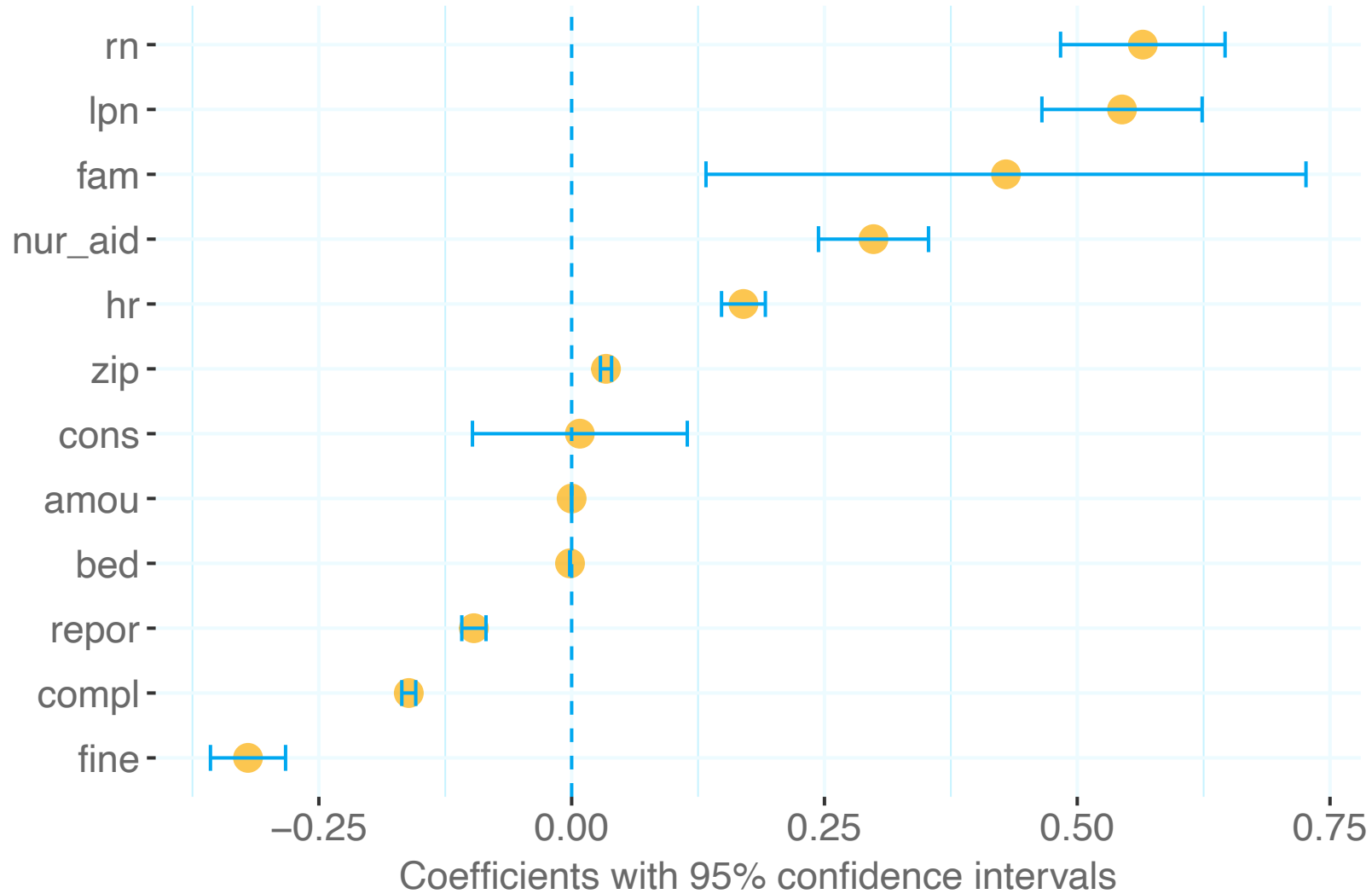
The Abuse Icon is a dummy used for Wald test of exogeneity with IV Probit. As a result, the null hypothesis that the endogeneity does not is failed to reject at 5% alpha level.

CAUSAL RELATIONSHIP FAILS TO ESTABLISH, THE
ONSTRATES THE LIMITS OF DATA SCIENCE. FOR THAT
APPROACH IS ADOPTED.

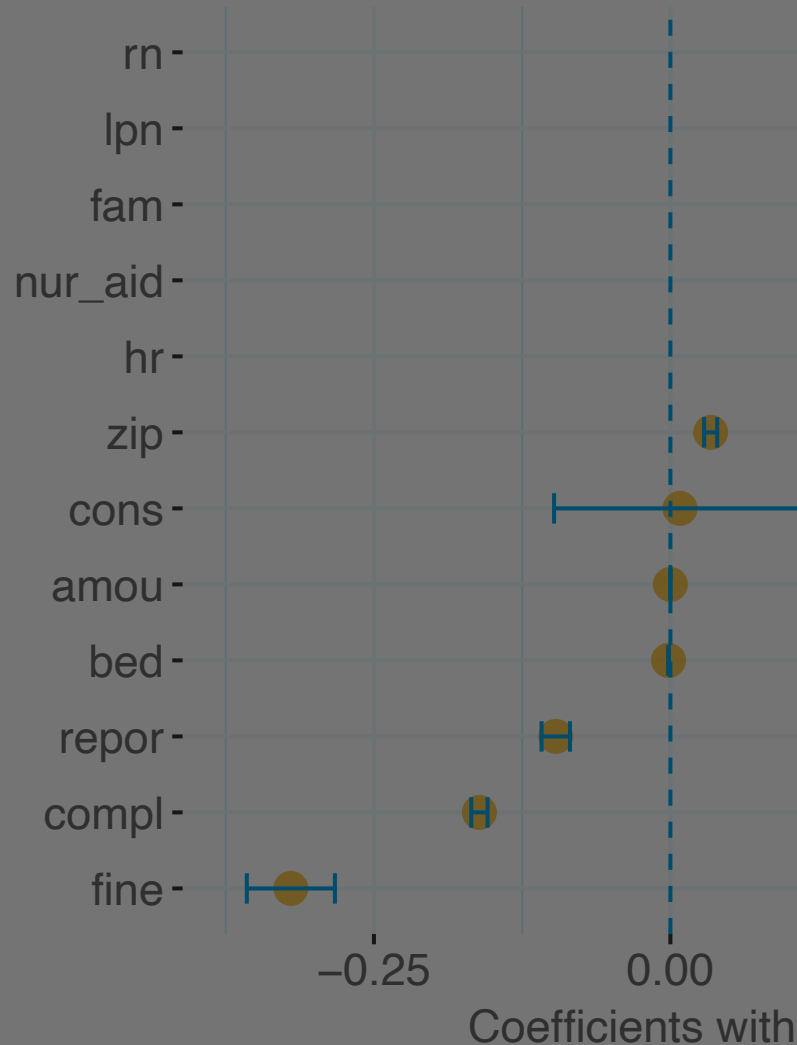
FOR THIS APPROACH IS THAT THE ADJUSTED R-
POSITIVE.

Y CAN BE ESTABLISHED WHEN MODEL IS SPECIFIED.
PTIONS ARE FUNDAMENTAL OF THIS APPROACH,
DEL MAY DEPART FROM SOME ASSUMPTIONS.

ACTIONABLE INSIGHTS



ACTIONABLE INSIGHTS

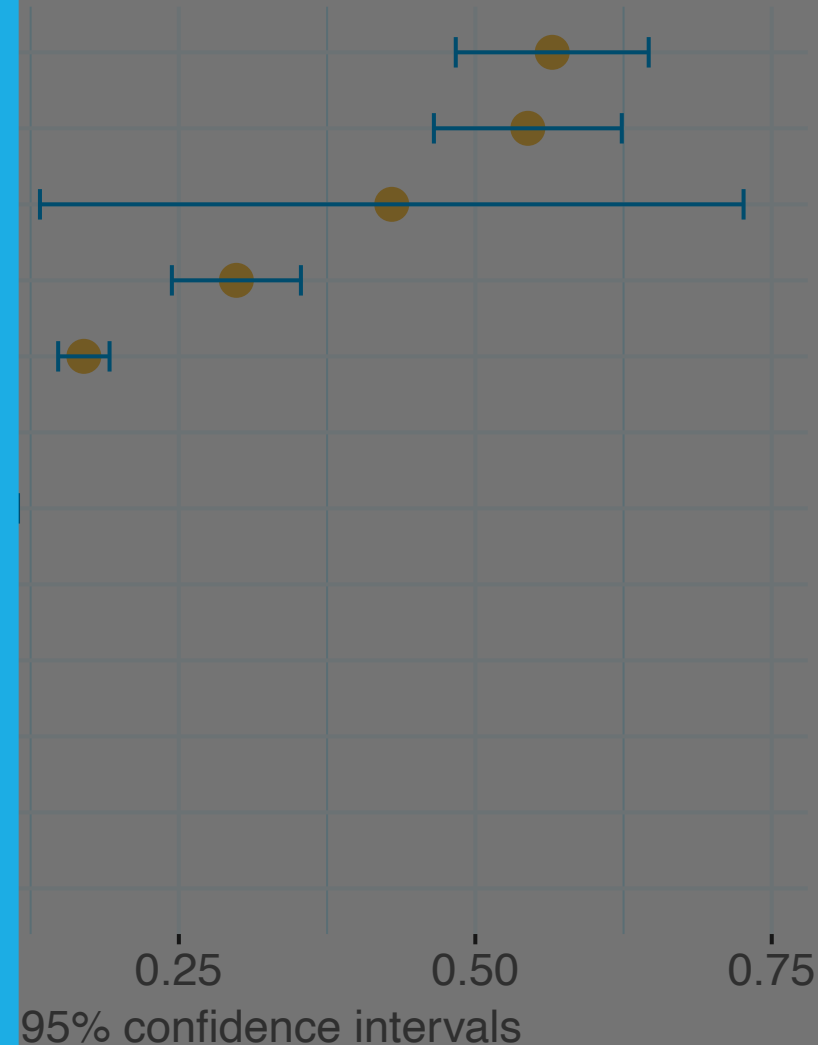


Key Results:

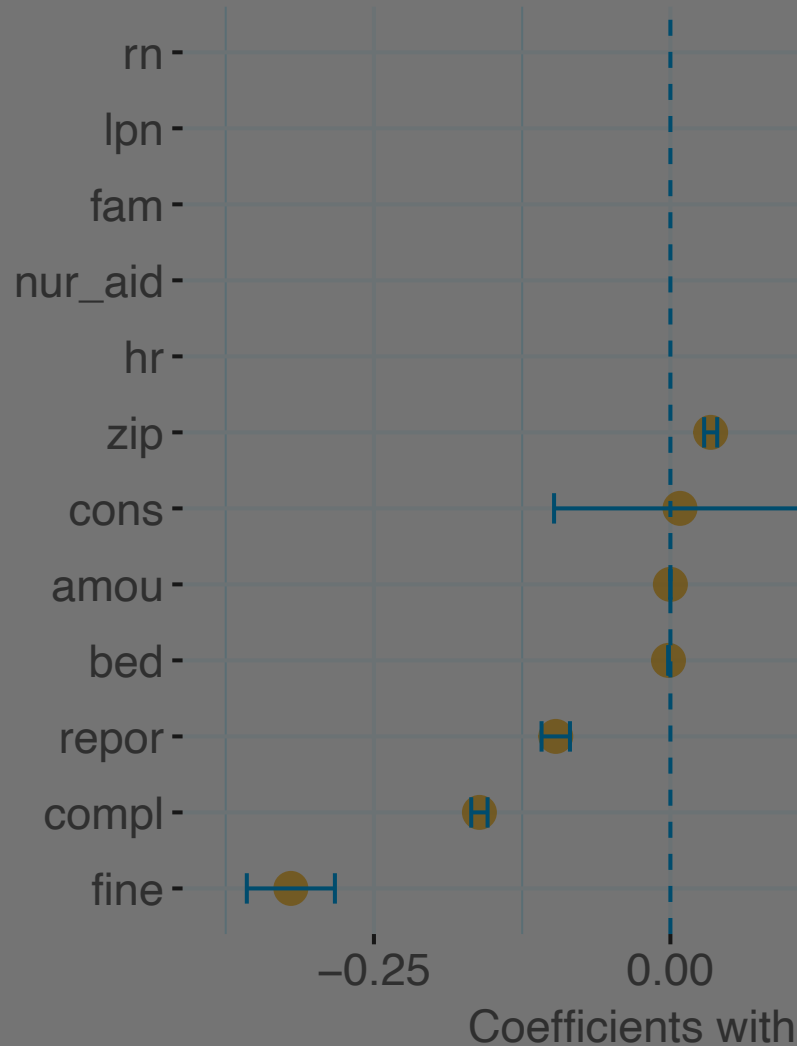
- Except for intercept, each coefficient of exogenous variables is statistically different from zero at 5% using Huber-White (HCO covariance type) robust standard errors.
- The pseudo-R-squared score is 32.54%, which is considerably high for social sciences.
- The p-value for LLR is below 5%, rejecting the null hypothesis that the fit of the intercept-only model and causal model is equal.
- The p-value for intercept is 0.881.

Key insights:

- When all variables are being controlled and the causal model is specified, the registered and license practical nurses have positive effects on health inspection rating as expected.
- With the nurse aides being present in facilities, their positive effect on health inspection rating is relatively less.
- When the interaction term is applied, things become more interesting.



ACTIONABLE INSIGHTS



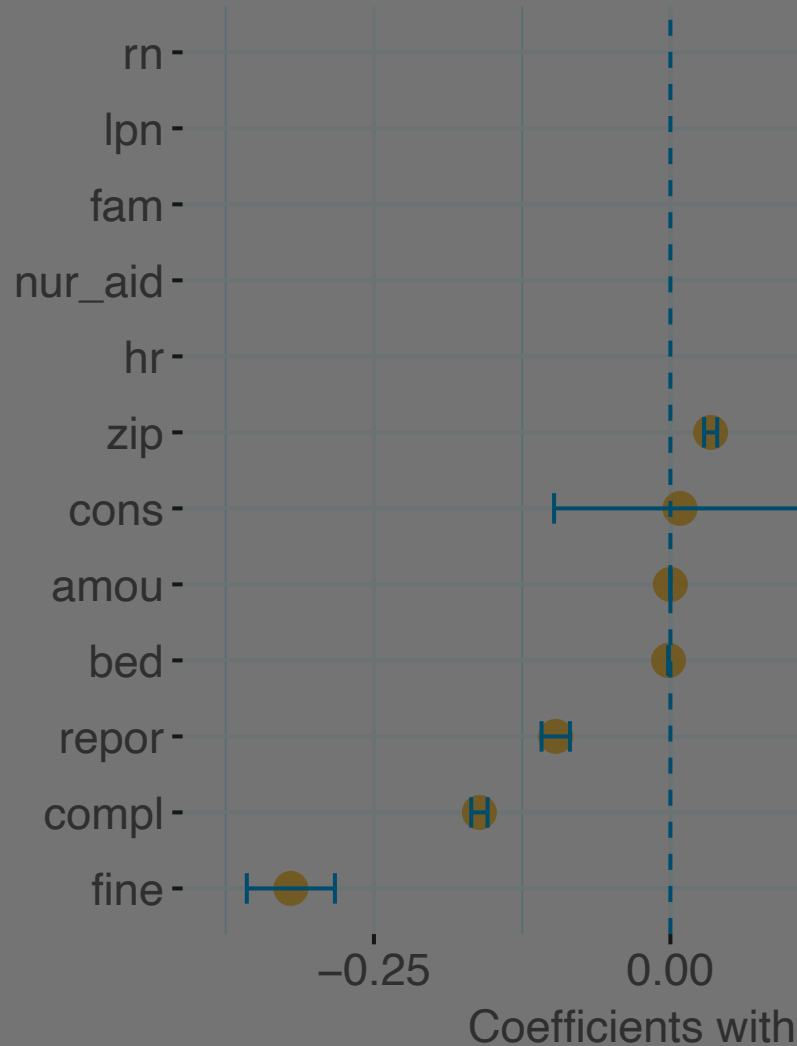
Before discussion goes further:

The concept of interaction term is worth explaining. This means that dummy variable is multiplied by independent variable.

Example:

Independent Variable	Dummy Variable	Interaction Term
32	1	32
13	0	0
20	1	20
53	1	53
8	0	0

ACTIONABLE INSIGHTS



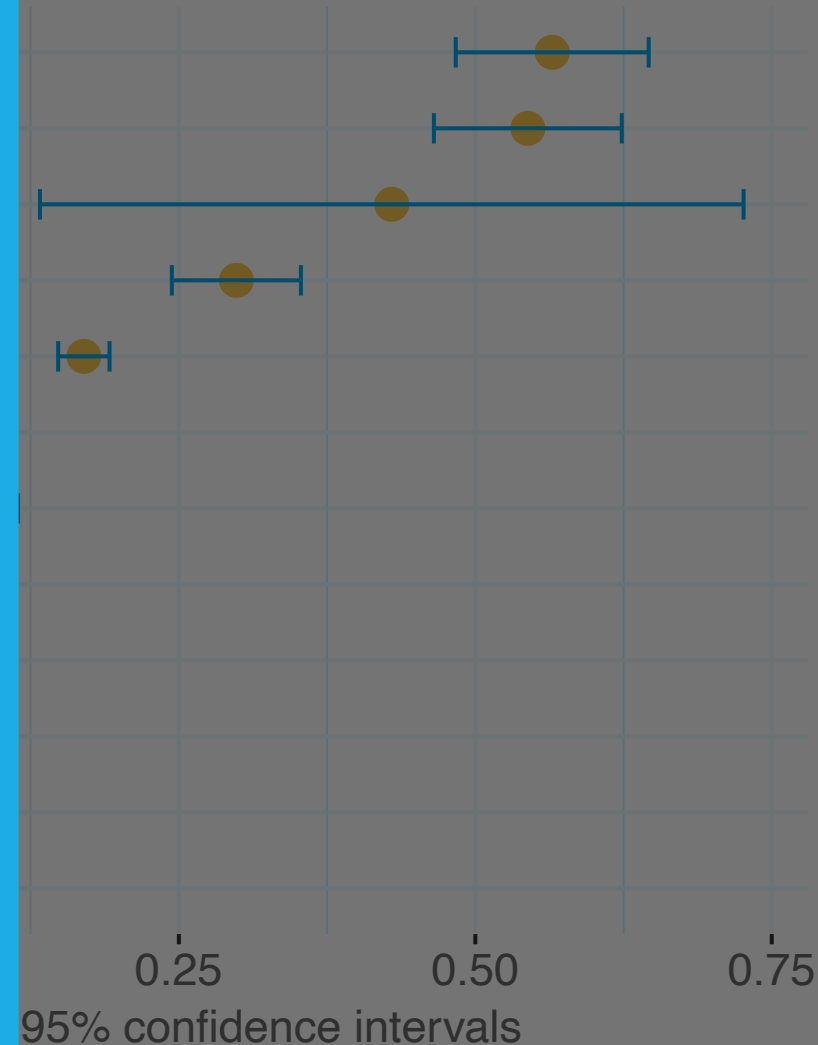
But why?:

The interaction term is to help understand more about nature of regressing y on x because the interaction term can dramatically change the coefficients. Without interaction term, the model assumes that there is no relationship between two independent variables.

Interaction Terms:

- When the h_i interacts with nur_aid_i , the increase in number of hours that the nurse aides spend with residents each day has little to no impact on health inspection rating.
- The registered nurses that interacts with h_i have a positive impact on this rating on the significance level.
- More surprisingly, when the h_i interacts with lpn_i , the coefficient is -0.0881 with margin of error equal 0.07, and that is, the incremental increase in the number of hours that license practical nurses spend has a negative impact on this rating on the significance level.

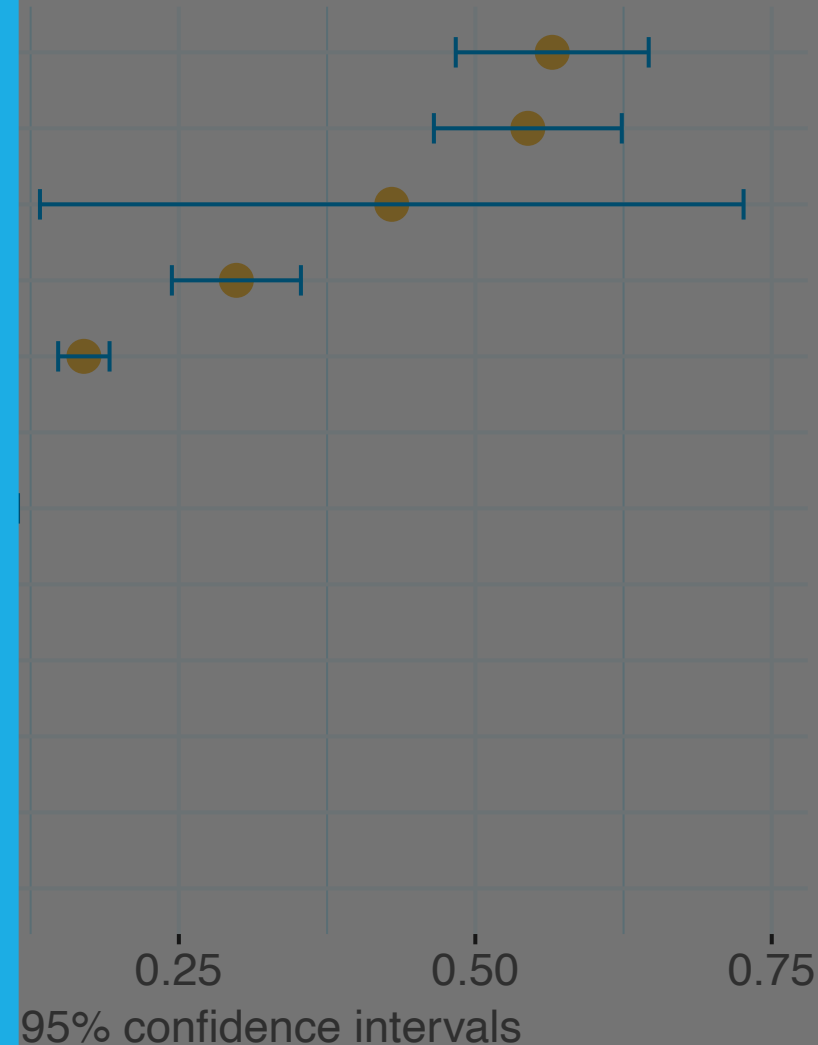
For example, the quality score for depressive-related code is higher while the score for catheter-related code is lower. There is potential linkage between the roles of nurses. The registered nurses administer medication and treatments while the license practical nurses comfort the residents and provide the basic care including the insertion of catheters.



Interaction Terms:

- When the h_i interacts with nur_aid_i , the increase in number of hours that the nurse aides spend with residents each day has little to no impact on health inspection rating.
- The registered nurses that interacts with h_i have a positive impact on this rating on the significance level.
- More surprisingly, when the h_i interacts with lpn_i , the coefficient is -0.0881 with margin of error equal 0.07, and that is, the incremental increase in the number of hours that license practical nurses spend has a negative impact on this rating on the significance level.

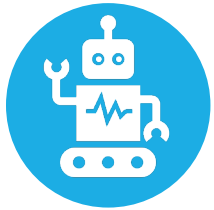
For example, the quality score for depressive-related code is higher while the score for catheter-related code is lower. There is potential linkage between the roles of nurses. The registered nurses administer medication and treatments while the license practical nurses comfort the residents and provide the basic care including the insertion of catheters.



FUTURE RESEARCH



THE CMS PILOT PROGRAM IS ESTABLISHED BY PROVIDING FINANCIAL AID FOR LPN-TO-RN CAREER PATHWAY OR PROMOTE AWARENESS OF EXISTING PROGRAMS. THE RCT SHOULD IDENTIFY REAL IMPACT OF THIS PROGRAM.



THE DIRECT SOLUTION IS BY TRAINING MACHINE OR DEEP LEARNING TO IDENTIFY INFLATION IN SELF-REPORT DATA THAT HELPS TO MAKE RATING SYSTEM MORE CREDIBLE ONLY IF DATA AUDIT IS UNDERTAKEN AND THIS DATA EXISTS.

REFERENCES

- "Technical Details." *Nursing homes including rehab services*, the Centers for Medicare and Medicaid Services, Sep. 2021. <https://data.cms.gov/provider-data/topics/nursing-homes/technical-details#health-inspections>
- Ahmed, S., Narasimhan, H., and Agarwal, S. "Bayes Optimal Feature for Supervised Learning with General Performance Measures." arXiv, 2015. <http://auai.org/uai2015/proceedings/papers/72.pdf>
- Boehmke, B. "Regularized Regression." *UC Business Analytics R Programming Guide*, University of Cincinnati, 2021. http://uc-r.github.io/regularized_regression#lasso
- Datta, A., Fredrikson, M., Ko, G., Mardziel, P., and Sen, S. "Proxy Discrimination in Data-Driven Systems." *Theory and Experiments with Learnt Programs*, arXiv, 25 Jul. 2017. <https://arxiv.org/abs/1707.08120>
- Freud, R.J., Wilson, W.J., and Mohr, D.L. "Inferences for Two or More Means." *Statistical Methods, Third Edition*, Academic Press, 2010. <https://doi.org/10.1016/B978-0-12-374970-3.00006-8>
- Goldstein, Nathan. "Lecture 1. Foundations of Microeconometrics." *Microeconometrics*, Zanvyl Krieger School of Arts and Sciences Johns Hopkins University, 2021.
- Kaynig-Fattkau, V., Blitzstein, J., and Pfister, H. "CS109 - Data Science." *Decision Trees*, Harvard University, 2021. <https://matterhorn.dce.harvard.edu/engage/player/watch.html?id=c22cbde8-94dd-42ad-86ef-091448ad02e4>
- Khandelwal, P. "Which algorithm takes the crown: Light GBM vs XGBOOST?" *Analytics Vidhya*, 12 Jun 2017. <https://analyticsvidhya.com/blog/2017/06/which-algorithm-takes-the-crown-light-gbm-vs-xgboost/>
- Park, Y. and Ho J.C. "PaloBoost." *An Overfitting-robust TreeBoost with Out-of-Bag Sample Regularization Techniques*, Emory University, 22 Jul 2018. <http://arxiv-export-lb.library.cornell.edu/pdf/1807.08383>
- Silver-Greenberg, Jessica, and Robert Gebeloff. "Maggots, Rape and Yet Five Stars: How U.S. Ratings of Nursing Homes Mislead the Public." *How U.S. Ratings of Nursing Homes Mislead the Public*, New York Times, 13 Mar. 2021. <https://www.nytimes.com/2021/03/13/business/nursing-homes-ratings-medicare-covid.html>
- Sinharay, S. "Continuous Probability Distributions." *The International Encyclopedia of Education*, Elsevier Science, 2010. <https://doi.org/10.1016/B978-0-08-044894-7.01720-6>