American Traveler PIC Math Project: Forecasting Demand for Traveling Nurses

Objective: Design a statistical model to predict how the demand for nurses will change over time in the next five years.

Navika Timsina and Daniel Walford

College of Science
Florida Atlantic University

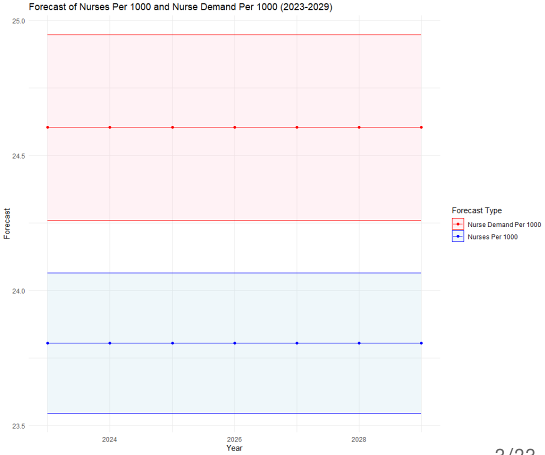
April 17, 2024

Statement

- Problem
  - Growing Nurse Shortage in USA
  - Projecting The Future Growth
- Motivation
  - Lessened Quality of Patient Care.
  - Overworking Nurses

Inconclusive Results

- ARIMA Prediction
- Real World Data



Approach

Approach

- Data Collection
- Supply and Demand Equations
- Forecast Equations
- Compare Supply and Demand to find surplus/deficit.

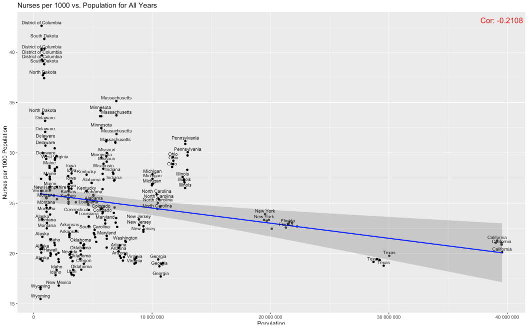
Data We Analyzed (2012-2022)

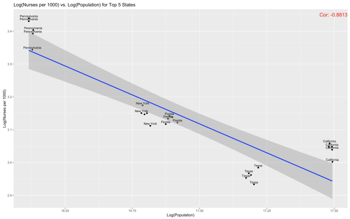
State	Year	NursesPer1000	Population	Over65_Percent	Median_Income	Median_OOPC
Alabama	2018	27.02199	4887871	16.90822	49940	1350
Alabama	2019	27.00490	4903185	17.33235	56200	1950
Alabama	2020	25.82998	5031362	17.40229	54690	1400
Alabama	2021	25.74930	5049846	17.67747	56930	1550
Alabama	2022	25.19364	5074296	18.02680	59910	1300
Alaska	2018	19.26942	737438	11.80981	68730	1200
Alaska	2019	20.43620	731545	12.51980	78390	1500
Alaska	2020	21.65302	732923	12.82809	74750	940
Alaska	2021	20.59435	734182	13.37516	81130	1100
Alaska	2022	23.39204	733583	13.85133	89740	1450
Arizona	2018	20.21433	7171646	17.51581	62280	1900

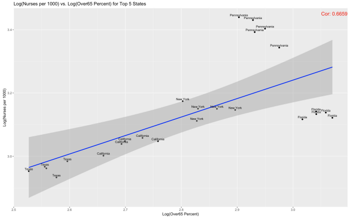
Correlations of Predictive Variabless

Florida Atlantic University, College of Science

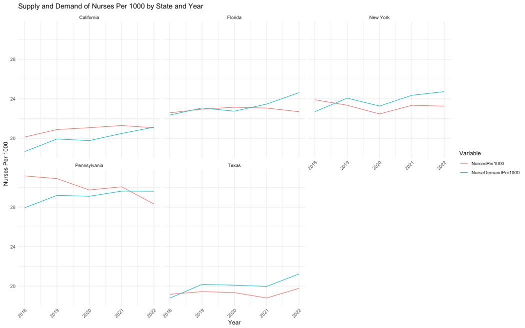
Population and Economic Indicators

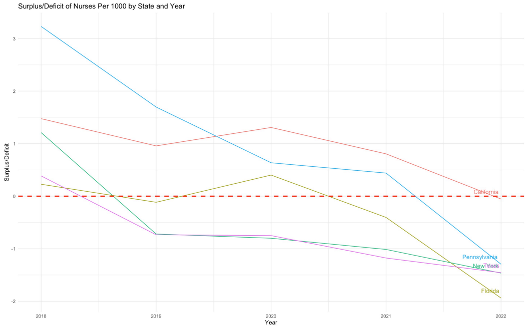






Supply and Demand Trends (5 Largest States)





Mathematics Concepts Utilized

Florida Atlantic University, College of Science

Demand Equation

- $\ln(\text{nurses per 1000 population}_{it}) = \beta_0 + \beta_1 \cdot \ln(\text{median income per state}_{it-1}) + \beta_2 \cdot \ln(\text{median income per state}_{it-4}) + \beta_3 \cdot \ln(\text{median income per state}_{it-5}) + \beta_4 \cdot \ln(\text{OOPPC}_{it-2}) + \beta_5 \cdot \ln(\%\text{Pop65}_{it-3}) + \mu_i + \xi_{it}$
    - β = coefficients [already found!]
    - i = population in states
    - t = time in years
    - μ_c = vector of state fixed effects
    - ξ_c = disturbance term

Citation: (Scheffler & Arnold, 2019), (Juraschek, Zhang, Ranganathan, & Lin, 2019)

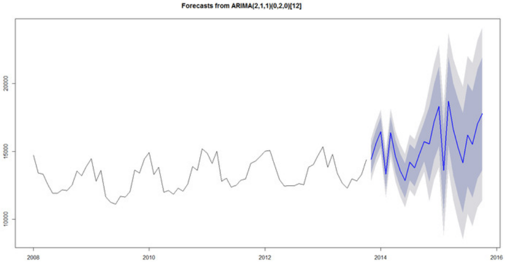
Supply Equation

- Nurses per 1000 population = $(\beta_0 + \beta_1 * t + \xi_t)$ [general equation]
  - ξ_t = random disturbance term
  - β = unknown parameters
- Future Prediction is National Data Only.
- NCLEX pass rates, Current Employment, Retirement data

Citation: (Scheffler & Arnold, 2019), (Juraschek, Zhang, Ranganathan, & Lin, 2019)

Understanding ARIMA Models

- Time Series Data
- An ARIMA model is characterized by 3 terms: p, d, q:
  - **p** – (AR).
  - **d** – (I).
  - **q** – (MA).



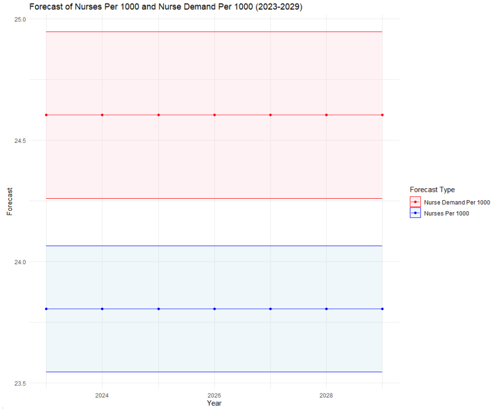
Example ARIMA Model

```r
library(forecast)
demanddata
# SUPPLY FORECAST
nurses_ts <- ts(demanddata$NursesPer1000, start = 2018, frequency = 1) # Convert to time series
fit <- auto.arima(nurses_ts) # Auto ARIMA to find suitable parameters
forecast_nurses <- forecast(fit, h = 7)


# DEMAND FORECAST
nursesd_ts <- ts(demanddata$NurseDemandPer1000, start = 2018, frequency = 1) # Convert to time series
fit <- auto.arima(nursesd_ts) # Auto ARIMA to find suitable parameter
forecast_nursesd <- forecast(fit, h = 7) # Forecasting for the next 7 years to reach 2029 (5 years from present)

# Print the forecasts
print(forecast_nurses)
print(forecast_nursesd)
```

```
> print(forecast_nurses)
     Point Forecast    Lo 80    Hi 80    Lo 95   Hi 95
2023       23.80579 23.63552 23.97606 23.54539 24.0662
2024       23.80579 23.63552 23.97606 23.54539 24.0662
2025       23.80579 23.63552 23.97606 23.54539 24.0662
2026       23.80579 23.63552 23.97606 23.54539 24.0662
2027       23.80579 23.63552 23.97606 23.54539 24.0662
2028       23.80579 23.63552 23.97606 23.54539 24.0662
2029       23.80579 23.63552 23.97606 23.54539 24.0662
```

```
> print(forecast_nursesd)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2023        24.6047 24.38002 24.82938 24.26108 24.94832
2024        24.6047 24.38002 24.82938 24.26108 24.94832
2025        24.6047 24.38002 24.82938 24.26108 24.94832
2026        24.6047 24.38002 24.82938 24.26108 24.94832
2027        24.6047 24.38002 24.82938 24.26108 24.94832
2028        24.6047 24.38002 24.82938 24.26108 24.94832
2029        24.6047 24.38002 24.82938 24.26108 24.94832
```

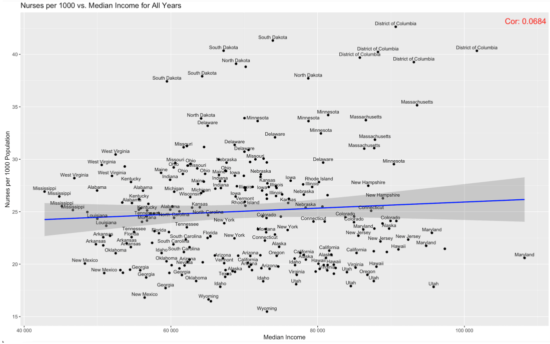
```r
# Plotting with ggplot2
ggplot(combined_forecast, aes(x = Year, y = Forecast, color = Type)) +
  geom_line() +
  geom_ribbon(aes(ymin = Lower, ymax = Upper, fill = Type), alpha = 0.2) +
  geom_point() +
  scale_color_manual(values = c("Nurses Per 1000" = "blue", "Nurse Demand Per 1000" = "red")) +
  scale_fill_manual(values = c("Nurses Per 1000" = "lightblue", "Nurse Demand Per 1000" = "pink")) +
  labs(title = "Forecast of Nurses Per 1000 and Nurse Demand Per 1000 (2023-2029)",
       y = "Forecast", x = "Year") +
  theme_minimal() +
  guides(color = guide_legend(title = "Forecast Type"), fill = guide_legend(title = "Forecast Type"))
```



Limitations

Real World Data

- Messy
- Not Strongly Correlated
- Sourcing Data
- Reliable Data

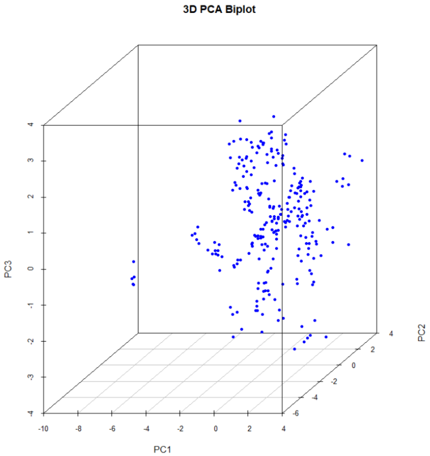


Other Limitations

- Legislations and Laws
- Personal Surveys
- COVID

Suggested Future Work

Principal Component Analysis (PCA)

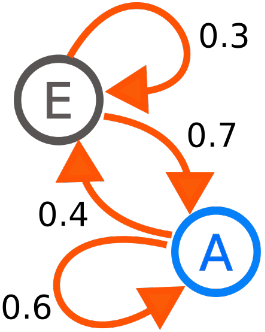


Markov Chain

1. **What is Markov Chain**
2. **Basic Notation and the Process**
   - $(X_t)_t$ of total number of nurses at time (t).
3. **Parameters**
   - $p \in [0,1]$ = rate of supply of nurses
   - $q \in [0,1]$ = rate of demand of nurses
   - $Q \in \mathbb{N}$ = total number of nurses (q-1 = nurses leave / q+1 = new nurse get certified)
4. **Utilized package:** `markovchain`

Markov Chain [Cont.]

- Likelihood function for a model: a quantitative measure of model fit.
- Scenario: imaginary small district with ”Q” available nurses and ”X” available positions in that district.



Model

- Model Parameters: $(p, q, Q) \in (0,1) \times (0,1) \times \mathbb{N}$
- State $X_t \in \{0, \ldots, Q\}$: Total number of jobs available, where:
  - x_0 is the number of available positions at the initial time, so that $X_0 = x_0$.
  - Since there cannot be more working nurses than job openings, $X_t \leq Q$.
  - As the number of nurses is positive, $X_t \geq 0$.
- Supply $Y_t \in \{0, 1\}$: Represents supply of nurses at time t, where:
  - Y_t follows a Bernoulli distribution with parameter p and q are mutually independent.
  - Once Q nurses are reached, no additional nurses can be employed: $P(Y_t = 1 | X_t = Q) = 0$.
- Departure $Z_t \in \{0, 1\}$: Represents nurses leaving or staying, where:
  - Z_t follow Bernoulli distributions with parameter q and are independent.
  - Once 0 nurses are reached, we need additional job openings: $P(Z_t = 1 | X_t = 0) = 0$.

Model [cont.]

- As stated in the article:
  - Incoming nurses are mutually independent: $\forall t, t' \in \mathbb{N}, t = t' \Rightarrow Y_t \perp\!\!\!\perp Y_{t'}$,
  - All nurses leaving are mutually independent: $\forall t, t' \in \mathbb{N}, t = t' \Rightarrow Z_t \perp\!\!\!\perp Z_{t'}$
  - Rates of flows are constant over time: $\forall t \in \mathbb{N}, P(Y_t = 1) = p$ and $P(Z_t = 1) = q$.
  - Accounting for the total number of nurses at time $t + 1$ is done by adding the number of new nurses and removing the number of nurses leaving at time t: $\forall t \in \mathbb{N}$, $X_{t+1} = X_t - Z_t + Y_t$.

$$\forall t \in \mathbb{N}, \begin{cases} X_0 = x_0 \\ X_{t+1} = X_t - Z_t + Y_t \\ 0 \leq X_t \leq Q \end{cases}$$

Acknowledgments

Acknowledgements

- PIC Math is a program of the Mathematical Association of America (MAA) and the Society for Industrial and Applied Mathematics (SIAM). Support is provided by the National Science Foundation (NSF grant DMS-1722275).
- Dr. Zvi Rosen (Instructor at FAU)
- Robbie Bok (VP at American Traveler)
- American Traveler (Industry Partner)

References

References

- Juraschek, S. P., Zhang, X., Ranganathan, V., & Lin, V. W. (2019). Republished: United States Registered Nurse Workforce Report Card and Shortage Forecast. *American Journal of Medical Quality, 34*(5), 473-481.
- Liu, J. X., Goryakin, Y., Maeda, A., Bruckner, T., & Scheffler, R. (2016). Global Health Workforce Labor Market Projections for 2030. *Health Nutrition and Population Global Practice Group*.
- National Council of State Boards of Nursing. (2022). NCLEX statistical brief: Q4 pass rates.
- Scheffler, R. M., & Arnold, D. R. (2019). Projecting shortages and surpluses of doctors and nurses in the OECD: what looms ahead. *Health Economics, Policy and Law, 14*(2), 274–290.
- Roussel, A., Biard, R., Deschamps, M., Diss, M. (2023, September). Modeling medical material shortage using Markov processes [Review of Modeling medical material shortage using Markov processes].
- Smiley, R. A., et al. (2022). The 2022 National Nursing Workforce Survey. *Journal of Nursing Regulation, 14*.
- State Health Access Data Assistance Center (SHADAC). (2022). Median Out-of-Pocket Medical Spending, by Insurance Coverage: 2018-2022.
- U.S. Bureau of Labor Statistics. (2022). Occupational Employment and Wages, Registered Nurses: 2018-2022.
- U.S. Census Bureau. (2022). State Population Totals: 2018-2022.

Thank you!

Navika Timsina and Daniel Walford

College of Science
Florida Atlantic University

April 17, 2024