Western Carolina UNIVERSITY

Modeling Undergraduate Enrollment with Markov Chains

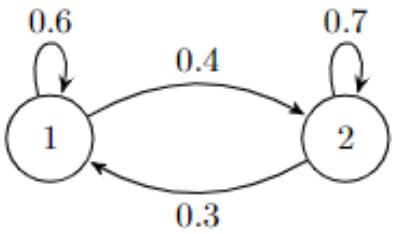
Nathan Hodges
04/03/2023

Western Carolina University

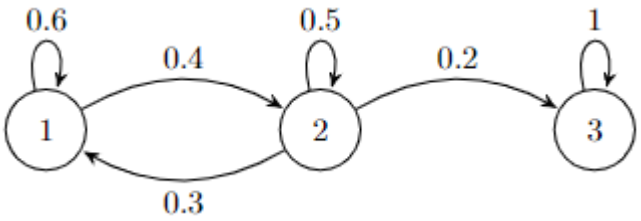
A Markov chain is a mathematical model used to describe a system that changes over time in a probabilistic way. The model consists of a set of states and a set of probabilities that describe how the system moves from one state to another.

At any given time, the system is in one of these states, and the probability of moving to any other state depends only on the current state and not on any previous states. This is called the "memoryless" property, and it allows us to model a wide variety of systems, including weather patterns, stock prices, and even the behavior of people. The Markov chain is named after the Russian mathematician Andrey Markov, who first described the concept in the early 20th century.

- Chat GPT (verified/unedited by actual human Nathan Hodges 20230331.)
https://setosa.io/ev/markov-chains/



Markov chain with transient states.



Markov chain with transient and absorbing states.

Western Carolina UNIVERSITY

Here is a real example of a transition diagram from fall 2019 to fall 2020 for students who were enrolled in fall 2019 with 0-30 credit hours at WCU.

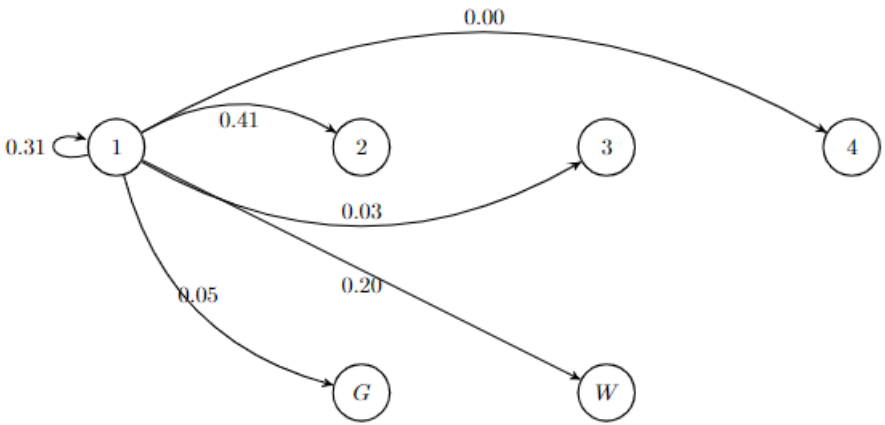
You can see that all possible transitions are shown.

The probability of transitioning from state 1 to a different state after one year is simply the proportion of students from state 1 that transitioned from state 1 to a different state in 2020.

As an example, if there were 100 students enrolled in state 1 in 2021 than we would expect in 2022.

- 31 students to continue at 1
- 41 students to continue to 2
- 3 students to continue to 3
- 5 students to graduate
- 20 students to withdraw

- state 1: 0 - 30 SCH
- state 2: 31 - 60 SCH
- state 3: 61 - 90 SCH
- state 4: > 90 SCH
- state G: Graduate
- state W: Withdraw

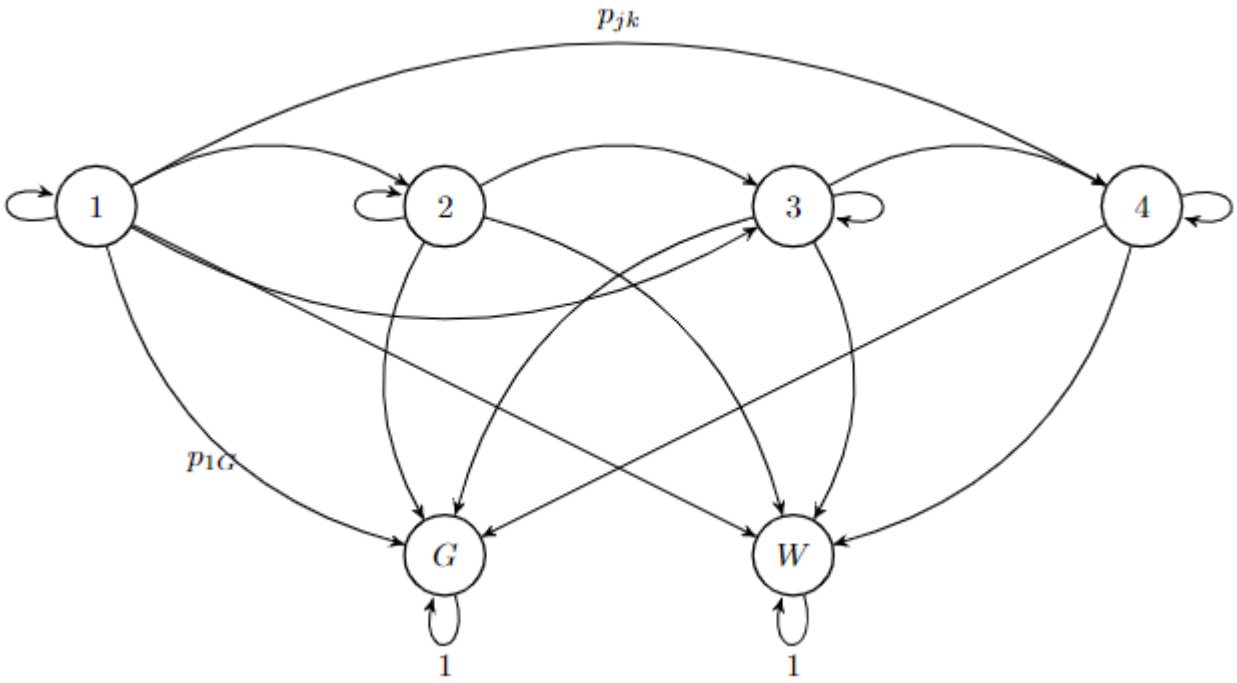


$$p_{11} + p_{12} + p_{13} + p_{14} + p_{1G} + p_{1W} = 1$$
$$0.31 + 0.41 + 0.03 + 0.00 + 0.05 + 0.20 = 1$$

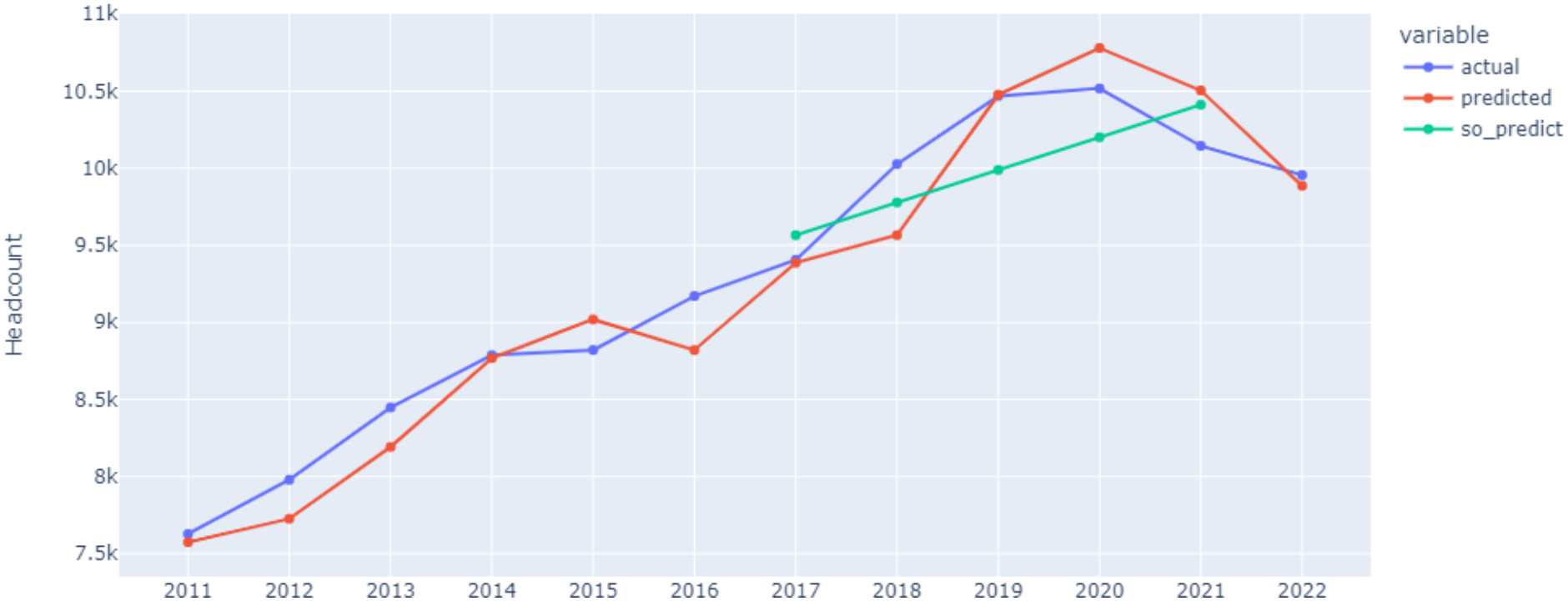
Western Carolina University

This diagram illustrates the model for all transitions. p_jk represents the proportion of students that transitioned from state j to state k.

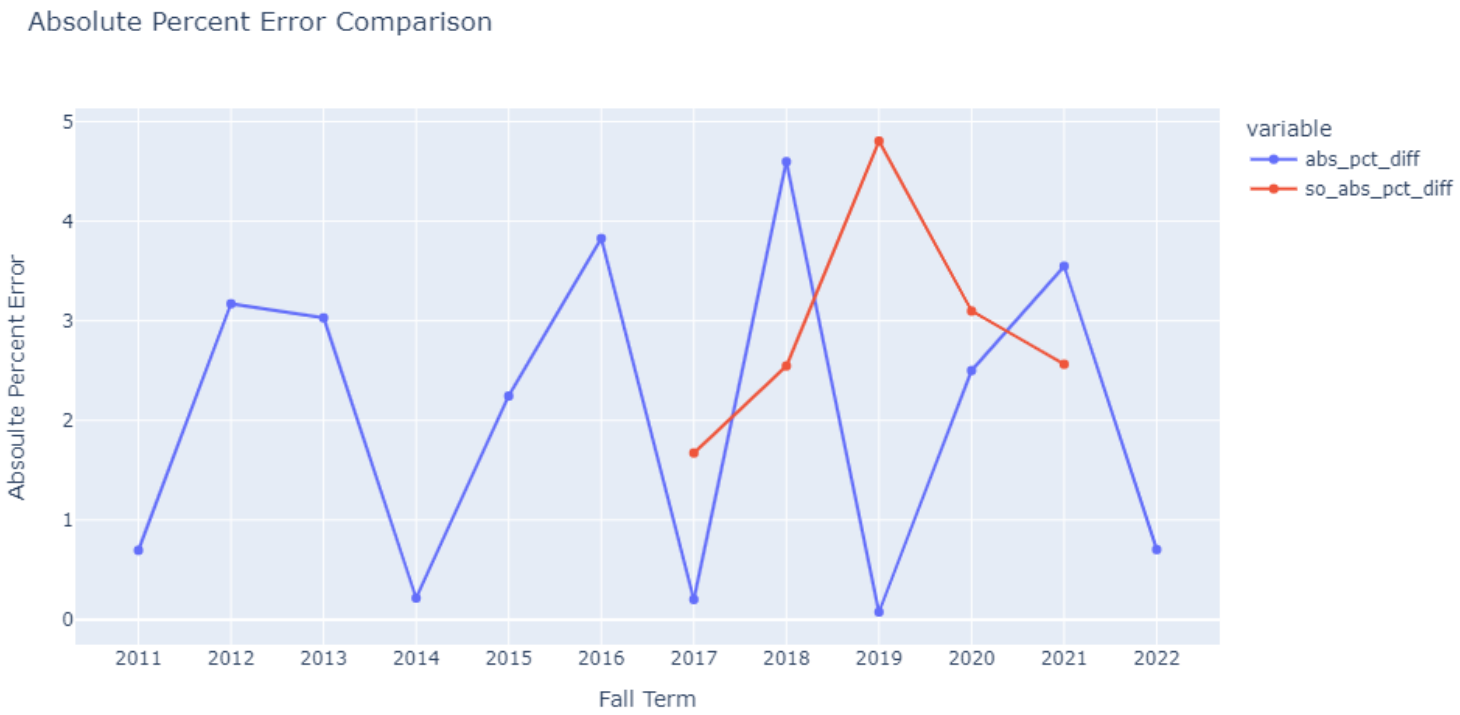
It is assumed that students who re-enroll after withdrawing will behave the same as other students in that same state.



Western Carolina UNIVERSITY



The red predicted line shows the results of the Markov Model. The green line shows the available years for comparison to the current model found on Insight.



That average absolute difference for the terms 2017 – 2021 for the Markov model was 2.185 and the existing enrollment model was 2.937 for undergraduates at WCU.

$$T_{(i,i+1),j,k} = \begin{bmatrix} p_{11} & p_{21} & p_{31} & p_{41} & 0 & 0 \\ p_{12} & p_{22} & p_{32} & p_{42} & 0 & 0 \\ p_{13} & p_{23} & p_{33} & p_{43} & 0 & 0 \\ p_{14} & p_{24} & p_{34} & p_{44} & 0 & 0 \\ p_{1G} & p_{2G} & p_{3G} & p_{4G} & 1 & 0 \\ p_{1W} & p_{2W} & p_{3W} & p_{4W} & 0 & 1 \end{bmatrix} \quad \bar{h}_{(i+1),j} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ 0 \\ 0 \end{bmatrix} \quad \bar{n}_{(i+1),j} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ 0 \\ 0 \end{bmatrix} \quad H_{(20,21),j,k} = \begin{bmatrix} 1769 & 1 & 0 & 0 & 0 & 0 \\ 1759 & 378 & 0 & 0 & 0 & 0 \\ 131 & 1272 & 173 & 0 & 0 & 0 \\ 0 & 110 & 823 & 179 & 0 & 0 \\ 258 & 565 & 573 & 877 & 1 & 0 \\ 1181 & 293 & 120 & 55 & 0 & 1 \end{bmatrix}$$

Element wise multiplication.

$$T_{20-21} \cdot \bar{h}_{21} = \begin{bmatrix} 5154 \\ 2219 \\ 1630 \\ 1142 \\ 0 \\ 0 \end{bmatrix} \odot \begin{bmatrix} 0.347 & 0 & 0 & 0 & 0 & 0 \\ 0.345 & 0.144 & 0 & 0 & 0 & 0 \\ 0.026 & 0.486 & 0.102 & 0 & 0 & 0 \\ 0.000 & 0.042 & 0.487 & 0.161 & 0 & 0 \\ 0.051 & 0.216 & 0.339 & 0.789 & 1 & 0 \\ 0.232 & 0.112 & 0.071 & 0.049 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1788 & 1 & 0 & 0 & 0 & 0 \\ 1778 & 320 & 0 & 0 & 0 & 0 \\ 132 & 1078 & 167 & 0 & 0 & 0 \\ 0 & 93 & 794 & 184 & 0 & 0 \\ 261 & 479 & 553 & 901 & 1 & 0 \\ 1194 & 248 & 116 & 57 & 0 & 1 \end{bmatrix} = \hat{H}_{(21,22),j,k}$$

Now you add the rows to get total headcount and then add the new students from the current year. You then add up the first four columns which represents the students that continued.

$$\hat{h}_{(22,j)} = \begin{bmatrix} 1789 \\ 2099 \\ 1377 \\ 1071 \\ 2194 \\ 1615 \end{bmatrix} + \begin{bmatrix} 3384 \\ 82 \\ 54 \\ 30 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 5173 \\ 2181 \\ 1431 \\ 1101 \\ 2194 \\ 1615 \end{bmatrix}$$

Expected headcount in 2022.

$$5173 + 2181 + 1431 + 1101 = 9886$$

Western Carolina University

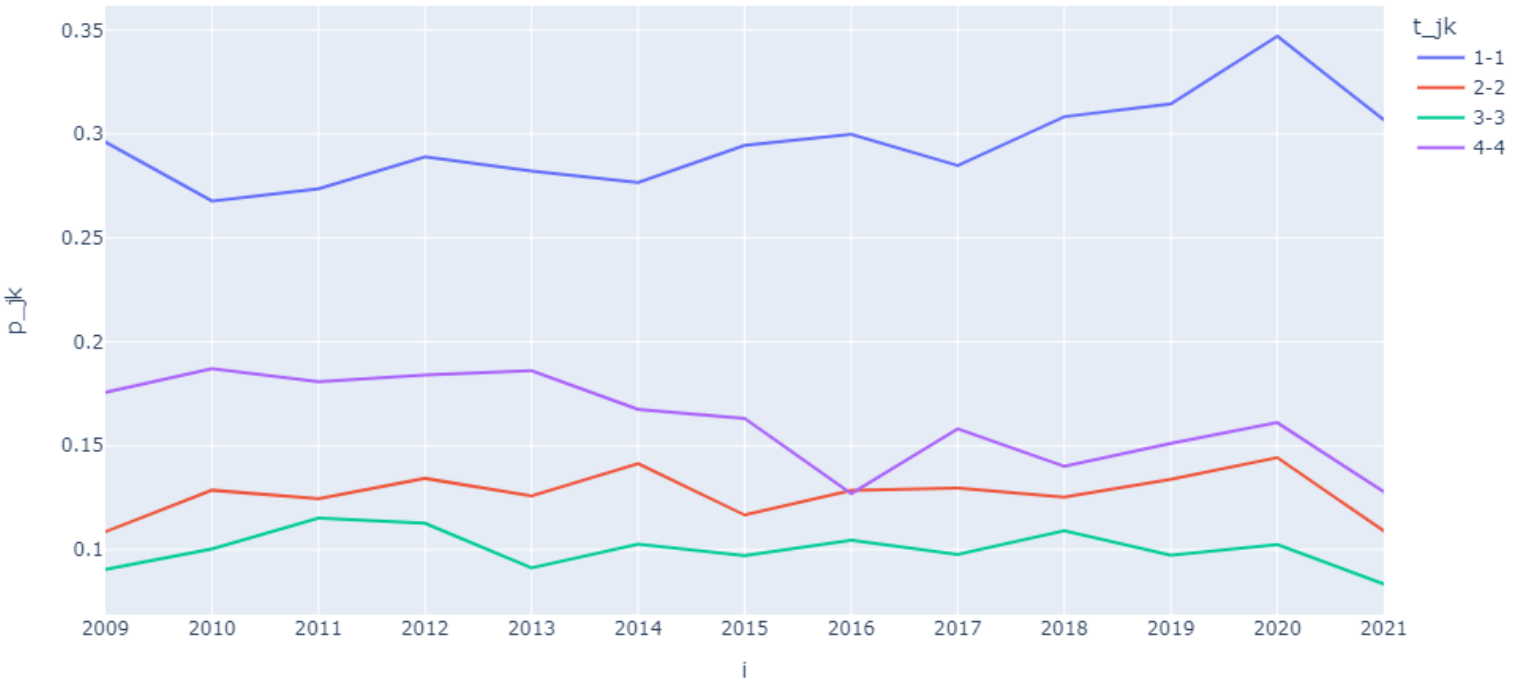
```sql
with base as (
SELECT
    substring(snapshot_term_code,1,4) term_code
    ,snapshot_term_type
    ,case when [career].[cum_inst_earned_hours] <= 30 then '1'
      when [career].[cum_inst_earned_hours] > 30 and [career].[cum_inst_earned_hours]<=60  then '2'
      when [career].[cum_inst_earned_hours] > 60 and [career].[cum_inst_earned_hours]<=90  then '3'
      when [career].[cum_inst_earned_hours] > 90  then '4'
      else '1'
      end as stdnt_level
      ,student_cid
  FROM [SRC_UNCInsight].[wcu_sdm].[career]
  where snapshot_type_code = '1'
  and primary_career_flag = 'Y'
  and career_code = 'U'
  and exclude_from_ipeds_flag = 'N'
  )
  select term_code
  ,snapshot_term_type
  ,stdnt_level
  ,count(distinct student_cid) headcount
  from base
  group by term_code, snapshot_term_type,stdnt_level
  order by term_code, snapshot_term_type,stdnt_level
```

106 %

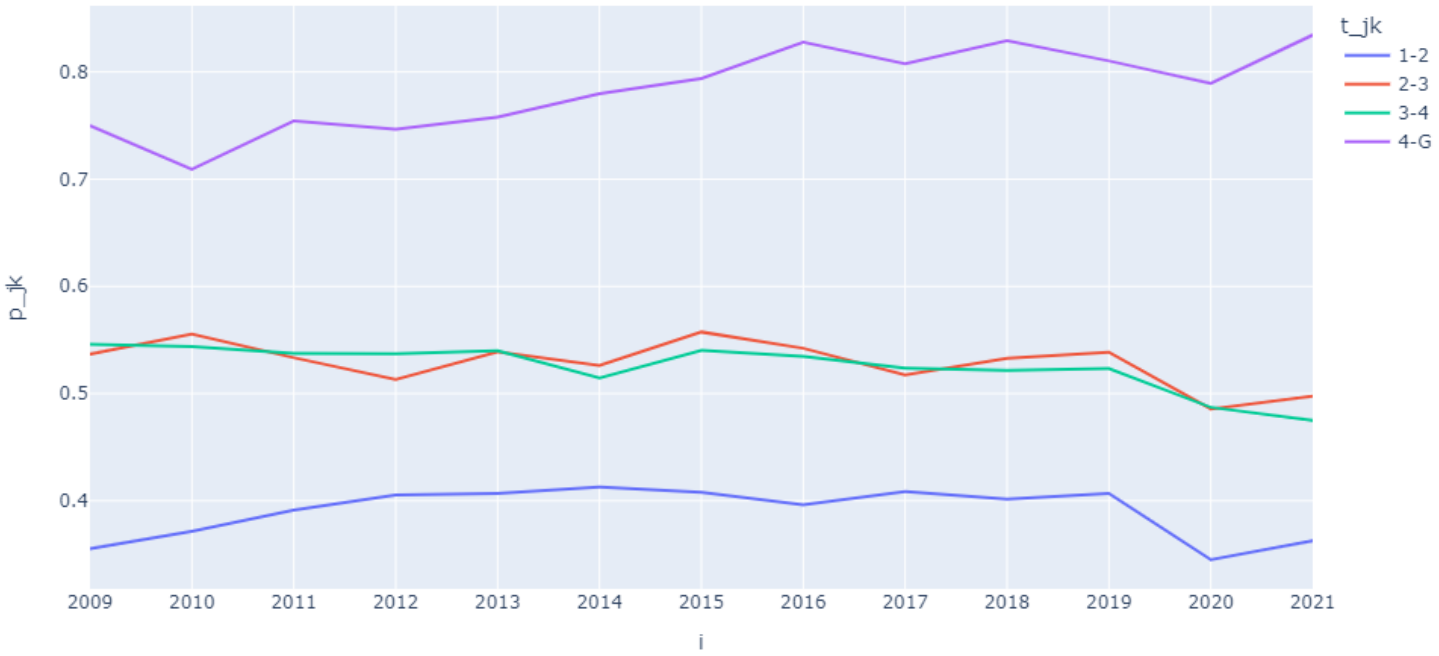
Results | Messages

	term_code	snapshot_term_type	stdnt_level	headcount
156	2019	Summer II	4	446
157	2020	Fall	1	5098
158	2020	Fall	2	2619
159	2020	Fall	3	1689
160	2020	Fall	4	1111

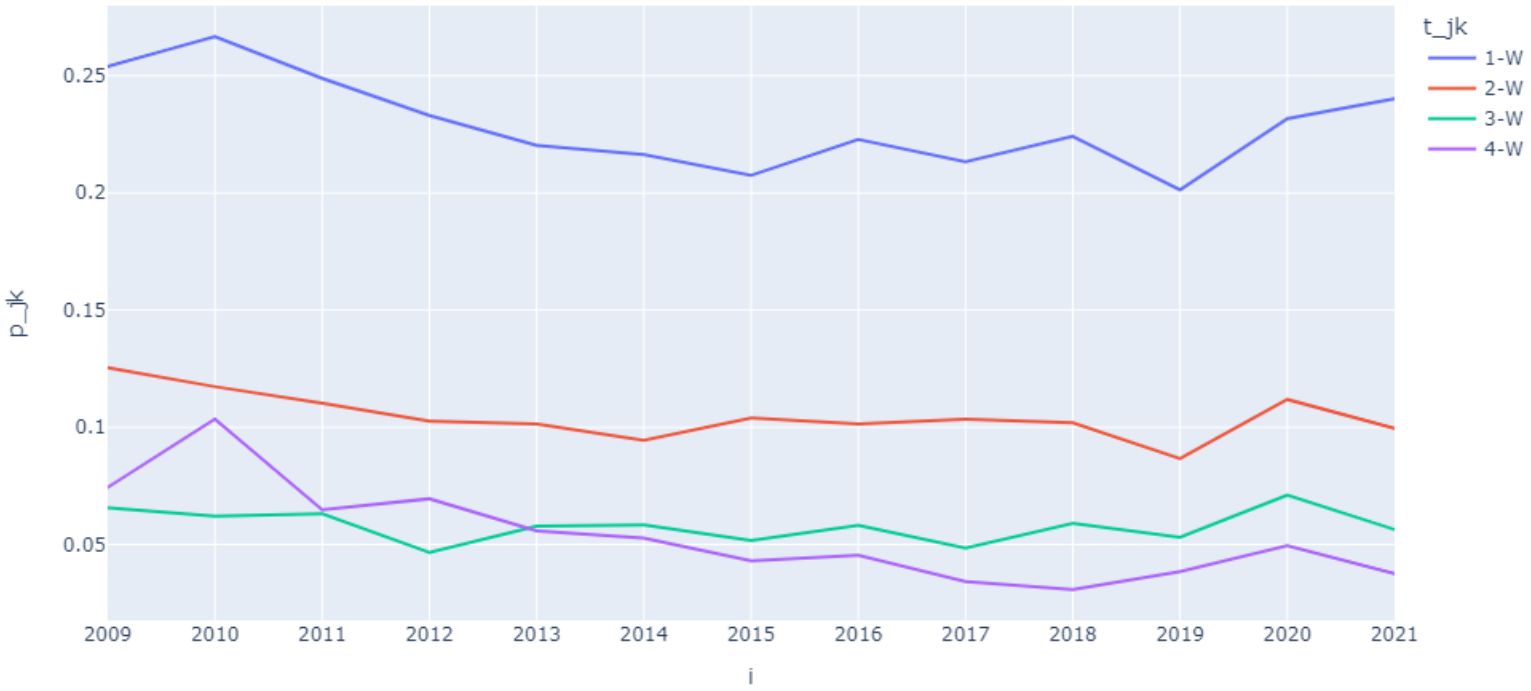
As you can see most people don’t have institutional credit. This model could possibly be improved by using xfer and institutional credit.



You can see that the proportion of students transitioning to the same state increased from 2020 to 2021, but fortunately went down from 2021 to 2022.

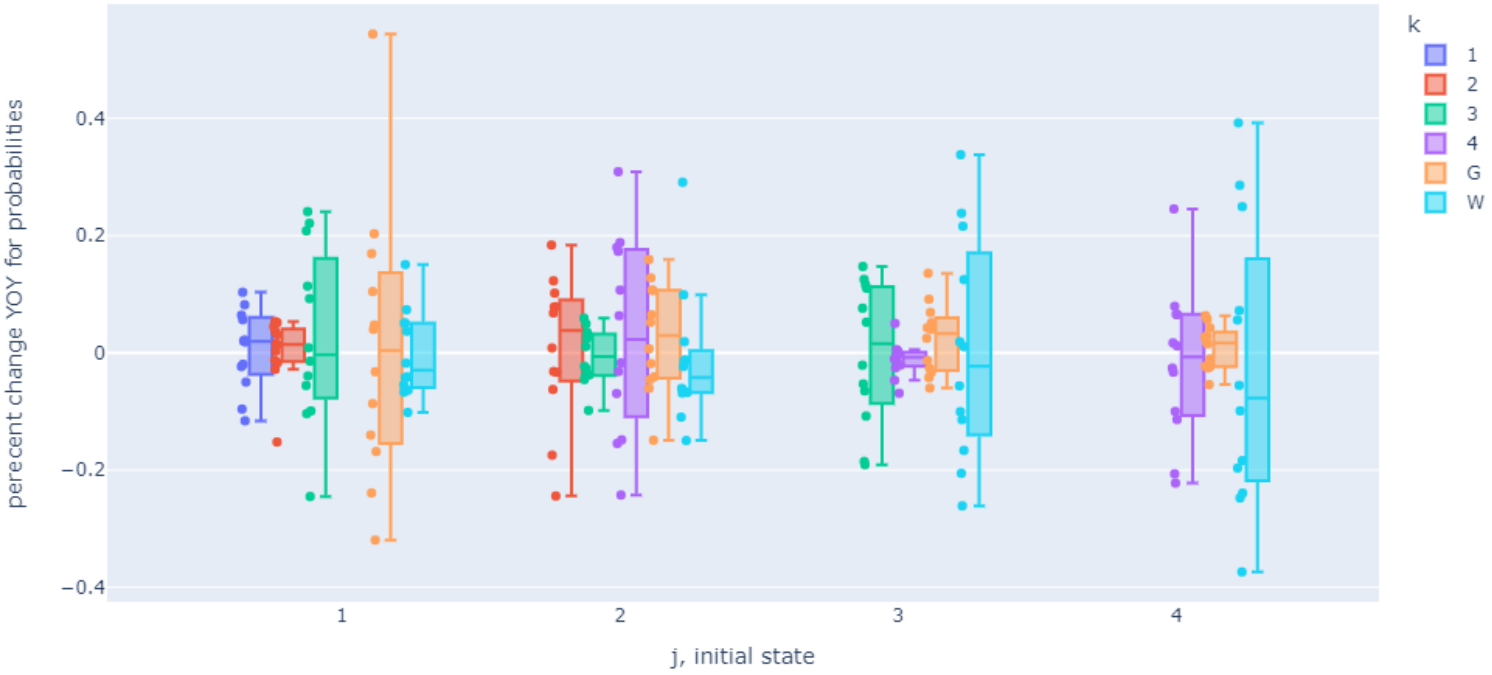


The proportion of students transitioning to the next SCH level decreased from 2020 to 2021.



You can see how withdraw rates in nearly all states decreased then went back up from 2019 to 2020.

A box and whiskers plot showing the distribution of transition probabilities from 2009 to 2021. j represents the student level during the incoming term and k represents what state they transitioned in to.



The best way to look at these probabilities is by looking at the difference between the next term, because after all the model is built on the premise that the same proportion are going to transition the same as last year. This is where you can see the "leaks" in your student profile.

The tables below show the starting point for creating the base table for calculating the probabilities. This will resemble to many the process for creating a retention dataset.

Term	ID	Credit Hours	Level
2014	1	15	1
2015	1	45	2
2014	2	30	1
2016	2	30	1
2014	3	15	1
2015	3	20	1
2016	3	61	3
2014	4	10	1
2015	4	20	1
2016	4	61	3

Table 4.1: Enrollment records ordered by ID and Term.

Term
2014
2015
2016

Table 4.2: Available terms.

Term	ID	Credit Hours	Level
2014	1	15	1
2015	1	45	2
2016	1	Null	Null
2014	2	30	1
2015	2	Null	Null
2016	2	30	1
2014	3	15	1
2015	3	20	1
2016	3	61	3
2014	4	10	1
2015	4	20	1
2016	4	61	3

Table 4.3: Joining these tables gives the needed withdraw & exited terms.

Term	Next Term	ID	j	k
2014	2015	1	1	2
2015	2016	1	2	W
2014	2015	2	1	W
2014	2015	3	1	1
2015	2016	3	1	3
2014	2015	4	1	1
2015	2016	4	1	3

Table 4.4: Join table 4.3 to itself by shifting the terms.

Term	Next Term	j	k	n_{jk}
2014	2015	1	1	2
2014	2015	1	2	1
2014	2015	1	W	1
2015	2016	1	3	2
2015	2016	2	W	1

Table 4.5: Group table 4.4 by term and state transition.

These tables show how to aggregate and pivot the base table.

Term	Next Term	j	k	p_{jk}
2014	2015	1	1	2/4
2014	2015	1	2	1/4
2014	2015	1	W	1/4
2015	2016	1	3	1
2015	2016	2	W	1

Table 4.6: Divide the number of n_{jk} by the total number of students in that state.

Term	Next Term	j	k_1	k_2	k_3	k_w
2014	2015	1	2/4	1/4	0	1/4
2015	2016	1	0	0	1	0
2015	2016	2	0	0	0	1

Table 4.7: To put this table in a matrix form pivot the table.

Western Carolina UNIVERSITY

Resources

[1] Rex Gandy, Lynne Crosby, Andrew Luna, Daniel Kasper, and Sherry Kendrick (2019) *Enrollment Projection Using Markov Chains: Detecting Leaky Pipes and the Bulge in the Boa*, AIR Professional File, Fall 2019 Article 147.

[2] Zachary T. Helbert (2015) *Modeling Enrollment at a Regional University using a Discrete-Time Markov Chain*, Undergraduate Honors Theses. Paper 281. https://dc.etsu.edu/honors/281

[3] Alenka Brezavšček, Mirjana Pejić Bach, Alenka Baggia (2017) *Markov Analysis of Students' Performance and Academic Progress in Higher Education*, Organizacija, Volume 50

Western Carolina UNIVERSITY

THE END

Nathan Hodges
nhodges@wcu.edu