

Research Topics

Statistics - Will Lieber, Wania Iftikhar Khan, AJ Tenser, Kami Akala

Research Topic 1

Our first topic of interest has to do with physical therapy outcomes depending on different approaches to doing rehabilitation exercises. While we don't currently have relevant data sets for this inquiry, we are aware there is extensive research done on ACL surgery recovery especially and that this is an area of interest in the medical community.

This is of interest as one of our team members is currently in rehab for an ACL injury and it's curious to consider if there are any specific ways for the patient to increase their likelihood of a successfully regaining strength in their injured limb.

Apart from our team mate, this research would be of interest to anyone currently undergoing or about to undergo the rehab process as well as physical therapist and other medical professionals. There are many different approaches to rehab to date, so the ability to investigate their success rates and potentially reveal new insight will be quite valuable.

Research Questions

1. Is there a difference in successful return to activity timelines for individuals who undergo more rigorous versus a more gradual physical therapy regimen?
2. Does more conscious/intentional guarding of the injury during the rehabilitation process lead to better long term outcomes? Or is there value in allowing the injury to experience more stress?
3. What is the optimal frequency of attending physical therapy? Is there such a thing as over training?

In order to answer these questions, large scale studies or an amalgamation of studies on individuals going through rehab for different injuries would be useful, especially if methods of rehabilitation are outlined so we can make conclusions about potential outcome differences. Seeing that Duke has a sport science clinic on campus, that could be a great place to start looking for publicly available data and studies.

Research Topic 2

Another possible topic is looking into housing loans in North Carolina. There is a large federal database which contains much of the information pertaining to loan transactions - from the loan details (type, amount, purchaser) to the demographic information of the borrower (location, race, gender, income). With over 180,000 observations per year, and data from 2007-2017, the Consumer Financial Protection Bureau database is a rich source to generate a regression model to gain insight and generate predictions.

We decided to look into this topic because banks are faced with the challenging task of either approving or denying a loan amount, and if banks were able to better compare a given loan to what would be expected (as suggested by a model) they would have the tools to properly address risk. Housing is a public good, so the data is easily available, and given the fact that everyone needs housing, there is a vast amount of data to analyze.

This model would be most useful for banks who would benefit from the risk analytics, but this model could also be used by the borrowers themselves to reason what would be the expected amount for them to ask to borrow. In short, this model would help both the lender and the borrower to make more data-informed decisions.

Research Questions

1. What is the total amount a borrower will request for a loan in North Carolina? Using location (county and/or city), property type, type of loan, race, gender, and median income, as well as possibly interaction effects, we envision creating a model to predict the expected amount a borrower will request in North Carolina, likely using the 2017 data.
2. How have demographic factors influenced the amount borrowers request for loans? This would use the aforementioned demographic variables but would use data from likely 2007 and 2017 to find what shifts have occurred.
3. How do loan characteristics vary by region? This would rely on the metropolitan area variable to create groups. It would also be interesting to look at the observations outside of these areas to compare the urban/rural loan requests (especially regarding FHA/VA types which signals information about the borrower).

[CFMD Data](#): this is a link to the data that we have referenced throughout this section - it contains the information about demographics and loan type. The wide range of variables and combination of both categorical and continuous variables will hopefully allow us to construct a strong model to create meaning.

Research Topic 3

Our final research topic is to investigate lead levels in water and find how it relates to the incarceration rates in the area. This will bring together multiple datasets and allow us to find new trends and possible correlations. In future investigations, we could also incorporate other datasets on other metrics to try to find significant correlations.

While each member in our group has different academic backgrounds, we all find environmental issues very important to analyze. One of the members in our group has designed their major around incarceration, and from previous research, they suggested there was a connection between the two aforementioned phenomena. Through data analysis and fitting a regression model, we hope to cement the correlation or reject its existence.

There are a variety of stakeholders who may find this research interesting. One group would be town and state government who may work towards improving water quality or at least learning about possible knock-on effects of water quality. Another group are social justice advocates who could use this data to focus their demands and have data to either support or dismiss the connection. The results may also be interesting for psychologists who may be interested in framing the actions which lead to incarceration as a function of water quality.

Research Questions

1. Is there a significant relationship between lead levels in water and incarceration?
2. How does average lead level in water supply vary by state and does this have any relationship to the per capita incarceration rate?
3. Is there an interaction effect between racial makeup in a county and incarceration rate?

In order to answer these questions, we would need to access county-level data on lead levels, incarceration rates, and demographic information. It appears that lead levels are published by the EPA, incarceration rates are available through the Bureau of Justice Statistics, and demographic information through the Census database.