Exploratory Study: Heavy Metal Correlations

**Table of Contents:**

Introduction……………………………………………………………………………………………………………………………………….……2

Background……………………………………………………………………………………………………………………………………………..2

Project Objectives……………………………………………………………………………………………………………………………………2

Methods

Task-relevant Data………………………………………………………………………………………………………………………2

Tools……………………………………………………………………………………………………………………………………………2

Data Preprocessing……………………………………………………………………………………………………………………..2

Data Analytics Methods………………………………………………………………………………………………………………3

Results…………………………………………………………………………………………………………………………………………………….3

Discussion………………………………………………………………………………………………………………………………………………11

Conclusion…………………………………………………………………………………………………………………………………………….12

Works Cited…………………………………………………………………………………………………………………………………………..13

Appendix

Other Info…………………………………………………………………………………………………………………………………14

Data-Dictionary…………………………………………………………………………………………………………………………14

**Introduction:**

The purpose of this study is to use various methods of data analysis, mainly regression, to find correlations between interesting measures and heavy metal levels in blood. Below you will find information about the data used in the study, results from analysis as well as anything of interest found. This is mainly an exploratory study; the goal is to try to find attributes that are correlated or that can predict heavy metal exposure. This can then be a jumping off point for subsequent studies.

**Background:**

The data for this project originated from ongoing research in Syracuse. The data is a cache of medical information both laboratory and psychological for children in the Syracuse area. The goal of the larger project is to find the effects of heavy metals on metabolites. The project and various data sets are headed by Dr. Isabelle Bichindaritz of SUNY Oswego.

**Project Objectives**:

Three main questions emerged after data preprocessing. These are:

1). How correlated are race or gender with other attributes and the heavy metals measured (cadmium, lead and mercury).

2). Are aspirin, ibuprofen or daily vitamins correlated with heavy metal levels.

3). Are any of the attributes good predictors for heavy metal levels in blood?

Another more general goal of the project is to discover avenues for further research with any correlation found that was unexpected.

**Methods:**

**Task-relevant Data:**

The database used for this study has 298 rows and 41 attributes. The source data had over 600 attributes. The data was cleaned and stored as a Microsoft Excel sheet. This is a small dataset with a size of 92 kilobytes.

**Tools:**

Data analysis in this project was done using SPSS a data analytics tool from IBM. Algorithms used include binary logistic regression and linear regression. This tool and these algorithms seemed like a good fit for exploratory analysis to find interesting correlations.

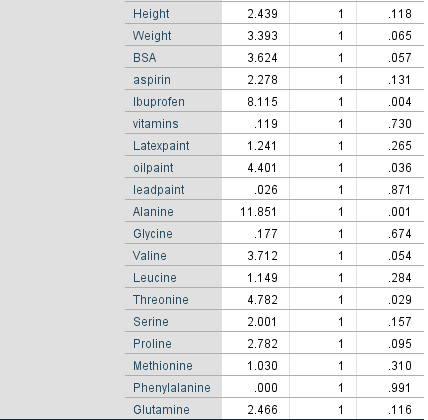
**Data Preprocessing:**

Data about metal exposure and various health and lifestyle measures were taken in Syracuse between 2013 and 2017 from children at Upstate Medical Center. This data comes from datasets acquired from Dr. Isabelle Bichindaritz in Syracuse and is available to students and faculty at request. The dataset used in this study has only 298 tuples, but a massive amount of information was gathered for each subject of the study. Some data is missing for certain subjects. The Dataset has over 600 attributes, but the dataset lacks a comprehensive data dictionary so many attributes which have unknown meaning have been discarded. The data has been cut down to 41 attributes. For the full Data Dictionary of the 41 attributes used please see the Appendix.

**Data Analytics Methods:**

To find answers to the first two questions posed in the project objectives section binary logistic regression was used. For the third question from project objectives linear regression was used. Explanation can be found in the results section.

**Results:**

A screenshot of a cell phone

Description automatically generatedFirst binary logistic regression was done using gender as the dependent variable. From the below table we can see that when not considering the attributes together, the attributes that are correlated with gender are high density lipoprotein, Ibuprofen use, the presence of oil paint and the levels of alanine and threonine. The attributes we are really interested in, the cd, pb and hg levels in blood, do not correlate significantly with gender on their own.

On the next page you can see results when considering attributes together. The attributes with significance with gender are child age, heart rate and ibuprofen use.

So, neither result correlated with heavy metals or tells us anything particularly interesting and a few of the correlations are almost certainly meaningless such as gender being correlated with child age for this data.

A screenshot of a cell phone

Description automatically generatedA close up of a piece of paper

Description automatically generated

The next analysis done was binary logistic regression again but this time with race as the dependent variable. We can see from the below table that socioeconomic score, bmipct, lead levels, systolic blood pressure, heart rate, many of the left ventricle measurements and many other attributes are correlated with race when considered independent of each other. The main correlation to look at is of course lead. This gives some evidence to suggest greater lead risk for children based on their demographic.

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

On the next page is the table for attributes considered together for correlation with race. When considered together lead stops being correlated with race and cadmium becomes correlated. Interestingly again vitamins and ibuprofen use are strongly correlated with race in both cases, this does not have pertinence to the main goal of the study, but it is an interesting observation.

A picture containing cellphone

Description automatically generatedA close up of text on a white background

Description automatically generated

A quick look at aspirin, ibuprofen and vitamins was done to see if use of these common supplements/ drugs were correlated with lead, cadmium or mercury levels. Again, the algorithm used was binary logistic regression. Each of the following three figures only show the results for the metals. No correlation was found.

Aspirin:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Ibuprofen:

A screenshot of a cell phone

Description automatically generated

A picture containing clock

Description automatically generated

Vitamins:

A screenshot of a cell phone

Description automatically generated

**A screenshot of a cell phone

Description automatically generated**

A Linear regression was done with each of the metals as the dependent variable.

A screenshot of a cell phone

Description automatically generatedStarting with cadmium we can see two tables for the regression and its fit below. Note that in the ANOVA table the F is higher than 1, but the significance is not less than 0.05 so the model is not a good fit. Even if it was a good fit no attributes significantly predict cadmium.

**A screenshot of a cell phone

Description automatically generated**

**A screenshot of a cell phone

Description automatically generated**

Next was lead and we can see that the model was a better fit than the one achieved with cadmium but still not significant enough. Lead did have a significant predictor in socioeconomic score if we accept the model.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Finally, with mercury, we can see that the model was a very bad fit with a F score less than one and significance very likely due to random chance. The coefficients table has been omitted as nothing was a significant predictor and the fit is too bad to be useful.

A screenshot of a cell phone

Description automatically generated

**Discussion:**

When considering outside work done around heavy metal toxicity and exposure there is a wealth of work already done as this has been an ongoing struggle for modern society and one that scientists have focused on for decades. In that way this project is likely retreading water with finding correlation that probably already exist somewhere in scientific literature. To that end finding information that could point in the direction of some novel insight was desirable.

One longer term goal is the identification of at-risk people. There was a good summary available on the CDC website, for the link please check the works cited section. This outlined known at-risk populations and further research could augment this understanding or even add to this list. The goal would be to find hidden markers for risk that might be overlooked, this was attempted in the study with looking at many broad health measures like different left ventricle measures and amino acid levels. Much is already known as to why these populations are at risk and even what needs to be done to reduce exposure, but perhaps there are missing pieces.

There was an interesting article on high-output cardiac failure available through the NCBI website. This highlighted the rarer form of heart failure indicative of high output that causes the inability to provide blood for the body’s oxygen demand. The article states the following two categories for why this failure occurs: *“(1) there is an increase in the body’s demand for blood from increased metabolism or (2) there is a bypass of the arteriolar and capillary bed causing an increased flow into venous circulation from a lack of resistance, leading to increased oxygen consumption and a low systemic vascular resistance, respectively.”* (Singh, 2019) Consider now an article on Web-MB talking about linking lead levels to blood pressure, follow the link in the works cited section to view. It points out that elevated lead levels have been linked to increased risk for hypertension. Further study could be done to try and map different heavy metals effects on heart health and perhaps various dangers of heart disease. Our study was targeted at children and lots of heart data was taken as shown in the attributes recorded. Now no correlation was found of any heart measures with increased levels of heavy metals, but a study over time could be interesting. It would be hard to orchestrate but testing some of the same subjects many years down the line to see how heart health faired over their lifetime with varying levels of metal exposure could bear fruit.

**Conclusion:**

A few interesting things came out of this study. Gender did not appear correlated with any of the heavy metals. Race was correlated with lead when considering attributes alone and was correlated with cadmium when considering attributes together. The use of aspirin, ibuprofen and vitamins was found to have no correlation with heavy metal levels in blood. Perhaps the most interesting thing was that socioeconomic score was found to be predictive of lead levels although the model for it did not have good enough significance so this point may be moot.

This gives several possibilities for further study, looking at SES scores as well as race in relation to the danger of increased likelihood of exposure could be useful and could be pursued with further research. A larger dataset and a more targeted study could be a good place to start. Once the indicators for exposure are identified then at-risk people can be targeted for intervention.

Works Cited

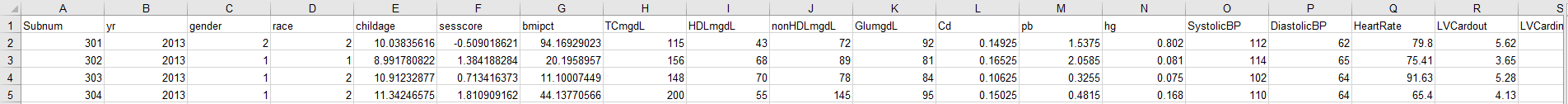
At-Risk Populations. (2019, July 30). Retrieved from <https://www.cdc.gov/nceh/lead/prevention/populations.htm>

Singh, S., & Sharma, S. (2019, November 14). High-Output Cardiac Failure. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK513337/>

Lead Levels Linked to Blood Pressure. (2003, March 25). Retrieved from <https://www.webmd.com/hypertension-high-blood-pressure/news/20030325/lead-levels-linked-to-blood-pressure#>

Appendix:

**Other Info:**

* For the PowerPoint Presentation please look for the PowerPoint file in the zip folder
* The data used in the study can be found as an excel file in the zip folder and a sample of the data format is provided below

**Data-Dictionary:**

Alanine-

Description: Measure of amino acid Alanine

Type: Float

Sample: 451.7324

aspirin-

Description: Aspirin use yes or no. 2 = yes, 1 = no

Type: Integer

Sample: 1

bmipct-

Description: Body mass index by percentile for age and gender (from the CDC standard norms)

Type: Float

Sample: 60.452003

BSA-

Description: Body surface area

Type: Float

Sample: 1.43

Cd-

Description: Cadmium level in blood

Type: Float

Sample: 0.14925

childage-

Description: Age of subject

Type: Float

Sample: 10.53442244

DiastolicBP-

Description: Diastolic blood pressure, a measure of the pressure exerted against your artery walls when your heart is resting between beats

Type: Integer

Sample: 61

gender-

Description: Gender of the Patient, 1 = Male, 2 = female

Type: Integer

Sample: 1

GlumgdL-

Description: Glucose in the blood measured milligrams per deciliter

Type: Integer

Sample: 94

Glutamine-

Description: Measure of amino acid Glutamine

Type: Float

Sample: 14121.1912291

Glycine-

Description: Measure of amino acid Glycine

Type: Float

Sample: 310.203423

HDLmgdL-

Description: high density lipoprotein in milligrams per deciliter. This is considered good cholesterol

Type: Integer

Sample: 53

HeartRate-

Description: Heart rate

Type: Float

Sample: 79.34

Height-

Description: Recorded Height

Type: Integer

Sample: 153

hg-

Description: Mercury level in blood

Type: Float

Sample: 0.075

Ibuprofen-

Description: Ibuprofen use yes or no, 2 = yes, 2 = no

Type: Integer

Sample: 1

Latexpaint-

Description: Latex paint at home 1 = yes, 0 = no

Type: Integer

Sample: 1

leadpaint-

Description: lead paint at home 1 = yes, 0 = no

Type: Integer

Sample: 0

Leucine-

Description: Measure of amino acid Leucine

Type: Float

Sample: 115.4168902

LVCardindex-

Description: Cardiac index, relates the cardiac output from the left ventricle to body surface area. Is a measure of heart performance, normal range is 2.5 to 4 liters per minute

Type: Float

Sample: 3.09

LVCardout-

Description: Left ventricle cardiac output in liters per minute

Type: Float

Sample: 5.28

LVMass-

Description: Left ventricle mass, established measure for predicting bad cardiovascular events

Type: Float

Sample: 102.09

LVMassHt-

Description: Left ventricle mass height

Type: Float

Sample: 61.02

LVMassIndex-

Description: Left ventricle mass index

Type: Float

Sample: 139.5

Methionine-

Description: Measure of amino acid Methionine

Type: Float

Sample: 16.834530203

nonHDLmgdL-

Description: This is Total Cholesterol Minus HDL cholesterol. All non HDL cholesterol considered bad so the lower this number the better

Type: Integer

Sample: 79

oilpaint-

Description: Oil paint at home 1 = yes, 0 = no

Type: Integer

Sample: 0

pb-

Description: Lead level in blood

Type: Float

Sample: 1.0695

Phenylalanine-

Description: Measure of amino acid Phenylalanine

Type: Float

Sample: 64.342903423

Proline-

Description: Measure of amino acid Proline

Type: Float

Sample: 218.34204503

race-

Description: Race of the patient, 1 = African-American, 2 = Caucasian

Type: Integer

Sample: 2

Serine-

Description: Measure of amino acid Serine

Type: Float

Sample: 202.672342

sesscore-

Description: SES score (Socioeconomic Status) is a combined measure of the family’s social position relative to others. Higher scores mean higher socioeconomic status.

Type: Float

Sample: 1.38234

Subnum-

Description: Patient ID

Type: Integer

Sample: 34

SystolicBP-

Description: Systolic blood pressure, a measure of the force exerted on your artery’s walls when the heart beats

Type: Integer

Sample: 108

TCmgdL-

Description: Total Cholesterol in milligrams per deciliter. According to Medlineplus.gov a healthy level of cholesterol for persons under the age of 19 is less than 170mg/dl total cholesterol.

Type: Integer

Sample: 124

Threonine-

Description: Measure of amino acid Threonine

Type: Float

Sample: 64.3425205

Valine-

Description: Measure of amino acid Valine

Type: Float

Sample: 238.5067

vitamins-

Description: Daily vitamins taken yes or no. yes = 2, no = 1

Type: Integer

Sample: 1

Weight-

Description: Subject recorded weight

Type: Float

Sample: 41.9

yr-

Description: Year the study was done

Type: Integer

Sample: 2014