EXCEL Homework 1

***Data Project Name:***

Predicting child abuse fatalities or maltreatment:

1. Center for Applied Research in Economics (CARE) Predictive Risk Modeling (PRM) – New Zealand
2. Los Angeles County Department of Children and Family Services, Approach to Understanding Risk Assessment (AURA)
3. Similar program for the Department of Human Services in Pennsylvania's Allegheny County

***Describe the Algorithm:***

Based on a child’s “risk score” social workers could more efficiently determine which cases need immediate attention to make better use of limited resources.

The risk score would be based on factors such as: age of parents, request for benefits before age 2 (age 5), change in home care givers, child’s protection history, care giver’s protection and benefit history.

***Links that describe results and its data sources (up to 3):***

Program data sources:

Hospital records, medical services, state and county benefits, police reports, birth certificates, school records, child protective services (reports of suspicion and reports of interactions)

New Zealand Report:

<https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/vulnerable-children/auckland-university-can-administrative-data-be-used-to-identify-children-at-risk-of-adverse-outcome.pdf>

Los Angeles County program:

<http://www.scpr.org/news/2015/01/13/49191/can-an-algorithm-predict-child-abuse-la-county-chi/>

Allegheny County article:

<http://www.vice.com/read/how-child-protection-agencies-are-trying-to-predict-which-parents-will-abuse-kids>

Commission to Eliminate Child Abuse Neglect and Fatalities Final Report

(Chapter 6 includes a recommendation to use predictive analytics and outlines some difficulties.)

<http://www.acf.hhs.gov/cb/resource/cecanf-final-report>

***Provide one sentence descriptions to 3 problems they faced with data, and give a one sentence solution they found for each one.***

1. Privacy (protecting the privacy of individuals) - One solution is to only provide the social worker with a “risk factor”, a numeric value with no specific information.
2. Sharing data among public programs (cost and sharing authority) - “In the past few years, new methods have emerged to facilitate the electronic exchange of selected pieces of information between systems without sharing complete case files.”
3. Differences in how programs categorize data - Set universal definitions for what is categorized as a “life-threatening injury” etc. and implement standardized data formats for dates etc.

***Provide a one sentence example of why or why it did not work or is or is not working overall.***

It seems none of these programs have been implemented. They are in the idea stage. However, Los Angeles County Department of Children and Family Services checked the analytics against past cases and found the “risk factor” would have been effective in identifying abuse 76% of the time.

To test the algorithm, DCFS gave SAS 2013 county records for families who were investigated for serious abuse, along with demographic information, such as the ages of the family members. These records were "de-identified" which means names were kept anonymous, and replaced with numbers.

The algorithm gave each child a score ranging from zero to 1,000, reflecting each child's risk of abuse. Children who score 800 or above would be considered at highest risk of serious abuse, giving social workers a clear indicator to act quickly.

LeRue said SAS's algorithm accurately identified children in the 800 or higher range - the most serious cases of abuse — 76 percent of the time.