

# INFO 370 Project Proposal

Rosemary Adams, Jessica Libman, Calvin Korver, Anushree Shukla

## **What is the purpose of your research project? How does it fit into the broader family of research in this area?**

Mental illness affects an estimated 18% of employees aged 15 to 54 in the United States, potentially affecting not only an employee's health but also work performance and culture (2010). By investigating and analyzing the possible variables that correlate with mental illness, we can create tools and solutions that attempt to decrease the number of employees affected by mental illness. This is the purpose of our research.

Studies in the broader family of research in this area usually focus on either workplace productivity or mental health illness in general. For example, "What benefits maximize workplace productivity?". Or, "How does family health history correlate with mental illness?" This research project will specifically focus on discovering and analyzing correlation that may exist between workplace conditions, employee habits, and health history with mental health illness. Below, we highlight the specific hypotheses that we are going to test.

## **What specific hypothesis are you going to test?**

What factors are most influential towards the social acceptance of mental health in the workplace?

## **What is the dataset you'll be working with? Please include background on who collected the data, where you accessed it, and any additional information we should know about how this data came to be.**

We will be using data obtained from [Kaggle](#) that was originally sourced from OSMI, an organization for [mental health data](#). More specifically, Open Sourcing Mental Illness (OSMI) is a "non-profit, 501(c)(3) corporation dedicated to raising awareness, educating, and providing resources to support mental wellness in the tech and open source communities". (Source: <https://osmihelp.org/about/about-osmi>). The survey itself is a 2014 Mental Health in Tech survey that has 1260 responses to questions about the employee's attitudes towards mental health, options available for treatment through their employer, their level of comfort talking to their superiors about seeking treatment. It does not actually specifically ask about if the employee has a mental illness, which is why our question is

about the social acceptance of mental health in the workplace, not about actual mental health prevalence.

### **What statistical and machine learning methods do you plan on using to test your hypothesis?**

To test our hypothesis we will need to run regressions on several different combinations of covariates, and then determine which model best fits the data. Because we are going to be looking at variables where the responses are either binary or responses that are on a scale we are going to use a combination of logistic regression on our binary outcomes and linear regression on our scaled responses. We will be comparing our models against each other using chi-squared tests, and our outcome statistic of interest is an odds ratio that indicates the likelihood of social acceptance of mental health illness based on the occurrence of our tested covariates. We aren't trying to predict anything beyond that, so we will not be using machine learning.

### **Who is your target audience? Depending on the domain of your data, there may be a variety of audiences interested in using the dataset. You should hone in on one of these audiences.**

There are a variety of audiences that can benefit from the our statistical project. Human Resources and Consultancy firms looking to improve workplace culture can use our results to express the **need and urgency** to introduce better mental health awareness programs and even request the appropriate groups such as the Office of the CEO and/or Finance Department to invest more resources into Mental Illness.

### **What should your audience learn from your resource? Please list out at least 3 specific questions that your project will answer for your audience.**

Our audience should be able to grasp the extent to which mental health is prioritized in the workplace. Specifically they should be able to find answers to the questions below:

1. What elements of a workplace make employees more likely to seek out assistance for mental health illness?
2. What elements of a workplace make employees less likely to seek out assistance for mental health illness?
3. What can a company do to increase openness and awareness of mental health issues in the workplace?

## **Technical Description**

### **What will be the format of your final web resource (Shiny app, HTML page or slideshow compiled with KnitR, etc.)?**

The final format of our project will be a Shiny app that will allow users to interactively explore our dataset and allow us to easily show graphics to help explain our findings.

### **Do you anticipate any specific data collection / data management challenges?**

Our raw dataset is organized as a series of responses to questions - so each column is the question that was asked in its entirety. We will need to neatly summarize the goal of the question and rename our columns for sake of simplicity. There are also lots of missing values in our set, which are denoted with an empty string instead of an NA, which will require some custom wrangling. As far as managing the data, as we make changes to it we will be tracking those changes using Github.

### **What new technical skills will need to learn in order to complete your project?**

While everyone has had at least some exposure to R, Shiny is a skill that needs to be refreshed for most. There will also be a learning curve for any new packages we might want to use to help with the statistical modeling, as well as for plotting our results. However, these are skills that can be Googled, and our members are confident in their skills in that area.

### **How will you conduct you analysis? Please include a detailed description of your intended modeling approach.**

First we will clean the data. This means that we will have to figure out what to do with our null value responses. There are lots of people who responded to some of the questionnaire, but not all of it, and determining what null values are tolerable and which ones aren't will be a challenge. We will also go through and neatly name all of the columns to be manipulated as covariates in our regressions.

After that we will select the covariates of interest. This will likely bleed into running the regressions, because it will be easier to determine if a covariate is interesting if it seems to influence our outcome. We will separate the responses into responses that are on a scale (never, rarely, sometimes, always) and responses that are binary (yes, no). We may decide to alter how many null values are tolerated based on if the response is on a scale or binary. We will also have to decide what "accepted" means to us. This will likely be by aggregating a couple responses to the survey.

We will then run logistic regressions on our binary responses and linear regressions on our scale responses, using our "accepted" as the outcome and each covariate of interest as the

predictor. Our outcome of interest will be an odds ratio, and we can use chi-squared tests to compare across our different regressions if necessary.

Ultimately we hope to identify the strength of the relationships between “accepted” and our different covariates, and determine what the strongest predictors of social acceptance mental health in the workplace are.

### **What major challenges do you anticipate?**

The biggest challenge we anticipate having is accurately conducting the statistical research we want to do, because there’s no easy way to tell if you are “right” or “wrong” here. Our current approach is to use two regressions on different types of responses, and comparing across those will be challenging. We also anticipate running into limitations in our dataset that will likely influence what kinds of conclusions we can draw about the problem space, specifically in dealing with null values.

## Works Cited

"Mental health problems in the workplace." Harvard Health Publishing, Harvard Medical School, Feb. 2010,  
[www.health.harvard.edu/newsletter\\_article/mental-health-problems-in-the-workplace](http://www.health.harvard.edu/newsletter_article/mental-health-problems-in-the-workplace).