How Mental Health is Addressed in the Tech Workplace

Natalie Alamo A11400731

Hannah Faulstick A11497701

Jimmy Lozano A12638975

Rita Navarro A12031872

University of California, San Diego

Abstract

We are looking at a data set from a survey concerning mental health, that was given to people employed in the tech industry. They were asked a variety of questions about mental health, and how it is addressed in their workplace. Specific questions inquired if their health care benefits covered mental health care, how comfortable they felt talking to supervisors or employees about mental health, and many other questions of that regard. The data was found on the website kaggle.com, and the data set was uploaded by Open Sourcing Mental Health. We hypothesized that as more resources are provided for employees' mental health they will become more comfortable discussing their mental health with coworkers. We use logistic regression to predict whether an employee would be willing to discuss their mental health based off of the health care provided. Due to vague answers (e.g., I don't know, uncertain, etc.) to questions we were unable to find any significant results that supported the hypothesis. This suggests that perhaps employees should be better informed of the mental health benefits that they receive as a part of their overall health care insurance.

Introduction

Often times mental health issues are overlooked, and that trend can especially be seen in the workplace. Employers tend to sympathize with physical conditions because they can be easily seen and diagnosed. However mental health issues may not receive the same attention because they are still stigmatized, making it difficult to seek help. An important question to address when looking at this data set is whether there is a correlation between the mental health benefits provided to employees and how open they are to talking about mental health problems with coworkers and supervisors. This question is important and worth investigating because it can make the workplace, for all industries, a happier and healthier place to be. If employees are given the support and proper care that they need for all of their health issues (not just the physical ones) it can minimize the time that they have to miss work due to untreated mental health conditions, that may interfere with their work and productivity. Additionally, it is

important to recognize how comfortable an employee is discussing their mental health with their colleagues, the more they are willing to talk about it the more likely they are to receive support in the workplace. We hypothesize that as more resources are provided for employees' mental health they will become more comfortable discussing their mental health with coworkers.

Materials and Methods

fa

Mental Health in Tech Workplace Survey (variables used)	
Variable	***Question***
benefits	Does your employer provide mental health benefits?
wellness_program	Has your employer ever discussed mental health as part of an employee wellness program?
seek_help	Does your employer provide resources to learn more about mental health issues and how to seek help?
mental_health_consequence	Do you think that discussing a mental health issue with employer would have negative consequence?
coworkers	Would you be willing to discuss a mental health issue with your coworkers?
supervisor	Would you be willing to discuss a mental health issue with your direct supervisor(s)?

We primarily focused on the variables that were closely related to mental health. More logistically based variables (e.g., no_employees, state, age, etc.) were left out because they would negatively impact the prediction. The variables that were used, consisted of categorical variables represented by strings (e.g., 'yes', 'no', 'I don't know', etc.). This screening of the variables ensured relativity to our overall research as well as the regressions that were fit onto the specific data. It should also be noted that certain logistical variables were not analyzed in their entirety and there may be some fluctuation and inaccuracies in certain regressions. An example of this is that location of businesses, while

mainly in the United States, had locations in other countries which may have an effect on the prediction. Moreover, gender and age may also play a role. Even so, leaving such variables out were optimal because it allowed for a simpler model while still retaining the fundamental accuracy of our methods and prediction.

We used logistic regression to predict whether an employee would be willing to discuss their mental health (variables: supervisor, coworkers) based on whether or not they receive mental health benefits (variable: benefits) as part of their health care package. This method is best because it is a simple prediction of an employee's willingness to disclose such information. The method will utilize multiple related variables in order to make an accurate prediction. Each variable is run through the method in order to see if one variable has a better effect than another, or if two variables in conjuction have a larger effect then this will also be shown and addressed.

To begin, we had to convert our data from strings to categorical representation. While performing this there were many responses from participants that were "uncertain" of their mental health care benefits (variable: benefits). This became a problem because it would not fit our binary "yes" (1) or "no" (0). To address this problem we took many different approaches. We began by expelling the "uncertain" responses from the data analysis, however this severely depleted the data that was available to work with. Our next tactic was to code the "uncertain" variables as "no," (0) for the variable: benefits. For the variables: coworkers and supervisor, vague answers such as "some of them" were coded as "yes" (1).

However, when running the logistic regression with the variables: benefits and seek_help, there were no significant findings. Then we ran a multiple logistic regression with the variables benefits, seek_help, and coworkers and there were still not any significant findings.

When we saw little correlation with the variable coworkers, we moved on to look at the variable supervisor. Again we saw little correlation in the data. However, when we singled out the variable seek_help, we got a p-value just below the threshold that suggests that the variable is statistically relevant.

Then we used a confusion matrix to gain better understanding of the relationship between this variable and the variable supervisor. Our predictor ended up predicting true every time, thus suggesting that the variable seek_help did not give us significant insight to whether or not an employee would feel comfortable talking to their supervisor (variable: supervisor) if the company provided resources for people seeking help for their mental health (variable: seek_help). We did not use cross validation because the data did not produce any statistically significant values meaning that we could not make any predictions from this data set.

After there were not any significant findings using variables that would have supported our hypothesis, we ran a logistic regression using the variables supervisor and mental_health_consequence and found that there was a statistical significance between the variables. We then calculated the confusion matrix and this time our variable was significant enough to produce predictions of both yes and no.

Results:

Using logistic regression, the variables did not produce statistically significant results, contradictory to what we had expected. We began with looking at the the logistic regression with the parameters benefits, seek_help, and wellness_program and a response of coworkers. When we looked at the t-statistic and the

```
logit(Coworkers) ~ 1 + Benefits + Wellness_Program + Seek_Help
Distribution = Binomial
```

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	1.2775	0.088045	14.509	1.0578e-47
Benefits	-0.048933	0.16331	-0.29964	0.76445
Wellness_Program	0.29955	0.24338	1.2308	0.2184
Seek_Help	0.2053	0.23666	0.8675	0.38567

p-value for this logistic regression we saw these correlations were not statistically relevant.

We then moved on to perform logistic regression using the same parameters, benefits, seek_help, and wellness_program, but this time with the response variable of supervisors. We again saw statistically irrelevant results according to our t-statistics and p-values except for seek_help.

```
logit(Supervisor) ~ 1 + Benefits + Wellness_Program + Seek_Help
Distribution = Binomial

Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	0.66858	0.076846	8.7003	3.3098e-18
Benefits	-0.032431	0.14353	-0.22594	0.82125
Wellness_Program	0.33577	0.2136	1.572	0.11596
Seek_Help	0.43811	0.21025	2.0837	0.037187

Because we saw that seek_help's p-value and t-statistic were statistically relevant we wanted to see its relationship alone with the response variable supervisor. So we ran a logistic regression on just those two variables.

```
logit(Supervisor) ~ 1 + Seek_Help
Distribution = Binomial

Estimated Coefficients:
```

	Estimate	SE	tStat	pVal
(Intercept)	0.68128	0.066651	10.222	1.5882e-24
Seek_Help	0.60785	0.16752	3.6285	0.00028503

```
confmat =

     0   393
     0   866
```

After seeing the p-value and t-statistic we tried creating a confusion matrix using this logistic regression as a predictor function. It ended up always predicting that an employee would feel comfortable talking to their supervisor.

After not finding correlations that supported our hypothesis we decided to look into the variable mental_health_consequence and if it was correlated with variables supervisor or coworkers. This variable is not significant to our hypothesis because whether or not an employee believes discussing a mental health issue with your employer (variable: supervisor) would have negative consequences (variable:

mental_health_consequence) does not relate to mental health care benefits being provided by an employer

```
logit(Supervisor) ~ 1 + Consequence
Distribution = Binomial

Estimated Coefficients:
                   Estimate       SE        tStat       pValue
                   ________    ________    _______    __________

    (Intercept)      1.376      0.080147     17.168    4.5893e-66
    Consequence     -2.121        0.1487    -14.263    3.7147e-46
```

(variable: benefits). Regardless, there was a large correlation between these two variables according to the t-statistic and the p-value.

We then chose to calculate a confusion matrix using this logistic regression to predict the response variable. This yielded a confusion matrix with a sensitivity = 0.8915 and a specificity = 0.5038. So it enabled us to predict when someone will be willing to talk to their supervisor but did less for our ability to predict when someone will not want be comfortable talking to their supervisor. This may be a direction to follow for future research.

```
confmat =

   198   195
    94   772
```

Discussion & Conclusion

Based on the analysis, there weren't any results that made us lean towards a certain idea or answer our overall question of willingness to discuss mental health based on benefits provided. This may have been impacted by the overwhelming amount of vague answers by surveyees such as "don't know" or "sometimes." It was difficult to analyze this data as it was not constricted to a binary 0 or 1 answer.

Switching the vague answers to similar binary answers ("don't know" to "no" (0)), did not seem to have a huge impact on the logistic regression either. This perhaps informs us that most workers are unaware of their own mental health benefits and related aspects of the healthcare options available to them. This result contrasts our initial hypothesis that marking yes on any of our predetermined variables would most likely result in a yes to comfortably being able to discuss mental health with coworkers and/or supervisors. It's difficult to discuss mental health when their isn't a concrete understanding of it, or the health care benefits available. If this topic were to be further explored, we would highly recommend being more direct in the survey. For instance, keeping the answers more simple rather than allowing "I don't know" or "maybe" as a response because they do not provide any meaningful feedback. It would also be optimal to provide certain information regarding mental health prior to the survey. This would aid the surveyee in understanding what exactly constitutes as a mental health issue, treatment for said issue, anonymity, etc. The overwhelming amount of vague answers created confounding variables and insignificant variables. Overall, it is clear that mental health is still a hard topic to address in the workplace and requires more research and accessible information to increase discussions had about the topic.

Mental Health in Tech Workplace Survey (Full survey)	
Variable	***Question***
Timestamp	
Age	
Gender	
Country	
state	If you live in the United States, which state or territory do you live in?

self_employed	Are you self-employed?
family_history	Do you have a family history of mental illness?
treatment	Have you sought treatment for a mental health condition?
work_interfere	If you have a mental health condition, do you feel that it interferes with your work?
no_employees	How many employees does your company or organization have?
remote_work	Do you work remotely (outside of an office) at least 50% of the time?
tech_company	Is your employer primarily a tech company/organization?
benefits	Does your employer provide mental health benefits?
care_options	Do you know the options for mental health care your employer provides?
wellness_program	Has your employer ever discussed mental health as part of an employee wellness program?
seek_help	Does your employer provide resources to learn more about mental health issues and how to seek help?
anonymity	Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?
leave	How easy is it for you to take medical leave for a mental health condition?
mental_health_consequence	Do you think that discussing a mental health issue with employer would have negative consequence?
phys_health _consequence	Do you think that discussing a physical health issue with your employer would have negative consequences?
coworkers	Would you be willing to discuss a mental health issue with your coworkers?
supervisor	Would you be willing to discuss a mental health issue with your direct supervisor(s)?
mental_health_interview	Would you bring up a mental health issue with a potential employer in an interview?
phys_health_interview	Would you bring up a physical health issue with a potential employer in an interview?

mental_vs_physical	Do you feel that your employer takes mental health as seriously as physical health?
obs_consequences	Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?
comments	Any additional notes or comments