**Model building**

The goal of the model that we are building is to predict if an employee is going to seek mental health treatment, and this is going to depend on the circumstances at an employee work place.

The type of model that we are going to build is a logit model. This is because we are predicting a binary output. Will the person seek mental health treatment or not.

The method of model building that we are going to use is start with all the data, then drop variables that are not significant, build the model again and repeat.

Right off the at we will drop the field state from the dataset, since over 50% of the values for state are not populated. We will also drop the Timestamp field as this is just the timestamp of when the user filled out the survey.

**Model 1**

This model will predict mental if someone will seek mental health care based on all of the variables in the dataset with the exception of state.

Here are the coeficents of this model: Coefficients: (4 not defined because of singularities)

Estimate Std. Error z value Pr(>|z|)

(Intercept) -8.216e+01 3.595e+01 -2.285 0.02228 \*

Timestamp 5.606e-08 2.548e-08 2.200 0.02779 \*

Age 2.490e-02 1.324e-02 1.880 0.06014 .

Genderfemale\_trans -7.312e-01 1.185e+00 -0.617 0.53710

Genderfluid -1.635e+00 1.252e+00 -1.306 0.19148

Gendergenderqueer 8.869e-01 1.221e+00 0.727 0.46751

Gendermale\_cis -7.113e-01 2.480e-01 -2.868 0.00413 \*\*

Genderunknown 1.483e+01 3.956e+03 0.004 0.99701

CountryAustria -1.695e+01 1.679e+03 -0.010 0.99194

CountryBelgium -1.735e+00 1.624e+00 -1.068 0.28550

CountryBosnia and Herzegovina -1.841e+01 3.956e+03 -0.005 0.99629

CountryBrazil 1.656e-01 1.431e+00 0.116 0.90791

CountryBulgaria 3.516e+00 2.829e+00 1.243 0.21387

CountryCanada 1.155e-01 7.498e-01 0.154 0.87758

CountryChina -1.706e+01 3.956e+03 -0.004 0.99656

CountryColombia -1.760e+01 2.781e+03 -0.006 0.99495

CountryCosta Rica -1.357e+01 3.956e+03 -0.003 0.99726

CountryCroatia 1.558e+01 2.705e+03 0.006 0.99540

CountryCzech Republic -2.031e+01 3.956e+03 -0.005 0.99590

CountryDenmark 1.645e+01 2.797e+03 0.006 0.99531

CountryFinland 1.534e-01 2.869e+00 0.053 0.95735

CountryFrance 7.050e-01 1.435e+00 0.491 0.62335

CountryGeorgia -1.738e+01 3.956e+03 -0.004 0.99649

CountryGermany 1.789e-01 7.962e-01 0.225 0.82218

CountryGreece -1.277e+01 2.759e+03 -0.005 0.99631

CountryHungary -1.813e+01 3.956e+03 -0.005 0.99634

CountryIndia 1.125e+00 1.325e+00 0.849 0.39585

CountryIreland -8.187e-02 8.529e-01 -0.096 0.92354

CountryIsrael -1.642e+01 1.469e+03 -0.011 0.99108

CountryItaly -1.327e+00 1.366e+00 -0.972 0.33121

CountryJapan 1.606e+01 3.956e+03 0.004 0.99676

CountryLatvia -1.240e+01 3.956e+03 -0.003 0.99750

CountryMexico 1.777e+00 2.461e+00 0.722 0.47034

CountryMoldova 1.379e+01 3.956e+03 0.003 0.99722

CountryNetherlands -4.686e-01 8.721e-01 -0.537 0.59104

CountryNew Zealand 3.601e-02 1.130e+00 0.032 0.97458

CountryNigeria -1.279e+01 3.956e+03 -0.003 0.99742

CountryNorway -1.456e+01 3.956e+03 -0.004 0.99706

CountryPhilippines -1.835e+01 3.956e+03 -0.005 0.99630

CountryPoland 5.475e-02 1.141e+00 0.048 0.96174

CountryPortugal -1.389e+01 2.627e+03 -0.005 0.99578

CountryRomania -1.182e+01 3.956e+03 -0.003 0.99762

CountryRussia -1.602e+01 2.022e+03 -0.008 0.99368

CountrySingapore -1.338e+00 1.501e+00 -0.892 0.37245

CountrySlovenia 1.843e+01 3.956e+03 0.005 0.99628

CountrySouth Africa 1.690e+00 1.925e+00 0.878 0.37983

CountrySpain -1.340e+01 3.956e+03 -0.003 0.99730

CountrySweden -1.610e+00 1.311e+00 -1.228 0.21931

CountrySwitzerland 1.180e+00 1.427e+00 0.827 0.40828

CountryThailand -1.456e+01 3.956e+03 -0.004 0.99706

CountryUnited Kingdom 7.107e-01 6.862e-01 1.036 0.30030

CountryUnited States 1.793e-01 6.642e-01 0.270 0.78720

CountryUruguay -1.227e+01 3.956e+03 -0.003 0.99753

self\_employedYes -2.977e-01 3.543e-01 -0.840 0.40084

family\_historyYes 1.002e+00 1.919e-01 5.221 1.78e-07 \*\*\*

work\_interfereNo Response -2.534e+00 5.709e-01 -4.439 9.05e-06 \*\*\*

work\_interfereOften 3.571e+00 3.743e-01 9.540 < 2e-16 \*\*\*

work\_interfereRarely 2.539e+00 3.050e-01 8.325 < 2e-16 \*\*\*

work\_interfereSometimes 2.960e+00 2.715e-01 10.904 < 2e-16 \*\*\*

no\_employees100-500 2.208e-01 4.111e-01 0.537 0.59126

no\_employees26-100 2.780e-01 3.750e-01 0.741 0.45855

no\_employees500-1000 1.611e-01 5.637e-01 0.286 0.77505

no\_employees6-25 1.093e-01 3.514e-01 0.311 0.75582

no\_employeesMore than 1000 -1.209e-01 4.134e-01 -0.292 0.76992

remote\_workYes -2.171e-01 2.168e-01 -1.001 0.31679

tech\_companyYes -1.188e-01 2.538e-01 -0.468 0.63964

benefitsNo 2.747e-01 2.890e-01 0.951 0.34184

benefitsYes 8.175e-01 2.862e-01 2.857 0.00428 \*\*

care\_optionsNot sure -1.570e-01 2.475e-01 -0.634 0.52592

care\_optionsYes 7.424e-01 2.562e-01 2.898 0.00375 \*\*

wellness\_programNo -1.667e-01 3.197e-01 -0.522 0.60197

wellness\_programYes -5.120e-01 3.772e-01 -1.357 0.17473

seek\_helpNo -6.722e-01 2.681e-01 -2.507 0.01216 \*

seek\_helpYes -8.626e-01 3.310e-01 -2.606 0.00917 \*\*

anonymityNo -1.645e-01 4.286e-01 -0.384 0.70107

anonymityYes 5.586e-01 2.472e-01 2.259 0.02387 \*

leaveSomewhat difficult 4.963e-01 3.313e-01 1.498 0.13420

leaveSomewhat easy -2.752e-01 2.440e-01 -1.128 0.25936

leaveVery difficult 3.938e-01 3.743e-01 1.052 0.29278

leaveVery easy 1.501e-01 2.998e-01 0.501 0.61652

mental\_health\_consequenceNo -6.074e-02 2.595e-01 -0.234 0.81493

mental\_health\_consequenceYes -1.733e-01 2.666e-01 -0.650 0.51580

phys\_health\_consequenceNo 8.976e-02 2.462e-01 0.365 0.71547

phys\_health\_consequenceYes -2.202e-03 4.845e-01 -0.005 0.99637

coworkersSome of them 3.955e-01 2.589e-01 1.528 0.12658

coworkersYes 1.078e+00 3.795e-01 2.840 0.00451 \*\*

supervisorSome of them -4.168e-01 2.569e-01 -1.623 0.10467

supervisorYes -3.558e-01 2.983e-01 -1.193 0.23295

mental\_health\_interviewNo 3.469e-01 3.089e-01 1.123 0.26140

mental\_health\_interviewYes 9.515e-01 6.762e-01 1.407 0.15943

phys\_health\_interviewNo 1.385e-01 2.156e-01 0.642 0.52056

phys\_health\_interviewYes 4.248e-01 3.029e-01 1.403 0.16071

mental\_vs\_physicalNo -1.025e-01 2.425e-01 -0.423 0.67258

mental\_vs\_physicalYes -2.825e-02 2.585e-01 -0.109 0.91298

obs\_consequenceYes 2.960e-01 2.843e-01 1.041 0.29768

continentAmericas NA NA NA NA

continentAsia NA NA NA NA

continentEurope NA NA NA NA

continentOceania NA NA NA NA

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The AIC of the Model is 1115.3 and the area under the ROC curve is .9256

**Model 2**

To build this model we are going to start with model1 and drop all of the variables that have a p-value of greater than .1. The variables for this model are: Age, Gender, work\_interfere (amount of work interference of the mental health issue) + family\_history (If the person has a family history of mental health issues) , benefits (employer provides mental health benefits), care\_options (if the person know the care options) , anonymity (if the person can stay anonymous), coworkers ( If the person has coworkers they can talk to)

The coefficients for this model are:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -6.452e-01 4.227e-01 -1.526 0.126963

Age 9.715e-03 9.906e-03 0.981 0.326717

Genderfemale\_trans 1.454e+01 5.739e+02 0.025 0.979793

Genderfluid 2.844e-01 1.209e+00 0.235 0.814105

Gendergenderqueer 5.714e-01 8.891e-01 0.643 0.520458

Gendermale\_cis 6.527e-02 1.816e-01 0.359 0.719289

Genderunknown 1.474e+01 1.455e+03 0.010 0.991917

work\_interfereNo Response 1.164e-01 2.219e-01 0.524 0.600021

work\_interfereOften 7.010e-01 3.004e-01 2.334 0.019599 \*

work\_interfereRarely 2.341e-01 2.615e-01 0.895 0.370669

work\_interfereSometimes 3.504e-01 2.127e-01 1.647 0.099500 .

family\_historyYes -1.402e-01 1.604e-01 -0.874 0.382237

benefitsNo 2.260e+00 2.511e-01 9.000 < 2e-16 \*\*\*

benefitsYes 5.132e-01 1.874e-01 2.738 0.006183 \*\*

care\_optionsNot sure -6.478e-01 1.695e-01 -3.823 0.000132 \*\*\*

care\_optionsYes 2.465e-01 2.219e-01 1.111 0.266631

anonymityNo 1.669e+00 6.234e-01 2.678 0.007411 \*\*

anonymityYes 8.418e-01 1.877e-01 4.486 7.26e-06 \*\*\*

coworkersSome of them 2.160e-01 1.823e-01 1.185 0.235920

coworkersYes 3.495e-01 2.412e-01 1.449 0.147304

The AIC is 1225 and the area under the curve is .7959



**Model 3**

For model 3, I am going to take the variables with a pvalue of .1 or higher from model1 and use them to predict if an employee will seek treatment. They are the amount of interference from work that the issue causes, employer provides mental health benefits, know the care options, and stay anonymous.

The coefecients of this model are:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.1029 0.1968 -0.523 0.601285

work\_interfereNo Response 0.1120 0.2210 0.507 0.612212

work\_interfereOften 0.6785 0.2917 2.326 0.020016 \*

work\_interfereRarely 0.1974 0.2565 0.770 0.441492

work\_interfereSometimes 0.2932 0.2026 1.447 0.147773

benefitsNo 2.2535 0.2493 9.039 < 2e-16 \*\*\*

benefitsYes 0.4985 0.1843 2.704 0.006846 \*\*

care\_optionsNot sure -0.6467 0.1680 -3.849 0.000119 \*\*\*

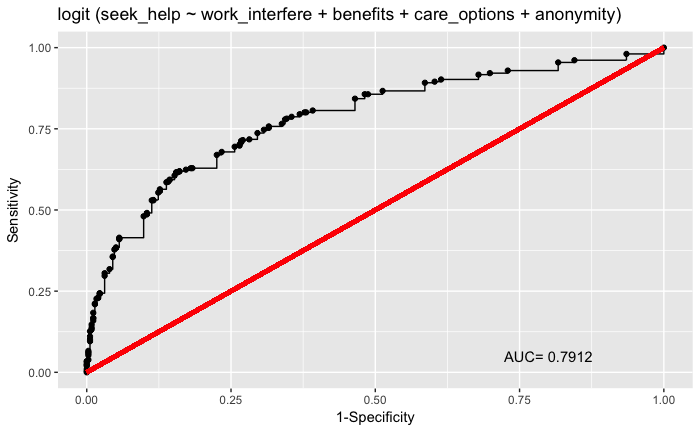
care\_optionsYes 0.2738 0.2199 1.245 0.213138

anonymityNo 1.6010 0.6244 2.564 0.010345 \*

anonymityYes 0.8375 0.1861 4.499 6.83e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1



The AIC is 1214.9 and the Area under the curve is .7912

**Conclusions.**

There is not much difference between model 2 and model 3. Model 1 the model with all the variables is the best model as it has the lowest AIC and the highest area under the curve. Therefore, this is a very well-designed survey for predicting if a person is going to seek mental health care. Also if you are an employer and you want to encourage you employees to seek mental health care the things you can do are the following

1. Provide Benefits
2. Make sure the person knows the care options
3. Make sure the person can stay anonymous

Also the most personally influential factor if someone is going to seek help is if the mental illness interferes with work.

The biggest surprise in the data, is that if the employer provides resource or a wellness program they do not predict very well if the person is going to seek treatment. This is probably because people want to stay anonymous if the seek mental health treatment.