Project Description: Regression

Subject: STA 9890

Submission to: Prof. Kamiar Rahnama Rad

Submission by: Tanay Mukherjee

Date: 20th April 2020

Submit a proposal on the Discussion Board on Blackboard

(a) Describe the response variable and the predictors. How was the data collected?

Dataset:

Financial Distress

Source:

https://www.kaggle.com/shebrahimi/financial-distress

Context:

This data set deals with the financial distress prediction for a sample of companies.

Content:

First column: Company represents sample companies.

Second column: Time shows different time periods that data belongs to. Time series length varies between 1 to 14 for each company.

Third column: The target variable is denoted by "Financial Distress" if it is greater than -0.50 the company should be considered as healthy (0). Otherwise, it would be regarded as financially distressed (1).

Fourth column to the last column: The features denoted by x1 to x83, are some financial and non-financial characteristics of the sampled companies. These features belong to the previous time period, which should be used to predict whether the company will be financially distressed or not (classification). Feature x80 is a categorical variable.

For example, company 1 is financially distressed at time 4 but company 2 is still healthy at time 14.

This data set is imbalanced (there are 136 financially distressed companies against 286 healthy ones i.e., 136 firm-year observations are financially distressed while 3546 firm-year observations are healthy) and skewed, so f-score should be employed as the performance evaluation criterion.

Objective:

To predict the possibility of distress in a financial set up and how it can lead to possible bankruptcy.

(b) Impute missing data-points with their mean. What is n and p?

There are no missing data-points. For the given dataset:

n = 86

p = 3672

```
> # Read the file
> fd <- read.csv("C://Users//its_t//Downloads//FinancialDistress//Project//Financial Distress.csv")

> # Summary of all the columns - predictors and response variables
> glimpse(fd)

Observations: 3,672
Variables: 86

$ Company           <int> 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 4, 4, ...
$ Time              <int> 1, 2, 3, 4, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1, 1...
$ Financial.Distress <dbl> 0.01063600, -0.45597000, -0.32539000, -0.56657000, 1.35730000, ...
$ x1                <dbl> 1.28100, 1.27000, 1.05290, 1.11310, 1.06230, 1.05580, 0.97059, ...
$ x2                <dbl> 0.0229340, 0.0064542, -0.0593790, -0.0152290, 0.1070200, 0.0819...
$ x3                <dbl> 0.87454, 0.82067, 0.92242, 0.85888, 0.81460, 0.87949, 0.90677, ...
$ x4                <dbl> 1.21640, 1.00490, 0.72926, 0.80974, 0.83593, 0.68673, 0.80980, ...
$ x5                <dbl> 0.0609400, -0.0140800, 0.0204760, 0.0760370, 0.1999600, 0.14263...
$ x6                <dbl> 0.188270, 0.181040, 0.044865, 0.091033, 0.047800, 0.043102, -0....
$ x7                <dbl> 0.52510, 0.62288, 0.43292, 0.67546, 0.74200, 0.77198, 0.73660, ...
$ x8                <dbl> 0.018854, 0.006423, -0.081423, -0.018807, 0.128030, 0.119280, 0...
$ x9                <dbl> 0.182790, 0.035991, -0.765400, -0.107910, 0.577250, 0.679730, 0...
$ x10               <dbl> 0.00644890, 0.00179510, -0.05432400, -0.06531600, 0.09407500, 0...
$ x11               <dbl> 0.85822, 0.85152, 0.89314, 0.89581, 0.81549, 0.81574, 0.81340, ...
$ x12               <dbl> 2.005800, -0.486440, 0.412200, 0.994900, 3.014700, 2.364600, 2....
$ x13               <dbl> 0.125460, 0.179330, 0.077578, 0.141120, 0.185400, 0.120510, 0.0...
$ x14               <dbl> 6.97060, 4.57640, 11.89000, 6.08620, 4.39380, 7.29780, 9.72670,...
$ x15               <dbl> 4.6512, 3.7521, 2.4884, 1.6382, 1.6169, 1.2609, 1.5041, 1.5652,...
$ x16               <dbl> 0.0501000, -0.0140110, 0.0280770, 0.0939040, 0.2392100, 0.20769...
$ x17               <dbl> 2.19840, 2.45750, 1.39570, 2.05880, 3.03110, 2.90890, 3.49660, ...
$ x18               <dbl> 0.0182650, 0.0275580, 0.0125950, 0.0116010, 0.0068143, 0.012307...
$ x19               <dbl> 0.0249780, 0.0288040, 0.0681160, 0.0943850, 0.0793460, 0.087834...
$ x20               <dbl> 0.0272640, 0.0411020, 0.0148470, 0.0144150, 0.0088763, 0.015929...
$ x21               <dbl> 1.41730, 1.18010, 0.81652, 0.90391, 1.02510, 0.84185, 0.99557, ...
$ x22               <dbl> 9.5554, 7.2952, 7.1204, 7.9828, 4.7463, 3.8821, 4.5406, 4.9847,...
$ x23               <dbl> 0.148720, 0.056026, 0.065220, 0.125160, 0.266020, 0.239880, 0.2...
$ x24               <dbl> 0.66995, 0.67048, 0.84827, 0.80478, 0.76770, 0.77264, 0.83805, ...
$ x25               <dbl> 214.760, 38.242, -498.390, -75.867, 1423.100, 1748.000, 938.500...
$ x26               <dbl> 12.641, 12.877, 13.225, 13.305, 11.575, 12.048, 12.193, 12.242,...
```

```
$ x27  <dbl> 6.4607, 5.5506, 16.2540, 8.8950, 17.4880, 15.9330, -32.8530, 19...
$ x28  <dbl> 0.0438350, 0.2654800, 0.4165700, 0.0837740, 0.6207700, 0.604710...
$ x29  <dbl> 0.2045900, 0.1501900, 0.0741490, 0.0540980, 0.0469070, 0.106850...
$ x30  <dbl> 0.35179, 0.41763, 0.36723, 0.54360, 0.56963, 0.59646, 0.61731, ...
$ x31  <dbl> 8.3161e+00, 9.5276e+00, 9.3513e+00, 7.0909e+00, 9.4861e+00, 5.5...
$ x32  <dbl> 0.28922, 0.41561, 0.50356, 0.67133, 0.68143, 0.86856, 0.76230, ...
$ x33  <dbl> 0.76606, 0.81699, 0.91962, 0.93701, 0.94242, 0.87851, 0.92421, ...
$ x34  <dbl> 2.5825, 2.6033, 1.4931, 2.3533, 4.1296, 3.8269, 4.6169, 4.8221,...
$ x35  <dbl> 77.400, 95.947, 144.670, 219.750, 222.650, 285.500, 239.350, 23...
$ x36  <dbl> 0.0267220, 0.0075797, -0.0664830, -0.0170000, 0.1312300, 0.1004...
$ x37  <dbl> 1.6307000, 0.8375400, 0.9557900, 0.3833500, 0.2530100, 0.886600...
$ x38  <dbl> 0.0150160, 0.0274250, 0.0172700, 0.0143270, 0.0081518, 0.017922...
$ x39  <dbl> 0.0054783, 0.0454340, 0.0280590, 0.2033700, 0.3530100, 0.318290...
$ x40  <dbl> 0.12730, 0.13774, 0.10242, 0.10143, 0.17612, 0.17690, 0.17835, ...
$ x41  <dbl> 9.6951, 5.6035, 9.4003, 5.7379, 4.5088, 5.6983, 8.6864, 4.8161,...
$ x42  <dbl> -0.736220, -0.643850, -14.032000, 0.722050, -0.113390, 0.228280...
$ x43  <dbl> 0.98559, 1.30190, 0.75746, 1.39120, 1.05270, 0.68127, 0.52272, ...
$ x44  <dbl> 0.180160, 0.046857, -0.579760, -0.150130, 0.607660, 0.463080, 0...
$ x45  <dbl> 1.50060, 1.00950, 0.57832, 0.64508, 0.25782, 0.35765, -0.26440,...
$ x46  <dbl> 0.0262240, 0.0078645, -0.0643730, -0.0177310, 0.1313800, 0.0931...
$ x47  <dbl> 7.05130, 4.60220, 11.98800, 6.11140, 4.41510, 7.33560, 9.78460,...
$ x48  <dbl> 1174.90, 1062.50, 651.15, 703.04, 2465.40, 2571.60, 1150.30, 23...
$ x49  <dbl> 5.33990, 3.73890, 10.93400, 5.70280, 4.14080, 6.41120, 8.98950,...
$ x50  <dbl> 0.85128, 0.94397, 0.93478, 0.87484, 0.73398, 0.76012, 0.75734, ...
$ x51  <dbl> 12.837, 12.881, 12.909, 13.094, 11.396, 11.672, 11.982, 12.110,...
$ x52  <dbl> 0.06173700, -0.00056494, 0.04162500, 0.10840000, 0.25031000, 0....
$ x53  <dbl> 0.180900, 0.056298, 0.047562, 0.101350, 0.222370, 0.164730, 0.1...
$ x54  <dbl> 209.87, 250.14, 280.55, 413.74, 315.34, 412.79, 592.09, 698.67,...
$ x55  <dbl> -0.582550, -0.474770, -1.000000, 0.565000, -0.060101, 0.102450,...
$ x56  <dbl> 0.47101, 0.38599, 0.48844, 0.34408, 0.20242, 0.17945, 0.17540, ...
$ x57  <dbl> 0.109900, 0.369330, 0.053299, 0.073356, 1.229100, 0.611770, 0.1...
$ x58  <dbl> 0.00000000, 0.00000000, 0.00378540, 0.00003660, -0.00249100, -0...
$ x59  <dbl> 0.00000000, 0.00000000, 0.00519070, 0.00004530, -0.00298000, -0...
$ x60  <dbl> 0.220090, 0.000000, 0.000000, 0.000000, 0.226880, 0.224450, 0.2...
$ x61  <dbl> 7.1241000, 7.4166000, 3.6373000, 5.1442000, 7.1241000, 7.416600...
$ x62  <dbl> 15.38100, 7.10500, 7.02130, 9.90990, 15.38100, 7.10500, 7.02130...
$ x63  <dbl> 3.27020, 14.32100, 1.15380, 2.04080, 3.27020, 14.32100, 1.15380...
$ x64  <dbl> 17.8720, 18.7700, 9.8951, -1.4903, 17.8720, 18.7700, 9.8951, -1...
$ x65  <dbl> 34.6920, 124.7600, 6.4467, -21.9070, 34.6920, 124.7600, 6.4467,...
$ x66  <dbl> 30.087, 26.124, 30.245, 34.285, 30.087, 26.124, 30.245, 34.285,...
$ x67  <dbl> 12.8, 11.8, 10.3, 11.5, 12.8, 11.8, 10.3, 11.5, 11.3, 10.5, 10....
$ x68  <dbl> 7991.4, 8322.8, 8747.0, 9042.5, 7991.4, 8322.8, 8747.0, 9042.5,...
$ x69  <dbl> 364.9500, 0.1896, 11.9460, -18.7480, 364.9500, 0.1896, 11.9460,...
$ x70  <dbl> 15.8, 15.6, 15.2, 10.4, 15.8, 15.6, 15.2, 10.4, 11.9, 18.4, 25....
$ x71  <dbl> 61.4760, 24.5790, 20.7000, 47.4290, 61.4760, 24.5790, 20.7000, ...
$ x72  <dbl> 4, 0, 0, 4, 4, 0, 0, 4, 4, 2, 0, 2, 4, 2, 0, 2, 2, 1, 4, 4, 0, ...
$ x73  <dbl> 36, 36, 35, 33, 36, 36, 35, 33, 31, 29, 28, 28, 27, 22, 28, 32,...
$ x74  <dbl> 85.437, 107.090, 120.870, 54.806, 85.437, 107.090, 120.870, 54....
$ x75  <dbl> 27.07, 31.31, 36.07, 39.80, 27.07, 31.31, 36.07, 39.80, 44.53, ...
$ x76  <dbl> 26.102, 30.194, 35.273, 38.377, 26.102, 30.194, 35.273, 38.377,...
$ x77  <dbl> 16.000, 17.000, 17.000, 17.167, 16.000, 17.000, 17.000, 17.167,...
$ x78  <dbl> 16.0, 16.0, 15.0, 16.0, 16.0, 16.0, 15.0, 16.0, 14.0, 12.0, 12....
$ x79  <dbl> 0.2, 0.4, -0.2, 5.6, 0.2, 0.4, -0.2, 5.6, 2.1, -6.4, -13.4, 1.2...
$ x80  <int> 22, 22, 22, 22, 29, 29, 29, 29, 29, 29, 29, 29, 29, 29, 29, 29,...
$ x81  <dbl> 0.06039000, 0.01063600, -0.45597000, -0.32539000, 1.25100000, 1...
$ x82  <int> 30, 31, 32, 33, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19...
$ x83  <int> 49, 50, 51, 52, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38,...
```

(c) Standardize the numerical predictors using equation (6.6) in the ISLR book.

The equation 6.6 in the ISLR as seen in page 217 is to standardize the predictors.

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}}$$

```
> # Standardize the predictors
> fd_std <- fd %>% select(3:86)

> out <- sapply(fd_std, function(x) x/sd(x))

> out_df <- as.data.frame((out))
> glimpse(out_df)
```

```
Observations: 3,672
Variables: 84

$ Financial.Distress <dbl> 4.010215e-03, -1.719197e-01, -1.226856e-01, -2.136205e-01, 5.11...
$ x1                 <dbl> 0.8816692, 0.8740983, 0.7246756, 0.7661093, 0.7311453, 0.726671...
$ x2                 <dbl> 0.19109596, 0.05377917, -0.49477138, -0.12689458, 0.89173669, 0...
$ x3                 <dbl> 4.915792, 4.612989, 5.184926, 4.827767, 4.578869, 4.943616, 5.0...
$ x4                 <dbl> 2.1267524, 1.7569660, 1.2750374, 1.4157485, 1.4615391, 1.200678...
$ x5                 <dbl> 0.48789461, -0.11272655, 0.16393387, 0.60876342, 1.60090921, 1....
$ x6                 <dbl> 0.8941623, 0.8598244, 0.2130801, 0.4323486, 0.2270195, 0.204707...
$ x7                 <dbl> 0.5080272, 0.6026280, 0.4188443, 0.6534984, 0.7178750, 0.746880...
$ x8                 <dbl> 4.378859e-06, 1.491748e-06, -1.891057e-05, -4.367943e-06, 2.973...
$ x9                 <dbl> 0.5280891, 0.1039797, -2.2112772, -0.3117572, 1.6677029, 1.9637...
$ x10                <dbl> 0.0464022372, 0.0129164130, -0.3908814110, -0.4699729445, 0.676...
$ x11                <dbl> 4.248913, 4.215743, 4.421797, 4.435015, 4.037364, 4.038601, 4.0...
$ x12                <dbl> 1.366156e-06, -3.313156e-07, 2.807506e-07, 6.776292e-07, 2.0533...
$ x13                <dbl> 0.7052099, 1.0080129, 0.4360655, 0.7932347, 1.0421323, 0.677386...
$ x14                <dbl> 2.9321418, 1.9250357, 5.0014584, 2.5601241, 1.8482261, 3.069776...
$ x15                <dbl> 4.280014e-05, 3.452666e-05, 2.289815e-05, 1.507464e-05, 1.48786...
$ x16                <dbl> 4.067801e-05, -1.137604e-05, 2.279674e-05, 7.624407e-05, 1.9422...
$ x17                <dbl> 3.650053e-05, 4.080242e-05, 2.317312e-05, 3.418272e-05, 5.03260...
$ x18                <dbl> 0.40249590, 0.60728070, 0.27754919, 0.25564495, 0.15016303, 0.2...
$ x19                <dbl> 7.670271e-06, 8.845163e-06, 2.091713e-05, 2.898385e-05, 2.43656...
$ x20                <dbl> 0.11801393, 0.17791257, 0.06426617, 0.06239623, 0.03842162, 0.0...
$ x21                <dbl> 1.4630877, 1.2182246, 0.8428987, 0.9331120, 1.0582172, 0.869047...
$ x22                <dbl> 0.042045871, 0.032100492, 0.031331333, 0.035126083, 0.020884768...
$ x23                <dbl> 0.8688246, 0.3273048, 0.3810163, 0.7311867, 1.5540931, 1.401382...
$ x24                <dbl> 3.671578, 3.674482, 4.648838, 4.410497, 4.207284, 4.234357, 4.5...
$ x25                <dbl> 0.11783134, 0.02098205, -0.27344925, -0.04162558, 0.78080544, 0...
$ x26                <dbl> 7.832094, 7.978314, 8.193928, 8.243494, 7.171623, 7.464684, 7.5...
$ x27                <dbl> 0.0013573227, 0.0011661206, 0.0034147885, 0.0018687427, 0.00367...
$ x28                <dbl> 0.15332407, 0.92858386, 1.45705958, 0.29302088, 2.17130104, 2.1...
$ x29                <dbl> 2.20010532, 1.61510249, 0.79737822, 0.58175521, 0.50442514, 1.1...
$ x30                <dbl> 2.0608316, 2.4465309, 2.1512812, 3.1844796, 3.3369667, 3.494140...
$ x31                <dbl> 1.004853e-06, 1.151241e-06, 1.129939e-06, 8.568094e-07, 1.14622...
$ x32                <dbl> 2.769167e-05, 3.979301e-05, 4.821387e-05, 6.427719e-05, 6.52442...
$ x33                <dbl> 5.442665, 5.804510, 6.533670, 6.657222, 6.695658, 6.241594, 6.5...
$ x34                <dbl> 3.383305e-05, 3.410555e-05, 1.956094e-05, 3.083033e-05, 5.41014...
$ x35                <dbl> 3.310875e-05, 4.104245e-05, 6.188428e-05, 9.400062e-05, 9.52411...
```

```
$ x36  <dbl> 0.09063223, 0.02570785, -0.22548845, -0.05765840, 0.44508896, 0...
$ x37  <dbl> 3.23323451, 1.66061399, 1.89507157, 0.76007877, 0.50165000, 1.7...
$ x38  <dbl> 3.573840e-05, 6.527209e-05, 4.110297e-05, 3.409857e-05, 1.94014...
$ x39  <dbl> 1.068221e-06, 8.859234e-06, 5.471260e-06, 3.965538e-05, 6.88338...
$ x40  <dbl> 0.6886061, 0.7450794, 0.5540223, 0.5486671, 0.9526890, 0.956908...
$ x41  <dbl> 3.0420484, 1.7582199, 2.9495485, 1.8003909, 1.4147340, 1.787965...
$ x42  <dbl> -0.051515591, -0.045052176, -0.981862445, 0.050524072, -0.00793...
$ x43  <dbl> 0.007230654, 0.009551221, 0.005557007, 0.010206359, 0.007722997...
$ x44  <dbl> 0.0030858595, 0.0008025872, -0.0099303835, -0.0025714925, 0.010...
$ x45  <dbl> 1.9633980, 1.3208385, 0.7566789, 0.8440282, 0.3373339, 0.467952...
$ x46  <dbl> 0.03682692, 0.01104429, -0.09040038, -0.02490002, 0.18449974, 0...
$ x47  <dbl> 1.546800e-04, 1.009556e-04, 2.629733e-04, 1.340620e-04, 9.68513...
$ x48  <dbl> 0.4012883, 0.3628979, 0.2224009, 0.2401240, 0.8420598, 0.878332...
$ x49  <dbl> 2.4698763, 1.7293621, 5.0573284, 2.6377293, 1.9152538, 2.965387...
$ x50  <dbl> 4.973193, 5.514689, 5.461001, 5.110831, 4.287924, 4.440635, 4.4...
$ x51  <dbl> 7.287428, 7.312406, 7.328301, 7.433324, 6.469387, 6.626069, 6.8...
$ x52  <dbl> 5.104473e-05, -4.670977e-07, 3.441594e-05, 8.962613e-05, 2.0695...
$ x53  <dbl> 1.3897747, 0.4325126, 0.3653978, 0.7786272, 1.7083704, 1.265547...
$ x54  <dbl> 0.01453212, 0.01732055, 0.01942625, 0.02864878, 0.02183523, 0.0...
$ x55  <dbl> -1.3982853, -1.1395827, -2.4002837, 1.3561603, -0.1442594, 0.24...
$ x56  <dbl> 3.366785, 2.759061, 3.491375, 2.459488, 1.446900, 1.282711, 1.2...
$ x57  <dbl> 0.0042868101, 0.0144062562, 0.0020790054, 0.0028613579, 0.04794...
$ x58  <dbl> 0.000000000, 0.000000000, 0.142486923, 0.001377667, -0.09376418...
$ x59  <dbl> 0.000000e+00, 0.000000e+00, 1.788190e-05, 1.560580e-07, -1.0266...
$ x60  <dbl> 2.459281, 0.000000, 0.000000, 0.000000, 2.535153, 2.508000, 2.5...
$ x61  <dbl> 1.816156037, 1.890723441, 0.927261598, 1.311417566, 1.816156037...
$ x62  <dbl> 2.21199246, 1.02179354, 1.00975637, 1.42517548, 2.21199246, 1.0...
$ x63  <dbl> 0.37424882, 1.63892646, 0.13204339, 0.23355360, 0.37424882, 1.6...
$ x64  <dbl> 1.4504151, 1.5232930, 0.8030440, -0.1209464, 1.4504151, 1.52329...
$ x65  <dbl> 0.79110500, 2.84498615, 0.14700843, -0.49956005, 0.79110500, 2....
$ x66  <dbl> 4.894583, 4.249879, 4.920287, 5.577518, 4.894583, 4.249879, 4.9...
$ x67  <dbl> 13.43327, 12.38379, 10.80958, 12.06895, 13.43327, 12.38379, 10....
$ x68  <dbl> 0.7970593, 0.8301130, 0.8724225, 0.9018956, 0.7970593, 0.830113...
$ x69  <dbl> 3.754941060, 0.001950779, 0.122911429, -0.192896657, 3.75494106...
$ x70  <dbl> 2.219300, 2.191208, 2.135023, 1.460805, 2.219300, 2.191208, 2.1...
$ x71  <dbl> 3.44708956, 1.37819660, 1.16069285, 2.65944451, 3.44708956, 1.3...
$ x72  <dbl> 2.7225782, 0.0000000, 0.0000000, 2.7225782, 2.7225782, 0.000000...
$ x73  <dbl> 9.691718, 9.691718, 9.422504, 8.884075, 9.691718, 9.691718, 9.4...
$ x74  <dbl> 5.114087, 6.410192, 7.235035, 3.280577, 5.114087, 6.410192, 7.2...
$ x75  <dbl> 0.4186740, 0.4842514, 0.5578712, 0.6155607, 0.4186740, 0.484251...
$ x76  <dbl> 0.4056294, 0.4692197, 0.5481482, 0.5963849, 0.4056294, 0.469219...
$ x77  <dbl> 7.842554, 8.332714, 8.332714, 8.414571, 7.842554, 8.332714, 8.3...
$ x78  <dbl> 5.656412, 5.656412, 5.302886, 5.656412, 5.656412, 5.656412, 5.3...
$ x79  <dbl> 0.02441209, 0.04882417, -0.02441209, 0.68353840, 0.02441209, 0....
$ x80  <dbl> 2.929978, 2.929978, 2.929978, 2.929978, 3.862244, 3.862244, 3.8...
$ x81  <dbl> 2.264569e-02, 3.988401e-03, -1.709845e-01, -1.220182e-01, 4.691...
$ x82  <dbl> 3.1692682, 3.2749105, 3.3805528, 3.4861951, 0.7394959, 0.845138...
$ x83  <dbl> 3.572844, 3.645760, 3.718675, 3.791590, 1.968710, 2.041625, 2.1...
```