OPIM 5604 B15 – Predictive Modeling Assignment Meghana Kasula (Net ID=mek15120)

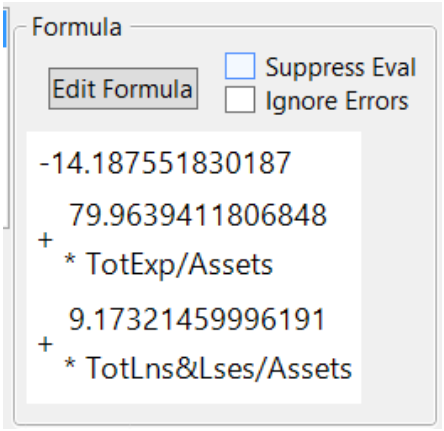
"The work contained and presented here is my work and my work alone."

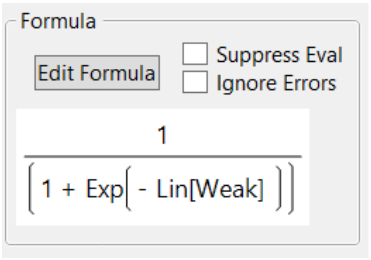
10.1 Financial Condition of Banks.

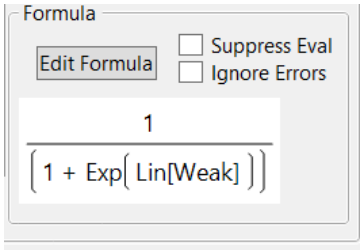
a. Write the estimated equation that associates the financial condition of a bank with its two predictors in two forms:

i. The logit as a function of the predictors.

ii. The probability as a function of the predictors.

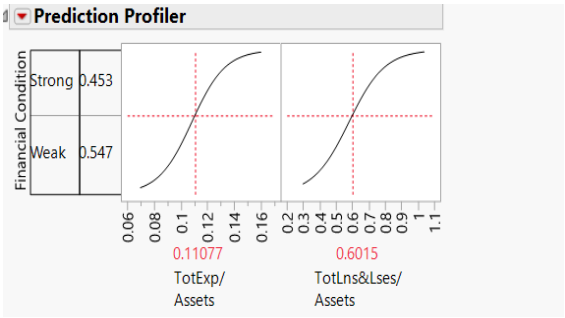






The logit as a function is the formula on the extreme right. Above are the two Probability formulas.

b. Consider a new bank whose total loans and leases/assets ratio = 0.6 and total expenses/assets ratio = 0.11. From your logistic regression model, estimate the following quantities for this bank (save the probability formula to the data table, and enter these values in a new row): the logit, the probability of being financially weak, and the classification of the bank. Confirm the probability using the *Profiler*.



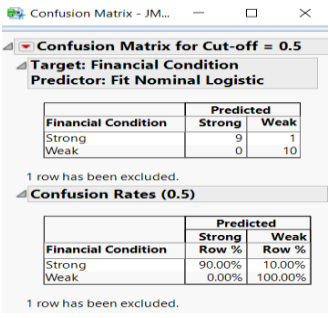
19	19	Strong	8.3	0.09	0.51	-2.312457678	0.0900964634	0.9099035366	Strong
20	20	Strong	20.6	0.13	0.79	3.4546000573	0.9693680285	0.0306319715	Weak
21	•		•	0.11	0.6	0.1124104597	0.5280730599	0.4719269401	Weak

As we enter the values, we get the above values. Also, from the profiler we can see that we got nearest values if we set the values correctly in the profiler.

c. The cutoff probability value of 0.5 is used to classify banks as being financially weak or strong. What is the misclassification rate for weak banks that are incorrectly classified as strong? Use the *Alternate Cutoff Confusion Matrix* add-in to find the cutoff value that minimizes the is classification rate for banks that are financially weak. What is this cutoff value and the misclassification rate?

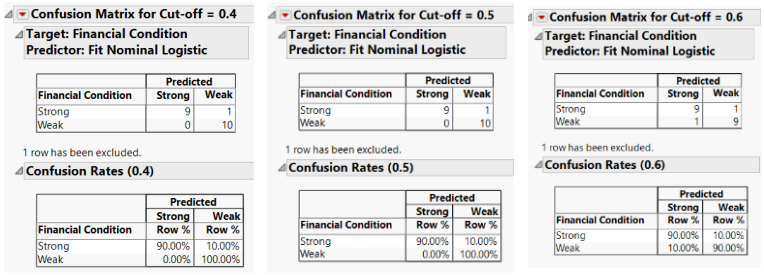
From the table, misclassification 0.0500,

From Alternate Cutoff confusion matrix -its 0.00%



Misclassification Rate	0.0500	∑(p[j]≠pMax)/n
N	20	n

d. When a bank that is in poor financial condition is misclassified as financially strong, the misclassification cost is much higher than when a financially strong bank is misclassified as weak. To minimize the expected cost of misclassification, should the cutoff value for classification (which is currently at 0.5) be increased or decreased?



Confusion Matrix for Cut-off = 0.4
Target: Financial Condition
Predictor: Fit Nominal Logistic

Financial Condition	Predicted Strong	Predicted Weak
Strong	9	1
Weak	0	10

1 row has been excluded.

Confusion Rates (0.4)

Financial Condition	Predicted Strong Row %	Predicted Weak Row %
Strong	90.00%	10.00%
Weak	0.00%	100.00%

Confusion Matrix for Cut-off = 0.5
Target: Financial Condition
Predictor: Fit Nominal Logistic

Financial Condition	Predicted Strong	Predicted Weak
Strong	9	1
Weak	0	10

1 row has been excluded.

Confusion Rates (0.5)

Financial Condition	Predicted Strong Row %	Predicted Weak Row %
Strong	90.00%	10.00%
Weak	0.00%	100.00%

Confusion Matrix for Cut-off = 0.6
Target: Financial Condition
Predictor: Fit Nominal Logistic

Financial Condition	Predicted Strong	Predicted Weak
Strong	9	1
Weak	1	9

1 row has been excluded.

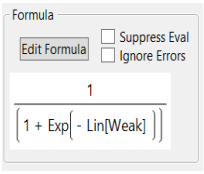
Confusion Rates (0.6)

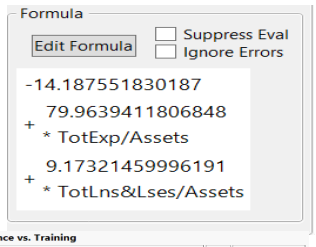
Financial Condition	Predicted Strong Row %	Predicted Weak Row %
Strong	90.00%	10.00%
Weak	10.00%	90.00%

Increase the cutoff and decrease to check, we notice that when we decrease its least and constant but when we increase it's increasing, the misclassification for weak increases. Hence we should decrease the cut-off.

e. Interpret the estimated coefficient for the total loans and leases to total assets ratio (TotLns&Lses/Assets) in terms of the odds of being financially weak.

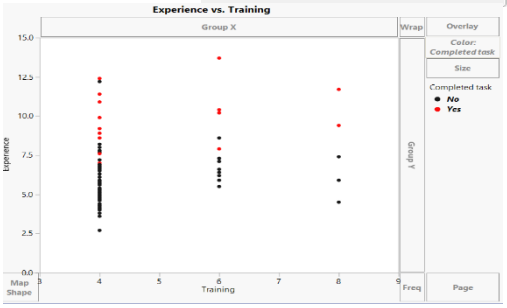
We can see, the coefficient is 9.1732. Hence it will increase with this co-efficient with 100%. That is 9.17 * 100% holding others constant.





10.2 Identifying Good System Administrators.

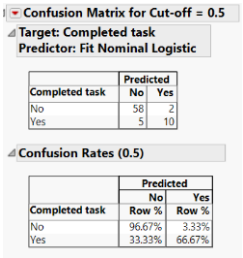
a. Create a scatterplot of Experience versus Training using color or symbol to differentiate programmers who complete the task from those who did not complete it. Which predictor(s) appear(s) potentially useful for classifying task completion?



As we can see, the as the experience increases in all number of training credits, the probability of jobs being completed is more. Very less or almost nil below 7.5 years of experience and above &.5, most of them are in "Yes" category.
I think, it is the experience which is the useful predictors since we can be sure that task will be most probably done after certain number of years.

b. Run a logistic regression model with both predictors using the entire dataset as training data. Model the probability of Completed Task = Yes. Among those who complete the task, what is the percentage of programmers who are incorrectly classified as failing to complete the task?

The Percentage is 33.33%.

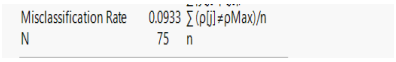


Confusion Matrix for Cut-off = 0.5
Target: Completed task
Predictor: Fit Nominal Logistic

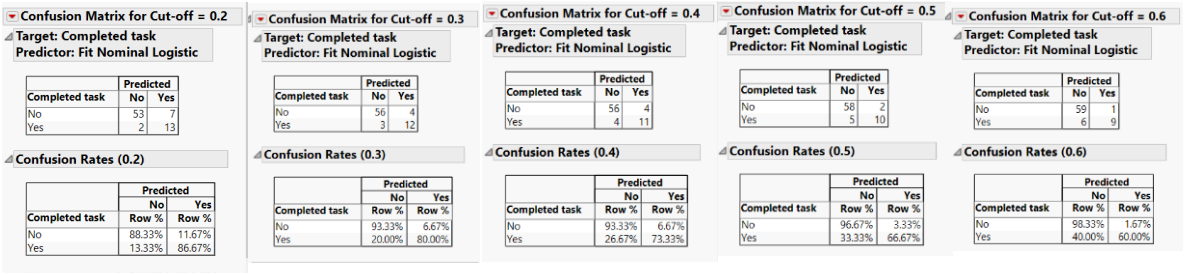
Completed task	Predicted No	Predicted Yes
No	58	2
Yes	5	10

Confusion Rates (0.5)

Completed task	Predicted No Row %	Predicted Yes Row %
No	96.67%	3.33%
Yes	33.33%	66.67%

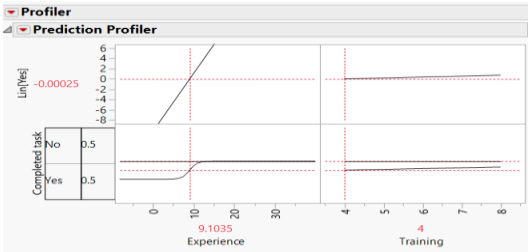


c. To decrease the percentage in part (b), should the cutoff probability be increased or decreased?



As we decrease the cut-off from 0.5 to below the percentage of misclassification of Yes to No decreases and vice versa. Hence we should decrease the cut-off.

d. How much experience must be accumulated by a programmer with four years of training before his or her estimated probability of completing the task exceeds 50%?



With Lin[Yes] as almost 0; and probability of Yes and No as 0.5 and training as 4 credits; we get the experience as 9.1035 years

10.3 Sales of Riding Mowers.

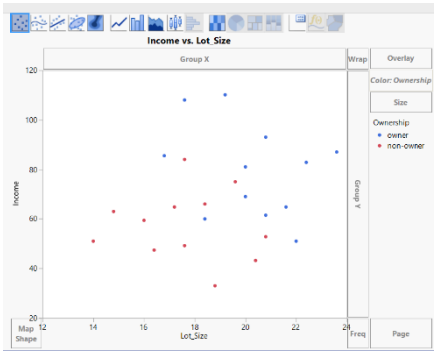
a. What percentage of households in the study were owners of a riding mower?

Total records=24

Owners=12; There fore the probability

12/24*100=50%

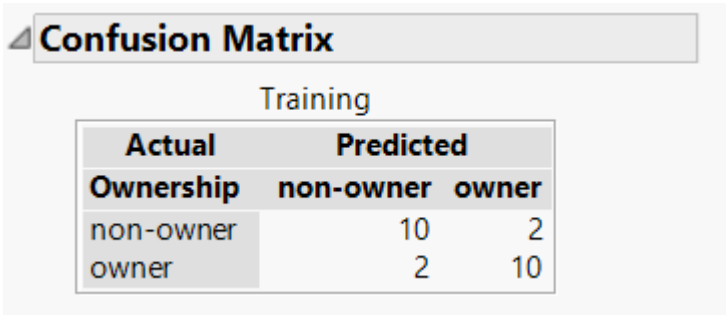
b. Create a scatterplot of Income versus Lot Size using color or symbol to differentiate owners from nonowners. From the scatterplot, which class seems to have the higher average income, owners or nonowners?



We can as the income and Lot size increases the owners increases. Hence it is directly dependent on both.

Owners have average higher income.

c. Among nonowners, what is the percentage of households classified correctly?



10/12*100=83.33%

d. To increase the percentage of correctly classified nonowners, should the cutoff probability be increased or decreased?

Confusion Matrix for Cut-off = 0.3

Target: Ownership
Predictor: Fit Nominal Logistic

Ownership	Predicted non-owner	Predicted owner
non-owner	11	1
owner	2	10

Confusion Rates (0.3)

Ownership	Predicted non-owner Row %	Predicted owner Row %
non-owner	91.67%	8.33%
owner	16.67%	83.33%

Confusion Matrix for Cut-off = 0.4

Target: Ownership
Predictor: Fit Nominal Logistic

Ownership	Predicted non-owner	Predicted owner
non-owner	11	1
owner	2	10

Confusion Rates (0.4)

Ownership	Predicted non-owner Row %	Predicted owner Row %
non-owner	91.67%	8.33%
owner	16.67%	83.33%

Confusion Matrix for Cut-off = 0.5

Target: Ownership
Predictor: Fit Nominal Logistic

Ownership	Predicted non-owner	Predicted owner
non-owner	10	2
owner	2	10

Confusion Rates (0.5)

Ownership	Predicted non-owner Row %	Predicted owner Row %
non-owner	83.33%	16.67%
owner	16.67%	83.33%

Confusion Matrix for Cut-off = 0.6

Target: Ownership
Predictor: Fit Nominal Logistic

Ownership	Predicted non-owner	Predicted owner
non-owner	9	3
owner	1	11

Confusion Rates (0.6)

Ownership	Predicted non-owner Row %	Predicted owner Row %
non-owner	75.00%	25.00%
owner	8.33%	91.67%

The cut-off of non owners probability must be decreased to increase the percentage of correctly classified nonowners.

e. What are the odds that a household with a $60K income and a lot size of 20,000 ft2 is an owner?

25	60	20		-0.011133832	0.4972165708	0.5027834292	non-owner	non-owner

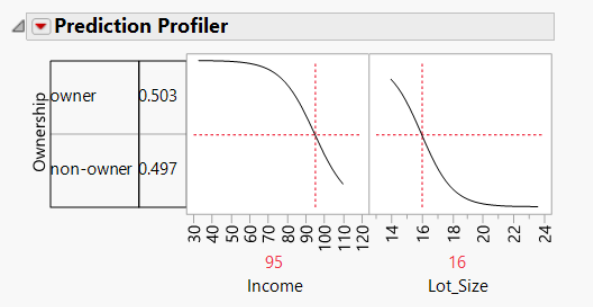
Odds= -0.011133832

f. What is the classification of a household with a $60K income and a lot size of 20,000 ft2?

25	60	20		-0.011133832	0.4972165708	0.5027834292	non-owner

Non-Owner

g. What is the minimum income that a household with 16,000 ft2 lot size should have before it is classified as an owner?



The minimum income at Logit -0 and probability of owners and non-owners as 0.5 approx is $95k.