Predicting Whether a Startup Will Fail or be Successful

Anyone who creates a startup is hoping that it will be successful. Investors also hope that the startups they invest in will be successful, and it can be interesting to see which startups were thought to be successful but ultimately failed. For this analysis, I looked at several news articles written between 2010-2012 about which startups people should keep an eye on. Forbes and Business Insider published several articles between those years discussing startups that they thought were going to become highly successful and suggested that people watch these startups carefully. I made a list of the startups that were discussed in these articles, and then used Crunchbase to figure out whether these startups were still active and successful, or if they had "failed."

Because of how I selected my data, it can be considered a prospective study. I compiled a list of 31 startups and then determined whether they had failed, rather than selecting startups which had failed and startups which had become successful. I used three predictor variables: **Total Funding Rounds**, **Funding After One Round**, and the number of **Months Before Raising Funds**. Some startups go through just one round of funding, while others have gone through upwards of 20. I had thought that perhaps the more funding rounds, the more likely the startup was to become successful. However, this relationship was not inherently obvious since a company like Kickstarter seemed to have only gone through one round of funding (and it is a rather successful company).

Funding After One Round refers to the amount of money a startup had received after one round of funding; to me, it seemed like it would make a lot of sense that a company that receives a lot of money in seed funding had more potential of being successful than a company that only had an initial investment of less than $50,000. However, like Total Funding Rounds, I could not be so certain because, for example, TaskRabbit had initially only received $25,000 in funding which is relatively low.

Finally, Months Before Raising Funds refers to the number of months after being founded that it took for a startup to receive seed funding. I thought that if it takes a substantial number of months for a startup to receive funding, it may indicate that that startup did not seem like it had much potential to become something valuable. On the other hand, if a startup received funding from an investor or venture capital group after one month of being founded, it may indicate that the startup idea had a lot of potential to succeed.

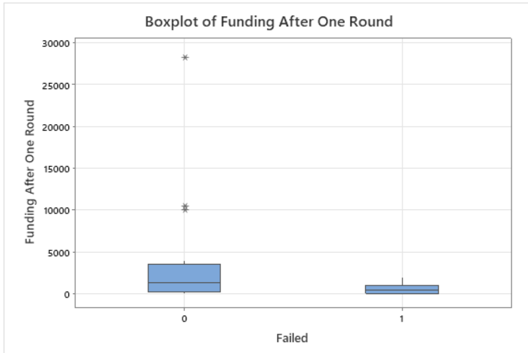
The data for all three variables was taken from either Crunchbase or Angel.co, since sometimes Angel.co had more information regarding a startup's financial history than did Crunchbase. The articles that I used to compile the list of observations included: "The 10 Greatest Industry-Disrupting Startups of 2012" (*Forbes*, 2012), "The 25 Hot New York City Startups You Need to Watch" (*Business Insider*, 2011), "The 20 Hot Silicon Valley Startups You Need to Watch" (*Business Insider,* 2011), and "Top 10 Startups of 2010" (*ReadWrite*, 2010).

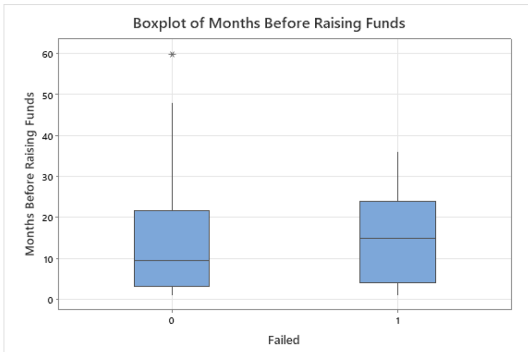
Startup	Failed?	Funding After One Round (Thousands)	Total Funding Rounds	Months Before Raising Funds

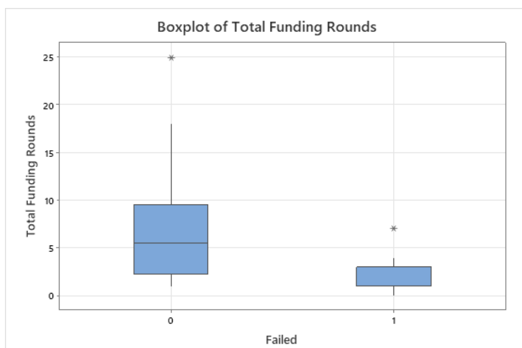
Kickstarter	0	10,000	1	23
Getaround	0	3,400	18	24
Storenvy	0	1,500	2	12
Noodle	0	28,300	3	60
Farmigo	1	2,000	3	1
ClickFox	0	4,000	7	48
BrainStorm	0	206.91	10	24
Recyclebank	0	100	6	12
Flipboard	0	10,500	5	6
Indinero	0	17	7	7
ZilloPay	1	25	1	4
PostRocket	1	610	1	19
HotelTonight	0	3,585	11	5
Pinterest	0	492.9	25	1
Dekko	0	1,900	2	12
iDoneThis	0	380	2	3
Singly	0	3,000	3	2
SwipeGood	1	500	1	24
Speakergram	1	50	1	12
Votizen	0	1,500	2	18
VigLink	0	800	5	3
TaskRabbit	0	25	7	16
Cruisewise	0	40	3	4
Artsicle	1	390	1	15
TutorSpree	1	1,000	4	15
Jibe	1	875	7	3
SignPost	0	1,100	8	3
SkillShare	0	550	11	6
RiotVine	1	0	0	13
Solasta	1	1,000	1	24
Sponty	1	0	0	36

A value of 1 indicates that the startup failed, while a 0 indicates that the startup was successful/didn’t fail.

I began by looking at the boxplots for the different predictor variables. MiniTab prodcuced the following:







We can see that there is separation between Total Funding Founds for the startups that failed and the startups that continue to run successfully, but the variables Months Before Initial Funding Rounds and Funding After One Round seem to have less predictive power, with very weak separation. The boxplots also indicate that there are outliers, however, we will look further into these later.

Here is the logistic regression output for the data:

Binary Logistic Regression: Failed versus Funding After One Round, Months Before Raising Funds, Total Funding Rounds

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	31

Response Information

Variable	Value	Count
Failed	1	11 (Event)
	0	20
	Total	31

Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))
Y' = 1.65 - 0.000686 Funding After One Round + 0.0315 Months Before Raising Funds - 0.511 Total Funding Rounds

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	1.65	1.27	1.31	0.192	
Funding After One Round	-0.000686	0.000563	-1.22	0.223	1.01
Months Before Raising Funds	0.0315	0.0622	0.51	0.612	1.03
Total Funding Rounds	-0.511	0.226	-2.26	0.024	1.02

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
Funding After One Round	0.9993	(0.9982, 1.0004)
Months Before Raising Funds	1.0320	(0.9137, 1.1657)
Total Funding Rounds	0.6001	(0.3854, 0.9344)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
42.59%	35.15%	31.15	32.69	36.89	0.9045

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	27	23.15	0.677
Pearson	27	25.35	0.555
Hosmer-Lemeshow	8	10.81	0.213

Analysis of Variance

				Likelihood Ratio	
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	3	17.1751	5.7250	17.18	0.001
Funding After One Round	1	5.2236	5.2236	5.22	0.022
Months Before Raising Funds	1	0.2499	0.2499	0.25	0.617
Total Funding Rounds	1	9.6837	9.6837	9.68	0.002
Error	27	23.1491	0.8574		
Total	30	40.3242			

Measures of Association

Pairs	Number	Percent	Summary Measures	Value
Concordant	199	90.5	Somers’ D	0.81
Discordant	21	9.5	Goodman-Kruskal Gamma	0.81
Ties	0	0.0	Kendall’s Tau-a	0.38
Total	220	100.0		

The test statistic for the Analysis of Variance is 17.18 with a *p*-value of 0.001, which is statistically significant at the $\alpha = 0.05$ level. We can reject the null hypothesis that indicates there is no relationship. However, we must note that while Funding After One Round and Total Funding Rounds are significant, the Months Before Raising Funds variable is not so much significant. The VIF values are fairly low, suggesting that there is no sign of collinearity in the model, however, that does not exactly mean that all of the variables need to stay in the model. Additionally, the Somers’ D of 0.81 is indication of excellent separation. The model seems to fit the data well, however, it has the potential of being simplified given that one of the predictor variables does not even seem to be significant.

I ran best subsets, and the model(s) which stuck out the most was a simplified model containing the variables Funding After One Round and Total Funding Rounds. This model minimizes Mallows Cp and AIC_c, and of course, is much simpler than a three-variable model. I ran the logistic regression containing just these two variables and received the following output:

Binary Logistic Regression: Failed versus Funding After One Round, Total Funding Rounds

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	31

Response Information

Variable	Value	Count
Failed	1	11 (Event)
	0	20
	Total	31

Regression Equation

```
P(1) = exp(Y')/(1 + exp(Y'))
Y'   = 2.11 - 0.537 Total Funding Rounds - 0.000691 Funding After One Round
```

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	2.11	1.00	2.11	0.035	
Total Funding Rounds	-0.537	0.230	-2.34	0.019	1.00
Funding After One Round	-0.000691	0.000589	-1.17	0.241	1.00

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
Total Funding Rounds	0.5843	(0.3723, 0.9171)
Funding After One Round	0.9993	(0.9982, 1.0005)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
41.97%	37.01%	29.40	30.29	33.70	0.9091

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	28	23.40	0.713
Pearson	28	23.83	0.691
Hosmer-Lemeshow	8	10.45	0.235

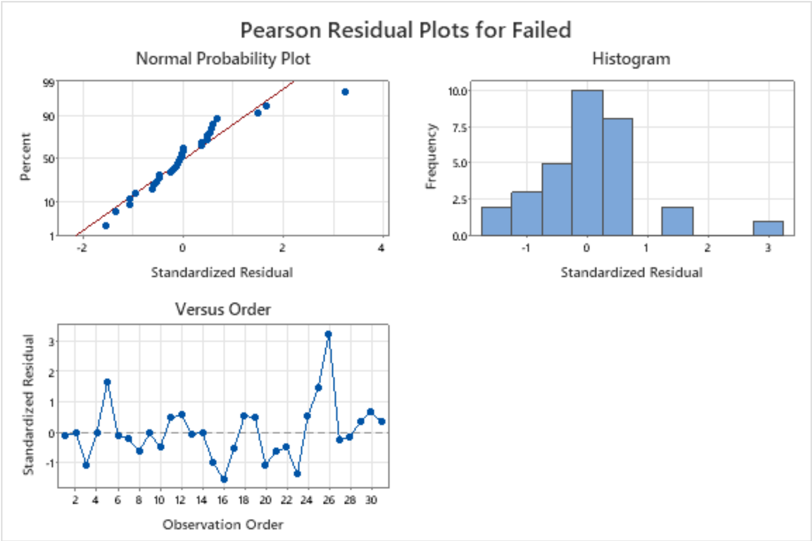
Analysis of Variance

				Likelihood Ratio	
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	16.925	8.4626	16.93	0.000
Total Funding Rounds	1	10.957	10.9571	10.96	0.001
Funding After One Round	1	5.241	5.2409	5.24	0.022
Error	28	23.399	0.8357		
Total	30	40.324			

Measures of Association

Pairs	Number	Percent	Summary Measures	Value
Concordant	200	90.9	Somers' D	0.82
Discordant	20	9.1	Goodman-Kruskal Gamma	0.82
Ties	0	0.0	Kendall's Tau-a	0.39
Total	220	100.0		

In this model, both the AIC_c and AIC are lower, though not by a considerable amount. Looking at the Anlysis of Variance output, we certainly see that the *p*-value of the model is now 0.000, which is slightly lower than what it was before, and both predictor variables remain significant. Moreover, the Somers' D is now 0.82, which is just short of the previous Somers' D by 0.01. These two models are effectively the same, however this one is simplified, so it is clearly the better model to go forward with.



In this model, there was one variable with an unusually high standardized residual of 3.25, and its Cook's Distance was also significantly high compared to the rest of the yielded Cook's Distances.

SPEARRES	HI	COOK
-0.07491	0.139555	0.000303
-0.00706	0.000761	0.000000
-1.06092	0.109276	0.046028
-0.00007	0.000001	0.000000
1.68292	0.148580	0.164749
-0.11346	0.057012	0.000259
-0.19165	0.093457	0.001262
-0.60766	0.167204	0.024712
-0.02010	0.013209	0.000002
-0.48161	0.179605	0.016927
0.48440	0.103550	0.009035
0.58776	0.087803	0.011084
-0.04370	0.012615	0.000008
-0.00294	0.000225	0.000000
-0.95316	0.162960	0.058958
-1.54088	0.084513	0.073062
-0.53159	0.264636	0.033898
0.56622	0.089009	0.010442

0.48833	0.102530	0.009081
-1.06092	0.109276	0.046028
-0.59767	0.091718	0.012024
-0.47993	0.178388	0.016670
-1.34860	0.116699	0.080095
0.54576	0.091205	0.009964
1.49438	0.074002	0.059489
3.24922	0.099493	0.388813
-0.23914	0.079952	0.001656
-0.12748	0.055474	0.000318
0.36570	0.096685	0.004772
0.67480	0.093923	0.015734
0.36570	0.096685	0.004772

This observation is startup company "Jibe," which managed to earn $875,000 in seed funding in just three months after being founded. Jibe also went through seven funding rounds, which is a considerable amount given that it ultimately failed. Even if it had been successful, seven funding rounds is still high compared to the successful startups in the set of data.

Jibe started off as a referral-based hiring platform, and it received a lot of positive attention in its earlier years. However, upon looking further into it, the company seemed to be dealing with management issues and encountered massive employee layoffs and struggled to deliver its service due to the fact that they did not even have any engineers. The startup was founded in 2009 and was acquired in 2019. It ceased operations. So, despite ultimately stopping operations, it was not necessarily a failure and had about 10 years of life.

I decided to remove this observation from the data and rerun the regression.

Binary Logistic Regression: Failed versus Funding After One Round, Total Funding Rounds

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	30

Response Information

Variable	Value	Count
Failed	1	10 (Event)
	0	20
	Total	30

Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))
Y' = 2.84 - 0.962 Total Funding Rounds - 0.000678 Funding After One Round

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	2.84	1.26	2.26	0.024	
Total Funding Rounds	-0.962	0.457	-2.11	0.035	1.01
Funding After One Round	-0.000678	0.000661	-1.03	0.305	1.01

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI

Total Funding Rounds	0.3820 (0.1559, 0.9358)
Funding After One Round	0.9993 (0.9980, 1.0006)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
55.10%	49.87%	23.15	24.07	27.35	0.9500

Goodness-of-Fit Tests

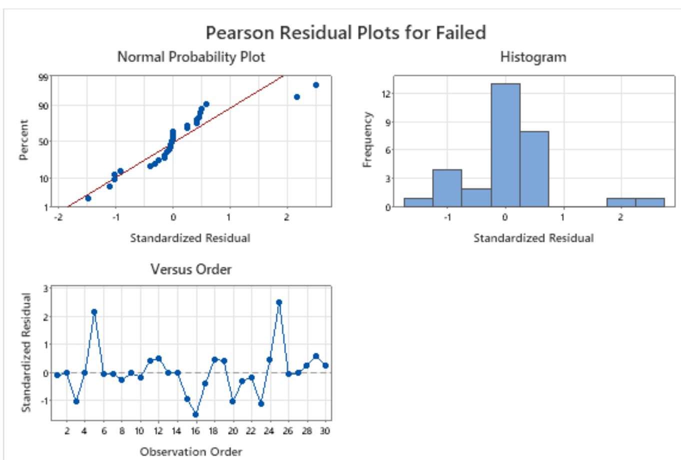
Test	DF	Chi-Square	P-Value
Deviance	27	17.15	0.927
Pearson	27	16.50	0.943
Hosmer-Lemeshow	8	6.68	0.572

Analysis of Variance

				Likelihood Ratio	
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	21.045	10.5224	21.04	0.000
Total Funding Rounds	1	15.289	15.2890	15.29	0.000
Funding After One Round	1	4.311	4.3111	4.31	0.038
Error	27	17.146	0.6350		
Total	29	38.191			

Measures of Association

Pairs	Number	Percent	Summary Measures	Value
Concordant	190	95.0	Somers' D	0.90
Discordant	10	5.0	Goodman-Kruskal Gamma	0.90
Ties	0	0.0	Kendall's Tau-a	0.41
Total	200	100.0		



In doing so, two more unusual observations appear. One is the startup TutorSpree (with standardized residual 2.51), which earned $1 million in initial seed funding only to end up shutting down just three years after being founded; the other is Farmigo (standardized residual 2.18) , which earned $2 million in initial seed funding in just one month after being founded. Farmigo went bankrupt seven years after being founded.

SPEARRES	HI	COOK
-0.10129	0.272194	0.001279
-0.00023	0.000003	0.000000
-1.01785	0.125745	0.049671
-0.00007	0.000001	0.000000

2.18526	0.150610	0.282247
-0.03700	0.011441	0.000005
-0.03162	0.013728	0.000005
-0.24680	0.182348	0.004528
-0.01067	0.004554	0.000000
-0.15090	0.116187	0.000998
0.41753	0.109720	0.007162
0.50638	0.100236	0.009522
-0.00618	0.000703	0.000000
-0.00002	0.000000	0.000000
-0.92085	0.185484	0.064367
-1.47833	0.114654	0.094341
-0.40113	0.222002	0.015305
0.48781	0.100071	0.008820
0.42089	0.108928	0.007219
-1.01785	0.125745	0.049671
-0.30864	0.148343	0.005531
-0.15042	0.115411	0.000984
-1.10344	0.235674	0.125144
0.47021	0.101008	0.008280
2.51159	0.145254	0.357327
-0.06160	0.027641	0.000036
-0.01732	0.005113	0.000001
0.25202	0.082281	0.001898
0.58195	0.112637	0.014329
0.25202	0.082281	0.001898

Both of these observations had unusually high standardized residuals and omitting both of them resulted in complete separation.

Binary Logistic Regression: Failed versus Funding After One Round, Total Funding Rounds

* WARNING * When the data are in the Response/Frequency format, the Residuals versus fits plot is unavailable.

* WARNING * The model could not be fit properly. Maximum likelihood estimates of parameters do not exist due to complete separation of data points. The results are not reliable. Please refer to help for more information about complete separation.

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	28

Response Information

Variable	Value	Count	
Failed	1	8	(Event)
	0	20	
	Total	28	

Regression Equation

```
P(1) = exp(Y')/(1 + exp(Y'))
Y'   = 17.5 - 10.72 Total Funding Rounds - 0.00146 Funding After One Round
```

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	17.5	11.4	1.54	0.125	
Total Funding Rounds	-10.72	6.95	-1.54	0.123	1.00
Funding After One Round	-0.00146	0.00367	-0.40	0.691	1.00

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
Total Funding Rounds	0.0000	(0.0000, 18.2282)
Funding After One Round	0.9985	(0.9914, 1.0058)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
99.81%	93.84%	6.06	7.06	10.06	1.0000

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	25	0.06	1.000
Pearson	25	0.03	1.000
Hosmer-Lemeshow	8	0.01	1.000

Analysis of Variance

Source	DF	Adj Dev	Adj Mean	Likelihood Ratio Chi-Square	Likelihood Ratio P-Value
Regression	2	33.4401	16.7201	33.44	0.000
Total Funding Rounds	1	25.8741	25.8741	25.87	0.000
Funding After One Round	1	5.6787	5.6787	5.68	0.017
Error	25	0.0630	0.0025		
Total	27	33.5031			

Measures of Association

Pairs	Number	Percent	Summary Measures	Value
Concordant	160	100.0	Somers' D	1.00
Discordant	0	0.0	Goodman-Kruskal Gamma	1.00
Ties	0	0.0	Kendall's Tau-a	0.42
Total	160	100.0		

So, we can try to run best subsets again.

Vars	R-Sq	R-Sq (adj)	PRESS	R-Sq (pred)	Mallows Cp	S	AICc	BIC	Cond No	Funding After One	Months Before Rai	Total Funding Rou

										Round	sing Funds	nds
1	24.1	21.2	5.5	4.4	6.1	0.40847	34.247	37.244	1.000			X
1	6.9	3.3	7.2	0.0	12.9	0.45242	39.969	42.966	1.000	X		
2	32.4	27.0	6.0	0.0	4.8	0.39297	33.722	37.311	1.114	X		X
2	24.3	18.2	6.0	0.0	8.0	0.41602	36.914	40.503	1.415		X	X
3	39.5	31.9	5.6	2.3	4.0	0.37962	33.631	37.565	4.074	X	X	X

The original model with all three variables now minimizes the Mallows Cp and AIC_c, so we will run the regression without the three outliers but with all three predicting variables.

However, this model also has complete separation.

Binary Logistic Regression: Failed versus Funding After One Round, Total Funding Rounds, Months Before Raising Funds

```
* WARNING * When the data are in the Response/Frequency format, the
Residuals versus fits
plot is unavailable.

* WARNING * The model could not be fit properly. Maximum likelihood
estimates of parameters
do not exist due to complete separation of data points. The results are not
reliable. Please
refer to help for more information about complete separation.
```

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	28

Response Information

Variable	Value	Count
Failed	1	8 (Event)
	0	20
	Total	28

Regression Equation

```
P(1) = exp(Y')/(1 + exp(Y'))
Y'   = 8.8 - 6.01 Total Funding Rounds - 0.0075 Funding After One Round
       + 0.51 Months Before Raising Funds
```

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	8.8	16.8	0.52	0.603	
Total Funding Rounds	-6.01	9.77	-0.62	0.539	2.07
Funding After One Round	-0.0075	0.0176	-0.43	0.670	6.54
Months Before Raising Funds	0.51	1.30	0.39	0.697	5.28

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
Total Funding Rounds	0.0025	(0.0000, 508508.6219)
Funding After One Round	0.9925	(0.9589, 1.0273)

Months Before Raising Funds 1.6583 (0.1295, 21.2380)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
99.84%	90.88%	8.05	9.79	13.38	1.0000

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	24	0.05	1.000
Pearson	24	0.03	1.000
Hosmer-Lemeshow	8	0.01	1.000

Analysis of Variance

				Likelihood Ratio	
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	3	33.4483	11.1494	33.45	0.000
Total Funding Rounds	1	20.1916	20.1916	20.19	0.000
Funding After One Round	1	4.4779	4.4779	4.48	0.034
Months Before Raising Funds	1	*	*	*	*
Error	24	0.0548	0.0023		
Total	27	33.5031			

Measures of Association

Pairs	Number	Percent	Summary Measures	Value
Concordant	160	100.0	Somers' D	1.00
Discordant	0	0.0	Goodman-Kruskal Gamma	1.00
Ties	0	0.0	Kendall's Tau-a	0.42
Total	160	100.0		

I also ran the regression using only one predictor variable, Total Funding Rounds, since it had a low Mallows Cp (although not the lowest); this model also has complete separation.

Binary Logistic Regression: Failed versus Total Funding Rounds

* WARNING * When the data are in the Response/Frequency format, the
Residuals versus fits
plot is unavailable.

* WARNING * The model could not be fit properly. Maximum likelihood
estimates of parameters
may not exist due to quasi-complete separation of data points. The results
might not be
reliable. Please refer to help for more information about quasi-complete
separation.

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	28

Response Information

Variable	Value	Count
Failed	1	8 (Event)
	0	20
	Total	28

Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))
Y' = 17 - 15 Total Funding Rounds

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF

Constant	17	272	0.06	0.950	
Total Funding Rounds	-15	272	-0.06	0.955	1.00

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
Total Funding Rounds	0.0000	(0.0000, 6.69970E+224)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
82.86%	79.88%	9.74	10.22	12.41	0.9812

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	26	5.74	1.000
Pearson	26	7.00	1.000
Hosmer-Lemeshow	6	0.00	1.000

Analysis of Variance

Source	DF	Adj Dev	Adj Mean	Likelihood Ratio Chi-Square	P-Value
Regression	1	27.761	27.7615	27.76	0.000
Total Funding Rounds	1	27.761	27.7615	27.76	0.000
Error	26	5.742	0.2208		
Total	27	33.503			

Measures of Association

Pairs	Number	Percent	Summary Measures	Value
Concordant	154	96.3	Somers' D	0.96
Discordant	0	0.0	Goodman-Kruskal Gamma	1.00
Ties	6	3.8	Kendall's Tau-a	0.41
Total	160	100.0		

So, I decided to just stick with the model with two predictor variables.

Binary Logistic Regression: Failed versus Funding After One Round, Total Funding Rounds

* WARNING * When the data are in the Response/Frequency format, the Residuals versus fits plot is unavailable.

* WARNING * The model could not be fit properly. Maximum likelihood estimates of parameters do not exist due to complete separation of data points. The results are not reliable. Please refer to help for more information about complete separation.

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	28

Response Information

Variable	Value	Count
Failed	1	8 (Event)
	0	20
	Total	28

Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))
Y' = 17.5 - 10.72 Total Funding Rounds - 0.00146 Funding After One Round

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	17.5	11.4	1.54	0.125	
Total Funding Rounds	-10.72	6.95	-1.54	0.123	1.00
Funding After One Round	-0.00146	0.00367	-0.40	0.691	1.00

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
Total Funding Rounds	0.0000	(0.0000, 18.2282)
Funding After One Round	0.9985	(0.9914, 1.0058)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
99.81%	93.84%	6.06	7.06	10.06	1.0000

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	25	0.06	1.000
Pearson	25	0.03	1.000
Hosmer-Lemeshow	8	0.01	1.000

Analysis of Variance

				Likelihood Ratio	
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	33.4401	16.7201	33.44	0.000
Total Funding Rounds	1	25.8741	25.8741	25.87	0.000
Funding After One Round	1	5.6787	5.6787	5.68	0.017
Error	25	0.0630	0.0025		
Total	27	33.5031			

Measures of Association

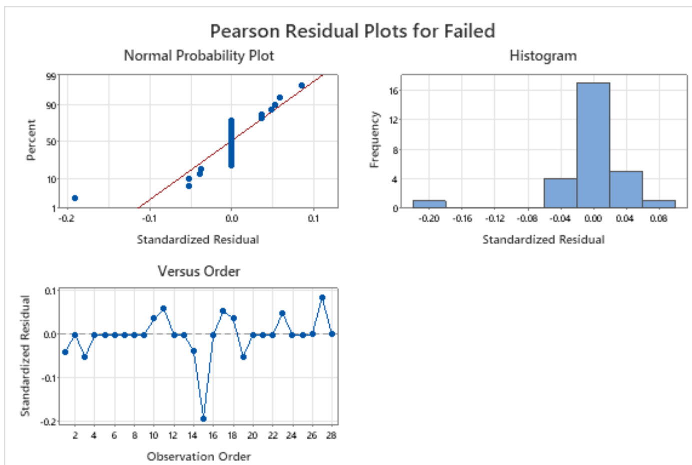
Pairs	Number	Percent	Summary Measures	Value
Concordant	160	100.0	Somers' D	1.00
Discordant	0	0.0	Goodman-Kruskal Gamma	1.00
Ties	0	0.0	Kendall's Tau-a	0.42
Total	160	100.0		

The test statistic for the Analysis of Variance is 33.44 with a *p*-value of 0.000, which is statistically significant at the $\alpha = 0.05$ level. We can reject the null hypothesis that indicates there is no relationship. The VIF values are low, so there is no sign of collinearity in the model. Both predictor variables are statistically significant, with Total Funding Rounds having a *p*-value of 0.000, and Funding After One Round having a *p*-value of 0.017.

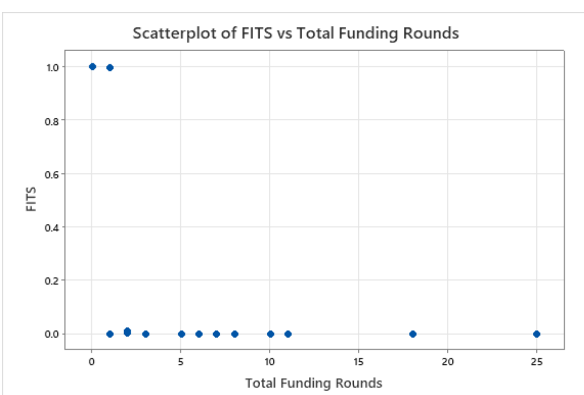
It is no surprise that the Somers' D is 1.00 since we are already aware that there is perfect separation in this model. What this means is that the predictor variables Funding After One Round and Total Funding Rounds which are being used to fit the model can very accurately differentiate between "Failed" startups and those that did not fail. The area under the ROC curve is also 1.00, which makes sense since it is equivalent to Somers' D. The Hosmer-Lemeshow test has a *p*-value of 1.000, indicating no lack of fit. The AIC_c in this model is 7.06, which is smaller than that of the one predictor model and the three-predictor model.

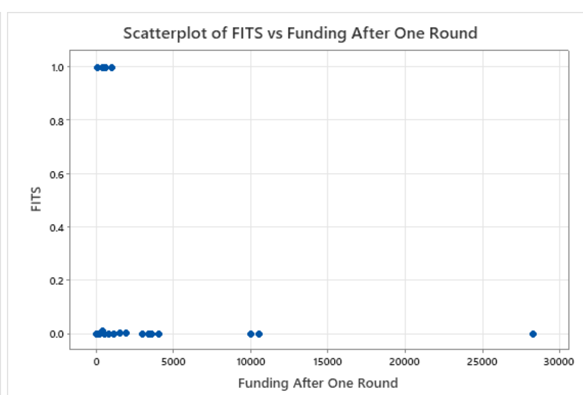
The confidence interval for the odds ratio of Total Funding Rounds (0.0000, 18.2282), which had an extremely small minimum and extremely large maximum, while the confidence interval for the odds ratio of Funding After One Round is (0.9914, 1.0058). The odds ratio for Total Funding Rounds is rather wide and large; however, the odds ratio is 0.0000, which means that a 1-unit change in Total Funding Rounds is associated with a 100% decrease in the odds of a startup failing, holding all else fixe. The odds ratio for Funding After One Round is 0.9985; each additional increase in one thousand dollars in Funding After One Round is associated with a 0.15% decrease in the odds of a startup failing, holding all else fixed. The negative slope coefficients for both predictor variables imply a decreasing probability of being a failed startup with increasing Total Funding Rounds and increasing Total Funding After One Round

This is understandable. If a startup is going through more rounds of funding, it may mean that the startup has been running for a long enough amount of time to have had many opportunities to receive funding, or that the startup is just that good and groundbreaking that many investors want to invest in it, or both. Additionally, if a startup is receiving a lot more funding compared to other startups in the market, it would make sense that it has more potential of lasting long since generally, if a startup is receiving a lot of funding from investors, it probably has some potential.



There does seem to be an observation very far out from the rest of the observations, however, its standardized residual is nothing alarming. There are no alarming patterns in any of these plots. Scatter plots of fitted values (estimated probabilities) and predictors are not that indicative of an s-shape but are indicative of the complete separation we are dealing with.





The following is the classification matrix, based on whether the estimated probability is above or below 0.1 (it seems that most of the successes had probabilities that were well below 0.1, so this seemed like an appropriate benchmark—this is likely because the sample size is small, the data is prospective, and there is an unequal number of successes and failures).

This matrix also accounts for the outliers which had been removed.

Tabulated Statistics: Failed, CLASSFIT

Rows: Failed   Columns: CLASSFIT

	0	1	All
0	20	3	23
	71.43	0.00	71.43
1	0	8	8
	0.00	28.57	28.57
All	20	11	31
	71.43	28.57	100.00

Cell Contents
Count
% of Total

According to this matrix, 90.3% of the startups were correctly classified. This is significantly higher than:

$$C_{pro} = [(23/31) * (20/31) + (8/31) * (11/31)] = 61.2\%$$

$$C_{max} = \max (23/31, 8/31) = 74.19\%$$

Because this was a prospective analysis, there is no need to inflate C_{pro}. It is evident that these two predictor variables do a good job at classifying startups as either failed or successful.

The following are the estimated probabilities provided by MiniTab when the three outliers were removed.

```
FITS
0.00041
0.00000
0.00218
0.00000
0.00000
0.00000
0.00000
0.00000
0.00000
0.99882
0.99724
0.00000
0.00000
0.00122
0.01106
0.00000
0.99765
0.99878
0.00218
```

```
0.00000
0.00000
0.00000
0.99800
0.00000
0.00000
1.00000
0.99514
1.00000
```

The following are the estimated probabilities provided by MiniTab when the outliers are included.

```
FITS
0.004805
0.000050
0.500637
0.000000
0.293134
0.011994
0.032225
0.235185
0.000398
0.159870
0.826207
0.760383
0.001882
0.000009
0.431967
0.684906
0.172051
0.773956
0.823713
0.500637
0.244967
0.159129
0.616342
0.786975
0.325953
0.095175
0.049984
0.015118
0.892212
0.707921
0.892212
```

For the startups that failed, their probability of failing is much higher before the outliers were removed. This makes sense since those outliers were misclassified due to the fact that both their Total Funding Rounds and Funding After One Round were unusually large for a startup that ended up failing. When these outliers are omitted, the probabilities are much more “polarized,” so to say. Startups that failed had a predicted probability of failing near 1, with some even having a probability of 1, while the startups that were successful had a predicted probability of failing near 0.

Based on this analysis, we can say that there is certainly an association between Total Funding Rounds and Funding After One Round. Startups that had a lot of funding rounds and earned a lot of money in seed funding after one right, did on average, end up being successful, while the startups which received a low amount of funding during the first round and did not go through many funding rounds seemed to fail more often.

After removing the outliers, it appears that the number of Total Funding Rounds that marks the separation of failed startups and successful startups is 1. Among the observations, any startup that had more than 1 funding round happened to be successful. However, if I had not removed Farmigo, TutorSpree, and Jibe, then this statement would not hold as these startups had at least three funding rounds, which happens to be a common number of rounds for some successful startups. This potentially explains why these startups were misclassified.