The Flight of the Space Shuttle Challenger

On January 28, 1986, the space shuttle Challenger took off on the 25^{th} flight in NASA's space shuttle program. Less than 2 minutes into the flight, the spacecraft exploded, killing all on board. A Presidential Commission was formed to explore the reasons for this disaster.

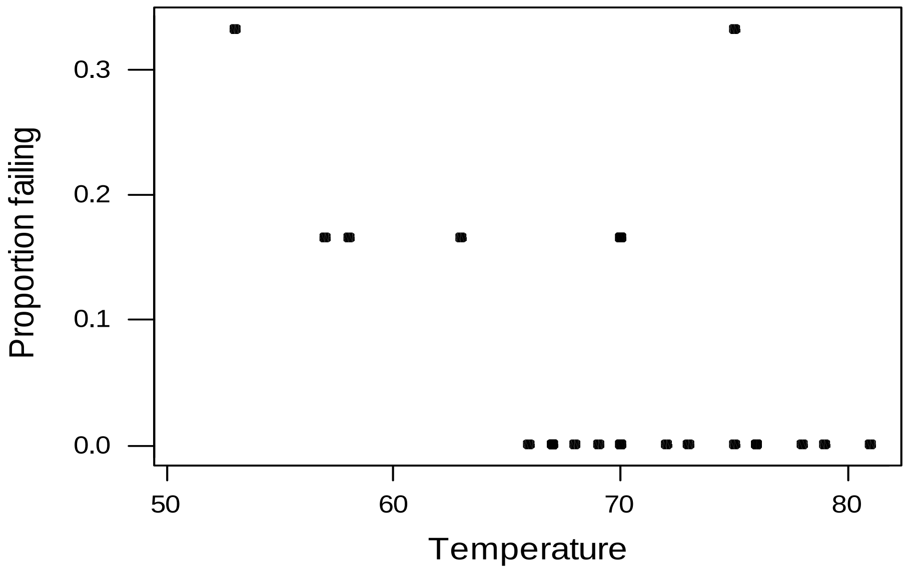
First, a little background information: the space shuttle uses two booster rockets to help lift it into orbit. Each booster rocket consists of several pieces whose joints are sealed with rubber O–rings, which are designed to prevent the release of hot gases produced during combustion. Each booster contains 3 primary O–rings (for a total of 6 for the orbiter). In the 23 previous flights for which there were data (the hardware for one flight was lost at sea), the O–rings were examined for damage.

One interesting question is the relationship of O–ring damage to temperature (particularly since it was (forecasted to be) cold — 31^o F — on the morning of January 28, 1986). There was a good deal of discussion among the Morton Thiokol engineers the previous day as to whether the flight should go on as planned or not (an important point is that no statisticians were involved in the discussions). A simplified version of one of the arguments made is as follows. There were 7 previous flights where there was damage to at least one O–ring. Consider the following table. The entry $\hat{p}$ is the frequency estimate of the probability of an O–ring failing for that flight.

Ambient temperature	$\hat{p}$
53^o	.333
57^o	.167
58^o	.167
63^o	.167
70^o	.167
70^o	.167
75^o	.333

If you look at the table above, there's no apparent relationship between temperature and the probability of damage; higher damage occurred at both lower and higher temperatures. Thus, the fact that it was going to be cold on the day of the flight doesn't imply that the flight should be scrubbed. (In fact, this table was not actually constructed the night of January 27th, but was rather given later by two Thiokol staff members as an example of the reasoning in the pre–launch debate. The actual charts faxed from the Thiokol engineers to NASA that night were considerably less informative than even this seriously flawed table.)

Unfortunately, this analysis is completely inappropriate. The problem is that it is ignoring the 16 flights where there was no O–ring damage, acting as if there is no information in those flights. This is clearly absurd! If flights with high temperatures **never** had O–ring damage, for example, that would certainly tell us a lot about the relationship between temperature and O–ring damage! In fact, here is a scatter plot of the frequency estimates of the probability of O–ring damage versus temperature for **all** of the flights:



The picture is very different now. With the exception of the one observation in the upper right of the plot, there is a clear inverse relationship between the probability of O–ring damage and the ambient temperature — lower temperature is associated with higher probability of failure (the unusual observation is the flight of the Challenger from October 30 through November 6, 1985; one way that it was different was that the two O–rings damaged in that flight suffered only "blow–by" [where hot gases rush past the O–ring], while in all of the other flights damaged O–rings suffered "erosion" [where the O–rings burn up], as well as (possibly) blow–by). A plot of this kind would certainly have raised some alarms as to the advisability of launching the shuttle. Unfortunately, such a plot was never constructed.

Here is the full set of data:

Row	Temp	Damaged	O-rings
1	53	2	6
2	57	1	6
3	58	1	6
4	63	1	6
5	66	0	6
6	67	0	6
7	67	0	6
8	67	0	6
9	68	0	6
10	69	0	6
11	70	0	6
12	70	0	6
13	70	1	6
14	70	1	6
15	72	0	6
16	73	0	6

17	75	0	6
18	75	2	6
19	76	0	6
20	76	0	6
21	78	0	6
22	79	0	6
23	81	0	6

Logistic regression can be used to analyze the relationship between temperature and the probability of O–ring failure more precisely. In this case, the number of failures is the target variable (which `MINITAB` calls *Success*, remember), and the program is told that the number of trials is given in a variable `O-rings` (which is 6 for each flight here). Here is the output of the logistic analysis:

```
Binary Logistic Regression

Link Function:  Logit

Response  Information

Variable  Value          Count
Damaged   Success            9
          Failure          129
O-rings   Total            138

Logistic Regression Table
                                                      Odds        95% CI
Predictor       Coef      StDev         Z      P     Ratio    Lower    Upper
Constant       5.085      3.052      1.67 0.096
Temperat    -0.11560    0.04702     -2.46 0.014       0.89     0.81     0.98

Log-Likelihood = -30.198
Test that all slopes are zero: G = 6.144, DF = 1, P-Value = 0.013

Goodness-of-Fit Tests

Method               Chi-Square    DF       P
Pearson                  13.572    14   0.482
Deviance                 11.956    14   0.610
Hosmer-Lemeshow           5.677     4   0.225

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

                        Group
Value      1     2     3     4     5     6   Total
Success
  Obs      0     2     2     0     2     3       9
  Exp      0.3   0.6   1.6   1.8   2.0   2.7
Failure
  Obs     18    22    34    30    16     9     129
  Exp     17.7  23.4  34.4  28.2  16.0   9.3

  Total   18    24    36    30    18    12     138
```

```
Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs           Number  Percent     Summary Measures
Concordant         759    65.4%     Somers' D                 0.38
Discordant         315    27.1%     Goodman-Kruskal Gamma     0.41
Ties                87     7.5%     Kendall's Tau-a           0.05
Total             1161   100.0%
```

The slope coefficient has the following natural interpretation: each increase in temperature by one degree Fahrenheit is associated with an estimated multiplication of the relative odds of an O-ring failure

$$\frac{P(\mathrm{O-ring\ fails})}{P(\mathrm{O-ring\ does\ not\ fail})}$$

by $\exp(-.1156) = 0.891$, or roughly an 11% decrease. This value is given in the output under `Odds Ratio`, along with a 95% confidence interval. If this interval does not contain 1, there is significant predictive power of the predictor on the probability of success (at a .05 level).

There are two other tests given related to the strength of the predictive power of temperature for probability of an O–ring failure. The z–statistic of -2.46 for `Temperature` corresponds to a t–statistic in linear regression, and is called a *Wald* statistic (it is equivalent to the odds ratio confidence interval comparison mentioned above). The G–statistic, given as testing that all slopes are zero, corresponds to the F–statistic for overall significance in linear regression. Note that for least squares linear regression these two tests are equivalent when there is one predictor, but here the tail probabilities are slightly different, demonstrating that the two tests are not exactly equivalent.

The three goodness–of–fit tests are designed to test whether the logistic model fits the data adequately. All three are based on a χ^2–test construction. For each value of temperature given in the data (there are $J = 16$ distinct values in these data), let $\hat{p}_j$ be the fitted probability of O–ring failure, let f_j be the observed number of O–rings that failed, and let n_j be the number of O–rings at risk for that temperature (6 for each flight at that temperature). Note that looking at the data this way means that all flights at a given ambient temperature are pooled together and treated as indistinguishable. These values can be obtained as `Storage` from a logistic fit, and are as follows:

Row	Temperature	NOCC1	NTRI1	EPRO1
1	53	2	6	0.260787
2	57	1	6	0.181787
3	58	1	6	0.165220
4	63	1	6	0.099940
5	66	0	6	0.072783
6	67	0	18	0.065357
7	67	*	*	*
8	67	*	*	*
9	68	0	6	0.058640
10	69	0	6	0.052575
11	70	2	24	0.047106
12	70	*	*	*
13	70	*	*	*
14	70	*	*	*
15	72	0	6	0.037749
16	73	0	6	0.033767
17	75	2	12	0.026985
18	75	*	*	*

19	76	0	12	0.024110
20	76	*	*	*
21	78	0	6	0.019229
22	79	0	6	0.017166
23	81	0	6	0.013671

Note that the n_j values range from 6 to 24. The sum of the n_j values is the total sample size, or here 132.

The Pearson goodness–of–fit statistic equals

$$X^2 = \sum_j \frac{(f_j - n_j\hat{p}_j)^2}{n_j\hat{p}_j},$$

while the deviance statistic equals

$$G^2 = 2\sum_j f_j \ln\left(\frac{f_j}{n_j\hat{p}_j}\right).$$

When the n_j values are reasonably large, each of these statistics follows a χ^2 distribution on $J-p-1$ degrees of freedom, where p is the number of predictors in the model, under the null hypothesis that the logistic regression model fits the data. Thus, a small tail probability suggests that the linear logistic regression model is not appropriate for the data. Here both tests have high tail probabilities, indicating no problem with the linear logistic model.

Unfortunately, these tests are not trustworthy when the n_j values are small (the $n_j = 6$ values here are marginal). This is the justification for the third goodness–of–fit test, the Hosmer–Lemeshow test. In this test, all of the 138 observations are ordered by estimated O–ring failure probability (of course for these data all of the O–rings for a given flight have the same value of `Temp`, and therefore the same estimated probability of O–ring failure). The observations are then divided into g roughly equisized groups; g is usually taken to be 10, except when that would lead to too few observations in each group (as is the case here, where $g = 6$). Based on this new categorization of the data there are values of f_j, n_j and $n_j\hat{p}_j$, all of which are given in the Hosmer–Lemeshow table in the output. Then, the Hosmer–Lemeshow goodness–of–fit test is the usual Pearson goodness–of–fit test based on the new categorization, which is compared to a χ^2 distribution on $g-2$ degrees of freedom. It can be seen that the Hosmer–Lemeshow test also does not indicate a lack of fit here. Even the Hosmer–Lemeshow test is suspect, however, when its expected counts for either group are too small (less than two or three, say), which is the case here.

The statistical significance and goodness–of–fit of this model are comforting, of course, but does temperature provide predictive power of any practical importance? Some guidance to answer this question is given in the output under `Measures of Association`. Consider the fitted logistic regression model, with resultant fitted probabilities of O–ring failure $\hat{p}$ for each of the $n = 138$ observations. There are $n_1 = 9$ observed O–ring failures, and $n_0 = 129$ observed non–failures. Consider each of the pairs (i, j) of observations where one observation is a failure (i) and the other is a non–failure (j). There are $9 \times 129 = 1161$ such pairs, each of which has a corresponding pair $(\hat{p}_i, \hat{p}_j)$. We would like the estimated probability of failure to be higher for the observed failure observation than for the observed non–failure observation; that is, $\hat{p}_i > \hat{p}_j$. Such pairs are called *concordant*. If for a given pair $\hat{p}_i < \hat{p}_j$, the pair is called *discordant*. We would like to have a high percentage of concordant pairs, and a low percentage of discordant pairs. Here there are 65.4% concordant pairs and 27.1% discordant ones, a reasonably good performance. There are no formal cutoffs for what constitutes a "good enough" performance here, but observed values can be compared for different possible models to assess relative practical performance. The statistics Somers' D, Goodman–Kruskal γ and Kendall's τ_a are different ways of summarizing these concordancies and discordancies, with higher values indicating more concordancy (e.g., D is the difference between concordant and discordant pairs).

Just as is true for other regression models, unusual observations can have a strong effect on a fitted logistic regression model. Among the diagnostics that are available for logistic regression are three that roughly correspond to the standardized residuals (here the standardized Pearson residuals), Cook's distance (here the standardized Delta–beta $[\Delta\beta]$) and leverage values. Here are the values here:

Row	Temperature	SPRE1	DSBE1	HI1
1	53	0.56258	0.29503	0.482442
2	57	-0.10892	0.00340	0.222683
3	58	0.01054	0.00002	0.180967
4	63	0.56529	0.02428	0.070609
5	66	-0.70527	0.02790	0.053119
6	67	-1.21983	0.27107	0.154100
7	67	*	*	*
8	67	*	*	*
9	68	-0.62752	0.02110	0.050853
10	69	-0.59239	0.01894	0.051209
11	70	0.94160	0.23361	0.208538
12	70	*	*	*
13	70	*	*	*
14	70	*	*	*
15	72	-0.49902	0.01443	0.054772
16	73	-0.47134	0.01322	0.056146
17	75	3.17765	1.33690	0.116919
18	75	*	*	*
19	76	-0.57995	0.04523	0.118540
20	76	*	*	*
21	78	-0.35377	0.00799	0.060045
22	79	-0.33389	0.00712	0.059995
23	81	-0.29729	0.00555	0.059061

There is an apparent outlier at row 17, corresponding to an ambient temperature of 75^o. Unfortunately, since there are two flights at that temperature, we can't tell for sure which is actually the outlier (of course, in this case we know what it is from the earlier graph, but in general the collapsing approach of `Minitab` makes it difficult to tell which observation is actually an outlier if there are replications in the data.

For this reason, it's a good idea to try to remove the collapsing effect by forcing each observation to have a unique set of predictor variable values, at least when looking at diagnostics. The way this is done is by "jittering" at least one predicting variable by adding a small amount of random noise to the variable. In fact, in this context we know that the temperature values are only given to the nearest integer value, so this is not at all unreasonable, but even if the values were exact, we need to do this if we want to examine diagnostics. Here are the results of a logistic fit using jittered temperature:

```
Binary Logistic Regression

Link Function:  Logit

Response  Information

Variable  Value         Count
Damaged   Success           9
          Failure         129
O-rings   Total           138

Logistic Regression Table
                                                        Odds        95% CI
Predictor       Coef       StDev          Z      P     Ratio     Lower    Upper
Constant       5.085       3.052       1.67 0.096
Temperat    -0.11561     0.04702      -2.46 0.014       0.89      0.81     0.98
```



```
Log-Likelihood = -30.198
Test that all slopes are zero: G = 6.145, DF = 1, P-Value = 0.013

Goodness-of-Fit Tests

Method              Chi-Square    DF      P
Pearson                 29.985    21  0.092
Deviance                18.085    21  0.644
Hosmer-Lemeshow          8.207     6  0.223

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

                              Group
Value      1     2     3     4     5     6     7     8   Total
Success
  Obs      0     0     2     2     0     0     2     3       9
  Exp      0.3   0.5   0.6   0.8   1.0   1.2   2.0   2.7
Failure
  Obs     18    18    16    16    18    18    16     9     129
  Exp     17.7  17.5  17.4  17.2  17.0  16.8  16.0   9.3

  Total   18    18    18    18    18    18    18    12     138

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs           Number  Percent    Summary Measures
Concordant         759    65.4%    Somers' D                 0.38
Discordant         315    27.1%    Goodman-Kruskal Gamma     0.41
Ties                87     7.5%    Kendall's Tau-a           0.05
Total             1161   100.0%
```

As would be expected, the output changes very little, with the exception of the Pearson and Deviance statistics (recall that their construction depends on how covariate patterns are defined). Here are the diagnostics:

Row	Temperature	SPRE2	DSBE2	HI2	EPRO2
1	53	0.56312	0.29509	0.482020	0.260696
2	57	-0.10927	0.00342	0.222782	0.181836
3	58	0.00959	0.00002	0.181253	0.165351
4	63	0.56538	0.02428	0.070600	0.099932
5	66	-0.70541	0.02792	0.053130	0.072810
6	67	-0.66515	0.02396	0.051372	0.065376
7	67	-0.66474	0.02392	0.051361	0.065302
8	67	-0.66465	0.02392	0.051359	0.065285
9	68	-0.62743	0.02109	0.050856	0.058624
10	69	-0.59263	0.01896	0.051208	0.052615
11	70	-0.55958	0.01722	0.052131	0.047136
12	70	-0.55956	0.01722	0.052132	0.047133
13	70	1.41935	0.11080	0.052134	0.047124

```
14          70   1.41904   0.11075   0.052132   0.047135
15          72  -0.49870   0.01442   0.054790   0.037702
16          73  -0.47110   0.01321   0.056161   0.033734
17          75  -0.42012   0.01096   0.058472   0.026950
18          75   4.77345   1.41487   0.058464   0.026976
19          76  -0.39706   0.00993   0.059268   0.024123
20          76  -0.39697   0.00993   0.059271   0.024112
21          78  -0.35357   0.00799   0.060046   0.019208
22          79  -0.33384   0.00711   0.059994   0.017161
23          81  -0.29737   0.00555   0.059064   0.013678
```

Now it's clear that the previously mentioned flight (number 18) is a very clear outlier, with 2 of 6 O–rings damaged when the estimated probability of O–ring damage was only .027. Here is output from the data with that flight omitted (I'm sticking with the jittered data):

```
Binary Logistic Regression

Link Function:  Logit

Response  Information

Variable  Value          Count
Damaged   Success            7
          Failure          125
O-rings   Total            132

Logistic Regression Table
                                                         Odds        95% CI
Predictor       Coef      StDev         Z      P    Ratio    Lower    Upper
Constant       8.664      3.635      2.38 0.017
Temperat    -0.17684    0.05870     -3.01 0.003      0.84     0.75     0.94

Log-Likelihood = -22.039
Test that all slopes are zero: G = 10.660, DF = 1, P-Value = 0.001

Goodness-of-Fit Tests

Method                 Chi-Square    DF       P
Pearson                    13.398    20   0.860
Deviance                    9.406    20   0.978
Hosmer-Lemeshow             3.908     6   0.689

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

                              Group
Value       1     2     3     4     5     6     7     8   Total
Success
  Obs       0     0     1     1     0     0     3     2       7
  Exp     0.1   0.2   0.3   0.4   0.6   0.8   2.7   2.0
Failure
  Obs      18    18    17    17    18    18    15     4     125
  Exp    17.9  17.8  17.7  17.6  17.4  17.2  15.3   4.0
```

Total	18	18	18	18	18	18	18	6	132

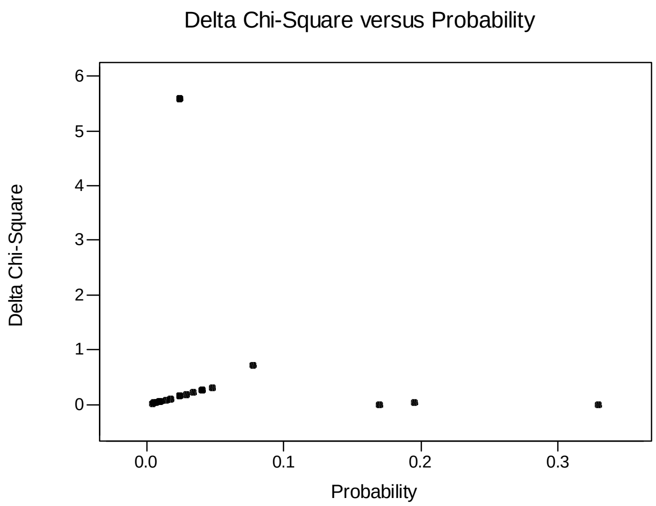
Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	671	76.7%	Somers' D	0.61
Discordant	137	15.7%	Goodman-Kruskal Gamma	0.66
Ties	67	7.7%	Kendall's Tau-a	0.06
Total	875	100.0%		

The strength of the relationship has gone up considerably once the outlier is removed, with there now being an estimated 16% reduction in the odds of an O–ring being damaged with each additional degree of temperature at launch. The goodness–of–fit tests suggest no lack of fit (remember, the Pearson and deviance tests are at least marginally valid here, since there are 6 replications for each flight). Here are diagnostics:

Row	Temperature	SPRE3	DSBE3	HI3	EPRO3
1	53	0.02865	0.001039	0.558615	0.329680
2	57	-0.20157	0.012003	0.228058	0.195330
3	58	-0.01796	0.000071	0.180636	0.169155
4	63	0.85067	0.060061	0.076638	0.077459
5	66	-0.56312	0.021926	0.064673	0.047104
6	67	-0.51494	0.017792	0.062877	0.039768
7	67	-0.51446	0.017753	0.062861	0.039698
8	67	-0.51435	0.017745	0.062857	0.039682
9	68	-0.47077	0.014490	0.061368	0.033508
10	69	-0.43096	0.011837	0.059913	0.028277
11	70	-0.39412	0.009620	0.058318	0.023798
12	70	-0.39410	0.009619	0.058318	0.023796
13	70	2.36664	0.346845	0.058314	0.023789
14	70	2.36606	0.346697	0.058318	0.023798
15	72	-0.32907	0.006236	0.054450	0.016779
16	73	-0.30092	0.004986	0.052192	0.014103
17	75	-0.25142	0.003125	0.047109	0.009939
18	76	-0.23015	0.002463	0.044429	0.008365
19	76	-0.23006	0.002460	0.044418	0.008359
20	78	-0.19204	0.001489	0.038797	0.005873
21	79	-0.17564	0.001153	0.036032	0.004932
22	81	-0.14686	0.000682	0.030671	0.003472

There are no extreme outliers, but the low temperature cases are possible leverage points (this is not surprising, given that most launches were at temperatures over 65^o). The noteworthy 70^o observations correspond to two 70^o flights where there was an O–ring failure. Omitting these two flights doesn't change things very much (strengthening the relationship further), although a plot of the change in the Pearson statistic versus estimated probability does show the two points as unusual.



Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Damaged	Success	5
	Failure	115
O-rings	Total	120

Logistic Regression Table

					Odds	95% CI	
Predictor	Coef	StDev	Z	P	Ratio	Lower	Upper
Constant	14.417	5.710	2.52	0.012			
Temperat	-0.28028	0.09864	-2.84	0.004	0.76	0.62	0.92

Log-Likelihood = -13.319
Test that all slopes are zero: G = 14.931, DF = 1, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	3.446	18	1.000
Deviance	2.780	18	1.000
Hosmer-Lemeshow	1.564	8	0.992

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

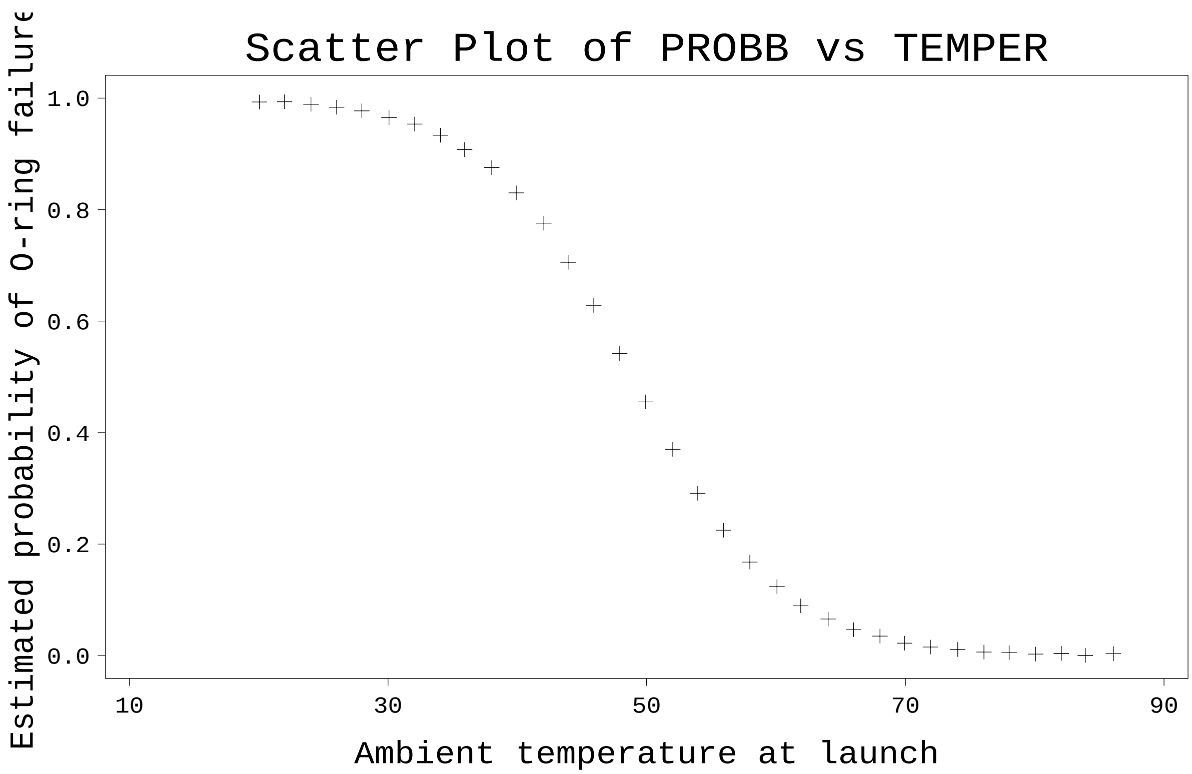
	Group										
Value	1	2	3	4	5	6	7	8	9	10	Total
Success											
Obs	0	0	0	0	0	0	0	0	2	3	5
Exp	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.2	1.1	3.4	
Failure											

Obs	12	12	12	12	12	12	12	12	10	9	115
Exp	12.0	12.0	12.0	12.0	11.9	11.9	11.8	11.8	10.9	8.6	
Total	12	12	12	12	12	12	12	12	12	12	120

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	525	91.3%	Somers' D	0.87
Discordant	27	4.7%	Goodman-Kruskal Gamma	0.90
Ties	23	4.0%	Kendall's Tau-a	0.07
Total	575	100.0%		

What about the morning of January 28, 1986? Here is a plot of the logistic curve for different values of temperature based on all flights except the October/November 1985 flight:



Substituting into the logistic function gives a probability estimate of O–ring failure for a temperature of 31^o of **.96**! (This is an extrapolation, but you get the idea.) Indeed, with the benefit of hindsight, it can be seen that the Challenger disaster was not at all surprising, **given data that were available at the time of the flight**. As a result of its investigations, one of the recommendations of the commission was that a statistician be part of the ground control team from that time on. A complete (and more correct) discussion of this material can be found in the paper "Risk Analysis of Space Shuttle: Pre–Challenger Prediction of Failure," by S.R. Dalal, E.B. Fowlkes and B.A. Hoadley, *Journal of the American Statistical Association*, **84**, 945–957 (1989). Chapter 2 of Edward R. Tufte's 1997 book *Visual Explanations: Images and Quantities, Evidence and Narrative* discusses the background of the disaster, and the charts used by the Thiokol engineers in their discussions with NASA.

By the way, an alternative way that these data might have been presented was as a set of 138 observations (one for each O–ring, rather than one for each flight), with a 0/1 target variable reflecting failure or non–failure of each O–ring. This is what the dataset would look like:

Row	Temp	Failed
1	53	1
2	53	1
3	53	0
4	53	0
5	53	0
6	53	0
7	57	1
8	57	0
9	57	0
10	57	0
11	57	0
12	57	0
13	58	1
14	58	0
15	58	0
16	58	0
17	58	0
18	58	0
19	63	1
20	63	0
21	63	0
22	63	0
23	63	0
24	63	0
25	66	0
26	66	0
27	66	0
28	66	0
29	66	0
30	66	0
31	67	0
32	67	0
33	67	0
34	67	0
35	67	0
36	67	0
37	67	0
38	67	0
39	67	0
40	67	0
41	67	0
42	67	0
43	67	0
44	67	0
45	67	0
46	67	0
47	67	0

48	67	0
49	68	0
50	68	0
51	68	0
52	68	0
53	68	0
54	68	0
55	69	0
56	69	0
57	69	0
58	69	0
59	69	0
60	69	0
61	70	0
62	70	0
63	70	0
64	70	0
65	70	0
66	70	0
67	70	0
68	70	0
69	70	0
70	70	0
71	70	0
72	70	0
73	70	1
74	70	0
75	70	0
76	70	0
77	70	0
78	70	0
79	70	1
80	70	0
81	70	0
82	70	0
83	70	0
84	70	0
85	72	0
86	72	0
87	72	0
88	72	0
89	72	0
90	72	0
91	73	0
92	73	0
93	73	0
94	73	0
95	73	0
96	73	0
97	75	0
98	75	0
99	75	0
100	75	0

101	75	0
102	75	0
103	75	1
104	75	1
105	75	0
106	75	0
107	75	0
108	75	0
109	76	0
110	76	0
111	76	0
112	76	0
113	76	0
114	76	0
115	76	0
116	76	0
117	76	0
118	76	0
119	76	0
120	76	0
121	78	0
122	78	0
123	78	0
124	78	0
125	78	0
126	78	0
127	79	0
128	79	0
129	79	0
130	79	0
131	79	0
132	79	0
133	81	0
134	81	0
135	81	0
136	81	0
137	81	0
138	81	0

Which representation is better? It turns out not to matter; if you analyze the data in this form, where `Failed` is chosen as the `Response` variable in the `Minitab` dialog box, the resultant output will be identical to that obtained using the data represented at the level of 23 different flights. There is one advantage to the earlier representation, however; since the natural way to view these data is at the flight level, rather than the O–ring level, jittering the data in the flight–level form is more natural.

MINITAB commands

Logistic regression modeling is obtained by clicking on `Stat` → `Regression` → `Binary Logistic Regression`. There are various ways that the data might be presented, which affect the command structure to the program. The two most common forms are as follows:

(1) The target variable is given as the number of successes out of the number of trials (or the number of items "at risk"). Enter the variable with the number of successes in the box next to `Success:`, and enter the variable with the number of trials in the box next to `Trial:`.

(2) The target variable is a 0/1 variable that represents success or failure for each observation. Enter the name of this variable in the box next to `Response:`.

The predicting variables for the model are entered under `Model:`. This includes both continuous variables and categorical ones. Categorical variables must also be entered under `Factors (optional):`. Interactions with (and between) factors are entered under `Model:` using the "multiplication" form as in ANOVA modeling.

Diagnostics, such as standardized Pearson residuals, leverages, and delta betas are obtained by clicking on `Storage`. This dialog box also allows storage of fitted success probabilities under `Event probability`, the number of successes for each distinct covariate pattern under `Number of occurrences of the event`, and the number of trials for each covariate pattern under `Number of trials`. Diagnostic plots, such as one of Delta Chi–Square versus Event Probability, are obtained by clicking on `Graphs`.

To jitter a predicting variable, follow the following steps:

(1) Go to `Calc` → `Random Data` → `Uniform`. In the box labeled *Generate rows of data*, enter the sample size (the number of observations, not the number of covariate patterns). Put in a new variable name under *Store in columns:*, such as `Jitter`. Under *Lower endpoint:* enter a negative number close to zero, such as $-.01$. Under *Upper endpoint:* enter the positive version of this number (.01, for example). The numbers you choose here should be smaller than the resolution of all of your predictors; so, for example, if one predictor is given to three decimal digits, use $-.0001$ and .0001 here.

(2) Go to the calculator, and create new versions of each predictor you wish to jitter as sums of the predictor with `Jitter`. So, for example, the variable `Newpred1` is determined as `Predict1` + `Jitter`.

(3) Fit the logistic regression model using the new predictors in place of the old ones.