Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure
Author(s): Siddhartha R. Dalal, Edward B. Fowlkes and Bruce Hoadley
Source: *Journal of the American Statistical Association*, Vol. 84, No. 408 (Dec., 1989), pp. 945-957
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2290069
Accessed: 26-01-2018 01:26 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

American Statistical Association, Taylor & Francis, Ltd. are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

Risk Analysis of the Space Shuttle: Pre-*Challenger* Prediction of Failure

SIDDHARTHA R. DALAL, EDWARD B. FOWLKES, and BRUCE HOADLEY*

The Rogers Commission report on the space shuttle *Challenger* accident concluded that the accident was caused by a combustion gas leak through a joint in one of the booster rockets, which was sealed by a device called an O-ring. The commission further concluded that O-rings do not seal properly at low temperatures. In this article, data from the 23 preaccident launches of the space shuttle is used to predict O-ring performance under the *Challenger* launch conditions and relate it to the catastrophic failure of the shuttle. Analyses via binomial and binary logistic regression show that there is strong statistical evidence of a temperature effect on incidents of O-ring thermal distress. In addition, a probabilistic risk assessment at 31°F, the temperature at which *Challenger* was launched, yields at least a 13% probability of catastrophic field-joint O-ring failure. Postponement to 60°F would have reduced the probability to at least 2%. To assess uncertainty in estimates and for any future prediction under the *Challenger* scenario, a postanalysis prior distribution of the probability of a catastrophic failure is derived. The mean and median for this distribution for 31°F are at least .16 and .13, and for 60°F they are at least .004 and .02, respectively. The analysis of this article demonstrates that statistical science can play an important role in the space-shuttle risk-management process.

KEY WORDS: Catastrophic failure; Data analysis; O-rings; Probability risk assessment; Statistical science.

1. INTRODUCTION AND SUMMARY

On the night of January 27, 1986, the night before the space shuttle *Challenger* accident, there was a three-hour teleconference among people at Morton Thiokol (manufacturer of the solid rocket motor), Marshall Space Flight Center [NASA (National Aeronautics and Space Administration) center for motor design control], and Kennedy Space Center. The discussion focused on the forecast of a 31°F temperature for launch time the next morning, and the effect of low temperature on O-ring performance. A data set, Figure 1a, played an important role in the discussion. Each plotted point represents a shuttle flight that experienced thermal distress on the field-joint O-rings; the X axis shows the joint temperature at launch and the Y axis shows the number of O-rings that experienced some thermal distress. The O-rings seal the field joints of the solid rocket motors, which boost the shuttle into orbit. Based on the U configuration of points (identified by the flight number), it was concluded that there was no evidence from the historical data about a temperature effect.

Nevertheless, there was a debate on this issue, and some participants recommended that the launch be postponed until the temperature rose above 53°F—the lowest temperature experienced in previous launches—because the corresponding flight had the highest number of distressed O-rings. Some participants believed, based on the physical evidence, that there was a temperature effect on O-ring performance; for example, one of the participants, Roger Boisjoly, stated: "temperature was indeed a discriminator." In spite of this, the final recommendation of Morton Thiokol was to launch the *Challenger* on schedule. The recommendation transmitted to NASA stated that "Temperature data [are] not conclusive on predicting primary O-ring blowby." The same telefax stated that "Colder O-rings will have increased effective durometer ('harder'), and 'Harder' O-rings will take longer to 'seat'" [Presidential Commission Report, Vol. 1 (PC1), p. 97 (Presidential Commission on the Space Shuttle *Challenger* Accident 1986)].

After the accident a commission was appointed by President R. Reagan to find the cause. The commission was headed by former Secretary of State William Rogers and included some of the most respected names in the scientific and space communities. The commission determined the cause of the accident to be the following: "A combustion gas leak through the right Solid Rocket Motor aft field joint initiated at or shortly after ignition eventually weakened and/or penetrated the External Tank initiating vehicle structural breakup and loss of the Space Shuttle *Challenger* during mission 51-L" (PC1, p. 70). This is the type of failure that was debated the night before the *Challenger* accident.

The Rogers Commission (PC1, p. 145) noted that a mistake in the analysis of the thermal-distress data (Fig. 1a) was that the flights with zero incidents were left off the plot because it was felt that these flights did not contribute any information about the temperature effect (see Fig. 1b). The Rogers Commission concluded that "A careful analysis of the flight history of O-ring performance would have revealed the correlation of O-ring damage in low temperature" (PC1, p. 148).

This article aims to give more substance to this quote and show how statistical science could have provided valuable input to the launch decision process. Clearly, the key question was What would have constituted proof that it was unsafe to launch? Since the phenomenon is sto-

* Siddhartha R. Dalal is District Manager, Statistics and Econometrics Research Group, and the late Edward B. Fowlkes was Distinguished Member of Professional Staff, Statistics Research Group, Bellcore, Morristown, NJ 07960. Bruce Hoadley is Division Manager, Analytical Methods and Software Division, Navesink Research and Engineering Center, Red Bank, NJ 07701. The authors thank R. Gnanadesikan and J. R. Kettenring for encouraging them to work on this problem, and D. Duffy, J. Landwehr, C. Mallows, the editor, an associate editor, and the referees for helpful comments. This article is dedicated to the memory of Edward B. Fowlkes.

© 1989 American Statistical Association
Journal of the American Statistical Association
December 1989, Vol. 84, No. 408, Applications & Case Studies

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

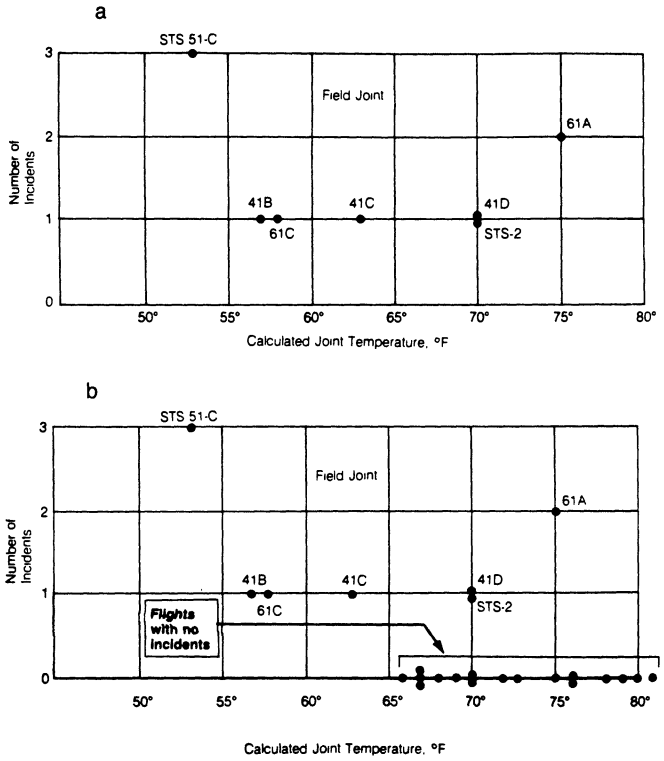

Figure 1. Joint Temperature Versus Number of O-Rings Having Some Thermal Distress Identified by Flight Number. Panel b includes flights with no incidents.

chastic, the answer is necessarily probabilistic. As in the teleconference, a good start would have been an examination of the thermal-distress data (Fig. 1b) for the presence of a temperature effect. Nevertheless, the most important question was What is the probability of catastrophic field-joint failure if we launch tomorrow morning at 31°F? We address both of these issues.

Specifically, based on the qualitative and quantitative knowledge available before the *Challenger*'s last launch (summarized in Sec. 2, with a brief description of the shuttle subsystems), we do the following.

1. In Section 3 we show that the thermal-distress data contains strong statistical evidence of a temperature effect on O-rings. This fact alone could have had an impact on the discussion during the teleconference and the subsequent decision to launch.

2. Using the analysis in Section 3 as a blueprint, in Section 5 we provide a probabilistic risk assessment of a catastrophic field-joint O-ring failure under *Challenger*'s launch conditions. We quantify the partial-degradation data and the engineering knowledge and show that the physical probability of a shuttle failure under the *Challenger*'s launch scenario was at least .13. Postponement of the flight to 60°F would have reduced the risk to at least .02.

3. In Section 5 we assess uncertainty in our estimates (given in Sec. 5.2) by deriving a postanalysis prior distribution for the probability of failure of a future flight under the *Challenger*-type design. The mean and median of this distribution for 31°F are at least .16 and .13, and that for 60°F are at least .004 and .02, respectively.

4. We conduct many other statistical and diagnostic analyses to study (a) nozzle-joint O-ring performance, (b) leak-check pressure effects, (c) influential observations, and (d) possible lack of fit. The analysis of the nozzle-joint O-ring thermal-incidents data in Section 4 is relevant, because the nozzle and field joints use the same primary O-ring.

Some of the work on this article was done while one of the authors, Bruce Hoadley, was a member of the National Research Council's Shuttle Criticality Review Hazard Analysis Audit Committee (SCRHAAC). This committee was called for in the third recommendation of the Rogers Commission.

SCRHAAC's recommendations (SCRHAAC 1988), which were influenced by this work, have already received substantial press coverage and have influenced NASA. NASA has begun to build a staff skilled in statistical science and begun to conduct probabilistic risk assessments of major subsystems. This article is intended to act as a blueprint for such future NASA analyses.

2. THE SHUTTLE SYSTEM AND EVENTS LEADING TO THE *CHALLENGER* ACCIDENT

Figure 2 gives a diagram of the shuttle system. The system consists of four subsystems, marked 1–4. Subsystem 1 is the orbiter that houses the crew and the controls of the system. Subsystem 2 is the external liquid-fuel tank for the space shuttle's main engines on the orbiter. Subsystems 3 and 4 are the solid rocket motors manufactured by Morton Thiokol. We concentrate on the solid rocket motors because the Rogers Commission Report (PC1) concluded that the O-rings used to seal the motor field joints were the cause of the *Challenger* accident.

Each of the solid rocket motors is shipped in four pieces and assembled at the Kennedy Space Center. The corresponding three joints, indicated by the arrows in Figure 2, are referred to as *field joints*. A similar joint, with a nozzle, is known as a *nozzle joint*. After each launch, the rocket motors are recovered from the ocean for inspection and possible reuse. There were 24 launches prior to *Challenger*. For one flight the motors were lost at sea, so motor data were available for 23 flights.

Figure 3 shows a cross-section of a typical field joint. The joint is formed where the tang segment fits into the clevis segment. To seal the small space left between the two segments, O-rings are used. For redundancy, two O-rings (primary and secondary) 37.5 feet in diameter and .28 inches thick are employed. (In the figure their cross-sections are denoted by two dots.)

At ignition of the solid rocket motor, pressure and heat build up rapidly inside the motor. The O-rings erode under intense heat, so putty is used to protect them. The pressure displaces the putty toward the O-rings. This dislacement causes air pressure to build up behind the primary O-ring, which energizes it to seal the joint.

The leak test port, shown in Figure 3, was a design innovation to allow a pressure test of the O-rings after assembly. Originally, the test was conducted at 50 pounds

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

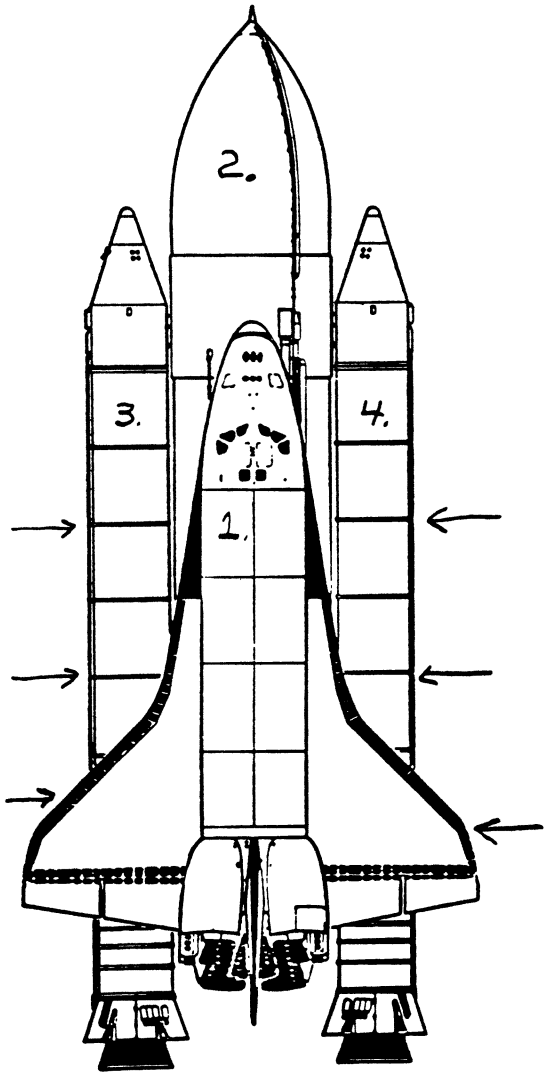

Figure 2. Space Shuttle: Orbiter, External Tank, Solid Rocket Motors, and Field Joints.

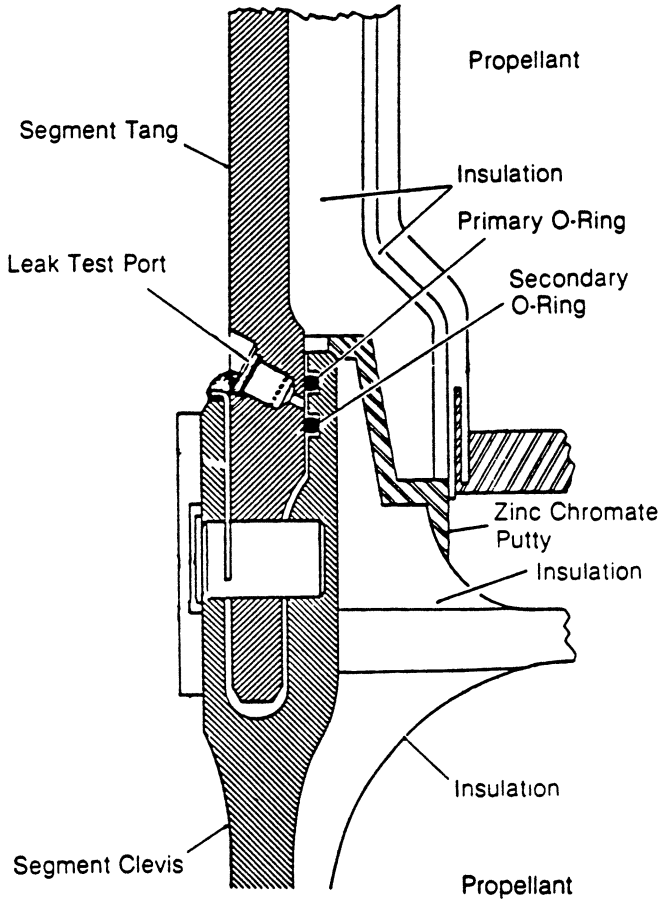

Figure 3. Solid Rocket Motor Cross Section: Tang, Clevis, and O-Rings.

per square inch (psi). But it was discovered that the putty alone could hold 50 psi; hence the test pressure was increased to 100 psi and finally 200 psi. During the late 1970s and early 1980s various developmental solid rocket motors were used to qualify the design; the engine was qualified to 40°F.

During the years prior to the *Challenger* flight (even before the first shuttle flight in 1981), evidence of field-joint O-ring reliability problems was accumulating. For example, in 1977 NASA discovered field-joint rotation. Soon after ignition, the combustion pressure would cause the sides of the motor's metal casings to bulge out a bit like a balloon. The maximum bulge was halfway between the joints. This bulging would cause the tang to rotate relative to the clevis, which would make the gap—which had to be sealed by the O-rings—grow larger.

Even with joint rotation, the primary O-ring would normally seal the gap. Since the secondary O-ring is not energized when the primary O-ring is sealed, however, the joint rotation could cause the secondary O-ring to lose contact with the tang. Then, if the primary O-ring should fail (e.g., because of erosion) the hot gases could escape through the gap between the tang and secondary O-ring and cause catastrophic failures. Following the occurrence of joint rotation, various NASA memos (PC1, p. 123) stated that design change was "mandatory to prevent hot gas leaks and resulting catastrophic failure" (1/9/78) and "forcing the seal to function in a way which violates industry and government O-ring application practices" (1/19/79). Because of the joint-rotation possibility, in 1982 NASA reclassified the secondary O-rings as a nonredundant component.

Soon after the shuttle flights began in 1981, thermal distress on O-rings began to appear. For the thermal-distress data in Figure 1b, there were two kinds of distress: erosion and blowby. Erosion is caused by excessive heat burning up the O-ring. Blowby happens when gases rush by the O-ring. Blowby can occur before the O-ring seals, or after it seals and subsequently fails. After ignition, it takes a short time for the O-ring to be energized and seal. During that time, hot gases can "blow by" the O-ring if they have leaked through the putty. After the O-ring seals, it can erode because of contact with the hot gas. If it erodes too much, the seal will fail. Then, blowby will take place again. One suspected cause of erosion was "blow holes" in the putty. The increase of leak test pressure from 50 psi to 200 psi (mentioned earlier) could have caused some of these blow holes.

To understand erosion, deterministic detailed engineer-

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

ing models were developed to predict erosion depth. According to PC1 (p. 133), Thiokol calculated that the maximum possible impingement erosion was .09 inches.

On January 24, 1985, flight 51-C was launched at 53°F—the coldest temperature to date. The field joints experienced severe primary O-ring erosion and blowby and secondary O-ring erosion. At the February flight-readiness review for flight 51-D, it was stated that "low temperature enhanced probability of blowby" and that "the condition is not desirable but is acceptable" (PC1, p. 136).

On April 29, 1985 (PC1, p. 137), the nozzle-joint primary O-ring for flight 51-B eroded .171 inches and never sealed; the secondary O-ring eroded .032 inches. And this happened at 75°F. After this event, a memo mentioned a new and significant event . . . that we certainly did not understand" (PC1, p. 137). A July 1985 Thiokol memo stated that "If the same scenario should occur in a field joint (and it could), then it is a jump ball. . . . The result would be a catastrophe" (PC1, p. 139).

In June 1985 Thiokol conducted a bench test of O-ring resiliency, a hardware simulation of O-ring squeeze in the normal stacked position, followed by joint rotation. The results were the following: "At 100°F, the O-ring maintained contact. . . . At 50°F, the O-ring did not reestablish contact" (PC1, p. 137). The Rogers Commission reiterated this result: "O-ring resiliency is directly related to its temperature. A warm O-ring that has been compressed will return to its original shape much quicker than will a cold O-ring when the compression is relieved [thus a warm O-ring will seal a joint appropriately and a cold O-ring may not]" (PC1, pp. 70–71).

At an August 1985 NASA briefing on the O-ring problem (Presidental Commission Report, Vol. 2, p. H-84), it was stated that the "qualitative probability" of secondary O-ring failure, given erosion penetration of the secondary O-ring, is "high" after 330 milliseconds of the ignition.

All of this shows that the pre-*Challenger* teleconference was conducted in an atmosphere of doubt about field-joint O-ring reliability—particularly at low temperatures.

3. DATA ANALYSIS FOR FIELD–JOINT O–RINGS

3.1 Model Fitting

In this section an exploratory analysis of the thermal-distress data plotted in Figure 1b is considered. Detailed data are featured in PC1 (pp. 129–131) and the relevant portions are extracted in Table 1. We consider the relation of incidents of thermal distress to both temperature and leak-check pressure.

Since in Figure 1b there is only one incident of secondary O-ring damage, for exploratory analysis we consider thermal distress in only the primary O-rings. Recall that there are six primary O-rings per shuttle. Figure 4 shows a plot of the number of primary O-ring incidents of thermal distress versus temperature for 23 shuttle flights. Inspection of this plot suggests that aside from one point (75, 2), there is a tendency for the number of incidents to decrease with increasing temperature.

A statistical model appropriate for this data, conditional on the temperature, t, and pressure, s, is the binomial model with $n = 6$. Specifically, if $p(t, s)$ denotes the probability per joint of some thermal distress for a given t and s and X denotes the number of thermally distressed O-rings, then X is assumed to be a binomial variable with $n = 6$ and $p = p(t, s)$. The model assumes that at temperature t and pressure s each of the six O-rings would suffer damage independently with the same probability. Finally, to link $p(t, s)$ to t and s a logistic regression model is employed:

$$\log[p(t, s)/(1 - p(t, s))] = \alpha + \beta t + \gamma s. \quad (3.1)$$

This model was fitted using maximum likelihood; estimates of the coefficients are $\hat{\alpha} = 2.520$, $\hat{\beta} = -.0983$, and $\hat{\gamma} = .00848$. Estimated asymptotic standard errors for these coefficients were $s_{\hat{\alpha}} = 3.487$, $s_{\hat{\beta}} = .045$, and $s_{\hat{\gamma}} = .0077$. The lack-of-fit statistic G^2, which is twice the log-likelihood ratio of the fully saturated model to that of the model under consideration (say G_1^2), was 16.546 with 20 df, based on an asymptotic chi-squared distribution. This indicates a good fit. Nevertheless, in analogy with the pitfalls of using R^2 alone to gauge the fit of the standard linear model, one should be cautious about using G^2 alone for interpreting the fit here. Incremental changes in G^2 from adding or removing variables are perhaps more pertinent. A logistic regression model containing only temperature would be of the form

$$\log[p(t)/(1 - p(t))] = \alpha + \beta t. \quad (3.2)$$

The coefficients were found to be $\hat{\alpha} = 5.085$ and $\hat{\beta} = -.1156$, and their asymptotic standard errors were $s_{\hat{\alpha}} = 3.052$ and $s_{\hat{\beta}} = .047$. The lack-of-fit $G^2 = G_2^2$ for this model was 18.086 with 21 df (again a good fit). Temperature appears to be the most important variable, whereas pressure does not seem to have much effect. To test the hypothesis of no pressure effect, we look at the incremental G^2 by going from Models (3.1) to (3.2), which is asymptotically chi squared with 1 df. The increment $18.086 - 16.546 = 1.54$ is not significant but indicates that there may be a very weak pressure effect. 90% bootstrap confidence intervals for the expected numbers of incidents were calculated for each temperature, first holding pressure constant at 50 psi and next setting the pressure at 200 psi. The intervals for the two pressures overlapped greatly, which strengthens the argument that a pressure effect may be present, but it cannot be estimated with enough precision to include it in the model. If erosion and blowby are analyzed separately, there appears to be no pressure effect. The decision was made to drop the pressure variable from the model and examine the adequacy of the resulting model (3.2) using diagnostic methods.

For Model (3.2) we plotted the contours of the log-likelihood function. The contours were elliptical, suggesting that the data were not leading to ill-conditioned computations. Figure 4 shows a plot of the original data (asterisks) versus the estimated number of incidents (solid line) from Model (3.2). The fit looks reasonable.

A second model for the occurrence of O-ring incidents can be derived by considering a binary response variable

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

Table 1. O-Ring Thermal-Distress Data

		Field			Nozzle				Leak-check pressure	
Flight	Date	Erosion	Blowby	Erosion or blowby	Erosion	Blowby	Erosion or blowby	Joint temperature	Field	Nozzle
1	4/12/81							66	50	50
2	11/12/81	1		1				70	50	50
3	3/22/82							69	50	50
5	11/11/82							68	50	50
6	4/04/83				2		2	67	50	50
7	6/18/83							72	50	50
8	8/30/83							73	100	50
9	11/28/83							70	100	100
41-B	2/03/84	1		1	1		1	57	200	100
41-C	4/06/84	1		1	1		1	63	200	100
41-D	8/30/84	1		1	1	1	1	70	200	100
41-G	10/05/84							78	200	100
51-A	11/08/84							67	200	100
51-C	1/24/85	2, 1*	2	2		2	2	53	200	100
51-D	4/12/85				2		2	67	200	200
51-B	4/29/85				2, 1*	1	2	75	200	100
51-G	6/17/85				2	2	2	70	200	200
51-F	7/29/85				1			81	200	200
51-I	8/27/85				1			76	200	200
51-J	10/03/85							79	200	200
61-A	10/30/85		2	2	1			75	200	200
61-B	11/26/85				2	1	2	76	200	200
61-C	1/12/86	1		1	1	1	2	58	200	200
61-I	1/28/86							31	200	200
	Total	7, 1*	4	9	17, 1*	8	17			

*Secondary O-ring.

Y, which is 1 if there was an incident in that flight and 0 otherwise. Y is thus a binary random variable with the probability $p^*(t, s)$ of at least one O-ring incident. It was felt that this binary model would be more robust to the actual count of the number of incidents and still model $p^*(t, s)$. Furthermore, unlike the earlier binomial model, the statistical independence of each joint is not required. It is also useful for assessing the need for overdispersion models such as the beta binomial model. The information loss from considering the binary model is not serious, since most of the data correspond to 0 or 1 incidents. The contribution of pressure was again negligible, and consequently we removed the pressure variable. For the binary model with the temperature alone, the estimated coefficients were $\hat{\alpha}^* = 15.043$ and $\hat{\beta}^* = -.2322$. Further, since $Y = 0$ iff $X = 0$, p and p^* can be compared under the binomial assumptions by the relation $p^*(t) = 1 - (1 - p(t))^6$. Figure 4 gives a side-by-side comparison of the binary and binomial models using the linear logistic function (3.2). Specifically, the figure shows the expected number of O-ring incidents as a function of temperature for the two cases. The two fits compare quite closely, giving more confidence in the binomial-logistic model (3.2). The figure suggests that for 31°F about four to five of the six O-rings will be damaged.

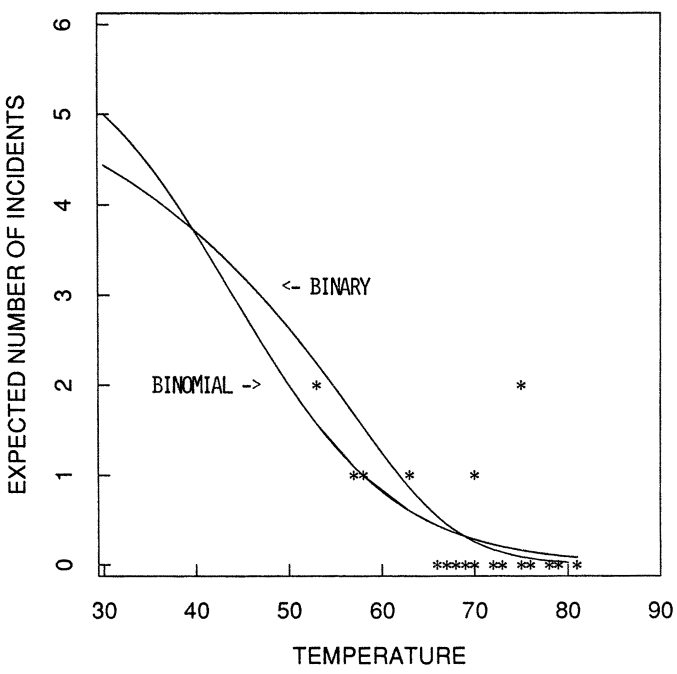

Figure 4. O-Ring Thermal-Distress Data: Field-Joint Primary O-Rings, Binomial-Logit Model, and Binary-Logit Model.

3.2 Confidence Intervals for Quantities Computed From Models

To assess the reliability of the results, we construct confidence intervals for the estimated parameters of the model and the probabilities, given the model fit.

Instead of using the asymptotic theory to construct confidence intervals, we use the parametric bootstrap procedure of Efron (1979). The key idea is to take the estimated $\hat{\alpha}_0$ and $\hat{\beta}_0$ as fixed, to generate random samples repeatedly with replacement from the corresponding binomial linear logistic model, and to use the new estimated β values to form the bootstrap distribution of $\hat{\beta}$. The confidence interval is taken as percentage points of this bootstrap distribution. An analogous procedure can be carried out for estimating the expected number of incidents of thermal distress. This procedure is slightly different from the basic bootstrap procedure in that the sample is taken from the fitted model rather than from the original data.

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

The procedure tends to reduce variability but is highly dependent on the model being of the correct form. Figure 5 gives a plot of the fitted-model 90% bootstrap intervals for the expected number of incidents as a function of temperature. The asterisks denote the estimated expected values, $m_i(t)$, and the lower and upper dots denote (respectively) the 5% and 95% points of the bootstrap distribution. Notice that the intervals are wide for temperatures less than 65°F and short for temperatures greater than 65°F. This should be expected, because most of the data have temperatures greater than 65°F. The estimated interval for temperature equal to 30°F is about (1, 6); this illustrates that we have high variability for the expected number of incidents.

3.3 The Effect of Data Perturbations

In this section we investigate the sensitivity of the chosen model to individual data points. For this we eliminate each data point in turn and estimate the model for each of the 23 samples of size 22. Each of these estimates is compared to the overall parameter estimate. Specifically, the following statistic is calculated: $\delta_i = (\hat{\beta}(\text{all}) - \hat{\beta}(-i))/S(\hat{\beta})$, where $\hat{\beta}(\text{all})$ is $\hat{\beta}$ when all data are used, $\hat{\beta}(-i)$ is the estimate of $\hat{\beta}$ with observation i removed, and $S(\hat{\beta})$ is the standard deviation of the 23 values, $\hat{\beta}(-i)$. δ_i's are plotted versus the observation number. Figure 6 shows a solid-line plot of δ_i versus i for Model (3.2). The plot shows a very dramatic result for the 21st flight. The standardized change in the $\hat{\beta}$ values is about four standard deviations when point 21 is removed. This corresponds to point (75, 2) from Figure 4. Further bootstrap experiments found that this event had a probability under the model of less than .001 from chance alone. Other large but not extreme values occur at time points 2, 11, and 14. Each of the large values occurs at a nonzero count of incidents.

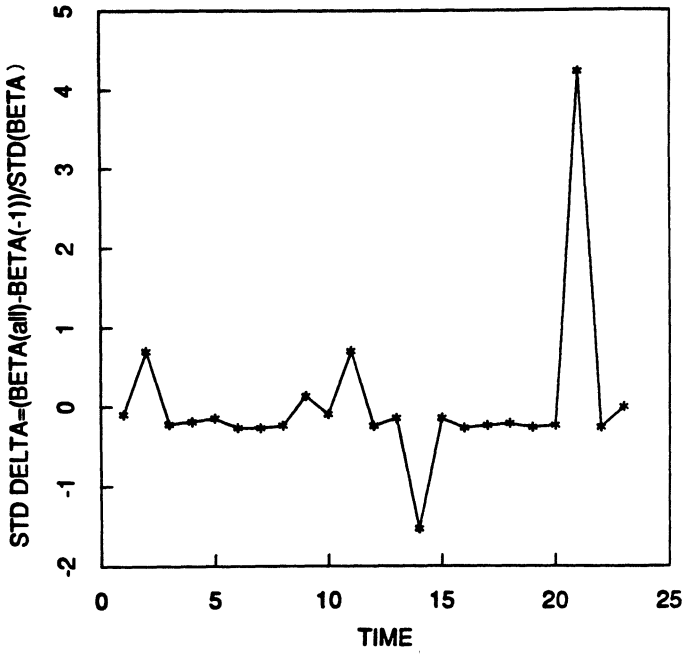

Figure 6. Delta Beta Plot and Binomial-Logit Model: N = 6.

Having identified an extreme point, we now investigate that point's influence. A direct method is to compare the probabilities of at least one O-ring failure when the point in question is omitted and retained. The binomial model was employed in the fitting process. Figure 7 gives a plot of temperature versus the probability of at least one incident without point 21 and a plot of the same probability when point 21 is included. The two sets of probabilities are close, though the relationship is stronger without point 21. After close study of the data we concluded that point 21, consisting of two blowby incidents, does not provide strong evidence that the fitted model is implausible and does not change our inference.

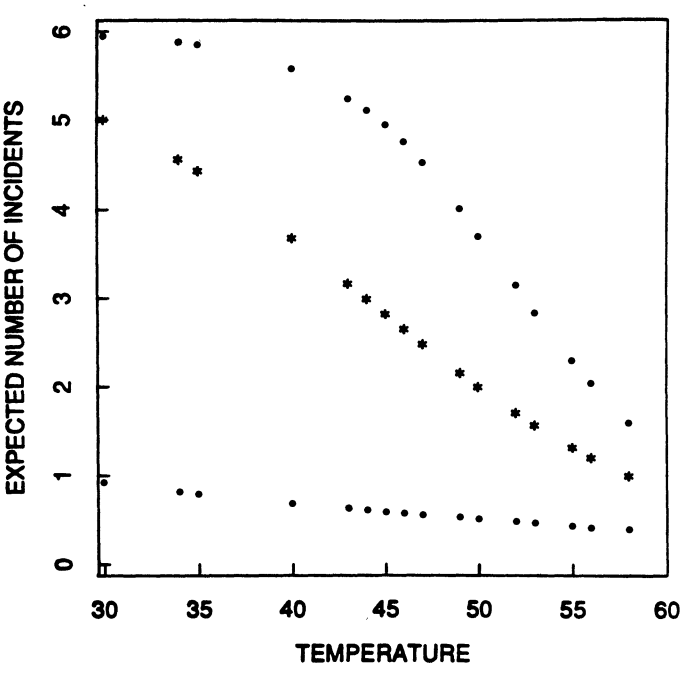

Figure 5. 90% Bootstrap Confidence Limits, Field-Joint Distress: Primary O-Rings and Binomial-Logit Model: N = 6.

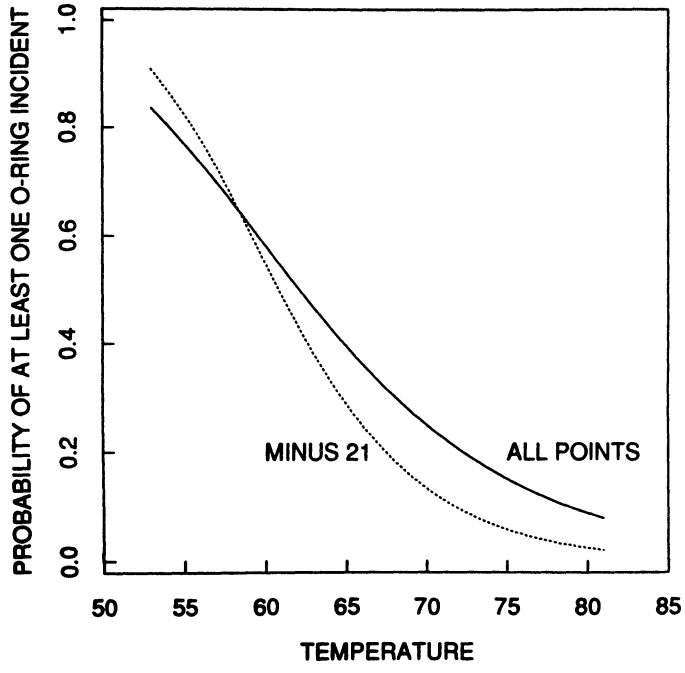

Figure 7. Binomial-Logit Model Based on All Data After Removing Observation 21.

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

3.4 The Form of the Model

We have concentrated on a logit model of the probabilities of the binomial distribution and assumed that it is linearly related to the covariates. Usually, there has been only one covariate, temperature. The logit transformation has numerous properties that make it appropriate for binary data. Other choices include the probit and the complementary log log [see McCullagh and Nelder (1983) for a further discussion of these transformations]. Another crucial question is whether the transform is linearly related to the covariates. A polynomial or some general nonlinear relationship is certainly a possibility. Although we desire to explore such choices, we would like to keep any selected models reasonably simple, since there are only 23 data points [recall that according to the lack-of-fit criterion G^2 the simple linear logit model (3.2) was adequate].

For checking linearity we added a squared term in temperature to the linear model. Specifically, we fitted the model $\log[p(t)/(1 - p(t))] = \alpha + \beta(t - \bar{t}) + \gamma(t - \bar{t})^2$. The coefficients of this fit were found to be $\hat{\alpha} = -3.1156$, $\hat{\beta} = -.0748$, and $\hat{\gamma} = .0041$. The likelihood ratio statistic for the linear model was $G_1^2 = 18.086$ with residual degrees of freedom equal to 21, whereas for the quadratic model $G_2^2 = 17.592$. For testing the significance, notice that $\delta G^2 = G_1^2 - G_2^2 = .494$ with 1 df. If we relate δG^2 to the upper 95% point of a chi-squared distribution, $\chi^2_{.95,1} = 3.84$, we find the curvature term highly nonsignificant.

A more general way of exploring the nonlinearity is to carry out diagnostics on a nonparametric estimate of the relationship between probability and temperature. For this we smoothed the data estimates of probability versus temperature using a method devised by Cleveland and Devlin (1988). Their method requires the specification of a fraction of the data, f, to be used in constructing smoothed values using a moving window. We tested values of f in the range .1(.1).9. The choice of f is basically a trade-off between bias and variance. If f is too small, there is excessive variability in the smoothed values. If f is too large, the smooth becomes a very biased estimate of the true relationship. If one has to err in the choice, it is the opinion of the authors that it is better to take f too small than too large. With this criterion as a guide, after some experimentation we selected $f = .4$ as a compromise choice. As a consequence of the trade-off, we are forced to accept some smoothed values that are 0. The corresponding smoothed values are given in Figure 8, where the binomial response is plotted for comparison. The linear logit fit again looks good.

The investigation of nonlinearity can be performed differently using the smoothed values and the techniques developed and detailed in Fowlkes (1987). Figure 9 shows a plot of the approximately standardized residual $\chi_i^s = (\hat{p}_i^s - \hat{p}_i)/\hat{\sigma}(\hat{p}_i^s)$ versus temperature, where $\hat{p}_i^s$ is the smoothed estimate of probability and $\hat{p}_i$ is the maximum likelihood estimate from Model (3.2). There are two far outliers in the plot. These correspond to point 9, for which the smoothed value is 0, and point 21, which Figure 7 shows to be a highly influential point for Model (3.2).

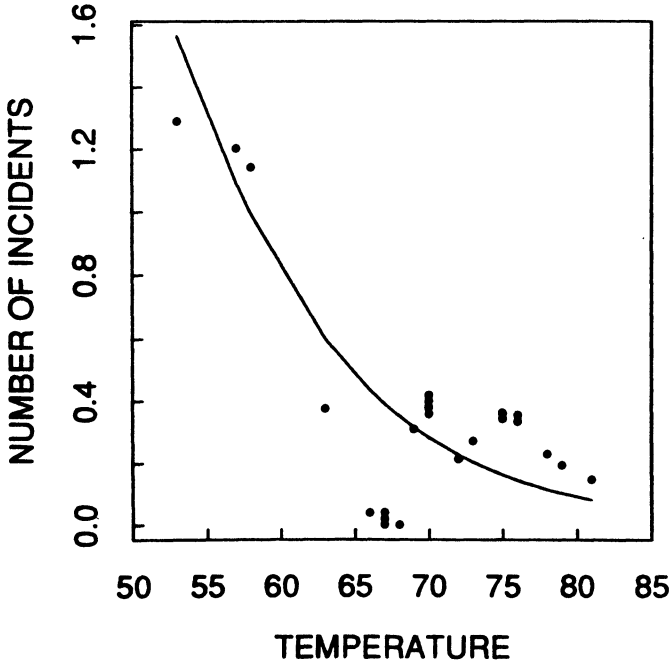

Figure 8. Smooth Values and the Binomial-Logit Model. The frequency is .4.

Aside from these outliers, there appears to be no relationship between temperature and the standardized residual. This suggests that there is no quadratic or other nonlinear relationship between the logit of the probability and temperature. A further indication of the adequacy of (3.2) can be seen in Figure 10, which shows a local deviance plot. Such a plot was originally introduced by Landwehr, Pregibon, and Shoemaker (1984). We use a modification of Fowlkes (1987). It is constructed by calculating a local estimate of deviance based on the smoothed values for each of the 23 points (see Fowlkes

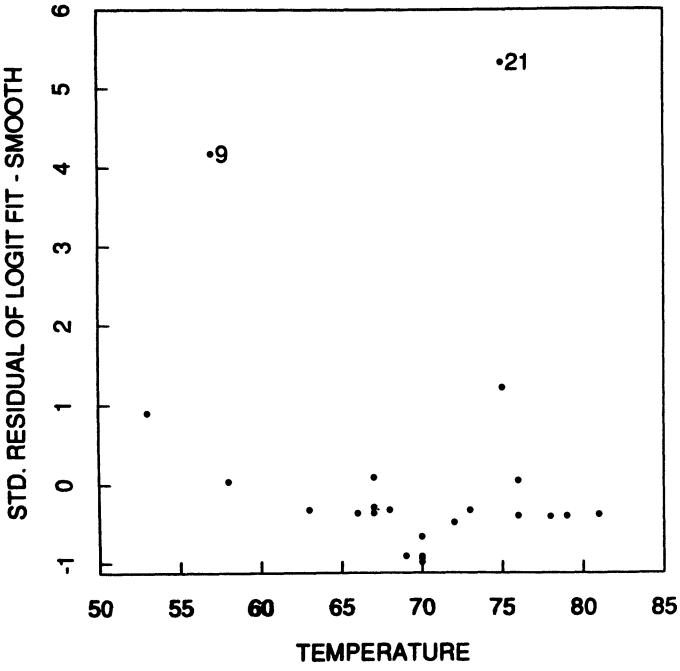

Figure 9. Standardized Residual of (Logit fit − Smooth fit): Smoothing Parameter, f = .4.

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

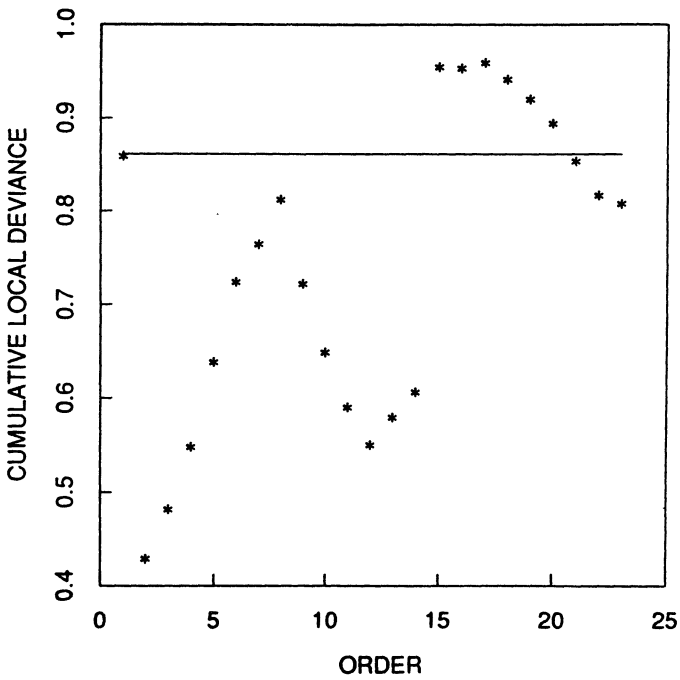

Figure 10. Cumulative Local Deviance Plot: Window Proportion, f = .4.

1987). These deviances are then ordered according to a measure of size of the smoothing region, cumulated, and normalized by an estimate of cumulative degrees of freedom. In addition, a horizontal line is plotted for the global deviance (G^2) of Model (3.2). Point configurations that level off to the right well below the global deviance give evidence of significant lack of fit of the chosen model. Such is not the case here; the point configuration tends to approach and remain near the global deviance. There is an additional peculiarity in the plot that should be considered. A large gap in the local deviances occurs between the 14th and 15th point. This is again caused by point 21, which has extremely high leverage. If this point is removed and plot reconstructed, the cumulative local deviance still tends to approach the global deviance as the order becomes large (both deviances are divided by cumulative degrees of freedom).

Finally, the linearity of the relationship between the logit transformation of the probability and temperature can be investigated using the alternating conditional expectations (ACE) algorithm of Breiman and Friedman (1985). To accomplish this, the logit transformation of the smoothed estimates of probability are regressed on temperature using appropriate weights (see Cox 1977). The regression was carried out using ACE; the algorithm returns nonparametric estimates of the transformation of the response as well as the explanatory variables. The corresponding plot of the estimated transformation versus the logit of the estimated probabilities was quite linear. This is strong evidence of the appropriateness of the logit transformation.

Investigation of the model containing temperature and pressure was also considered. Many of the diagnostic procedures used for the model containing temperature alone could not be carried out for several reasons. Smoothing of the observed counts of thermal incidents according to both temperature and pressure was deemed unwise because of the highly discrete nature of the pressure variable. In addition, the linear equations involved in the Newton–Raphson algorithm became singular when observation 2 was omitted. Observation 2 was the only observation having pressure equal to 50 that had a thermal incident. The limited number of diagnostics that could be made suggested that Model (3.2) was adequate.

4. ANALYSIS OF NOZZLE–JOINT DATA

From the sources quoted in the Rogers Commission Report and from other engineering sources, it is clear that in certain aspects the nozzle joints and the field joints are different. For example, there is an opinion that a nozzle joint's failure is not as catastrophic as a field joint's failure, and that it is less likely to fail, since it is not as severely affected by the joint rotation. To examine these issues, we created a data base of the number of nozzle-joint heat incidents on the primary O-rings, along with the corresponding temperature and pressure. These are given in Table 1. Recall that there are two nozzle joints per shuttle, and that by a heat incident we mean that there is some erosion or blowby on the primary O-ring.

For this data set we fitted a binomial regression model, with the number of heat incidents a response variable in which the temperature and the pressure entered linearly. The fit was poor. As discussed in Section 3, the binomial fit can be seriously affected by outliers, as well as by a lack of independence between the heat incidents in a given flight. To guard against these, we decided to fit a binary model of the type described in Section 3, where the response variable is 1 or 0, depending on whether there was at least one nozzle-joint heat incident. The logistic model using the pressure and temperature linearly fit well with

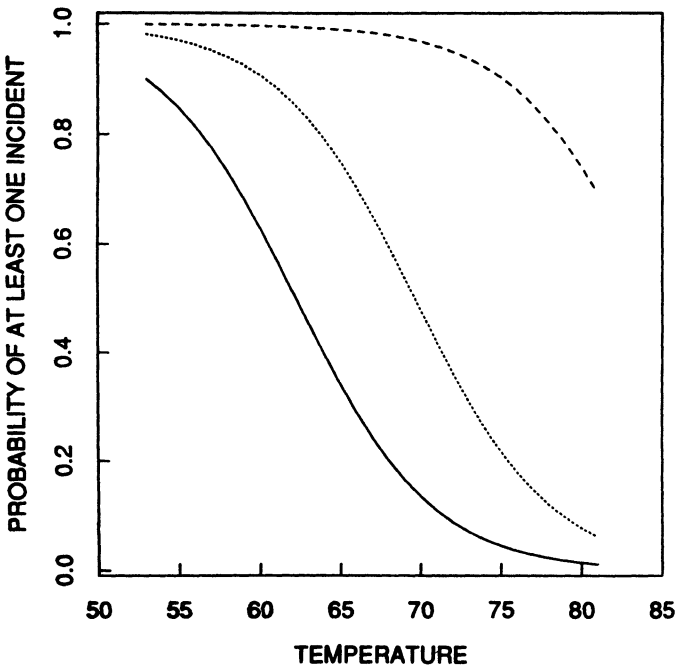

Figure 11. Nozzle-Joint Binary-Logit Model: —, 50 psi; ···, 100 psi; – –, 200 psi.

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

$G^2 = 18$ based on 20 df. The pressure and the temperature variables were both important, as evidenced by the changes in G^2 from the partial models to the full model. The fitted equation was $\log[p^N(t, s)/(1 - p^N(t, s))] = 13.08 - .238t + .35s$, where $p^N(t, s)$ denotes the probability of at least one nozzle-joint heat incident.

It is interesting to note that the temperature effect in this model is similar to that given in Section 3 for field joints. Unlike that model, however, the pressure effect, besides being significant, is more important than the temperature effect.

To understand the magnitude of these effects, in Figure 11 we have plotted the estimated probability of at least one heat incident on the primary nozzle O-ring against temperature for varying pressures. From the plot it is obvious that primary nozzle O-rings at 100 and 200 psi are far more prone to heat damage than the corresponding field O-rings. In fact, since the time of our analysis, in field tests of nozzle joints conducted by Morton Thiokol there have been some serious incidents that may delay the shuttle relaunching program.

5. PROBABILISTIC RISK ASSESSMENT

5.1 Introduction

The approach outlined in Section 3 addressed the central issue of the temperature effect in the context of the data set that was prominently discussed during the prelaunch teleconference described in Section 1. This data, consisting of the number of field-joint primary O-ring erosion or blowby incidents, indicated a strong temperature effect that was quantified by a logistic regression model. A battery of diagnostic checks confirmed the appropriateness of the logistic model. Thus this analysis would indicate that the decision makers were wrong in their conclusion that there was no temperature effect and that quantitative assessments of the effects of changing temperature could not be made. In fact, their launch/no launch decision should have been strongly influenced by our assessment (from Fig. 4) that five of the six primary O-rings would have been expected to suffer erosion or blowby damage at 31°F. Nevertheless, this analysis cannot answer the question directly relevant to the decision makers—What is the probability of a catastrophic failure?—since a primary O-ring failure does not imply total failure.

This section develops a model for assessing the probability of a complete field-joint failure, based on the information available prior to the *Challenger* launch, and comes up with an explicit estimate of the probability of shuttle failure under the *Challenger* launch scenario on January 27, 1986. Admittedly, this is "water over the dam," but it does point out the important role statistical science can play at NASA in avoiding such future mishaps. Since there was no total field-joint failure before this, we use available partial-failure data and engineering knowledge to create a sequence of conditional models. These are partly validated or estimated from the various sources of data in Section 5.2.

For estimating some of the conditional models, the data are scanty, and hence uncertainty about the estimates derived for some of the models is somewhat higher than in the analysis of Section 3. To incorporate this uncertainty as well as to provide a quick way of appending additional information, we carry out a Bayesian analysis and derive an approximate posterior distribution of the probability of a field-joint failure.

5.2 Probability Risk Assessment: Probability of a Mission Failure Via Field Joints

Consider the failure mode of a complete failure of a particular field joint. From engineering considerations (see Sec. 1) it is known that three events accompany a total failure: (a) primary O-ring erosion, (b) primary O-ring blowby, and (c) secondary O-ring erosion. The final event of the total failure is (d) secondary O-ring failure. (Note that in Sec. 3 we modeled the probability of the event $a \cup b$.) Let p_a, p_b, p_c, and p_d denote the corresponding probabilities of events (a)–(d), conditional on all of the preceding events. These conditional probabilities are modeled as functions of explanatory variables such as temperature, pressure, and so forth. The probability of the complete failure of a joint, p_F, is equal to the product $p_a \cdot p_b \cdot p_c \cdot p_d$. Assuming that joint failures are independent and identically distributed, the probability of at least one field-joint failure is $p_{FF} = 1 - (1 - p_F)^6$. This is reasonable because the binary and binomial models for the field joints give similar probabilities in Section 3. Since there were several other failure modes on the shuttle criticality list (failure modes that could lead to catastrophic failure) that were of concern, the probability of catastrophic shuttle failure could be much greater than the probability of at least one field-joint failure. The principal objective in this section is to assess p_{FF}. Much of the data for this analysis come from Table 1.

We now estimate p_a, p_b, p_c, and p_d. Figure 12 gives the raw data for the number of eroded primary rings, coded by pressure as a function of the temperature. The binomial-logistic model, linear in pressure and temperature, gave a good fit. The fitted equation was $\log[p(t, s)/(1 - p(t, s))] = 7.8 - .17t + .0024s$. The parametric 90% bootstrap confidence intervals for the coefficients were, respectively, $(-.1, 15.7)$, $(-.28, -.06)$, $(-.012, .016)$. From these results, the pressure variable is clearly not significant. We also applied some of the diagnostic procedures outlined in Section 3. Even though the pressure was found to be statistically nonsignificant, we decided to use the logistic regression model, based on temperature and pressure for prediction purposes. This is because we wanted to use the best linear predictor based on all available covariates that NASA engineers thought were important. On the basis of this model, the predicted probability of erosion per field joint at 31°F and 200 psi (the conditions under which the *Challenger* was launched), p_a, is .95. For assessing variability, 90% parametric bootstrap intervals for the regression curve are given in Figure 13. Transforming the results of Figure 13 from the number-of-incidents scale to the probability scale indicates that at 31°F and 200 psi a 90% confidence interval for p_a is $(\frac{1}{2}, 1)$.

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

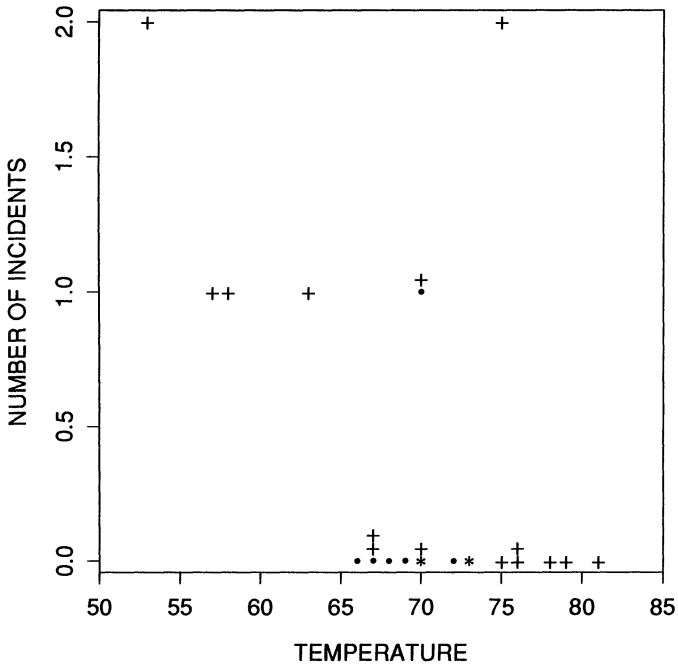

*Figure 12. Number of Eroded Primary O-Rings Versus Temperature for Field Joints (N = 6), Coded by Pressure: ·, 50 psi; *, 100 psi; +, 200 psi.*

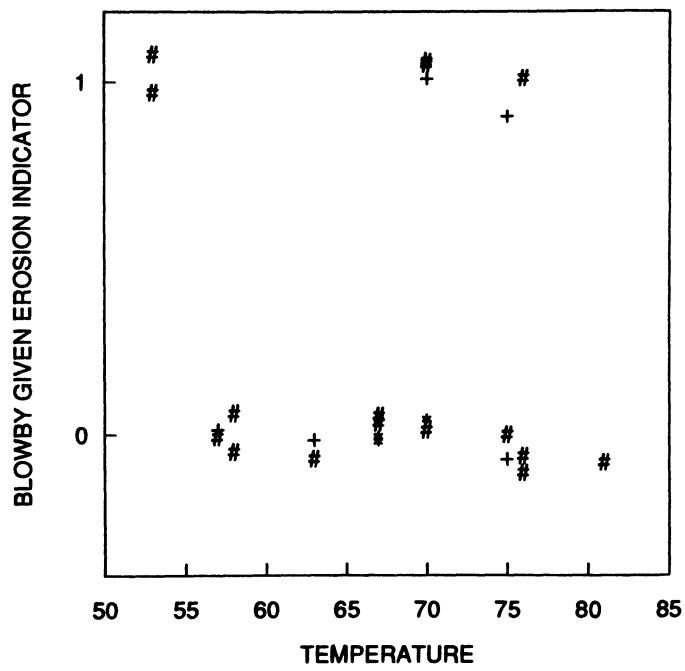

*Figure 14. Indicator of Field- and Nozzle-Joint Primary O-Rings Having Blowby Versus Temperature, Conditional on Having Erosion, Coded by Pressure: *, 50 psi; +, 100 psi; #, 200 psi.*

Now, consider the conditional probability, p_b, of a blowby, given an erosion. Of the seven eroded primary field joints, two had blowby. Since there were so little data, we pooled the field-joint data with the data for blowby, conditional on erosion for nozzle joints (17 incidents). In Figure 14 the combined data, coded by pressure, is plotted versus temperature. No obvious relationship is seen; a formal analysis using the linear logistic model gave the same conclusion. Furthermore, the relative frequency of blowby, conditional on erosion for the field joints, is 2/7 = .286, and that for the nozzle joints is 5/17 = .294. These values are very similar; this provides another rationale for pooling the data. Thus we estimated p_b from the pooled data as $\hat{p}_b = 7/24 = .292$ (independent of any covariates).

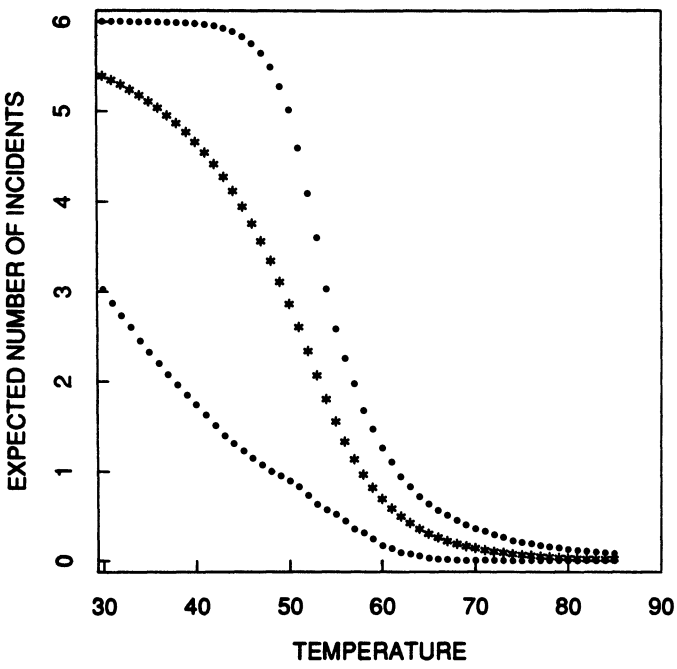

Figure 13. Binomial-Logit Model (Temperature and Pressure). Pressure: 200 psi.

Now, consider p_c, the probability of erosion in the secondary O-ring, given blowby and erosion in the corresponding primary O-ring. The total numbers of incidents involving primary erosion and blowby for the field as well as for the nozzle joints were two and five, respectively. Of these, one of each type of joint suffered a secondary erosion. The corresponding plot of the number of incidents versus temperature for the pooled data is given in Figure 15. There does not seem to be any temperature or pressure effect. Nevertheless, the relative frequencies for these two joints are different. For the field it is $\frac{1}{2} = .50$, and for the nozzle it is $\frac{1}{5} = .20$. This difference might be due to joint rotation, because the nozzle joint is designed differently and does not experience joint rotation. Since there are very little data and no statistically significant difference, however, we disregard this information, act optimistically, and estimate p_c by the pooled probability, 2/7 = .286.

Finally, p_d, the probability of a second O-ring failure, given the primary erosion, blowby, and the secondary erosion, was calculated. We quantify engineering judgment probabilistically to complete the analysis, since this never happened before.

The engineering knowledge of secondary O-ring reliability is summarized in Section 2. Two significant items are the following.

1. The secondary O-ring was not considered a redundant part because of joint rotation. In NASA's official critical failure-mode rationales for retention (PC1), the

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

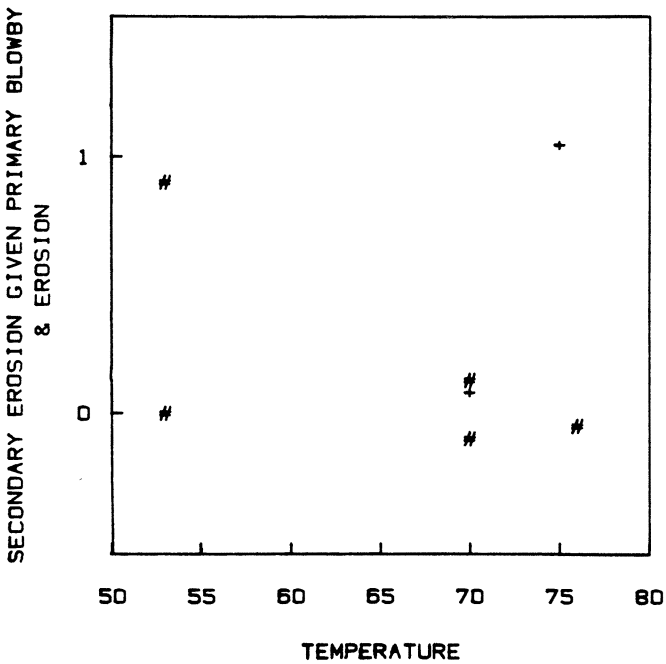

Figure 15. Indicator of Field- and Nozzle-Joint Secondary O-Rings Having Erosion Versus Temperature, Conditional on Primary O-Rings Having Erosion and Blowby, Coded by Pressure: +, 100 psi; #, 200 psi.

secondary O-ring was considered useful only for conducting the leak check on the primary O-ring.

2. Beyond 330 milliseconds after ignition, the secondary O-ring was considered to have a high probability of failure, given erosion penetration of this secondary O-ring.

We quantify this with the assumption that $p_d \geq p_b$, and consequently $\hat{p}_d \geq \hat{p}_b = .292$. The logic is that once blowby starts on the secondary O-ring, it will fail. With respect to blowby, given erosion, the secondary O-ring is less reliable than the primary O-ring.

Under the aforementioned assumptions, our estimate of the probability of a complete field-joint failure is $\geqq \hat{p}_a \cdot \hat{p}_b^2 \cdot \hat{p}_c = .023$ at 31°F and 200 psi. Thus the estimated probability of at least one complete joint failure at 31°F and 200 psi is $\geq 1 - (1 - .023)^6 = .13$. One may question whether this number is too high or whether it reflects an acceptably low risk for a shuttle mission. We cannot say. Nevertheless, to shed more light on the matter, one may ask a slightly different question. Suppose that the mission was postponed until the temperature went up to 60°F: What would have been the risk? Carrying out the previous analysis at 60°F with 200 psi leak-check pressure, the probability of at least one complete joint failure would have been $\geq .019$. The difference between these probabilities is quite large; the risk at 31°F is 600% higher than that at 60°F. This way of looking at the situation may have made the launch-delay option look very attractive.

At this juncture, one may try to see the extent to which the optimistic assumption about p_c, where we pooled the field- and nozzle-joint data, influences our estimates. Suppose that instead we had estimated p_c to be $\frac{1}{2}$ on the basis of the field-joint data alone. In that case, by carrying out the previous calculations we get $P_{FF}(31) \geq .218$ and $P_{FF}(60) \geq .032$, again giving rise to a 600% increase in risk.

5.3 Uncertainty Assessment

In the previous section we derived point estimates for various conditional probabilities of events making up the field-joint failure scenario. Obviously, at several steps there were very little data. Thus we would like to quantify the uncertainty involved in our estimates. Furthermore, we present our results so that any additional information can be appended. Finally, in our interactions with engineers we have found that they have a better understanding of uncertainty about parameter values than of sampling

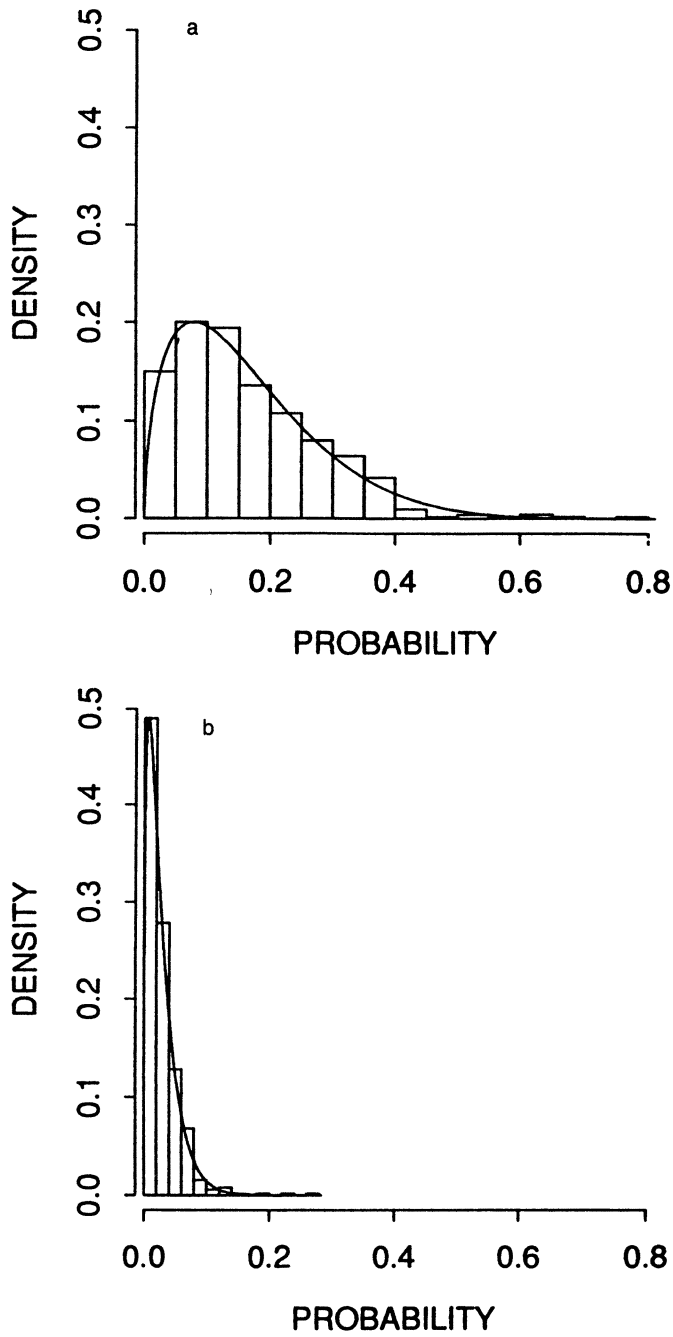

Figure 16. Histogram and Density Estimate of the Probability of a Shuttle Failure Via Field Joint at (a) 31°F and (b) 60°F. Pressure: 200 psi.

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

distributions. For all of these reasons, we decided to cast the problem of determining the uncertainty into a Bayesian setting. We express our findings in terms of posterior distributions of parameters by using Bayesian bootstrap ideas (without explicitly carrying out a detailed Bayesian analysis).

We assume that the parameters of interest are p_a, p_b, p_c, and p_d, and functions of their products. Under the assumptions made in the previous section, we derive the posterior distributions of p_a, p_b, p_c, and p_d. The arguments leading to this derivation are approximate; no attempt is made to make these more precise.

Since p_a is approximately a location parameter for $\hat{p}_a$, it can be shown that the posterior distribution p_a is approximately equal to the sampling distribution of $\hat{p}_a$, given p_a. Furthermore, the sampling distribution of $\hat{p}_a$ is approximately equal to the bootstrap distribution of $\hat{p}_a$, and thus it follows that the posterior distribution of p_a is approximately equal to the bootstrap distribution of $\hat{p}_a$. This kind of argument was used in a nonparametric bootstrap context by Rubin (1981) and made more precise by Lo (1985).

Since p_b does not depend on the covariates, we assume that x_b, the number of primary joints with blowbys, conditional on primary erosion, is binomial (n_b, p_b), where n_b is the number of eroded primary joints. Taking the uniform distribution on [0, 1] as the noninformative prior for p_b, it follows that the conditional distribution, $p_b \mid x_b, n_b$, is beta(α, β), with $\alpha = x_b + 1$ and $\beta = n_b + 1 - x_b$. A similar analysis gives another beta distribution for p_c. Since we are in effect assuming $p_d \equiv p_b$, the posterior of p_d is the same as that of p_b.

Now, for the rest of the analysis of the probability of a joint failure, we do not differentiate between $p_F = p_a \cdot p_b \cdot p_c \cdot p_d$ and its lower bound $p_a \cdot p_b^2 \cdot p_c$. Thus the posterior distribution of the probability of the failure of a field joint (p_F) is obtained through $p_F = p_a \cdot p_b^2 \cdot p_c$. Although we could obtain an analytic expression for the distribution of $p_b^2 \cdot p_c$, it is complicated, and it would still not give an analytical expression for p_F. Instead, we simulated the distributions of p_a, p_b, and p_c. For p_a we used the 500 observations obtained from the earlier bootstrap experiment, and for p_b and p_c we drew 500 observations from the appropriate beta distributions. These samples were used to generate the samples from the posterior of $p_a \cdot$

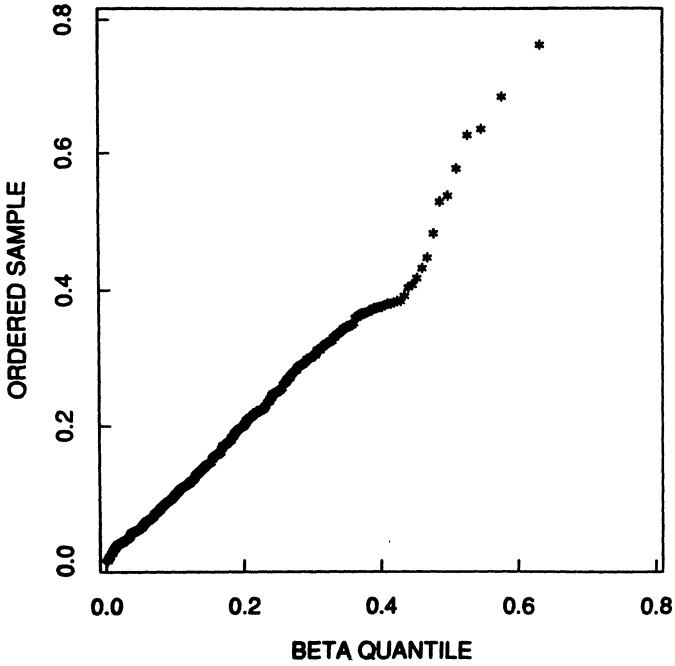

Figure 17. Q–Q Plot: Empirical Quantiles Versus Beta Quantiles for the Probability of Shuttle Failure Via Field Joint at 31°F. α = 1.598; β = 8.068.

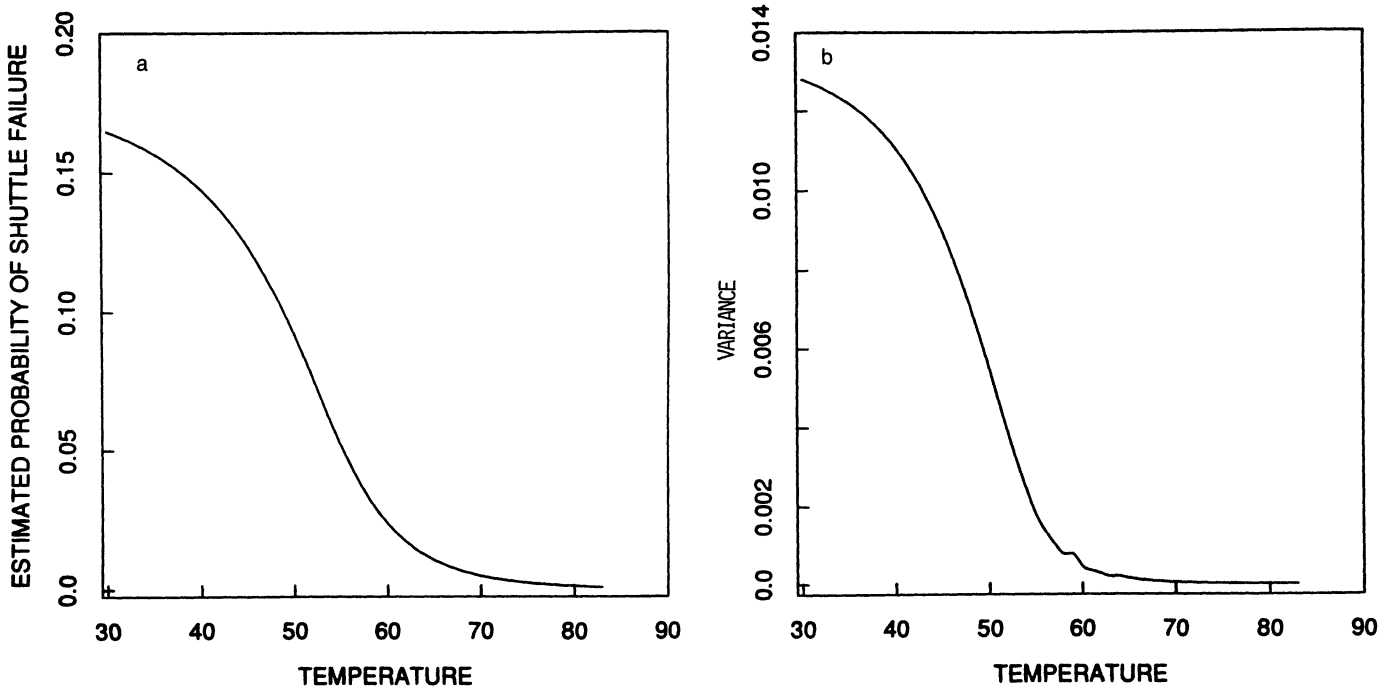

Figure 18. (a) Means and (b) Variances of Fitted Beta Distributions Versus Temperature.

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms

$p_b^2 \cdot p_c$ and hence that from the posterior of p_{FF}. In Figure 16a, based on the aforementioned 500 derived observations we have plotted the histogram of the posterior of the probability of total failure via field joints: $p_{FF}(31)$, for 31°F at 200 psi. The median of this distribution is .1378, and the 90% posterior interval is (.0275, .3704). To understand the significance of these numbers, a similar analysis was carried out for 60°F at 200 psi. The posterior of $P_{FF}(60)$ is plotted in Figure 16b. The corresponding median is .0205 and the 90% posterior interval is (.0031, .0764). We again see a marked difference between the risks. One may wish to use the posterior means rather than the medians as estimates, since they are "gambling probabilities." The means for 31°F and 60°F are .1641 and .0048, respectively.

For convenience in analytically adding new information, we decided to examine the extent to which posterior distributions of p_{FF}, as functions of the varying joint temperature, may be approximated by beta distributions. Throughout this analysis we held the leak-check pressure constant at 200 psi, the pressure used for the last 16 flights. For fitting beta models to $P_{FF}(t)$, we used the maximum likelihood algorithm of Gnanadesikan, Pinkham, and Hughes (1967). To assess the adequacy of the beta fit, in Figure 17 we constructed Q–Q plots of the empirical posterior distributions against the corresponding fitted beta distributions at 31°F. This plot, based on 500 observations, shows a reasonably linear configuration, except in the tails. A plot for 60°F was similar. The corresponding histograms, superimposed with the fitted beta densities, are given in Figure 16. Since these fits are reasonably good, we fitted betas for all temperatures in the range 30°(1)82°F. The corresponding means ($\alpha/\alpha + \beta$) and variances ($\alpha\beta/(\alpha + \beta + 1)$) are plotted against temperature in Figure 18. If necessary, one can easily obtain the corresponding α and β, solving the moment equations for the mean and the variance.

[*Received September 1988. Revised May 1989.*]

REFERENCES

Breiman, L., and Friedman, J. H. (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation" (with discussion), *Journal of the American Statistical Association,* 80, 580–619.

Cleveland, W. S., and Devlin, S. J. (1988), "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association,* 83, 596–610.

Cox, D. R. (1977), *Analysis of Binary Data,* London: Chapman & Hall.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics,* 7, 1–26.

Fowlkes, E. B. (1987), "Some Diagnostics for Binary Logistic Regression Via Smoothing," *Biometrika,* 75, 503–515.

Gnanadesikan, R., Pinkham, R. S., and Hughes, L. P. (1967), "Maximum Likelihood Estimation of the Parameters of the Beta Distribution From Smallest Order Statistics," *Technometrics,* 9, 607–620.

Landwehr, J. M., Pregibon, D., and Shoemaker, A. C. (1984), "Graphical Methods for Assessing Logistic Regression Models," *Journal of the American Statistical Association,* 79, 61–83.

Lo, A. K. (1985), "A Large Sample Study of the Bayesian Bootstrap," *The Annals of Statistics,* 15, 360–375.

McCullagh, P., and Nelder, J. A. (1983), *Generalized Linear Models,* London: Chapman & Hall.

Presidential Commission on the Space Shuttle *Challenger* Accident (1986), *Report of the Presidential Commission on the Space Shuttle Challenger Accident* (Vols. 1 & 2), Washington, DC: Author.

Rubin, D. R. (1981), "The Bayesian Bootstrap," *The Annals of Statistics,* 9, 130–134.

Shuttle Criticality Review Hazard Analysis Audit Committee (1988), *Post-Challenger Evaluation of Space Shuttle Risk Assessment and Management,* Washington, DC: National Academy of Sciences Press.

This content downloaded from 142.58.11.141 on Fri, 26 Jan 2018 01:26:12 UTC
All use subject to http://about.jstor.org/terms