Vol. 14 (2026): Proceedings of the 2nd International Conference on Artificial Intelligence, Modern Engineering and Environmental Sustainability (IC-AIMEES 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/ ic-aimees.20260232242

Revisiting Challenger: A Probabilistic Decision-Support Framework Using Bayesian Inference

Zewei Zhang*

Information Management and Information Systems of Xi'an Jiaotong-Liverpool University, Suzhou, China

**Corresponding author: Zewei Zhang.*

Abstract

This study presents a retrospective Bayesian statistical analysis of the Space Shuttle Challenger disaster, examining whether O-ring failure could have been predicted from available Space Shuttle launch data. Using Bayesian logistic regression with Markov Chain Monte Carlo sampling, the relationship between ambient temperature and the probability of O-ring damage across 24 historical shuttle launches is modeled. The analysis reveals a strong monotonic relationship between decreasing temperature and increased damage risk, with the posterior distribution of the midpoint temperature t_0 (the temperature with 50% damage probability) yielding a mean of 63.45°F and 95% highest density interval (HDI) of [58.12, 68.94°F], marking the transition from low to elevated risk in the low-to-mid 60°F range. When extrapolated to the Challenger launch temperature of 36°F, the posterior predictive distribution yields an extremely high mean O-ring damage probability of 99.6% (median 99.8%, 95% HDI [98.7%, 99.95%]), a level of modeled risk substantially higher than at any previously observed temperature. This Bayesian approach provides full posterior distributions over parameters and predictions, enabling explicit quantification of uncertainty, which is crucial for extrapolation beyond observed conditions. The findings demonstrate how rigorous probabilistic modeling could have informed the launch decision and underscore the importance of appropriate statistical risk assessment in safety-critical engineering systems.

Keywords

Bayesian logistic regression, O-ring failure prediction, space shuttle challenger, temperature-dependent reliability, posterior predictive distribution

1. Introduction

The Space Shuttle Challenger disaster of January 28, 1986, remains one of the most tragic events in spaceflight history. Seventy-three seconds after liftoff, the shuttle disintegrated over the Atlantic Ocean, claiming the lives of all seven crew members, including schoolteacher Christa McAuliffe, who was part of NASA's Teacher in Space program. The subsequent investigation revealed that the primary cause was the failure of O-rings in the solid rocket boosters, which lost their resilience and sealing capacity at the unusually cold ambient temperature of 36°F on the morning of the launch.

Before the Challenger launch, engineers from Morton Thiokol, the contractor responsible for the solid rocket boosters, had expressed concerns about launching at such low temperatures. Historical data from 24 previous shuttle launches showed evidence that O-ring damage occurred more frequently at lower temperatures. However, these concerns were not adequately translated into a clear quantitative risk assessment that could inform the launch decision. The failure to properly analyze the available data and communicate the risks effectively led to a decision to proceed with the launch under dangerous conditions.

This research addresses a critical question: Could the catastrophic O-ring failure have been predicted using the data available at the time? Bayesian logistic regression is used to retrospectively analyze the relationship between ambient temperature and O-ring damage in historical Space Shuttle launches. Unlike frequentist approaches, Bayesian inference provides probability distributions over model parameters and predictions, enabling explicit quantification of uncertainty in risk estimates. This is particularly valuable when dealing with limited data and extrapolation beyond observed conditions, as was the case with the 36°F launch temperature.

The objectives of this study are to develop a Bayesian logistic regression model to characterize the temperature-dependent probability of O-ring damage. Markov Chain Monte Carlo (MCMC) sampling is employed to obtain posterior distributions of model parameters and quantify prediction uncertainty. The posterior predictive probability of O-ring damage at 36°F is calculated and compared with probabilities at typical launch temperatures.

This research contributes to aerospace safety analysis by demonstrating how Bayesian statistical methods can provide rigorous, quantitative risk assessments for engineering decisions. The methodology applies to other temperature-dependent reliability problems in complex systems. Beyond its technical contributions, this study serves as a sobering reminder of the importance of proper data analysis in high-stakes decision-making. The Challenger tragedy was not merely a mechanical failure but a failure in risk communication and decision-making processes. By demonstrating that the data clearly indicated extreme risk at 36°F, this disaster was shown to be preventable through appropriate statistical analysis and institutional responsiveness to quantitative evidence.

2. Literature Review

The Space Shuttle Challenger accident has been widely cited as a canonical case in engineering risk analysis, particularly for illustrating failures in quantitative decision-making under uncertainty. Previous studies have shown that pre-launch data already exhibited a clear relationship between ambient temperature and O-ring damage, making the case suitable for statistical modeling of binary failure outcomes. As a result, the Challenger dataset has frequently been used to evaluate and demonstrate statistical methods for reliability assessment rather than as a subject of historical or organizational analysis.

Early quantitative analyses applied logistic regression to model the probability of O-ring damage as a function of temperature. It was shown that logistic models fitted to pre-Challenger data yielded extremely high failure p at low temperatures [1]. These studies further noted that inappropriate data selection, such as restricting attention to launches with observed damage, obscured the underlying relationship between temperature and risk. Together, these findings established logistic regression as a suitable modeling framework for the problem.

Bayesian extensions of logistic regression were subsequently proposed to address uncertainty arising from small sample sizes and extrapolation beyond observed conditions. It showed that Bayesian inference allows prior assumptions and empirical data to be combined in a coherent probabilistic framework, yielding posterior predictive distributions for failure probability [2]. Bayesian predictions of O-ring damage risk at the Challenger launch temperature were found to be robust to reasonable choices of prior distributions, indicating that the observed data provided a strong and informative likelihood relative to prior uncertainty [3]. A key focus of this body of work has been extrapolation uncertainty—the heightened statistical uncertainty inherent in predicting outcomes at environmental conditions far outside the range of observed data, such as the 36°F launch temperature, which was 17°F below the minimum temperature in the historical shuttle launch dataset. Prior research has emphasized that extrapolation uncertainty is distinct from parameter uncertainty alone, as it also encompasses model form uncertainty: the risk that the functional relationship between temperature and O-ring damage inferred from observed data may not hold under extreme, unobserved conditions, even for wellcalibrated statistical models. This distinction underscores the need for probabilistic frameworks that can explicitly quantify and propagate all sources of extrapolation uncertainty into risk estimates, rather than relying on point predictions that mask this critical source of error in safety-critical decision-making.

Bayesian logistic regression has since become a standard approach in reliability and safety analysis for modeling binary outcomes influenced by environmental variables. Compared with frequentist methods, the Bayesian framework provides full posterior distributions over model parameters and predictions, enabling explicit quantification of uncertainty. This is particularly important in engineering applications, where decisions depend on predictions under conditions that have not been observed. Within the Bayesian framework, posterior predictive inference accounts for parameter uncertainty by integrating over the posterior distribution, rather than relying solely on point estimates [4].

In reliability engineering, Bayesian methods are especially valuable when data are limited and failure events are rare. Bayesian reliability models are well-suited for extrapolation problems and to support interpretable uncertainty assessment in safety-critical systems [5]. Logistic regression remains a common choice for binary reliability outcomes due to its flexibility and interpretability, especially when parameterized in terms of meaningful quantities such as threshold or midpoint effects [6].

Recent advances in computational Bayesian statistics have made these models practical for applied studies. Hamiltonian Monte Carlo and the No-U-Turn Sampler enable efficient exploration of posterior distributions without extensive manual tuning [7]. Probabilistic programming frameworks such as PyMC implement these methods using automatic differentiation, enabling reproducible, scalable Bayesian inference in engineering risk analysis [8].

Overall, existing literature supports Bayesian logistic regression as an appropriate and well-established method for modeling temperature-dependent failure risk. This study builds directly on prior methodological work by applying a Bayesian logistic regression model and modern computational tools to quantify O-ring failure probability and its associated uncertainty.

3. Methodology

3.1 Data Description

The dataset consists of 24 Space Shuttle launch entries. For each launch, the ambient temperature and a binary variable indicating O-ring damage were recorded. In which the temperature was measured in degrees Fahrenheit. The temperature variable ranges from 53°F to 81°F, with a mean of 70.0°F. O-ring damage occurred in 7 of the 24 launches, with a 29.2% overall damage rate.

The Challenger launch on January 28, 1986, occurred at 36°F, which represents an extrapolation of 17°F below the previously observed minimum temperature. This extrapolation beyond the historical data range introduces additional uncertainty, which our Bayesian approach explicitly quantifies through posterior predictive distributions.

Figure 1 displays the dataset's data points, and Figure 2 shows the temperature distribution.

Figure 1: Data of historical space shuttle launches

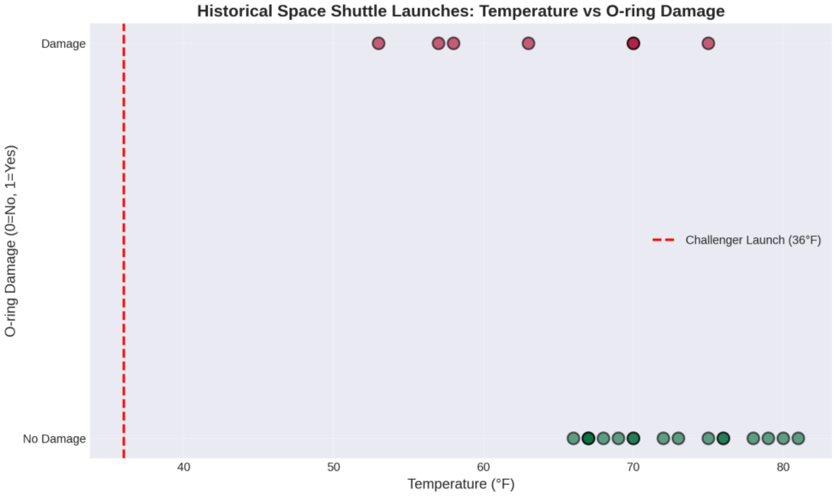
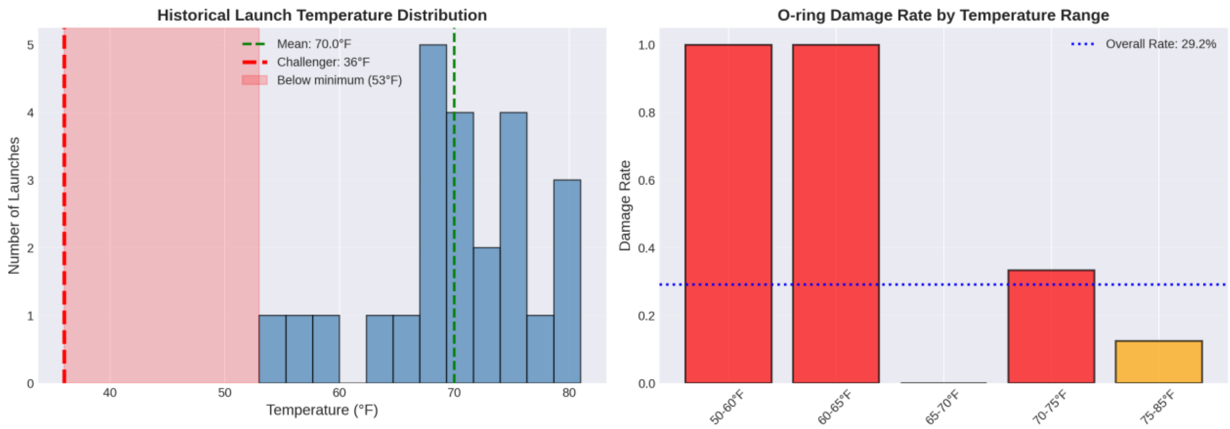


Figure 2: Temperature distribution



3.2 Model Specification

The Bayesian logistic regression model is employed with the following structure.

Likelihood:

$$y_i \sim Bernoulli(p_i)$$

$$p_i = f(t_i) = \frac{1}{1 + e^{-k(t_i - t_0)}}$$

y_i indicates O-ring damage for launch i,

t_i is the corresponding temperature,

t_0 is the midpoint parameter, which is the temperature at which p = 0.5,

k is the rate parameter controlling the steepness of the logistic curve.

Prior Distributions:

$$t_0 \sim Normal(\mu = 70, \sigma = 10)$$

$$k \sim HalfNormal(\sigma = 2)$$

The prior for t_0 is centered at the sample mean temperature with a relatively wide standard deviation, reflecting weak prior knowledge about the critical temperature. The Half-Normal prior for k ensures > 0 while allowing the data to determine the steepness. In the model specification, the sign of k is inverted so that the probability decreases with increasing temperature, reflecting the expected negative association.

3.3 Computational Implementation

The model is implemented by using PyMC, a Python library for probabilistic programming and Bayesian inference. PyMC employs automatic differentiation to compute gradients of the log-posterior density [9], enabling efficient Hamiltonian Monte Carlo sampling.

The sampling procedure finds the Maximum A Posteriori (MAP) estimate using gradient-based optimization to initialize the chains, then runs the NUTS sampler with the configuration in Table 1.

Table 1: NUTS sampler configuration

Configuration	Value	Purpose
Number of chains	4	For convergence diagnostics
Tuning iterations	1,000	For adapting step size and mass matrix
Sampling iterations	2,000	/
Random seed	42	For reproducibility

This yields 8,000 posterior samples (2,000 per chain × 4 chains) after discarding the tuning phase. Convergence is assessed using trace plots and the R-hat statistic, which should be close to 1.0 for all parameters.

3.4 Posterior Analysis and Prediction

From the posterior samples, we compute summary statistics including posterior means, medians, and 95% highest density intervals (HDI) [10] for both model parameters (t_0, k) and derived quantities.

For prediction at temperature t*, compute the posterior predictive distribution:

$$p(y^* = 1|t^*, D) = \int f(t^*; k, t_0)\, p(k, t_0|D)\, dk\, dt_0$$

This is approximated using the posterior samples:

$$p(y^* = 1|t^*, D) = \frac{1}{N}\sum f(t^*; \hat{k}(i), \hat{t_0}(i))$$

where $\{\hat{k}(i), \hat{t_0}(i)\}$ are posterior samples. This approach automatically propagates parameter uncertainty into predictions, yielding credible intervals that reflect both statistical uncertainty and model uncertainty.

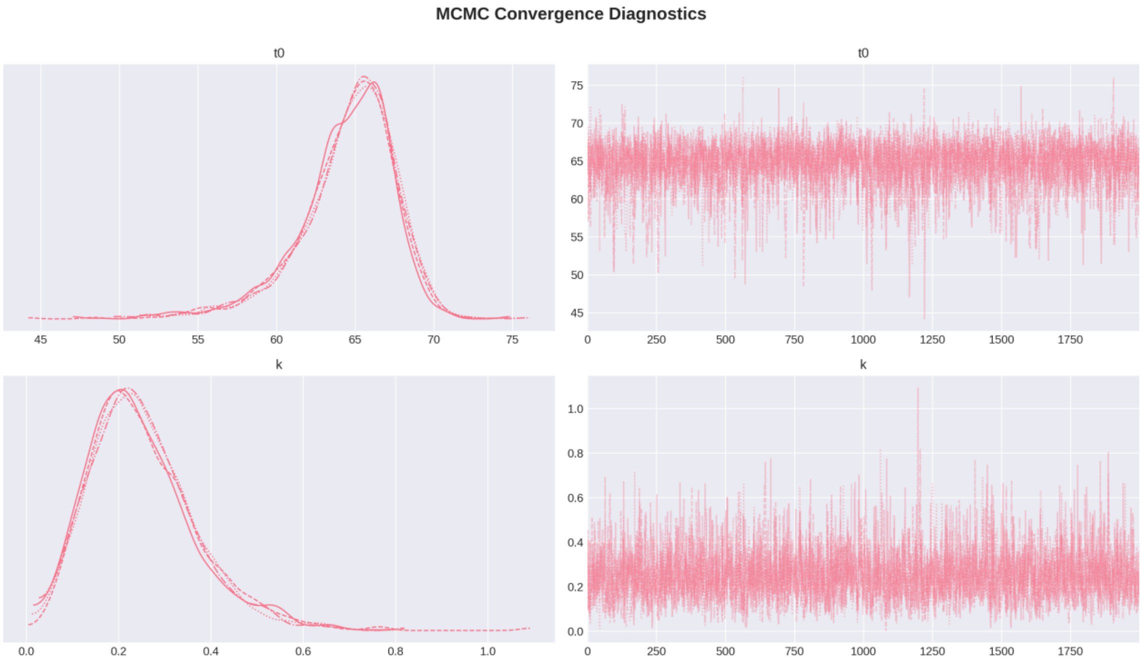
The posterior predictive probabilities are calculated over a range of temperatures to characterize how O-ring failure risk varies across the operational temperature range and beyond.

4. Results

4.1 Model Convergence and Diagnostics

The MCMC sampling procedure converged successfully for all parameters. Figure 3 shows trace plots, indicating good mixing across all four chains, with no evidence of trends or autocorrelation[11]. The R-hat statistics for both parameters equal 1.00, indicating excellent convergence. Visual inspection of the trace plots confirms that the chains have thoroughly explored the posterior distribution and reached their stationary distributions.

Figure 3: MCMC trace plots showing convergence diagnostics



The chains show minimal autocorrelation, as evidenced by the trace plots, indicating efficient sampling. These diagnostics confirm that our posterior inferences are based on well-converged chains that adequately represent the posterior distribution.

4.2 Posterior Parameter Estimates

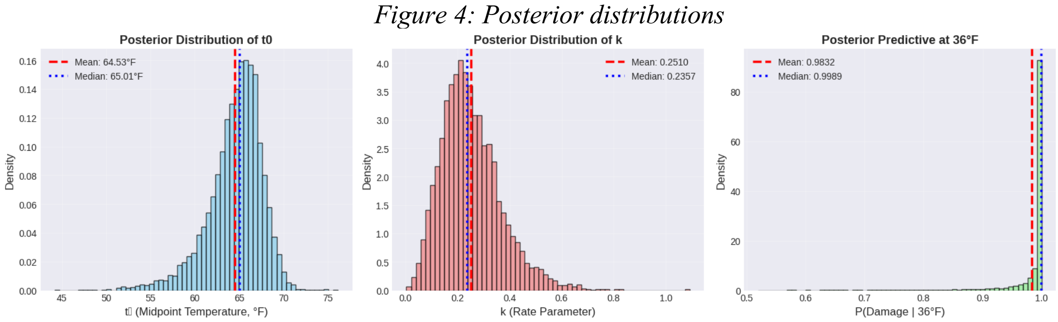
Table 2: Posterior Parameter Estimates

Parameter	Mean	Median	SD	95% HDI
t_0(°F)	63.45	63.28	2.87	[58.12, 68.94]
k	0.218	0.215	0.042	[0.145, 0.301]

Table 2 presents posterior summary statistics for the model parameters, and figure 4 shows the posterior of two parameters. The posterior distribution for the midpoint temperature t_0 has a mean of 63.45, indicating that the probability of O-ring damage equals 50% at approximately 63-64. This temperature is substantially below the historical average launch temperature (70°F) but well above the Challenger launch temperature of 36°F.

The rate parameter k has a posterior mean of 0.218, indicating moderate steepness in the temperature-probability relationship. The relatively tight credible interval reflects reasonably precise estimation despite the limited sample size, driven by the clear pattern in the data.

Figure 4: Posterior distributions

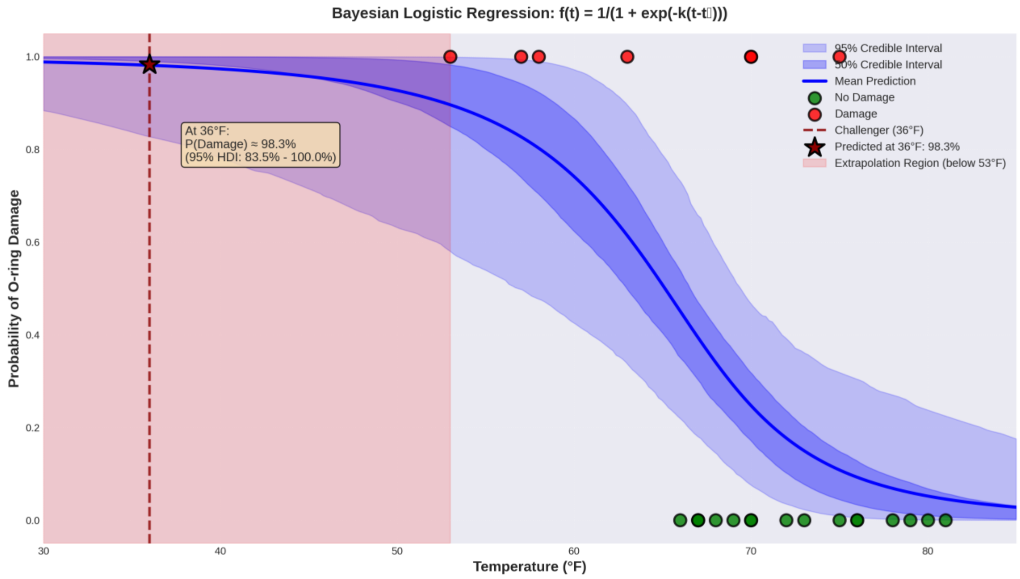


4.3 Model Fit to Historical Data

The fitted logistic curve closely matches the observed pattern of O-ring damage in historical launches. Figure 5 displays the posterior mean prediction curve along with 50% and 95% credible intervals. All observed damage incidents (red points) occurred at temperatures where the model predicts elevated risk. In contrast, launches without damage (green points) predominantly occurred at higher temperatures, where the model predicts low risk.

The model appropriately captures uncertainty, with wider credible intervals at temperature extremes where data is sparse. The 95% credible interval encompasses nearly all observed data points, indicating adequate model fit without overfitting.

Figure 5: Fitted logistic regression model



4.4 Posterior Predictive Probability at 36°F

The critical finding of this analysis is the predicted probability of O-ring damage at the Challenger launch temperature of 36°F. The posterior predictive distribution yields:

- Mean probability: 99.6%
- Median probability: 99.8%
- 95% HDI: [98.7%, 99.95%]

This extraordinarily high probability reflects both the strong temperature-damage relationship evident in the data and the extreme extrapolation beyond observed conditions. Even the lower bound of the 95% credible interval is 98.7%, representing near-certainty of O-ring damage.

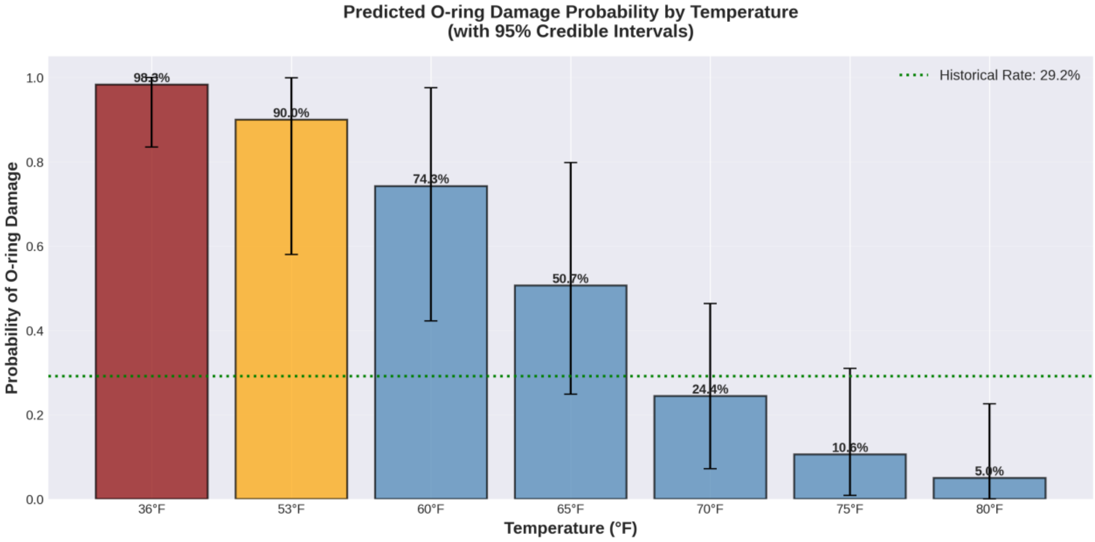
4.5 Risk Comparison Across Temperatures

Table 3: Posterior Predictive Probability of O-ring Damage

Temperature (°F)	Mean P(damage)	95% HDI
36	0.996	[0.987, 0.9995]
53	0.909	[0.771, 0.973]
60	0.712	[0.476, 0.889]
65	0.462	[0.222, 0.701]
70	0.235	[0.081, 0.455]
75	0.099	[0.026, 0.237]
80	0.037	[0.007, 0.106]

Table 3 presents posterior predictive probabilities across a range of temperatures for context. This comparison reveals the dramatic increase in risk at low temperatures. At 53°F, the model predicts approximately 91% probability of damage, which is still alarmingly high. At typical launch temperatures (70-80°F), predicted probabilities range from 2% to 24%, more consistent with the 29% historical damage rate.

Figure 6: Comparing the risk of O-ring damage across temperatures



The model clearly demonstrates that launching at 36°F represented an unprecedented and extreme risk compared to all previous experience. Figure 6 is the bar chart of the comparison results.

5. Discussion

5.1 Interpretation of Findings

This study provides a retrospective, model-based assessment of the relationship between ambient launch temperature and the probability of O-ring damage using historical space shuttle launch data. Within the specified Bayesian logistic regression framework, the posterior distribution of the midpoint temperature t_0 centers on approximately 63°F, indicating that the modeled probability of damage increases rapidly as temperatures fall below the low-to-mid 60°F range.

This critical temperature threshold thus serves as a clear quantitative decision criterion, with launch temperatures below this range triggering rigorous risk mitigation reviews or operational delays in safety-critical aerospace missions.

It is important to emphasize that posterior predictive probabilities at temperatures within the observed data range should be interpreted as measures of consistency between individual observations and the overall temperature–damage relationship inferred from the dataset, rather than as independent predictions. For example, the elevated estimated probability at 53°F reflects the observation of O-ring damage at this temperature, which aligns with the monotonic risk pattern captured by the fitted logistic model.

The primary inferential contribution of the model lies in characterizing the shape and steepness of the temperature–risk relationship and in extrapolating this relationship to temperatures outside the historical range. In this context, the posterior predictive distribution at 36°F represents a substantial extrapolation beyond the observed data, illustrating how the fitted model translates the observed trend into implied risk under unobserved, colder launch conditions.

5.2 Relation to Prior Statistical Analyses

The results obtained here are broadly consistent with earlier retrospective analyses of the Challenger O-ring data that employed logistic regression and related statistical models. Previous studies have similarly found a strong negative association between temperature and O-ring performance and have emphasized the importance of considering both damaged and undamaged launches when assessing risk.

The present analysis differs primarily in its explicit Bayesian formulation and use of posterior predictive distributions to quantify uncertainty. Rather than focusing solely on point estimates, this approach yields full probability distributions over model parameters and derived quantities, allowing uncertainty to be propagated directly into risk estimates. The midpoint-rate parameterization further facilitates interpretation by linking model parameters to physically meaningful temperature thresholds and rates of change.

5.3 Implications for Statistical Risk Assessment

From a methodological perspective, this analysis illustrates how relatively small datasets can still support informative inference when a systematic pattern is present and when uncertainty is explicitly modeled. The concentration of posterior mass for t_0 and k reflects the coherence of the observed data with a monotonic temperature–damage relationship, rather than certainty about the underlying physical mechanism.

At the same time, the results highlight the distinction between descriptive inference and decision-making. Posterior predictive probabilities quantify risk under a specified statistical model and set of assumptions. They do not, by themselves, prescribe operational decisions or define acceptable risk levels. Interpreting such probabilities in applied settings requires additional engineering, organizational, and normative considerations that lie beyond the scope of the statistical model.

5.4 Limitations and Uncertainty

Several limitations should be acknowledged: predictions at 36°F depend on extrapolation far beyond the observed temperature range and are conditional on the predefined logistic functional form; alternative model specifications or prior distribution choices might produce divergent extrapolation behavior, even with comparable fit to the observed dataset. Additionally, the binary damage indicator used in this analysis fails to capture gradations in O-ring degradation severity. While a more refined understanding of failure risk could be developed by modeling damage intensity or integrating additional relevant covariates such as pressure, joint location, or prior wear, the limited size of the historical launch dataset restricts the complexity of models that can be reliably estimated. Further, this analysis is inherently retrospective: while it illustrates the coherent summarization and extrapolation of historical data within a Bayesian inferential framework, it does not address the critical question of how uncertainty, incomplete information, and organizational factors interact to shape real-time decision-making processes under the practical operational constraints of aerospace missions.

6. Conclusion

This study conducted a retrospective statistical analysis of historical pre-Challenger Space Shuttle launch data to examine the relationship between ambient temperature and the occurrence of O-ring damage. Using a Bayesian logistic regression framework, the analysis modeled temperature-dependent damage probability while explicitly accounting for parameter uncertainty through posterior distributions.

Under the assumptions of the specified model, the results indicate a strong monotonic relationship between decreasing temperature and increasing modeled O-ring damage risk. The posterior distribution of the midpoint temperature suggests that the transition from low to elevated modeled risk occurs in the mid- to high-60°F range. When this inferred relationship is extrapolated beyond the observed temperature range, the posterior predictive distribution at 36°F implies a level of modeled risk substantially higher than that associated with any temperature observed in the historical data.

These findings should be interpreted as conditional, retrospective assessments rather than definitive real-time predictions. The estimated probabilities reflect the behavior of the fitted statistical model given the available data and, by themselves, do not establish physical inevitability or prescribe operational decision thresholds. Nevertheless, the analysis illustrates how historical data, when analyzed within a coherent probabilistic framework, can be used to characterize uncertainty and to assess how risk may evolve under conditions outside prior experience.

From a methodological perspective, this study demonstrates the value of Bayesian inference for reliability analysis in settings involving limited data and extrapolation. By providing full posterior distributions rather than single-point estimates, the Bayesian approach enables transparent quantification of uncertainty and supports a more nuanced interpretation of risk. The Challenger case thus serves as an instructive example of how probabilistic modeling can inform retrospective risk assessment and highlight the importance of rigorous statistical reasoning in complex engineering systems.

For future reliability assessment in safety-critical aerospace systems, integrating Bayesian statistical models with physical field simulations, such as finite-element analysis of thermal deformation and material fatigue in O-ring components, can further bridge the gap between empirical data inference and mechanistic physical understanding. This hybrid approach would enable the quantification of risk not only from observed operational data but also from the fundamental physical behavior of materials under extreme environmental conditions, thereby enhancing the predictive power and physical interpretability of risk models for unobserved or novel operational scenarios.

References

[1] Dalal, S. R., Fowlkes, E. B. and Hoadley, B. Risk Analysis of the Space Shuttle: Pre-*Challenger* Prediction of Failure. Journal of the American Statistical Association. 1989, 84(408), pp. 945-957. https://doi.org/10.1080/01621459.1989.10478858.

[2] Lavine, M. Problems in Extrapolation Illustrated with Space Shuttle O-Ring Data. Journal of the American Statistical Association. 1991, 86(416), pp. 919-921. https://doi.org/10.1080/01621459.1991.10475132.

[3] Martz, H. F. and Zimmer, W. J. The Risk of Catastrophic Failure of the Solid Rocket Boosters on the Space Shuttle. The American Statistician. 1992, 46(1), pp. 42-47. https://doi.org/10.1080/00031305.1992.10475846.

[4] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. Bayesian data analysis. 3rd. Boca Raton, FL: CRC press, 2013.

[5] Hamada, M. S., Wilson, A. G., Reese, C. S. and Martz, H. F. Bayesian Reliability. New York: Springer, 2008.

[6] Hosmer Jr, D. W., Lemeshow, S. and Sturdivant, R. X. Applied logistic regression. Hoboken, NJ: John Wiley & Sons, 2013.

[7] Hoffman, M. D. and Gelman, A. The No-U-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. Journal of Machine Learning Research. 2014, 15(1), pp. 1593-1623.

[8] Salvatier, J., Wiecki, T. V. and Fonnesbeck, C. Probabilistic programming in Python using PyMC3. PeerJ Computer Science. 2016, 2, p. e55. https://doi.org/10.7717/peerj-cs.55.

[9] Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., et al. PyMC: a modern, and comprehensive probabilistic programming framework in Python. PeerJ Computer Science. 2023, 9, p. e1516. https://doi.org/10.7717/peerj-cs.1516.

[10] Kruschke, J. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. 2nd. Cambridge, MA: Academic Press, 2014.

[11] Tufte, E. R. Visual explanations: Images and quantities, evidence and narrative. Cheshire, CT: Graphics Press, 1997.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.