Challenger Analysis

What's good and wrong with this analysis

In this review, I will find out what parts of the Challenger analysis showed some bias or errors that have prevented the scientists from finding out about the risks and complications leading to the catastrophe of Challenger.

1.1

In 1.1, the researcher adds in all the data that shows every important data to test potential joint failure and temperature. The only things I would probably do are the following:

Limit testing: Which is the art of going to extremes where we are sure that within our confidence intervals, there will always be a malfunction, however here we went with a less broad range of temperatures, implying that any launch should always come around these temperatures.

Use a better metric: While this is indeed America, using a more universal metric is always recommended, or at least give the reminder to convert F to C, but it's not inheritely a bad thing.

Order by temperature: Instead of having the data in chronological order (which holds little value) I would order it by temperature, that way it's better to see a pattern if too cold or too hot causes malfunctions.

1.2

In 1.2, I can find the first biasness in the analysis. The researcher did not take into count every variable and only assumed that if a given temperature didn't cause a malfunction, that it was never gonna do so, which is a bias selection of data, which is often seen as lazy and careless.

The graphic already displays that mistake as we calculate the probability of malfunction for each temperature, but only we have a single test per temperature, but we have a very limited of discrete data to work with.

1.3

At that point on, they determine a logistical model.

They define: **Frequency = Malfunction / Count**

In the code, we see `family=Binomial(logit)).fit()` which implies it uses a binomial model. However such a model is for counts of success and failures over n trials, but here we only have a count of 6 O-rings. However let's assume a 0.5 failure rate, in a model with 6 counts is 3 out 6, however in a much larger count let's say 20, 10 out 20 is much more informative.

A binomial model always expects a number of failures and a number of trials, however the number of trials are forgotten once the frequency is calculated.

A vulgarised example would be: let's say a restaurant has a 3 star rating, and we asked if there were 2 or 2000 reviews, and i told you it doesn't matter. That's a methodological fallacy.

Moreover, only having such a small sample size makes the confidence intervals extremely untrustworthy as they explode, the p-values become meaningless and any regression is unstable as we do not see the full picture.

1.4

In this case, they tried to create a linear regression of all our probabilities and data which gave us a nearly straight line, assuming near no impact in failure based on temperature, however I spotted a handful of issues:

1. The person plot a curve without making any confidence intervals, no uncertainty range and only using a single probability value as said earlier.
2. This is a more vulgarised approach, but saying that temperature has no effect after this does not imply "safety" therefore we have a logical fallacy. Correlation does not imply causation. The absence of evidence does not prove the evidence of absence, especially with how lazy the effort was in this testing and analysis.
3. We extrapolated a conclusion outside of the observed temperature range.