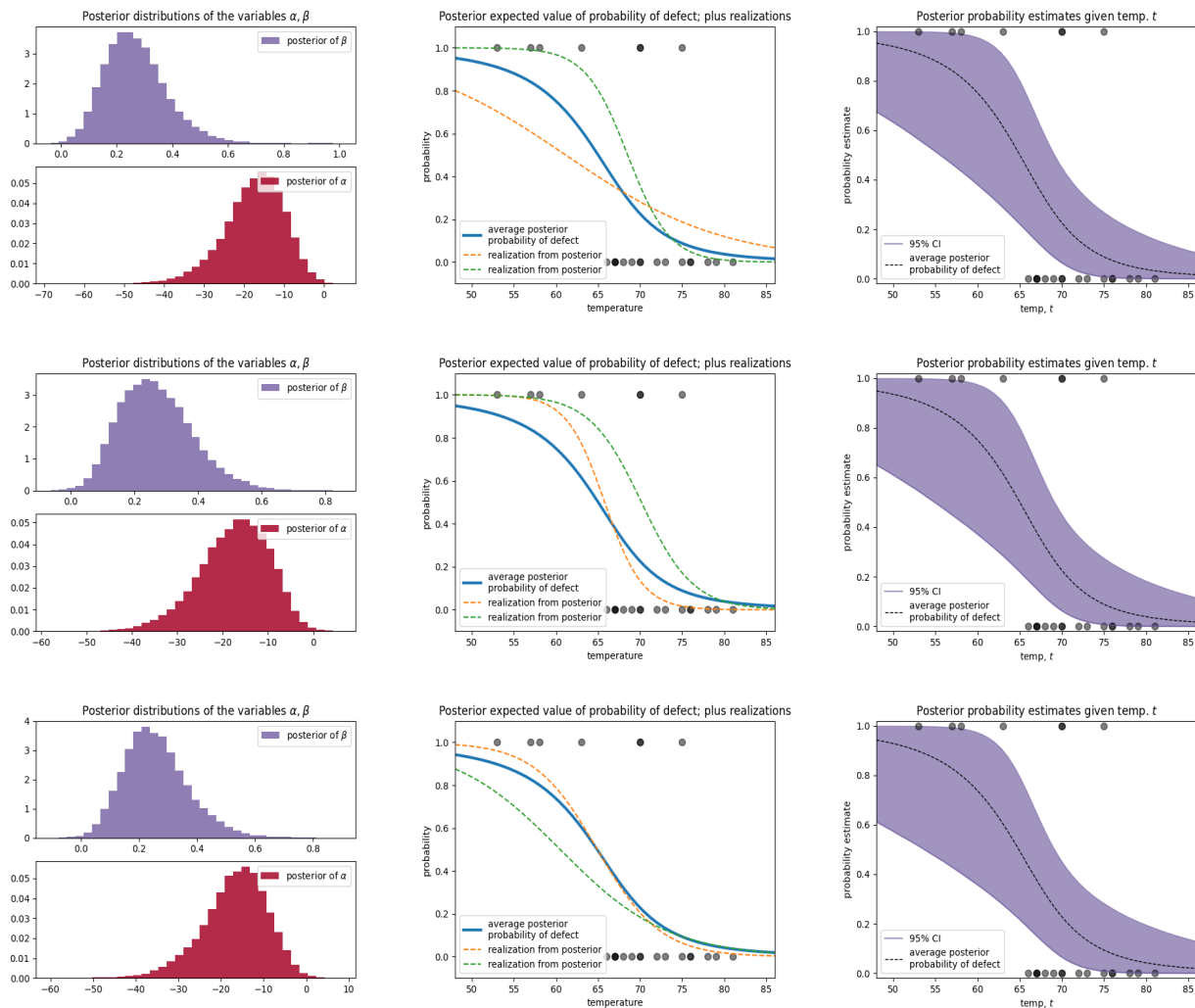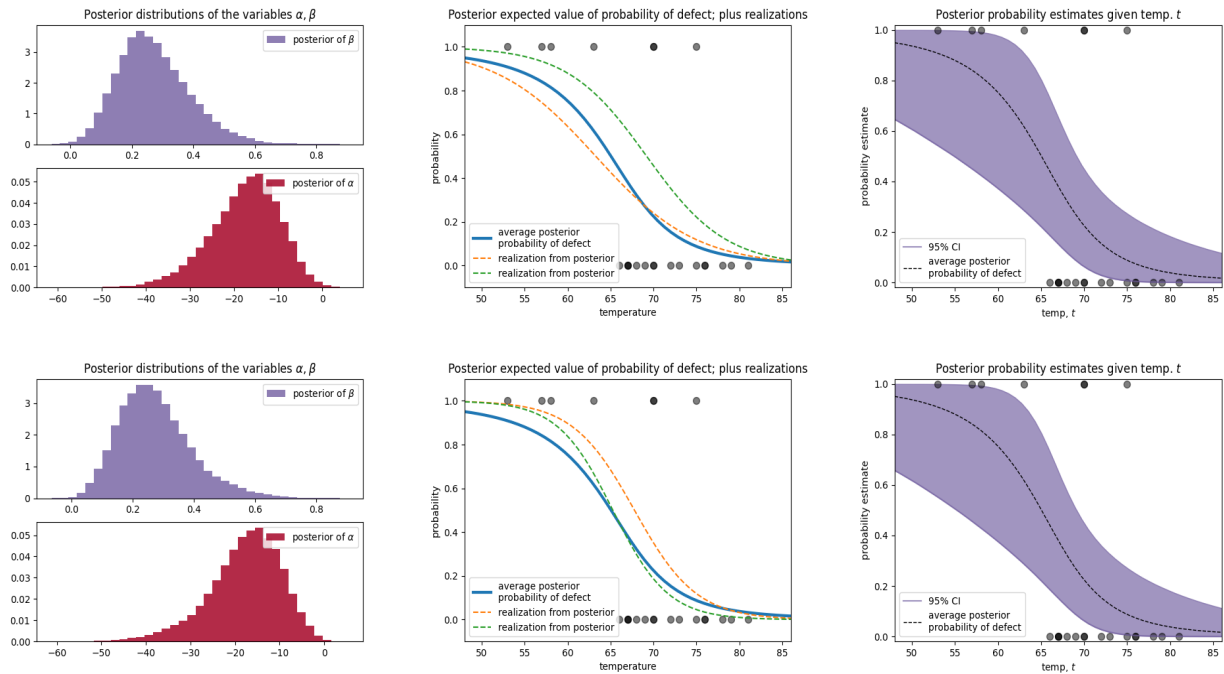Jack Greff

Dr. Grissom

Foundations of Data Science

8/27/2022

Question 1: Try running your simulation at least five times without changing anything. Are the results consistent?
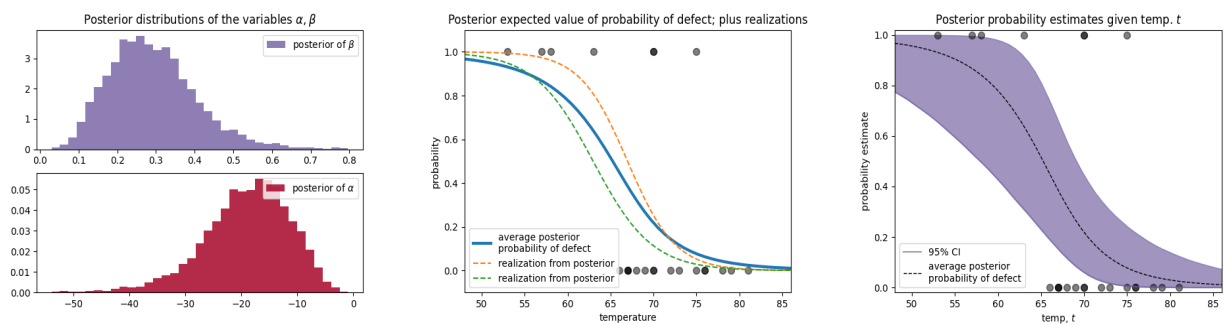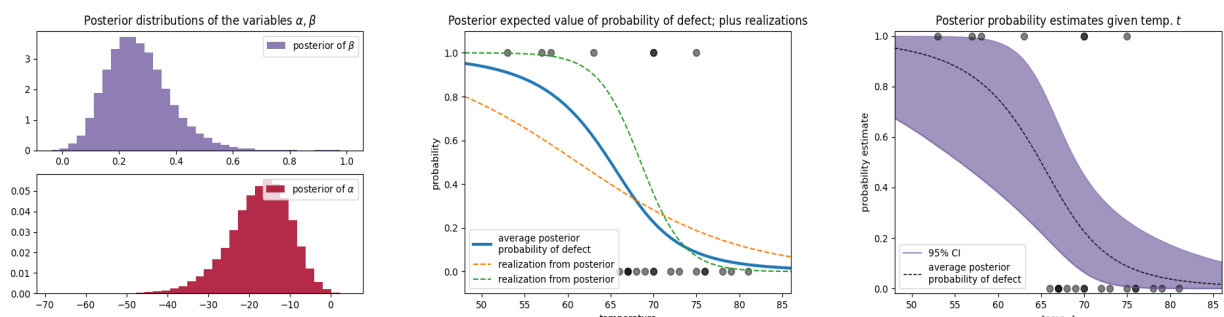
The distributions for the Alpha and Beta values look amost identical . As well, the average posterior probabiility of defect and credibility intervals are almost identical throughout all five. There is definitely some variance (like the floor of the credibility intervals in the probability model) which is to be expected, but overall, yes, the results are consistent.

Question 2:. Try changing the number of samples. Do you find that this changes your results?
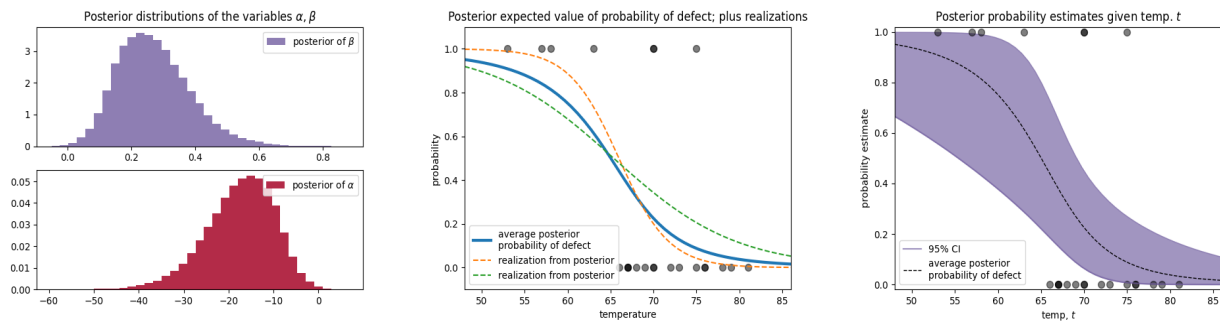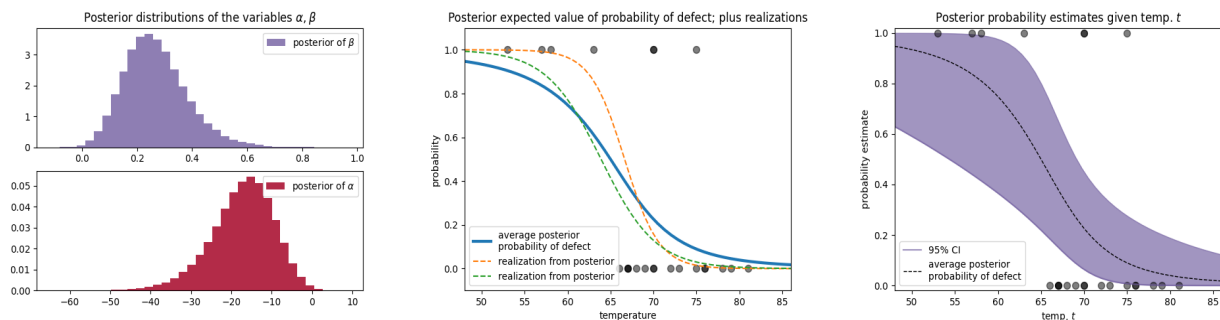
30,000 trials:

300,000 trials (1 of 5)

600,000 trials:

1,200,000 trials:

There are numerous differences to point out with more trials. To start, the alpha and beta values lose variance as more trials occur. Since these values are generated from normal distribution, more samples means more certainty. As well, the credibility intervals gain more variance as more trials occur (i.e at 55 degrees, the floor of probability moves from near .8 to near .6 from 30,000 trials to 1,200,000 trials) . The unpredictability of the data due to it's little quanity allows for more variance, which is amplified with more trials. With less data, the CI is more likely to dismiss outlying data, but more trials brings more credibility to these values.

Question 3: Try removing the first half and the last half of the Challenger data. How does this affect your results?

Data with only the first 13 launches:
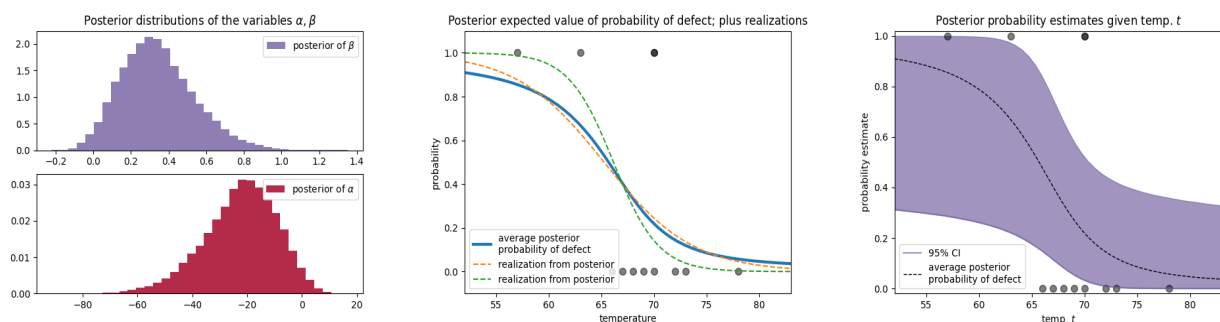
Data with only the first 13 launches:
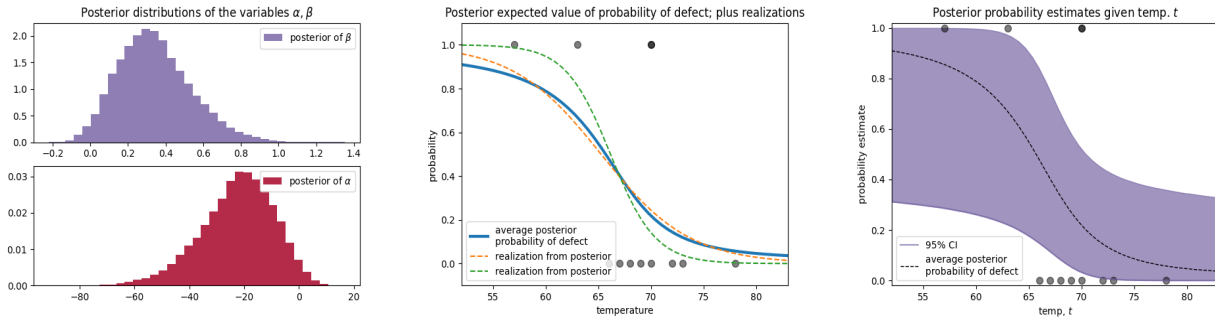
Our results have noticable changes. The largest difference is the confidence intervals have significantly more variance then before. With the full data, at 55 degrees, the probabiliity of an accident is between 65% and 100% and 0% and 15% at 85 degrees. However, with the first half, the probability can be as low as 30% at 55 degrees and as high as 30%, while with only the second half, the probability can be as low as 35% at 55 degrees and as high as 20%. Consistent among both, the variance has grown with less data. This is consistent with the notion that when more data is provided, our certainty of our results and models grows.

Question 4: Based on your experiments, argue for reasonable conclusions (with plots and analysis) about the relationship between temperature and failure.

30,000 trials (1 of 5):
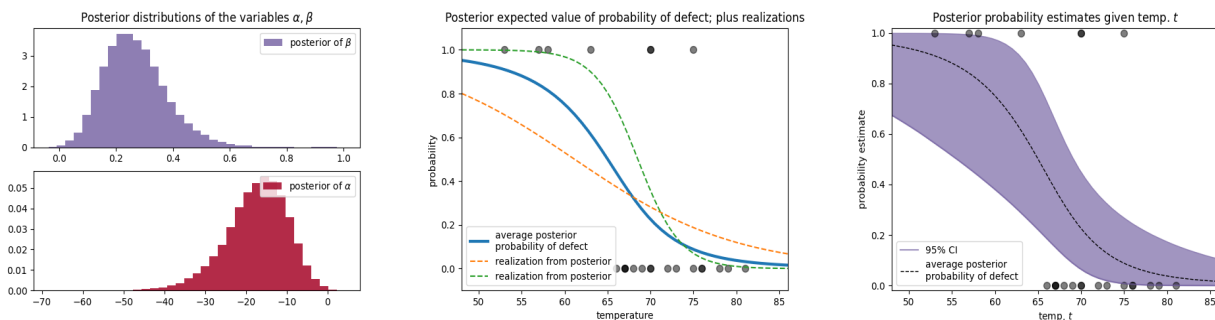
With only 26 pieces of data, making confident statements about the likelihood of failure isn't responsible or fair. For example, at 65 degrees, the likelihood of failure can be as high as roughly 85%, but as low as 20%. That ranges from extremely likely to strongly unlikely. Our distrubutions for alpha and beta are also not very certain either. We could be shifting our model (controlled by alpha) as much as 40 degrees and flattening the curve with a factor differece of .6. It's also important to bring up that as more trials occured, more variance was provided. The strongest conclusion is that roughly the data will follow this shape and we can expect the trends shown above to be somewhat accurate, but more data is needed to clear much of the uncertainty that remains.