Questions for Discussion

4.1

> A 1 in 334 chances of occurring naturally seems really unlikely. However the distribution of scaled residual looks pretty Gaussian, and the scatter plot above looks pretty random; what is going on? Why does it make sense that the largest scaled residual is about 2.75 sigma even if the data are distributed randomly? Why wouldn't we consider this to be a really flaring episode? (Use numbers to back up your explanations.)

While a 1 in 334 chance of this occurring is rather small, we've made 651 observations in this case so it's actually rather likely that this would occur at least once.

4.2

> Given the way that analysis was done, and that we have observed for 651 weeks, what level of statistical significance might we want to use as a threshold for a flaring episode?

We'd want a level of statistical significance such that it would be very unlikely to observe a certain datapoint given a batch of 651 observations (e.g., something with a 1 in 10,000 chance of occurring).

5.1

> In contrast to the case for the Vela pulsar, here, the chance for an outlier this size occurring randomly is tiny, even given the fact that we are considering over 600 weeks of data. In fact we'd have to observe for something like 2×10^{11} = 200 billion weeks to expect a random fluctuation of this size. On the other hand, the change in significance only went from 2.75σ to 6.81σ. Explain, in your own words, what this means in terms of how we should think about significance in terms of σ as compared to significance in terms of p-values.

When thinking about significance in terms of σ instead of p-values, we must realize that the significance increases much faster than σ does. For example, there is a more than 100

times decrease in the survival fraction at 4σ when compared to 2σ.

5.2

> Does this change your opinions about the conventions used to define statistical significance? If someone reports a result with a p-value of $p < 0.05$ what are some follow up questions you might ask them?

One question I might ask them was the number of observations they made. If the made a copious number of observations, then I'd expect them to use a lower p-value threshold than if they only made a few observations.