SSW533 HW 2

Names: Keidon Roettger, Markell Torres, Jon Cucci, William Baltus

I pledge my honor that I have abided by the Stevens Honor System. **WB, KR, JC, MT**

Summary:
　　In this assignment, the group was provided with data on 15 projects, their lines of code, and the defects their project had. Using this data, a linear model was created to determine the relationship between defects and lines of code. The trendline was strictly non-decreasing, therefore indicating that defects increase with lines of code, though there were also some outliers! Moreover, using a different metric (Defects/KLOC), a frequency distribution was plotted. The standard deviation and mean were determined, and a clear right skew was shown. Moreover, the group learned the importance of good data, whether it be with the outlier(s) in the trendline, or the lack of data in the histogram. Having that sort of 'ideal' data would make the models much more accurate. Though that is not always the case, and so different modeling and analysis techniques are better or worse for certain tasks.
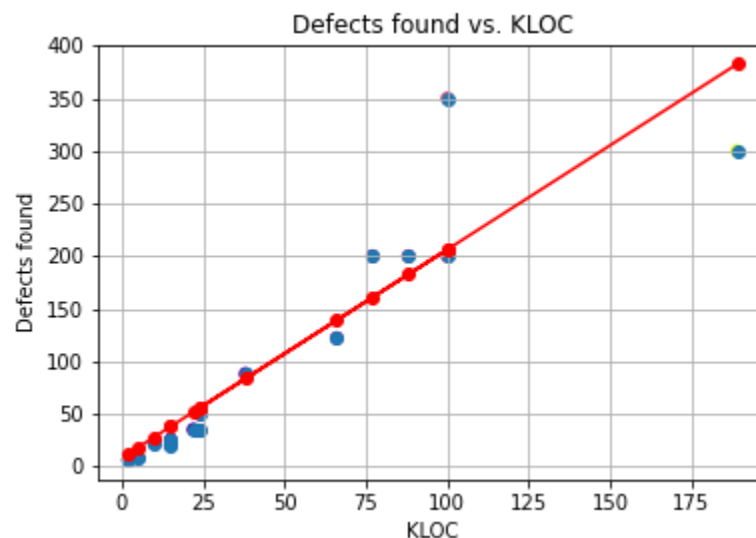
1.

2. Based on the linear model above, the predicted amount of defects for 50K LOC is approximately 103!

Histogram of Defects / KLOC: $\mu = 2.07$, $\sigma = 0.608$
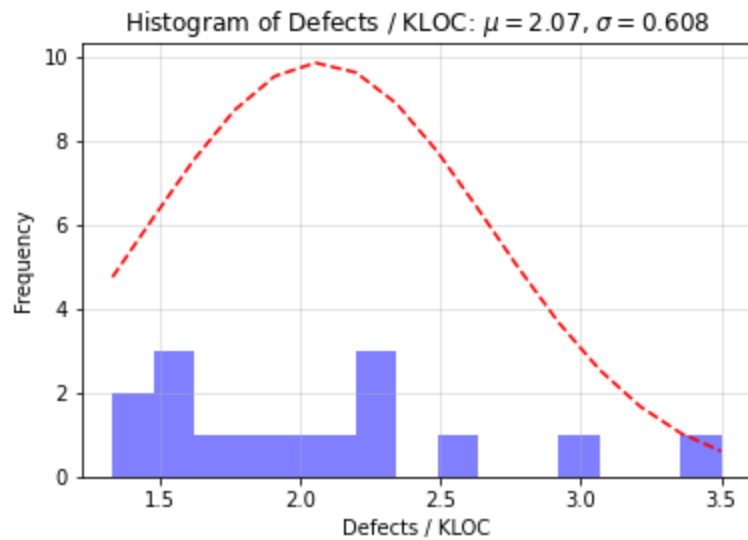
3.

Mean: 2.07
Standard Deviation: 0.608

4. Assuming the above distribution was theoretically a normal distribution, then the expected range of defects for 95% of the data would be within two standard deviations. The two standard deviations would be approximately 1.216 defects/kloc from the center. The figure below indicates what this range would be from ~0 - ~3.286.

Histogram of Defects / KLOC: $\mu = 2.07$, $\sigma = 0.608$
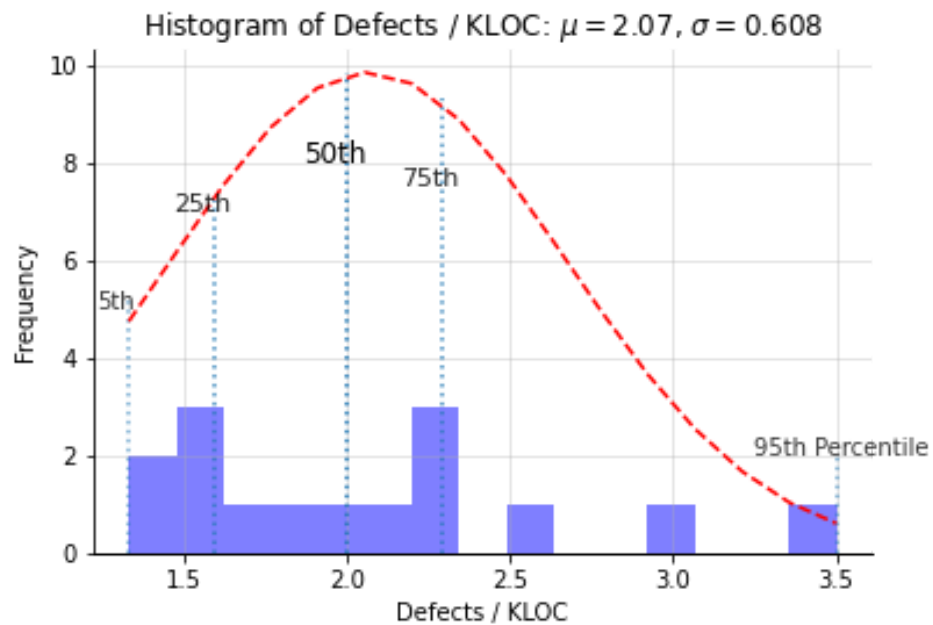
Mean: 2.07
Standard Deviation: 0.608

5. A normal distribution is clearly not reasonable for the given data. There is a clear skew as most projects fell under 2.5 Defects/KLOC. Therefore this distribution is right skewed.

Observations:

The group determined to generate random data based on the trendline initially made. While the trendline was not perfect, it still provided insight on what a 'more normal' distribution might have looked like. The figure below shows the distribution the team discovered.

Histogram of Defects / KLOC: $\mu = 2.07$, $\sigma = 0.608$