Homework 1 (METR 5803)

| | Mean | Median | Std Dev | IQR | MAD | p0.10 | p0.25 | p0.75 | p0.90 |
|---|---|---|---|---|---|---|---|---|---|
| Jan | 27.5 | 20.5 | 34.6 | 28.5 | 14.5 | 2.5 | 6.8 | 35.3 | 54.5 |
| Feb | 35.5 | 29.0 | 31.2 | 40.0 | 17.5 | 7.5 | 12.0 | 52.0 | 65.5 |
| Mar | 86.2 | 71.0 | 50.1 | 50.8 | 23.0 | 41.5 | 56.0 | 106.8 | 135.5 |
| Apr | 176.4 | 169.0 | 70.1 | 77.5 | 38.0 | 97.0 | 131.8 | 209.3 | 245.5 |
| May | 290.8 | 270.5 | 104.6 | 139.0 | 71.5 | 170.5 | 207.0 | 346.0 | 453.0 |
| Jun | 262.0 | 266.0 | 79.5 | 101.8 | 53.0 | 151.0 | 217.3 | 319.0 | 344.0 |
| Jul | 147.5 | 138.5 | 53.4 | 67.8 | 30.5 | 84.5 | 110.0 | 177.8 | 225.5 |
| Aug | 96.0 | 95.5 | 35.1 | 39.3 | 22.5 | 55.0 | 72.3 | 111.5 | 140.0 |
| Sep | 71.6 | 54.5 | 49.9 | 58.0 | 22.5 | 32.0 | 39.8 | 97.8 | 113.5 |
| Oct | 48.8 | 43.0 | 27.4 | 39.5 | 21.0 | 18.0 | 27.0 | 66.5 | 84.0 |
| Nov | 50.5 | 34.0 | 44.4 | 50.0 | 22.0 | 8.5 | 19.8 | 69.8 | 115.0 |
| Dec | 28.5 | 21.0 | 29.4 | 35.5 | 16.0 | 2.0 | 5.8 | 41.3 | 71.5 |

Points of interest:

1) Note the asymmetry in June's mean/median numbers compared to the rest of the year. All the other months show mean>median (August is essentially even). In particular, compare June's values to May. The means differ by ~10%, with the medians being equal. The 90$^{th}$ %ile is much higher for May, leading to the mean being much higher. Big Junes aren't as big as big Mays.

2a) The standard deviation in winter is > the mean (or approximately equal). That implies that the normal distribution is not a particularly good fit for monthly winter tornadoes. (It may not be a good fit any other time, but it looks bad from just a quick look at the parameters in April.)

2b) For a normal distribution, IQR~1.34*(2*StdDev). Feb, May-July, and Oct are roughly there. One could argue that the annual cycle looks like May-July are close to normal, and the rest of the year has too small of an IQR for the standard deviation. This implies that really extreme values (to which the StdDev is sensitive) occur more often than in a normal distribution outside of May-July. Perhaps the normal distribution doesn't work well there.

3) Look at the annual cycle of the lower three quantiles (p0.10, p0.25, and median) compared to p0.75 and p0.95, particularly for November. The lower quantiles have no secondary maxima, p0.75's Nov is greater than Oct and p0.90 Nov is much greater than Oct. Apparently, when there are tornadoes in November, there are a lot. It happens in several years, but not very often. As a result, the median for Nov fits nicely between Oct and Dec, but the mean Nov is greater than Oct. We don't see similar behavior (at least when calendar months are considered) in the late winter/early spring.

4) If we were looking at larger sample sizes, we could consider normality by looking at the fraction of the sample inside the 2StdDev confidence bounds. For a normal distribution, 95% of the distribution is within 2 StdDev of the mean. If we had p0.025 and p0.975, we could look at that, but with a sample size of 56, we'd essentially be looking at the second largest and second smallest values in the distribution. That's sufficiently out in the tails that we shouldn't have much confidence in the robustness of those values.