**Q1: How to minimize lost days and restricted days.**

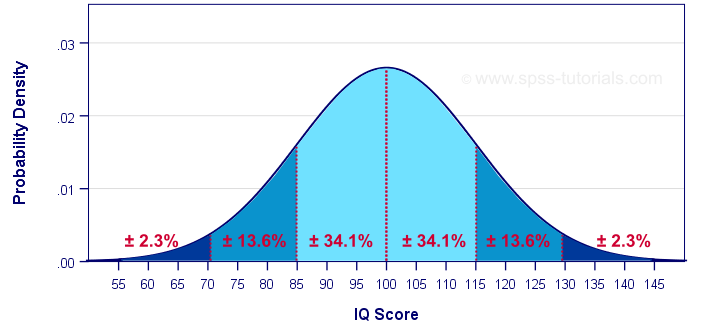
**Q2: How much we can save and how much we need to cost in order to do that?**

I extracted **one** year’s **WC** data about both lost days and restricted days. There are **410 rows** of records in total.

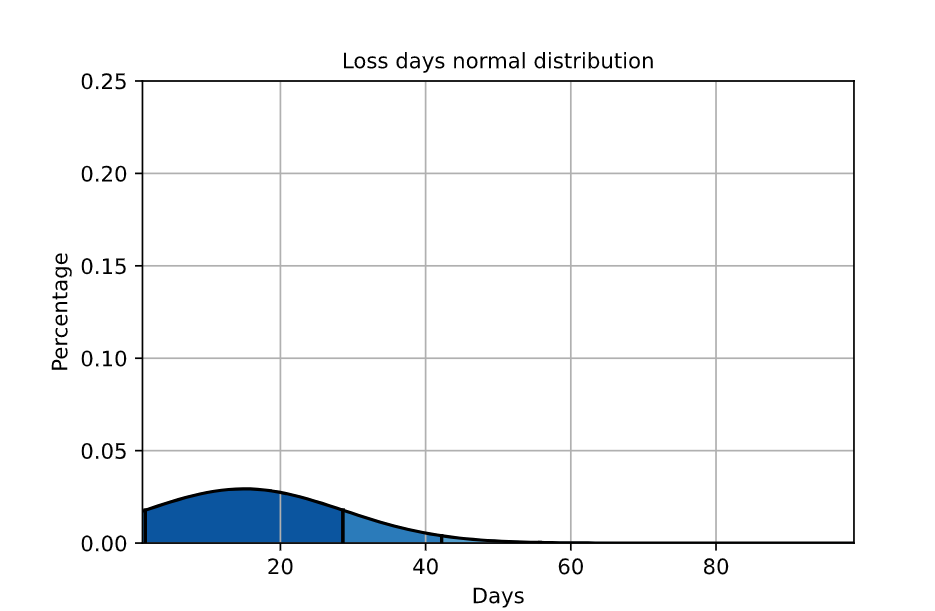
**Lost days only**: We have 163 individuals and 164 records within last year. There is one person who claimed 2 lost days records (‘WC511620006939’).

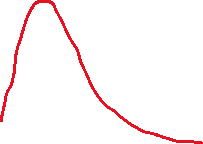
* The mean lost day is: 15
* The lost day’s standard deviation is: 13.6
* Min loss days: 1
* 25% : 8.5
* 50% ：12
* 75%：16
* Max：99

I am expecting to see a normal distribution as below:



However, we got this:





Clear, our loss days accident data is highly skewed. The problem is we have too much **large** loss days and because of that, our **standard deviation is too high**, so that this chart is stretched.

**So how can we make this happen and how much we can save by doing it?**

I assume the average hourly salary is $25(I need some help to get the accurate average hourly salary). For now, we are losing 2459 hours in this category, which means we lose $61,475 in total.

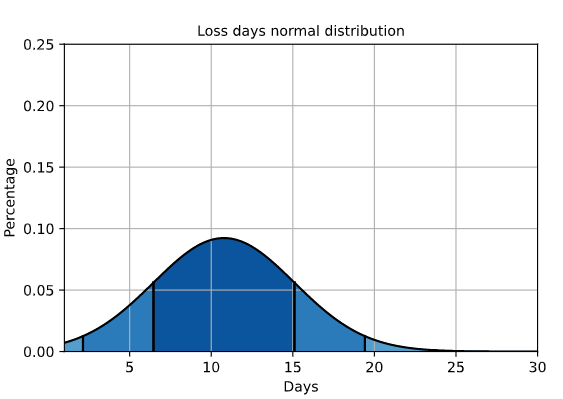
We should have 84.1% cases in the dark blue area but for now we have 91.5% in this area. It’s not a good thing in this situation, our standard deviation is too big. Then I realize we have 84.7% of records which total loss days <=19 days.

I decide to split out loss days data file into 2 part:

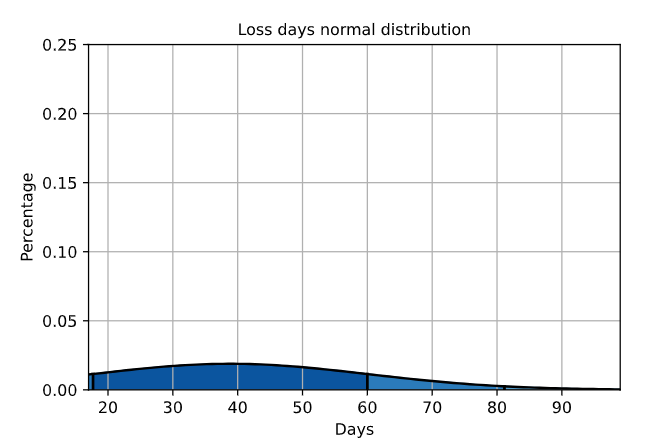
* + Part1: loss days<=19 days
    - count 138.000000
    - mean 10.782609
    - std 4.327369
    - min 1.000000
    - 25% 8.000000
    - 50% 11.000000
    - 75% 14.000000
    - max 19.000000
  + Part2: loss days>19
    - count 25.000000
    - mean 38.840000
    - std 21.153565
    - min 20.000000
    - 25% 22.000000
    - 50% 31.000000
    - 75% 46.000000
    - max 99.000000

And here is the new distribution we got:

**Part1 chart**



**Part2 chart:**



Then we get a nice-looking normal distribution chart in part 1 and an ugly chart in part 2. Part 2 should occupy 2.5% (4 records) and now we have 15.3% (25 records out of 163). If we can successfully reduce the number of outliers, we can save $14,810 or 24% in Loss days category. From now on, I have changed the original question into **“How can we reduce the long loss days”.** The new direction would be we need to find out all records longer than 19 days and dig in from there.