**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?**

**Some time they can be assigned to a restricted positions. If we create a program.**

**EX: you can go back to work but you can’t lift more than 10 lb in the following 30 days.**

**What’s used to happen is he would stay at home for more than 30 days but for now I want to get this guy to work somewhere else. The main goal is to turn lost days into restricted days.**

**My function is to extract data from Origami Risk, and we want to reduce the loss days rate by, let’s say 25%, how much we can save and how much we need to cost.**

**The Main goal is to look at the claim and interpretation.**

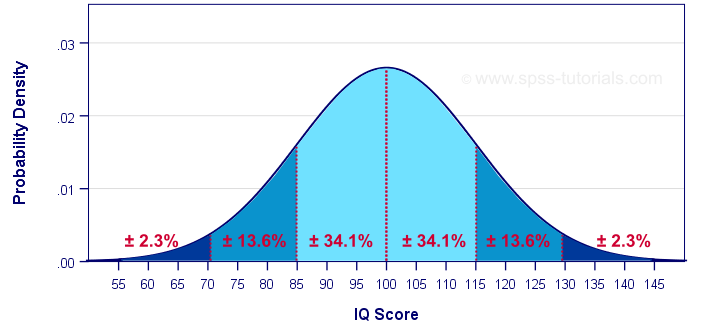
I extracted one year’s WC data about lost days or restricted days.

There are 410 rows of records in total.

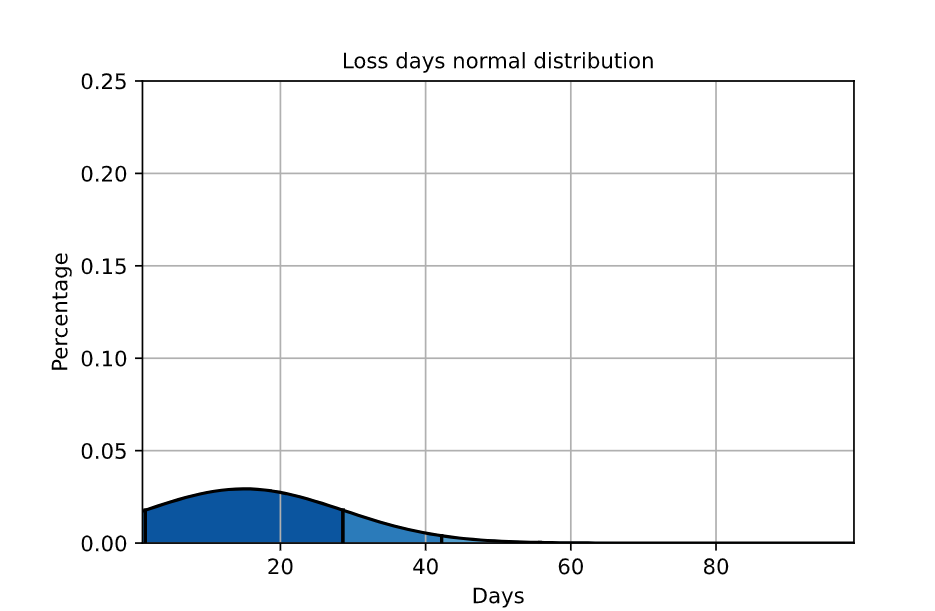
**Lost days only**: We have 163 individuals and 164 records within last year. There is one person who claimed 2 lost days records (‘WC511620006939’).

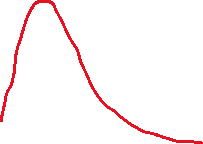
* The mean lost day is: 15
* The lost day’s standard deviation is: 13.6
* Min loss days: 1
* 25% : 8.5
* 50% ：12
* 75%：16
* Max：99

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 here is we have too much large numbers 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. Right now, we are losing 2459 hours, so in total we lose $61475 in loss days.

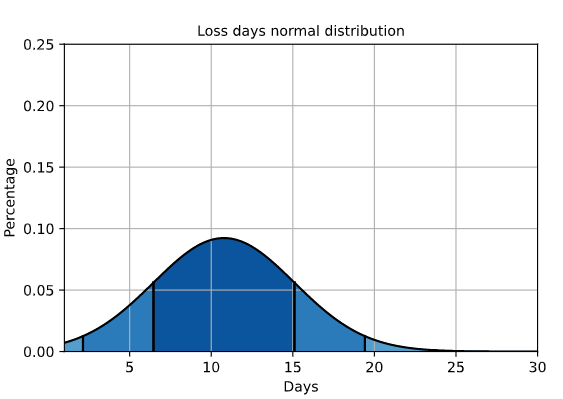
We should have 84.1% cases in the dark blue area but for now we only have 91.5% in this area. It’s not a good thing in this situation, our standard deviation is too big. Then I realize we have 84.7% of records which total loss days <=19 days.

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

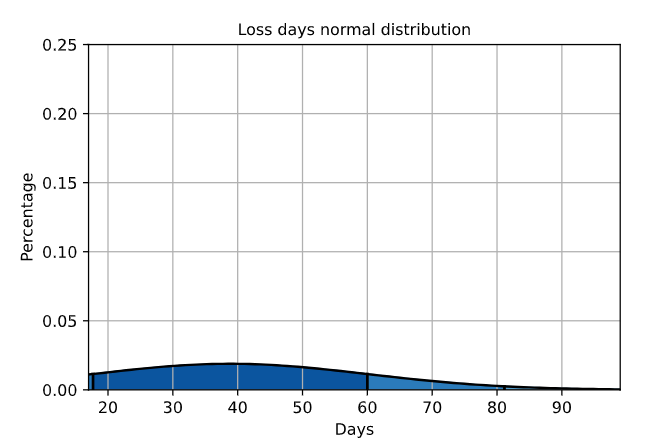
* + Part1: loss days<=19 days
    - count 138.000000
    - mean 10.782609
    - std 4.327369
    - min 1.000000
    - 25% 8.000000
    - 50% 11.000000
    - 75% 14.000000
    - max 19.000000
  + Part2: loss days>19
    - count 25.000000
    - mean 38.840000
    - std 21.153565
    - min 20.000000
  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

**Restricted days only**: 166 individuals and 173 records. There were 7 person with 2 restricted days records ('WC250520005462', 'WC267420006767', 'WC513520005639', 'WC275120006147', 'WC267220007763', 'WC266220007047', 'WC507320005772')

-03