**Central Limit Theorem and the Poisson Distribution: an Example**

At a certain intersection, we have a red light camera that helps the Department of Transportation fine reckless drivers. Imagine that X, the number of violations recorded per week, has a Poisson distribution, and that in a typical week, there are 25 violations. That is, average number of violations per week = λ = 25. Because in a Poisson distribution, it implies that , and .

Let’s ask 2 questions:

1. What’s the probability that there are 30 violations in a particular week if the average number of violations is 25?

**Answer:** This is exactly like the problem we saw before. We can calculate this probability using the **dpois(x, λ**) formula in R:

**dpois(30,25)**

and obtain 0.0454.

1. Recall that if we have a population where , it means that . Now, imagine that we get a sample of 64 weeks of data and see how many violations there are in each week. Here are the data:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **17** | **23** | **26** | **20** | **21** | **20** | **21** | **22** |
| **23** | **20** | **26** | **19** | **18** | **25** | **19** | **25** |
| **26** | **17** | **20** | **28** | **38** | **28** | **29** | **28** |
| **24** | **26** | **22** | **25** | **14** | **22** | **22** | **22** |
| **17** | **24** | **26** | **16** | **20** | **21** | **20** | **19** |
| **28** | **16** | **20** | **27** | **17** | **20** | **31** | **14** |
| **19** | **24** | **30** | **22** | **18** | **20** | **26** | **24** |
| **23** | **26** | **26** | **17** | **20** | **28** | **24** | **24** |

We can calculate the average of the sample:

violations/week

The question is, how likely are we to draw a sample of 64 weeks with mean number of violations of violations/week from our population, where violations/week?

**Answer:** Because we have 64 weeks, and 64 > 30, the Central Limit Theorem applies. That is, even though our sample of 64 weeks came from a Poisson Distribution, we can still assume that the random variable will have a normal distribution, with violations/week and violations/week. That way, we can re-express the distance between 22.55 violations/week and 25 violations/week (i.e., 2.45 violations/week) in terms of standard deviations.

We do this as follows:

In other words, 22.55 is 3.92 standard deviations below 25 weeks. The probability of being (at most) 3.92 standard deviations below the mean is calculated using the function **pnorm(-3.92)** in R. We get 0.00004427. In other words, it’s very unlikely to see a sample of 64 weeks where the mean is (at most) 22.55 violations/week if the true population mean is 25 violations/week.

And if we had a sample of say, 10 weeks, we couldn’t use the Central Limit Theorem, because you need n > 30 for it to work.