The graph clearly shows that set 1 data has more error when compared to set 2 data. This is how we estimate the standard error.

BOOTSTRAP:

One easy and effective way to estimate the sampling distribution of a statistic is to draw additional samples, with replacement, from the sample itself and recalculate the statistics or model for each resample. This procedure is called the bootstrap and it does not necessarily involve any assumptions about the data.

Conceptually, you can imagine the bootstrap as replicating the original sample thousands or millions of times so that you have a hypothetical population.

In practice, it is not necessary to actually replicate the sample a huge number of times. We simply replace each observation after each draw; that is we sample with replacement.

In this way we effectively create an infinite population.

Semester V Subject Statistics for AIDS Academic Year 2023-2024

The algorithm for a bootstrap resampling of the mean is as follows:

* Draw a sample value, record, replace it.
* Repeat n times.
* Record the mean of the n resampled values.
* Repeat step 1-3 N times.
* Use the N results to
  (a) Calculate their standard deviation
  (b) Produce a histogram or boxplot.
  (c) Find a confidence interval.

The more iterations you do, the more accurate the estimate of the standard error. (or) confidence interval.
The R package boot combines these steps in one function.
For example, the following applies the bootstrap to the incomes of people taking out loans:

```
library (boot)
stat_fun ← function(x, idx) median (x[idx])
boot_obj ← boot(loans_income, R=1000,
                statistic = stat_fun)
```

The function stat_fun computes the median for a given sample specified by the index idx. The result is as follows:

Semester: V Subject: Statistics for AI&DS Academic Year: 2023 2024

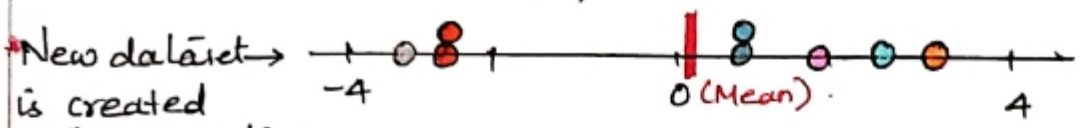
Bootstrap statistics:

	Original	Bias	Std. Error
t1*	62000	-70.5595	209.1515

The original estimate of the median is $62,000. The bootstrap indicates that the estimate has a bias of about -$70 and a standard error of $209.

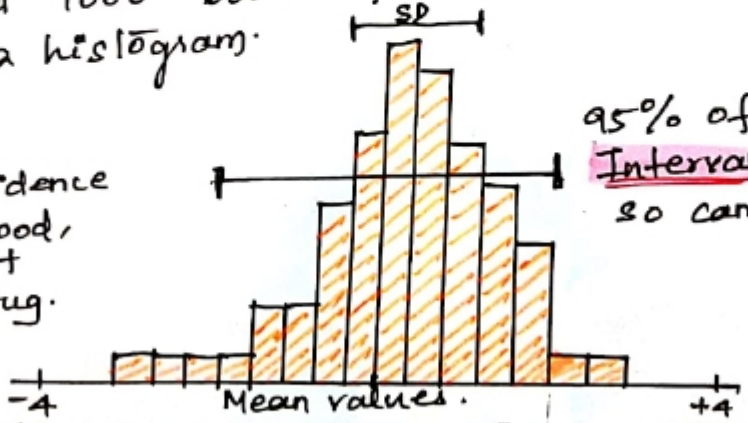
Example:- (Refer PPT)

Consides this example to check a drug on a sample of 8 people. The graph shows 5 people are feeling better and 3 are feeling worse.



is created using sampling with replacement with same values as original - **Bootstraped Dataset**

We get different mean. Let us continue to do the same with around 1000 bootstraped Dataset and plot the means in a histogram.



Conclusion:-

Since the confidence interval is good, so we cannot reject this drug.