

The background of the slide is a light gray gradient. It is decorated with numerous realistic water droplets of various sizes. Some droplets are large and prominent, while others are small and subtle. They are scattered across the slide, with a higher concentration in the top-left and bottom-right corners. Each droplet has a highlight and a shadow, giving it a three-dimensional appearance.

STANDARD ERROR AND BOOTSTRAP

PROF. SARALA MARY

STANDARD ERROR

- THE STANDARD ERROR QUANTIFIES THE VARIATION IN THE MEANS FROM MULTIPLE SET OF MEASUREMENTS.

$$SE = \frac{s}{\sqrt{n}}$$

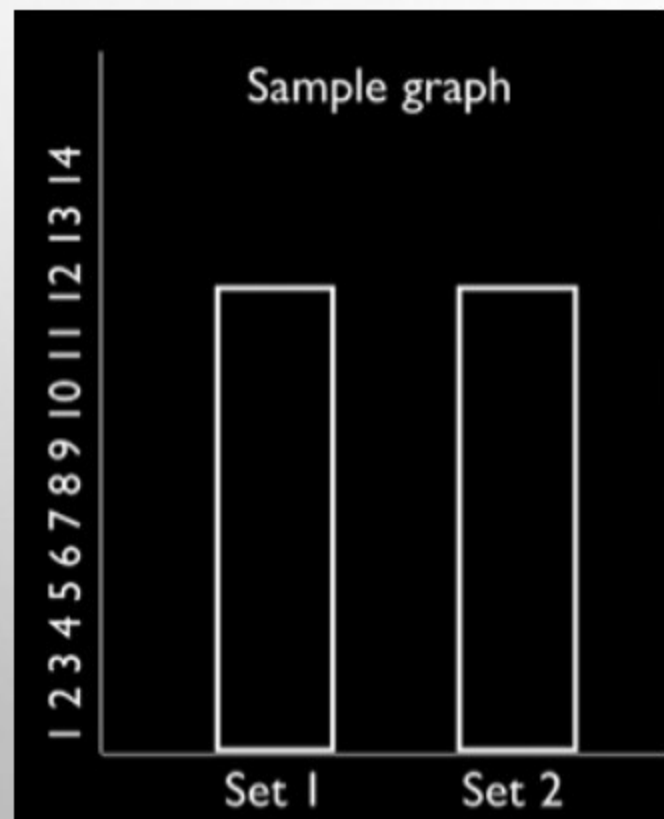
STANDARD ERROR

- CONSIDER THE FOLLOWING APPROACH TO MEASURE STANDARD ERROR:
 1. COLLECT A NUMBER OF BRAND-NEW SAMPLES FROM THE POPULATION.
 2. FOR EACH NEW SAMPLE, CALCULATE THE STATISTIC (E.G., MEAN).
 3. CALCULATE THE STANDARD DEVIATION OF THE STATISTICS COMPUTED IN STEP 2; USE THIS AS YOUR ESTIMATE OF

BOOTSTRAP

- A SAMPLE TAKEN WITH REPLACEMENT FROM AN OBSERVED DATA SET.
- WE SIMPLY REPLACE EACH OBSERVATION AFTER EACH DRAW CALLED AS SAMPLE WITH REPLACEMENT.

STANDARD ERROR - EXAMPLE



Prof. Sarala Mary - APSIT

Set 1

9

15

Set 2

10.9

11.9

12.2

12.2

12.9

12.6

12.3

12.3

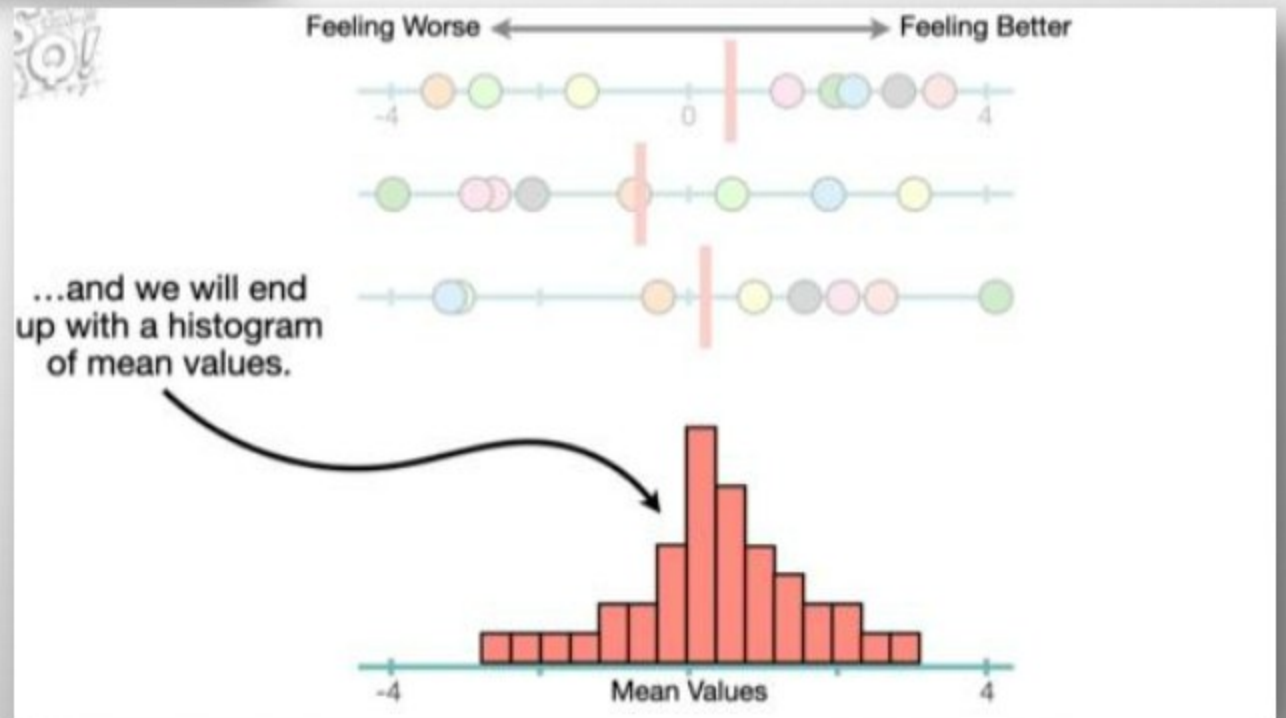
12.5

10.2

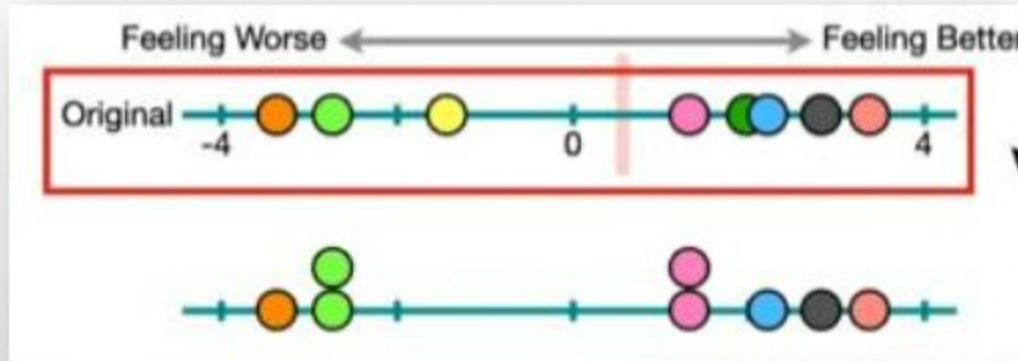
BOOTSTRAP ALGORITHM

- THE ALGORITHM FOR A BOOTSTRAP RESAMPLING OF THE MEAN IS AS FOLLOWS,
 1. DRAW A SAMPLE VALUE, RECORD IT, AND REPLACE IT.
 2. REPEAT N TIMES.
 3. RECORD THE MEAN OF THE N RESAMPLED VALUES.
 4. REPEAT STEPS 1–3 R TIMES.
 5. USE THE R RESULTS TO:
 - A. CALCULATE THEIR STANDARD DEVIATION (THIS ESTIMATES SAMPLE MEAN STANDARD ERROR).
 - B. PRODUCE A HISTOGRAM OR BOXPLOT.
 - C. FIND A CONFIDENCE INTERVAL.

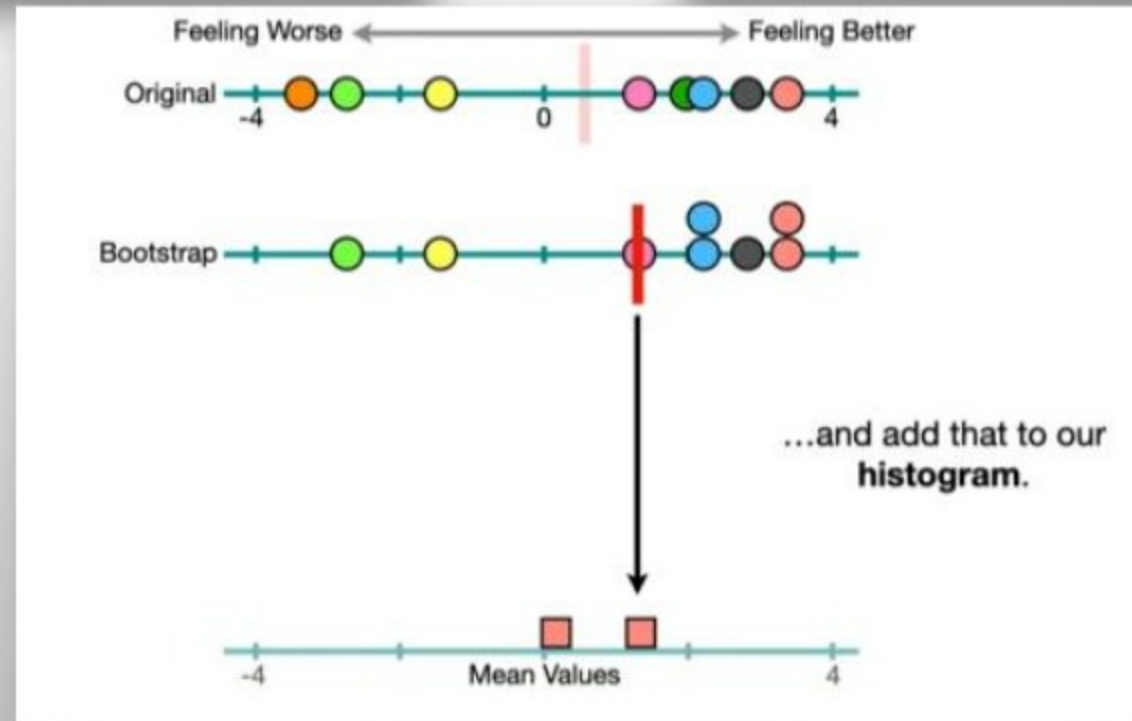
BOOTSTRAPPING - EXAMPLE



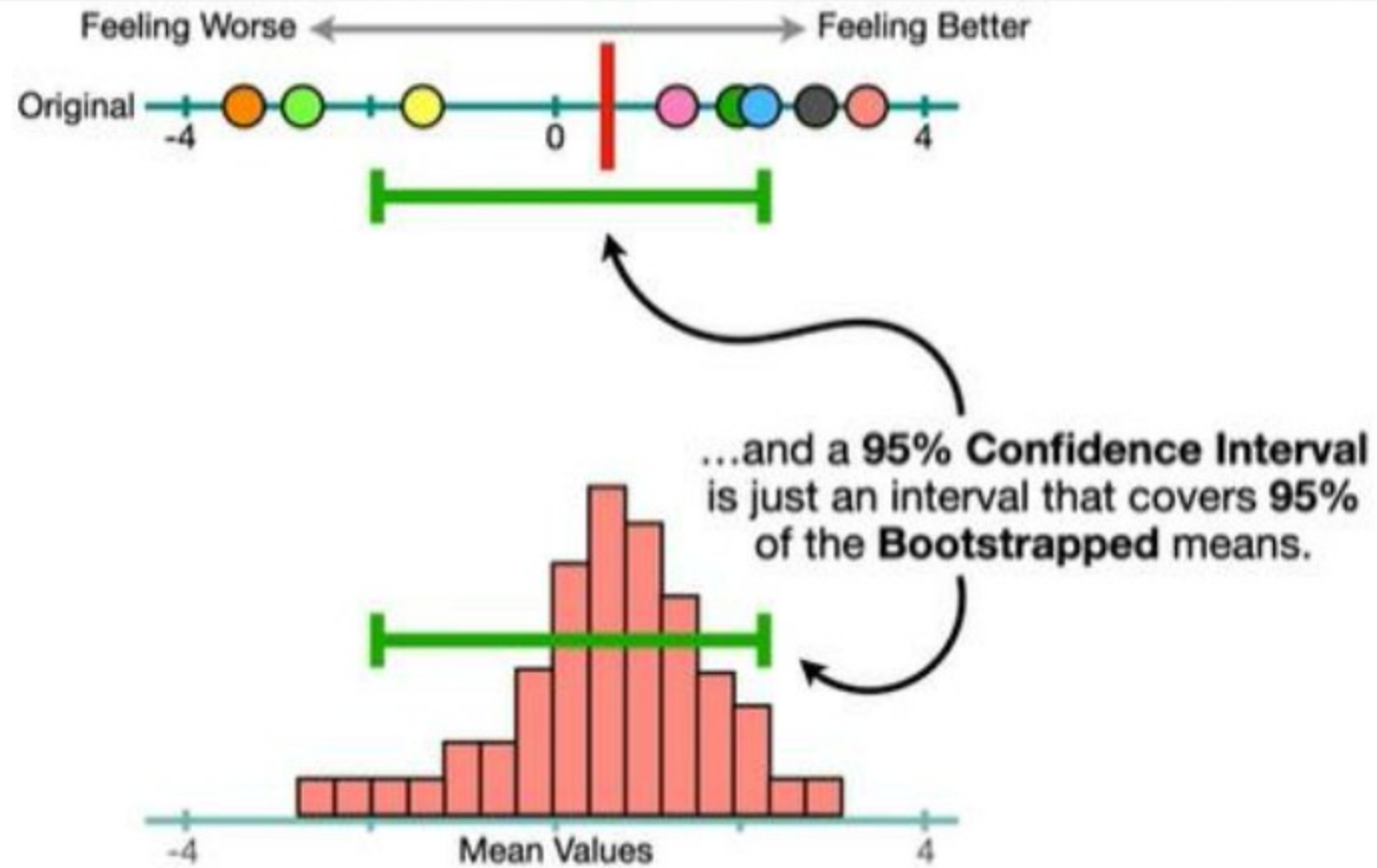
BOOTSTRAPPING - EXAMPLE



Randomly selecting data and allowing for duplicates is called **Sampling With Replacement**.



BOOTSTRAPPING - EXAMPLE



BOOTSTRAP USING R

- 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:

```
Bootstrap Statistics :
  original    bias    std. error
t1*      62000 -70.5595    209.1515
```

- THE ORIGINAL ESTIMATE OF THE MEDIAN IS \$62,000. THE BOOTSTRAP DISTRIBUTION INDICATES THAT THE ESTIMATE HAS A BIAS OF ABOUT -\$70 AND A STANDARD ERROR OF \$209.

RESAMPLING VERSUS BOOTSTRAPPING

- SOMETIMES THE TERM RESAMPLING IS USED SYNONYMOUSLY WITH THE TERM BOOTSTRAPPING, AS JUST OUTLINED.
- MORE OFTEN, THE TERM RESAMPLING ALSO INCLUDES TEST, WHERE MULTIPLE SAMPLES ARE COMBINED AND THE SAMPLING MAY BE DONE WITHOUT REPLACEMENT.
- IN ANY CASE, THE TERM BOOTSTRAP ALWAYS IMPLIES SAMPLING WITH REPLACEMENT FROM AN OBSERVED DATA SET.