

1 Lesson 2 Example 4

The “bootstrap” is a statistical method for generating a new data set that is like an existing one. Suppose we have a data set consisting of 6 observations: $x_1, x_2, x_3, x_4, x_5, x_6$. To generate a “bootstrap” data set, we sample from the original data set with replacement, meaning that it is possible for each observation to be sampled more than once. Examples of bootstrap data sets include:

$$x_4, x_2, x_4, x_3, x_2, x_4$$

$$x_3, x_1, x_6, x_1, x_1, x_2$$

$$x_2, x_1, x_4, x_3, x_6, x_5$$

2 Answer

To determine the probability that a bootstrap sample contains each observation exactly once, we proceed as follows:

2.1 Total Number of Possible Bootstrap Samples:

Since each of the 6 positions in the bootstrap sample can be filled by any of the 6 observations (with replacement), the total number of possible bootstrap samples is:

$$6^6$$

2.2 Number of Favorable Outcomes (Each Observation Appears Exactly Once):

If each observation appears exactly once, we are simply permuting the 6 observations. The number of such permutations is:

$$6!$$

$$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

2.3 Probability Calculation:

The probability that no observation occurs more than once in the bootstrap sample is the ratio of the number of favorable outcomes to the total number of possible bootstrap samples:

$$P(\text{each observation appears exactly once}) = \frac{6!}{6^6}$$

Simplifying the expression:

$$P(\text{each observation appears exactly once}) = \frac{720}{46656} = \frac{5}{324}$$

Therefore, the probability that a bootstrap sample contains each observation exactly once is:

$$P(\text{each observation appears exactly once}) = \frac{5}{324} \approx 0.0154$$