

1. Problem

I have a sample X_1, \dots, X_{60} . I generate one naive bootstrap sample X_1^*, \dots, X_{60}^* .

Let N be the number of times the first observation will be copied in the bootstrap sample.

Find the probability $\mathbb{P}(N = 2)$.

2. Problem

I have a sample X_1, \dots, X_{60} . I generate one naive bootstrap sample X_1^*, \dots, X_{60}^* .

Let L be the number of initial observations missing in the bootstrap sample.

Find the expected value $\mathbb{E}(L)$.

3. Problem

We have data of an AB experiment: $\bar{X}_a = 5.1$, $\bar{X}_b = 5.6$, $n_a = 18$, $n_b = 12$, $\sum (X_i^a - \bar{X}_a)^2 = 840$, $\sum (X_i^b - \bar{X}_b)^2 = 820$.

Calculate the estimate of variance of $\bar{X}_a - \bar{X}_b$ for the Welch test.

4. Problem

Variables X_1, X_2, \dots, X_{20} are independent and normally distributed $\mathcal{N}(5; 3)$.

Calculate the variance

$$\text{Var} \left(\sum_{i=1}^{20} (X_i - \bar{X})^2 \right).$$

5. Problem

I have results of two runners A and B: 2 results for A and 3 results for B. The running time for both runners are continuously distributed and their distribution are equal.

What is the probability that the maximal rank of running times of the runner A will be equal to 4?

6. Problem

I have results of two runners A and B: 2 results for A and 3 results for B. The running times for both runners are continuously distributed.

Consider the Mann-Whitney test statistic U_a that positively depends on the rank sum for the runner A.

Find the expected value $\mathbb{E}(A)$ under the null-hypothesis of equal running time distributions.

7. Problem

Consider three variables: target variable y_i , predictor x_i and indicator of treatment $z_i \in \{0, 1\}$.

The treatment z_i was designed to be independent of x_i , but in fact $x_i = f_i \cdot (1 + 0.08z_i)$.

We suppose that z_i are Bernoulli with $p = 0.4$, $f_i \sim \mathcal{N}(900; 9)$ and they are independent.

Find the probability limit

$$\text{plim} \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{n - 1}.$$

8. Problem

Consider three variables: target variable y_i , predictor x_i and indicator of treatment $z_i \in \{0, 1\}$. The treatment z_i was assigned independently of x_i , total $n = 100$.

The sample covariance matrix C is provided. The order of variables is y, x and z . For example the sample covariance of x and z is 0.26.

$$C = \begin{pmatrix} 7.09 & 2.64 & 0.75 \\ 2.64 & 1.17 & 0.26 \\ 0.75 & 0.26 & 0.25 \end{pmatrix}$$

Consider CUPED with first regression given by $\hat{y}_i = \hat{\alpha}_1 + \hat{\alpha}_2 x_i$ with residuals $r_i = y_i - \hat{y}_i$. What is the sample covariance of r and z ?

9. **Problem**

Vasiliy uses knn algorithm with 1 neighbor to match observations. Here z is treatment indicator, x is predictor and y is target variable.

id	y	z	x
1	28	0	5
2	20	0	11
3	21	0	19
4	28	0	37
5	20	1	48
6	22	1	12
7	20	1	4
8	25	1	17

Which individual will be matched with individual number 2?

10. **Problem**

I ate five M&Ms: red, green, green, red and yellow. Only these three colors are possible. I assume that $p_{yellow} = 2p_{green}$.

What is the estimated probability of yellow color?