Part I. Only numeric answer is checked, you can attempt all questions but the maximal grade is 4 points.

1. I have a sample $X_1, \ldots, X_{90}$. I generate one naive bootstrap sample $X_1^*, \ldots, X_{90}^*$.

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

   What is the probability that the first observation will be present in the bootstrap sample more than once?

2. We have data of an AB experiment: $\bar{X}_a = 5.4$, $\bar{X}_b = 6$, $n_a = 18$, $n_b = 15$, $\sum(X_i^a - \bar{X}_a)^2 = 890$, $\sum(X_i^b - \bar{X}_b)^2 = 800$.

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

3. I have five results of two runners A and B for the 5 km race:

   16:49 (B), 21:17 (A), 18:30 (B), 6:18 (B), 20:16 (A), 15:39 (B).

   Calculate Mann-Whitney statistic U_A that tests the null-hypothesis of equal distributions of time. (The statistic U_A should positively depend on the ranks of the runner A).

4. I have 30 hypothesis with independent statistics. The null hypothesis for all 30 cases is actually true, but I don't know this.

   I calculate all p-values. If the 4 lowest p-value are simultaneously lower than 0.01 I wrongly conclude that not all H_0 are true. Otherwise I correctly conclude that all H_0 are true.

   What is the probability that I will get the correct conclusion?

5. My target variable is binary and I wish minimal detectable effect equal to 0.04, probability of I-error not greater than 0.01, probability of II-error not greater than 0.3. The control and experimental group are of the same size equal to n.

   Which minimal value of n is sufficient in the worst case?

6. Vasiliy loves to eat shaurma. He has 6 local shaurma dealers. Vasiliy bought 5 shaurmas from each dealer and measured their weight. He would like to test the hypothesis that mean weight is the same for all dealers.

   Total sum of squares is 700, between sum of squares is 200 Calculate the F -statistic to test the hypothesis.

Part II. Full solution is required, you can attempt all questions but the maximal grade is 6 points.

7. Let random variables be iid uniform $U[0; 1]$. Consider the original sample Y_1, ..., Y_n and naive bootstrap sample Y_1^*, ..., Y_n^*.

   Let N_1 be the number of copies of Y_1 in the bootstrap sample.

   Find the distribution of N_1 when $n \to \infty$.

8. I have invented the GCA, *garbage component analysis*. It is completely similar to PCA with one exception: you *minimize* the variance of components. First you find the component with minimal sample variance. Than you find the second component that is orthogonal to the first one with minimal variance. And so on.

   How my new garbage components are related to classic principal components?

9. Winnie-the-Pooh simulteneously tests h null hypothesis using independent samples. All the null hypothesis are true but Winnie does not know it.

   a) What is the expected value of the second lowest P-value?

   b) What is the variance of the number of wrongly regected hypothesis if Winnie rejects all the hypothesis with P-value less 0.05?

10. There are three continuosly distributed samples of the same size n, X_1, ..., X_n, Y_1, ..., Y_n, Z_1, ..., Z_n. Imagine that the null hypothesis that all samples have the same distribution is true.

    Consider the random variable R_X — the sum of ranks of the X sample in the pooled sample.

    Find the correlation $\mathrm{Corr}(R_X, R_Y)$.