Rules: two parts, two hours in total, one A4 cheat sheet is allowed.

Part 1. Test part, only numerical answers are checked, 6 questions, each question gives 1 point but no more than 4 points in total. This part is very predictable :)

1. (bootstrap) I have a sample X_1, ..., X_{100}.

   I generate one naive bootstrap sample X_1^*, ..., X_{100}^*.

   What is the probability that the first observation will be present in the bootstrap sample 2 times or more?

2. (welch) We have data for an AB-experiment $\bar{X}_a = 10$, $\bar{X}_b = 12$, $n_a = 20$, $n_b = 30$, $\sum(X_i^a - \bar{X}_a)^2 = 100$, $\sum(X_i^b - \bar{X}_b)^2 = 200$.

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

3. (mw test) I have five results of two runners A and B for the 5 km race: 25:12 (A), 26:34 (B), 27:43 (A), 28:12 (A), 29:05 (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. (multiple comparison) I have 100 hypothesis with independent statistics. The null hypothesis for all 100 cases is actually true, but I don't know this.

   I calculate all p-values. If the two lowest p-value are both lower than 0.05 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. (sample size) My target variable is binary and I wish minimal detectable effect equal to 0.01, probability of I-error not greater than 0.02, probability of II-error not greater than 0.10, control and experimental group of the same size equal to n.

   What is minimal value of n?

6. (anova 1+2) Vasiliy loves to eat shaurma. He has three local shaurma dealers. Vasiliy bought 7 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 1000, between sum of squares is 500.

   Calculate the F-statistic to test the hypothesis.

Part 2. Open part, solutions are required, 4 problems, each problem gives 2 points but no more than 6 points in total. This part is almost unpredictable :)

1. Let random variables Y_1, ..., Y_n be iid uniform $U[0;1]$. Consider the naive bootstrap sample Y_1^*, ..., Y_n^*.
   Find $\text{Var}(Y_1^*)$, $\text{Cov}(Y_1^*, Y_2^*)$, $\text{Var}(\bar{Y}^*)$.

2. 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 probability that the highest P-value will be greater than 0.95?

   (b) What is the possible range for the probability in point (a) if exactly one null hypothesis is false?

3. The correlation matrix of standardized variables a, b and c is given by

$$C = \begin{pmatrix} 1 & 0.2 & 0 \\ & 1 & 0.2 \\ & & 1 \end{pmatrix}$$

   Let p_1, p_2 and p_3 be the principal components.

   (a) Express p_1 in terms of a, b and c.

   (b) Express b in terms of p_1, p_2 and p_3.

   (c) How would you restore the second observation of variable b if you know that first and second components for the second observation are equal to -1 and 2 respectively?

4. Consider the Mann-Whitney test with possible ties. The variables X_1, X_2, ..., X_{n_x} are iid Poisson with rate $\lambda = 1$. The variables Y_1, Y_2, ..., Y_{n_y} are iid Poisson with the same rate, independent from X sample.
   Let L be the number of all pairs (X_i, Y_j) such that $X_i > Y_j$.

   (a) Find $\mathbb{E}(L)$, $\text{Var}(L)$.

   (b) What is the probability that the ordered sequence of all X_i and Y_j will start with three or more members from X-sample?