1. A/B testing. (25 points)

   Consider two samples with sizes $n_x = n_y = 50$ from:

   - two normal distributions with unit variances, means $\mu_x = 0$ and varying μ_y;
   - two normal distributions with variances one and two, means $\mu_x = 0$ and varying μ_y;
   - any other distributions (not normal) of your choice. Here you should vary one parameter — location of the second distribution.

   We would like to test the hypothesis H_0: $\mu_x = \mu_y$ against H_a: $\mu_x \neq \mu_y$ at nominal significance level 5%.

   (a) For $\mu_y = 0$ calculate the real significance level of t-test, Welch test and Mann-Whitney test. Choose a grid of positive μ_y and calculate real power of these tests. Use 10000 simulations for each μ_y.

   (b) For each case provide a plot that describes how real significance level and real power do change with μ_y.

   (c) Which test is more appropriate in each case?

   (d) How will the answers change with $n_x = n_y = 1000$?

2. Bootstrap (25 points)

   Use the dataset `SIC33.csv` with the following variables:

   - *output* — Value added.
   - *labor* — Labor input.
   - *capital* — Capital stock.

   Obtain OLS estimates of the model

   $$\ln Y_i = \beta_1 + \beta_K \ln K_i + \beta_L \ln L_i + u_i.$$

   Use the paired bootstrap with $B = 10000$.

   (a) Obtain a bootstrap estimate of the standard error of $\hat{\beta}_K$.

   (b) Use this standard error estimate to test H_0: $\beta_K = 0$ against H_a: $\beta_K \neq 0$.

   (c) Provide three 95% CI for β_K: classic, heteroskedasticity robust (HC1), naive bootstrap.

   (d) Provide 95% bootstrap CI for the product $\beta_K \cdot \beta_L$.

   (e) Estimate the same model using median (absolute value) regression. The target function is

   $$\min_{\hat{\beta}} \sum_{i=1}^{n} \left| \ln Y_i - \ln \hat{Y}_i \right|.$$

   Provide 95% bootstrap CI for each β.

3. CUPED. (25 points)

The data generating process is:

$$\begin{cases} n = 200; \\ x_i \sim \mathcal{N}(1; 1); \\ (u_i \mid x_i) \sim \mathcal{N}(0; 1); \\ (d_i \mid x_i, u_i) \sim \text{Bernoulli}(1/2); \\ y_i = 3 + 2x_i^2 + \delta d_i + u_i \end{cases}$$

Here x_i is a characterestic of an individual, $d_i = 1$ for those in experimental group, and $d_i = 0$ for those in control group, y_i is the target variable. We are interested in estimating the effect δ. Consider three strategies:

- SIMPLE. Regression with a standard confidence interval for δ:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\delta} d_i;$$

- CUPED-A. Two stages:

Step I. Estimate regression and calculate residual $r_i = y_i - \hat{y}_i$.

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i;$$

Step II. Regress r_i on d_i and build a standard CI for δ:

$$\hat{r}_i = \hat{\alpha} + \hat{\delta} d_i.$$

- CUPED-B. Two stages:

Step I. Estimate regression and calculate quasi-residual $r_i = y_i - (\hat{\beta}_1 + \hat{\beta}_2 x_i)$.

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\delta} d_i;$$

Step II. Regress r_i on d_i and build a standard CI for δ:

$$\hat{r}_i = \hat{\alpha} + \hat{\delta} d_i.$$

Your task is:

(a) Simulate the data set 1000 times for each of these deltas: $\delta = 0$, $\delta = 0.1$, $\delta = 1$, $\delta = 10$.

(b) Calculate 95% nominal CI according to SIMPLE, CUPED-A and CUPED-B.

(c) For $\delta = 0$ provide the histogram of p-values to check H_0: $\delta = 0$ agains $\delta \neq 0$.

(d) What is the actual coverage probability for each method for each delta?

(e) For each delta for each method find: the average length of the interval, the number of times the interval was the shortest among three methods considered.