Home Assignment 2.

1. Set random seed to $M * 100 + D$, where M and D is your Month and Date of birth, respectively.

   Generate two samples X of size 20 and Y of size 200 s.t. $E(X) \sim Unif[0, 1], E(Y) \sim Unif[0, 1]$ from

   - two normal distributions with the same variance $\sigma_X^2 = \sigma_Y^2 = 0.1$;
   - two normal distributions with if M is odd $\sigma_X^2 = 0.1$ and $\sigma_Y^2 = 1$, else $\sigma_X^2 = 1$ and $\sigma_Y^2 = 0.1$
   - two normal distributions with 3 outliers added afterwards $Out_i \sim Exp(10) * (-1)^{U\{1,2\}}$;
   - two exponential distributions.

   (a) Conduct t-test, Welch test, Mann-Whitney test 10000 times.

   (b) Calculate nominal type I error rate and power (to calculate type I error fix mean for Y the same as for X).

   (c) For each case described above plot how nominal power changes for different values of parameters of the distribution.

   (d) Based on the plots, which test is the most appropriate in each of the cases. Will your answer change if both sample size are increased by 10.

2. Use dataset `SIC33.csv` with the following data:

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

   Consider the following model $\ln Y = \beta_0 + \beta_1 \ln(K) + \beta_1 \ln(L) + u$. Suppose we are interested in estimating the quantity $\gamma = \beta_1 + \beta_2$ from the data. Test $H_0 : \gamma = 1$ against $H_a : \gamma \neq 1$ using:

   (a) Jackknife estimator;

   (b) paired bootstrap;

   (c) wild bootstrap (bootstrapping the residuals).

   How the procedure should be changed in case u is heteroskedastic?

3. Consider K and L from the same dataset `SIC33.csv`.

   (a) Illustrate principal components on a scatter plot.

   (b) Calculate the explained variance by each component, the weight of K in the 1st component and the weight of 2nd component in the reconstruction of L.

   (c) Repeat the analysis for $\ln(K)$ and $\ln(L)$.

Deadline: **2022-12-18, 21:00**.