HSE and University of London Double Degree Programme in Data Science and Business Analytics

Statistical Methods for Market Research, 2023-2024

Home assignment 1.

Problem 1. Sample size, MDE, power

Consider a test for testing the null hypothesis that a proportion is equal to some constant $p_0 = 1/2$:

$$H_0 : p = 1/2$$

(a) Calculate minimum sample size needed to determine confidence interval with relative precision of 8%, i.e. $\{p \pm 0.08 * p\}$.

(b) Derive the formula for MDE as a function of power $(1 - \beta)$, type I error (α), number of observations (N). Calculate MDE in case the researcher has a sample of 500 observations, desired power is 60% and significance level is 5%.

(c) Based on the derivations from (b), find what will be the power of a test in case true value of p is known: $p = 1/4$. Sample size is 500 observations and significance level is 5%.

Problem 2. UoL exam. Sampling methods

Case Study:

Cryptocurrencies, such as Bitcoin and Ethereum, are digital currencies which employ encryption techniques to regulate the generation of units of currency and verify the transfer of funds, operating independently of a central bank. Volatile price movements in the past year have resulted in financial regulators leading calls for cryptocurrencies to be reined in for fear of contagion risks if they enter the financial mainstream.

Recently, some retail banks have banned customers from trading in cryptocurrencies using credit cards (due to the bank's liability in the event of default), but still permit transactions using debit cards (as a customer's own funds are used), although some banks are considering also banning the use of debit cards to trade as well to safeguard against customers realising large losses.

One bank is considering imposing a cap on debit card transactions rather than an outright ban (for fear of losing customers to competitors). The bank has invited you to devise an appropriate sampling scheme to research attitudes of its customers regarding the introduction of a cap, including the level of any cap.

Consider the following sampling techniques:

i. Simple random sampling.

ii. Quota sampling.

iii. Stratified sampling.

iv. Cluster sampling.

Explain how each of these sampling methods could be applied to this case study. Describe the merits and limitations of each of these methods in light of the given case·.

Problem 3. A/B testing. Power analysis simulation study

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

The following parameters should be **fixed** for the whole experiment: $N_x = 20 -$ sample size of X, $N_y = 20 -$ sample size of Y, generate means for sample X and sample Y from the uniform distribution $E(X) \sim Unif[0,0.5]$, $E(Y) \sim Unif[0.5,1]$ (use one value across the whole experiment).

You are to generate samples X and Y from different distributions

- (i) two normal distributions with the same variance $\sigma_X^2 = \sigma_Y^2 = 0.1$ with 3 outliers $Out_i \sim Exp(10) * (-1)^{U\{1,2\}}$ added afterwards;

- (ii) two normal distributions with: $-$ if M is odd $\sigma_X^2 = 0.1$ and $\sigma_Y^2 = 1$, $-$ if M is even $\sigma_X^2 = 1$ and $\sigma_Y^2 = 0.1$

- (iii) two exponential distributions (derive parameter based on $E(X), E(Y)$ generaed above).

Answer the following questions for each of the cases separately:

(a) Calculate nominal type I error rate (to calculate type I error set mean for Y the same as for X) for t-test, Welch test, Mann-Whitney test and bootstrap (use any appropriate bootstrap technique to test the hypothesis about equality of two means). *Is nominal type I error rate controlled at 5% significance level? Discuss the obtained results.*

(b) Calculate nominal power for the same 4 tests for 5% significance level. *Which method is better?*

(c) Calculate nominal power for the 4 same tests but control nominal type I error rate for all of the tests on the same level – exactly at 5% (you can use simple binary search). *Which method is better given that type I error is controlled?*

Your report should consist of
(1) the experiment setting with sample distribution parameters being specified
(2) (10 points) the following table:

	Nominal type I error			Nominal Power			Nominal Power given that type I error is controlled		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
t-test									
Welch									
MW									
Bootstrap									

(3) (10 points) Answers to questions posed in (a)-(c)
(4) (10 points) Summary of the results: which test is the most appropriate in each of the cases (i),(ii),(iii)?

Your code is NOT graded, only your report.

Problem 4. Difference-in-difference approach

This assignment asks you to replicate some of the findings from Sheridan and Ball [Sheridan, Ball, 2005].

Source data file: $Inflation\_Targeting.xlsx$. Please note that there are two separate sheets (for developed countries and for developing countries).

(a) Compare average inflation rates in target countries before and after inflation targeting. Based on this result, can we conclude about the impact of inflation targeting on the inflation rate?

(b) Now apply the difference-in-differences method by estimating the parameters of equation (1) from the paper by Sheridan and Ball [Sheridan, Ball, 2005, pp. 249-276]. Estimate for three samples:

- complete selection of countries;
- the developed countries;
- developing countries.

Interpret the results, explain what can be said about the impact of the transition to inflation targeting on the inflation rate in the long term, based on the obtained parameter estimates?