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

Statistical Methods for Market Researh, 2023-2024

Classteacher: Ksenia Kasianova.

Class 7: Preparations

Task 1

UoL Exam

Suppose we are interested in estimating the mean of a population with a finite variance using a simple random sample of size n.

i. State a suitable estimator of the population mean as well as its sampling distribution. Mention any assumptions which you make.

ii. Explain how the sampling distribution derived in i. should be interpreted.

iii. Explain how to determine the minimum sample size necessary to estimate a population mean to within e units assuming the population standard deviation is known. If the population standard deviation was unknown, how would you deal with this?

iv. Explain the purpose of the finite population correction factor (including a formula) and when it should be used.

Task 2

Suppose we are interested in estimating the proportion of a population using a simple random sample of size n.

i. Explain what a sampling distribution is.

ii. State a suitable estimator of the population proportion as well as its sampling distribution. Mention any assumptions which you make.

iii. Explain statistically how to determine the minimum sample size necessary to estimate a population proportion to within d units.

iv. When constructing a confidence interval for a proportion, which level of confidence would you propose and why?

Task 3

i. Define the terms 'incidence rate' and 'completion rate'.

ii. Explain how you would adjust the statistically-determined sample size, n, in light of incidence and completion rates.

Task 4

A researcher analyzes the impact of a law prohibiting the sale of alcohol after 11 p.m. on alcohol consumption. The researcher has information on per capita alcohol consumption in eight regions in 2014 and 2015. In 2014, alcohol was sold without restrictions in all regions. In 2015, this law was introduced in regions A, B, C, D, but in other regions it was not applied. Data on alcohol consumption (liters per person per year) are given in the table.

Region	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***
2014	6	6	8	4	4	3	3	2
2015	6	8	9	5	6	5	5	4

a. The researcher uses a fixed effects model to estimate the impact of interest: $y_{it} = \beta \cdot x_{it} + \alpha_i + \varepsilon_{it}$, where α_i- is the fixed effect of the i-th region ; x_{it} is a dummy variable equal to one if there was a law restricting the sale of alcohol in the i-th region in year t, and equal to zero otherwise; y_{it} - alcohol consumption per capita in the i-th region in the year t. Using the within-group transformation, find the estimate of the parameter β and interpret the result.

b. Now estimate the effect of the alcohol restriction law using the difference-in-differences method. Interpret the result. Provide a graphic illustration of the solution (do not forget to indicate in the figure the coordinates of all key points, as well as the magnitude of the impact effect).

c. What could cause such a discrepancy in estimates?