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 4: Sampling

Task 1

UoL Exam

(a)

Transport for London is the local government body responsible for most aspects of the transport system in London. Its role is to implement the transport strategy and to manage transport services across London.

TfL launched the Night Tube service on the London Underground in August 2016. Currently, the Night Tube operates on Friday and Saturday nights across five lines (the Central, Jubilee, Northern, Piccadilly and Victoria lines). TfL management would consider expansion of the service to more lines and/or other nights of the week, subject to consumer demand.

Some 1.34 billion passenger journeys are made each year on London Underground (albeit with many individuals making multiple journeys), however there is no sampling frame with full coverage of all distinct passengers available. Despite this limitation, TfL management wants to conduct market research to assess consumer appetite for an extension to the Night Tube service, which may have fare implications. At present Night Tube fares are classified as off-peak, although a special night-off-peak fare could be introduced, depending on passengers' price elasticities.

TfL has invited you to devise an appropriate sampling scheme to research attitudes towards developing the Night Tube. Explain in detail how each of the following sampling methods could be applied to the overall sampling strategy. Make sure you describe the merits and limitations of each as well as how each would be applied in practice.

i. Convenience sampling.

ii. Simple random sampling.

iii. Systematic sampling.

iv. Stratified sampling.

(b)

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

(a)

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. Explain in detail how each of the following sampling methods could be applied to the overall sampling strategy. Make sure you describe the merits and limitations of each as well as how each would be applied in practice.

i. Quota sampling.

ii. Snowball sampling.

iii. Stratified sampling.

iv. Cluster sampling.

(b)

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

(a)

A government is considering the need for additional airport capacity due to projections of increased demand for air travel in the years ahead. The government is deciding how the expected demand can be met in the long term. Short-listed options for increasing airport capacity include expansion of one of two existing airports, A or B. The decision to expand either site involves numerous trade-offs.

You have been asked to devise an appropriate sampling scheme of airport A and airport B users (passengers as well as non-passengers, such as staff and local residents to the airport) to research their views of expanding one or other site. Explain how each of the following sampling methods could be applied to the overall sampling strategy for this study. Make sure you describe the merits and limitations of each as well as how each would be applied in practice.

Variant A

i. Convenience sampling.

ii. Quota sampling.

iii. Stratified sampling.

iv. Cluster sampling.

Variant B

i. Judgemental sampling.

ii. Snowball sampling.

iii. Systematic sampling.

iv. Cluster sampling.

(b)

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

Variant A

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

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

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

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

Variant B

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

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

iii. Explain what the term '95% confidence interval' means and give a market research example.

iv. Explain what the finite population correction is (including a formula) and when it should be used.

Task 4

(a)

A travel agency offers customers a range of ways to make holiday bookings – in store, online and through their call centres. Revenues are generated through different commission rates on holidays paid by tour operators (the actual suppliers of the holiday products). The company prides itself on delivering customer satisfaction. Therefore, management is keen to research customer satisfaction levels across the different booking methods and by tour operators (some tour operators offer higher commission margins).

To determine the level of customer satisfaction, the company's management has decided to use a survey of all types of customers and have asked you to devise an appropriate sampling scheme. Explain in detail how each of the following sampling methods could be applied to the overall sampling strategy for this study. Make sure you describe the merits and limitations of each as well as how each would be applied in practice.

i. Quota sampling

ii. Snowball sampling

iii. Systematic sampling

iv. Stratified sampling.

(b)

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

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

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

iii. Provide a practical marketing example of a 95proportion.

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

Task 5

(a)

As countries around the world embark on rolling out their respective Covid-19 vaccination programmes, many 'physical' businesses (i.e. those not online) are eager for customers to return. A global cinema chain which operates in many countries worldwide recognises that, despite the vaccines, some people may remain concerned about watching a new film release in one of their cinemas.

The company has a database of members of its loyalty scheme, many of whom have not visited a cinema for over a year. While the company wants these customers to return, it also wants to win back 'casual customers', i.e. those who might visit a cinema only occasionally.

To better understand potential customer anxieties, the company's management has decided to use a survey of all types of customers and has asked you to devise an appropriate sampling scheme. Explain in detail how each of the following sampling methods could be applied to the overall sampling strategy for this study. Make sure you describe the merits and limitations of each as well as how each would be applied in practice.

Variant A

i. Convenience sampling.

ii. Snowball sampling.

iii. Systematic sampling.

iv. Stratified sampling.

Variant B

i. Judgemental sampling.

ii. Quota sampling.

iii. Simple random sampling.

iv. Cluster sampling.

(b)

Suppose we are interested in estimating the proportion of a population using a simple random sample of size n. In your own words, answer the following.

Variant A

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 e units.

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

Variant B

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

ii. Explain statistically how to determine the minimum sample size necessary to estimate a population mean to within e units with 90% confidence.

iii. When several parameters are being estimated, what is the procedure for determining the sample size?

iv. Explain the difference between the incidence rate and the completion rate in survey sampling.