IAQS | INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Subject: Actuarial Model

Chapter: Unit 3 & 4

Category: Practice Paper Solutions

Q1.

1. a) minimum expenditure of time and money.

2. c) blocking

3. d) Studying the data collection process and the collected data.

4. a) Stratified sampling

5. c) Censored

6. a) efficiency

7. c) cross-sectional

8. d) big data

9. c) unbiased

10. a) Longitudinal

Note: These are indicative solutions, you should answer in detail of the question asks for.

11. Other aspects of the data which are determined by the collection process, and which affect the way it is analysed include the following:

i. Cross-sectional data involves recording values of the variables of interest for each case in the sample at a single moment in time.

ii. Longitudinal data involves recording values at intervals over time.

iii. Censored data occurs when the value of a variable is only partially known, for example, if a subject in a survival study withdraws, or survives beyond the end of the study: here a lower bound for the survival period is known but the exact value isn't.

iv. Truncated data occurs when measurements on some variables are not recorded so are completely unknown. For example: in case of claims above a benchmark which are passed to the reinsurer , the insurer only knows that the claim amount is above the benchmark however it does not know the exact amount. This is a truncated data.

v. Large or Big data. The term big data is not well defined but has come to be used to describe data with characteristics that make it impossible to apply traditional methods of analysis A data can be classified as large data on different grounds like size- eg large observations, many attributes, speed –eg collected every second, variety – eg data from different sources in different structural variety. When a data is extremely large, sampling is suitable.

12. Experimental design is a way to carefully plan experiments in advance so that your results are both objective and valid.

a) Ideally, your experimental design should:

i. Describe how participants are allocated to experimental groups. A common method is completely randomized design, where participants are assigned to groups at random. A second method is randomized block design, where participants are divided into homogeneous blocks (for example, age groups) before being randomly assigned to groups.

ii. Minimize or eliminate confounding variables (outside influences) which can offer alternative explanations for the experimental results.

iii. Allow you to make inferences about the relationship between independent variables and dependent variables.

iv. Reduce variability, to make it easier for you to find differences in treatment outcomes.

b) Design of experiments involves:

i. The systematic collection of data

ii. A focus on the design itself, rather than the results

iii. Planning changes to independent (input) variables and the effect on dependent variables or response variables

iv. Ensuring results are valid, easily interpreted, and definitive.

13.

i. In general, this is accomplished by solving an optimization problem in which the objective function (the function being minimized or maximized) relates the response variable and the functional part of the model containing the unknown parameters in a way that will produce parameter estimates that will be close to the true, unknown parameter values.

ii. The unknown parameters are, loosely speaking, treated as variables to be solved for in the optimization, and the data serve as known coefficients of the objective function in this stage of the modelling process. In theory, there are as many different ways of estimating parameters as there are objective functions to be minimized or maximized.

iii. However, a few principles have dominated because they result in parameter estimators that have good statistical properties. The two major methods of parameter estimation for process models are maximum likelihood and least squares. Both of these methods provide parameter estimators that have many good properties.

iv. Both maximum likelihood and least squares are sensitive to the presence of outliers, however. There are also many newer methods of parameter estimation, called robust methods, that try to balance the efficiency and desirable properties of least squares and maximum likelihood with a lower sensitivity to outliers.

v. Although robust techniques are valuable, they are not as well developed as the more traditional methods and often require specialized software that is not readily available. Maximum likelihood also requires specialized algorithms, in general.

14.

1. Define the problem and the questions to be addressed. Before data collection begins, specific questions that the researcher plans to examine must be clearly identified. In addition, a researcher should identify the sources of variability in the experimental conditions. One of the main goals of a designed experiment is to partition the effects of the sources of arability into distinct components in order to examine specific questions of interest.

2. Define the population of interest. A population is a collective whole of people, animals, plants, or other items that researchers collect data from. Before collecting any data, it is important that researchers clearly define the population. The designed experiment should designate the population for which the problem will be examined. The entire population for which the researcher wants to draw conclusions will be the focus of the experiment.

3. Determine the need for sampling. A sample is one of many possible sub-sets of units that are selected from the population of interest. In many data collection studies, the population of interest is assumed to be much larger in size than the sample. The results from a sample are then used to draw valid inferences about the population. A random sample is a sub-set of units that are selected randomly from a population. A random sample represents the general population or the conditions that are selected for the experiment because the population of interest is too large to study in its entirety. Using techniques such as random selection after stratification or blocking is often preferred.

4. Define the experimental design. A clear definition of the details of the experiment makes the desired statistical analyses possible, and almost always improves the usefulness of the results.

15. The most important principles are:

i. Randomization: the assignment of study components by a completely random method, like simple random sampling. Randomization eliminates bias from the results

ii. Replication: the experiment must be replicable by other researchers. This is usually achieved with the use of statistics like the standard error of the sample mean or confidence intervals.

iii. Blocking: controlling sources of variation in the experimental results

16.

i. In the design of any investigation, consideration of issues related to data security, privacy and complying with relevant regulations should be paramount. Data security means protecting digital data, such as those in a database, from destructive forces and from the unwanted actions of unauthorized users, such as a cyberattack or a data breach.

ii. Data privacy relates to how a piece of information—or data— should be handled based on its relative importance. Data privacy has always been important.

iii. For instance A single company may possess the personal information of millions of customers—data that it needs to keep private so that customers' identities stay as safe and protected as possible, and the company's reputation remains untarnished.

iv. Another point to be aware of is that just because data has been made available on the internet, doesn't mean that that others are free to use it as they wish. This is a very complex area and laws vary between jurisdictions.

17. Stratified sampling will be used.

i. Stratified sampling refers to a type of sampling method. With stratified sampling, the analyst divides the population into separate groups, called strata. In our case each acreage group is a strata

ii. Then, a probability sample (often a simple random sample) is drawn from each group.

iii. Stratified sampling has several advantages over simple random sampling.

iv. For example, using stratified sampling, it may be possible to reduce the sample size required to achieve a given precision. Or it may be possible to increase the precision with the same sample size.

18.

i. Truncated data occurs when measurements on some variables are not recorded so are completely unknown.

ii. Truncated data are data for which measurements are only reported if they fall above a lower bound, below an upper bound, or between a lower and upper bound. Truncated data may be modeled in Stan using truncated distributions.

iii. In the given case, the individual's exact income is unknown. Only the bounds or range where the income is falling is known. Truncation can be modeled using some truncated distributions.

iv. The truncated count data estimators and their properties are reviewed. Statistics of interest are therefore unbiased by the specification search, and the prediction results allow comparison of the statistical models' robustness.

v. The new estimators are found to be more appropriate for estimating and predicting demand and social benefits than the alternative estimators based on a variety of criteria.