IAQS | INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Subject: Actuarial Model

Chapter: Unit 1 & 2

Category: Practice Paper Solutions

1. d) All of the above

2. a) deterministic

3. b) data

4. c) correlated

5. a) broad range of functions can be fit

6. c) inefficient

7. a) potential outliers

8. d) Sensitivity to outlier

9. b) False

10. a) Normal

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

11)

i. Nonlinear least squares regression extends linear least squares regression for use with a much larger and more general class of functions. Almost any function that can be written in closed form can be incorporated in a nonlinear regression model.

ii. Nonlinear regression is a regression in which the dependent variables are modelled as a non-linear function of model parameters and one or more independent variables. The reason that these models are called nonlinear regression is because the relationships between the dependent and independent parameters are not linear.

iii. As the name suggests, a nonlinear model is any model of the basic form,

$Y = f(x^{\rightarrow}\,;\, \beta^{\rightarrow}) + \varepsilon,$

in which,

a) the functional part of the model is not linear with respect to the unknown parameters, β0,β1,…, and

b) the method of least squares is used to estimate the values of the unknown parameters.

12)

i. Predictive analysis extends the principles behind inferential analysis in order for the user to analyze past data and make predictions about future events.

ii. It achieves this by using an existing set of data with known attributes (also known as features), known as the training set in order to discover potentially predictive relationships. Those relationships are tested using a different set of data, known as the test set, to assess the strength of those relationships.

iii. A typical example of a predictive analysis is regression analysis. The simplest form of this is linear regression where the relationship between a scalar dependent variable and an explanatory or independent variable is assumed to be linear and the training set is used to determine the slope and intercept of the line.

iv. A practical example might be the relationship between a car's braking distance against speed.

13) In assessing the suitability of a model for a particular exercise it is important to consider the following:

- The objectives of the modelling exercise.
- The validity of the model for the purpose to which it is to be put.
- The validity of the data to be used.
- The validity of the assumptions.
- The possible errors associated with the model or parameters used not being a perfect representation of the real-world situation being modelled.
- The impact of correlations between the random variables that 'drive' the model.
- The extent of correlations between the various results produced from the model.
- The current relevance of models written and used in the past.
- The credibility of the data input and results output.
- The dangers of spurious accuracy.
- The ease with which the model and its results can be communicated.
- Regulatory requirements.

14)

i. Proxy model is a mathematically or statistically defined function that replicates or approximates the response of the full-scale simulation model output for selected input parameters. They are the substitute to the complex numerical simulation by producing a meaningful representation of the complex system in a very short time.

ii. Proxy models, combined with design-of-experiment techniques, are used widely for sensitivity analysis. Application scenarios include the traditional one-parameter-at-a-time approach for linear-sensitivity analyses and advanced experimental designs that are capable of resolving correlation and higher-order effects.

iii. For probabilistic forecasting, proxy models are used routinely as input to a Monte Carlo sampling process. The high computational efficiency of proxy models enables exhaustive sampling rates.

15)

i. The assumption about the random error term is that its probability distribution remains the same for all observations of X and in particular that the variance of each error term is the same for all values of the explanatory variables, i.e the variance of errors is the same across all levels of the independent variables.

ii. This assumption is known as the assumption of homoscedasticity or the assumption of constant variance of the error term. If the assumption of homoscedastic disturbance (Constant Variance) is not fulfilled, following are the consequence:

a) We cannot apply the formula of the variance of the coefficient to conduct tests of significance and construct confidence intervals

b) The prediction would be inefficient, because of the variance of prediction includes the variance of error and of the parameter estimates which are not minimal due to the incidence of heteroscedasticity i.e. The prediction of Y for a given value of X based on the estimates β^'s from the original data, would have a high variance.

c) The estimates of the coefficients also would be inefficient. The presence of any drift, outliers or any other trends will distort the assumption of zero mean and constant variance.

16)

i. In order to build a statistical model, we need to be very careful in selecting the method. There are more general approaches and more competing techniques available for model building. There is often more than one statistical tool that can be effectively applied to a given modeling application.

ii. The large menu of methods applicable to modeling problems means that there is both more opportunity for effective and efficient solutions and more potential to spend time doing different analyses, comparing different solutions and mastering the use of different tools.

iii. In the process of developing the model we will often come across situations where we build a first basic model and then run the model, perform calculations and try to improvise.

17)

i. There are three main parts to every model. These are

a) the response variable, usually denoted by y, Y – sales of the company

b) the mathematical function, usually denoted as, $f(x^{\rightarrow}\ ;\ \beta^{\rightarrow})$, x – expenditure, β – slope of regression line

c) the random errors, usually denoted by €.

ii. The general or basic form is given as; Most of the models have this general form. $Y = f(x^{\rightarrow}\ ;\ \beta^{\rightarrow}) + \varepsilon$

In our case, using the expenditure of company we can predict the sales; the function will be

$Y(sales) = \alpha + \beta * x(expenses)$

iii. The random errors that are included in the model make the relationship between the response variable and the predictor variables a "statistical" one, rather than a perfect deterministic one. This is because the functional relationship between the response and predictors holds only on average, not for each data point.

Finally, $Y(sales) = \alpha + \beta * x(expenses) + \varepsilon(errors)$

18)

i. Data analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

ii. Random sampling is a part of the sampling technique in which each sample has an equal probability of being chosen. A sample chosen randomly is meant to be an unbiased representation of the total population. An unbiased random sample is important for drawing conclusions If for some reasons, the sample does not represent the population, the variation is called a sampling error. Data best reflects the population Via unbiased sampling Given that we can never determine what the actual random errors in a particular data set are, representative samples of data are best obtained by randomly sampling data from the population.

iii. Random sampling ensures that the act of data Random sampling is a part of the sampling technique in which each sample has an equal probability of being chosen. A sample chosen randomly is meant to be an unbiased representation of the total population. An unbiased random sample is important for drawing conclusions If for some reasons, the sample does not represent the population, the variation is called a sampling error.

iv. Data best reflects the population Via unbiased sampling Given that we can never determine what the actual random errors in a particular data set are, representative samples of data are best obtained by randomly sampling data from the population. Random sampling ensures that the act of data.

Sample survey has the following limitations:

a) Sample survey is not suitable if higher order accuracy is required.

b) If the items of the sample are not selected without any bias, the conclusions

may not be correct.

c) The investigator's personal bias regarding the choice of units and drawing of

sample may lead to false conclusions.

d) Sample investigation method is not suitable if the information is required about

each individual of the population.

e) Sample survey is a specialized technique and everybody cannot use it. Its use

requires specialized knowledge and trained personnel.