# Level 5 Diploma in Foundations of Data Science Statistical Methods using Excel (951) 177 Credits

| | |
|---|---|
| **Unit:** Sampling, Hypothesis Testing and Analysis of Variance<br><br>**Exam Paper No.:** 4 | **Guided Learning Hours:** 300<br><br>**Number of Credits:** 30 |
| **Prerequisites:** Business terms and Excel knowledge. | **Corequisites:** A pass or higher in Diploma in Analytics or equivalence. |

**Aim:** Sampling is the process of selecting a subset of observations from a population to make an inference about various population parameters such as mean, proportion and standard deviation. Sampling is necessary because even when the entire population is available, using the entire population for estimation of a population parameter may not be feasible. Confidence Interval is the interval estimate of a population parameter estimated from a sample using a specified confidence level.

In statistics, hypothesis is a claim made by a person/organisation, usually about population parameters such as mean or proportion. A **hypothesis** is a claim or belief and **hypothesis testing** is the statistical process of either **rejecting** or **retaining** the claim (by seeking evidence against the claim from a sample). *Analysis of Variance (ANOVA)* is used to understand the differences in population means among more than two populations.

On completion of the course, learners will be able to describe:
- Sampling and estimation
- Confidence interval
- Hypothesis
- Analysis of Variance (ANOVA)

| | |
|---|---|
| **Required Materials:** Recommended Learning Resources. | **Supplementary Materials:** Lecture notes and tutor extra reading recommendations. |
| **Special Requirements:** The unit requires a combination of lectures, demonstrations, discussions, and hands-on labs. | |

| Intended Learning Outcomes: | Assessment Criteria: |
|---|---|
| 1. Understand sources of data variations, sampling error and reasons why it is impossible to sample the entire population. | 1.1 Describe differences between sample and population.<br>1.2 Be able to explain related formulas.<br>1.3 Demonstrate using sampling in Excel.<br>1.4 Describe point estimate in statistics.<br>1.5 Describe Central Limit Theorem.<br>1.6 Demonstrate applying Central Limit Theorem in Excel. |
| 2. Understand the issues behind random sampling and how to deal with uncertainty in relation to creating confidence interval. | 2.1 Demonstrate the confidence interval formula.<br>2.2 Describe confidence interval for a proportion and the associated formula.<br>2.3 Demonstrate the motivation for creating confidence interval.<br>2.4 Be able to calculate confidence interval.<br>2.5 Be able to interpret confidence interval.<br>2.6 Explore real life examples of confidence interval implementation. |
| 3. Understand the importance of hypothesis testing, possible outcomes of the test and how to conduct the test. | 3.1 Define statistically significant, levels of confidence and levels of significance.<br>3.2 Describe null and alternative hypothesis.<br>3.3 Describe Type I and II errors.<br>3.4 Demonstrate performing different hypothesis tests in Excel. |

| | | 3.5 | Explore examples of real-life hypothesis testing examples. |
| | | 3.6 | Describe directional and non-directional hypothesis. |
| | | 3.7 | Demonstrate how to calculate the p-value of a hypothesis test t-score result. |
| 4. Understand Analysis of Variance (ANOVA) in determining significance differences between *means* of different variables. | | 4.1 | Describe one-way ANOVA. |
| | | 4.2 | Describe two-way ANOVA. |
| | | 4.3 | Demonstrate using ANOVA in Excel. |
| | | 4.4 | Explain ANOVA assumptions. |
| | | 4.5 | Describe repeated measures ANOVA. |
| | | 4.6 | Demonstrate conducting a nested ANOVA. |
| | | 4.7 | Describe ANOVA, ANCOVA, MANOVA and MANCOVA statistical methods. |
| | | 4.8 | Describe real life applications of ANOVA technique. |

**Methods of Evaluation:** A 2½ hour essay written paper with 5 questions, each carrying 20 marks. Candidates are required to answer all questions. Candidates also undertake project/coursework in **Sampling, Hypothesis Testing and Analysis of Variance** with a weighting of 100%.

## Recommended Learning Resources:  Sampling, Hypothesis Testing and Analysis of Variance

| **Text Books** | • Sampling in Statistics by Stephanie Glen. ISBN-13 : 979-8416512354<br>• Hypothesis Testing by Jim Frost.  ISBN-13 : 978-1735431154<br>• Making Sense of Z-Scores, T-Tests and ANOVA for Excel by Michael R. O'Brien.  ISBN-13 : 978-1976394485 |
| **Study Manuals** | BCE produced study packs |
| **CD ROM** | Power-point slides |
| **Software** | Excel |