Data Mining & Predictive Analytics (CSE 4859)
Quiz 3

Date: December 14, 2025 **Section: 19**
Maximum Marks: 12 **Time: 8-9 AM**

Use of mobile phones is strictly prohibited. For computations, only use scientific calculators.

1. What do you mean by a parameter? Point out the key differences between population parameters and sample statistics. [2]

2. Construct a flow-chart of the hypothesis testing mechanism for an unknown population parameter μ with the significance level α. Assume that necessary sample data are supplied as inputs. [2]

3. The duration of customer service calls to an insurance company is normally distributed, with mean 20 minutes, and standard deviation 5 minutes. For a sample of size 100, construct a 95% confidence interval for the population mean duration of customer service calls. [2]

4. Suppose that we partitioned the churn data set (**Ref. Chapter 3**) into a training set of 2529 records and a test set of 804 records. The summary statistics for customer service calls are given as follows: [2]

	Sample size	No. of VMail subscribers	Proportion of VMail subscribers
Training set	2529	707	0.2796
Test set	804	215	0.2674

   Test whether the proportion of VMail subscribers differs between the training set and the test set, given that $P(Z < 0.6736) = 0.7497$.

5. Of 3333 customers who received promotional materials for a marketing campaign, 483 responded to the promotion. Construct a 99% confidence interval for the population proportion who would respond to the promotion. (Given the standard normal variate Z satisfies the followings: $P(Z < 1.645) = 0.95, P(Z < 1.96) = 0.975, P(Z < 2.576) = 0.995$). [2]

6. A sample of 100 donors to a charity has a mean donation amount of \$ 55 with a sample standard deviation of \$ 25. Test using $\alpha = 0.05$ whether the population mean donation amount exceeds \$ 50. [2]