This course had 6 mini projects.
All the projects were done using R. The problem statements and questions along with the code and the reports are available in this folder.

Both the project and the exams covered very similar topics.
The course included Probability, Random Variables(Continuous and discrete) and their respective distribution functions(Cumulative distribution function, Probability density function), Monte Carlo Methods. Method of Moments, Method of Maximum Likelihood(Log Likelihood function), Confidence intervals for one sample, two samples, with known/unknown variance, in two samples(equal and unequal variance). Z distributions for Normal known variance and unknown non-normal but large population(Central Limit theorem CLT). In two samples, Paired and Independent sample. Paired used the difference between the values to construct a single sample and find CIs. For Normal unknown population variance, we used Student's T distribution with sample variance. CIs, generally, are taken with 90,95,99 percent confidence, with 99% having a wider range to be sure to capture the known fixed mean.

Hypothesis testing, Type I & II errors, Setting Null and alternative hypotheses even before looking at data(or else Data Snooping). In general, Null: Mean = Mean_0, in alternative, Mean != Mean_0, Mean>Mean_0, Mean<Mean_0. Based on these we calculate a test statistic and if the value falls in the rejection region, we reject null and accept alternatives. (|T_obs/Z_obs |>T/Z, Z_obs>Z, Z_obs<Z) for the three alternative hypotheses. Here we use p-value and if p-value is greater than alpha(says 5% confidence interval), alpha=0.05, then we accept null and reject alternatives and vice versa.

Regression analysis: How to observe the relationship between two variables and how to predict the response with single/multiple predictors.
R: lm(response~predictor), anova(analysis of variance)(how to read the table and analyse values like p-value, F-statistic, R-Square, Adjusted R square and so on). Again based on this p-value, we accept/reject the null hypothesis(that the predictor has usual insight into the response variable). The predictor variable can include both normal and categorical variables.

There are three types of selection methods used for choosing the right set of predictors; forward, backward and step. This used the R function step() after using lm. See more syntax of step() on how to implement the three methods. These methods are selected using AIC and BIC, where BIC is considered to be more practical and chooses a lesser number of variables. The lesser values of AIC and BIC are preferred.

If the response variable had categorical variables, then chi-square testing was used.

Also, pearson.test, shapiro-wilk test, and other test were used to check of a variable follows a normal distribution.

Chi square testing for independence between two categorical variables in regression analysis, and chi square testing for homogeneity for the distribution of a categorical variable across two different populations.