Exercise 1:
Suppose 40% of the employees in a very large corporation are women. A random sample of 30 employees are chosen from the corporation. Let, X be the number of women in the sample.

i) [**R**] Find exact value of $\mathbb{P}[X > 5]$.

ii) [**R**] Estimate $\mathbb{P}[X > 5]$ on the basis of random observations from X and compare it with the exact value.

Exercise 2:

i) Draw 10000 random observations from a binomial(35,0.45) distribution.

ii) If the observed value for the i^{th} sample is denoted by X_i, then for each of the samples, calculate the value of the statistic $T_i = \sqrt{35}\left(\frac{X_i}{35} - 0.45\right),\ i = 1, 2, \ldots, 10000.$

iii) [**R**] Prepare a box-plot for the data $T_i,\ i = 1, 2, \ldots, 10000$ and comment on the symmetry.

iv) [**R**] Also prepare a histogram for $T_i,\ i = 1, 2, \ldots, 10000$ considering 10 class intervals of equal width. What can be said about the nature of the distribution of T_i.

v) [**R**] Assume that the above 10000 observations are from a binomial(35,p) distribution with unknown p. Plot the likelihood function and identify the maximum likelihood estimate of p.

Exercise 3:
The following data give the nicotine(X) content in different brands of cigarettes along with carbon monoxide(CO) emission(Y) for each:

Brand	Nicotine(mg)	CO(mg)
Alpine	0.86	13.6
Benson & Hedges	1.06	16.6
Bull Durham	2.03	23.5
Camel lights	0.67	10.2
Carlton	0.4	5.4
Chesterfield	1.04	15
Golden Lights	0.76	9
Kent	0.95	12.3
Kool	1.12	16.3
L&M	1.02	15.4
Lark Lights	1.01	13
Marlboro	0.9	14.4

i) [**R**] Draw a scatterplot of the data and comment on what you can infer about the relationship of the two datasets.

ii) [**R**] Examine the residual plot to check if it contains some symmetric pattern.

iii) [**R**] We want to predict CO emission on the basis of Nicotine content. What do you think will be a good model to fit the data for prediction(Can you say something based on the scatterplot you obtained?)?

iv) [**R**] Set up a regression model to predict Y based on X(Write the model, write the estimates of the parameters in the model).

v) [**R**] Write the coefficient of determination between Y and X.

vi) [**R**] Assuming a bivariate normal setup with parameters well approximated by the corresponding statistics, find $\mathbb{P}[\frac{Y}{X} < 1]$.