Practical-6

Aim:-

Produce stratified Boxplots, Histograms and Scatterplots using different datasets.

Theory:-

Stratified Boxplot in R

A boxplot is a graphical representation of groups of numerical data through their quartiles. Box plots are non-parametric that they display variation in samples of a statistical population without making any assumptions of the underlying statistical distribution. The spacings between the different parts of the box in a bloxplot indicate the degree of dispersion and skewness in the data and show outliers. Boxplot can be drawn either vertically or horizontally. Boxplot got their name from the box in the middle. Stratified boxplots are used to examine the relationship between a categorical and a numeric variable, between strata or groups defined by a third categorical and a numeric variable, between strata or groups defined by a third categorical variable. Stratified Boxplots are useful when it comes to comparing categorical variables.

Syntax:-

boxplot(formula, data = NULL, ..., subset, na.action = NULL, xlab = mklab(y_var = horizontal), ylab = mklab(y_var = !horizontal), add = FALSE, ann = !add, horizontal = FALSE, drop = FALSE, sep = ".", lex.order = FALSE)

boxplot(x, ..., range = 1.5, width = NULL, varwidth = FALSE, notch = FALSE, outline = TRUE, names, plot = TRUE, border = par("fg"),

col = NULL, log = "", pars = list(boxwex = 0.8, staplewex = 0.5, outwex = 0.5), ann = !add, horizontal = False, add = False, at = N

	Parameter	Description
1)	Formula	a formula
2)	data	a data.frame / list from which the variable in the formula should be taken.
3)	subset	An optional vector specifying a subset of observations to be used for plotting.
4)	na.action	A function which indicates what should happen when the data contains NAs.
5)	xlab, ylab	x-axis and y-axis annotation. Can be suppressed by ann = FALSE.
6)	add	logical, if true add boxplot to the curr plot.
7)	ann	logical indicating if axes should be annotated (by xlab and ylab).
8)	horizontal	logical indicating if the boxplots shou horizontal; default FALSE means vertica boxes.
9)	x	Either a numeric vector or a single li containing such vectors.
10)	range	this determines how far the plot whi extend out from the box.
11)	width	a vector giving the relative width of boxes making up the plot.
12)	varwidth	if varwidth is TRUE, the boxes are with widths proportional to the square- of the number of observations in the g
13)	notch	if the notch is TRUE, a notch is dra in each side of the boxes.

Histograms in R

A histogram contains a rectangular area to display the statistical information which is proportional to the frequency of a variable and its width in successive numerical intervals. A graphical representation that manages a group of data points into different specified ranges. It has a special feature which shows no gaps between the bars and is similar to a vertical bar graph. We can create histogram in R programming language using hist() function.

Syntax:- hist(v, main, xlab, xlim, ylim, breaks, col, border)

Parameters:-

v:- This parameter contains numerical values used in histogram.

main:- This parameter main is the title of the chart.

col:- This parameter is used to set color of the bars.

xlab:- This parameter is the label for horizontal axis.

border:- This parameter is used to set border color of each bar.

xlim:- This parameter is used for plotting values of x-axis.

ylim:- This parameter is used for plotting values of y-axis.

breaks:- This parameter is used as width of each bar.

Scatter plots in R

A scatter plot is a set of dotted points to represent individual pieces of data in the horizontal and vertical axis. A graph in which the values of two variables are plotted along X-axis and Y-axis, the pattern of the resulting points reveals a correlation between them.

R- Scatter plots

We can create a scatter plot in R Programming Language using the plot() function.

Syntax:- plot(x, y, main, xlab, ylab, xlim, ylim, axes)

Parameters:-

- x:- This parameter sets the horizontal coordinates.
- y:- This parameter sets the vertical coordinates.
- xlab:- This parameter is the label for horizontal axis.
- ylab:- This parameter is the label for vertical axis.
- main:- This parameter main is the title of the chart.
- xlim:- This parameter is used for plotting values of x.
- ylim:- This parameter is used for plotting values of y.
- axes:- This parameter indicates whether both axes should be drawn on the plot.

Creating a Scatterplot Graph

In order to create a scatterplot graph:-
We are using the required parameters to plot the graph.
In this "xlab" describes the X-axis and "ylab" describes the Y-axis.

Scatterplot Matrices

When we have two or more variables and we want to correlate between one variable and others so we use a scatterplot matrix.
pairs() function is used to create matrices of scatterplots.
Syntax:- pairs(formula, data)

Parameters:-
formula:- This parameter represents the series of variables used in pairs.

data:- This parameter represents the data set from which the variables will be taken.

Scatterplot with fitted values

In order to create scatterplot chart:-
We are using the ggplot2 package provides ggplot() and geom_point() function for creating a scatterplot.
Also we are using the columns "cut" and "mpg" in mtcars.

3D Scatterplots

Here we will use scatterplot3D package to create 3D scatterplots, this package can plot scatterplot in 3D using scatterplot3D() methods.