How can I use the summary() function in R? Can you provide some examples?

Authored by

stats writer

May 5, 2024

The summary() function in R is a powerful tool that provides a concise overview of a dataset or statistical model. It can be used to quickly understand the distribution, central tendency, and spread of data, as well as identify any outliers or missing values. To use the summary() function, simply pass the desired dataset or model as the argument. This function is particularly useful for exploring and summarizing large datasets, as it can efficiently display key information such as mean, median, and quartiles. Additionally, it can provide a summary of regression models, including coefficients, p-values, and R-squared values. Some examples of using the summary() function in R include summarizing a data frame to see the mean and standard deviation of numerical variables, and summarizing a linear regression model to identify significant predictors. Overall, the summary() function is a valuable tool for quickly understanding and analyzing data in R.

Use summary() Function in R (With Examples)

The summary() function in R can be used to quickly summarize the values in a vector, data frame, regression model, or ANOVA model in R.

This syntax uses the following basic syntax:

summary(data)

The following examples show how to use this function in practice.

Example 1: Using summary() with Vector

The following code shows how to use the summary() function to summarize the values in a vector:

```
#define vector
x <- c(3, 4, 4, 5, 7, 8, 9, 12, 13, 13, 15, 19, 21)

#summarize values in vector
summary(x)

Min. 1st Qu. Median Mean 3rd Qu. Max.
3.00 5.00 9.00 10.23 13.00 21.00
```

The summary() function automatically calculates the following summary statistics for the vector:

Min: The minimum value1st Qu: The value of the 1st quartile (25th percentile)Median: The median value3rd Qu: The value of the 3rd quartile (75th percentile)Max: The maximum value

Note that if there are any missing values (NA) in the vector, the summary() function will automatically exclude them when calculating the summary statistics:

```
#define vector
x <- c(3, 4, 4, 5, 7, 8, 9, 12, 13, 13, 15, 19, 21, NA, NA)

#summarize values in vector
```

```
summary(x)

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
3.00 5.00 9.00 10.23 13.00 21.00 2
```

Example 2: Using summary() with Data Frame

The following code shows how to use the summary() function to summarize every column in a data frame:

```
#define data frame
df <- data.frame(team=c('A', 'B', 'C', 'D', 'E'),
points=c(99, 90, 86, 88, 95),
assists=c(33, 28, 31, 39, 34),
rebounds=c(30, 28, 24, 24, 28))

#summarize every column in data frame
summary(df)

team points assists rebounds
Length:5 Min. :86.0 Min. :28 Min. :24.0
Class :character 1st Qu.:88.0 1st Qu.:31 1st Qu.:24.0
Mode :character Median :90.0 Median :33 Median :28.0
Mean :91.6 Mean :33 Mean :26.8
3rd Qu.:95.0 3rd Qu.:34 3rd Qu.:28.0
Max. :99.0 Max. :39 Max. :30.0
```

Example 3: Using summary() with Specific Data Frame Columns

The following code shows how to use the summary() function to summarize specific columns in a data frame:

```
#define data frame
df <- data.frame(team=c('A', 'B', 'C', 'D', 'E'),
points=c(99, 90, 86, 88, 95),
assists=c(33, 28, 31, 39, 34),
rebounds=c(30, 28, 24, 24, 28))

#summarize every column in data frame
summary(df)

points rebounds
Min. :86.0 Min. :24.0
1st Qu.:88.0 1st Qu.:24.0
Median :90.0 Median :28.0
Mean :91.6 Mean :26.8
3rd Qu.:95.0 3rd Qu.:28.0
Max. :99.0 Max. :30.0
```

Example 4: Using summary() with Regression Model

The following code shows how to use the summary() function to summarize the results of a linear regression

model:

#define data
df <- data.frame(y=c(99, 90, 86, 88, 95, 99, 91),
x=c(33, 28, 31, 39, 34, 35, 36))

#fit linear regression model
model <- lm(y~x, data=df)

#summarize model fit
summary(model)

Call:
lm(formula = y ~ x, data = df)

Residuals:
1 2 3 4 5 6 7
6.515 -1.879 -6.242 -5.212 2.394 6.273 -1.848

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 88.4848 22.1050 4.003 0.0103 *
x 0.1212 0.6526 0.186 0.8599

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Residual standard error: 5.668 on 5 degrees of freedom
Multiple R-squared: 0.006853, Adjusted R-squared: -0.1918
F-statistic: 0.0345 on 1 and 5 DF, p-value: 0.8599

Example 5: Using summary() with ANOVA Model

The following code shows how to use the summary() function to summarize the results of an ANOVA model in R:

```
#make this example reproducible
set.seed(0)

#create data frame
data <- data.frame(program = rep(c("A", "B", "C"), each = 30),
weight_loss = c(runif(30, 0, 3),
runif(30, 0, 5),
runif(30, 1, 7)))

#fit ANOVA model
model <- aov(weight_loss ~ program, data = data)

#summarize model fit
```

summary(model)

Df Sum Sq Mean Sq F value Pr(>F)

program 2 98.93 49.46 30.83 7.55e-11 ***

Residuals 87 139.57 1.60

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

The following tutorials offer more information on calculating summary statistics in R: