R-Lab: Effectively Create Analytic Data Sets

STA 321 Topics in Advanced Statistics

West Chester University

Contents

1 Intorduction

In this note, we continue to introduce a few more base R functions that are also commonly used in data management.

2 with() and within() Functions

with() and **within()** are two closely related yet different base R functions that useful in data management.

2.1 The with() Function

`with()` function enables us to define a new variable based on the variables in a **data frame** using basic **R expressions** that include mathematical and logical operations. We can add the newly defined variables to the existing data frame as usual.

with() Syntax

```
with(data-frame, R-expression)
```

Example 1

```
Num <- c(1400,1200,1100,1700,1500)
Cost <- c(1200,1300,1400,1500,1600)
##
dataA <- data.frame(Num,Cost,stringsAsFactors = FALSE)
##
product <- with(dataA, Num*Cost)
quotient <- with(dataA, Cost/Num)
logical <- with(dataA, Num > Cost)
pander(cbind(product = product, quotient = quotient, logical = logical))
```

product	quotient	logical
1680000	0.8571	1
1560000	1.083	0
1540000	1.273	0
2550000	0.8824	1
2400000	1.067	0

```
## add the new variables to data frame dataA
dataA$product = product
dataA$quotient = quotient
dataA$logical = logical
##
pander(dataA)
```

Num	Cost	product	quotient	logical
1400	1200	1680000	0.8571	TRUE
1200	1300	1560000	1.083	FALSE
1100	1400	1540000	1.273	FALSE
1700	1500	2550000	0.8824	TRUE
1500	1600	2400000	1.067	FALSE

The **within()** Function

within() function allows us to create a copy of the data frame and add a column that would eventually store the result of the R expression.

```
Num <- c(1400,1200,1100,1700,1500)
Cost <- c(1200,1300,1400,1500,1600)
##
dataA <- data.frame(Num,Cost,stringsAsFactors = FALSE)
##
dataB <- within(dataA, Product <- Num*Cost)   # defined Product and added to dataA simultaneously
dataC <- within(dataB, Quotient <- Cost/Num)
dataD <- within(dataC, Logical <- Num > Cost)
pander(dataD)
```

Num	Cost	Product	Quotient	Logical
1400	1200	1680000	0.8571	TRUE
1200	1300	1560000	1.083	FALSE
1100	1400	1540000	1.273	FALSE
1700	1500	2550000	0.8824	TRUE
1500	1600	2400000	1.067	FALSE

3   Markdown Tables

Num	Cost	Product	Quotient	Logical
1400	1200	1680000	0.8571	TRUE
1200	1300	1560000	1.083	FALSE
1100	1400	1540000	1.273	FALSE
1700	1500	2550000	0.8824	TRUE
1500	1600	2400000	1.067	FALSE

Num	Cost	Product	Quotient	Logical
1400	1200	1680000	0.8571	TRUE
1200	1300	1560000	1.083	FALSE
1100	1400	1540000	1.273	FALSE
1700	1500	2550000	0.8824	TRUE
1500	1600	2400000	1.067	FALSE