User support over the years

Dave (Data analyst)

How can I adjust p-values for multiple comparisons in R?
User support over the years

Ask a colleague
User support over the years

Ask a colleague
User support over the years
Mailing lists vs. Stack Exchange

December 2003 Archives by thread

- Messages sorted by: [subject] [author] [date]
- More info on this list...

Starting: Mon Dec 1 00:27:40 CEST 2003
Messages: 1159

- [R] fitting a theoretical distribution with truncated tails
  - Piyush Sharma
  - [R] fitting a theoretical distribution with truncated tails
    - Spencer Graves
- [R] Discovering methods
  - Gabor Grothendieck
  - [R] Discovering methods
    - Duncan Murdoch
  - [R] Discovering methods
    - Thomas Lumley
  - [R] Discovering methods
    - Duncan Murdoch
    - [R] Discovering methods
      - Thomas Lumley
- [R] hdf library for windows
  - Toby.Patterson at csiro.au
  - [R] hdf library for windows
    - Mulholland, Tom
- [R] strptime Usage
  - Ko-Kang Kevin Wang
  - [R] strptime Usage
    - Prof Brian Ripley
- [R] Discovering methods
  - Prof Brian Ripley
  - [R] significance in difference of proportions: What problema
    - Torsten Hothorn
- [R] index ANOVA table
  - Pascal A. Niklaus
- [R] wilcoxon-pratt signed rank test in R - drug-efficacy
  - Niels Steen Krogh
  - [R] wilcoxon-pratt signed rank test in R - drug-efficacy
    - Peter Dalgaard
    - [R] wilcoxon-pratt signed rank test in R - drug-efficacy
      - Knut M. Wittkowski
- [R] Changing distance scale in pffclust()
  - Mike White

Historically, coordinating dev't and user support activities

Changing

collaborate, learn and communicate among
themselves and with their users

Fastest Q&A site

Good technical solutions
Research goal

How did knowledge sharing in the community change with the emergence of StackExchange Q&A sites?
Hello,

I am having some problems to use the 'merge' function. I'm not sure if I got its working right.

What I want to do is:

1) Suppose I have a dataframe like:

   height    width
   1         1.1
   2         2.1
   3         1.8
   4         1.6
   5         1.8

   height    width
   3         1.8
   4         1.6
   5         1.8

   [...]

4) So, I want to merge those dataframes, so that the new variable, color, is bound to the first dataframe. Of course some cases won't have value for it, since I generated this variable in a smaller dataframe. In those cases I want the value to be NA. The result dataframe should be:

   height    width    color
   1         2.3       red
   2         2.5       NA
   3         1.9       red
   4         2.1       NA
   5         2.4       blue

I have written some codes, but they're not working properly. The new variable has its values mixed up, and they do not correspond to its row.names.

# Generate first dataframe
data1 <- data.frame(height=rnorm(20,3,0.2),width=rnorm(20,2,0.5))
# Sample a smaller dataframe from data1
data2 <- data1[sample(1:20,15,replace=F),]
# Generate new variable
color <- sample(c("red","blue"),15,replace=T)
# Bind the new variable to data2
data2 <- cbind(data2, color)
# Merge the data1 and data2$color by row.names, and force it to have values that data1. Next it generates a new dataframe where column 1 is t.row.name, and then sort it by the row.name from data1.
data.frame(merge(data1, data2$color, by=0, all.x=T),row.names=1)[row.names(data1),]

I'm not sure what am I doing wrong. Can anyone see where the mistake is, please?
Hello,

I am having some problems to use the 'merge' function. I'm not sure if I got its working right.

What I want to do is:

1) Suppose I have a dataframe like:

<table>
<thead>
<tr>
<th>height</th>
<th>width</th>
<th>color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1</td>
<td>red</td>
</tr>
<tr>
<td>2</td>
<td>2.1</td>
<td>red</td>
</tr>
<tr>
<td>3</td>
<td>1.8</td>
<td>red</td>
</tr>
<tr>
<td>4</td>
<td>1.6</td>
<td>red</td>
</tr>
<tr>
<td>5</td>
<td>1.8</td>
<td>red</td>
</tr>
</tbody>
</table>

What you offered would be the fractional variation, but if you multiplied by 100 you get the "percent variation":

Hello,

I am having some problems to use the `merge` function. I'm not sure if I got its working right.

What I want to do is:

1) Suppose I have a dataframe like:

```
height            width         color
1        1.1                2.3            red
2        2.1                2.5            red
3        1.8                2.1            red
4        1.6                1.9            red
5        1.8                2.4            blue
```

I have written some codes, but they're not working properly. The new variable has its values mixed up, and they do not correspond to its row.names.

```
# Generate the first dataframe
data1 <- data.frame(height=rnorm(20,3,0.2),width=rnorm(20,2,0.5))
# Sample a smaller dataframe from data1
data2 <- data1[sample(1:20,15,replace=F),]
# Generate the new variable
color <- sample(c("red","blue"),15,replace=F)
# Bind the new variable to data2
data2 <- cbind(data2,color)
# Merge the data1 and data2 color by row.names, and force it to have the same values that data1. Next it generates a new dataframe where column 1 is the row.name, and then sort it by the row.name from data1.
data.frame(merge(data1,data2,color,by=0,all.x=T),row.names=1)
```

2) And I generate a second dataframe sampled from this one, like:

```
data1 <- data.frame(height=rnorm(20,3,0.2),width=rnorm(20,2,0.5))
data2 <- cbind(data2, color)
data2 <- data1[sample(1:20,15,replace=F),]
```

3) Next, I add a new variable from this dataframe:

```
data1 <- data.frame(height=rnorm(20,3,0.2),width=rnorm(20,2,0.5))
data2 <- cbind(data2, color)
data2 <- data1[sample(1:20,15,replace=F),]
```

4) So, I want to merge those dataframes, so that the new variable, color, is binded to the first dataframe. Of course some cases won't have value for it, since I generated this variable in a smaller dataframe. In those cases I want the value to be NA. The result dataframe should be:

```
height            width         color
1        1.1                2.3            red
2        2.1                2.5            red
3        1.8                2.1            red
4        1.6                1.9            red
5        1.8                2.4            blue
```

Since I generated this variable in a smaller dataframe. In those cases I want the value to be NA. The result dataframe should be:

```
height            width         color
1        1.1                2.3            red
2        2.1                2.5            red
3        1.8                2.1            red
4        1.6                1.9            red
5        1.8                2.4            blue
```

I'm not sure what am I doing wrong. Can anyone see where the mistake is?
What I want to do is:

3) Next, I add a new variable from this dataframe:

I am having some problems to use the `merge` function. I'm not sure if I got its working right.

Hello,

I need to calculate percent variation for consecutive values:

```r
pv <- vector("numeric", length(x))
pv[1] <- 0
for(i in 2:length(x)) {
  pv[i] <- (x[i] - x[i-1])/x[i-1]
}
```

So, as you can see, I have to use both the `x[i]` element, but also the `x[i-1]` element. By using the `apply` functions, I just see how to use the `x[1]`. Is there anyway I can avoid the `for` loops?

I have written some codes, but they’re not working properly. The new variable has its values mixed up, and they do not correspond to its row.names.

```
# Generate the first dataframe
data1 <- data.frame(height=rnorm(20,3,0.2),width=rnorm(20,2,0.5))
# Sample a smaller dataframe from data1
data2 <- data1[sample(1:20,15,replace=F),]
# Generate the new variable color <- sample(c("red","blue"),15,replace=T)
# Bind the new variable to data2 data2 <- cbind(data2, color)
# Merge the data1 and data2$color by row.names, and force it to has the same values
# that data1. Next it generates a new dataframe where column 1 is the row.name, and then sort by it the row.name from data1.
data.frame(merge(data1,data2$color, by=0, all.x=T),row.names=1){row.names(data1),}
```

I'm not sure what am I doing wrong. Can anyone see where the mistake is?
Research goal

How did knowledge sharing in the community change with the emergence of StackExchange Q&A sites?
Research goal

How did knowledge sharing in the community change with the emergence of StackExchange Q&A sites?

Transition to StackExchange?

“Young experts don’t want to have to monitor email all day to be part of the discussion. Their answers belong on a website with a normal content management system, with good search functions and user interactions. Go [to Stack Exchange and] sign up.”

[2010 blog entry]
Research goal

How did knowledge sharing in the community change with the emergence of StackExchange Q&A sites?
Who are they?
Today

1. Who are they?
2. Do they behave any differently?
Today

1. Who are they?
2. Do they behave any differently?

Mining repositories + User survey (112 resp.)
Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk
Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk
Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk
Laurent Gautier - s010592@student.dtu.dk
Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk
Laurent Gautier - s010592@student.dtu.dk
Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk
Laurent Gautier - s010592@student.dtu.dk
Laurent - lgautier@gmail.com
Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk
Laurent Gautier - s010592@student.dtu.dk
Laurent - lgautier@gmail.com
Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk
Laurent Gautier - s010592@student.dtu.dk
Laurent - lgautier@gmail.com
- lgautier@altern.org

9 email addresses, 3 names

Identity merging is far from trivial!
Mining challenges

Laurent Gautier - laurent@cbs.dtu.dk
Laurent Gautier - s010592@student.dtu.dk
Laurent - lgautier@gmail.com
- lgautier@altern.org

...
The R community

> September 2008

~16k

~3k (20%)

Stack Exchange

stack overflow ~10k

Cross Validated ~2k

Who are they?
The R community

- Users
- Package maintainers
- Peripheral
- Core

Who are they?

- Have write access to the R source (~20)
- Contribute code, bug fixes and documentation (~40)
- Author or maintain R packages (~2600)
- Others

Who are they?

@b_vasilescu
users on Stack Exchange

Who are they?

Who are they?

Who are they?

Who are they?
Devs vs users: 1.6x-2x higher chance of being on StackExchange.

Who are they?

[@b_vasilescu](https://twitter.com/b_vasilescu)
users on StackExchange

Consistently more active

median 1 thread answered

median 4 threads answered

Who are they?

@b_vasilescu
Today

1
Who are they?
Mostly devs, consistently more active

2
Do they behave any differently?

@b_vasilescu
The same users answer faster on StackExchange.

Speed of answers for r-help participants active on StackExchange

- Median ~3h on r-help
- Median 47 mins on StackExchange

Do they behave any differently?

@b_vasilescu
The same users answer faster on StackExchange.

**Answer Badges**

<table>
<thead>
<tr>
<th>Badge</th>
<th>Description</th>
<th>Awarded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enlightened</td>
<td>First to answer and accepted with score of 10 or more</td>
<td>117k</td>
</tr>
<tr>
<td>Generalist</td>
<td>Provided non-wiki answers of 15 total score in 20 of top 40 tags</td>
<td>514</td>
</tr>
<tr>
<td>Guru</td>
<td>Accepted answer and score of 40 or more</td>
<td>26.4k</td>
</tr>
<tr>
<td>Nice Answer</td>
<td>Answer score of 10 or more</td>
<td>374.8k</td>
</tr>
<tr>
<td>Good Answer</td>
<td>Answer score of 25 or more</td>
<td>92.3k</td>
</tr>
<tr>
<td>Great Answer</td>
<td>Answer score of 100 or more</td>
<td>11.5k</td>
</tr>
</tbody>
</table>

Do they behave any differently?
Survey

Motivation to contribute

In case of Stack Exchange, the reputation ratings are a nice little incentive.

It’s a game, which also serves a good purpose.

Peer recognition/gamification within Stack overflow

Do they behave any differently?

@b_vasilescu
Research goal

How did knowledge sharing in the R community change with the emergence of Stack Exchange Q&A sites?
Transition to StackExchange

Activity of r-help answerers also active on Stack Exchange

- On r-help
- On StackExchange

Number of answers

Date

@b_vasilescu
Survey

Disengagement from the mailing list

Google is getting better at finding answers related to R so I use it more. I rely less on going directly to mailing lists now.

r-help used to be very helpful. But as the number of posts has gone up, I find that reading it is not as useful as it had been.
Research goal

How did knowledge sharing in the community change with the emergence of Stack Exchange Q&A sites?

The same users answer faster on Stack Exchange

Survey

Motivation to contribute

In case of Stack Exchange, the reputation ratings are a nice little incentive.

It's a game, which also serves a good purpose.