DPLYR Expand Dataframe with Function

Fan Wang

2020-04-01

Contents

Mx1 to MxQ Rows

Go back to fan's REconTools Package, R4Econ Repository, or Intro Stats with R Repository.

Case One: There is a dataframe with M rows, based on these m specific information, generate dataframes for each m. Stack these indivdiual dataframes together and merge original m specific information in as well. The number of rows for each m is Q_m, each m could have different number of expansion rows.

Generate a panel with M individuals, each individual is observed for different spans of times (*uncount*). Before expanding, generate individual specific normal distribution standard deviation. All individuals share the same mean, but have increasing standard deviations.

Generate Dataframe with M Rows. This is the first step, generate M rows of data, to be expanded. Each row contains the number of normal draws to make and the mean and the standard deviation for normal daraws that are m specific.

```
# Parameter Setups
it_M <- 3
it_Q_max <- 5
fl_rnorm_mu <- 1000
ar_rnorm_sd <- seq(0.01, 200, length.out=it_M)
ar_it_q <- sample.int(it_Q_max, it_M, replace=TRUE)

# N by Q varying parameters
mt_data = cbind(ar_it_q, ar_rnorm_sd)
tb_M <- as_tibble(mt_data) %>% rowid_to_column(var = "ID") %>%
                rename(sd = ar_rnorm_sd, Q = ar_it_q) %>%
                mutate(mean = fl_rnorm_mu)

# display
kable(tb_M) %>%
  kable_styling_fc_wide()
```

Random Normal Draw Expansion The steps are:

1. do anything
2. use ".$" sign to refer to variable names, or [['name']]
3. unnest
4. left_join expanded and original

Note these all give the same results

ID	Q	sd	mean
1	5	0.010	1000
2	4	100.005	1000
3	5	200.000	1000

Use dot dollar to get variables

```
# Generate $Q_m$ individual specific incomes, expanded different number of times for each m
tb_income <- tb_M %>% group_by(ID) %>%
  do(income = rnorm(.$Q, mean=.$mean, sd=.$sd)) %>%
  unnest(c(income))

# Merge back with tb_M
tb_income_full_dd <- tb_income %>%
  left_join(tb_M)
```

```
## Joining, by = "ID"
```

```
# display
kable(tb_income) %>%
  kable_styling_fc_wide()
```

```
kable(tb_income_full_dd) %>%
  kable_styling_fc_wide()
```

ID	income
1	999.9987
1	1000.0089
1	999.9985
1	1000.0033
1	999.9677
2	922.8170
2	1028.6563
2	877.9427
2	1043.4572

ID	income	Q	sd	mean
1	999.9987	5	0.010	1000
1	1000.0089	5	0.010	1000
1	999.9985	5	0.010	1000
1	1000.0033	5	0.010	1000
1	999.9677	5	0.010	1000
2	922.8170	4	100.005	1000
2	1028.6563	4	100.005	1000
2	877.9427	4	100.005	1000
2	1043.4572	4	100.005	1000
3	1160.0354	5	200.000	1000
3	967.2138	5	200.000	1000
3	1248.5838	5	200.000	1000
3	813.1230	5	200.000	1000
3	1078.7417	5	200.000	1000