Bootstrap

12/10/2017

```
# lib
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.4.2
```

```
## -- Attaching packages ---------------------------------- tidyverse 1.2.1 --
```

```
## <U+221A> ggplot2 2.2.1     <U+221A> purrr   0.2.4
## <U+221A> tibble  1.3.4     <U+221A> dplyr   0.7.4
## <U+221A> tidyr   0.7.2     <U+221A> stringr 1.2.0
## <U+221A> readr   1.1.1     <U+221A> forcats 0.2.0
```

```
## Warning: package 'tidyr' was built under R version 3.4.2
```

```
## Warning: package 'purrr' was built under R version 3.4.2
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
## -- Conflicts ------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(boot)

# load the data
data_bootstrap = read.csv("./data/data_no_outliers.csv") %>%
  select(-X, -postalcode,-facilityname,-facilityzip, -patientid, -visitid, -date) %>%
  mutate(is30dayreadmit = as.factor(is30dayreadmit), icu_flag = as.factor(icu_flag)) %>%
  select(log_length_of_stay,
         is30dayreadmit,
         cindex,
         evisit,
         age,
         gender,
         maritalstatus,
         temperature_cat,
         insurancetype,
         heartrate_transformed,
         respirationrate_transformed,
         mean_arterial_pressure)

# final model
# model_sw_0.10 = lm(log_length_of_stay ~ is30dayreadmit + cindex
#                       + evisit + age + gender + maritalstatus + temperature_cat
#                       + insurancetype + heartrate_transformed
#                       + respirationrate_transformed
#                       + mean_arterial_pressure)

# bootstrap to obtain the distribution

# code 1: raw code
# boot_res1 =
```

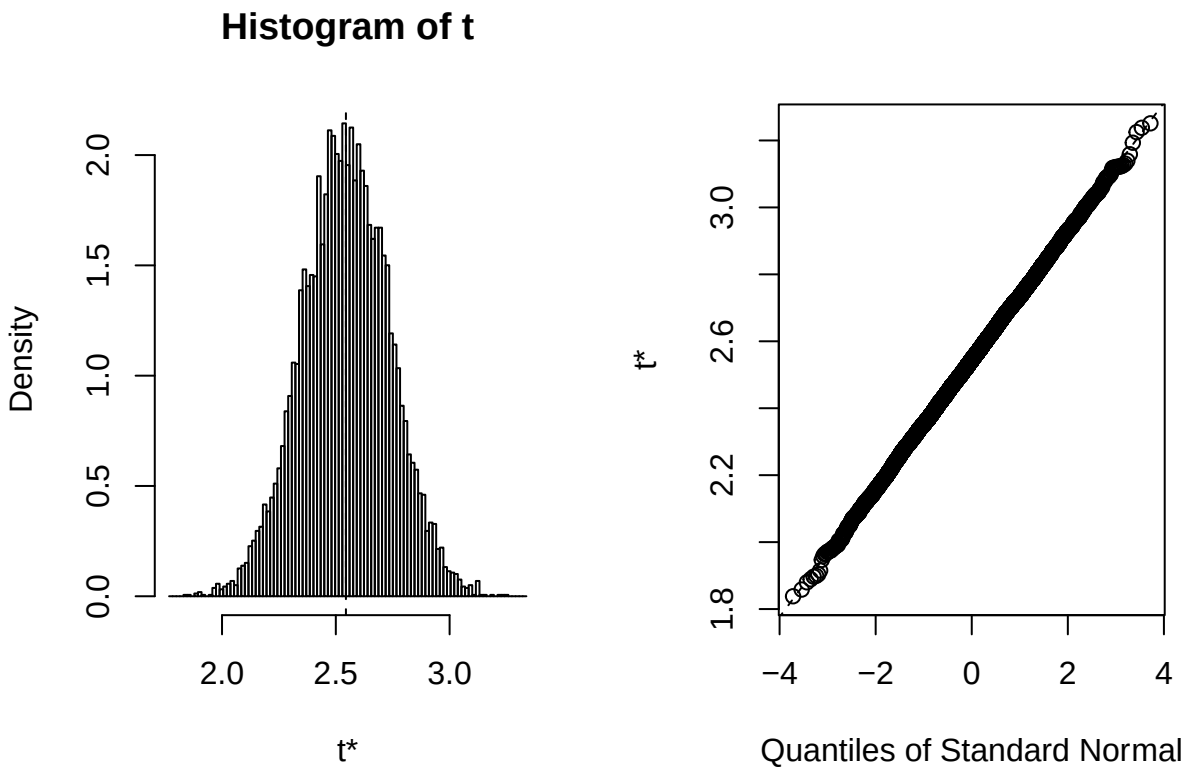
Table 1: Bootstrap

term	statistic	bias	std.error
(Intercept)	2.5444567	-0.0002083	0.1920116
is30dayreadmit1	0.1801712	-0.0002694	0.0421892
cindexmoderate	0.1323495	-0.0001276	0.0476890
cindexnormal	0.0251459	0.0000472	0.0345832
cindexsevere	0.1775025	-0.0006796	0.0428682
evisit	0.0642347	0.0001129	0.0095534
age	0.0096015	0.0000127	0.0009287
genderMale	0.0611153	0.0003365	0.0295831
maritalstatusNot Married	0.0885007	-0.0000767	0.0293569
temperature_catlow	-0.0593297	0.0005501	0.0906106
temperature_catnormal	-0.2456611	0.0003272	0.0584033
insurancetypeMedicare	-0.1447609	-0.0009221	0.0722351
insurancetypePrivate	-0.1855449	-0.0002085	0.0681488
heartrate_transformed	-2061.1985080	0.0079740	301.2852763
respirationrate_transformed	-178.7334811	-0.0165204	30.1096161
mean_arterial_pressure	-0.0058527	-0.0000044	0.0008342

```
#   data_bootstrap %>%
#   modelr::bootstrap(n = 10000) %>%
#   mutate(models = map(strap, ~lm(log_length_of_stay ~ ., data = .x) ),
#          results = map(models, broom::tidy)) %>%
#   select(-strap, -models) %>%
#   unnest() %>%
#   select(id = `.id`, term, estimate) %>%
#   spread(key = term, value = estimate)

# code 2: boot function
boot.fn = function(data, index){
  return(coef(lm(log_length_of_stay ~ ., data = data, subset = index)))
}
boot_res2 = boot(data_bootstrap, boot.fn, 10000)
# table for bootstrapping results
broom::tidy(boot_res2) %>%
  knitr::kable(format = "latex",
               caption = "Bootstrap",
               booktabs = TRUE) %>%
  kableExtra::kable_styling(latex_options = "scale_down")
```

```
# plot
plot(boot_res2)
```




```
# CI
boot.ci(boot.out = boot_res2, type = c("norm", "basic"))
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 10000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot_res2, type = c("norm", "basic"))
##
## Intervals :
## Level      Normal              Basic
## 95%   ( 2.168,  2.921 )   ( 2.167,  2.926 )
## Calculations and Intervals on Original Scale
```