OLS Analysis, Figures and Tables Used in Master Thesis (Rev. January 14, 2016)

Replication Procedure

Kentaro WADA

January 6, 2016

Basic Settings

In this document, you will get necessary R codes to replicate ols analysis, figures and tables used in my master thesis (you can download the document (master_thesis.pdf) at Kentaro WADA's GitHub Website (https://github.com/KentaroWADA/Master-Thesis)). In order to do so, you need to install necessary R packages as follows:

Please keep in mind that this document is created under *"Japanese"* language setting. If your laptop is set to use *"English"* or any other languages, you will face garbled issues. Therefore, if you need to replicate figures with Japanese, please change language setting to "*Japanese*".

Also, if you would like to see this document *html* style, you can get them by copy and paste next line.
file:///Users/WadaKen/Dropbox/Kobe_University/Master_Thesis/master_thesis_replication.html

(file:///Users/WadaKen/Dropbox/Kobe_University/Master_Thesis/master_thesis_replication.html)

```
## stop showing "this objects are masked" in advance
suppressMessages(library(dplyr))
suppressMessages(library(xtable))
suppressMessages(library(stargazer))
suppressMessages(library(psych))
suppressMessages(library(Hmisc))
suppressMessages(library(grid))
suppressMessages(library(coefplot))

## install necessary R packages
## create new variables with mutate(), and extract necessary v
ariables with filter()
library("dplyr")
## display high quality figure with ggplot()
library("ggplot2")
## export LaTeX style table with xtable()
library("xtable")
## export regression result with LaTeX style table with starga
zer()
library("stargazer")
## show basic statistics with describe()
library("psych")
## calculate correlations with significance levels with rcorr(
)
library("Hmisc")
## grid adds an nx by ny rectangular grid to an existing plot
with grid.newpage()
library("grid")
## show regression result with high quality figure with coefpl
ot()
library("coefplot")
```

After installing necessary packages, you need to prevent Japanese characters from garbling as follows:

```
quartzFonts(HiraKaku = quartzFont(rep("Hiragino Kaku Gothic Pro W3", 4)))
theme_set(theme_gray(base_size = 10.5, base_family = "HiraKaku"))
```

Then, go to Kentaro WADA's GitHub Website (https://github.com/KentaroWADA/Master-Thesis), download **dataset** folder that contains all necessary data set for making figures and tables. After you finish downloading, save it in the **dataset** folder within your project folder (or anywhere you want and change the path to the file accordingly).

Figures

Before plotting figure 1-1 and figure 1-2, create necessary data set as follows

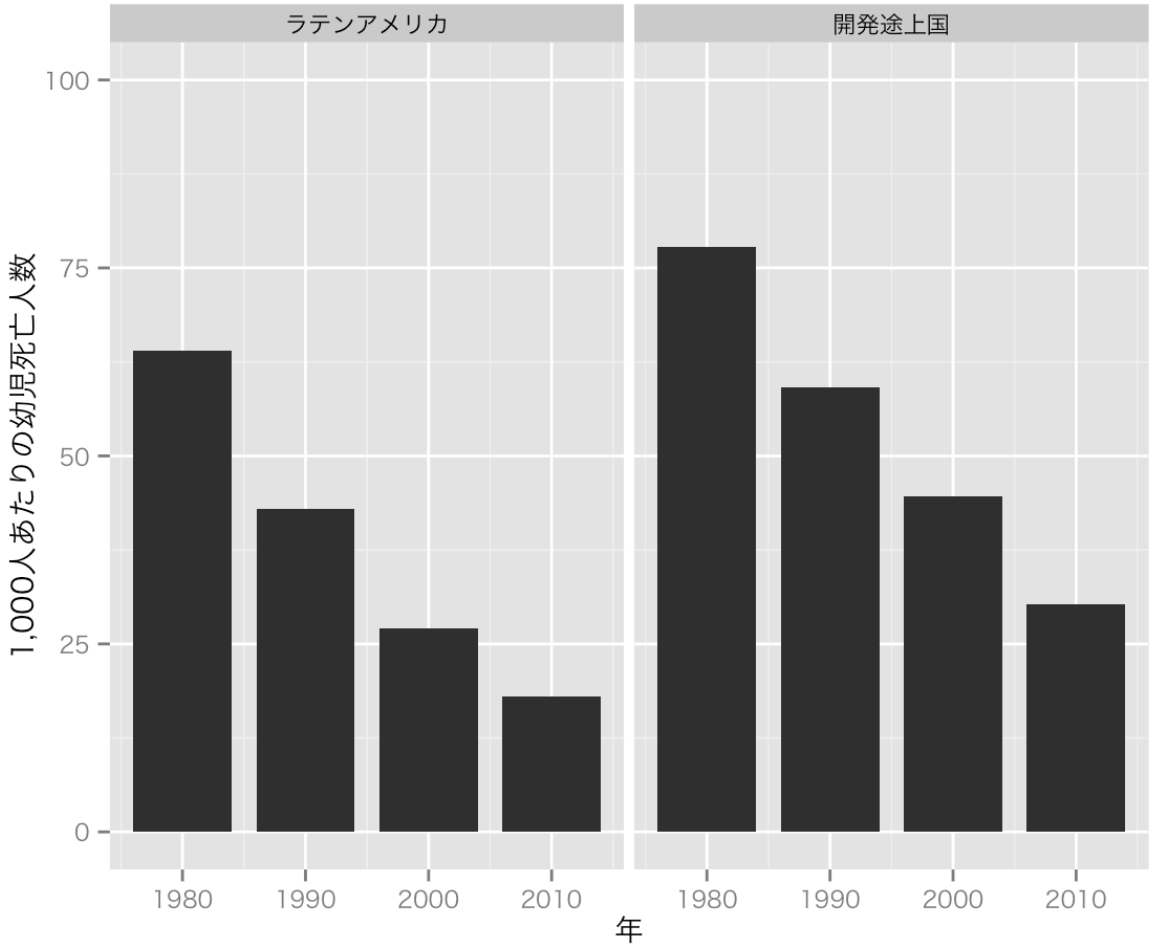
```
## import health related data
data <- read.csv("dataset/figure1-1_1-2.csv", skip = 3)

## extract Latin America
latin <- filter(data, region == "Latin America")
latin1980 <- filter(latin, year == 1980)
latin1990 <- filter(latin, year == 1990)
latin2000 <- filter(latin, year == 2000)
latin2010 <- filter(latin, year == 2010)

## infant moratality rate in Latin America
linf1980 <- latin1980$infant_mortality
linf1990 <- latin1990$infant_mortality
linf2000 <- latin2000$infant_mortality
linf2010 <- latin2010$infant_mortality

## immunization rate in Latin America
limm1980 <- latin1980$immunization
limm1990 <- latin1990$immunization
limm2000 <- latin2000$immunization
limm2010 <- latin2010$immunization

## extract designated year
data1980 <- filter(data, year == 1980)
data1990 <- filter(data, year == 1990)
data2000 <- filter(data, year == 2000)
```

```
data2010 <- filter(data, year == 2010)

## infant moratality rate in developing countries
inf1980 <- mean(data1980$infant_mortality, na.rm = T)
inf1990 <- mean(data1990$infant_mortality, na.rm = T)
inf2000 <- mean(data2000$infant_mortality, na.rm = T)
inf2010 <- mean(data2010$infant_mortality, na.rm = T)

## immunization rate in developing countries
imm1980 <- mean(data1980$immunization, na.rm = T)
imm1990 <- mean(data1990$immunization, na.rm = T)
imm2000 <- mean(data2000$immunization, na.rm = T)
imm2010 <- mean(data2010$immunization, na.rm = T)

## create vectors
年 <- c(1980, 1990, 2000, 2010) ## 年 means year
年 <- rep(年, 2)
type <- rep(c("ラテンアメリカ", "開発途上国"), c(4, 4))
infant <- c(linf1980, linf1990, linf2000, linf2010, inf1980, i
nf1990, inf2000, inf2010)
immunization <- c(limm1980, limm1990, limm2000, limm2010, imm1
980, imm1990, imm2000, imm2010)

## create data frame
df1 <- data.frame(年 = 年, type = type, infant = infant, immun
ization = immunization)
```

Figure 1-1

Figure 1-1 demonstrates infant mortality rate per 1,000 live births in Latin America and developing countries. I compared the mean of countries in Latin America and that of develping countries.

```
## plot bar chart with ggplot2
fig.11 <- ggplot(df1, aes(x = 年, y = infant))
fig.11 <- fig.11 + geom_bar(stat = "identity", width = 8.0) +
facet_grid(. ~ type) +
  labs(y = "1,000人あたりの幼児死亡人数", x = "年") + ylim(c(0, 1
00))
print(fig.11)
```

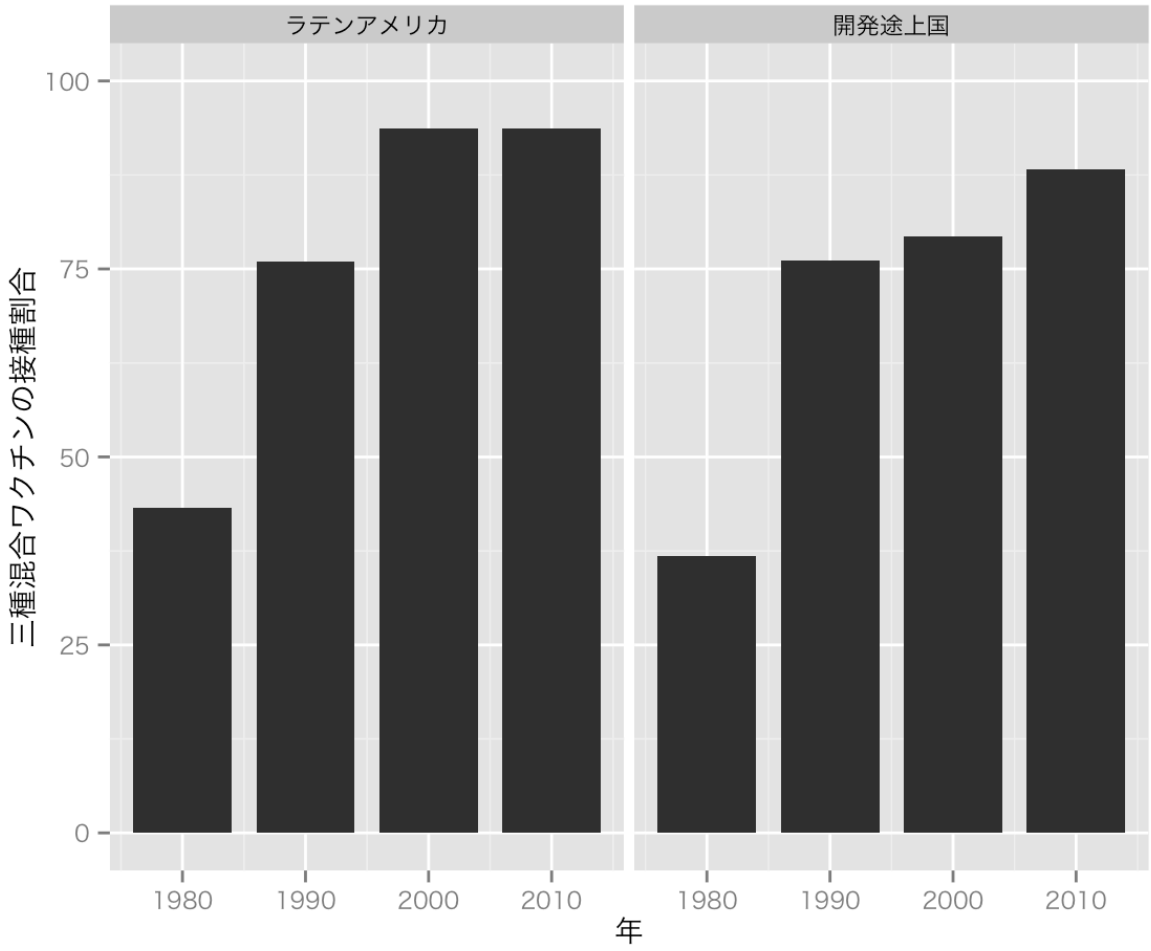



```
##====================== Note ====================##
## if you want to save figures in pdf, put
## quartz(file = "mortality.pdf", family = "sans",
##        type = "pdf", width = 10, height = 8)
## before print(fig.11), and put
## dev.off()
## after print(fig.11)
##====================== Note ====================##
```

Figure 1-2

Figure 1-2 represents measles immunization rate of children ages 12-23 months in Latin America and developing countries. I compared the mean of countries in Latin America and that of develping countries.

```
## plot bar chart with ggplot2
fig.12 <- ggplot(df1, aes(x = 年, y = immunization))
fig.12 <- fig.12 + geom_bar(stat = "identity", width = 8.0) +
facet_grid(. ~ type) +
  labs(y = "三種混合ワクチンの接種割合", x = "年") + ylim(c(0, 100
))
print(fig.12)
```




```
##====================== Note ====================##
## if you want to save figures in pdf, put
## quartz(file = "immunization.pdf", family = "sans",
##        type = "pdf", width = 10, height = 8)
## before print(fig.12), and put
## dev.off()
## after print(fig.12)
##====================== Note ====================##
```

Figure 1-3

Figure 1-3 shows social spending in Latin America from 1990 to 2010, using the mean of 18 countries in Latin America. You need to separately create two figures, then combine them into one figure. First, you create a line chart of social spending in Latin America.

```
## import data set
df2 <- read.csv(file = "dataset/figure1-3_percent.csv", skip=
1)
## confirm names of variables
head(df2)
```

```
##   year     type  value
## 1 1990   health 27.300
## 2 1995   health 21.050
## 3 2000   health 19.930
## 4 2005   health 20.700
## 5 2010   health 21.891
## 6 1990 security 32.880
```

```
## rename variables
df2 <- mutate(df2, type = factor(type, labels = c("健康保健支出
", "社会保障・扶助")))
colnames(df2) <- c("year","部門","value")
## reonfirm names of variables
head(df2)
```

```
##   year           部門  value
## 1 1990   健康保健支出 27.300
## 2 1995   健康保健支出 21.050
## 3 2000   健康保健支出 19.930
## 4 2005   健康保健支出 20.700
## 5 2010   健康保健支出 21.891
## 6 1990 社会保障・扶助 32.880
```

```
## create line chart
line <- ggplot(df2, aes(x=year, y=value)) + geom_line(aes(co
lour = 部門)) +
  geom_point(stat="identity", colour="blue",size=2) + ylim(15
, 40) +
  labs(y = "対社会支出に占める割合", x = "年")
```

Second, you create a bar chart of social spending per capita in Latin America.

```
## import dataset
df3 <- read.csv(file = "dataset/figure1-3_capita.csv", skip=1
)
## confirm names of variables
head(df3)
```

```
##   year     type value
## 1 1990   health   103
## 2 1995   health   107
## 3 2000   health   115
## 4 2005   health   129
## 5 2010   health   167
## 6 1990 security   119
```

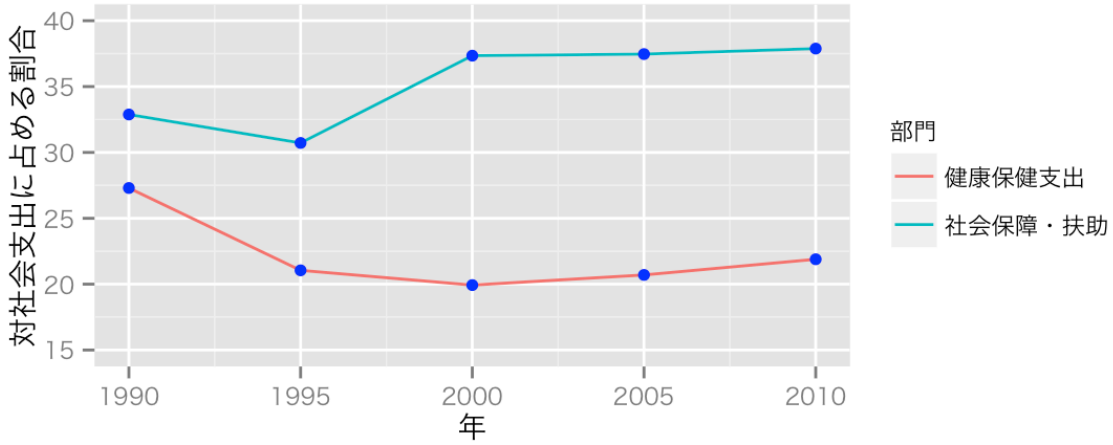
```
## rename variables
df3 <- mutate(df3, type = factor(type, labels = c("健康保健支出
", "社会保障・扶助")))
colnames(df3) <- c("year","部門","value")
## reconfirm names of variables
head(df3)
```

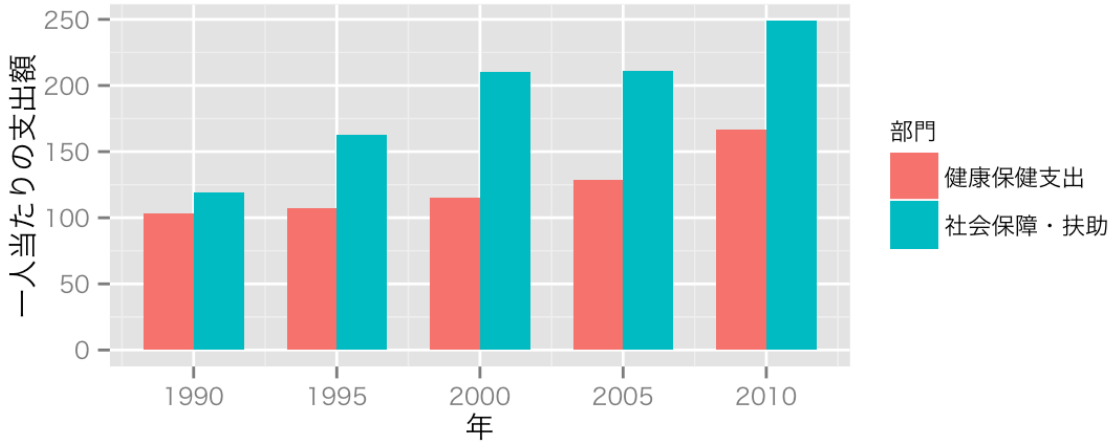
```
##   year            部門 value
## 1 1990   健康保健支出   103
## 2 1995   健康保健支出   107
## 3 2000   健康保健支出   115
## 4 2005   健康保健支出   129
## 5 2010   健康保健支出   167
## 6 1990 社会保障・扶助   119
```

```
## create bar chart
bar <- ggplot(df3, aes(x=year, y=value, fill = 部門)) +
  geom_bar(position = "dodge", stat = "identity", width = 3.5
) +
  labs(y = "一人当たりの支出額", x = "年")
```

Finally, you fit two charts into a figure.

```
## put two different graphs into one
grid.newpage()
pushViewport(viewport(layout=grid.layout(2, 2) ))
print(line, vp=viewport(layout.pos.row=1, layout.pos.col=1:2)
)
print(bar, vp=viewport(layout.pos.row=2, layout.pos.col=1:2))
```






```
##====================== Note ====================##
## since I used grid.newpage() function,
## you need to manually save the figure.
## I saved it as pdf with 8 x 11 inches
##====================== Note ====================##
```

In my master thesis, I mixed a line chart and bar chart into one figure, but you can output those charts in one figure as follows: (But not recommended)

```
fig.13 <- ggplot(df2, aes(x=year, y=value, fill = 部門))
fig.13 <- fig.13 + geom_bar(position = "dodge", stat = "ident
ity")
fig.13 + geom_line(data=df3, aes(x=year, y=value, colour=部門)
)
```

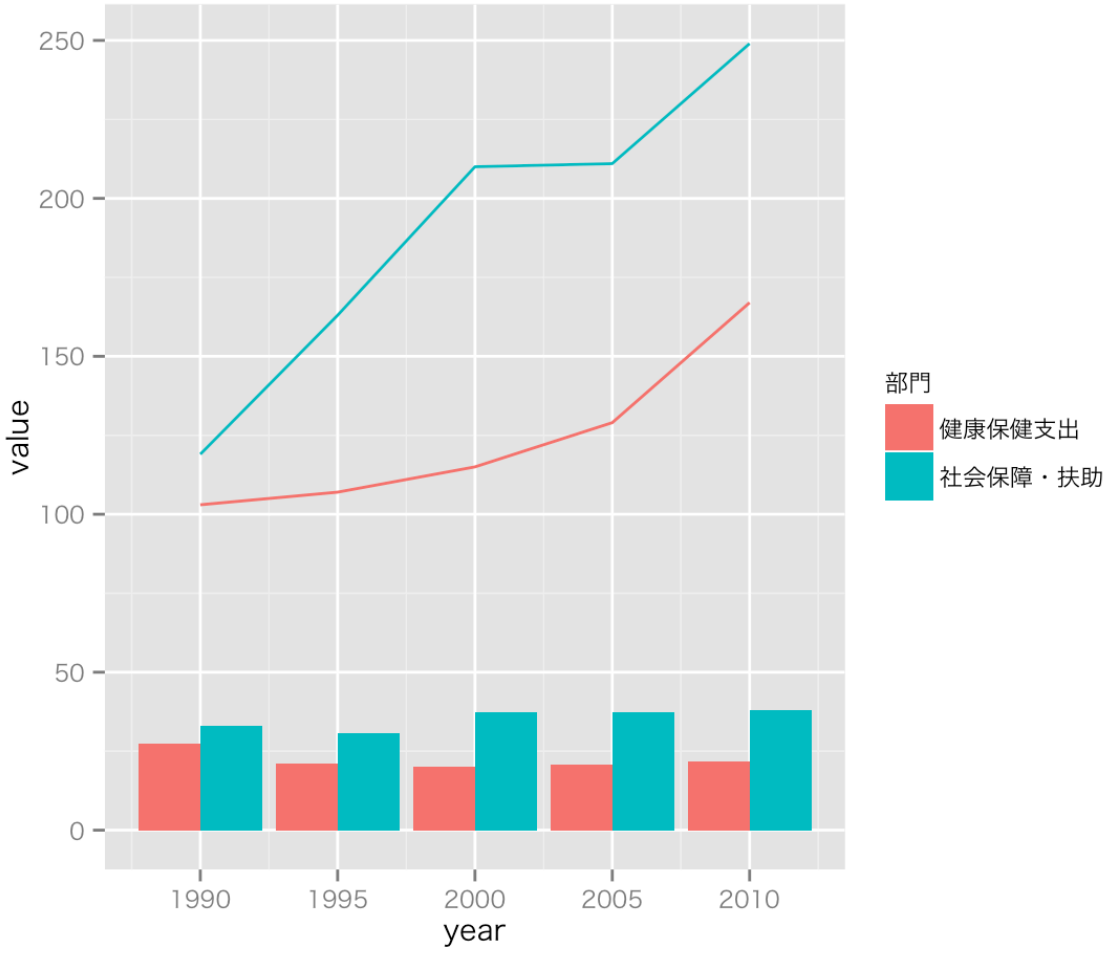
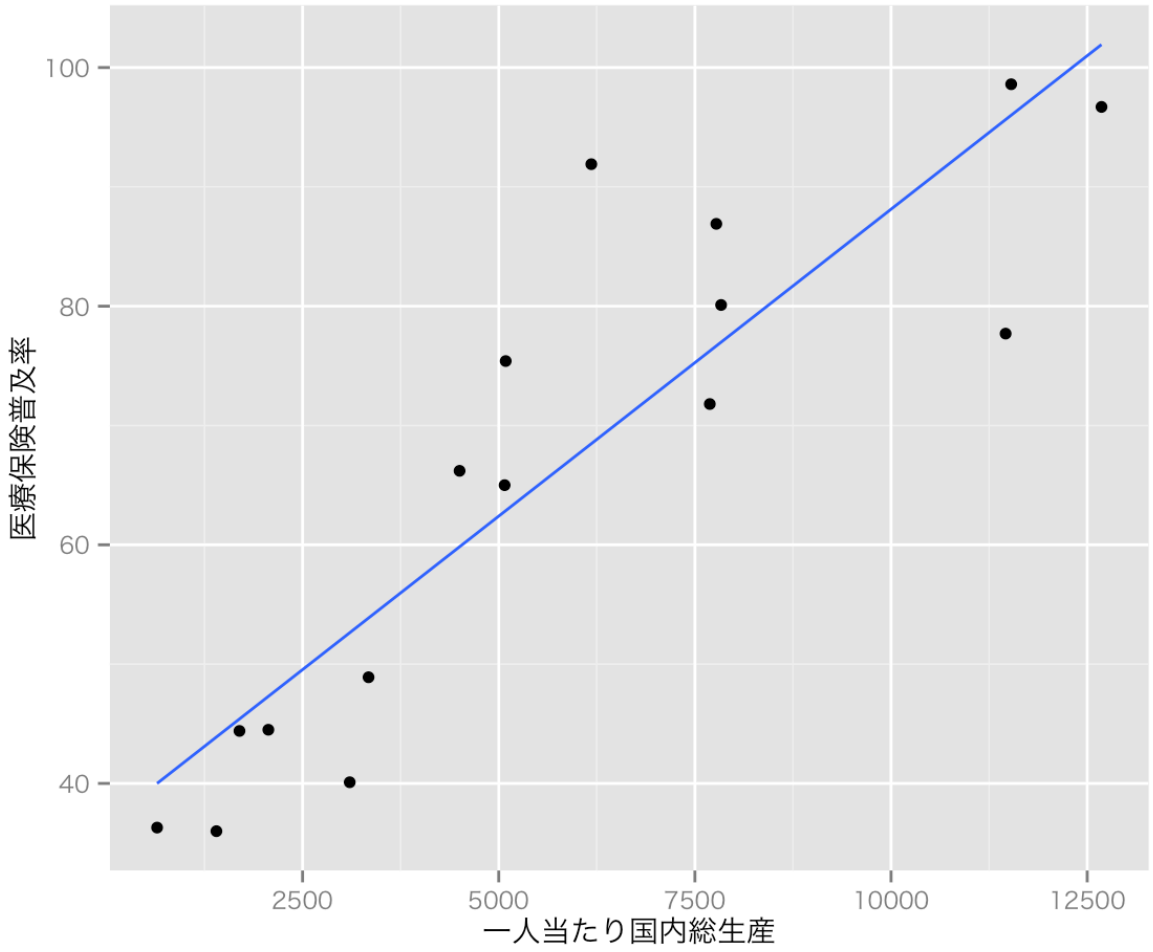


Figure 2-1

Figure 2-1 represents the relation between health insurance coverage and gdp per capita in Latin America.

```
## import data and reshape
rawdata <- read.csv("dataset/rawdata.csv", skip = 4)
fig.21.data <- filter(rawdata, year > 2003)
fig.21.data <- filter(fig.21.data, health_coverage_total !=
"NA")
fig.21.data <- fig.21.data[-6,]
fig.21.data <- fig.21.data[-11,]

## make scatter plot
fig.21 <- ggplot(fig.21.data, aes(x = gdp_per_capita, y =
health_coverage_total))
fig.21 <- fig.21 + geom_point() +
  geom_smooth(method = 'lm', se = F) +
  labs(x = "一人当たり国内総生産", y = "医療保険普及率")
print(fig.21)
```



```
##====================== Note ====================##
## if you want to save figures in pdf,  put
## quartz(file = "mortality.pdf", family = "sans",
##        type = "pdf", width = 8, height = 6)
## before print(fig.21), and put
## dev.off()
## after print(fig.21)
##====================== Note ====================##
```

Tables

In section three of my master thesis, I conducted linear regression analysis. Before the analysis, you need to manually calculate gini coeffient of helath insurance coverage by sector and income stratification. Note that his step can be ignored as long as your interest is focused on replicating the tables. If so, jump to Table 1-1.

(This step is rather notes to record data set creating procedure.)

```
## create function that calculates gini coefficient
gini <- function(y) {
  n <- length(y)
  y <- sort(y)
  y <- cumsum(y)
  y <- c(0, y / y[n])
  x <- seq(0, 1, len = n + 1)
  sum(x - y) / n * 2
}

## health insurance coverage based on sector
argentina <- c(95.2,  83.4, 80.4,   50.8,   45.6)
a1 <- gini(argentina)
bolivia <- c(83.4,  44.1,   23.3,   17.4,   15.5)
a2 <- gini(bolivia)
chile <- c(97.1,  96.9, 97, 93.9,   97.2)
a3 <- gini(chile)
colombia <- c(100,  95.6,   86.3,   82.3,   88.3)
a4 <- gini(colombia)
costarica <- c(99.7,  90.9, 80.2,   66.6,   76.5)
a5 <- gini(costarica)
dominica <- c(96.2,  81.7,  75, 32.6,   36.7)
```

```
a6 <- gini(dominica)
ecuador <- c(95.9,  74.9,   47.6,   24, 46.8)
a7 <- gini(ecuador)
salvador <- c(94.9,  65.9,  43.2,   8.7,    7.4)
a8 <- gini(salvador)
guatemala <- c(86.2,  60,   42.9,   9.7,    8.3)
a9 <- gini(guatemala)
honduras <- c(66,  61.5,    22.3,   1.7,    3.1)
a10 <- gini(honduras)
mexico <- c(82,  64.1,  52.7,   62.1)
a11 <- gini(mexico)
nicaragua <- c(79.6,  47.6, 14.8,   3.1,    2.3)
a12 <- gini(nicaragua)
panama <- c(97,  88.2,  71.2,   32, 47.6)
a13 <- gini(panama)
paraguay <- c(75.1,  47.9,  54.6,   14.1,   15)
a14 <- gini(paraguay)
peru <- c(89,  68.5,    44.6,   41.9,   51.1)
a15 <- gini(peru)
uruguay <- c(99.9,  99.1,   96.4,   95.1,   98)
a16 <- gini(uruguay)

sector <- c(a1, a2, a3, a4, a5, a6, a7, a8, a9, a10, a11, a12,
a13, a14, a15, a16)

## create vectors consisting health insurance coverage
## based on income stratification of health.coverage
argentina <- c(40.3, 65.5, 76, 86, 95)
a1 <- gini(argentina)
bolivia <- c(27.8  ,28.1    ,37.1   ,46.4   ,54.4)
a2 <- gini(bolivia)
chile <- c(97.4,  96.9, 96.6,   95.9,   97.3)
a3 <- gini(chile)
colombia <- c(83.2,  87.3,  90.2,   92, 95.7)
a4 <- gini(colombia)
costarica <- c(68,  78.2,   85.7,   89.5,   95.7)
a5 <- gini(costarica)
dominica <- c(58.1,  69.9,  72.7,   81.2,   84.4)
a6 <- gini(dominica)
ecuador <- c(33.2,  47.7,   59.7,   71.6,   86.7)
a7 <- gini(ecuador)
salvador <- c(8.8,  22.6,   38.3,   52.9,   75.8)
```

```
a8 <- gini(salvador)
guatemala <- c(9.1,  21,    36, 49.6,   69.2)
a9 <- gini(guatemala)
honduras <- c(1.2,  8.5,    22, 41.4,   59.5)
a10 <- gini(honduras)
mexico <- c(58.9,  65.2,    70.8,   74, 80.9)
a11 <- gini(mexico)
nicaragua <- c(9.4,  15.9, 28,  40.4,   57.3)
a12 <- gini(nicaragua)
panama <- c(41.3,  67.5,    75.9,   86.8,   90)
a13 <- gini(panama)
paraguay <- c(5.5,  18.8,   32.4,   41.3,   59.3)
a14 <- gini(paraguay)
peru <- c(64.7,  59.5,  58.2,   63.6,   73.3)
a15 <- gini(peru)
uruguay <- c(97.3,  97.9,   98.3,   99, 99.5)
a16 <- gini(uruguay)

quintile <- c(a1, a2, a3, a4, a5, a6, a7, a8, a9, a10, a11, a1
2, a13, a14, a15, a16)

##====================== Note ====================##
##  (1) open "rawdata.csv"
##  (2) lag outcome variable manually,
##      making sure causes precede results
##  (3) manually substitute most recent data for NA
##  (4) rename it "regress.csv"
##====================== Note ====================##

## merge rawdata, sector and quintile
ols <- read.csv("dataset/regress.csv")
ols <- filter(ols, health_coverage_total != "NA")
write.csv(ols, file = "regression_dataset.csv")

##====================== Note ====================##
##  (1) open regression_dataset.csv
##  (2) erase odd number lines
##  (3) add president dummy based on table 3-1
##  (4) add sector and quintile
##  (5) save it and import again
##====================== Note ====================##
rm(ols)
```

Table 1-1

Table 1-1 shows health insurance coverage by sector in 16 Latin American countries.

(Since I used LaTeX to write my master thesis, all tables are outputted by LaTeX style table using `xtable()`.)

```
## import health insurance coverage data set.
health.coverage <- read.csv("dataset/health_coverage.csv", sk
ip = 1)

## health insurance coverage by sector
sector.table <- health.coverage[, -4:-5]
sector.table <- sector.table[, -4:-8]

## output table with LaTeX style
table.11 <- xtable(sector.table, align = "lcccccccc", label =
"tb-ref",
       caption = "部門別に見た医療保険普及率")
print(table.11)
```

```
## % latex table generated in R 3.2.1 by xtable 1.7-4 package
## % Sun Jan 17 10:26:46 2016
## \begin{table}[ht]
## \centering
## \begin{tabular}{lcccccccc}
##   \hline
##  & state & year & total & public\_sector & private\_sector
& microenterprise\_high & microenterprise\_low & domestic \\
##   \hline
## 1 & Argentina & 2004 & 65.10 & 74.80 & 75.50 & 71.70 & 40.4
0 & 32.20 \\
##   2 & Argentina & 2011 & 77.70 & 95.20 & 83.40 & 80.40 & 50
.80 & 45.60 \\
##   3 & Bolivia & 2002 & 29.80 & 68.20 & 29.30 & 24.60 & 6.10
& 5.60 \\
##   4 & Bolivia & 2009 & 44.40 & 83.40 & 44.10 & 23.30 & 17.4
0 & 15.50 \\
##   5 & Brazil & 2001 &  &  &  &  &  &  \\
##   6 & Brazil & 2011 &  &  &  &  &  &  \\
```

```
##   7 & Chile & 2000 & 92.40 & 97.00 & 92.80 & 84.50 & 85.50
& 90.70 \\
##   8 & Chile & 2011 & 96.70 & 97.10 & 96.90 & 97.00 & 93.90
& 97.20 \\
##   9 & Colombia & 1999 & 53.20 & 94.90 & 46.40 & 79.50 &  &
25.20 \\
##   10 & Colombia & 2011 & 91.90 & 100.00 & 95.60 & 86.30 & 8
2.30 & 88.30 \\
##   11 & Costa Rica & 2002 & 83.70 & 99.30 & 88.90 & 84.40 &
58.40 & 68.80 \\
##   12 & Costa Rica & 2011 & 86.90 & 99.70 & 90.90 & 80.20 &
66.60 & 76.50 \\
##   13 & Dominican Republic & 2005 & 51.10 & 77.40 & 58.10 &
27.60 & 9.40 &  \\
##   14 & Dominican Republic & 2011 & 75.40 & 96.20 & 81.70 &
75.00 & 32.60 & 36.70 \\
##   15 & Ecuador & 2002 & 45.00 & 88.00 & 48.20 & 38.60 & 12.
10 & 10.90 \\
##   16 & Ecuador & 2011 & 66.20 & 95.90 & 74.90 & 47.60 & 24.
00 & 46.80 \\
##   17 & El Salvador & 1999 & 50.00 & 90.30 & 62.50 & 48.60 &
11.20 & 6.30 \\
##   18 & El Salvador & 2010 & 48.90 & 94.90 & 65.90 & 43.20 &
8.70 & 7.40 \\
##   19 & Guatemala & 2002 & 38.50 & 82.60 & 53.60 & 22.70 & 6
.50 & 3.30 \\
##   20 & Guatemala & 2006 & 44.50 & 86.20 & 60.00 & 42.90 & 9
.70 & 8.30 \\
##   21 & Honduras & 2006 & 36.00 & 62.80 & 41.20 & 57.10 & 31
.00 & 2.00 \\
##   22 & Honduras & 2010 & 37.20 & 66.00 & 61.50 & 22.30 & 1.
70 & 3.10 \\
##   23 & Mexico & 2002 & 53.80 & 85.40 & 71.50 & 49.20 & 17.8
0 & 10.40 \\
##   24 & Mexico & 2010 & 71.80 &  & 82.00 & 64.10 & 52.70 & 6
2.10 \\
##   25 & Nicaragua & 2001 & 28.80 & 60.60 & 36.80 & 24.90 & 4
.80 & 2.50 \\
##   26 & Nicaragua & 2005 & 36.30 & 79.60 & 47.60 & 14.80 & 3
.10 & 2.30 \\
##   27 & Panama & 2002 & 73.20 & 96.60 & 83.00 & 51.70 & 29.5
0 & 29.50 \\
```

```
##   28 & Panama & 2011 & 80.10 & 97.00 & 88.20 & 71.20 & 32.0
0 & 47.60 \\
##   29 & Paraguay & 2000 & 33.20 & 75.30 & 44.80 & 41.30 & 9.
10 & 7.30 \\
##   30 & Paraguay & 2011 & 40.10 & 75.10 & 47.90 & 54.60 & 14
.10 & 15.00 \\
##   31 & Peru & 2001 & 39.30 & 82.10 & 37.60 & 59.60 & 12.50
& 13.20 \\
##   32 & Peru & 2011 & 65.00 & 89.00 & 68.50 & 44.60 & 41.90
& 51.10 \\
##   33 & Uruguay & 2002 & 98.20 & 98.90 & 98.70 & 98.30 & 95.
50 & 98.40 \\
##   34 & Uruguay & 2011 & 98.60 & 99.90 & 99.10 & 96.40 & 95.
10 & 98.00 \\
##   35 & Venezuela & 2002 &  &  &  &  &  &  \\
##   36 & Venezuela & 2011 &  &  &  &  &  &  \\
##   37 & Latin America & 2002 & 54.40 & 83.20 & 61.50 & 54.00
& 26.90 & 27.10 \\
##   38 & Latin America & 2011 & 66.40 & 90.40 & 72.80 & 59.00
& 39.20 & 44.30 \\
##    \hline
## \end{tabular}
## \caption{部門別に見た医療保険普及率}
## \label{tb-ref}
## \end{table}
```

```
##=================== Note ==================##
## Don't forget to erase unnecessary lines.
##=================== Note ==================##
```

Table 1-2

Table 1-2 demonstrates health insurance coverage by income stratification in 16 Latin American countries.

```
## health insurance coverage by income stratification
quintile.table <- health.coverage[, -4:-5]
quintile.table <- quintile.table[, -9:-13]

## output table with LaTeX style
table.12 <- xtable(quintile.table, align = "lcccccccc", label
= "tb-ref",
       caption = "所得階層別に見た医療保険普及率")
print(table.12)
```

```
## % latex table generated in R 3.2.1 by xtable 1.7-4 package
## % Sun Jan 17 10:26:46 2016
## \begin{table}[ht]
## \centering
## \begin{tabular}{lcccccccc}
##   \hline
##  & state & year & total & quintile1 & quintile2 & quintile3
& quintile4 & quintile5 \\
##   \hline
## 1 & Argentina & 2004 & 65.10 & 20.30 & 45.10 & 63.00 & 77.2
0 & 90.80 \\
##   2 & Argentina & 2011 & 77.70 & 40.30 & 65.50 & 76.00 & 86
.00 & 95.00 \\
##   3 & Bolivia & 2002 & 29.80 & 9.00 & 9.80 & 18.50 & 26.10
& 45.30 \\
##   4 & Bolivia & 2009 & 44.40 & 27.80 & 28.10 & 37.10 & 46.4
0 & 54.40 \\
##   5 & Brazil & 2001 &  &  &  &  &  &  \\
##   6 & Brazil & 2011 &  &  &  &  &  &  \\
##   7 & Chile & 2000 & 92.40 & 89.40 & 91.00 & 91.60 & 93.00
& 95.00 \\
##   8 & Chile & 2011 & 96.70 & 97.40 & 96.90 & 96.60 & 95.90
& 97.30 \\
##   9 & Colombia & 1999 & 53.20 & 17.40 & 33.40 & 41.90 & 53.
60 & 73.40 \\
##   10 & Colombia & 2011 & 91.90 & 83.20 & 87.30 & 90.20 & 92
.00 & 95.70 \\
##   11 & Costa Rica & 2002 & 83.70 & 63.20 & 73.60 & 80.50 &
86.50 & 92.00 \\
##   12 & Costa Rica & 2011 & 86.90 & 68.00 & 78.20 & 85.70 &
```

```
89.50 & 95.70 \\
##   13 & Dominican Republic & 2005 & 51.10 & 34.80 & 43.00 &
48.80 & 52.70 & 64.60 \\
##   14 & Dominican Republic & 2011 & 75.40 & 58.10 & 69.90 &
72.70 & 81.20 & 84.40 \\
##   15 & Ecuador & 2002 & 45.00 & 18.10 & 26.90 & 36.30 & 48.
30 & 70.30 \\
##   16 & Ecuador & 2011 & 66.20 & 33.20 & 47.70 & 59.70 & 71.
60 & 86.70 \\
##   17 & El Salvador & 1999 & 50.00 & 7.10 & 18.00 & 37.70 &
54.70 & 77.70 \\
##   18 & El Salvador & 2010 & 48.90 & 8.80 & 22.60 & 38.30 &
52.90 & 75.80 \\
##   19 & Guatemala & 2002 & 38.50 & 10.20 & 17.50 & 32.60 & 4
4.70 & 64.00 \\
##   20 & Guatemala & 2006 & 44.50 & 9.10 & 21.00 & 36.00 & 49
.60 & 69.20 \\
##   21 & Honduras & 2006 & 36.00 & 0.60 & 5.60 & 21.00 & 40.2
0 & 57.10 \\
##   22 & Honduras & 2010 & 37.20 & 1.20 & 8.50 & 22.00 & 41.4
0 & 59.50 \\
##   23 & Mexico & 2002 & 53.80 & 16.60 & 35.00 & 53.00 & 63.5
0 & 79.10 \\
##   24 & Mexico & 2010 & 71.80 & 58.90 & 65.20 & 70.80 & 74.0
0 & 80.90 \\
##   25 & Nicaragua & 2001 & 28.80 & 7.10 & 13.90 & 21.20 & 30
.80 & 44.10 \\
##   26 & Nicaragua & 2005 & 36.30 & 9.40 & 15.90 & 28.00 & 40
.40 & 57.30 \\
##   27 & Panama & 2002 & 73.20 & 39.10 & 62.00 & 71.00 & 76.3
0 & 78.40 \\
##   28 & Panama & 2011 & 80.10 & 41.30 & 67.50 & 75.90 & 86.8
0 & 90.00 \\
##   29 & Paraguay & 2000 & 33.20 & 2.60 & 9.90 & 19.10 & 33.5
0 & 54.00 \\
##   30 & Paraguay & 2011 & 40.10 & 5.50 & 18.80 & 32.40 & 41.
30 & 59.30 \\
##   31 & Peru & 2001 & 39.30 & 11.50 & 15.20 & 25.60 & 41.80
& 57.50 \\
##   32 & Peru & 2011 & 65.00 & 64.70 & 59.50 & 58.20 & 63.60
& 73.30 \\
##   33 & Uruguay & 2002 & 98.20 & 98.10 & 97.40 & 97.80 & 98.
```

```
10 & 99.30 \\
##   34 & Uruguay & 2011 & 98.60 & 97.30 & 97.90 & 98.30 & 99.
00 & 99.50 \\
##   35 & Venezuela & 2002 &  &  &  &  &  &  \\
##   36 & Venezuela & 2011 &  &  &  &  &  &  \\
##   37 & Latin America & 2002 & 54.40 & 27.80 & 37.30 & 47.50
& 57.60 & 71.40 \\
##   38 & Latin America & 2011 & 66.40 & 44.00 & 53.20 & 61.10
& 69.50 & 79.60 \\
##    \hline
## \end{tabular}
## \caption{所得階層別に見た医療保険普及率}
## \label{tb-ref}
## \end{table}
```

```
##================== Note ==================##
## Don't forget to erase unnecessary lines.
##================== Note ==================##
```

Table 3-2

Table 3-2 represents correlation coefficient of all variables contained in data set for linear regression.

```
## import data set
ols <-read.csv("dataset/regression_dataset.csv")
## calculate correlation coefficient
cor <- ols[, -1:-5] ## remove "state" and "year" and responsiv
e variables
cor <- cor[, -6:-7] ## remove irrelevant explanatory variables
cor <- cor[, -8:-11] ## remove irrelevant explanatory variable
s

##============================ Note ============================
==##
## correlation test is conducted as follows:
## null hypothesis: correlation is equal to zero
## alternative hypothesis: correlation is not equal to zero
## if p value < 0.05, accept null hypothesis
## else reject null hypothesis
##============================ Note ============================
==##

## plot the correlation table
cor.table <- rcorr(as.matrix(cor), type = "pearson") ## type
can be pearson or spearman
print(cor.table)
```

```
##                gdp_per_capita age65 trade_openness polity d
emocracy
## gdp_per_capita           1.00  0.84          -0.24   0.48
0.42
## age65                    0.84  1.00          -0.30   0.48
0.45
## trade_openness          -0.24 -0.30           1.00   0.20
-0.28
## polity                   0.48  0.48           0.20   1.00
0.22
## democracy                0.42  0.45          -0.28   0.22
1.00
## total_left               0.19  0.23           0.15  -0.02
0.08
## president                0.05  0.36          -0.21   0.08
0.30
```

```
##                total_left president
## gdp_per_capita       0.19      0.05
## age65                0.23      0.36
## trade_openness       0.15     -0.21
## polity              -0.02      0.08
## democracy            0.08      0.30
## total_left           1.00      0.25
## president            0.25      1.00
##
## n= 16
##
##
## P
##                gdp_per_capita age65  trade_openness polity
democracy
## gdp_per_capita                0.0000 0.3726         0.0580
0.1075
## age65          0.0000                0.2532         0.0595
0.0795
## trade_openness 0.3726         0.2532                0.4467
0.2855
## polity         0.0580         0.0595 0.4467
0.4211
## democracy      0.1075         0.0795 0.2855         0.4211
## total_left     0.4736         0.3914 0.5722         0.9270
0.7791
## president      0.8516         0.1744 0.4347         0.7808
0.2553
##                total_left president
## gdp_per_capita 0.4736     0.8516
## age65          0.3914     0.1744
## trade_openness 0.5722     0.4347
## polity         0.9270     0.7808
## democracy      0.7791     0.2553
## total_left                0.3498
## president      0.3498
```

```
##============================ Note ============================
==##
## (1) copy and paste "cor.table" at Excel
## (2) exclude variable "democracy" and and reload it
## (3) save it as correlation_table.csv
##============================ Note ============================
==##

rm(cor.table)
cor.table <- read.csv("dataset/correlation_table.csv")
table.32 <- xtable(cor.table, align="lcccccccc", label="tb-ref
", caption = "説明変数の相関係数表")
print(table.32)
```

```
## % latex table generated in R 3.2.1 by xtable 1.7-4 package
## % Sun Jan 17 10:26:46 2016
## \begin{table}[ht]
## \centering
## \begin{tabular}{lccccccc}
##   \hline
##  & X & gdp\_per\_capita & age65 & trade\_openness & polity
& democracy & total\_left & president \\
##   \hline
## 1 & gdp\_per\_capita & 1.00 & 0.73 & -0.21 & 0.37 & 0.47 &
0.20 & 0.17 \\
##   2 & age65 & 0.73 & 1.00 & -0.30 & 0.48 & 0.45 & 0.23 & 0.
36 \\
##   3 & trade\_openness & -0.21 & -0.30 & 1.00 & 0.20 & -0.28
& 0.15 & -0.21 \\
##   4 & polity & 0.37 & 0.48 & 0.20 & 1.00 & 0.22 & -0.02 & 0
.08 \\
##   5 & democracy & 0.47 & 0.45 & -0.28 & 0.22 & 1.00 & 0.08
& 0.30 \\
##   6 & total\_left & 0.20 & 0.23 & 0.15 & -0.02 & 0.08 & 1.0
0 & 0.25 \\
##   7 & president & 0.17 & 0.36 & -0.21 & 0.08 & 0.30 & 0.25
& 1.00 \\
##    \hline
## \end{tabular}
## \caption{説明変数の相関係数表}
## \label{tb-ref}
## \end{table}
```

Table 3-3

Table 3-3 shows summary statistics of data set used for linear regression analysis.

```
## show basic statistics
basic <- ols[, -1:-2]
basic <- basic[, -13:-16]
basic <- basic[, -9:-10]
descriptive <- describe(basic)
## xtable(descriptive, align="lccccccccc",
##    caption="変数の記述統計")

##====================== Note ====================##
## For some reasons, I could not execute xtable()
## in this R markdown document. But, if you use
## xtable() in R studio or R, you can get basic
## statistics with LaTeX style. Also, you can
## get summary statistics using functions
## embedded in R as follows:
## mean(x), sd(x), min(x), max(x)
##====================== Note ====================##
```

Table 3-4

Table 3-4 shows the result of linear regression analysis.

```
## take the natural logarithm
ols$gdp_per_capita <- log(ols$gdp_per_capita)

## ols analysis
ols.1 <- lm(health_coverage_total ~ gdp_per_capita + trade_ope
nness +
              + polity + democracy + total_left + president, d
ata = ols)
ols.2 <- lm(health_coverage_total ~ age65 + trade_openness +
              + polity + democracy + total_left + president, d
ata = ols)

## sector regression analysis
ols.3 <- lm(sector_gini ~ gdp_per_capita + trade_openness +
              + polity + democracy + total_left + president, d
ata = ols)
ols.4 <- lm(sector_gini ~ age65 + trade_openness +
              + polity + democracy + total_left + president, d
```

```
ata = ols)

## income quintile regression analysis
ols.5 <- lm(quintile_gini ~ gdp_per_capita + trade_openness +
              + polity + democracy + total_left + president, d
ata = ols)
ols.6 <- lm(quintile_gini ~ age65 + trade_openness +
              + polity + democracy + total_left + president, d
ata = ols)

label_explanatory <- c("一人当たり国内総生産の自然対数",
                       "65歳以上人口割合", "市場開放度",
                       "民主主義度合", "民主主義年数", "左派政党の割
合",
                       "大統領ダミー", "定数項")

table.34 <- stargazer(ols.1, ols.2, ols.3, ols.4, ols.5, ols.
6,
          digits = 3, digits.extra = 0, align = TRUE,
          #star.cutoffs = NA, omit.table.layout = "n", ## this
line is important!
          #keep.stat = c("n", "adj.rsq", "f"),
          df = FALSE,
          covariate.labels = label_explanatory,
          dep.var.caption = "Outcome variable",
          #dep.var.labels = "Vote share (%)",
          title = "線形回帰分析の推定結果",
          label = "tbl:reg-res",
          out = "stargazer-reg-res.tex")
```

```
##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard
University. E-mail: hlavac at fas.harvard.edu
## % Date and time: 日,  1 17, 2016 - 10時26分47秒
## % Requires LaTeX packages: dcolumn
## \begin{table}[!htbp] \centering
##   \caption{線形回帰分析の推定結果}
##   \label{tbl:reg-res}
## \begin{tabular}{@{\extracolsep{5pt}}lD{.}{.}{-3} D{.}{.}{-3
} D{.}{.}{-3} D{.}{.}{-3} D{.}{.}{-3} D{.}{.}{-3} }
## \\[-1.8ex]\hline
```

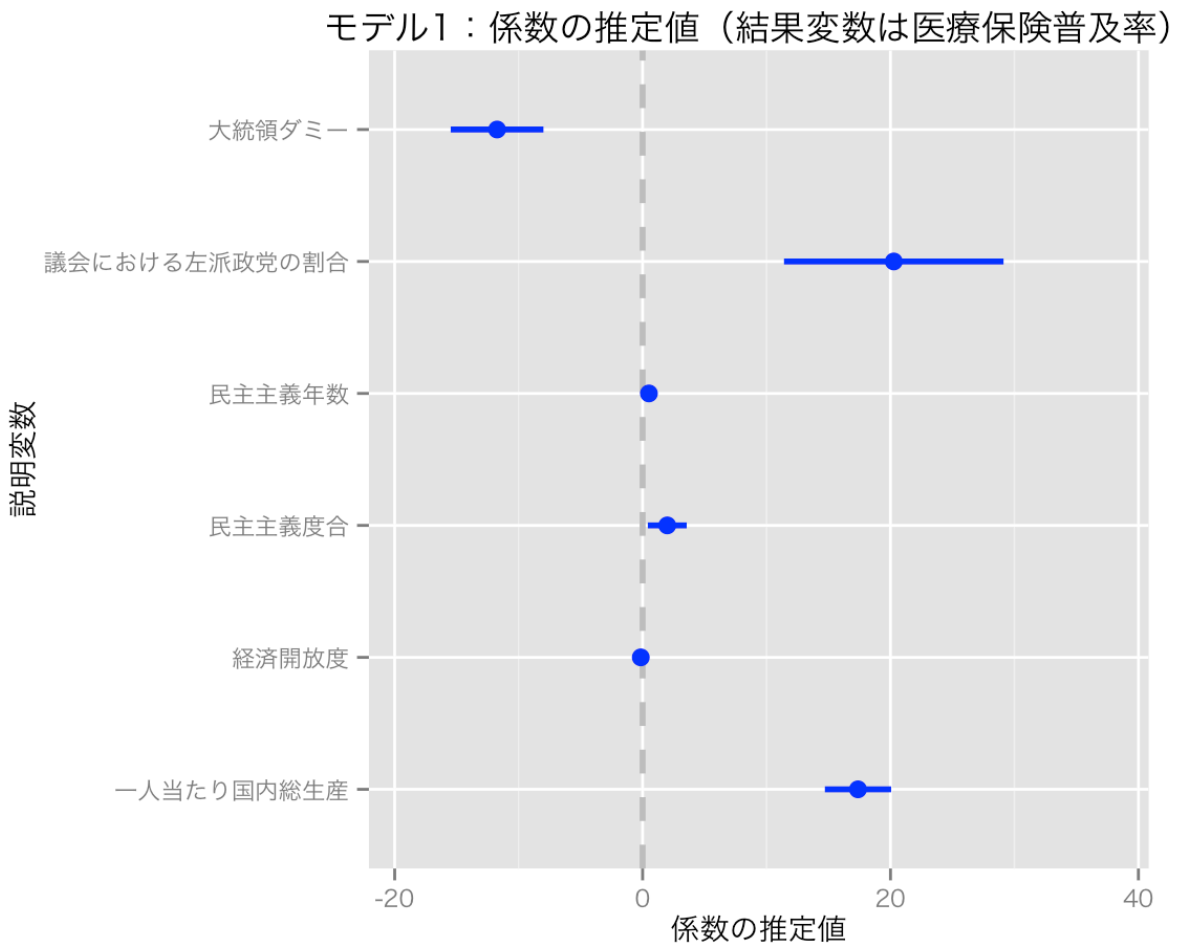
```
## \hline \\[-1.8ex]
##  & \multicolumn{6}{c}{Outcome variable} \\
## \cline{2-7}
## \\[-1.8ex] & \multicolumn{2}{c}{health\_coverage\_total} &
\multicolumn{2}{c}{sector\_gini} & \multicolumn{2}{c}{quintile
\_gini} \\
## \\[-1.8ex] & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(
2)} & \multicolumn{1}{c}{(3)} & \multicolumn{1}{c}{(4)} & \mul
ticolumn{1}{c}{(5)} & \multicolumn{1}{c}{(6)}\\
## \hline \\[-1.8ex]
##  一人当たり国内総生産の自然対数 & 17.377^{***} &  & -0.154^{***}
&  & -0.084^{*} &  \\
##   & (2.672) &  & (0.028) &  & (0.042) &  \\
##   & & & & & & \\
##  65歳以上人口割合 &  & 4.100^{*} &  & -0.023 &  & -0.007 \\
##   &  & (1.946) &  & (0.020) &  & (0.019) \\
##   & & & & & & \\
##  市場開放度 & -0.145^{*} & -0.127 & 0.001 & 0.001 & 0.002 & 0
.002 \\
##   & (0.074) & (0.156) & (0.001) & (0.002) & (0.001) & (0.00
1) \\
##   & & & & & & \\
##  民主主義度合 & 1.991 & 1.752 & -0.017 & -0.028 & -0.025 & -0
.037 \\
##   & (1.568) & (3.458) & (0.017) & (0.036) & (0.024) & (0.03
3) \\
##   & & & & & & \\
##  民主主義年数 & 0.511^{***} & 0.707^{**} & -0.002 & -0.004 &
-0.002 & -0.003 \\
##   & (0.143) & (0.270) & (0.002) & (0.003) & (0.002) & (0.00
3) \\
##   & & & & & & \\
##  左派政党の割合 & 20.264^{**} & 23.765 & -0.065 & -0.136 & -0.
114 & -0.170 \\
##   & (8.854) & (17.640) & (0.094) & (0.185) & (0.138) & (0.1
68) \\
##   & & & & & & \\
##  大統領ダミー & -11.739^{**} & -16.150^{*} & 0.049 & 0.080 &
0.032 & 0.046 \\
##   & (3.740) & (7.383) & (0.040) & (0.078) & (0.058) & (0.07
0) \\
##   & & & & & & \\
```

```
##  定数項 & -101.346^{***} & 13.261 & 1.646^{***} & 0.652^{**}
& 1.036^{***} & 0.500^{*} \\
##   & (20.398) & (23.286) & (0.217) & (0.245) & (0.318) & (0.
222) \\
##   & & & & & & \\
## \hline \\[-1.8ex]
## Observations & \multicolumn{1}{c}{16} & \multicolumn{1}{c}{
16} & \multicolumn{1}{c}{16} & \multicolumn{1}{c}{16} & \multi
column{1}{c}{16} & \multicolumn{1}{c}{16} \\
## R$^{2}$ & \multicolumn{1}{c}{0.944} & \multicolumn{1}{c}{0.
785} & \multicolumn{1}{c}{0.896} & \multicolumn{1}{c}{0.612} &
\multicolumn{1}{c}{0.677} & \multicolumn{1}{c}{0.537} \\
## Adjusted R$^{2}$ & \multicolumn{1}{c}{0.906} & \multicolumn
{1}{c}{0.641} & \multicolumn{1}{c}{0.826} & \multicolumn{1}{c}
{0.354} & \multicolumn{1}{c}{0.461} & \multicolumn{1}{c}{0.228
} \\
## Residual Std. Error & \multicolumn{1}{c}{6.703} & \multicol
umn{1}{c}{13.097} & \multicolumn{1}{c}{0.071} & \multicolumn{1
}{c}{0.138} & \multicolumn{1}{c}{0.104} & \multicolumn{1}{c}{0
.125} \\
## F Statistic & \multicolumn{1}{c}{25.110$^{***}$} & \multico
lumn{1}{c}{5.470$^{**}$} & \multicolumn{1}{c}{12.906$^{***}$}
& \multicolumn{1}{c}{2.368} & \multicolumn{1}{c}{3.137$^{*}$}
& \multicolumn{1}{c}{1.737} \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{6}{r}{$^{*}$p$<$0.1; $^{**}$
p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

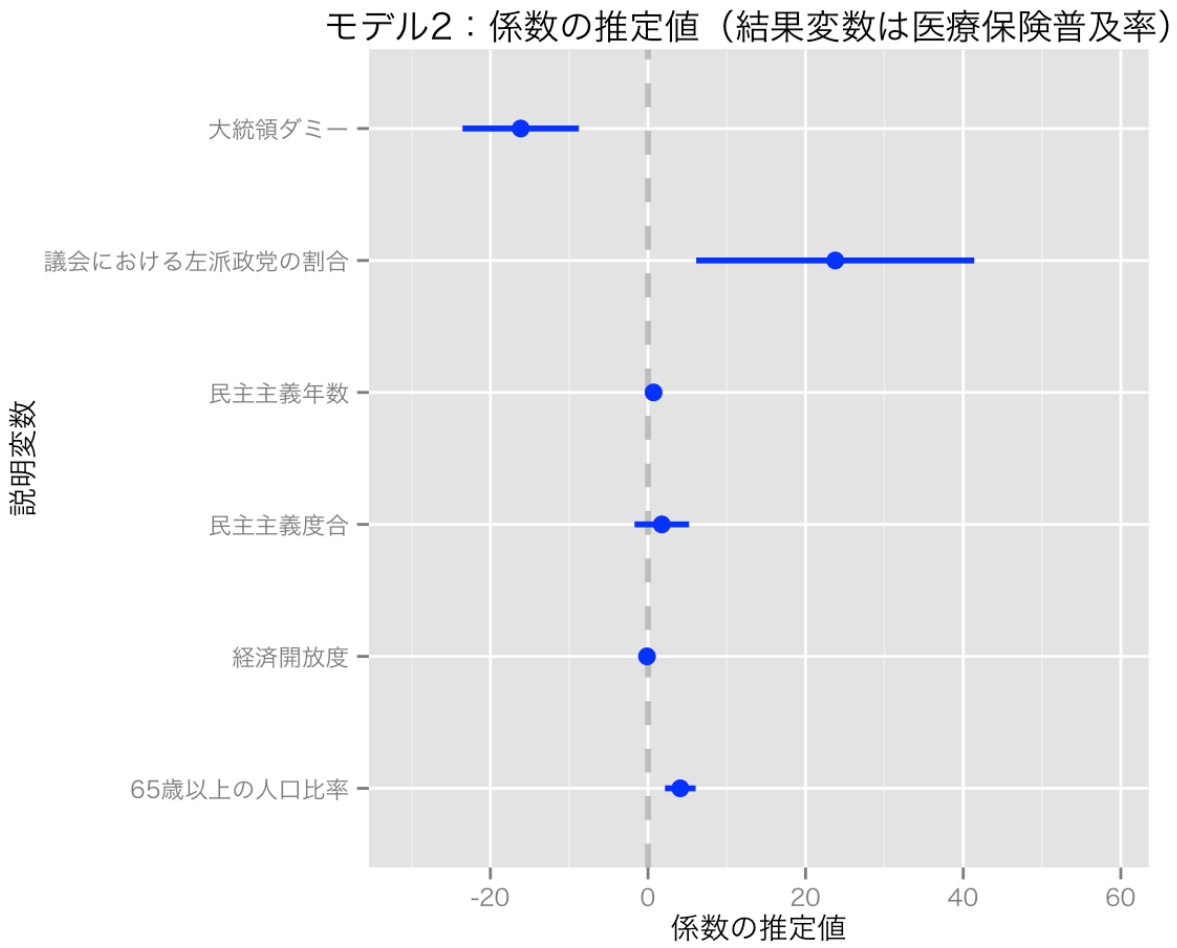
Showing Results of Linear Regression in Figures

In a general principle, results of statistical analysis are displayed in table. On the ohter hand, presenting the results of regression analysis with figures is much easier to intuitively understand because human beings have been very skillful to grasp the main points visually and geometrically. For this reason, I introduce *coefplot* to visualize the results of regression analysis.

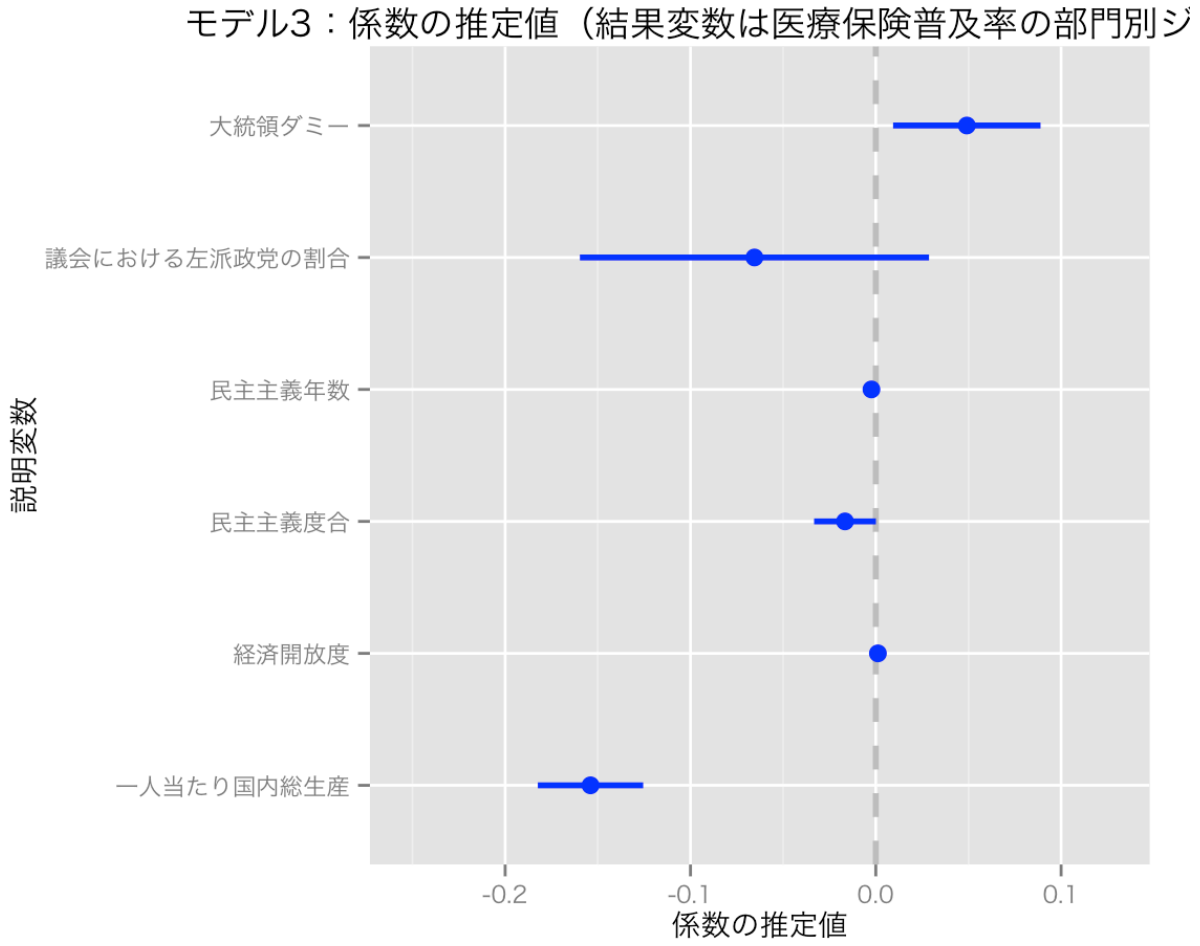
```
##========================== Note ==========================##
## (1) Points indicate estimated value
## (2) Lines show 95 percent confidence intervance
##========================== Note ==========================##

## display the result of regression using coef-plot
coefplot(ols.1, intercept = F,
         title = "モデル1：係数の推定値（結果変数は医療保険普及率）",
         xlab = "係数の推定値", ylab = "説明変数",
         newNames = c(gdp_per_capita = "一人当たり国内総生産", tra
de_openness = "経済開放度",
                      polity = "民主主義度合", democracy = "民主
主義年数",
                      total_left = "議会における左派政党の割合", pr
esident = "大統領ダミー"))
```



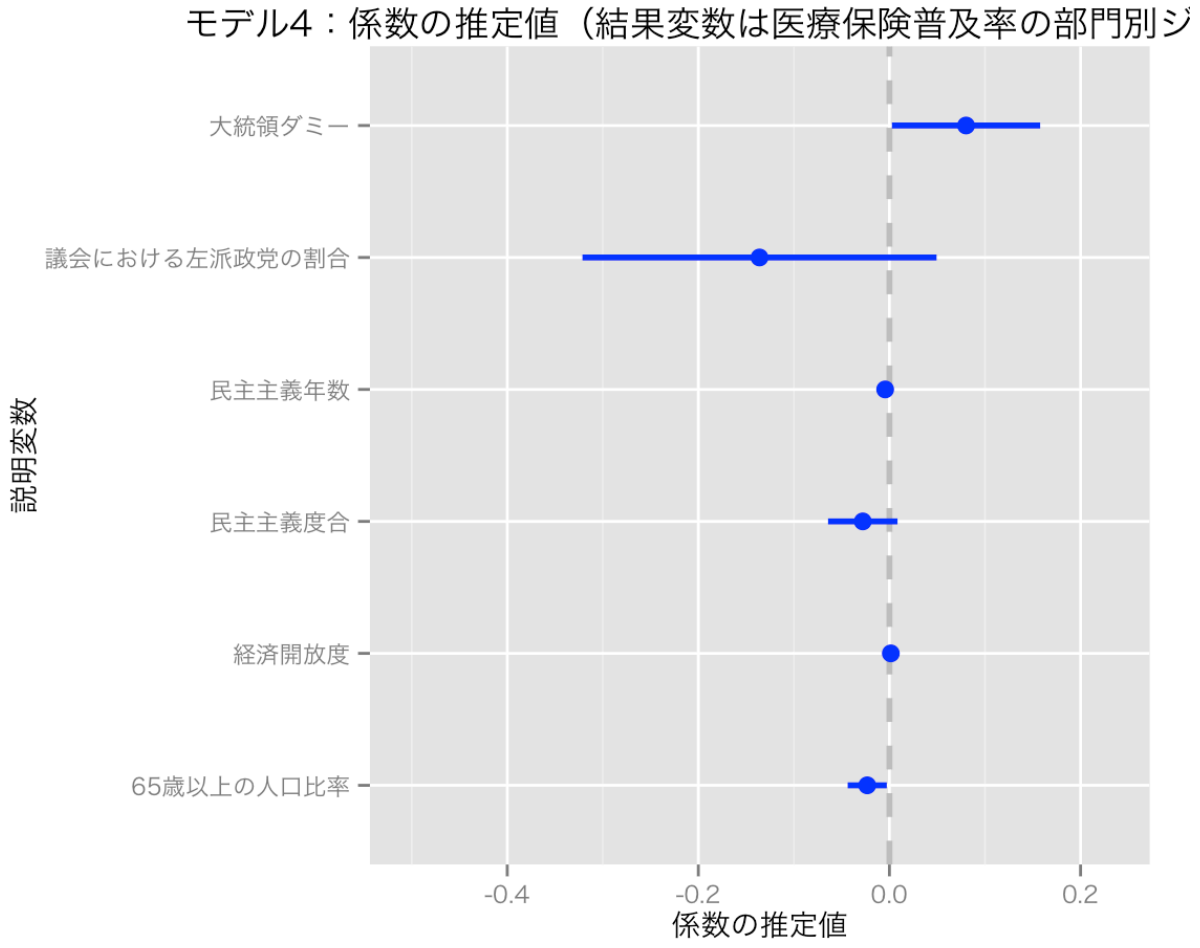
```
coefplot(ols.2, intercept = F,
         title = "モデル2：係数の推定値（結果変数は医療保険普及率）",
         xlab = "係数の推定値", ylab = "説明変数",
         newNames = c(age65 = "65歳以上の人口比率", trade_openness = "経済開放度",
                      polity = "民主主義度合", democracy = "民主主義年数",
                      total_left = "議会における左派政党の割合", president = "大統領ダミー"))
```



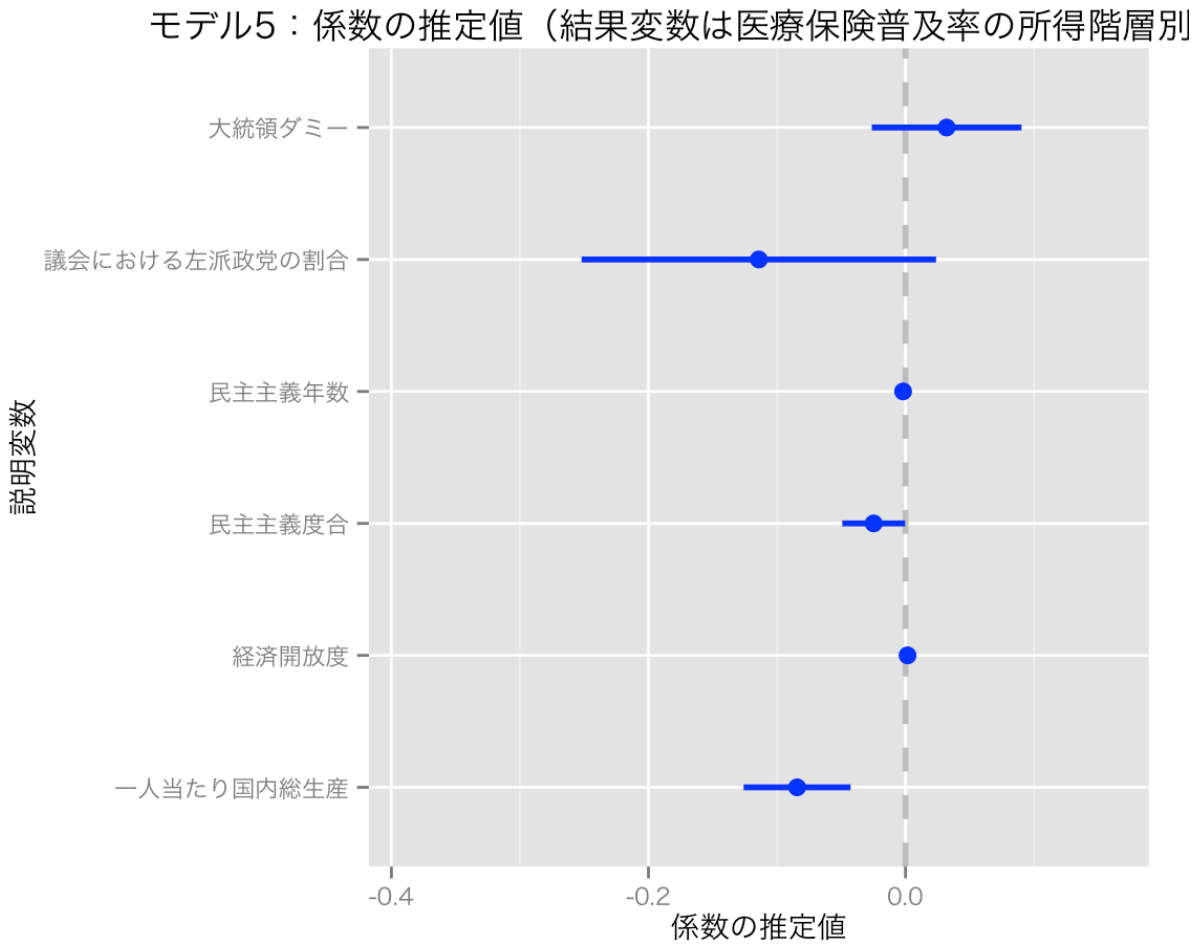
```
coefplot(ols.3, intercept = F,
         title = "モデル3：係数の推定値（結果変数は医療保険普及率の部門別ジニ係数）",
         xlab = "係数の推定値", ylab = "説明変数",
         newNames = c(gdp_per_capita = "一人当たり国内総生産", trade_openness = "経済開放度",
                      polity = "民主主義度合", democracy = "民主主義年数",
                      total_left = "議会における左派政党の割合", president = "大統領ダミー"))
```



```
coefplot(ols.4, intercept = F,
         title = "モデル4：係数の推定値（結果変数は医療保険普及率の部門別ジニ係数）",
         xlab = "係数の推定値", ylab = "説明変数",
         newNames = c(age65 = "65歳以上の人口比率", trade_openness = "経済開放度",
                      polity = "民主主義度合", democracy = "民主主義年数",
                      total_left = "議会における左派政党の割合", president = "大統領ダミー"))
```



```
coefplot(ols.5, intercept = F,
         title = "モデル5：係数の推定値（結果変数は医療保険普及率の所得階層別ジニ係数）",
         xlab = "係数の推定値", ylab = "説明変数",
         newNames = c(gdp_per_capita = "一人当たり国内総生産", trade_openness = "経済開放度",
                      polity = "民主主義度合", democracy = "民主主義年数",
                      total_left = "議会における左派政党の割合", president = "大統領ダミー"))
```



```
coefplot(ols.6, intercept = F,
         title = "モデル6：係数の推定値（結果変数は医療保険普及率の所得階層別ジニ係数）",
         xlab = "係数の推定値", ylab = "説明変数",
         newNames = c(age65 = "65歳以上の人口比率", trade_openness = "経済開放度",
                      polity = "民主主義度合", democracy = "民主主義年数",
                      total_left = "議会における左派政党の割合", president = "大統領ダミー"))
```