平成28年度社会医学実習-実施例

王　超辰

2016年6月23日

1　例:日本人肝がん罹患の年齢,出生コホート,時期効果分析

```
## Load XLConnect
library(XLConnect)
```

```
## Loading required package: XLConnectJars
```

```
## XLConnect 0.2-11 by Mirai Solutions GmbH [aut],
##   Martin Studer [cre],
##   The Apache Software Foundation [ctb, cph] (Apache POI, Apache Commons
##     Codec),
##   Stephen Colebourne [ctb, cph] (Joda-Time Java library)
```

```
## http://www.mirai-solutions.com ,
## http://miraisolutions.wordpress.com
```

```
## From a newly created file with sheet 4 (rate data) only
rate.all <- readWorksheetFromFile("cancer_incidence(1975-2011)rate.xls",
                                  sheet = 1)
## Change variable names
names(rate.all) <- gsub("X", "age", names(rate.all))
names(rate.all) <- gsub(" ", "", names(rate.all))
names(rate.all) <- gsub(" ", "plus", names(rate.all))
names(rate.all) <- gsub("  ", "Dia_yr", names(rate.all))
names(rate.all) <- gsub("\\.", "_", names(rate.all))
## Show data
head(rate.all)
```

```
##          ICD_10     Dia_yr         age0_4   age5_9 age10_14
## 1      1    C00-C96      1975 184.6549 13.24920 8.021910 7.944879
## 2      1    C00-C96      1976 185.0981 13.14640 7.311146 7.597889
## 3      1    C00-C96      1977 189.4090 13.86778 7.400872 7.299964
## 4      1    C00-C96      1978 194.5231 13.11493 7.108764 6.130980
## 5      1    C00-C96      1979 206.5175 13.38973 6.660657 5.638117
## 6      1    C00-C96      1980 214.4543 14.02163 7.107233 6.551611
##    age15_19 age20_24 age25_29 age30_34 age35_39 age40_44 age45_49 age50_54
## 1  8.882128 13.18414 24.54009 41.92178 71.88043 128.1969 211.0058 300.2575
## 2  8.891426 11.32447 25.69828 38.42129 70.01396 125.0030 205.6629 296.9398
```

```
## 3 10.551438 10.42322 26.78686 38.13225 72.73761 125.9141 206.6744 297.4311
## 4 10.116003 10.40606 23.60261 38.56832 75.24708 126.1025 205.0247 308.5769
## 5  9.655429 11.11250 22.93809 41.57780 84.30307 127.7256 211.8893 326.0116
## 6  8.643361 12.09025 20.82652 43.88338 81.21430 127.6162 213.8217 327.8601
##   age55_59 age60_64 age65_69 age70_74 age75_79 age80_84 age85plus
## 1 430.9053 639.0219 884.1888 1173.113 1377.386 1360.574  1087.513
## 2 417.5249 618.8536 876.7602 1147.714 1389.089 1407.351  1121.981
## 3 405.1523 625.2349 895.6720 1114.264 1376.906 1431.072  1137.963
## 4 406.5789 621.4057 903.8451 1125.717 1400.840 1400.000  1208.894
## 5 438.9733 656.5278 910.9551 1170.201 1436.000 1442.207  1258.635
## 6 456.2925 656.3355 911.2209 1235.173 1448.579 1500.307  1314.581
```

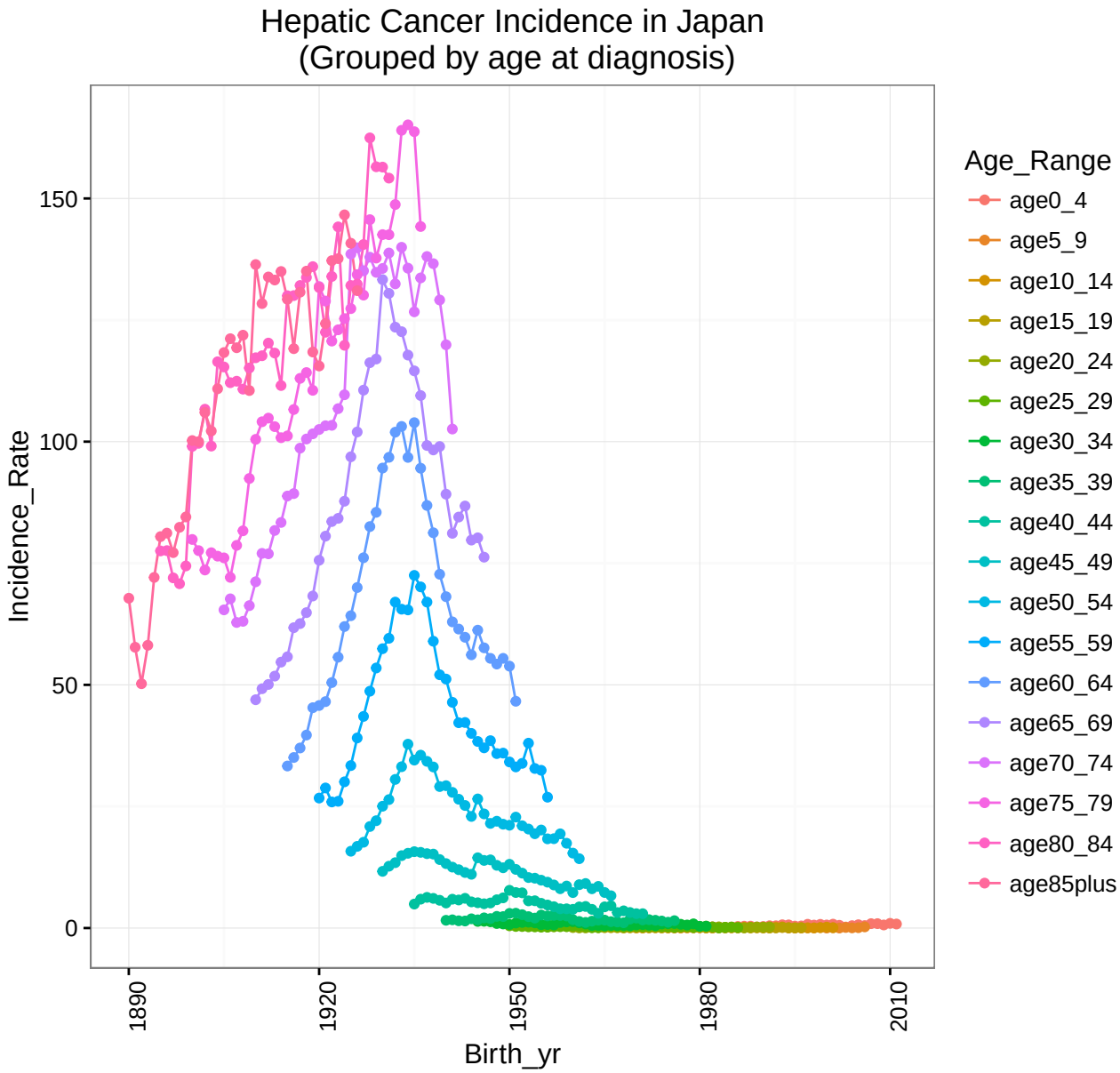
1.1 Graphing hepatic cancer data

```
## Extract all-sex data hepatic cancer mortality data
rate.hepatic <- subset(rate.all,   == " " &   == "  ")
## Change to long format
library(reshape2)
rate.hepatic.melt <- melt(data          = rate.hepatic,
                ##id.vars       = c(),
                measure.vars  = names(rate.hepatic)[grep("age", names(rate.hepatic))],
                variable.name = "Age_Range",
                value.name    = "Incidence_Rate"
                          )
names(rate.hepatic.melt$Age_Range) <- gsub("_", "-",
                                           as.character(rate.hepatic.melt$Age_Range))

## Regroup calendar year of death by five year intervals
rate.hepatic.melt$Cal_yr5 <- cut(rate.hepatic.melt$Dia_yr,
                                 breaks = seq(from = 1974, to = 2015, by = 5))

## Create a variable representing the lowest age in the interval
rate.hepatic.melt$age <- seq(from = 0, to = 85,
                             by = 5)[rate.hepatic.melt$Age_Range]

## Calculate the year of birth
rate.hepatic.melt$Birth_yr <- with(rate.hepatic.melt,
                                   Dia_yr - age)

## Create the year of birth categories
rate.hepatic.melt$Birth_yr5 <- cut(rate.hepatic.melt$Birth_yr,
                                   breaks = seq(from = 1889, to = 2015, by = 5))
rate.hepatic.melt$Birth_yr30 <- cut(rate.hepatic.melt$Birth_yr,
                                    breaks = seq(from = 1870, to = 2030, by = 30))

## Check first 20 rows
```

```r
head(rate.hepatic.melt, 20)
```

```
##         ICD_10     Dia_yr        Age_Range Incidence_Rate
## 1       8      C22      1975  9.679323    age0_4      0.6699593
## 2       8      C22      1976 10.232036    age0_4      0.7111653
## 3       8      C22      1977 10.302749    age0_4      0.6767309
## 4       8      C22      1978 11.002483    age0_4      0.5081630
## 5       8      C22      1979 11.981952    age0_4      0.3948111
## 6       8      C22      1980 12.930932    age0_4      0.2818418
## 7       8      C22      1981 14.055343    age0_4      0.2070898
## 8       8      C22      1982 15.245212    age0_4      0.2383940
## 9       8      C22      1983 16.551309    age0_4      0.1801106
## 10      8      C22      1984 18.210172    age0_4      0.1572533
## 11      8      C22      1985 19.452466    age0_4      0.2010923
## 12      8      C22      1986 20.603754    age0_4      0.3281378
## 13      8      C22      1987 22.210135    age0_4      0.3770423
## 14      8      C22      1988 22.936400    age0_4      0.3876525
## 15      8      C22      1989 23.603099    age0_4      0.2227171
## 16      8      C22      1990 26.152168    age0_4      0.3234304
## 17      8      C22      1991 27.076901    age0_4      0.4731861
## 18      8      C22      1992 28.361939    age0_4      0.4672144
## 19      8      C22      1993 28.694175    age0_4      0.6729033
## 20      8      C22      1994 28.142745    age0_4      0.6448413
##        Cal_yr5 age Birth_yr   Birth_yr5          Birth_yr30
## 1  (1974,1979]   0     1975 (1974,1979] (1.96e+03,1.99e+03]
## 2  (1974,1979]   0     1976 (1974,1979] (1.96e+03,1.99e+03]
## 3  (1974,1979]   0     1977 (1974,1979] (1.96e+03,1.99e+03]
## 4  (1974,1979]   0     1978 (1974,1979] (1.96e+03,1.99e+03]
## 5  (1974,1979]   0     1979 (1974,1979] (1.96e+03,1.99e+03]
## 6  (1979,1984]   0     1980 (1979,1984] (1.96e+03,1.99e+03]
## 7  (1979,1984]   0     1981 (1979,1984] (1.96e+03,1.99e+03]
## 8  (1979,1984]   0     1982 (1979,1984] (1.96e+03,1.99e+03]
## 9  (1979,1984]   0     1983 (1979,1984] (1.96e+03,1.99e+03]
## 10 (1979,1984]   0     1984 (1979,1984] (1.96e+03,1.99e+03]
## 11 (1984,1989]   0     1985 (1984,1989] (1.96e+03,1.99e+03]
## 12 (1984,1989]   0     1986 (1984,1989] (1.96e+03,1.99e+03]
## 13 (1984,1989]   0     1987 (1984,1989] (1.96e+03,1.99e+03]
## 14 (1984,1989]   0     1988 (1984,1989] (1.96e+03,1.99e+03]
## 15 (1984,1989]   0     1989 (1984,1989] (1.96e+03,1.99e+03]
## 16 (1989,1994]   0     1990 (1989,1994] (1.96e+03,1.99e+03]
## 17 (1989,1994]   0     1991 (1989,1994] (1.99e+03,2.02e+03]
## 18 (1989,1994]   0     1992 (1989,1994] (1.99e+03,2.02e+03]
## 19 (1989,1994]   0     1993 (1989,1994] (1.99e+03,2.02e+03]
## 20 (1989,1994]   0     1994 (1989,1994] (1.99e+03,2.02e+03]
```

```r
## Load ggplot2
library(ggplot2)

## Plot by calendar year, grouped by age of diagnosis
ggplot(data = rate.hepatic.melt,
       mapping = aes(x = Birth_yr, y = Incidence_Rate,
                     color = Age_Range)) +
  geom_line() +
  geom_point() +
    labs(title =
           "Hepatic Cancer Incidence in Japan\n (Grouped by age at diagnosis)") +
    theme_bw() +
    theme(legend.key = element_blank(),
          axis.text.x = element_text(angle=90, vjust=1))
```


Hepatic Cancer Incidence in Japan
(Grouped by age at diagnosis)
Incidence_Rate
150
100
50
0
1890
1920
1950
1980
2010
Birth_yr
Age_Range
age0_4
age5_9
age10_14
age15_19
age20_24
age25_29
age30_34
age35_39
age40_44
age45_49
age50_54
age55_59
age60_64
age65_69
age70_74
age75_79
age80_84
age85plus