平成28年度社会医学実習--実施例

王　超辰

2016年6月23日

# 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("歳", "", names(rate.all))  
names(rate.all) <- gsub("以上", "plus", names(rate.all))  
names(rate.all) <- gsub("診断年", "Dia\_yr", names(rate.all))  
names(rate.all) <- gsub("\\.", "\_", names(rate.all))  
## Show data  
head(rate.all)

## コード 部位 ICD\_10 性別 Dia\_yr 粗率 age0\_4 age5\_9 age10\_14  
## 1 1 全部位 C00-C96 男女計 1975 184.6549 13.24920 8.021910 7.944879  
## 2 1 全部位 C00-C96 男女計 1976 185.0981 13.14640 7.311146 7.597889  
## 3 1 全部位 C00-C96 男女計 1977 189.4090 13.86778 7.400872 7.299964  
## 4 1 全部位 C00-C96 男女計 1978 194.5231 13.11493 7.108764 6.130980  
## 5 1 全部位 C00-C96 男女計 1979 206.5175 13.38973 6.660657 5.638117  
## 6 1 全部位 C00-C96 男女計 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

## Graphing hepatic cancer data

## Extract all-sex data hepatic cancer mortality data  
rate.hepatic <- subset(rate.all, 部位 == "肝臓" & 性別 == "男女計")  
## 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  
head(rate.hepatic.melt, 20)

## コード 部位 ICD\_10 性別 Dia\_yr 粗率 Age\_Range Incidence\_Rate  
## 1 8 肝臓 C22 男女計 1975 9.679323 age0\_4 0.6699593  
## 2 8 肝臓 C22 男女計 1976 10.232036 age0\_4 0.7111653  
## 3 8 肝臓 C22 男女計 1977 10.302749 age0\_4 0.6767309  
## 4 8 肝臓 C22 男女計 1978 11.002483 age0\_4 0.5081630  
## 5 8 肝臓 C22 男女計 1979 11.981952 age0\_4 0.3948111  
## 6 8 肝臓 C22 男女計 1980 12.930932 age0\_4 0.2818418  
## 7 8 肝臓 C22 男女計 1981 14.055343 age0\_4 0.2070898  
## 8 8 肝臓 C22 男女計 1982 15.245212 age0\_4 0.2383940  
## 9 8 肝臓 C22 男女計 1983 16.551309 age0\_4 0.1801106  
## 10 8 肝臓 C22 男女計 1984 18.210172 age0\_4 0.1572533  
## 11 8 肝臓 C22 男女計 1985 19.452466 age0\_4 0.2010923  
## 12 8 肝臓 C22 男女計 1986 20.603754 age0\_4 0.3281378  
## 13 8 肝臓 C22 男女計 1987 22.210135 age0\_4 0.3770423  
## 14 8 肝臓 C22 男女計 1988 22.936400 age0\_4 0.3876525  
## 15 8 肝臓 C22 男女計 1989 23.603099 age0\_4 0.2227171  
## 16 8 肝臓 C22 男女計 1990 26.152168 age0\_4 0.3234304  
## 17 8 肝臓 C22 男女計 1991 27.076901 age0\_4 0.4731861  
## 18 8 肝臓 C22 男女計 1992 28.361939 age0\_4 0.4672144  
## 19 8 肝臓 C22 男女計 1993 28.694175 age0\_4 0.6729033  
## 20 8 肝臓 C22 男女計 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]

## 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))

