2007年12月8日

2007 年度共同研究集会 「データ解析環境Rの整備と利用」

疫学におけるRの活用

谷村 晋[*1]

[*1] 立命館アジア太平洋大学

1　疫学とは何か

- 人口集団の疾病を調べて、疾病のリスク要因を同定するための科学
- 疫学研究の段階
  1. 記述疫学（人、時間、場所について記述する。病原体、感染源、伝搬様式、リスク要因を考える）
  2. 分析疫学（リスク要因の仮説を検証する）
  3. 介入研究 (研究者がリスク要因をコントロールする実験研究)

2　疫学の研究デザイン

- 観察研究
  - 記述疫学
    - 地域相関研究（生態学的研究）、横断的研究
  - 分析疫学
    - 症例対照研究，コホート研究
- 実験研究
  - ランダム化対照試験
  - ワクチン野外試験

3　疫学関連のRパッケージ

Epi	0.7.0
epibasix	0.1
epicalc	2.6.1.0
epitools	0.4-9

4 Epi

- Epi: A package for statistical analysis in epidemiology.
- Functions for demographic and epidemiological analysis in the Lexis diagram, i.e. register and cohort follow-up data, including interval censored data.
- Also some useful functions for tabulation and plotting.
- Contains some epidemiological datasets.

- Version: 0.7.0
- Date: 2007-01-10
- Author: Bendix Carstensen, Martyn Plummer, Esa Läärä, Mark Myatt, David Clayton et. al.
- Maintainer: Bendix Carstensen
- License: GPL Version 2 or newer.

項目	説明
Icens	Fits a regression model to interval censored data.
Lexis	Split follow-up time in cohort studies.
Lexis.diagram	Plot a Lexis diagram
Lexis.lines	Draw life lines in a Lexis diagram.

項目	説明
Life.lines	Compute dates/ages for life lines in a Lexis diagram
ROC	Function to compute and draw ROC-curves.
Relevel	Reorder and combine levels of a factor
S.typh	Salmonella Typhimurium outbreak 1996 in Denmark.
apc.fit	Fit an Age-Period-Cohort model to tabular data.
apc.frame	Produce an empty frame for display of parameter-estimates from Age-Period-Cohort-models.
apc.lines	Plot APC-estimates in an APC-frame.
apc.plot	Plot the estimates from a fitted Age-Period-Cohort model
bdendo	A case-control study of endometrial cancer

項目	説明
bdendo11	A 1:1 subset of the endometrial cancer case-control study
births	Births in a London Hospital
blcaIT	Bladder cancer mortality in Italian males
brv	Bereavement in an elderly cohort
cal.yr	Functions to convert character, factor and various date objects into a number, and vice versa.
ccwc	Generate a nested case-control study
ci.cum	Compute cumulative sum of estimates.
ci.lin	Compute linear functions of parameters with s.e.
ci.pd	Compute confidence limits for a difference of two independent proportions.
count	Special functions for use in stat.table

項目	説明
detrend	Projection of a model matrix on to the orthogonal complement of a trend.
diet	Diet and heart data
epi.eff	Function to calculate effects
ewrates	Rates of lung and nasal cancer mortality, and total mortality.
ex1	Split follow-up time along a timescale
expand.data	Function to expand data for regression analysis of interval censored data.
fcut	Cuts follow-up time at multiple failure times.
fcut1	Cut follow-up time at a failure time.
fit.add	Fit an addive excess risk model to interval censored data.
fit.baseline	Fit a piecewise contsnt intesity model for interval censored data.

項目	説明
fit.mult	Fits a multiplicative relative risk model to interval censored data.
float	Calculate floated variances
ftrend	Fit a floating trend to a factor in generalized linear model
gmortDK	Population mortality rates for Denmark in 5-years age groups.
hivDK	hivDK: seroconversion in a cohort of Danish men
icut	Function to cut the follow-up in cohort at a point in time.
isec	Determine the intersection between follow-up intervals and a fixed interval.
lep	An unmatched case-control study of leprosy incidence
lungDK	Male lung cancer incidence in Denmark

項目	説明
mh	Mantel-Haenszel analyses of cohort and case-control studies
mortDK	Population mortality rates for Denmark in 1-year age-classes.
ncut	Function to group a variable in intervals.
nice	Nice breakpoints
nickel	A Cohort of Nickel Smelters in South Wales
pctab	Create percentages in a table
plotEst	Plot estimates with confidence limits
plotevent	Plot Equivalence Classes
projection.ip	Projection of columns of a matrix.
rateplot	Functions to plot rates from a table classified by age and calendar time (period)
stat.table	Tables of summary statistics
tabplot	Graphical display of a 2-way contingency table
thoro	Thorotrast Study

項目	説明
twoby2	Analysis of a two by two table

5 epibasix

- epibasix: Elementary Epidemiological Functions for a Graduate Epidemiology Biostatistics Course
- This package contains elementary tools for analysis of common epidemiological problems, ranging from sample size estimation, through 2x2 contingency table analysis and basic measures of agreement (kappa, sensitivity/specificity).

- Appropriate print and summary statements are also written to facilitate interpretation wherever possible.
- Version: 0.1
- Date: 2007-07-26
- Author: Michael A Rotondi
- Maintainer: Michael A Rotondi
- License: GPL (version 2 or later)

項目	説明
corXY	Correlation of Two Vectors

項目	説明
diffDetect	Mean Difference Detetion Tool
epi2x2	Epidemiological 2x2 Contingency Table Analysis Tool
epiKappa	Computation of the Kappa Statistic for Agreement Between Two Raters
epiTTest	Epidemiological T-Test Function
mcNemar	Pair-Matched Analysis Tool
n4means	Number of Subjects Required for a Randomized Trial with a Continuous Outcome
n4props	Number of Subjects Required for a Randomized Trial with Binary Outcomes
sensSpec	Sensitivity and Specificity Analysis of a 2x2 Matrix
univar	Univariate Analysis of a Single Variable

6 epicalc

- epicalc: Epidemiological calculator
- Functions making R easy for epidemiological calculation.
- Version: 2.6.1.0
- Date: 2007-12-05
- Author: Virasakdi Chongsuvivatwong
- Maintainer: Virasakdi Chongsuvivatwong
- License: GPL (>= 2)

項目	説明
ANCdata	Dataset on effect of new antenatal care method on mortality
ANCtable	Dataset on effect of new ANC method on mortality (as a table)
Attitudes	Dataset from an attitude survey among hospital staff
BP	Dataset on blood pressure and determinants
Bang	Dataset from 1988 Bangladesh Fertility Survey
Compaq	Dataset on cancer survival
DHF99	Dataset for exercise on predictors for mosquito larva infestation
Decay	Dataset on tooth decay and mutan streptococci

項目	説明
Ectopic	Dataset of a case-control study looking at history of abortion as a risk factor for ectopic pregnancy
Familydata	Dataset of a hypothetical family
HW93	Dataset from a study on hookworm prevalence and intensity in 1993
Hakimi	Dataset on effect of training personnel on neonatal mortality
Marryage	Dataset on age at marriage
Montana	Dataset on arsenic exposure and respiratory deaths
Oswego	Dataset from an outbreak of food poisoning in US
Outbreak	Dataset from an outbreak of food poisoning on a sportsday, Thailand 1990.

項目	説明
Planning	Dataset for practicing cleaning, labelling and recoding
SO2	Dataset on air pollution and deaths in UK
Sleep3	Dataset on sleepiness in a workshop
Suwit	Hookworm infection and blood loss: SEAJTM 1970
Timing	Dataset on time going to bed, waking up and arrival at the workshop
VC1to1	Datasets on a matched case-control study of esophageal cancer
adjust	Adjusted and standardize mean, proportion and rate
aggregate.numeric	Compute summary statistics of a numeric variable
alpha	Cronbach's alpha

項目	説明
be2ad	Change year in B.E. to A.D.
cc	Odds ratio calculation and graphing
ci	Confidence interval of probabilty, mean and incidence
codebook	Codebook of a data frame
des	Desription of a data frame or a variable
detachAllData	Detach all data frames
dotplot	Dot plot
expand	Expand an aggregated data frame
followup.plot	Longitudinal followup plot
kap	Kappa statistic
keepData	Keep a subset of variables or records
label.var	Variable manipulation
logistic.display	Tables for multivariate odds ratio, incidence density etc
lookup	Recode several values of a variable

項目	説明
lroc	ROC curve
lrtest	Likelihood ratio test
lsNoFunction	List non-function objects
matchTab	Matched tabulation
n.for.survey	Sample size calculation
poisgof	Goodness of fit test for modeling of count data
power.for.2means	Power calculation for two sample means and proportions
pyramid	Population pyramid
recode	Recode variable(s)
rename	Rename variable(s) in the default data frame
setTitle	Setting language of Epicalc graph title
shapiro.qqnorm	Qqnorm plots with Shapiro-Wilk's test
summ	Summary with graph

項目	説明
tab1	One-way tabulation
tableStack	Tabulation of variables in a stack form
tabpct	Two-way tabulation
titleString	Replace commonly used words in Epicalc graph title
unclassDataframe	Unclass factor(s) in the default data frame
use	Quick command to read in data
zap	Remove and detach all

7 epitools

- epitools: Epidemiology Tools
- EpiTools: R Package for Epidemiologic Data and Graphics
- Version: 0.4-9
- Date: 2007-06-27
- Author: Tomas Aragon
- Maintainer: Tomas Aragon
- License: GPL version 2 or newer

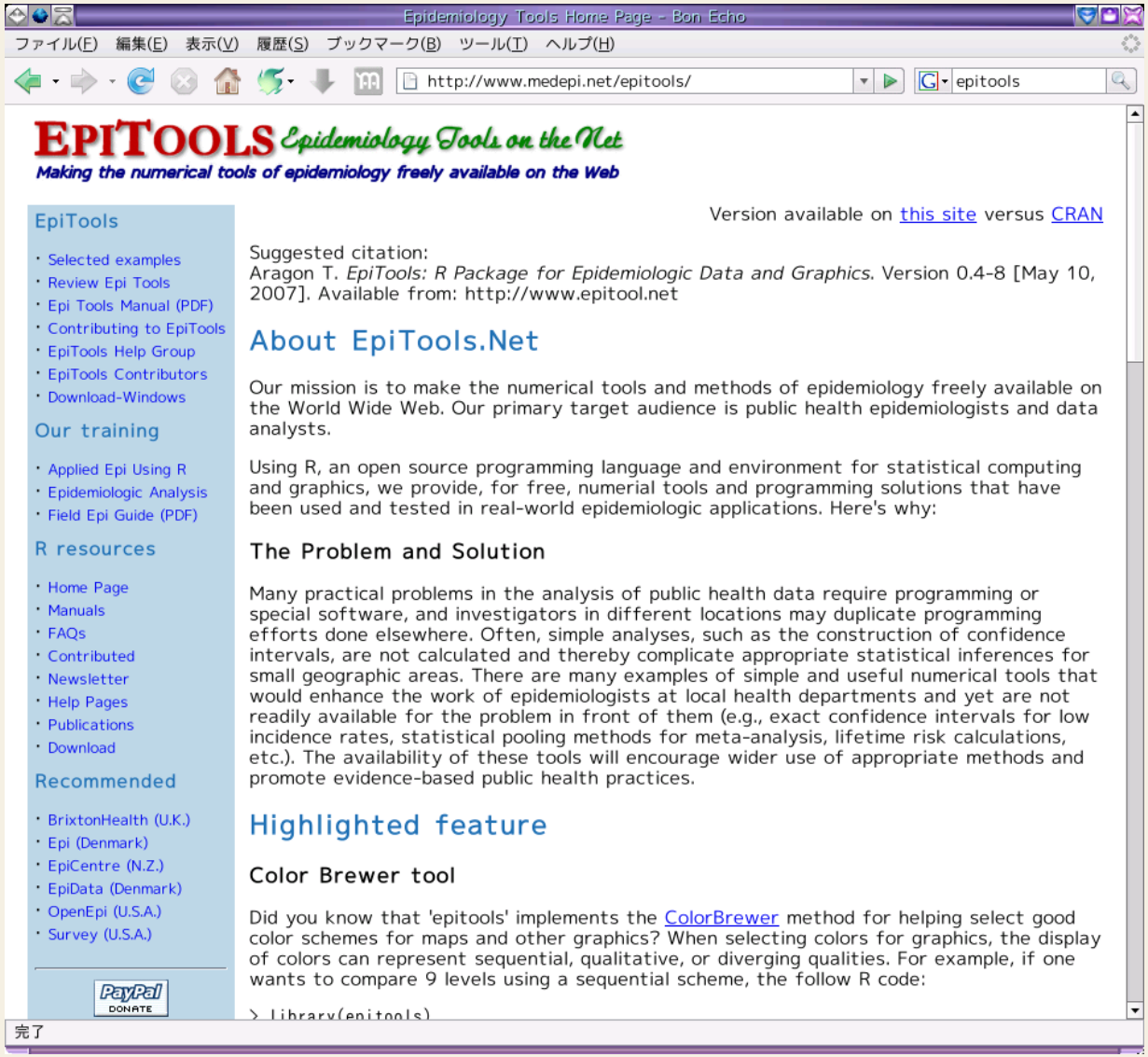
項目	説明
ageadjust.direct	Age standardization by direct method, with exact confidence intervals
ageadjust.indirect	Age standardization by indirect method, with exact confidence intervals
as.hour	Convert date-time object into hour units
as.month	Convert dates into months of the year for plotting epidemic curves
as.week	Convert dates object in 'disease week' for plotting epidemic curves
binom.exact	Confidence intervals for binomial counts or proportions
colorbrewer.display	Display and create ColorBrewer palettes
colors.plot	Plots R's 657 named colors for selection
epicurve.hours	Construct an epidemic curve
epidate	Convert dates into multiple legible formats

項目	説明
epitab	Epidemiologic tabulation for a cohort or case-control study
epitable	Create r x c contigency table (exposure levels vs. binary outcome)
expand.table	Expand contingency table into individual-level data set
expected	Expected values in a table
julian2date	Convert a julian date into standard a date format
kapmeier	Implements product-limit (Kaplan-Meier) method
oddsratio	Odds ratio estimation and confidence intervals
or.midp	Odds ratio estimation and confidence intervals using mid-p method

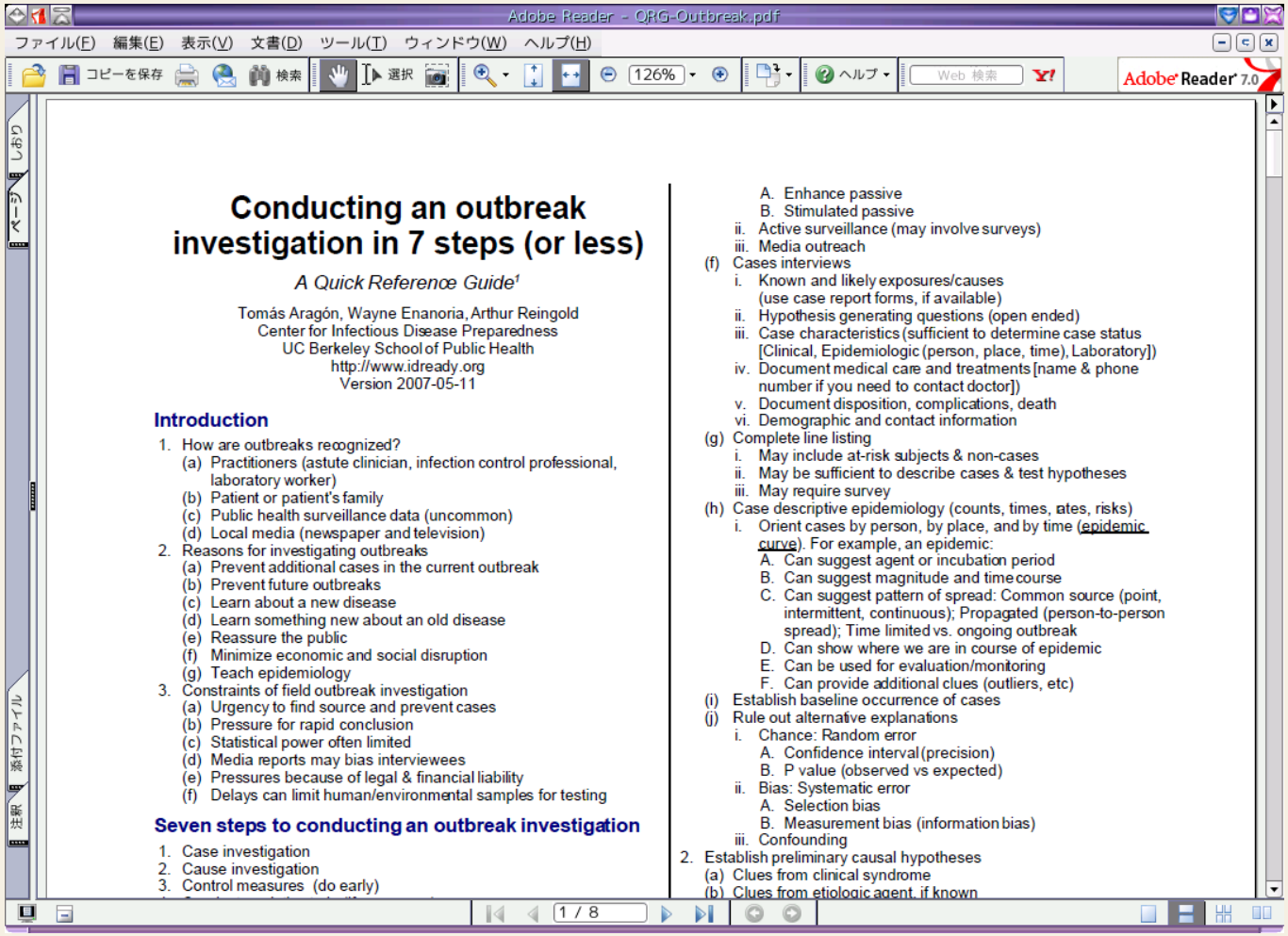
項目	説明
ormidp.test	odds ratio test for independence (p value) for a 2x2 table
oswego	Outbreak of Gastrointestinal Illness in Oswego County, 1940
pois.exact	Confidence intervals for Poisson counts or rates
rate2by2.test	Comparative tests of independence in rx2 rate tables
rateratio	Rate ratio estimation and confidence intervals
ratetable	Create r x 2 count and person-time table for calculating rates
riskratio	Risk ratio estimation and confidence intervals
tab2by2.test	Comparative tests of independence in rx2 contigency tables

項目	説明
table.margins	Marginal totals of a table
wcgs	Western Collaborative Group Study data
wnv	West Nile Virus human cases reported in California, USA, as of December 14, 2004

8 epitools


Epidemiology Tools Home Page - Bon Echo
ファイル(F) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)
http://www.medepi.net/epitools/
epitools
EPITOOLS Epidemiology Tools on the Net
Making the numerical tools of epidemiology freely available on the Web
EpiTools
Selected examples
Review Epi Tools
Epi Tools Manual (PDF)
Contributing to EpiTools
EpiTools Help Group
EpiTools Contributors
Download-Windows
Our training
Applied Epi Using R
Epidemiologic Analysis
Field Epi Guide (PDF)
R resources
Home Page
Manuals
FAQs
Contributed
Newsletter
Help Pages
Publications
Download
Recommended
BrixtonHealth (U.K.)
Epi (Denmark)
EpiCentre (N.Z.)
EpiData (Denmark)
OpenEpi (U.S.A.)
Survey (U.S.A.)
PayPal DONATE
Version available on this site versus CRAN
Suggested citation:
Aragon T. EpiTools: R Package for Epidemiologic Data and Graphics. Version 0.4-8 [May 10, 2007]. Available from: http://www.epitool.net
About EpiTools.Net
Our mission is to make the numerical tools and methods of epidemiology freely available on the World Wide Web. Our primary target audience is public health epidemiologists and data analysts.
Using R, an open source programming language and environment for statistical computing and graphics, we provide, for free, numerial tools and programming solutions that have been used and tested in real-world epidemiologic applications. Here's why:
The Problem and Solution
Many practical problems in the analysis of public health data require programming or special software, and investigators in different locations may duplicate programming efforts done elsewhere. Often, simple analyses, such as the construction of confidence intervals, are not calculated and thereby complicate appropriate statistical inferences for small geographic areas. There are many examples of simple and useful numerical tools that would enhance the work of epidemiologists at local health departments and yet are not readily available for the problem in front of them (e.g., exact confidence intervals for low incidence rates, statistical pooling methods for meta-analysis, lifetime risk calculations, etc.). The availability of these tools will encourage wider use of appropriate methods and promote evidence-based public health practices.
Highlighted feature
Color Brewer tool
Did you know that 'epitools' implements the ColorBrewer method for helping select good color schemes for maps and other graphics? When selecting colors for graphics, the display of colors can represent sequential, qualitative, or diverging qualities. For example, if one wants to compare 9 levels using a sequential scheme, the follow R code:
> library(epitools)
完了

9 Quick reference for outbreaks



Conducting an outbreak investigation in 7 steps (or less)

A Quick Reference Guide[1]

Tomás Aragón, Wayne Enanoria, Arthur Reingold
Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
http://www.idready.org
Version 2007-05-11

Introduction

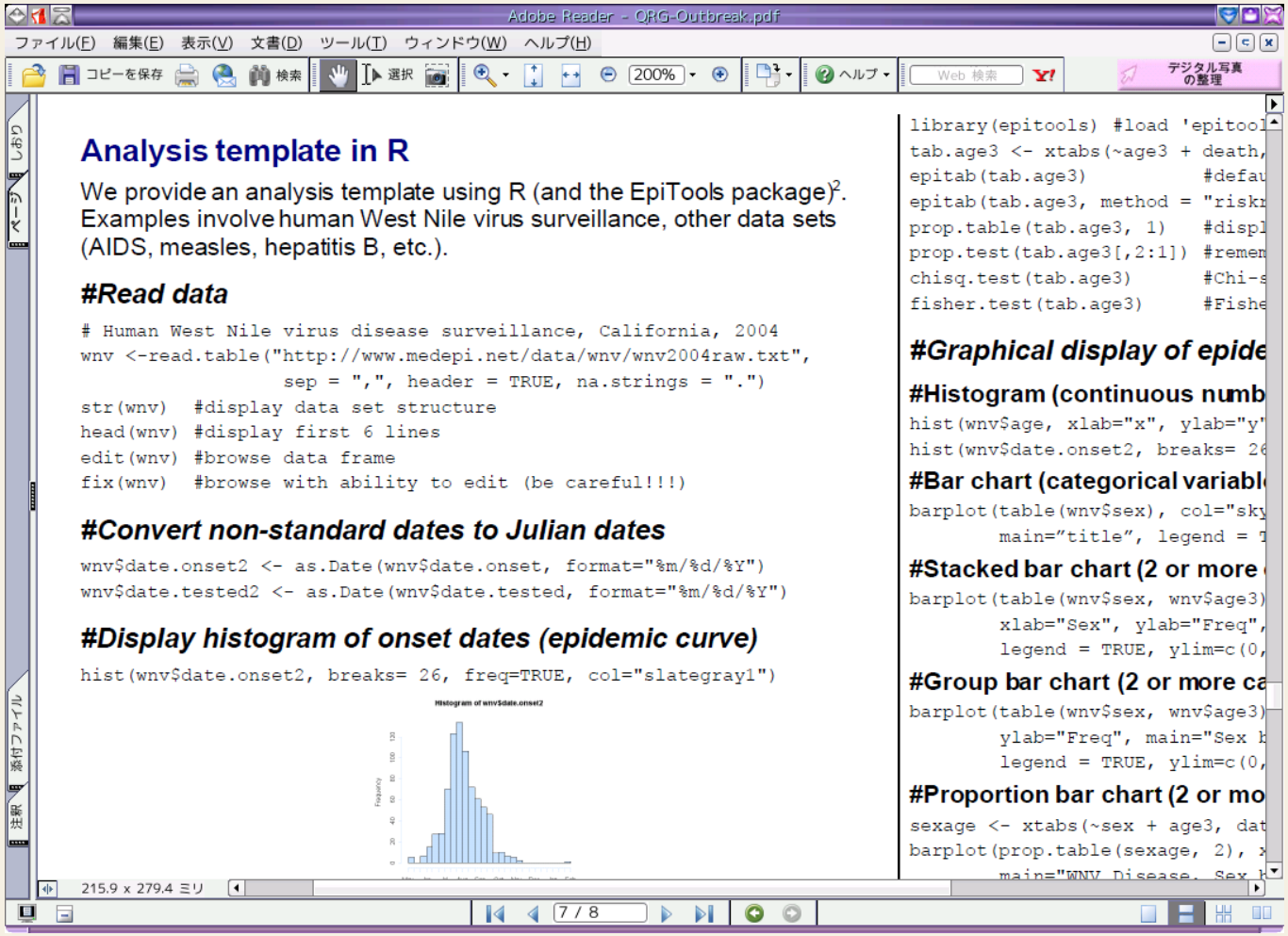
1. How are outbreaks recognized?
   (a) Practitioners (astute clinician, infection control professional, laboratory worker)
   (b) Patient or patient's family
   (c) Public health surveillance data (uncommon)
   (d) Local media (newspaper and television)
2. Reasons for investigating outbreaks
   (a) Prevent additional cases in the current outbreak
   (b) Prevent future outbreaks
   (c) Learn about a new disease
   (d) Learn something new about an old disease
   (e) Reassure the public
   (f) Minimize economic and social disruption
   (g) Teach epidemiology
3. Constraints of field outbreak investigation
   (a) Urgency to find source and prevent cases
   (b) Pressure for rapid conclusion
   (c) Statistical power often limited
   (d) Media reports may bias interviewees
   (e) Pressures because of legal & financial liability
   (f) Delays can limit human/environmental samples for testing

Seven steps to conducting an outbreak investigation

1. Case investigation
2. Cause investigation
3. Control measures (do early)
   A. Enhance passive
   B. Stimulated passive
   ii. Active surveillance (may involve surveys)
   iii. Media outreach
   (f) Cases interviews
   i. Known and likely exposures/causes (use case report forms, if available)
   ii. Hypothesis generating questions (open ended)
   iii. Case characteristics (sufficient to determine case status [Clinical, Epidemiologic (person, place, time), Laboratory])
   iv. Document medical care and treatments [name & phone number if you need to contact doctor])
   v. Document disposition, complications, death
   vi. Demographic and contact information
   (g) Complete line listing
   i. May include at-risk subjects & non-cases
   ii. May be sufficient to describe cases & test hypotheses
   iii. May require survey
   (h) Case descriptive epidemiology (counts, times, rates, risks)
   i. Orient cases by person, by place, and by time (epidemic curve). For example, an epidemic:
   A. Can suggest agent or incubation period
   B. Can suggest magnitude and time course
   C. Can suggest pattern of spread: Common source (point, intermittent, continuous); Propagated (person-to-person spread); Time limited vs. ongoing outbreak
   D. Can show where we are in course of epidemic
   E. Can be used for evaluation/monitoring
   F. Can provide additional clues (outliers, etc)
   (i) Establish baseline occurrence of cases
   (j) Rule out alternative explanations
   i. Chance: Random error
   A. Confidence interval (precision)
   B. P value (observed vs expected)
   ii. Bias: Systematic error
   A. Selection bias
   B. Measurement bias (information bias)
   iii. Confounding
2. Establish preliminary causal hypotheses
   (a) Clues from clinical syndrome
   (b) Clues from etiologic agent, if known

A quick reference for outbreak investigation

9 Quick reference for outbreaks



Analysis template in R

We provide an analysis template using R (and the EpiTools package)[2]. Examples involve human West Nile virus surveillance, other data sets (AIDS, measles, hepatitis B, etc.).

#Read data

```
# Human West Nile virus disease surveillance, California, 2004
wnv <-read.table("http://www.medepi.net/data/wnv/wnv2004raw.txt",
                 sep = ",", header = TRUE, na.strings = ".")
str(wnv)  #display data set structure
head(wnv) #display first 6 lines
edit(wnv) #browse data frame
fix(wnv)  #browse with ability to edit (be careful!!!)
```

#Convert non-standard dates to Julian dates

```
wnv$date.onset2 <- as.Date(wnv$date.onset, format="%m/%d/%Y")
wnv$date.tested2 <- as.Date(wnv$date.tested, format="%m/%d/%Y")
```

#Display histogram of onset dates (epidemic curve)

```
hist(wnv$date.onset2, breaks= 26, freq=TRUE, col="slategray1")
```

```
library(epitools) #load 'epitool
tab.age3 <- xtabs(~age3 + death,
epitab(tab.age3)          #defau
epitab(tab.age3, method = "riskr
prop.table(tab.age3, 1)   #displ
prop.test(tab.age3[,2:1]) #remem
chisq.test(tab.age3)      #Chi-s
fisher.test(tab.age3)     #Fishe
```

#Graphical display of epide

#Histogram (continuous numb

```
hist(wnv$age, xlab="x", ylab="y"
hist(wnv$date.onset2, breaks= 26
```

#Bar chart (categorical variabl

```
barplot(table(wnv$sex), col="sky
        main="title", legend = T
```

#Stacked bar chart (2 or more

```
barplot(table(wnv$sex, wnv$age3)
        xlab="Sex", ylab="Freq",
        legend = TRUE, ylim=c(0,
```

#Group bar chart (2 or more ca

```
barplot(table(wnv$sex, wnv$age3)
        ylab="Freq", main="Sex b
        legend = TRUE, ylim=c(0,
```

#Proportion bar chart (2 or mo

```
sexage <- xtabs(~sex + age3, dat
barplot(prop.table(sexage, 2), x
```

R commands for outbreak investigation

10　疫学教育のおけるRの活用

- UC Berkeley School of Publich Health
- Master of Public Health Management, APU

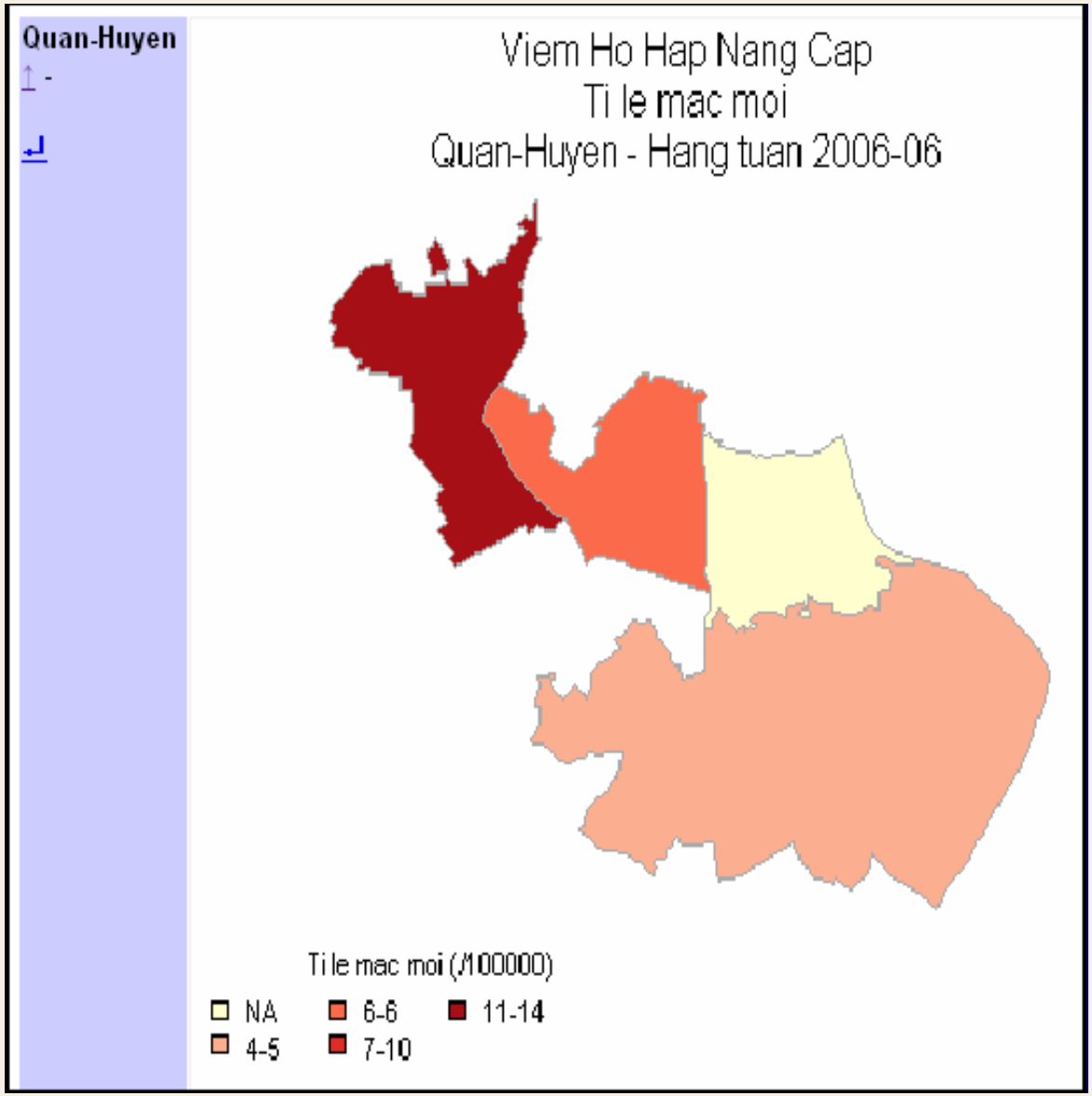
11　疫学研究におけるRの活用

- Rを使用した疫学論文をたまに見るが、使用した統計ソフトで学術論文を検索するすべがなく、定量的に評価できない
- 疫学研究者の中にも少なくない数のユーザ

12　疾病対策におけるRの活用

- 現場レベルでは、EpiInfoやExcelが中心
- バックエンドとしてRの利用：WHOのサーベイランスシステム


Close application
Administration
Places management
Data entry
Reports
Search
Cross Query Extractor
Query extractor
Import / Export
Events
MOH
Viet Nam
Current language
English
V 1.36
Backup database
World Health Organization


Quan-Huyen
Viem Ho Hap Nang Cap
Ti le mac moi
Quan-Huyen - Hang tuan 2006-06
Ti le mac moi (/100000)
NA
4-5
6-6
7-10
11-14

13　アウトブレイクにおける活用

- アウトブレイクが発生すると疫学者の出番
- データ：1990年8月25日にタイ王国スパンブリ県で発生した急性胃腸炎

14　データの収集

- データ入力にはEpiInfoが使われ、イベント参加者が摂取した食品ごとに、0=Yes 1=No 9=missing/unknown と入力された
- 性年齢、食べた時間、発症の有無、症状なども収集された

```
> des()
No. of observations = 1094
   Variable       Class            Description
1  id             numeric
2  sex            numeric
3  age            numeric
4  exptime        AsIs
```

```
5   beefcurry     numeric
6   saltegg       numeric
7   eclair        numeric
8   water         numeric
9   onset         AsIs
10  nausea        numeric
11  vomiting      numeric
12  abdpain       numeric
13  diarrhea      numeric
```

15　データの内容

```
> summ()
No. of observations = 1094
   Var. name obs. mean    median  s.d.    min.    max.
1  id        1094 547.5   547.5   315.95  1       1094
2  sex       1094 0.66    1       0.47    0       1
3  age       1094 23.69   18      19.67   1       99
4  exptime
5  beefcurry 1094 0.95    1       0.61    0       9
6  saltegg   1094 0.96    1       0.61    0       9
7  eclair    1094 11.48   2       27.75   0       90
8  water     1094 1.02    1       0.61    0       9
9  onset
10 nausea    1094 0.4     0       0.49    0       1
11 vomiting  1094 0.38    0       0.49    0       1
12 abdpain   1094 0.35    0       0.48    0       1
13 diarrhea  1094 0.21    0       0.41    0       1
```

16　症例の定義

- 症例を悪心、嘔吐、腹痛、下痢のいずれかの症状を有する者とする

```
case <- (nausea==1)|(vomiting==1)|
        (abdpain==1)|(diarrhea==1)
label.var(case,"diseased")
```

- caseが14番目の変数として加えられた

17　曝露の時刻

```
> exptime[1:3]
[1] "25330825180000" "25330825180000" "
     25330825180000"
```

- 14桁で表現されており、最初の4桁は仏歴（タイの暦で543を引くと西暦になる）、5–6桁は月、7–8桁は日、9-10桁は時間、11–12桁は分、13–14桁は秒になっている

```
> day.exptime <- substr(exptime,7,8)
> tab1(day.exptime)
day.exptime :
        Frequency %(NA+) %(NA-)
25 1055 96.4 100
<NA> 39 3.6 0
  Total 1094 100.0 100
```

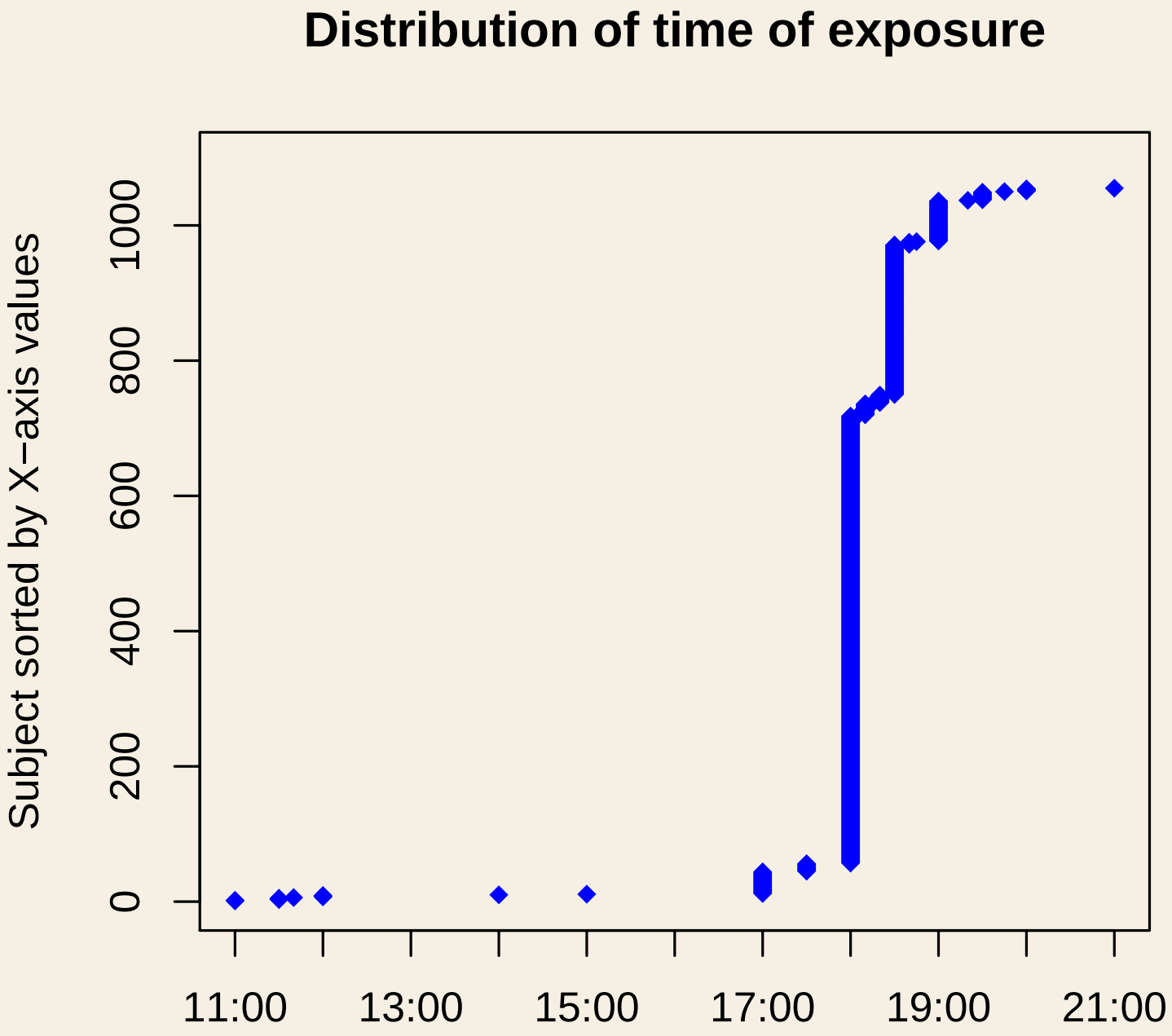
- 全てが25日で、missingが39件あることが分かる
- 時間、分でも同様に調べる

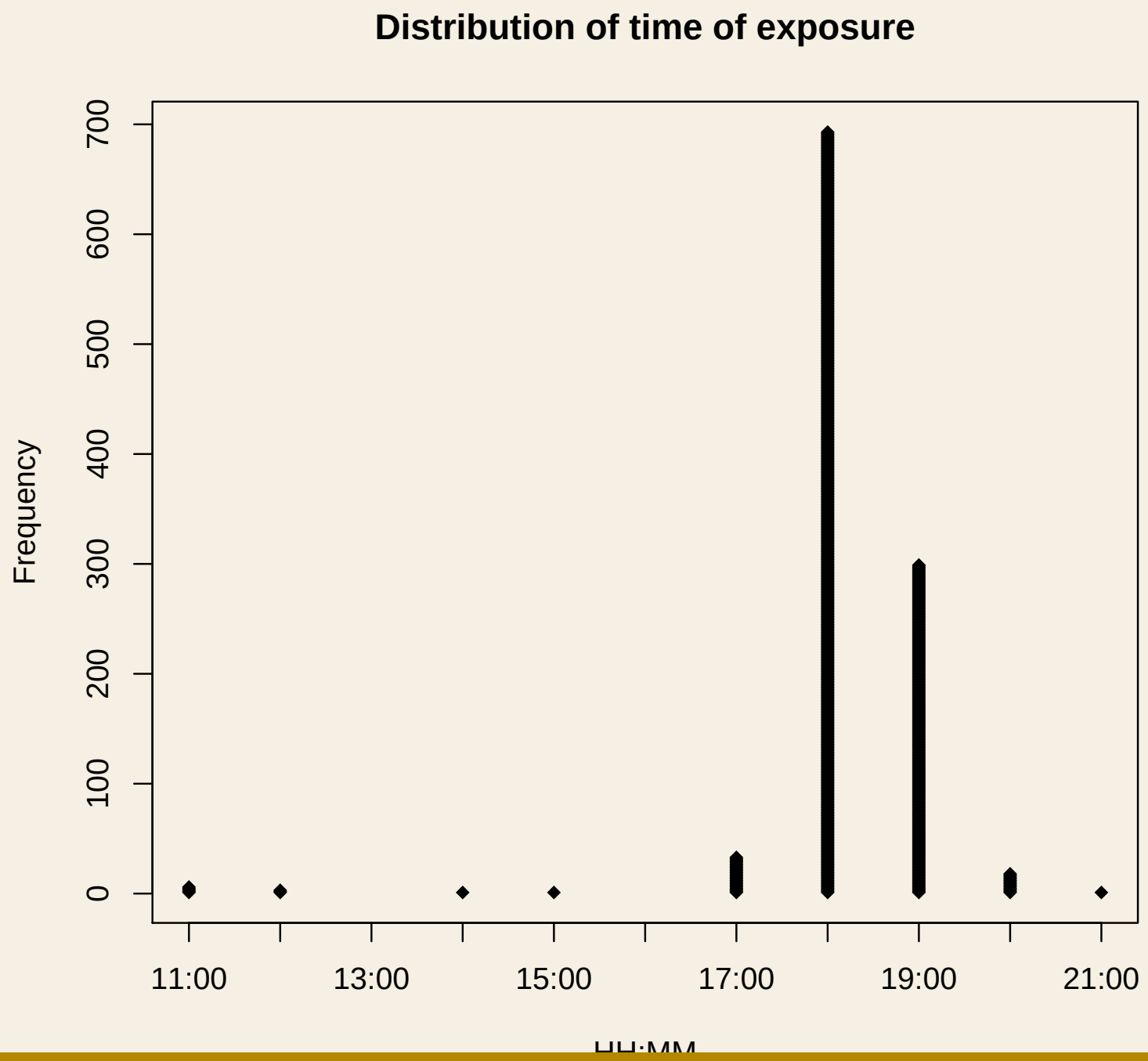
```
> hr.exptime <- substr(exptime,9,10)
> tab1(hr.exptime)
hr.exptime :

        Frequency   %(NA+)   %(NA-)
11              6      0.5      0.6
12              3      0.3      0.3
14              1      0.1      0.1
15              1      0.1      0.1
17             45      4.1      4.3
18            920     84.1     87.2
19             74      6.8      7.0
20              4      0.4      0.4
21              1      0.1      0.1
<NA>           39      3.6      0.0
  Total      1094    100.0    100.0
```

```
> min.exptime <- substr(exptime,11,12)
> tab1(min.exptime)
min.exptime :
        Frequency   %(NA+)   %(NA-)
00            767     70.1     72.7
05              1      0.1      0.1
10             17      1.6      1.6
15              1      0.1      0.1
20             13      1.2      1.2
30            249     22.8     23.6
40              5      0.5      0.5
45              2      0.2      0.2
<NA>           39      3.6      0.0
  Total      1094    100.0    100.0
```

```
> time.expose <- ISOdatetime(year=1990,month=8,day=day.exptime,
+ hour=hr.exptime,min=min.exptime,sec=0)
> label.var(time.expose,"time of exposure")
> summ(time.expose)
             Min.            Median              Mean
                                Max.
1990-08-25 11:00 1990-08-25 18:00 1990-08-25 18:06
       1990-08-25 21:00
```

- time.exposeをデータフレームに入れた


Distribution of time of exposure
Subject sorted by X−axis values
0
200
400
600
800
1000
11:00
13:00
15:00
17:00
19:00
21:00


Distribution of time of exposure
Frequency
0
100
200
300
400
500
600
700
11:00
13:00
15:00
17:00
19:00
21:00
HH:MM

18　日付操作

```
> (lctime.old <- Sys.getlocale("LC_TIME"))
[1] "ja_JP.eucJP"
> Sys.setlocale("LC_TIME","C")
[1] "C"
> dates <- c("11-02-1959","1959Nov02","November␣2,␣1959")
> jdates<-as.Date(dates, format=c("%m-%d-%Y","%Y%b%d","%B␣%d
      ,␣%Y"))
> jdates; julian(jdates)
[1] "1959-11-02" "1959-11-02" "1959-11-02"
[1] -3713 -3713 -3713
attr(,"origin")
[1] "1970-01-01"
```

18　日付操作

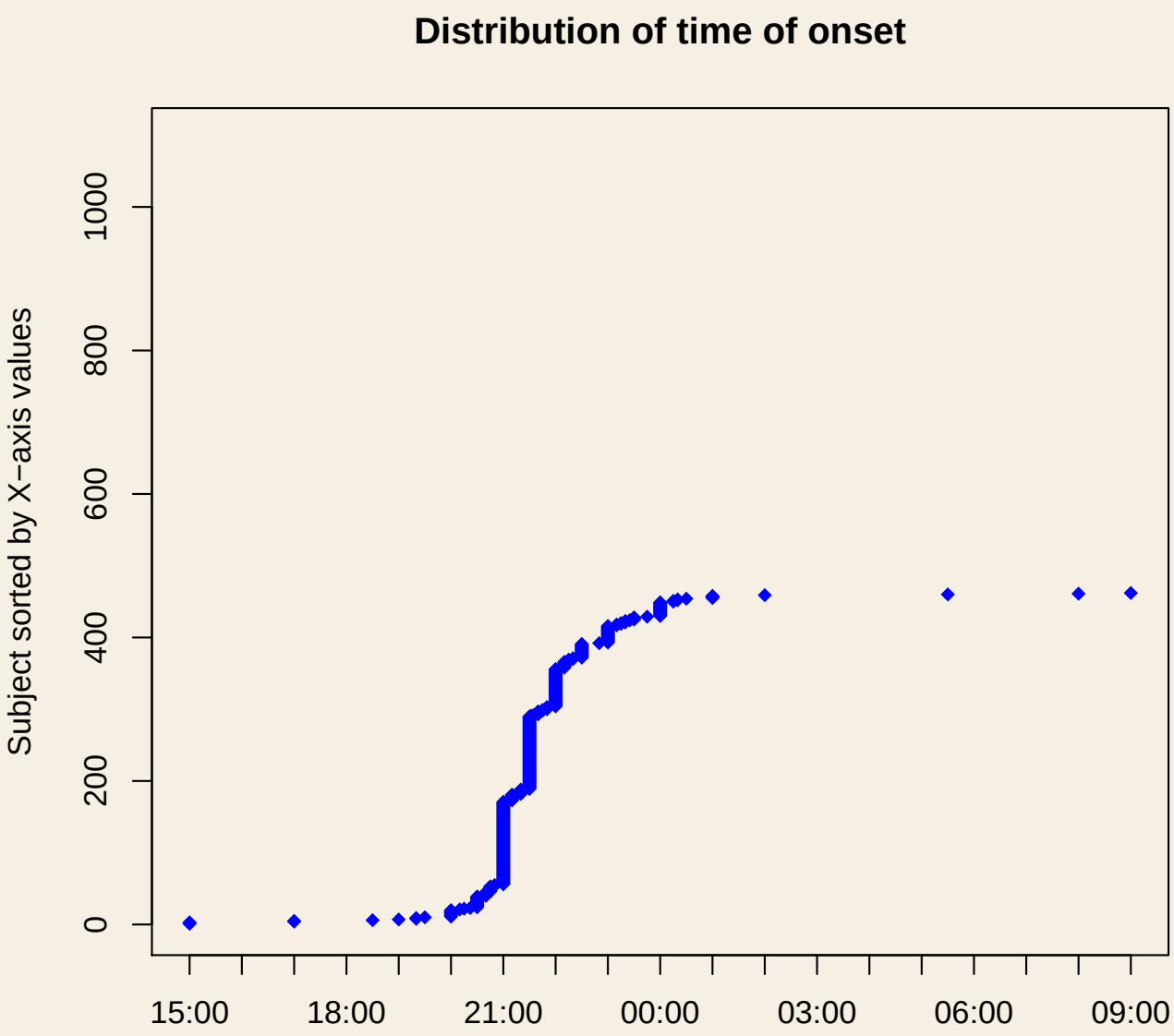
```
> dtim <- c("4/19/1940␣12:30␣AM", "4/18/1940␣9:45␣PM")
> std.dt <- strptime(dtim, format="%m/%d/%Y␣%I:%M␣%p")
> std.dt
[1] "1940-04-19␣00:30:00" "1940-04-18␣21:45:00"
```

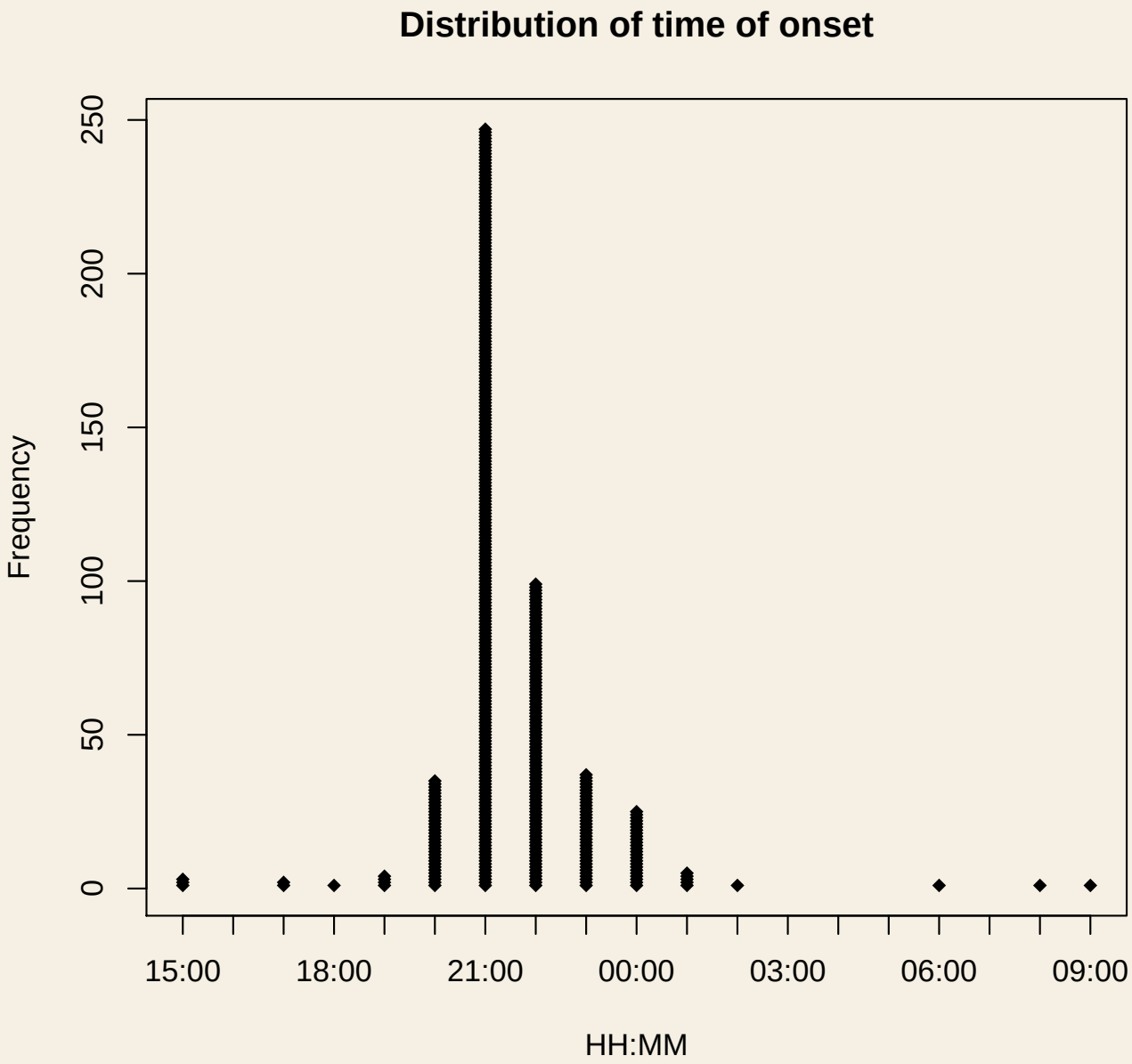
19　発症時刻

```
> onset[!case] <- NA
> day.onset <- substr(onset,7,8)
> tab1(day.onset)
day.onset :
        Frequency   %(NA+)   %(NA-)
25            429     39.2     92.9
26             33      3.0      7.1
<NA>          632     57.8      0.0
  Total      1094    100.0    100.0
```

```
hr.onset <- substr(onset,9,10)
tab1(hr.onset)
min.onset <- substr(onset,11,12)
tab1(min.onset)
```

```
> time.onset <- ISOdatetime(year=1990,month=8,day=day.onset,
+ hour=hr.onset,min=min.onset,sec=0,tz="")
> label.var(time.onset,"time of onset")
> summ(time.onset)
            Min.                Median                Mean
                                   Max.
1990-08-25 15:00 1990-08-25 21:30 1990-08-25 21:40
      1990-08-26 09:00
```

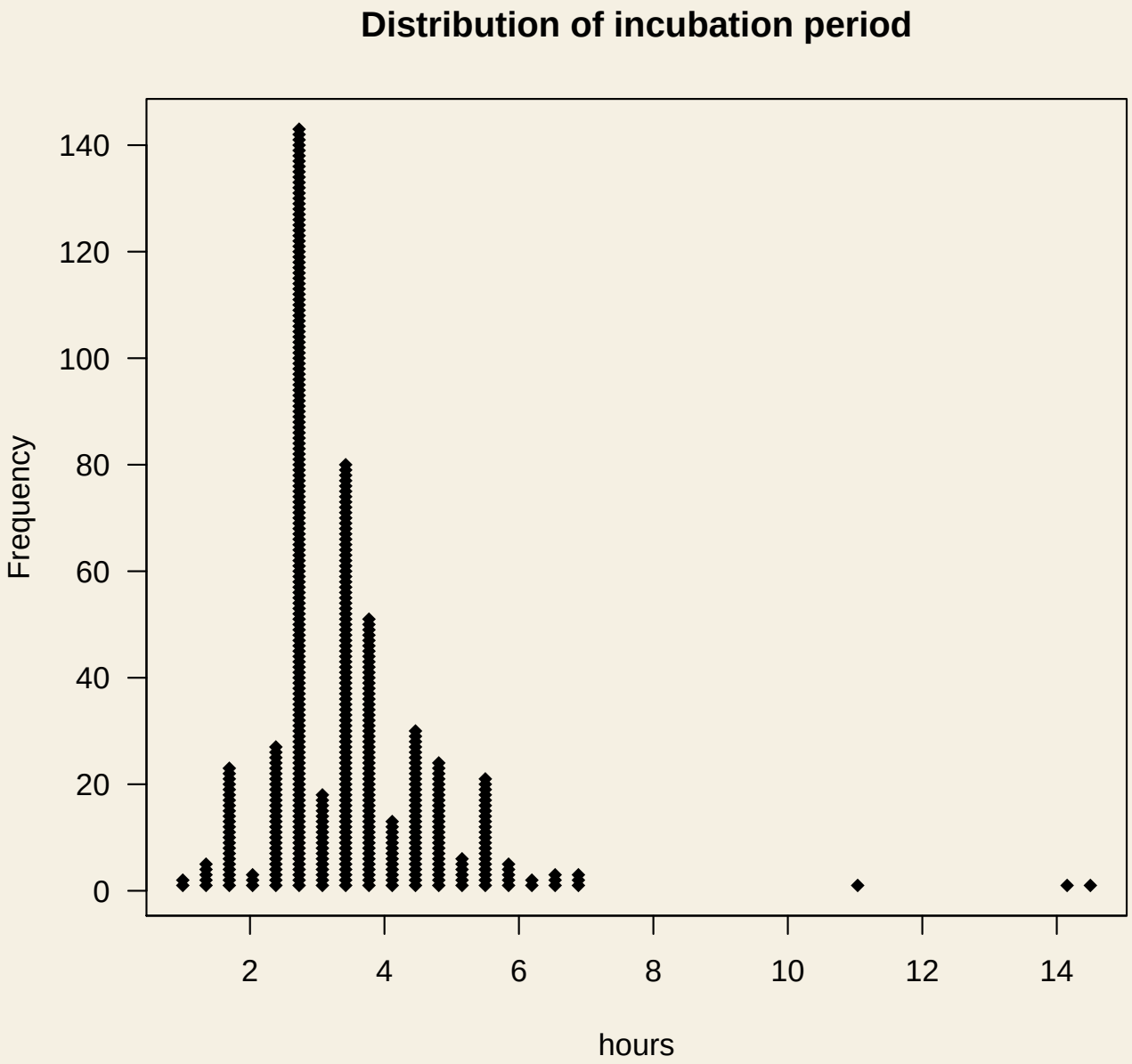

Distribution of time of onset
Subject sorted by X−axis values
0
200
400
600
800
1000
15:00
18:00
21:00
00:00
03:00
06:00
09:00


Distribution of time of onset
Frequency
0
50
100
150
200
250
15:00
18:00
21:00
00:00
03:00
06:00
09:00
HH:MM

20　潜伏期間

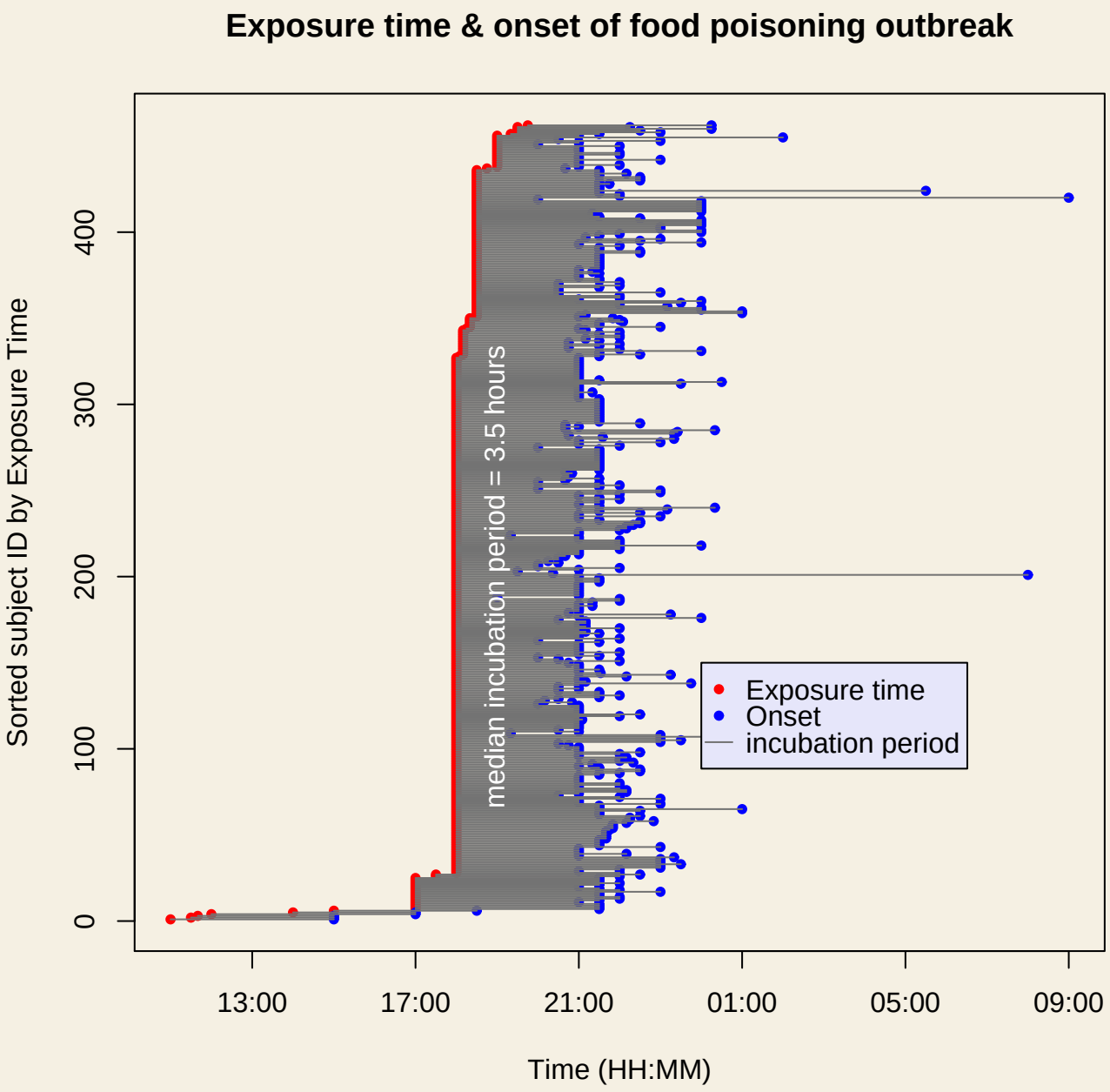
```
> incubation.period <- time.onset - time.expose
> label.var(incubation.period,"incubation␣period")
> summ(incubation.period)
  obs.  mean    median  s.d.    min.    max.
  462   3.631   3.5     1.28    1       14.5
> dotplot(incubation.period,las=1)
```

- 潜伏期間の中央値は3.5時間


Distribution of incubation period
Frequency
0
20
40
60
80
100
120
140
2
4
6
8
10
12
14
hours

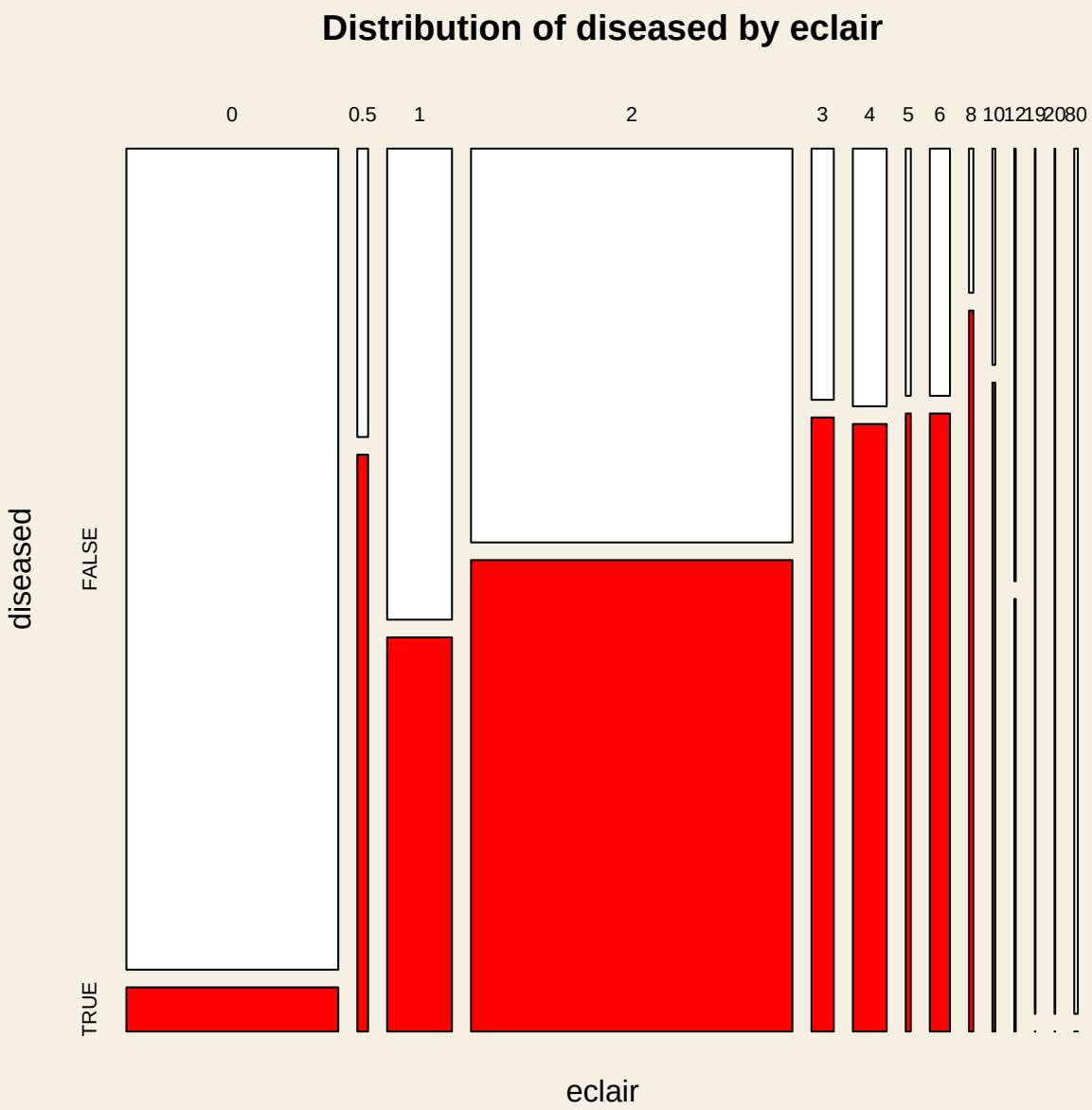
21　Paired plot

- 曝露時刻と発症時刻を同時に視覚化する
- 曝露時刻でソートしてから表示した方が、より多くの情報を得られる

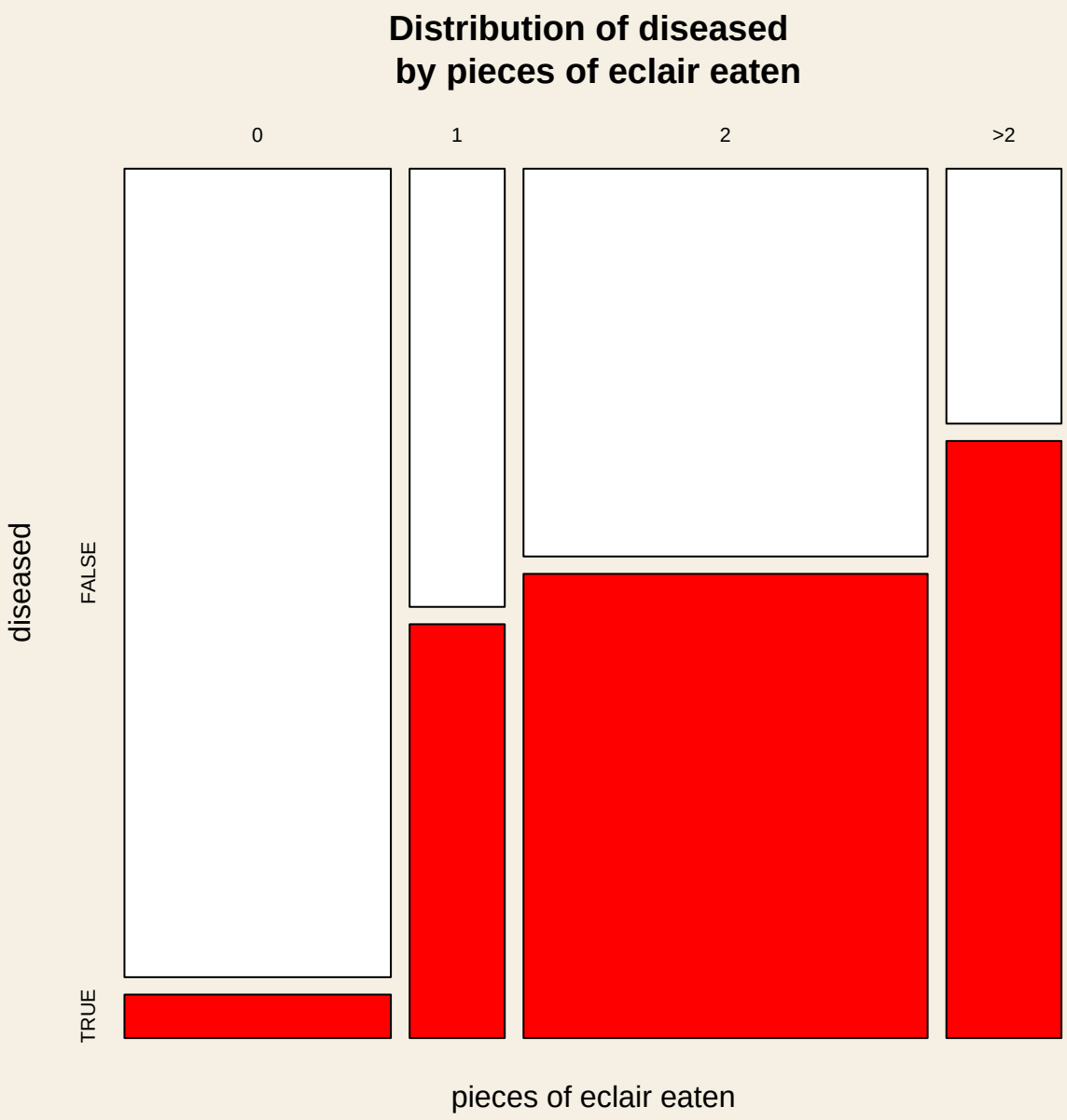

Exposure time & onset of food poisoning outbreak
Sorted subject ID by Exposure Time
0
100
200
300
400
median incubation period = 3.5 hours
Exposure time
Onset
incubation period
13:00
17:00
21:00
01:00
05:00
09:00
Time (HH:MM)

22　原因食品を見つける

- 食品摂取の有無と発症の有無で四分割表を作成して、食べた人の発症リスクと食べなかった人の発症リスクを計算し、その比の相対リスクを調べる
- 食品摂取の程度を調べてある場合は、四分割表よりも複雑になる


Distribution of diseased by eclair
0
0.5
1
2
3
4
5
6
8
10
12
19
20
80
diseased
FALSE
TRUE
eclair

```
> eclairgr <- cut(eclair,breaks=c(0,0.4,1,2,79),
+ include.lowest=TRUE,labels=c("0","1","2",">2"))
> label.var(eclairgr,"pieces␣of␣eclair␣eaten")
> tabpct(eclairgr,case)
[line omitted]
Row percent
                          diseased
pieces of eclair eaten    FALSE     TRUE   Total
                       0    279       15     294
                          (94.9)    (5.1)  (100)
                       1     54       51     105
                          (51.4)   (48.6)  (100)
                       2    203      243     446
                          (45.5)   (54.5)  (100)
                       >2    38       89     127
                          (29.9)   (70.1)  (100)
[line omitted]
```


Distribution of diseased
by pieces of eclair eaten
0
1
2
>2
diseased
FALSE
TRUE
pieces of eclair eaten

23　リスクの計算

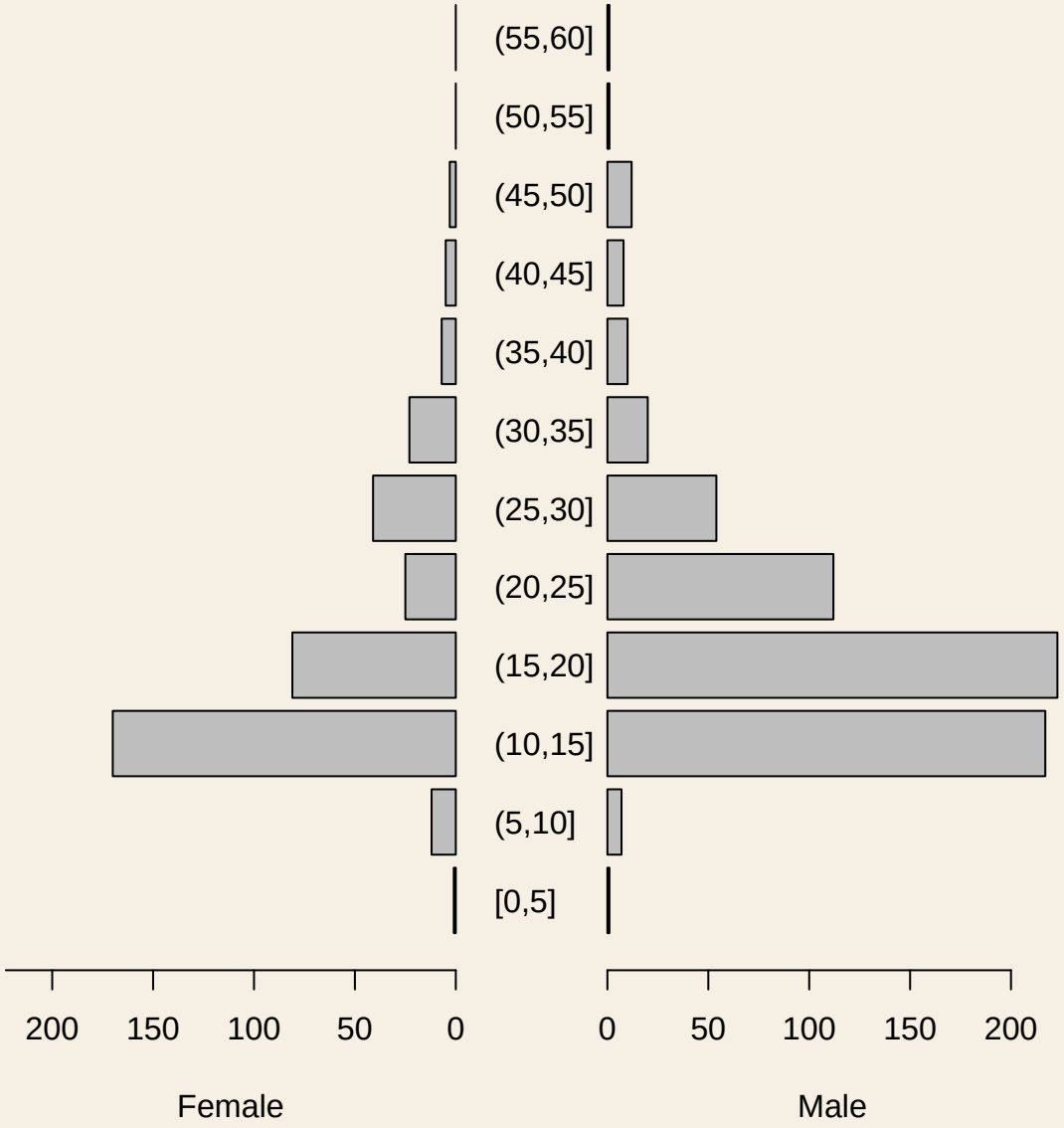
```
> eclair.eat <- eclair > 0
> label.var(eclair.eat,"eating␣eclair")
> cs(case,eclair.eat)
       eating eclair
case     FALSE TRUE Total
  FALSE 279    300   579
  TRUE  15     383   398
  Total 294    683   977

         Rne    Re    Rt
  Risk   0.05   0.56  0.41
                                                Estimate Lower95ci
                                                       Upper95ci
 Risk difference (attributable risk)            0.51      0.44
        0.58
 Risk ratio                                     10.99     8
        15.1
 Attr. frac. exp. -- (Re-Rne)/Re                0.91
```

```
Attr. frac. pop. -- (Rt-Rne)/Rt*100 %   87.48
```

24　性年齢の検討

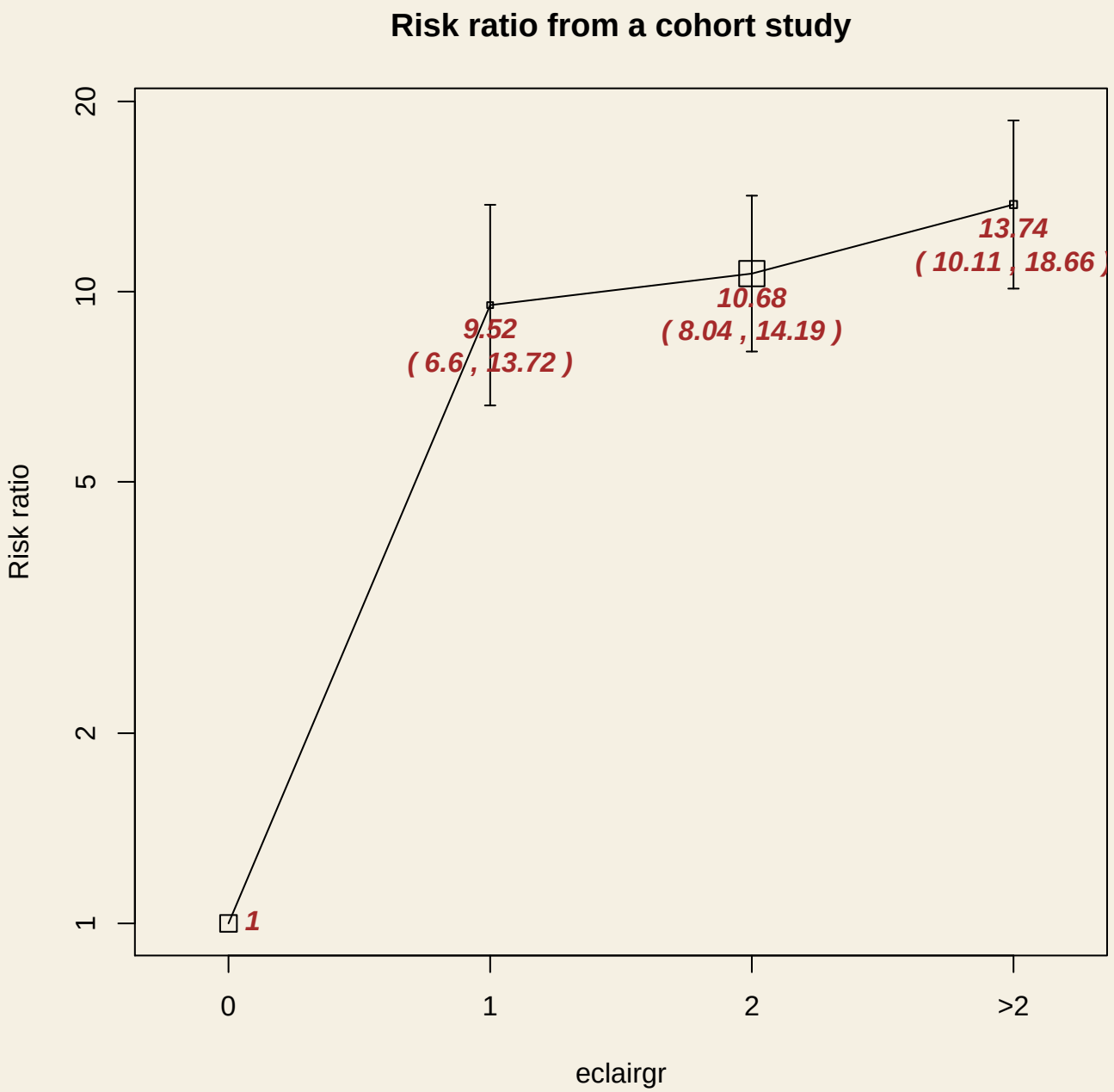

Distribution of age
Frequency
0
50
100
150
200
0
10
20
30
40
50
60


Distribution of age by Sex
Male
Female
0
10
20
30
40
50
60


(55,60]
(50,55]
(45,50]
(40,45]
(35,40]
(30,35]
(25,30]
(20,25]
(15,20]
(10,15]
(5,10]
[0,5]
200
150
100
50
0
0
50
100
150
200
Female
Male

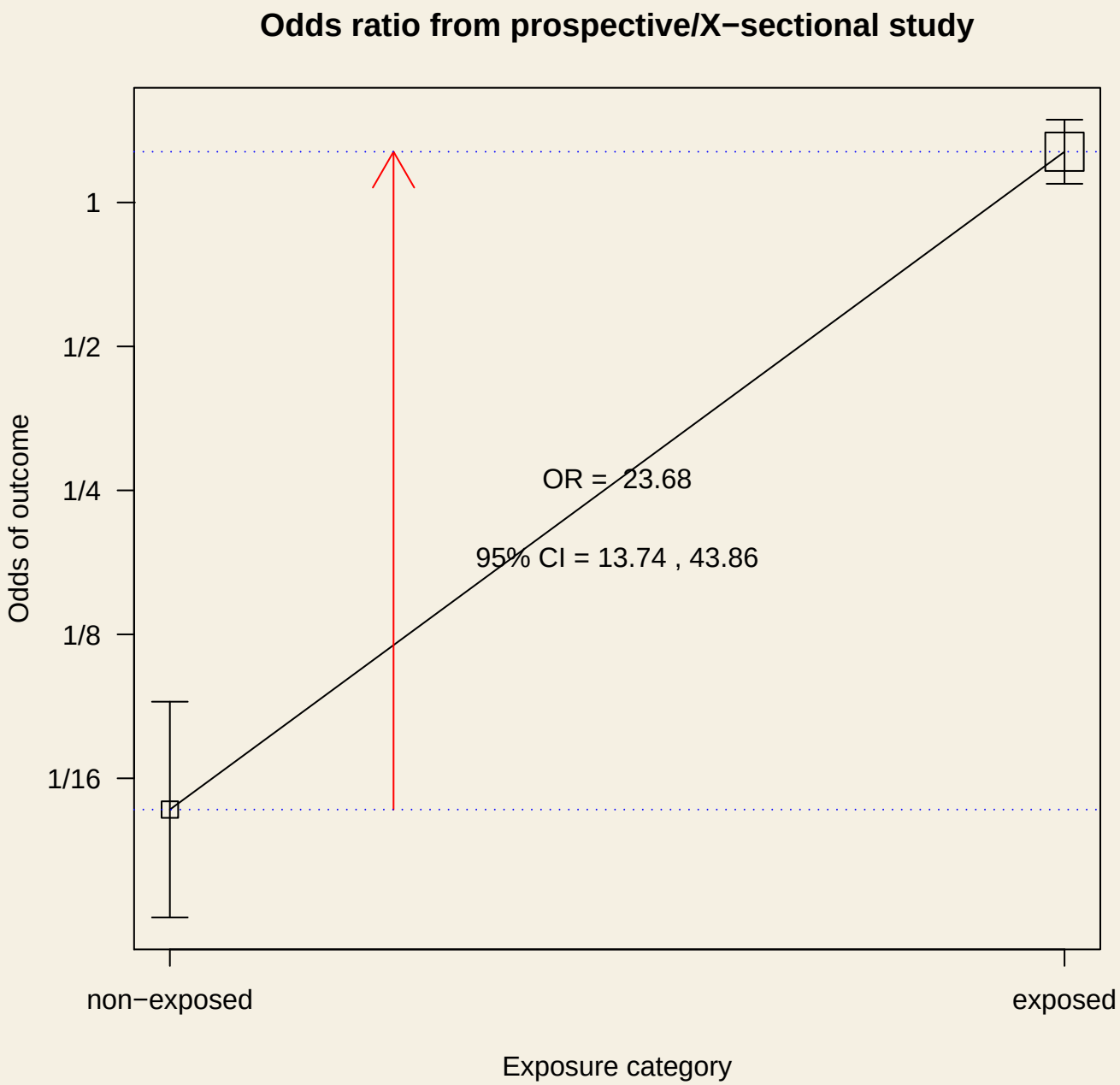
25　量・反応の関係

```
> cs(case,eclairgr)
                 eclairgr
case              0     1      2      >2
  FALSE           279   54     203    38
  TRUE            15    51     243    89

  Absolute risk   0.05 0.49   0.54   0.7
  Risk ratio      1     9.52   10.68 13.74
  lower 95% CI          6.6    8.04   10.11
  upper 95% CI          13.72 14.19 18.66

Chi-squared = 237.121 , 3 d.f., P value = 0
Fisher's exact test (2-sided) P value = 0
```


Risk ratio from a cohort study
Risk ratio
20
10
5
2
1
1
9.52
(6.6 , 13.72)
10.68
(8.04 , 14.19)
13.74
(10.11 , 18.66)
0
1
2
>2
eclairgr

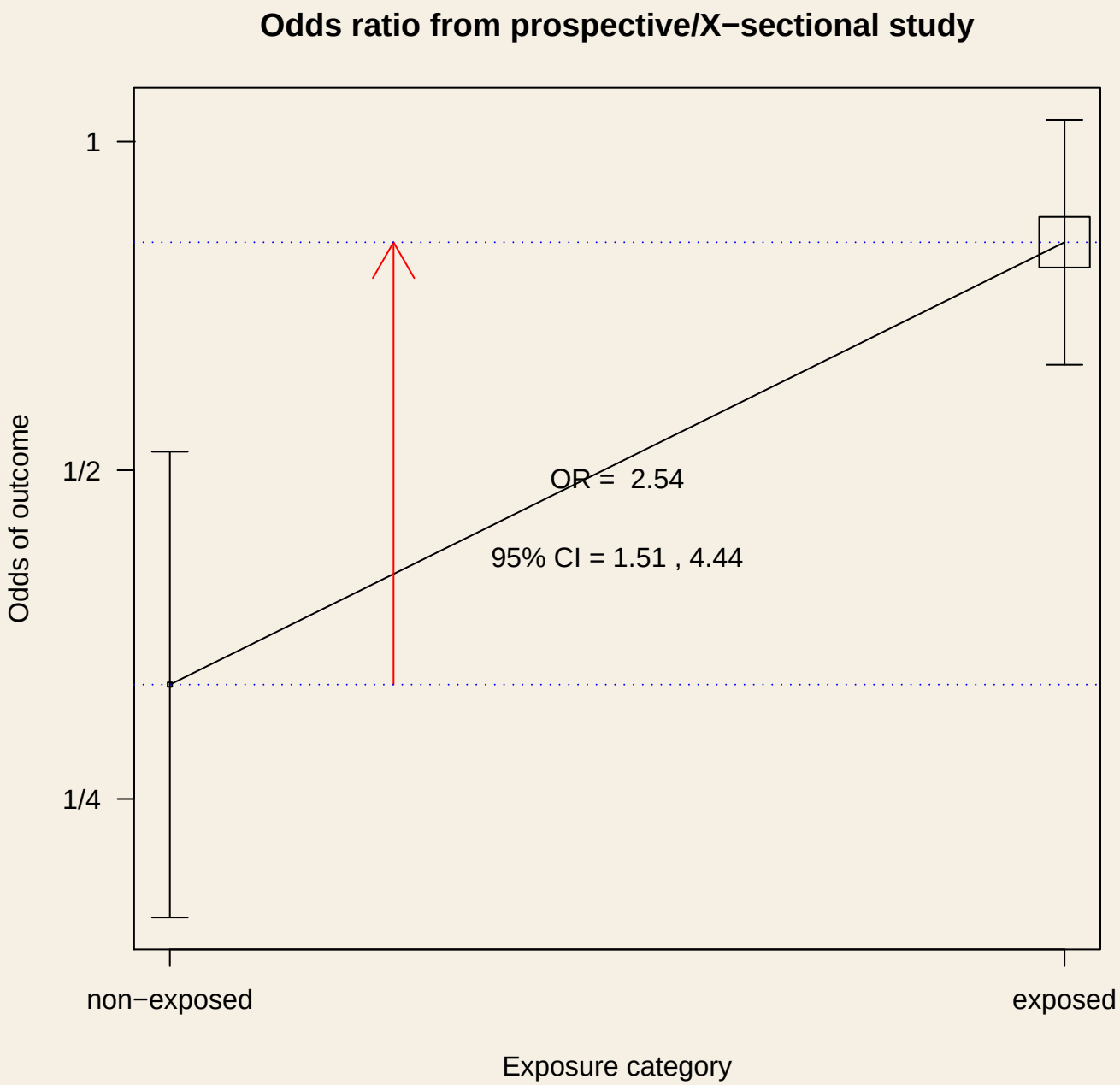
26　オッズ比の計算

```
> cc(case,eclair.eat)

        eating eclair
case     FALSE TRUE Total
  FALSE    279  300   579
  TRUE      15  383   398
  Total    294  683   977

OR =  23.68
95% CI = 13.74 43.86
Chi-squared = 221.21 ,  1 d.f. , P value = 0
Fisher's exact test (2-sided) P value = 0
```


Odds ratio from prospective/X−sectional study
Odds of outcome
1
1/2
1/4
1/8
1/16
OR = 23.68
95% CI = 13.74 , 43.86
non−exposed
exposed
Exposure category

27　交絡因子

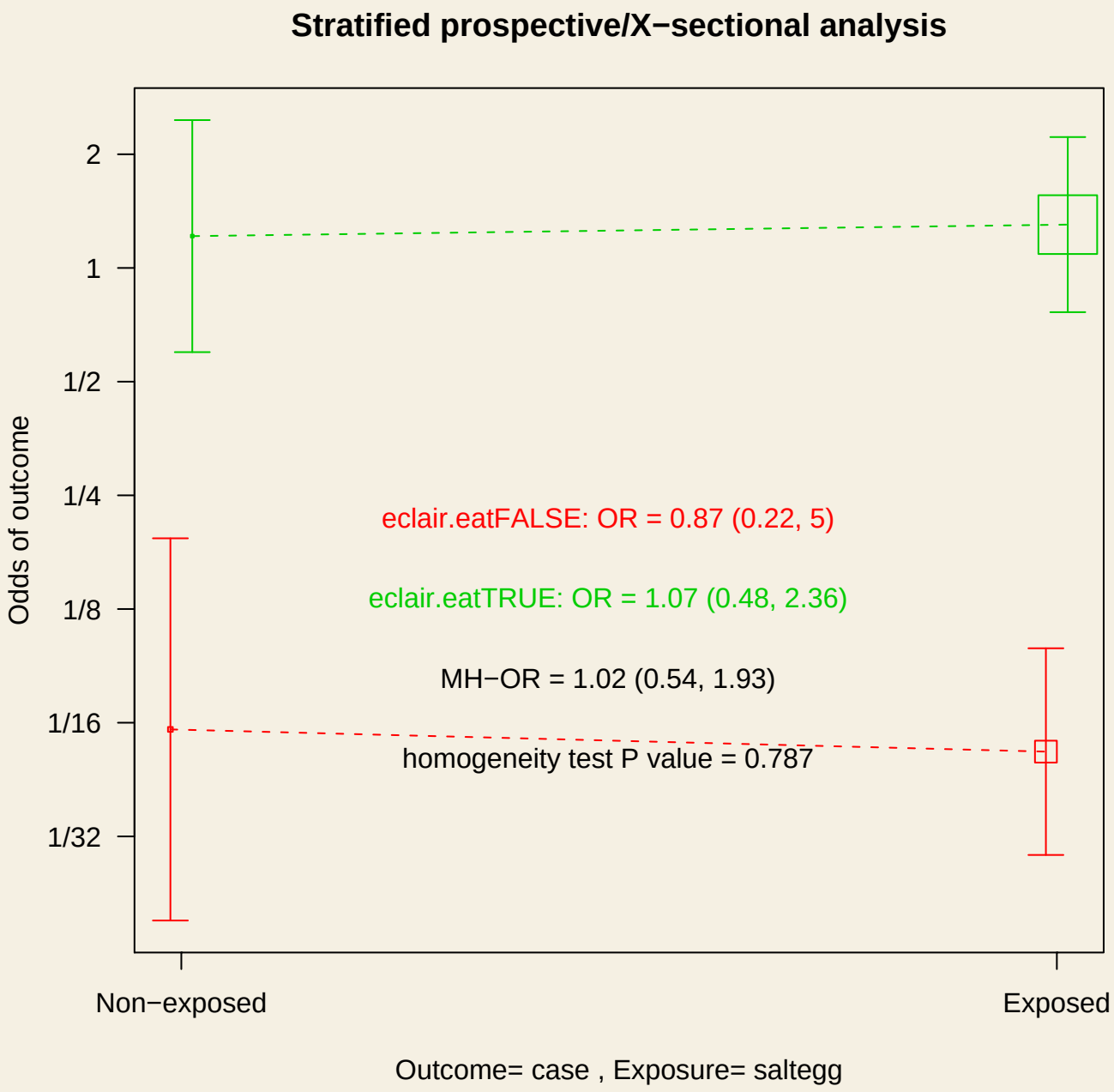
```
> cc(case,saltegg)

        saltegg
case        0    1 Total
  FALSE    66  554   620
  TRUE     21  448   469
  Total    87 1002  1089

OR =  2.54
95% CI = 1.51 4.44
Chi-squared = 13.82 ,  1 d.f. , P value = 0
Fisher's exact test (2-sided) P value = 0
```


Odds ratio from prospective/X−sectional study
1
1/2
1/4
Odds of outcome
OR = 2.54
95% CI = 1.51 , 4.44
non−exposed
exposed
Exposure category

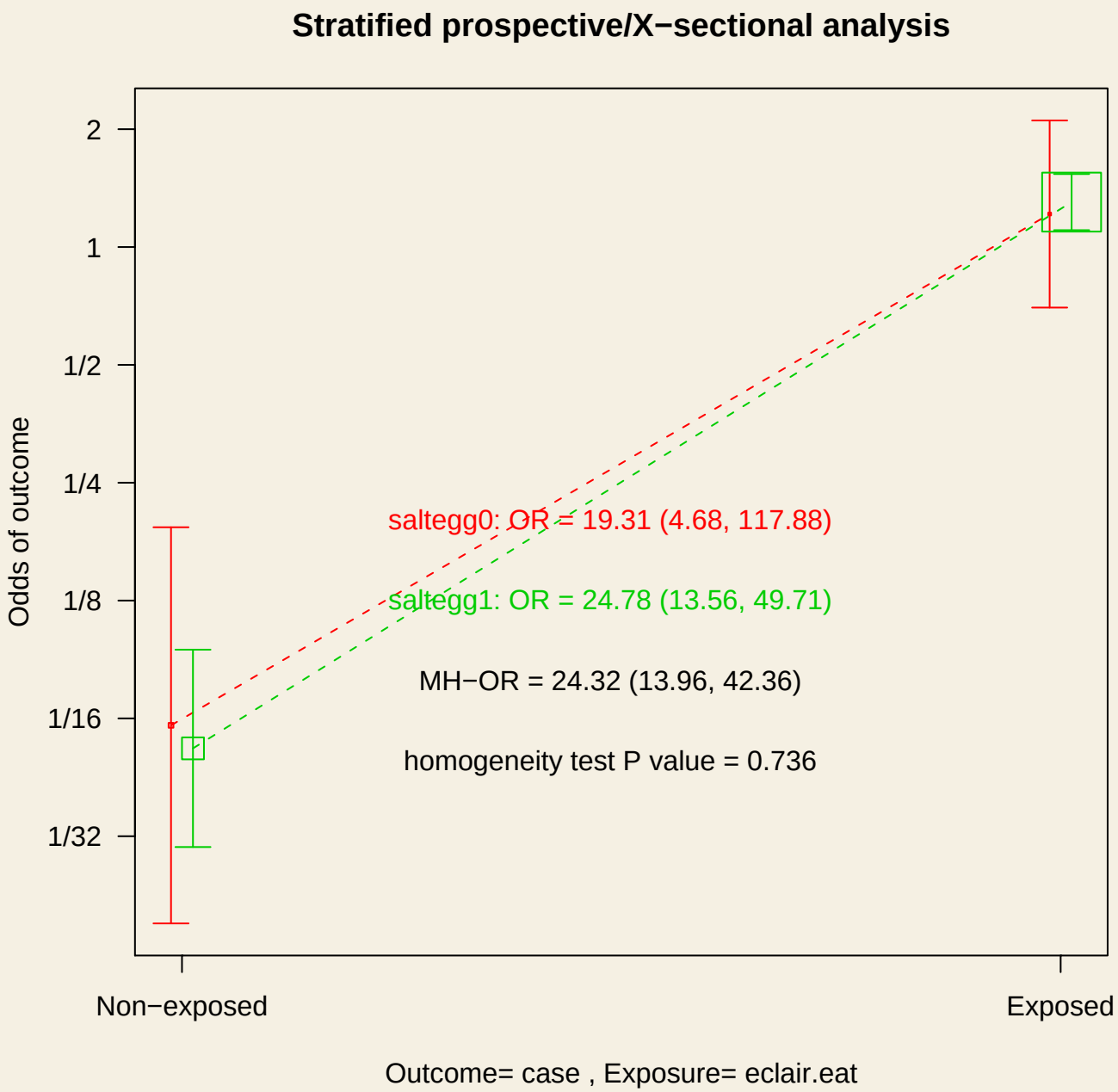
```
> cc(saltegg,eclair.eat,graph=FALSE)
        eating eclair
saltegg FALSE TRUE Total
  0        53   31    84
  1       241  647   888
  Total   294  678   972

OR =  4.58
95% CI = 2.81 7.58
Chi-squared = 47.02 ,  1 d.f. , P value = 0
Fisher's exact test (2-sided) P value = 0
```

28　層化

```
> mhor(case,saltegg,eclair.eat)

Stratified analysis by  eclair.eat
                      OR lower lim. upper lim. P value
eclair.eat FALSE  0.874      0.224       5.00   0.739
eclair.eat TRUE   1.073      0.481       2.36   0.855
M-H combined      1.023      0.541       1.93   0.944

M-H Chi2(1) = 0 , P value = 0.944
Homogeneity test, chi-squared 1 d.f. = 0.07 , P value =
      0.787
```

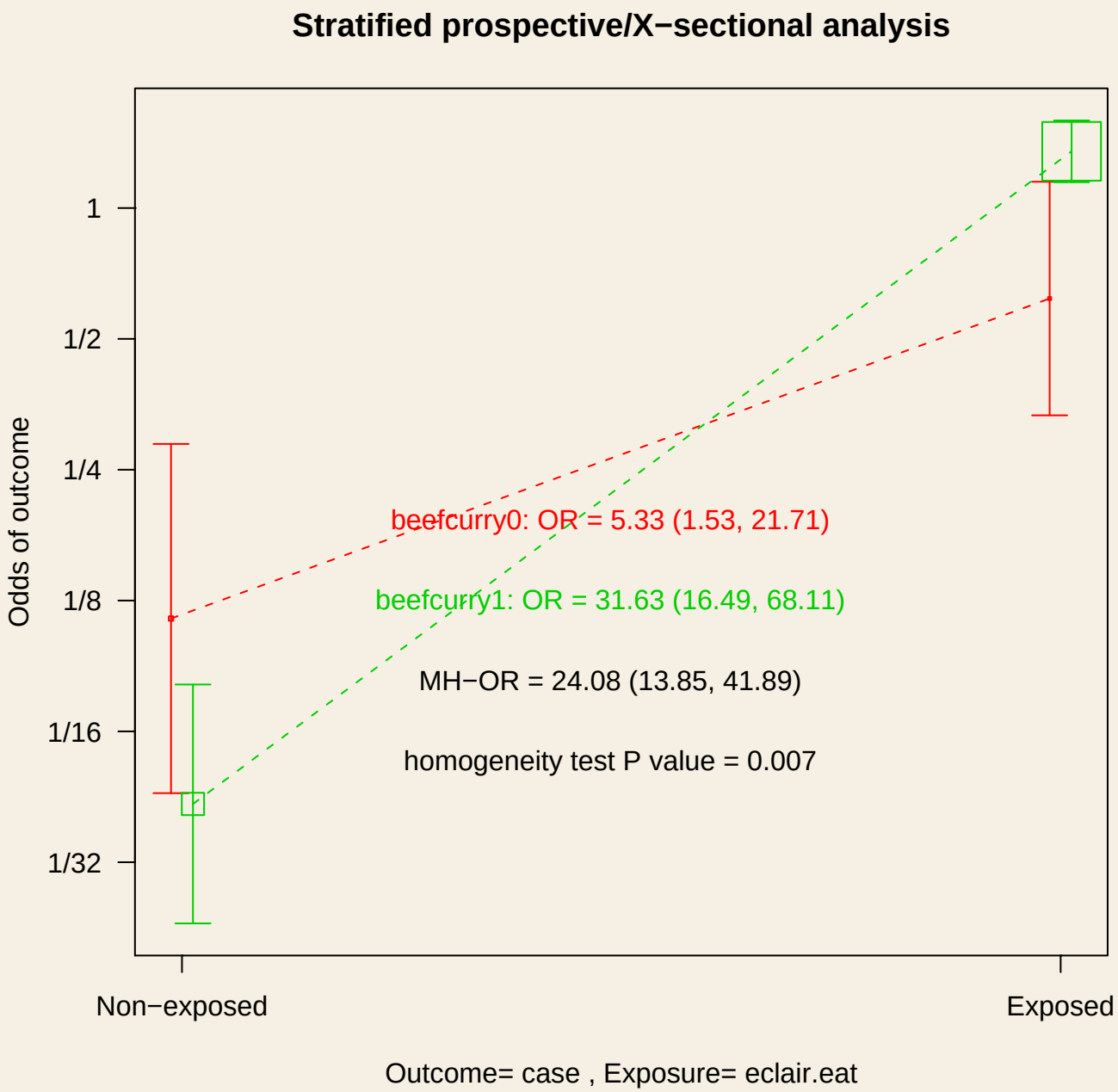

Stratified prospective/X−sectional analysis
Odds of outcome
2
1
1/2
1/4
1/8
1/16
1/32
eclair.eatFALSE: OR = 0.87 (0.22, 5)
eclair.eatTRUE: OR = 1.07 (0.48, 2.36)
MH−OR = 1.02 (0.54, 1.93)
homogeneity test P value = 0.787
Non−exposed
Exposed
Outcome= case , Exposure= saltegg

```
> mhor(case,eclair.eat,saltegg)

Stratified analysis by  saltegg
                OR lower lim. upper lim.  P value
saltegg 0    19.3       4.68      117.9 6.06e-07
saltegg 1    24.8      13.56       49.7 2.42e-51
M-H combined 24.3      13.96       42.4 8.12e-49

M-H Chi2(1) = 215.63 , P value = 0
Homogeneity test, chi-squared 1 d.f. = 0.11 , P value =
      0.736
```

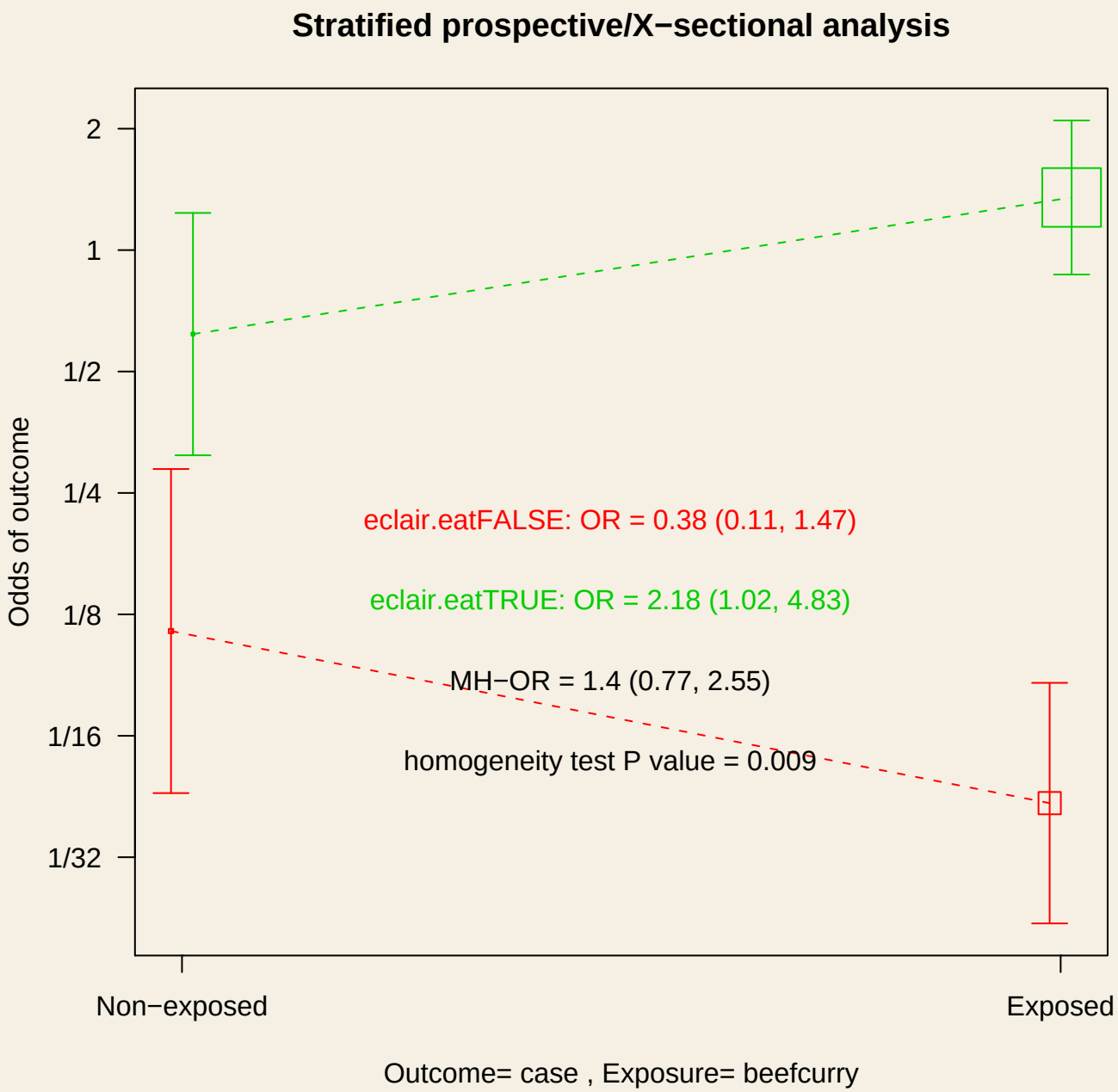

Stratified prospective/X−sectional analysis
Odds of outcome
2
1
1/2
1/4
1/8
1/16
1/32
saltegg0: OR = 19.31 (4.68, 117.88)
saltegg1: OR = 24.78 (13.56, 49.71)
MH−OR = 24.32 (13.96, 42.36)
homogeneity test P value = 0.736
Non−exposed
Exposed
Outcome= case , Exposure= eclair.eat

29　交互・修飾作用

```
> mhor(case,eclair.eat,beefcurry)

Stratified analysis by  beefcurry
                OR lower lim. upper lim.  P value
beefcurry 0   5.33       1.53       21.7 3.12e-03
beefcurry 1  31.63      16.49       68.1 4.79e-56
M-H combined 24.08      13.85       41.9 1.39e-48

M-H Chi2(1) = 214.56 , P value = 0
Homogeneity test, chi-squared 1 d.f. = 7.23 , P value =
      0.007
```


Stratified prospective/X−sectional analysis
Odds of outcome
1
1/2
1/4
1/8
1/16
1/32
beefcurry0: OR = 5.33 (1.53, 21.71)
beefcurry1: OR = 31.63 (16.49, 68.11)
MH−OR = 24.08 (13.85, 41.89)
homogeneity test P value = 0.007
Non−exposed
Exposed
Outcome= case , Exposure= eclair.eat

```
> mhor(case,beefcurry,eclair.eat)

Stratified analysis by  eclair.eat
                      OR lower lim. upper lim. P value
eclair.eat FALSE  0.376      0.111       1.47  0.1446
eclair.eat TRUE   2.179      1.021       4.83  0.0329
M-H combined      1.401      0.769       2.55  0.2396

M-H Chi2(1) = 1.38 , P value = 0.24
Homogeneity test, chi-squared 1 d.f. = 6.78 , P value =
      0.009
```


Stratified prospective/X−sectional analysis
Odds of outcome
2
1
1/2
1/4
1/8
1/16
1/32
eclair.eatFALSE: OR = 0.38 (0.11, 1.47)
eclair.eatTRUE: OR = 2.18 (1.02, 4.83)
MH−OR = 1.4 (0.77, 2.55)
homogeneity test P value = 0.009
Non−exposed
Exposed
Outcome= case , Exposure= beefcurry

30　症例対照研究

- ある疾病の患者群と対照（疾病に罹患していない者）群を設定し、過去に遡って仮説的要因の曝露群を比較して、要因と疾病の関連を調べる研究デザイン
- 相対リスクを計算できないので、近似的にオッズ比で代用
- まれな疾患で特に有効
- 交絡因子を調整するためにマッチドペア法がよく用いられる

31　マッチドペア

- 患者に対して性・年齢・その他の属性（兄弟、近隣など）が合致する対照を選ぶ
- 患者と対照は1:1または1:複数にする
- McNemar検定など対応のある検定を用いる

32 Rの活用の実際

- 食道がんの症例に対して、それぞれの近隣住民から性・年齢が一致した対照群が選ばれた

```
> data(VC1to6)
> use(VC1to6)
> head(.data)
  matset case smoking rubber alcohol
1      1    1       1      0       0
2      1    0       1      0       0
3      2    1       1      0       1
4      2    0       1      1       0
5      3    1       1      1       0
6      3    0       1      1       0
```

```
> tail(.data, n=10)
    matset case smoking rubber alcohol
110     25    1       0      0       0
111     25    0       1      1       0
112     25    0       1      1       1
113     25    0       1      0       1
114     25    0       1      0       0
115     26    1       1      0       1
116     26    0       0      0       0
117     26    0       1      1       0
118     26    0       0      0       0
119     26    0       1      1       1
```

- マッチする対照群を集めた結果、1:1から、最高で1:6になった

33　オッズ比の計算

```
> matchTab(case,smoking,strata=matset)

Exposure status: smoking = 1

Total number of match sets in the tabulation = 26

Number of controls = 1
                      No. of controls exposed
No. of cases exposed 0 1
                    0 0 0
                    1 0 3

Number of controls = 2
                      No. of controls exposed
No. of cases exposed 0 1 2
                    0 0 0 1
                    1 1 1 0
```

```
Number of controls = 5
                    No. of controls exposed
No. of cases exposed 0 1 2 3 4 5
                   0 0 1 0 0 0 0
                   1 0 1 0 1 0 0

Number of controls = 6
                    No. of controls exposed
No. of cases exposed 0 1 2 3 4 5 6
                   0 0 0 0 1 0 0 0
                   1 0 0 0 0 0 1 2

Odds ratio by Mantel-Haenszel method = 1.988

Odds ratio by maximum likelihood estimate (MLE) method =
      2.066
 95%CI= 0.678 , 6.301
```

- `matchTab()` により、マッチドペアの対照数ごとに6つの表が出力され、2種類の方法でオッズ比が計算された
- 95%信頼区間が1を含むので、食道がんに対する喫煙の効果は統計的に有意ではない

34　条件付きロジスティック回帰

```
> library(survival)
> colgit3 <- (clogit(case ~ smoking + alcohol + rubber +
      strata(matset)))
> colgit2 <- (clogit(case ~ alcohol + rubber + strata(matset
      )))
> colgit1 <- (clogit(case ~ alcohol + strata(matset)))
```

34　条件付きロジスティック回帰

```
> summary(colgit3)
Call:
coxph(formula = Surv(rep(1, 119L), case) ~ smoking + alcohol
      + rubber + strata(matset), method = "exact")
  n= 119
          coef exp(coef) se(coef)      z      p
smoking  0.440     1.552    0.646  0.681 0.5000
alcohol  1.667     5.295    0.595  2.800 0.0051
rubber  -0.457     0.633    0.647 -0.706 0.4800

        exp(coef) exp(-coef) lower .95 upper .95
smoking     1.552      0.644     0.437      5.51
alcohol     5.295      0.189     1.649     17.00
rubber      0.633      1.580     0.178      2.25

Rsquare= 0.096   (max possible= 0.471 )
Likelihood ratio test= 12  on 3 df,   p=0.00738
Wald test            = 9.18  on 3 df,   p=0.027
```

```
Score (logrank) test = 11.2  on 3 df,   p=0.0105
```

34　条件付きロジスティック回帰

```
> summary(colgit2)
Call:
coxph(formula = Surv(rep(1, 119L), case) ~ alcohol +
      rubber + strata(matset), method = "exact")
  n= 119
          coef exp(coef) se(coef)      z     p
alcohol  1.747     5.740    0.589  2.968 0.003
rubber  -0.393     0.675    0.645 -0.609 0.540

        exp(coef) exp(-coef) lower .95 upper .95
alcohol     5.740      0.174     1.810     18.20
rubber      0.675      1.481     0.191      2.39

Rsquare= 0.092   (max possible= 0.471 )
Likelihood ratio test= 11.5  on 2 df,   p=0.00314
Wald test            = 8.98  on 2 df,   p=0.0112
Score (logrank) test = 10.9  on 2 df,   p=0.00428
```

34　条件付きロジスティック回帰

```
> summary(colgit1)
Call:
coxph(formula = Surv(rep(1, 119L), case) ~ alcohol +
      strata(matset), method = "exact")
  n= 119
        coef exp(coef) se(coef)    z      p
alcohol 1.73      5.66    0.581 2.98 0.0029

        exp(coef) exp(-coef) lower .95 upper .95
alcohol      5.66      0.177      1.81      17.7

Rsquare= 0.089   (max possible= 0.471 )
Likelihood ratio test= 11.1  on 1 df,   p=0.000843
Wald test            = 8.9  on 1 df,   p=0.00286
Score (logrank) test = 10.7  on 1 df,   p=0.00105
```

34　条件付きロジスティック回帰

```
> lrtest(colgit3,colgit2)

Likelihood ratio test for Cox regression & conditional
      logistic regression
Chi-squared 1 d.f. =  0.4743344 , P value =  0.491
```

34　条件付きロジスティック回帰

```
> lrtest(colgit2,colgit1)

Likelihood ratio test for Cox regression & conditional
      logistic regression
Chi-squared 1 d.f. =  0.383735 , P value =  0.5356
```

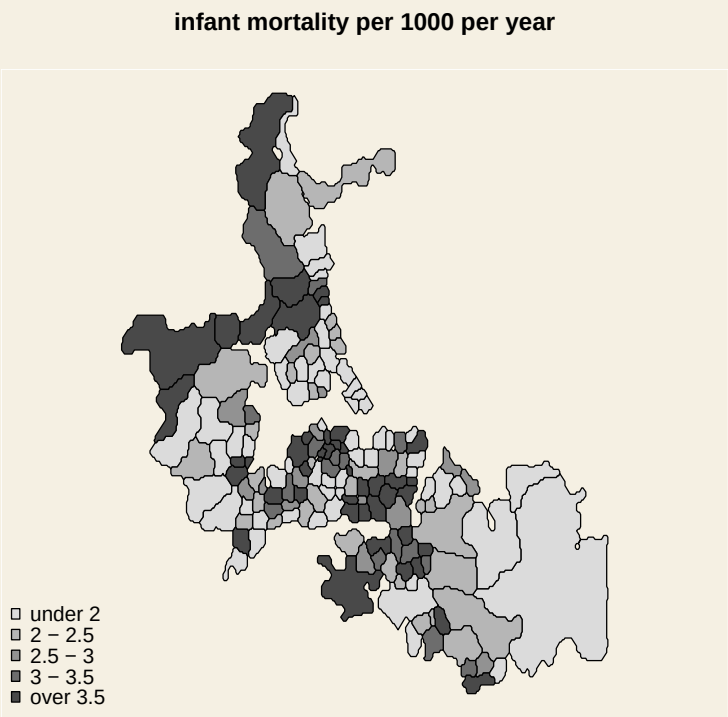
34　条件付きロジスティック回帰

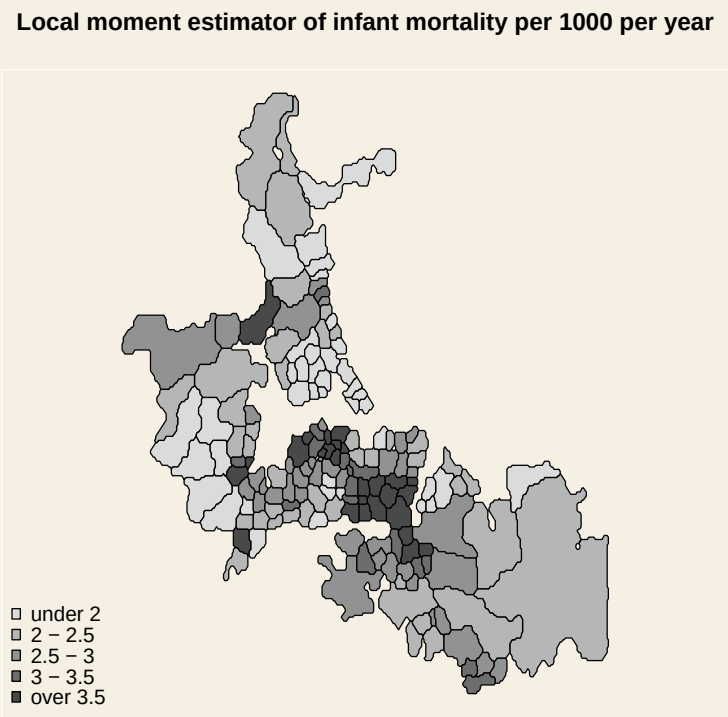
```
> anova(colgit3, colgit2,colgit1)
Analysis of Deviance Table

Model 1: Surv(rep(1, 119L), case) ~ smoking + alcohol +
       rubber + strata(matset)
Model 2: Surv(rep(1, 119L), case) ~ alcohol + rubber +
       strata(matset)
Model 3: Surv(rep(1, 119L), case) ~ alcohol + strata(matset)
  Resid. Df Resid. Dev  Df Deviance
1       116     63.788
2       117     64.262  -1   -0.474
3       118     64.646  -1   -0.384
```

35　空間疫学

- 空間的な位置関係に基づいて疫学像に迫る
- 詳しくはThe R Bookを参照

36　経験ベイズ平滑化


infant mortality per 1000 per year
under 2
2 − 2.5
2.5 − 3
3 − 3.5
over 3.5


Local moment estimator of infant mortality per 1000 per year
under 2
2 − 2.5
2.5 − 3
3 − 3.5
over 3.5