

Appendix 4

Data structure and R commands

Data structure and variable details used in manuscript I and manuscript II are described in this appendix. All commands for manipulating data, creating graphs and analyzing data from R programme are also illustrated. All of the details show in this appendix will be useful for any other statistical analyzer or researcher to follow and understand this thesis. Moreover, those who are interested in doing the same statistical analysis with similar data can also follow these methods.

4.1 Manuscript I

4.1.1 Model for HIV/AIDS

4.1.1.1 Data structure

Variable	Descriptive	Type	Size	Labels
supdist	Super-district	Integer	2	The codes for each super-district are shown in appendix 1.
sex	Gender	Byte	1	1 = Male, 2 = Female

Variable	Descriptive	Type	Size	Labels
inhos	Place of death	Byte	1	1 = in-hospital , 2 = Outside hospital
season	di-month period Should this be di-month or bi-month- generally??	Byte	1	1 = Jan – Feb, 2 = Mar – Apr, 3 = May – Jun, 4 = July – Aug, 5 = Sep – Oct, 6 = Nov - Dec
dmonth	The code for di-month period. The codes are -5, -4, -3, -2, -1, 0, 1, 2, ... 30.	Numeric	3	
pop	Number of population	Integer	6	
death	Number of deaths from HIV/AIDS	Integer	3	

4.1.1.2 Commands in R

```
#####
# Commands in R for forecasting demographic time series counts for
# HIV/AIDS mortality
# Data file:          hiv.txt
# Determinants:      sex, inhos, season, supdist, lag 1-3, lag1OD,
#                   lag1FtoM, lag1MtoF
# Outcome:          HIV/AIDS incidence mortality rate
# Statistical Method: Negative binomial
# Graph:           Residual plot
#####
setwd("C:/phdthesis")
library(MASS)

hiv <- read.table("hiv.txt", header=T)

nInhos <- length(unique(hiv$inhos))

nSex <- length(unique(hiv$sex))

nMonth <- length(unique(hiv$dmonth))

nSupdist <- length(unique(hiv$supdist))

nObs <- nInhos*nSex*nMonth*nSupdist

# calculate popAllD, popAllA, nAllD, nAllA for computing
# lag1, lag1, lag2, lag3, lag1OD, log1OA

# compute laged incidence rates for each combination of district,
```

```

# gender, in-hospital and period

k <- 500000

hiv$y <- log(1+k*hiv$death/hiv$pop)

hiv$lag1 <- c(0,hiv$y)[1:nObs]

hiv$lag1 <- ifelse(hiv$dmonth==5,0, hiv$lag1)

hiv$lag2 <- c(0,0,hiv$y)[1:nObs]

hiv$lag2 <- ifelse(hiv$dmonth<=4,0, hiv$lag2)

hiv$lag3 <- c(0,0,0,hiv$y)[1:nObs]

hiv$lag3 <- ifelse(hiv$dmonth<=3,0, hiv$lag3)

# Compute the prevalence in other districts for the same gender,

# in-hospital and time

# populations by gender, in-hospital and district
tPop <- tapply(hiv$pop,list(hiv$supdist,hiv$sex),sum)/(nMonth*nInhos)
popAllD <- colSums(tPop)
popAllS <- rowSums(tPop)

nDeathAD <- tapply(hiv$death,list(hiv$sex),sum)
nDeathAS <- tapply(hiv$death,list(hiv$supdist),sum)

hiv <- hiv[order(hiv$sex,hiv$in hos,hiv$supdist,hiv$dmonth),]

hiv$popAllD <- rep(popAllD,each=nSupdist*nMonth)
hiv$popAllS <- rep(popAllS,each=nSex*nMonth)

# ... and by gender, in-hospital and month

tND <- tapply(hiv$death,list(hiv$dmonth,hiv$sex,hiv$in hos),sum)

dim(tND) <- c(nMonth*nSex*nInhos)

```

```

hiv <- hiv[order(hiv$sex,hiv$inhos,hiv$supdist,hiv$dmonth),]

hiv$nAllD <- rep(tND,nSupdist)

# ... and by district, in-hospital and month

tNS <- tapply(hiv$death,list(hiv$dmonth,hiv$supdist,hiv$inhos),sum)

dim(tNS) <- c(nMonth*nSupdist*nInhos)

hiv <- hiv[order(hiv$sex,hiv$inhos,hiv$supdist,hiv$dmonth),]

hiv$nAllS <- rep(tNS,nSex)

# compute lag 1 incidence rate for other districts

hiv$lag1OD <- log(1+500000*(hiv$nAllD-hiv$death)/(hiv$popAllD-
hiv$pop))

hiv$lag1OD <- c(0,hiv$lag1OD)[1:nObs]

hiv$lag1OD <- ifelse(hiv$dmonth== -5,0,hiv$lag1OD)

# Compute lag 1 incidence rate for female to male

hiv$lag1FtoM <- log(1+500000*(hiv$nAllS-hiv$death)/(hiv$popAllS-
hiv$pop))

hiv$lag1FtoM <- c(0,hiv$lag1FtoM)[1:nObs]

hiv$lag1FtoM <- ifelse(hiv$dmonth== -5,0,hiv$lag1FtoM)

hiv$lag1FtoM <- ifelse(hiv$sex==2,0,hiv$lag1FtoM)

# Compute lag 1 incidence rate for male to female

hiv$lag1MtoF <- log(1+500000*(hiv$nAllS-hiv$death)/(hiv$popAllS-
hiv$pop))

hiv$lag1MtoF <- c(0,hiv$lag1MtoF)[1:nObs]

hiv$lag1MtoF <- ifelse(hiv$dmonth== -5,0,hiv$lag1MtoF)

```

```

hiv$lag1MtoF <- ifelse(hiv$sex==1,0,hiv$lag1MtoF)

hiv <- subset(hiv,dmonth>-3)

hiv.out <- glm.nb(data=hiv,
death~offset(log(pop/100000))+factor(supdist)+factor(season)+
factor(sex)+factor(inhos)+lag1+lag2+lag3+lag1OD+lag1FtoM+lag1MtoF)

summary(hiv.out,corr=F)

anova(hiv.out)

sum(hiv.out$fitted)

# Compute and plot deviance residuals
resDev <- resid(hiv.out,type="deviance")

qqnorm(resDev,cex=1,main="",ylab="",xlab="",xaxp=c(-4,4,8),yaxp=c(-
4,4,8), xlim=c(-4.2,4.2),ylim=c(-4.2,4.2),las=1,tcl=-
0.3,cex.axis=1.3)

abline(0,1,col="red")

minR <- min(resDev)

maxR <- max(resDev)

mtext("HIV/AIDS mortality among persons aged 15+", side=3,
adj=0.5,cex=1.5, line=2.5,font=2)

mtext("Deviance Residuals",side=3,adj=-0.1,cex=1.3,line=0.6)

mtext("Normal Scores",side=1,adj=0.5,cex=1.3,line=2.5)

polygon(c(0.3,0.3,0.3,3.7,3.7),c(minR+0.1,minR+0.1,minR+1.5,minR+1.5,m
inR+0.1),col="white")

text(0.5,(minR+1.1),paste("deviance =", round(hiv.out$deviance,1)),
cex=0.9,adj=0)

```

```
text(0.5, (minR+0.6), paste("resid.df =", hiv.out$df.residual),
cex=0.9, adj=0)

theta <- round(hiv.out$theta, 3)

thetaSE <- round(hiv.out$SE.theta, 3)

text(-3.0, maxR-1, paste("theta (SE)
=", theta, "(", thetaSE, ")"), cex=0.9, adj=0)

# calculate coeff and residual deviance

hiv.out <- glm.nb(data=hiv,
death~offset(log(pop/100000))+factor(supdist)+factor(season)+
factor(sex)+factor(inhos))

summary(hiv.out, corr=F)

hiv.out <- glm(data=hiv,
death~offset(log(pop/100000))+factor(supdist)+factor(season)+
factor(sex)+factor(inhos), family=negative.binomial(5.041))

summary(hiv.out, corr=F)
```

4.1.2 Model for other infectious diseases

4.1.2.1 Data structure

Variable	Descriptive	Type	Size	Labels
supdist	Super-district	Integer	2	The codes for each super-district are shown in appendix 1.
season	di-month period	Byte	1	1 = Jan – Feb, 2 = Mar – Apr, 3 = May – Jun, 4 = July – Aug, 5 = Sep – Oct, 6 = Nov - Dec
diagnosis	Principal diagnosis	Byte	1	1 = Unspecified septicaemia 2 = Other infectious disease
dmonth	The code for di-month period. The codes are -5, -4, -3, -2, -1, 0, 1, 2, ... 30.	Numeric	3	
pop	Number of population	Integer	6	

Variable	Descriptive	Type	Size	Labels
death	Number of deaths from infectious disease excluded HIV/AIDS.	Integer	3	

4.1.2.2 Commands in R

```
#####
# Commands in R for forecasting demographic time series counts for
# hiv/aids mortality
# Data file:          sep.txt
# Determinants:      supdist, season, diagnosis, lag 1-3, lag10D
# Outcome:           other infectious disease incidence
#                   mortality rate
# Statistical Method: Negative binomial
# Graph:             Residual plot
#####

setwd("C:/phdthesis")

rm(list=ls())

library(MASS)

sep <- read.table("sep.txt", header=T)

nSep <- length(unique(sep$a419))

nMonth <- length(unique(sep$dmonth))

nSupdist <- length(unique(sep$supdist))
```

```

nAge <- length(unique(sep$agegr))

nObs <- nSep*nAge*nMonth*nSupdist

# calculate popAllD,popAllA,nAllD,nAllA for compute
lag1,lag1,lag2,lag3,lag1OD,log1OA

# Compute laged incidence rates for each combination of district,
gender, in-hospital and period

k <- 1000000

sep <- sep[order(sep$agegr,sep$supdist,sep$a419,sep$dmonth),]

sep$y <- log(1+k*sep$death/sep$pop)

sep$lagg1 <- c(0,sep$y)[1:nObs]

sep$lagg1 <- ifelse(sep$dmonth==5,0, sep$lagg1)

sep$lagg2 <- c(0,0,sep$y)[1:nObs]

sep$lagg2 <- ifelse(sep$dmonth<=4,0, sep$lagg2)

sep$lagg3 <- c(0,0,0,sep$y)[1:nObs]

sep$lagg3 <- ifelse(sep$dmonth<=3,0, sep$lagg3)

a <- sep[,c("lag1","lagg1","lag2","lagg2","lag3","lagg3")]

# First compute the prevalence in other districts for the same age

# group, diagnosis and time

# Populations by age group, diagnosis and district

tPop <- tapply(sep$pop,list(sep$supdist,sep$agegr),sum)/(nMonth*nSep)

popAllD <- colSums(tPop)

popAllA <- rowSums(tPop)

nDeathAD <- tapply(sep$death,list(sep$agegr),sum)

```

```

nDeathAA <- tapply(sep$death, list(sep$supdist), sum)

sep <- sep[order(sep$agegr, sep$supdist, sep$a419, sep$dmonth), ]

sep$popAllD <- rep(popAllD, each=nSupdist*nMonth*nSep)

sep$popAllA <- rep(popAllA, each=nAge*nMonth*nSep)

# ... and by age group, disease group and month

tND <- tapply(sep$death, list(sep$dmonth, sep$a419, sep$agegr), sum)

dim(tND) <- c(nMonth*nAge*nSep)

sep <- sep[order(sep$supdist, sep$agegr, sep$a419, sep$dmonth), ]

sep$nAllD <- rep(tND[c(1:72, 73:144, 145:216)], nSupdist)

# ... and by district, in-hospital and month

tNA <- tapply(sep$death, list(sep$dmonth, sep$a419, sep$supdist), sum)

dim(tNA) <- c(nMonth*nSupdist*nSep)

sep <- sep[order(sep$supdist, sep$agegr, sep$a419, sep$dmonth), ]

sep$nAllA <- rep(tNA, nAge)

# compute lag 1 incidence rate for other districts

sep$lagg1OD <- log(1+1000000*(sep$nAllD-sep$death)/(sep$popAllD-
sep$pop))

sep$lagg1OD <- c(0, sep$lagg1OD)[1:nObs]

sep$lagg1OD <- ifelse(sep$dmonth== -5, 0, sep$lagg1OD)

# Compute lag 1 incidence rate for other age group

sep$lagg1OA <- log(1+1000000*(sep$nAllA-sep$death)/(sep$popAllA-
sep$pop))

```

```

sep$lagg10A <- c(0, sep$lagg10A) [1:nObs]

sep$lagg10A <- ifelse(sep$dmonth== -5, 0, sep$lagg10A)

sep <- subset(sep, dmonth>-3 & agegr==3)

sep.out <- glm(family=gaussian, lnincK1m~factor(supdist)+factor(a419)+
factor(season)+lag1+lag2+lag3+lag10D)

summary(sep.out, corr=F)

sep.out <-
glm(family=poisson, death~offset(log(pop/100000))+factor(supdist)+
factor(a419)+ factor(season)+lag1+lag2+lag3+lag10D)

summary(sep.out, corr=F)

sum(sep.out$fitted)

anova(sep.out)

sep.out <- glm.nb(death~offset(log(pop/100000))+factor(supdist)+
factor(a419)+factor(season)+lag1+ lag2+lag3+lag10D)

summary(sep.out, corr=F)

anova(sep.out)

sum(sep.out$fitted)

par(oma=c(1, 1, 1, 0.5))

# compute and plot deiance residuals

resDev <- resid(sep.out, type="deviance")

qqnorm(resDev, cex=1, main="", ylab="", xlab="", xaxp=c(-4, 4, 8), yaxp=c(-
4, 4, 8), xlim=c(-4.2, 4.2), ylim=c(-4.2, 4.2), las=1, tcl=-0.3,
cex.axis=1.3)

abline(0, 1, col="red")

```

```
minR <- min(resDev)

maxR <- max(resDev)

mtext("Hospital infectious disease mortality among persons aged 40+",
      side=3, adj=0.5, cex=1.5, line=2.1, font=2)

mtext("Deviance Residuals", side=3, adj=-0.1, cex=1.3, line=0.6)

mtext("Normal Scores", side=1, adj=0.5, cex=1.3, line=2.3)
```

Prince of Songkla University
Pattani Campus

4.2 Manuscript II

4.2.1 Logistic model

4.2.1.1 Data structure

Variable	Descriptive	Type	Size	Labels
PdxSexAge	The combination of principal diagnosis, gender and age group with 54 categories. The codes are 111, 112, 113, 121, 122, 123, 211, 212, 213, 221, 222, 223, ..., 923.	Integer	3	Pdx: 1 = Injuries, 2 = CVD, 3 = Septicaemia, 4 = Other ID, 5 = DD, 6 = COPD, 7 = RI, 8 = Cancer, 9 = Other Sex: 1 = Male, 2 = Female Age: 1 = < 60 years, 2 = 60 – 74 years, 3 = 75+ years

Variable	Descriptive	Type	Size	Labels
ProvHosGrp	The combination of province, region and hospital size with 19 categories. The codes are 11, 12, 13, 21, 22, 23, 31, 32, 33, ..., 73.	Integer	2	Prov: 1 = North, 2 = Surat Thani, 3 = North West, 4 = Nakhon Si Thammarat, 5 = South West, 6 = Central, 7 = South East HosGrp: 1 = < 61 beds 2 = 61 – 499 beds 3 = 500 + beds
Long	Number of days stay in hospital at least 7 days.	Integer	4	
Short	Number of days stay in hospital less than 7 days.	Integer	4	

4.2.1.2 Commands in R

```
#####

# Commands in R for analyzing length of hospital stay data

# Data file:           LOSLG.txt

# Determinants:       PdxSexAge and ProvHosGrp

# Outcome:           Long and Short

# Statistical Method:  Logistic regression

# Graph:             Residual and Confidence interval plot

#####

# Setting up working directory and reading dataset into R

setwd("C:/phdthesis/")

LOSLG <- read.table("LOSLG.txt",h=T,sep="\t")

# Changing variable type from integer to factor

LOSLG$PdxSexAge <- as.factor(LOSLG$PdxSexAge)

LOSLG$ProvHosGrp <- as.factor(LOSLG$ProvHosGrp)

# Choosing the category with the largest number of cases as the
referent group

LOSLG$ProvHosGrp <- relevel(LOSLG$ProvHosGrp,ref="43")

# Creating logistic model for analysis the data together with

# calculating deviance

# residual
```



```
model <- glm(data=LOSLG,
cbind(Long, Short)~factor(PdxSexAge)+factor(ProvHosGrp),
family=binomial)

rez <- summary(model)

res <- residuals(model, type="deviance")

# Creating new window for plotting the graphs

windows(8,4)

par(mfrow=c(1,2), mar=c(4,3,2,0.5), las=1)

LOSLG$Total <- LOSLG$Short+LOSLG$Long

fv <- LOSLG$Total*model$fitted.values

obs <- LOSLG$Long

xymax <- max(sqrt(fv), sqrt(obs))

xymin <- min(sqrt(fv), sqrt(obs))

# Creating residual plot and observed versus fitted values plot

plot(sqrt(fv), sqrt(obs), cex=1, cex.axis=1,
ylim=c(0, xymax), xlim=c(0, xymax), main="", ylab="", xlab="", xaxt="n",
yaxt="n")

abline(0,1)

xlab <- c(0,4,20,50,100,200)

ylab <- xlab

xTicks <- sqrt(c(0,4,20,50,100,200))

yTicks <- xTicks

axis(side=1, at=xTicks, labels=xlab, cex.axis=1)

axis(side=2, at=yTicks, labels=ylab, cex.axis=1)
```

```

mtext("Observed counts",side=3,adj=-0.1,cex=1,line=0.3)

mtext("Fitted values", side=1, adj=0.5, cex=1, line=2)

qqnorm(res, cex=0.5, cex.axis=1, main="",
ylab="",xlab="",cex.lab=1,pch=16)

mtext("Deviance residuals",side=3,adj=-0.1,cex=1,line=0.3)

mtext("Normal quantiles", side=1, adj=0.5, cex=1, line=2)

abline(0,1,col="red")

# Get coefficients and SEs of medical Demographic effects
n <- length(unique(LOSLG$PdxSexAge))

Coef <- c(0,rez$coef[2:n,1])

SE <- c(0,rez$coef[2:n,2])

# Get proportions of death by medical demographic factors
Short <- tapply(LOSLG$Short,LOSLG$PdxSexAge,sum)
Long <- tapply(LOSLG$Long,LOSLG$PdxSexAge,sum)

ni <- as.numeric(Short)+as.numeric(Long)

pS <- as.numeric(Long)/ni

total <- sum(pS*ni)

overallPrev <- 100*total/(sum(Short)+sum(Long))

# Use Newton-Raphson method to estimate constant to match observed

# and fitted

# values for years

ExpAi <- exp(-Coef)

delta <- 0.5 # Marquardt damping constant

```

```

epsilon <- 0.00005      # convergence criterion

nit <- 20                # maximum number of iterations

b0 <- 0                  # initial value of constant

b1 <- 0

it <- 0

bDiff <- 1

while ( (abs(bDiff)>epsilon) && ((it <- it+1) < nit) ) {

  adjpShsy <- 1/(1+exp(-(b1+Coef)))

  F0 <- sum(ni*adjpShsy) - total

  DF0 <- sum(ni*ExpAi*adjpShsy^2)*exp(-b1)

  b1 <- b0-delta*(F0/DF0)

  bDiff <- b1-b0

  b0 <- b1

  cat(it,b0,"\n")

}

b0 <- -1.861986      # from Newton-Raphson method

adjpShsyM <- 100/(1+exp(-(b0 + Coef[c(1:3, 7:9, 13:15, 19:21, 25:27,
31:33, 37:39, 43:45, 49:51)]))))

CILBxM <- 100/(1+exp(-(b0 + Coef[c(1:3, 7:9, 13:15, 19:21, 25:27,
31:33, 37:39, 43:45, 49:51)]-1.96*SE[c(1:3, 7:9, 13:15, 19:21, 25:27,
31:33, 37:39, 43:45, 49:51)]))))

CIUBxM <- 100/(1+exp(-(b0 + Coef[c(1:3, 7:9, 13:15, 19:21, 25:27,
31:33, 37:39, 43:45, 49:51)]+1.96*SE[c(1:3, 7:9, 13:15, 19:21, 25:27,
31:33, 37:39, 43:45, 49:51)]))))

```

```

adjpShsyF <- 100/(1+exp(-(b0 + Coef[c(4:6, 10:12, 16:18, 22:24,
28:30, 34:36, 40:42, 46:48, 52:54)])))

CILBxF <- 100/(1+exp(-(b0 + Coef[c(4:6, 10:12, 16:18, 22:24, 28:30,
34:36, 40:42, 46:48, 52:54)]-1.96*SE[c(4:6, 10:12, 16:18, 22:24,
28:30, 34:36, 40:42, 46:48, 52:54)])))

CIUBxF <- 100/(1+exp(-(b0 + Coef[c(4:6, 10:12, 16:18, 22:24, 28:30,
34:36, 40:42, 46:48, 52:54)]+1.96*SE[c(4:6, 10:12, 16:18, 22:24,
28:30, 34:36, 40:42, 46:48, 52:54)])))

# Creating new window for plotting confidence interval

windows(6,6)

layout(height=c(2.8,0.40,2.8),rbind(1,2,3))

par(mar=c(1,1,1,1), mgp=c(1.1,0.2,0), oma=c(2,2.5,2,0), las=1,
tcl=0.2, lwd=1)

xcoordM <- c(1:3,10:12,19:21,28:30,37:39,46:48,55:57,64:66,73:75)
xcoordF <- c(5:7,14:16,23:25,32:34,41:43,50:52,59:61,68:70,77:79)

# Creating confidence interval plot

plot(xcoordM,adjpShsyM,cex=1.5,pch=16,ylim=c(1,70),xlim=c(0,80),cex.a
xis=1,cex.lab=1,main="",ylab="",xlab="",xaxt="n",xaxs="i")

points(xcoordF,adjpShsyF,cex=1.5,pch=21,ylim=c(1,70),xlim=c(0,80),
cex.axis=1,cex.lab=1,main="",ylab="",xlab="",xaxt="n",xaxs="i",col="m
agenta")

abline(v=8.5+(0:7)*9,col="darkgrey")

abline(h=overallPrev,lty="13")

# ...with 95% confidence intervals

for (i in (1:27)) {

```

```

lines(c(0*(1:2))+xcoordM[i],c(CILBxM[i],CIUBxM[i]))

lines(c(0*(1:2))+xcoordF[i],c(CILBxF[i],CIUBxF[i]),col="magenta")

}

disn <- c("Injuries","CVD","Septicaemia","Other ID", "DD", "COPD",
"RI", "Cancer", "Other")

age1 <- rep(c(1:3),9)

mtext(age1,at=xcoordM, line=-1.5,side=1,cex=0.6,adj=0.3)

mtext(age1,at=xcoordF, line=-1.5,side=1,cex=0.6,adj=0.3)

axis(1,at=xcoordM,labels=F)

axis(1,at=xcoordF,labels=F)

mtext(disn,at=c(4.5,13.5,22.5,31.5,40.5,49.5,58.5,67.5,76.5), side=1,
cex=0.7, adj=0.5)

mtext("Length of stay at least 7 days",side=3,adj=0.5,cex=0.8,line=2)
mtext("Percent",side=3,adj=-0.05,cex=0.7,line=0.3)

mtext("Principal diagnosis group by increasing age and
gender",side=1,adj=0.5,cex=0.8,line=1.5)

legend(13.6,69,c("Male","Female"),col=c("black","magenta"),
pt.cex=1.5, cex=1, lty=c(1,1), pch=c(16,21), inset=0.02, bg="white")

legend(1,69,pch=c("1","2","3"), c("< 60 years", "60-74 years", "75+
years"), pt.cex=1, cex=1, bg="white")

plot(1,col="white",xaxt="n", yaxt="n",xlab="",ylab="",frame.plot=F)

# Get coefficients and SEs of medical Geographic effects

n <- length(unique(LOSLG$ProvHosGrp))

Coef <- c(rez$coef[55:65,1],0,rez$coef[66:72,1])

```

```

SE <- c(rez$coef[55:65,2],0,rez$coef[66:72,2])

# Get proportions of death by medical demogrphic factors

tapply(LOSLG$Short,LOSLG$ProvHosGrp,sum) -> Short

tapply(LOSLG$Long,LOSLG$ProvHosGrp,sum) -> Long

ni <- as.numeric(Short)+as.numeric(Long)

pS <- as.numeric(Long)/ni

total <- sum(pS*ni)

overallPrev <- 100*total/(sum(Short)+sum(Long))

# Use Newton-Raphson method to estimate constant to match observed

# and fitted

# values for years

ExpAi <- exp(-Coef)

delta <- 0.5          # Marquardt damping constant

epsilon <- 0.00005   # convergence criterion

nit <- 20            # maximum number of iterations

b0 <- 0              # initial value of constant

b1 <- 0

it <- 0

bDiff <- 1

while ( (abs(bDiff)>epsilon) && ((it <- it+1) < nit) ) {

  adjpShsy <- 1/(1+exp(-(b1+Coef)))

  F0 <- sum(ni*adjpShsy) - total

  DF0 <- sum(ni*ExpAi*adjpShsy^2)*exp(-b1)

```

```

b1 <- b0-delta*(F0/DF0)

bDiff <- b1-b0

b0 <- b1

cat(it,b0,"\n")

}

b0 <- -0.6836682      # from Newton-Raphson method

adjpShsy <- 100/(1+exp(-(b0+Coef)))

CILBx <- 100/(1+exp(-(b0+Coef-1.96*SE)))

CIUBx <- 100/(1+exp(-(b0+Coef+1.96*SE)))

xcoord <- c(1:3,5:7,9:11,13:15,17:18,21:23,25:26)

plot(xcoord, adjpShsy,cex=1.5, pch=16, ylim=c(1,70), xlim=c(0,28),
     cex.axis=1, cex.lab=1,main="", ylab="", xlab="", xaxt="n", xaxs="i")

abline(v=4+(0:5)*4,col="darkgrey")
abline(h=overallPrev,lty="13")

# ...with 95% confidence intervals

for (i in (1:19)) {

  lines(c(0*(1:2))+xcoord[i],c(CILBx[i],CIUBx[i]))

}

xPos <- c(2,6,10,14,18,22,26)

disn <- c("North","Surat Thani","North West","Nakhon","South
West","Central","South East")

bedl <- rep(c("S","M","L"),4)

bedl <- c(bedl,"S","M","S","M","L","S","M")

```

```
mtext (bed1,at=xcoord, side=1, cex=0.6,adj=0.5, line=-1.5)

axis(1,at=xcoord,labels=F)

mtext (disn,at=xPos,side=1,cex=0.7,adj=0.5)

mtext ("Percent",side=3,adj=-0.05,cex=0.7,line=0.3)

mtext ("Province group by hospital
size",side=1,adj=0.5,cex=0.8,line=1.3)

legend(0.35,69,pch=c("S","M","L"), c("< 61 beds", "61-499 beds",
"500+ beds"), pt.cex=1, cex=1, bg="white")
```

Prince of Songkla University
Pattani Campus

4.2.2 Linear model

4.2.2.1 Data structure

Variable	Descriptive	Type	Size	Label
PdxSexAge	The combination of principal diagnosis, gender and age group with 54 categories. The codes are 111, 112, 113, 121, 122, 123, 211, 212, 213, 221, 222, 223, ..., 923.	Integer	3	<p>Pdx: 1 = Injuries, 2 = CVD, 3 = Septicaemia, 4 = Other ID, 5 = DD, 6 = COPD, 7 = RI, 8 = Cancer, 9 = Other</p> <p>Sex: 1 = Male, 2 = Female</p> <p>Age: 1 = < 60 years, 2 = 60 – 74 years, 3 = 75+ years</p>

Variable	Descriptive	Type	Size	Label
ProvHosGrp	The combination of province, region and hospital size with 19 categories. The codes are 11, 12, 13, 21, 22, 23, 31, 32, 33, ..., 73.	Integer	2	Prov: 1 = North, 2 = Surat Thani, 3 = North West, 4 = Nakhon Si Thammarat, 5 = South West, 6 = Central, 7 = South East HosGrp: 1 = < 61 beds 2 = 61 – 499 beds 3 = 500 + beds
dayslos	Number of days stay in hospital.	Integer	4	

4.2.2.2 Commands in R

```
#####

# Commands in R for analyzing length of hospital stay data

# Data file:           LOSLN.txt

# Determinants:       PdxSexAge and ProvHosGrp

# Outcome:            dayslos

# Statistical Method:  Linear regression

# Graph:              Histogram, residual and confidence interval plot

#####

# Setting up working directory and reading dataset into R

setwd("C:/phdthesis/")

LOSLN <- read.table("LOSLN.txt",h=T,sep="\t")

windows(5,3)

par(mar=c(4,3,2,0))

LOSLN$lnlos <- log(LOSLN$dayslos+0.4)

h <- hist(LOSLN$lnlos,breaks=40,col="lightblue", freq=T, las=1,
xlim=c(-2,8),ylab="",main="", xlab="")

xfit<-seq(min(-2.5),max(LOSLN$lnlos),length=40498)

yfit<-dnorm(xfit,mean=mean(LOSLN$lnlos),sd=sd(LOSLN$lnlos))

abline(h=0,xlim=c(-2.4,8))

lines(xfit,yfit * 40498 * (h$breaks[2]-h$breaks[1]), col="red")

mtext("Frequency",adj=-0.1,line=0.5)
```

```
mtext("ln(LOS+0.4)", side=1, adj=0.5, line=2)

# Transforming "dayslos" by taking natural logarithm and plus 0.4

LOSLN$lnlos <- log(LOSLN$dayslos+0.4)

# Changing variable type from integer to factor

LOSLN$ProvHosGrp <- as.factor(LOSLN$ProvHosGrp)

# Choosing the category with the largest number of cases as the
referent group

LOSLN$ProvHosGrp <- relevel(LOSLN$ProvHosGrp, ref="43")

# Creating logistic model for analysis the data together with
calculating deviance

# residual

model <- lm(data=LOSLN, lnlos~factor(PdxSexAge)+factor(ProvHosGrp))

rez <- summary(model)
res <- rstudent(model)

windows(4,4)

par(mfrow=c(1,1), mar=c(4,3,2,0.5), las=1)

sym <- ifelse(LOSLN$dayslos==0,16,ifelse(LOSLN$dayslos==1,16,16))

cx <- ifelse(LOSLN$dayslos==0,0.5,ifelse(LOSLN$dayslos==1,0.7,0.5))

qqnorm(res, cex=cx, cex.axis=1, main="",
ylab="", xlab="", cex.lab=1, pch=sym)

mtext("Standardized residuals", side=3, adj=-0.1, cex=1, line=0.3)

mtext("Normal quantiles", side=1, adj=0.5, cex=1, line=2)

abline(0,1,col="red")
```

```

points(-0.6,-0.8,pch=21,cex=8)

mtext("LOS = 1",at=0.8,line=-4,side=1,cex=0.8)

na <- length(unique(LOSLN$PdxSexAge))

Coefa <- c(0,rez$coef[2:na,1])

SEa <- c(0,rez$coef[2:na,2])

nb <- length(unique(LOSLN$ProvHosGrp))

Coefb <- c(rez$coef[55:65,1],0,rez$coef[66:72,1])

SEb <- c(rez$coef[55:65,2],0,rez$coef[66:72,2])

alpha <- as.numeric(c(0,rez$coef[2:54,1]))

a <- data.frame(table(LOSLN$PdxSexAge))

a <- a$Freq

alphaNi <- sum(a*alpha)

amu <- (sum(LOSLN$lnlos)-alphaNi)/length(LOSLN$lnlos)

beta <- as.numeric(c(rez$coef[55:65,1],0,rez$coef[66:72,1]))

b <- as.data.frame(table(LOSLN$ProvHosGrp))

b <- b$Freq

betaNi <- sum(b*beta)

bmu <- (sum(LOSLN$lnlos)-betaNi)/length(LOSLN$lnlos)

adjalpha <- exp(alpha+amu+(1.377527^2/2))-0.4

adjbeta <- exp(beta+bmu+(1.422631767^2/2))-0.4

CILBa <- exp(alpha+amu+(1.377527^2/2)-1.96*SEa)-0.4

CIUBa <- exp(alpha+amu+(1.377527^2/2)+1.96*SEa)-0.4

```

```

adjalphaM <- adjalpha[c(1:3, 7:9, 13:15, 19:21, 25:27, 31:33,
37:39, 43:45, 49:51)]

adjalphaF <- adjalpha[c(4:6, 10:12, 16:18, 22:24, 28:30, 34:36,
40:42, 46:48, 52:54)]

CILBxM <- CILBa[c(1:3, 7:9, 13:15, 19:21, 25:27, 31:33, 37:39, 43:45, 49:51)]

CIUBxM <- CIUBa[c(1:3, 7:9, 13:15, 19:21, 25:27, 31:33, 37:39, 43:45, 49:51)]

CILBxF <- CILBa[c(4:6, 10:12, 16:18, 22:24, 28:30, 34:36, 40:42,
46:48, 52:54)]

CIUBxF <- CIUBa[c(4:6, 10:12, 16:18, 22:24, 28:30, 34:36, 40:42,
46:48, 52:54)]

mean <- mean(LOSLN$dayslos)

CILBb <- exp(beta+bmu+(1.42263182/2)-1.96*SEb)-0.4

CIUBb <- exp(beta+bmu+(1.42263182/2)+1.96*SEb)-0.4

##### plot CI #####

windows(6,6)

layout(height=c(2.8,0.40,2.8),rbind(1,2,3))

par(mar=c(1,1,1,1), mgp=c(1.1,0.2,0), oma=c(2,2.5,2,0), las=1,
tcl=0.2, lwd=1)

xcoordM <- c(1:3,10:12,19:21,28:30,37:39,46:48,55:57,64:66,73:75)

xcoordF <- c(5:7,14:16,23:25,32:34,41:43,50:52,59:61,68:70,77:79)

plot(xcoordM, adjalphaM, cex=1.5, pch=16, ylim=c(0,27), xlim=c(0,80),
cex.axis=1, cex.lab=1, main="", ylab="", xlab="", xaxt="n", xaxs="i")

points(xcoordF, adjalphaF, cex=1.5, pch=21, main="", col="magenta")

abline(v=8.5+(0:7)*9, col="darkgrey")

```

```

abline(h=mean,lty="13")

# ...with 95% confidence intervals

for (i in (1:27)) {

  lines(c(0*(1:2))+xcoordM[i],c(CILBxM[i],CIUBxM[i]))

  lines(c(0*(1:2))+xcoordF[i],c(CILBxF[i],CIUBxF[i]),col="magenta")

}

disn <- c("Injuries", "CVD", "Septicaemia", "Other ID", "DD", "COPD",
"RI", "Cancer", "Other")

age1 <- rep(c(1:3),9)

mtext(age1, at=xcoordM, line=-1.5, side=1, cex=0.6, adj=0.3)

mtext(age1, at=xcoordF, line=-1.5, side=1, cex=0.6, adj=0.3)

axis(1,at=xcoordM,labels=F)

axis(1,at=xcoordF,labels=F)

mtext(disn,at=c(4.5, 13.5, 22.5, 31.5, 40.5, 49.5, 58.5, 67.5, 76.5),
side=1, cex=0.7, adj=0.5)

mtext("Reported Hospital deaths in Southern Thailand 2000 - 2003",
side=3, adj=0.5, cex=0.8, line=2)

mtext("Length of stay (days)",side=3,adj=-0.05,cex=0.7,line=0.3)

mtext("Principal diagnosis group by increasing age and
gender",side=1,adj=0.5,cex=0.8,line=1.5)

legend(13.6,27, c("Male","Female"), col=c("black","magenta"),
pt.cex=1.5, cex=1, lty=c(1,1), pch=c(16,21), inset=0.02, bg="white")

legend(1,27, pch=c("1", "2", "3"), c("< 60 years", "60-74 years",
"75+ years"), pt.cex=1, cex=1, bg="white")

```

```

plot(1,col="white", xaxt="n", yaxt="n", xlab="", ylab="", frame.
plot=F)

# Get coefficients and SEs of medical Geographic effects

xcoord <- c(1:3,5:7,9:11,13:15,17:18,21:23,25:26)

plot(xcoord, adjbeta, cex=1.5, pch=16, ylim=c(0,27), xlim=c(0,28),
cex.axis=1, cex.lab=1, main="", ylab="", xlab="", xaxt="n", xaxs="i")

abline(v=4+(1:5)*4,col="darkgrey")

abline(h=mean,lty="13")

# ...with 95% confidence intervals

for (i in (1:19)) {

  lines(c(0*(1:2))+xcoord[i],c(CILBb[i],CIUBb[i]))

}

xPos <- c(2,6,10,14,18,22,26)
disn <- c("North", "Surat Thani", "North West", "Nakhon", "South
West", "Central", "South East")

bedl <- rep(c("S","M","L"),4)

bedl <- c(bedl,"S","M","S","M","L","S","M")

mtext(bedl,at=xcoord, side=1, cex=0.6,adj=0.5, line=-1.5)

axis(1,at=xcoord,labels=F)

mtext(disn,at=xPos,side=1,cex=0.7,adj=0.5)

mtext("Length of stay (days)",side=3,adj=-0.05,cex=0.7,line=0.3)

mtext("Province group by hospital
size",side=1,adj=0.5,cex=0.8,line=1.3)

```



```

legend(0.5,27,pch=c("S","M","L"), c("< 61 beds","61-499
beds","500+ beds"),pt.cex=1,cex=1,bg="white")

#####

# Calculating the cost

nd <- table(LOSLN$PdxSexAge)

total <- sum(LOSLN$dayslos)

demop <- nd*adjalpha

dp <- demop*100/total

dp1 <- tapply(LOSLN$dayslos,LOSLN$PdxSexAge,sum)*100/total

ng <- table(LOSLN$ProvHosGrp)

geop <- ng*adjbeta

gp <- geop*100/total

gp1 <- tapply(LOSLN$dayslos,LOSLN$ProvHosGrp,sum)*100/total

cbind(nd,adjalpha,dp)

cbind(ng,adjbeta,gp)

```