eAsia Summer School. Time-series regression in Public Health.

Practical 6. Two Stage Design.

In this practical, students will learn how to conduct a multicity analysis using a two-stage design. First, students will get the city-specific exposure-response relationship using time-series regression, and next, they will combine them using univariate and multivariate meta-analysis using R. Finally, we will introduce alternative designs to analyze multicity data to obtain a countrywide exposure-response curve. Students will encounter some given questions. Please try to answer them, which could help you better understand the exercise. Let's get started. Enjoy!

First, install and load the necessary libraries and functions to do this practical.

```
# Remove all previous objects.
rm(list = ls())

# Install packages.
install.packages("dplyr","tsModel","splines","dlnm","Epi","metafor","mixmeta")

# Load packages.
library(dplyr); library(tsModel); library(splines); library(dlnm); library(Epi)
library(metafor); library(mixmeta)

# Load functions.
source("qAIC.R"); source("findmin.R")
```

In this practical, you will analyze a time-series dataset for 10 regions of England and Wales, between 1993-2006. The dataset includes daily mortality counts, mean temperature (ºC), and relative humidity. Heat will be the exposure of interest, defined as the mortality risk between the 99th percentile of the temperature distribution and the minimum mortality temperature (MMT).

```
# Load and inspect the dataset.
data <- read.csv("EngWales.csv")
names(data)
head(data)
tail(data)

# Split the dataset in 10 data frames.
dlist <-split(data, data$regnames)
summary(dlist)
```

1. Two stage design

1.1. Frist-stage

Before start the region-specific analysis, let us generate vectors and matrices to store the risk estimates and the reduced crossbasis parameters to be pooled in the second-stage using univariate and multivariate meta-analysis, respectively.

```
# Vector to store risk estimates for univariate meta-analysis.
regions <- names(dlist)
logRR <- logRRse <- vector("numeric",10)

# Matrix to store reduced crossbasis parameters for multivariate meta-analysis.
coef <- matrix(NA, nrow=length(regions), ncol=3+1,
               dimnames=list(regions, paste0("b",seq(4))))
vcov <- vector("list", length(regions)))
```

Now, run the loop which fits a time-series regression model for each region and displays the region-specific exposure-response curves centered at the MMT.

```
par(mex=0.8, mfrow=c(3,4))

# Loop for region specific analysis.
for(i in seq(regions)){
  cat(i,"")
  sub <- dlist[[i]]
  pct <- quantile(sub$tmean,prob=c(.01,.10,.25,.50,.75,.90,.99),na.rm=T)
  varknot <- pct[c(3:5)]
  cb.temp <- crossbasis(sub$tmean, lag=14,
                        argvar=list(fun="ns", knots=varknot),
                        arglag=list(fun="ns", knots=c(2,5)))
  model <- glm(all ~ cb.temp+ns(time,df=10*14)+dow, sub, family=quasipoisson)
  pred.temp <- crosspred(cb.temp, model, by=1)

  # Predicted exposure-response curve centered at the MMT.
  mmt <- findmin(cb.temp, model)
  pred.heat <- crosspred(cb.temp, model, cen=mmt, by=1)

  # Plot region specific exposure-response curve.
  plot(pred.heat, "overall", ylim=c(0.9,1.8), col=2, lwd=2,
       xlab="Temperature", ylab="RR", main=regions[i])

  # Store risk estimates for univariate meta-analysis.
  target <- as.character(round(pct[7]))
  logRR[i]   <- pred.heat$allfit[target]
  logRRse[i] <- pred.heat$allse[target]

  # Store reduced crossbasis parameters for multivariate meta-analysis.
  cr <- crossreduce(cb.temp, model)
  coef[i,]  <- coef(cr)
  vcov[[i]] <- vcov(cr)
}

layout(1)
```

Check the region-specific risk estimates for heat.

```
# List risk estimates by region for univariate meta-analysis.
cbind(logRR, logRRse)
```

Questions. Do you think that the exposure-response curves between regions are similar or show different shapes? Are the estimated heat effects similar between regions?

1.2. Second-stage using univariate meta-analysis

Let us combine the region-specific risk estimates for heat using a univariate random-effects meta-analysis.

```
# Random effects meta-analysis.
uni <- rma(y=logRR, sei=logRRse, slab=regions, measure="RR")
summary(uni)
ci.exp(uni)

# Forest plot.
forest(uni, transf=exp, refline=1, pch=23, bg=4, col=2,
       main="Heat effect")
```

Questions. Do you consider the estimated heat effects between regions to be homogeneous? How much heterogeneous is the data? How do you interpret the combined estimator for the effect of heat?

However, you can attempt to explain the heterogeneity in the meta-analysis using variables at the regional level. For example, let us use the regions' latitude as a meta-predictor in a random effects meta-regression model.

```
# Generate latitude data.
lat <- c(54.84815, 53.58832, 53.72352, 52.85539, 52.53304, 52.03734, 51.50583,
         51.24213, 51.05361, 52.02615)

# Meta-regression by latitude.
res <- rma(y=logRR, sei=logRRse, mods=lat)
summary(res)

# Bubble plot.
preds <- predict(res, newmods = cbind(51:55), transf = exp)
wi <- 1/sqrt(logRRse)
size <- 0.5 + 3 * (wi - min(wi))/(max(wi) - min(wi))
plot(lat, exp(logRR), xlim=c(51,55), ylim=c(0.95,1.2), pch=19,
     cex=size, xlab="Latitude", ylab="Relative Risk", las=1, bty="l", log="y")
lines(51:55, preds$pred)
lines(51:55, preds$ci.lb, lty="dashed")
```

```
lines(51:55, preds$ci.ub, lty="dashed")
abline(h=1, lty = "dotted")
```

Questions. Does the latitude explain the heterogeneity among the effects of heat? How do you interpret the slope of the meta-regression model?

1.3. Second-stage with multivariate meta-analysis

Instead of pooling the region-specific risk estimates for heat derived from, you can pool the region-specific exposure-response curves using a random effects multivariate model.

```
# Multivariate mixed-effects meta-analysis.
mv <- mixmeta(coef~1, vcov, method="ml")
summary(mv)

# Generate the temperature distribution for prediction.
bound <- rowMeans(sapply(dlist, function(x) range(x$tmean)))
xvar <- seq(bound[1], bound[2], by=0.1)

# Predicted exposure-response curve from meta-analysis estimates.
argvar=list(fun="ns", knots=quantile(xvar, prob=c(.25,.50,.75)))
bvar <- do.call(onebasis, c(list(x=xvar), argvar))
pred.pool <- crosspred(bvar, coef=coef(mv), vcov=vcov(mv),
                       model.link="log", by=0.1)

# Center the exposure-response curve at the MMT.
mmt <- pred.pool$predvar[which.min(pred.pool$allRRfit)]
pred.pool <- crosspred(bvar, coef=coef(mv), vcov=vcov(mv),
                       model.link="log", by=0.1, cen=mmt)

# Plot pooled exposure-response curve.
plot(pred.pool, type="l", ci="n", ylab="RR", ylim=c(.95,1.3), lwd=2,
     xlab="Temperature (C)", main="Pooled and first-stage")

# Get risk estimate for heat exposure.
xvar.heat <- quantile(xvar, prob=c(.99))
target <- as.character(round(xvar.heat))
cbind(pred.pool$allRRfit[target], pred.pool$allRRlow[target],
pred.pool$allRRhigh[target])
```

Questions. How much heterogeneous are the exposure-response curves between regions? Is the heat effect derived from the pooled exposure-response curve similar to that obtained in the univariate meta-analysis?

Moreover, you improve the visualization by plotting the pooled exposure-response curve overlapping the region-specific response-curves.
```
# Predicted region-specific curve from meta-analysis estimates.
pred.reg <- lapply(seq(regions),
```

```
                    function(i) crosspred(bvar, coef=coef[i,], vcov=vcov[[i]],
                                          model.link="log", cen=mmt))

# Plot pooled exposure-response curve.
for(i in seq(regions)) lines(pred.reg[[i]], col="grey")
lines(pred.pool, lwd=3)
```

Finally, let us explore heterogeneity calculating the redicted pooled exposure-response curve by latitude fitting a multivariate meta-regression model.

```
# Multivariate meta-regression by latitude.
mixlat <- mixmeta(coef~lat, vcov, method="ml")
print(summary(mixlat), digits=3)

# Predicted pooled exposure-response curve by latitude.
predlat <- predict(mixlat, data.frame(lat=range(lat)), vcov=T)
predmin <- crosspred(bvar, coef=predlat[[1]]$fit, vcov=predlat[[1]]$vcov,
                     model.link="log", cen=mmt)
predmax <- crosspred(bvar, coef=predlat[[2]]$fit, vcov=predlat[[2]]$vcov,
                     model.link="log", cen=mmt)

# Plot pooled exposure-response curve by latitude.
plot(predmax, type="l", ci.arg=list(density=50,col=4), col=4, lwd=3,
     ylab="RR", ylim=c(.95,1.3),
     xlab="Temperature (C)",
     main="Effect modification by latitude")
lines(predmin, ci="area", ci.arg=list(density=50,col=2), col=2, lwd=3)
legend("top", c("High (north)","Low (south)"),
       lty=1, col=c(4,2), inset=0.05,
       title="Latitude")
```

Questions. Does the latitude explain the heterogeneity among the effects of heat? Can you think of other meta-predictor variables that could be useful to explain the heterogeneity of the heat effects between regions?

2. Pooled design

Now let us use a pooled design to estimate the temperature-mortality association in England and Wales, and compare it with the exposure-response curve previously estimated using the two-stage design.

```
# Load and inspect the dataset.
data <- read.csv("EngWales.csv")
names(data)
head(data)

# Define crossbasis for temperature.
pct <- quantile(data$tmean, prob=c(.01,.10,.25,.50,.75,.90,.99), na.rm=T)
varknot <- pct[c(3:5)]
cb.temp <- crossbasis(data$tmean, lag=21,
```

```
                    argvar=list(fun="ns", knots=varknot),
                    arglag=list(fun="ns", knots=c(2,5)))

# Fit regression model.
model <- glm(all ~ cb.temp + ns(time, df = 10*14) + dow + factor(region), data,
family=quasipoisson)
pred <- crosspred(cb.temp, model, by=1)

# Center the exposure-response curve at the MMT.
mmt <- findmin(cb.temp, model)
pred.pooled <- crosspred(cb.temp, model, cen=mmt, by=1)

# Figure for exposure-response curve.
par(mex=0.8,mfrow=c(1,2))

# Plot pooled design exposure-response curve.
plot(pred.pooled, "overall", col=2, lwd=2,
     ylim=c(0.9,1.6), xlim=c(-10,35),
     xlab="Temperature", ylab="RR", main="Pooled design")

# Plot two-stage design exposure-response curve.
plot(pred.pool, "overall", col=2, lwd=2,
     ylim=c(0.9,1.6), xlim=c(-10,35),
     xlab="Temperature", ylab="RR", main="Two-stage design")

layout(1)
```

Questions. Is the exposure-response curve of the pooled design similar to that obtained with the two-stage design? If there is a difference, why could it be? In which situations could the pooled design be used instead of the two-stage design?

3. Summarized design

Finally, try a simple summarized design to estimate the temperature-mortality association, and compare it with the exposure-response curve previously estimated using the two-stage design.

```
# Define crossbasis for temperature.
pct <- quantile(data$tmean, prob=c(.01,.10,.25,.50,.75,.90,.99),na.rm=T)
varknot <- pct[c(3:5)]
cb.temp <- crossbasis(data$tmean, lag=14,
                    argvar=list(fun="ns", knots=varknot),
                    arglag=list(fun="ns", knots=c(2,5)))

# Fit regression model.
model <- glm(all ~ cb.temp + ns(time, df = 10*14) + dow , data, family=quasipois
son)
pred <- crosspred(cb.temp, model, by=1)

# Center the exposure-response curve at the MMT.
mmt <- findmin(cb.temp, model)
```

```
pred.summ <- crosspred(cb.temp, model, cen=mmt, by=1)

par(mex=0.8,mfrow=c(1,2))

# Plot summarized design exposure-response curve.
plot(pred.summ, "overall", col=2, lwd=2,
     ylim=c(0.9,1.6), xlim=c(-10,35),
     xlab="Temperature", ylab="RR", main="Summarized design")

# Plot two-stage design exposure-response curve.
plot(pred.pool, "overall", col=2, lwd=2,
     ylim=c(0.9,1.6), xlim=c(-10,35),
     xlab="Temperature", ylab="RR", main="Two-stage design")

layout(1)
```

Questions. Is the exposure-response curve of the summarized design similar to that obtained with the two-stage design? If there is a difference, why could it be? In which situations could the summarized design be used instead of the two-stage design?