Supplementary information for the paper

Title: Determinants of Antibiotic Consumption - Development of a Model using Partial Least Squares Regression based on Data from India

Authors: Ashok J. Tamhankar, Shreyasee S. Karnik, Cecilia Stålsby Lundborg

The general underlying model of PLS regression

$X = TP^T + E$ Equation S1

$Y = UQ^T + F$

where,

$\boldsymbol{X}$ = *n (rows) x m (columns)* matrix of determinants; $\boldsymbol{Y}$ = *n (rows) x p (columns)* matrix of responses (antibiotic consumption); $\boldsymbol{T}$ and $\boldsymbol{U}$ = *n (rows) x l(columns)*matrices that are, respectively, projections of $\boldsymbol{X}$ (the X score, component or factor matrix) and projections of $\boldsymbol{Y}$ (the Y scores); $\boldsymbol{P}$ and $\boldsymbol{Q}$ = *m (rows) x l (columns)* and *p (rows) x l (columns)* orthogonal loading matrices; matrices $\boldsymbol{E}$ and $\boldsymbol{F}$ are the error terms, assumed to be independent and identically distributed random normal variables. The decompositions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are made so as to maximise the covariance between $\boldsymbol{T}$ and $\boldsymbol{U}$.

R code for building the prediction model to estimate human antibiotic consumption in India

Note: Comments to explain each line of code are in italics starting with "#"

```
# Import the data for model building
ind.subset = read.csv(file.choose(), header=T)

# 75% of the sample size
set.seed(896)
smp_size<- floor(0.75 * nrow(ind.subset))

# set the seed to make your partition reproducible
set.seed(457)
train_ind<- sample(seq_len(nrow(ind.subset)), size = smp_size)

# Create training set
train.subset<- ind.subset[train_ind, ]

# Create test set
test.subset<- ind.subset[-train_ind, ]

# Call the library pls to run PLSR
library(pls)

# Regression command
mod.subset<- plsr(ab_use~., ncomp=6, data = train.subset, validation = "LOO")
```

```
summary(mod.subset)


# Command to plot the RMSEP graph

plot(RMSEP(mod.subset))


# 2 components plot with loading values

par(mar = c(10,4,2,2), mgp = c(7, 1, 0))

plot(mod.subset, plottype="loadings", comps = 2, ylim=c(-1,1.5), xlab="Determinants", ylab="", xaxt="n")

axis(1, at=1:38, labels=prednames(mod.subset), las=3)

title(ylab = "loading value", mgp = c(2.5, 1, 0))

abline(h = 0)


# Predict the antibiotic consumption values for test set

predict(mod.subset, ncomp = 2, newdata = test.subset)


# Command to plot the predicted vs. measured value of the test data

par(mar = c(10,4,2,2), mgp = c(4, 1, 0))

predplot(mod.subset, ncomp=2, newdata = test.subset, asp=1, line=TRUE, xlab="Measured value")

title(ylab = "Predicted value", mgp = c(2.5, 1, 0))
```