Recalculation of leverages based on formula 5.21 from Hosmer et al 2013

Johann Popp

1 November 2017

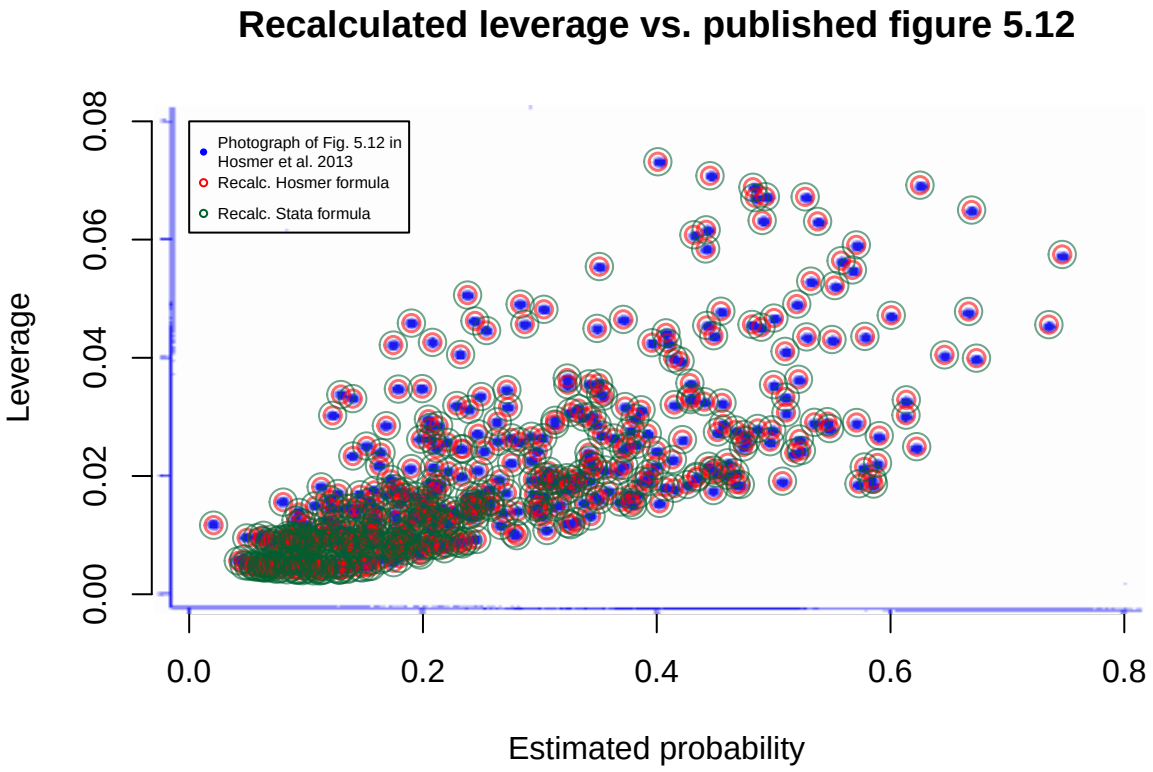
This code recalculates leverages based on formula 5.12 from Hosmer et al 2013[1] to reproduce figure 5.12 from the same source. It also calcultates leveraged based given the formula of Stata[2] which was used to draw the graphs for Hosmer et al 2013.

```
# Example data
glow <- aplore3::glow500

# Recode RATERISK
glow$raterisk3 <- cut(as.numeric(glow$raterisk), 2)
levels(glow$raterisk3) <- c("less/same", "greater")


# Logistic model from Table 4.16
model <- glm(fracture ~ age + height + priorfrac + momfrac +
               armassist + raterisk3 + age:priorfrac + momfrac:armassist,
             data = glow, family = "binomial")


########################

# Extract non-aggregated data
datN <- model$model

# convert to covariate pattern
cp <- unique(datN[,-1])

# Indicator for covariate patterns
datN$cpID <- apply(datN[,-1], 1, function(x) paste(x, collapse = ""))
cp$cpID <- apply(cp, 1, function(x) paste(x, collapse = ""))

# Size of covariate patterns
cp$m <- tapply(datN[,1], datN$cpID, length)[rank(cp$cpID)]

# Number of cases per covariate pattern
cp$Y <- tapply(model$y, datN$cpID, sum)[rank(cp$cpID)]

# Estimated probability
cp$P <- fitted.values(model)[rownames(cp)]
```

[1] Hosmer, David W., Stanley Lemeshow, and Rodney X. Sturdivant. Applied Logistic Regression. 3rd ed. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley, 2013.

[2] "Logistic Postestimation - Postestimation Tools for Logistic." In Base Reference Manual, Release 15. College Station, Texas: Stata Press, 2017. https://www.stata.com/bookstore/base-reference-manual/.

```r
########################################
# Leverage

# Diagonal of variances
V <- diag(cp$m * cp$P * (1 - cp$P))

# Design matrix
X <- unique(model.matrix(model)) # This makes the difference to epiR::epi.cpresids

H <- V^(1/2) %*% X %*% solve(t(X) %*% V %*% X) %*% t(X) %*% (V^(1/2))
h <- diag(H)



####################################
# Leverage based on Stata formula

# Extract variance-covariance matrix from the model
V2 <- vcov(model)
# Calculate the raw hat values
rawH <- diag(X %*% V2 %*% t(X))
# Adjust for covariate pattern
h2 <- rawH * cp$m * cp$P * (1-cp$P)



#####################
# Plot
library(png)
img512 <- readPNG("HLS5-12new.png")
plot(1:2, xlim = c(0, 0.8), ylim = c(0, 0.08), bty = "n",
     xlab = "Estimated probability", ylab = "Leverage",
     main = "Recalculated leverage vs. published figure 5.12")
rasterImage(img512, xleft = -0.03, ybottom = -0.0045, xright = 0.817,ytop = 0.0827)
points(cp$P, h,
       col = "#FF000090", pch = 1, cex = 1.2, lwd = 1.5)
points(cp$P, h2,
       col = "#00603090", pch = 1, cex = 1.9, lwd = 1)
legend(0, 0.08, pch = c(16, 1, 1),
       legend = c("Photograph of Fig. 5.12 in\nHosmer et al. 2013",
                  "Recalc. Hosmer formula", "Recalc. Stata formula"),
       col = c("blue", "red", "#006030"), cex = 0.5)
```



This fits perfectly with the original graph. Leverages based on formula 5.12 from Hosmer et al (wich is taken from Pregibon 1981[3]) are exactly the same as those based on the Stata formula.

[3] "Logistic Postestimation - Postestimation Tools for Logistic." In Base Reference Manual, Release 15. College Station, Texas: Stata Press, 2017. https://www.stata.com/bookstore/base-reference-manual/.