

Automating model fits in R

JoAnn Alvarez

May 6, 2013

Suppose you have the following data:

```
n <- 246
x1 <- sample(0:5, size = n, replace = TRUE)
x2 <- rgamma(shape = 1, scale = 2, n = n)
y1 <- -3 + 2*x1 + 4*x2 + rnorm(n)
y2 <- 0 - 2*x1 - 2*x2 + rnorm(n)
y3 <- 1 + 1*x1 + 1*x2 + rnorm(n, sd = 2)
```

You have several response variables that have the mean model structure. You could fit three separate models by copying and pasting code.

What if you instead have 78 response variables? And then you decide to add another covariate?

Automate the model fits so that there is only one place to edit the mean model.

Pre-allocate the length of the object using a list:

```
outcomes <- c("y1", "y2", "y3")
mods <- vector(length = length(outcomes), mode = "list")
names(mods) <- outcomes
mods

$y1
NULL

$y2
NULL

$y3
NULL
```

Use `as.formula()` with `paste()` to automate many model fits:

```
for(i in seq_along(outcomes)){
  frmla <- as.formula(paste(outcomes[i], " ~ x1 + x2"))
  mods[[i]] <- lm(frmla)
}
```

```
mods[[2]]
```

```
Call:
```

```
lm(formula = frmla)
```

```
Coefficients:
```

```
(Intercept)      x1      x2
      0.117    -2.041    -1.991
```

Write a function to extract elements from the model fits:

```
latexmodeloutput <- function(mod, i, modlabels){
  jill <- data.frame(summary(mod[[i]])$coefficients[ , c("Estimate", "Pr(>|t|)"), confint(mod[[i]])
  jill$pval <- format.pval(jill$Pr...t..., eps = 0.001, digits = 4)
  latex(jill[ , c("Estimate", "X2.5..", "X97.5..", "pval")], digits = 3,
        colheads = c("$\\hat{\\beta}$", "Lower 95\\%", "Upper 95\\%", "$p$-value"),
        booktabs = TRUE,
        title = "",
        where = "!h",
        caption = paste("Estimated regression parameters from ", modlabels[i], " model", sep = ""),
        file = "")}
```

Use lapply to apply it over the list.

```
lapply(X = seq_along(mods), FUN = latexmodeloutput, mod = mods, modlabels = outcomes)
```

Table 1: Estimated regression parameters from y1 model

	$\hat{\beta}$	Lower 95%	Upper 95%	p -value
(Intercept)	-3.05	-3.30	-2.80	<0.001
x1	2.01	1.94	2.08	<0.001
x2	4.01	3.94	4.08	<0.001

Table 2: Estimated regression parameters from y2 model

	$\hat{\beta}$	Lower 95%	Upper 95%	p -value
(Intercept)	0.117	-0.152	0.386	0.393
x1	-2.041	-2.117	-1.965	<0.001
x2	-1.991	-2.066	-1.917	<0.001

```
[[1]]
[[2]]
[[3]]
```

Table 3: Estimated regression parameters from y3 model

	$\hat{\beta}$	Lower 95%	Upper 95%	<i>p</i> -value
(Intercept)	1.380	0.857	1.904	<0.001
x1	0.827	0.679	0.975	<0.001
x2	1.052	0.907	1.196	<0.001