Using R for Analytic Graphs: Learn How Data Visualization Can Improve Interpretation in Social Work Research

Saturday, January 18, 2015

Why Use R?

- Free
- Open Source
- Easy Collaboration
- Replicable Research

Why Wouldn't You Use R?

Steep(er) learning curve compared to, say, Excel or SPSS. This matters a lot if

- You run statistics rarely.
- You want a point and click interface.

Where Can you Get R?

- CRAN
- Our Thumb Drives

Where Are We Going Today?

- Graphing Model Results
- Graphing Other Things

Graphing Model Results (Basic Algorithm)

1. Choose a counterfactual x_c.
2. Estimate model parameters $\hat{\beta}$ and the variance-covariance matrix, $\hat{\boldsymbol{V}}$.
3. Draw several $\tilde{\beta}$ from $\mathcal{N}(\hat{\beta}, \hat{\boldsymbol{V}})$, where $\mathcal{N}$ is a mulivariate normal distribution.
4. Calculate expected outcomes based on model parameters for all draws from $\mathcal{N}$.
5. Calculate summary statistics for each level of x_c.

This approach will work for most models that social welfare researchers use.

A Practical Example - Background

- Research Question: How does a child's probability of exiting foster care vary by child characteristics?
- Multiple Permanency Outcomes: Requires that we estimate a mulinomial logistic regression model.
- Data: Children ($n = 619$) entering out-of-home care in late 2007. Children's parents' surveyed in 2007 and linked to administrative data to faciliate follow-up.
- NOTE: These data are simulated because to protect client confidentiality.

A practical example - Choose a counterfactual x_c.

Getting Data Into R

- R can import and export most formats: read.dta(), read.spss(), read.csv(), read.dbf()
- Use the foriegn() library to import SPSS, Stata, and DBase formats.

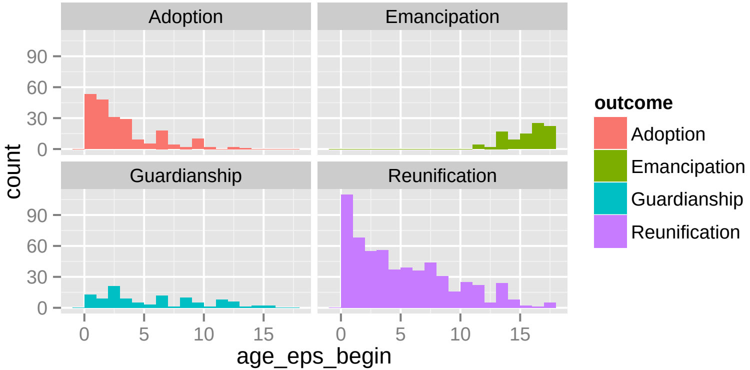
```
dat <- read.csv("dat.csv")
```

-To export use the write.csv() command. # A practical example - Choose a counterfactual x_c.

```
#looking at age of child at episode begin
require(ggplot2)
ggplot(dat, aes(x=age_eps_begin)) +
  geom_histogram(binwidth = 1)
```

A practical example - Choose a counterfactual x_c.

```
#looking at age of child at episode begin by outcome
ggplot(dat, aes(x=age_eps_begin, fill=outcome)) +
  geom_histogram(binwidth = 1) +
  facet_wrap(~ outcome)
```



A practical example - Estimate a model.

Need to estimate a statistical model to get

1. A vector of parameters $\hat{\beta}$, and
2. The associated variance-covariance matrix, $\hat{\boldsymbol{V}}$.

A practical example - Estimate a model.

Prep the data

```
# easy to load external packages
# install.packages("nnet") # install once
require(nnet)                # load every time

# recode data
levels(dat$outcome) <- c("Adoption", "Emancipation"
                         ,"Guardianship", "Reunification")

# relevel our outcome variable
dat$outcome_rl <- relevel(dat$outcome
                          , ref = "Emancipation")

# recode to numeric
dat$outcome_rl <- as.numeric(dat$outcome_rl)
```

A practical example - Estimate a model.

Run the model

```
# run the multinomial model
model <- multinom(outcome_rl ~ age_eps_begin +
                   eps_rank
                 ,data = dat
                 ,Hess = TRUE)
```

```
## # weights:  16 (9 variable)
## initial  value 1386.294361
## iter  10 value 931.103300
## iter  20 value 860.375750
## final  value 860.374425
## converged
```

A practical example - Estimate a model.

Display of summary the model

```
model
```

```
## Call:
## multinom(formula = outcome_rl ~ age_eps_begin + eps_rank
##     Hess = TRUE)
##
## Coefficients:
##   (Intercept) age_eps_begin    eps_rank
## 2   11.457365    -1.0280750 -0.10995325
## 3    9.797665    -0.8393067  0.05195097
## 4   11.597181    -0.8691345  0.07149574
##
## Residual Deviance: 1720.749
## AIC: 1738.749
```

A practical example - Estimate a model.

Extract a vector of parameters $\hat{\beta}$

```
#run the multinomial model
pe <- model$wts[c(6,7,8,10,11,12,14,15,16)]
pe[1:3]
```

```
## [1] 11.4573653 -1.0280750 -0.1099532
```

```
pe[4:6]
```

```
## [1]  9.79766546 -0.83930667  0.05195097
```

```
pe[7:9]
```

```
## [1] 11.59718150 -0.86913446  0.07149574
```

A practical example - Estimate a model.

Extract the associated variance-covariance matrix, $\hat{\boldsymbol{V}}$

```
#run the multinomial model
vc <- solve(model$Hess)
```

A practical example - Draw several $\tilde{\beta}$ from $\mathcal{N}(\hat{\beta},\ \hat{\boldsymbol{V}})$.

```r
#load a package which contains a multivariate normal
#sampling function
require(MASS)
#assign a variable for the number of simulations
sims <- 10000
#draw the indicates number of beta simulates
#using our extracted model data
simbetas <- mvrnorm(sims,pe,vc)
```

A practical example - Last two steps...

- Calculate expected values for all of your draws from $\mathcal{N}$, and
- Calculate summary statistics for each level of x_c.
- Specific calculations are beyond the scope of this presentation
- But the `simcf` package from Chris Adolph (political scientist at the University of Washington) will do them for us!

A practical example - Last two steps

Get data read for `simcf`

- Re-arrange simulates to array format

```
simb <- array(NA, dim = c(sims,3,3))
simb[,,1] <- simbetas[,1:3]
simb[,,2] <- simbetas[,4:6]
simb[,,3] <- simbetas[,7:9]
```

- Specify range of counterfactual values

```
agerange <- seq(0,17,by=0.1)
```

A practical example - Last two steps

Get data read for simcf

- Load simcf and use the cfFactorial() function to set specific values for simulation.

```
require(simcf)
xhyp <- cfFactorial(age = agerange
                    ,ep_rank = mean(dat$eps_rank))
```

- Run the simulation (this is where the last two steps are really performed).

```
test_sims <- mlogitsimev(xhyp,simb,ci=0.95)
```

Get the data ready to graph

```
y <- as.vector(test_sims$pe[,1:4])

x <- rep(1:length(agerange), 4)

lower <- as.vector(test_sims$lower[,1:4,])

upper <- as.vector(test_sims$upper[,1:4,])

Outcome <- c(rep("Adoption", length(agerange))
             ,rep("Guardianship"
                  ,length(agerange))
             ,rep("Reunification"
                  ,length(agerange))
             ,rep("Emancipation"
                  ,length(agerange)))
```

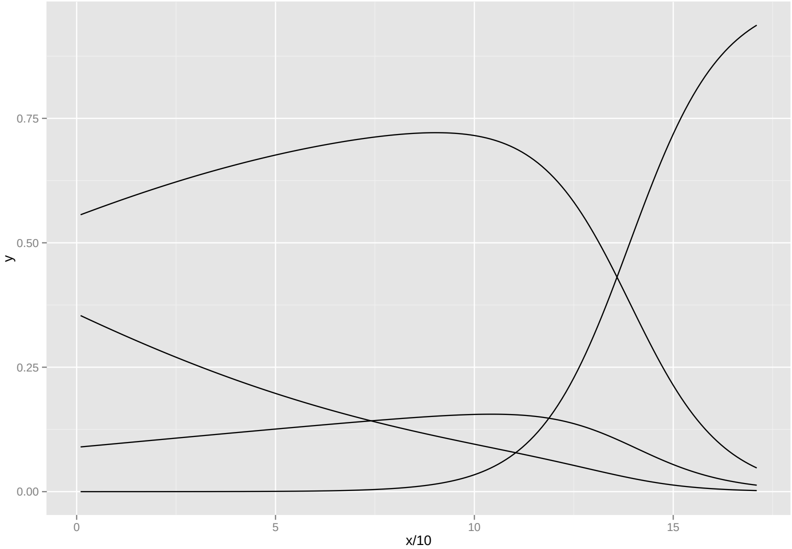
Get the data ready to graph

```
dat_sim_plot <- data.frame(y,x,lower,upper,Outcome)
```

Graph the data!

```
p1 <- ggplot(dat_sim_plot
        ,aes(x=x/10, y=y, group=Outcome)) +
         geom_line()
```

Graph the data!



Make it Pretty!

```
p2 <- ggplot(dat_sim_plot
       ,aes(x=x/10, y=y, group=Outcome)) +
        geom_line(size=1, alpha=.5) +
        geom_ribbon(aes(ymin=lower
                         ,ymax=upper
                         ,fill=Outcome), alpha=.5) +
        ylab("Pr(Outcome|Age,Prior Episodes)") +
        xlab("Age at Entry into Foster Care") +
        theme_bw()
```

Make it Pretty