Random effects ANOVA
Created by: AJ Richards
Last updated: August 7, 2014

Contents

1 Problem and data

1.1 About

This is an example from Marc Kery's book "Introduction to WinBUGS for ecologists [1]. Here we explore a random effects model with ANOVA (a t-test applied more than two groups). First we generate the data—five populations of snakes each with $n = 10$ and a single measured co-variate. See the function `generateData.R`.

```
#!/usr/bin/Rscript

### generate data for fixed effects ###
ngroups <- 5
nsample <- 10
popMeans <- c(50,40,45,55,60)
sigma <- 3

n <- ngroups * nsample
resid <- rnorm(n,0,sigma)
means <- rep(popMeans, rep(nsample,ngroups))
x <- rep(1:5,rep(nsample,ngroups))

## create design matrix and target variables
X <- as.matrix(model.matrix(~as.factor(x)-1))
y <- as.numeric(X%*% as.matrix(popMeans) + resid)

## plot the data
pdf("fe-svl-data.pdf",height=6,width=6)
boxplot(y~x,col="grey",xlab="population",ylab="SVL",main="",las=1)
dev.off()

## write to outfile
xyFrame <- data.frame(x,y)
write.csv(xyFrame,file="fe-svl.csv",row.names=FALSE)

### generate data for random effects ###
npop <- 10
nsample <- 12
n <- npop * nsample

popGrandMean <- 50
popSd <- 5
popMeans <- rnorm(n=npop,mean=popGrandMean,sd=popSd)
sigma <- 3
resid <- rnorm(n,0,sigma)
x <- rep(1:npop, rep(nsample, npop))

## create the design matrix and target variables
X <- as.matrix(model.matrix(~ as.factor(x)-1))
y <- as.numeric(X %*% as.matrix(popMeans)+resid)

# Plot of generated data
pdf("re-svl-data.pdf",height=6,width=6)
boxplot(y~x, col="grey", xlab="population", ylab="SVL",main="", las=1)
dev.off()

## write to outfile
xyFrame <- data.frame(x,y)
write.csv(xyFrame,file="re-svl.csv",row.names=FALSE)
print("done.")
```

Figure 1.1: The distributions of the five populations of snakes with respect to snout-vent-length

1.2 Fixed and random effects ANOVA

The means parameterization for the one-way ANOVA:

$$y_i = \alpha_{j(i)} + \epsilon_i \tag{1.1}$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{1.2}$$

$$\alpha_{j(i)} \sim \mathcal{N}(\mu, \tau^2) \tag{1.3}$$

y_i refers to the `svl` of snake i in population j. Eqn 1.3 is the key assumption that moves this from a fixed-effects to a random effects ANOVA. It is not always clear whether we should be using a fixed-effect or random-effects ANOVA, in fact statisticians have differing opinions. Random effects are often used for things like `year`, `month`, or `location`. The decision has to do with whether we want to generalize our conclusions to the larger (unsampled) population as well.

2 Fixed effects ANOVA

2.1 Maximum likelihood analysis

```
data = read.csv("fe-svl.csv")
print(anova(lm(data$y~as.factor(data$x))))
cat("\n")
print(summary(lm(data$y~as.factor(data$x)))$coeff,dig=3)
cat("Sigma:", summary(lm(data$y~as.factor(data$x)))$sigma, "\n")
```

```
[1] "..."
```

2.2 MCMC analysis

Figure 2.1: The MCMC chains

Figure 2.2: MCMC densities

3 Random effects ANOVA

Figure 3.1: The distributions of the 10 populations of snakes with respect to snout-vent-length

Bibliography

[1] M. Kery. *Introduction to WinBUGS for Ecologists*, Elsevier Academic Press, 2010.