**Random effects ANOVA**
Created by: AJ Richards
Last updated: August 6, 2014

# Contents

# 1 Problem and data

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

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

## plot the data
pdf("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="svl.csv",row.names=FALSE)
print("done.")
```
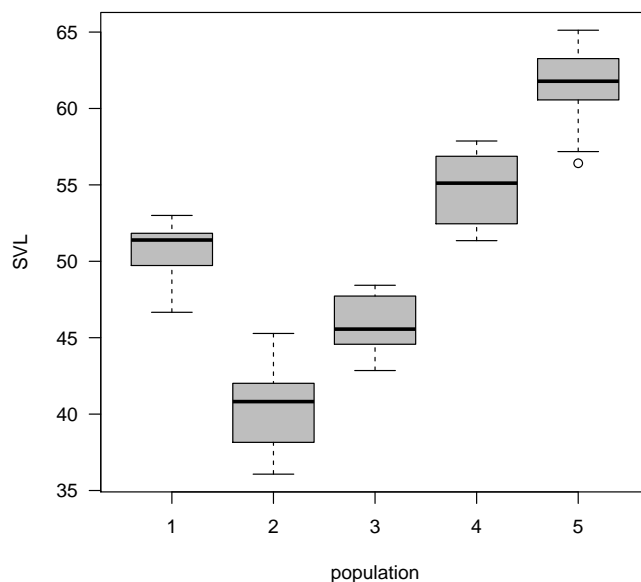


Figure 1.1: The distributions of the five populations of snakes with respect to snout-vent-length

## 1.1 Fixed 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 Analyses

## 2.1 Maximum likelihood analysis

```
data = read.csv("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] "..."
Analysis of Variance Table

Response: data$y
                  Df Sum Sq Mean Sq F value Pr(>F)
as.factor(data$x)  4 2640.49 660.12 112.32 < 2.2e-16 ***
Residuals         45 264.48 5.88
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1


                  Estimate Std. Error t value Pr(>|t|)
(Intercept) 50.54 0.767 65.92 2.07e-46
as.factor(data$x)2 -10.32 1.084 -9.51 2.42e-12
as.factor(data$x)3 -4.83 1.084 -4.45 5.54e-05
as.factor(data$x)4 4.29 1.084 3.95 2.70e-04
as.factor(data$x)5 10.77 1.084 9.93 6.42e-13
Sigma: 2.424334
```

## 2.2 MCMC analysis

```
Compiling model graph
    Resolving undeclared variables
    Allocating nodes
    Graph Size: 122

Initializing model

null device
          1
null device
          1
$BUGSoutput
Inference for Bugs model at "fe-anova.txt", fit using jags,
 3 chains, each with 5000 iterations (first 200 discarded), n.thin = 2
 n.sims = 7200 iterations saved
         mean  sd  2.5%   25%   50%   75% 97.5% Rhat n.eff
alpha[1] 50.5 0.8  48.9  50.0  50.5  51.0  52.0    1  7200
alpha[2] 40.2 0.8  38.7  39.7  40.2  40.7  41.7    1  3900
alpha[3] 45.7 0.8  44.1  45.1  45.7  46.2  47.2    1  7200
alpha[4] 54.8 0.8  53.2  54.3  54.8  55.3  56.4    1  7200
alpha[5] 61.3 0.8  59.7  60.7  61.3  61.8  62.8    1  7200
deviance 231.7 3.7 226.5 229.0 231.1 233.7 240.7  1  7200
effe2   -10.3 1.1 -12.5 -11.1 -10.3  -9.6  -8.1    1  2700
effe3    -4.8 1.1  -7.0  -5.6  -4.9  -4.1  -2.6    1  7200
effe4     4.3 1.1   2.0   3.5   4.3   5.0   6.5    1  7200
effe5    10.7 1.1   8.6  10.0  10.7  11.5  13.0    1  7200
sigma     2.5 0.3   2.0   2.3   2.5   2.7   3.1    1  7200
test1   -30.2 1.6 -33.4 -31.3 -30.2 -29.1 -27.0   1  6800
test2     2.2 2.0  -1.7   0.9   2.2   3.5   6.1    1  7200

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = var(deviance)/2)
pD = 6.9 and DIC = 238.7
DIC is an estimate of expected predictive error (lower deviance is better).
```
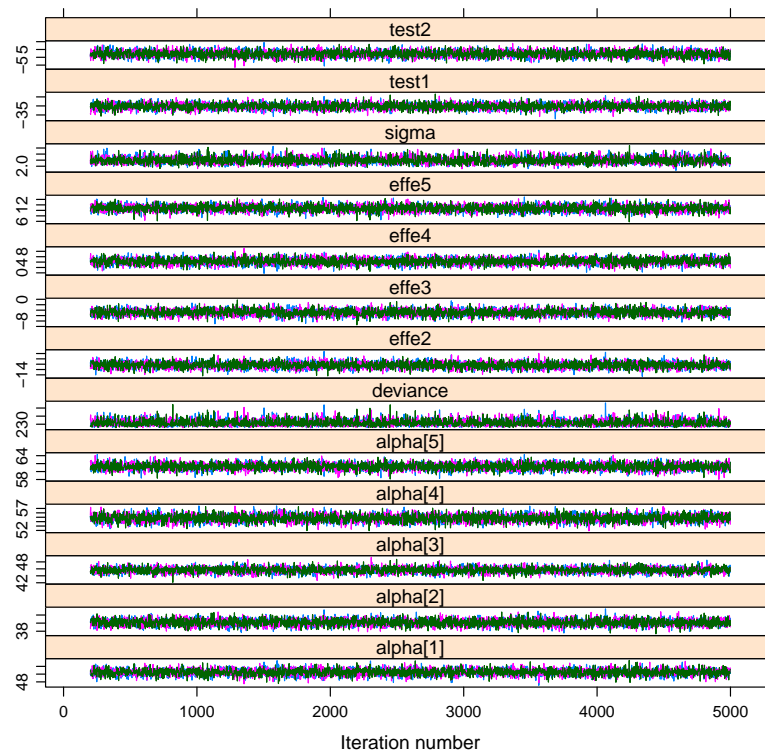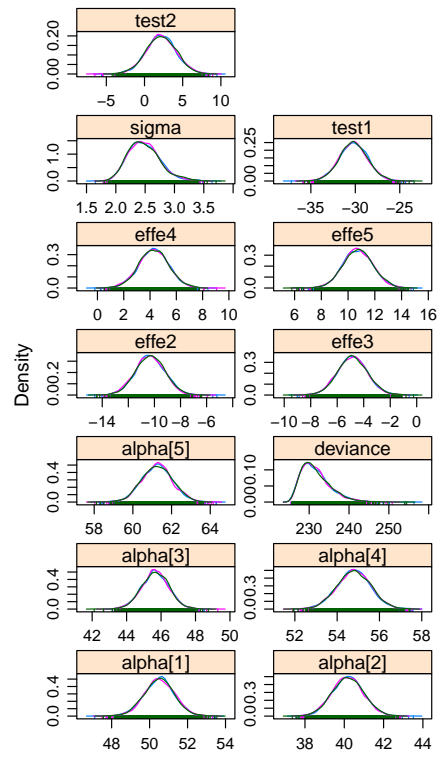
Figure 2.1: The MCMC chains

Figure 2.2: MCMC densities

# Bibliography

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