MixedModR2

Erika Mudrak

Thursday, January 08, 2015

Generate the Data

Generate 1000 data points from a population with one random effect:

20 levels of Factor A, each sampled 50 times

```
set.seed(39)

siga <- 50
sige <- 30
beta0 <- 100
beta1 <- 3
beta2 <- 1.2
Alevel <- rep(paste("A", 1:20, sep=""), each=50)
effA <- rep(rnorm(20, 0, siga), each=50)   #level a

x1 <- rnorm(1000, 0, 8) #obs level
eps <- rnorm(1000, 0, sige)

y <- beta0 + beta1*x1 + effA + eps
```

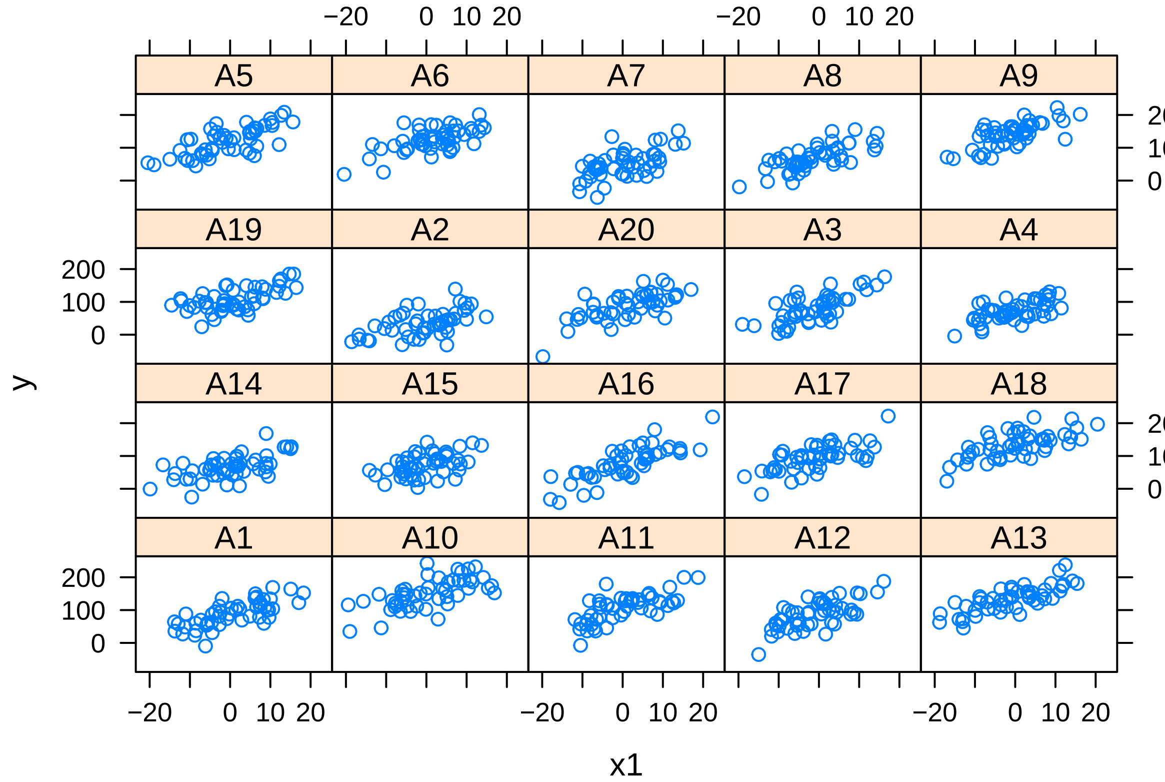
This matches the equation

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \alpha_j + \varepsilon_{ij}$$
$$\alpha_j + Gaussian(0, \sigma_\alpha^2)$$
$$\varepsilon_{ij} + Gaussian(0, \sigma_\varepsilon^2)$$

Where β_0 =`beta0`

```
dat <- data.frame(Alevel,effA, x1, y)
write.csv(dat, file="R2trial.csv")


library(lattice)
xyplot(y~x1|Alevel, data=dat)
```



Fit the models

Fit a model with x1 as the fixed effect (mod1) and a null model (mod0)

Model 1

$$
\begin{aligned}
y_{ij} = \beta_0 + \beta_1 x_{1ij} + \alpha_j + \varepsilon_{ij} \\
\alpha_j + Gaussian(0, \sigma^2_\alpha) \\
\varepsilon ij + Gaussian(0, \sigma^2_\varepsilon)
\end{aligned}
$$

```
library(lme4)

mod1 <- lmer(y ~ x1 + (1|Alevel), data=dat)
summary(mod1)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ x1 + (1 | Alevel)
   Data: dat

REML criterion at convergence: 9682.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
```

```
-2.9129 -0.6551  0.0473  0.6248  3.1789

Random effects:
 Groups   Name        Variance Std.Dev.
 Alevel   (Intercept) 956.5    30.93
 Residual             870.8    29.51
Number of obs: 1000, groups:  Alevel, 20

Fixed effects:
            Estimate Std. Error t value
(Intercept)  95.1638     6.9784   13.64
x1            3.3648     0.1181   28.48

Correlation of Fixed Effects:
   (Intr)
x1 0.001
```

Note that the Fixed effect estimates were recovered, more or less. The intercept was estimated to be 95.164 which is close to 100, and the coefficient for X1 was spot-on at 3.

Note I'm still trying to figure out how to recover siga and sige from this output...

Null Model

$$y_{ij} = \beta_0 + \alpha_j + \varepsilon_{ij}$$
$$\alpha_j + Gaussian(0, \sigma_\alpha^2)$$
$$\varepsilon ij + Gaussian(0, \sigma_\varepsilon^2)$$

```
mod0 <- lmer(y ~ 1 + (1|Alevel), data=dat)
summary(mod0)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ 1 + (1 | Alevel)
   Data: dat

REML criterion at convergence: 10271.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8979 -0.6763  0.0336  0.6533  3.5152

Random effects:
 Groups   Name        Variance Std.Dev.
 Alevel   (Intercept)  959.1   30.97
 Residual             1590.4   39.88
Number of obs: 1000, groups:  Alevel, 20

Fixed effects:
            Estimate Std. Error t value
(Intercept)   94.899      7.039   13.48
```

Options for R^2

Obtain the variances at each level

Most of the R^2 equations rely on the variance at the level of observation (residual) and at the level of the A group (`effA`). Extract the variance tables from each of the models, and save the pieces we want (under the `vcov`) heading in easier to handle names:

```
(vars0 <- as.data.frame(VarCorr(mod0)))
```

```
       grp        var1 var2      vcov    sdcor
1   Alevel (Intercept) <NA>  959.1121 30.96954
2 Residual        <NA> <NA> 1590.4469 39.88041
```

```
vara_0 <- vars0[1,4]
vare_0 <- vars0[2,4]
```

Here `vara_0` is the variance for level a for the null model $\sigma^2_{a0} = 959.112$ and `vare_0` is the variance for the residual $\sigma^2_{\varepsilon 0} = 1590.447$

Similarly we get the same information for the model with the fixed effect:

```
(vars1 <- as.data.frame(VarCorr(mod1)))
```

```
       grp        var1 var2     vcov    sdcor
1   Alevel (Intercept) <NA> 956.5491 30.92813
2 Residual        <NA> <NA> 870.7986 29.50930
```

```
vara_1 <- vars1[1,4]
vare_1 <- vars1[2,4]
```

Here `vara_1` is the variance for level a for the null model $\sigma^2_a = 956.549$ and `vare_1` is the variance for the residual $\sigma^2_\varepsilon = 870.7986$

Now we explore several Rˆ2 options outlined in Nakagawa & Schielzeth 2013

These equations were proposed by Snijders & Bosker (1994) for Linear mixed models with 1 random factor.

Nakagawa and Schielzeth 2013

Nakagawa and Schielzeth 2013 suggest new versios for R^2. They both rely on σ^2_f, the variance explained by the fixed effect commponents. This is estimated by multiplying the design matrix for the fixed effects by the vector of fixed effects estimates. This is also the same as predicting the data without worrying about the random effects.

$$\sigma^2_f = var(\beta_1 x_1)$$

```
varf_1 <- var(as.vector(lme4::fixef(mod1) %*% t(mod1@pp$X)))


#another way to get this is to get the yhats wihtout considering the random effects.

#varf_1 predict(mod1,  re.form=NA)
```

Marginal $\mathbf{R^2}$ (whole model)

$R^2_{LMM(m)}$ is the marginal R^2 for a linear mixed model, meaning that it is concerned with the variance explained by the fixed factors:

$$R^2_{LMM(m)} = \frac{\sigma^2_f}{\sigma^2_f + \sigma^2_\alpha + \sigma^2_\varepsilon}$$

```
(R2LMMm <- varf_1/(varf_1+vara_1+vare_1))
```

Conditional $\mathbf{R^2}$ (whole model)

$R^2_{LMM(c)}$ is the conditional R^2 for a linear mixed model, meaning that it is concerned with the variance explained by the fixed *and random* factors:

$$R^2_{LMM(m)} = \frac{\sigma^2_f + \sigma^2_\alpha}{\sigma^2_f + \sigma^2_\alpha + \sigma^2_\varepsilon}$$

```
(R2LMMc <- (varf_1+vara_1)/(varf_1+vara_1+vare_1))
```

```
[1] 0.6575735
```

Proportion change in variance at each level of variance

The authors also advocate for lookign at the proportion change in variance at each level, which looks at the ratio of variances at each level for the model with fixed-effects to the variances at each level for the null model.

$$PCV_\alpha = 1 - \frac{\sigma^2_\alpha}{\sigma^2_{\alpha 0}}$$
$$PCV_\varepsilon = 1 - \frac{\sigma^2_\varepsilon}{\sigma^2_{\varepsilon 0}}$$

```
(PCVa <- 1 - vara_1/vara_0)
```

```
[1] 0.002672253
```

```
(PCVe <- 1 - vare_1/vare_0)
```

```
[1] 0.4524819
```