# Homework 12

Steve Harms

April 19, 2018

## 1)

### a)

```
leaf<- read.table(file="LeafArea.txt",header=T)
leaf$ResearchStation <- factor(leaf$ResearchStation)
mod1 <- lmer(LeafArea ~ Dose + (1 + Dose | ResearchStation), data = leaf)
mod1

## Linear mixed model fit by REML ['lmerMod']
## Formula: LeafArea ~ Dose + (1 + Dose | ResearchStation)
##    Data: leaf
## REML criterion at convergence: 1333.905
## Random effects:
##  Groups          Name        Std.Dev. Corr
##  ResearchStation (Intercept) 3.238626
##                  Dose        0.007499 0.06
##  Residual                    1.987148
## Number of obs: 300, groups:  ResearchStation, 15
## Fixed Effects:
## (Intercept)         Dose
##     13.8577       0.0519

sige <- sigma(mod1)^2
sige

## [1] 3.948756
```

**From the model, we can quickly get $\sigma_e^2 = 1.987148^2 = 3.948756$.**

### b)

**We can use the random effects to get _{b} as below, using the standard covariance matrix formula for a multivariate normal distribution**

```
Sigmab <- matrix(c(3.238626^2, .06*3.238626*.007499, .06*3.238626*.007499,
.007499^2), nrow = 2, ncol = 2)
Sigmab

##               [,1]         [,2]
## [1,] 10.488698368 0.001457187
## [2,]  0.001457187 0.000056235
```
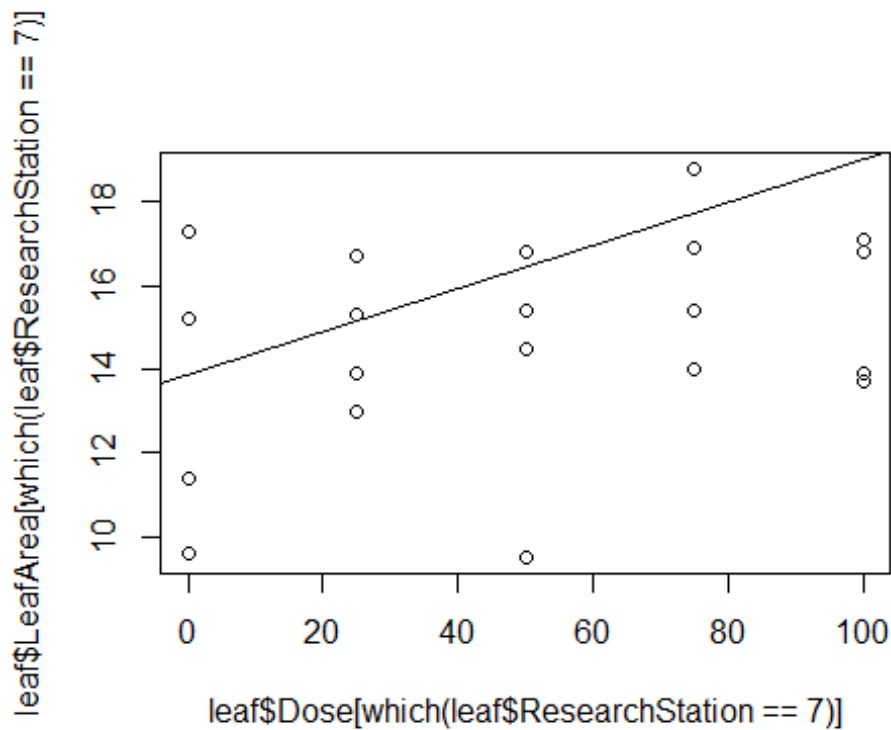
## c)

```
##c)
#get our regression coefficients using the model
b <- fixef(mod1)[1]
a <- fixef(mod1)[2]
c(a,b)

##          Dose (Intercept)
##       0.05190    13.85767

#plot data and the line together
plot(y=leaf$LeafArea[which(leaf$ResearchStation==7)], x =
leaf$Dose[which(leaf$ResearchStation==7)]) +
  abline(coef = c(b,a))
```



```
## integer(0)
```

## d)

```
##d)
B1b17 <- fixef(mod1)[1] + ranef(mod1)$ResearchStation[7,1]
B2b27 <- fixef(mod1)[2] + ranef(mod1)$ResearchStation[7,2]
c(B1b17,B2b27)

## (Intercept)        Dose
## 12.59070435   0.04378328
```

**The regression function is y = 12.5907 + .04378x**

## e)

```
station7 <- leaf[leaf$ResearchStation==7,]
mod7 <- lm(LeafArea ~ Dose, data=station7)
mod7

##
## Call:
## lm(formula = LeafArea ~ Dose, data = station7)
##
## Coefficients:
## (Intercept)          Dose
##     13.6500        0.0222
```

**The SLR function is y = 13.65 + .0222x**

## f)

```
plot(y=leaf$LeafArea[which(leaf$ResearchStation==7)], x =
leaf$Dose[which(leaf$ResearchStation==7)], pch = 20,
     ylab = "Leaf Area" , xlab = "Dose") +
  abline(coef = c(b,a), lwd = 2)   +
  abline(reg= mod7, col ="Blue", lwd = 2) +
  abline(coef = c(B1b17,B2b27), col = "Red", lwd = 2)
```
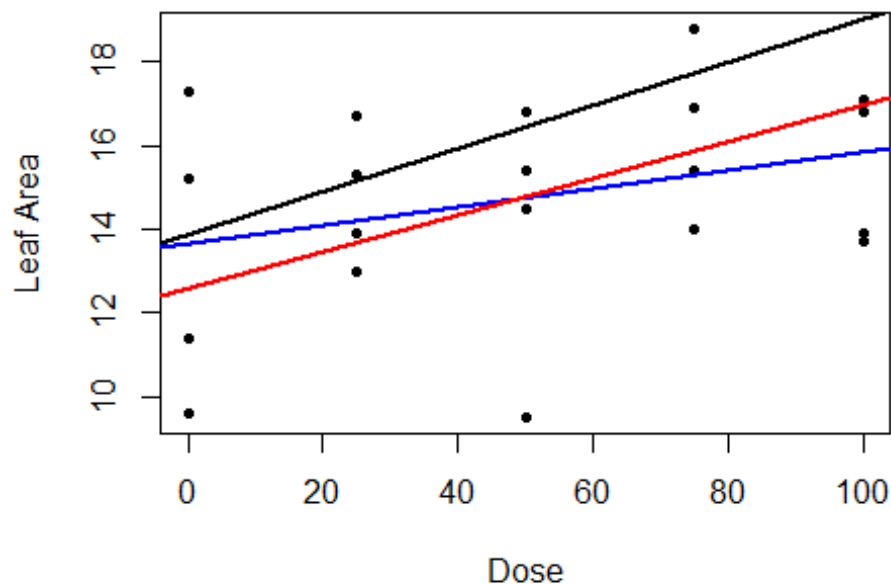


```
## integer(0)
```

## g)

```
#fit a reduced model where B2 = 0
mod0 <- lmer(LeafArea ~ 1 + (1 + Dose | ResearchStation), data = leaf)
#Calculate the Likelihood ratio statistic for reduced vs. full model
lrs <- -2*(logLik(mod0) - logLik(mod1))
lrs

## 'log Lik.' 30.73227 (df=5)
```

## h)

```
AIC(mod1)

## [1] 1345.905

-2*logLik(mod1) + 2*(6)

## 'log Lik.' 1345.905 (df=6)
```

## i)

```
#fit model with only a random effect on the intercept
modi <- lmer(LeafArea ~ Dose + (1 | ResearchStation), data = leaf)
AIC(modi)

## [1] 1342.693

-2*logLik(modi) + 2*4

## 'log Lik.' 1342.693 (df=4)
```

## j)

```
#without random effects for intercept or slope, it's just a simple linear
regression
modj <- lm(LeafArea ~ Dose, data = leaf)
AIC(modj)

## [1] 1650.107

-2*logLik(modj) + 2*3

## 'log Lik.' 1650.107 (df=3)
```

## k)

```
anova(modi,modj,mod1)

## refitting model(s) with ML (instead of REML)

## Data: leaf
## Models:
## modj: LeafArea ~ Dose
## modi: LeafArea ~ Dose + (1 | ResearchStation)
## mod1: LeafArea ~ Dose + (1 + Dose | ResearchStation)
```

```
##        Df    AIC    BIC  logLik deviance    Chisq Chi Df Pr(>Chisq)
## modj  3 1650.1 1661.2 -822.05   1644.1
## modi  4 1334.6 1349.4 -663.29   1326.6 317.5320      1    <2e-16 ***
## mod1  6 1338.0 1360.3 -663.02   1326.0   0.5419      2    0.7627
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#model from i is preferred with the smallest AIC
```

**We see that the model from part (i), which has no random slope effect, is the
preferred with the smallest AIC.**


## 2)


## 3)

```
donner <- read.table(file="Donner.txt", header=T)
mod3 <- glm(status~sex + age, family = binomial(link=logit), data= donner)
summary(mod3)

##
## Call:
## glm(formula = status ~ sex + age, family = binomial(link = logit),
##     data = donner)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -1.7445  -1.0441  -0.3029   0.8877   2.0472
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.23041    1.38686   2.329   0.0198 *
## sexMALE      -1.59729    0.75547  -2.114   0.0345 *
## age          -0.07820    0.03728  -2.097   0.0359 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 51.256  on 42  degrees of freedom
## AIC: 57.256
##
## Number of Fisher Scoring iterations: 4

mod3a <- glm(status~sex, family = binomial(link=logit), data= donner)
mod3b <- glm(status~age, family = binomial(link=logit), data= donner)
anova(mod3, mod3a, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ sex + age
## Model 2: status ~ sex
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        42      51.256
## 2        43      57.286 -1    -6.03  0.01406 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(mod3, mod3b, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: status ~ sex + age
## Model 2: status ~ age
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        42      51.256
## 2        43      56.291 -1  -5.0344  0.02485 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef(mod3))-1

## (Intercept)      sexMALE          age
## 24.29012009 -0.79755631 -0.07522431

exp(confint(mod3))-1

## Waiting for profiling to be done...

##                   2.5 %       97.5 %
## (Intercept)   1.3429692 617.1277279
## sexMALE      -0.9603899  -0.1772493
## age          -0.1499309  -0.0139673

1/exp(confint(mod3))

## Waiting for profiling to be done...

##                   2.5 %       97.5 %
## (Intercept)   0.4268089 0.001617789
## sexMALE      25.2460693 1.215434931
## age           1.1763750 1.014165144
```

From the results of my analysis, we can see that both age and sex have a significant effect on survival likelihood at a 5% significance level. The model that includes both variables is preferred over either variable on its own as a predictor.

We can see that females are 1/.2024437 = 4.9396 times as likely to survive than males, after controlling for age. However, the variance our estimate is large. We are 95% confident that the odds of females surviving vs. males is between 25.2460693

and 1.215434931 times as likely. It is clear that women at any age are more likely to survive than men, but the magnitude is uncertain based on our analysis.

After controlling for Sex, we can see that each additional year in age reduces survival probability by an estimated 7.5%. We can be 95% confident that the predicted reduction for each additional year is between 1.4% and 17%.

The plot below shows predicted survival probability for both males and females as age increases. We can see from the plot that males are much less likely to survive than females, and that older members are less likely to survive than younger members, regardless of sex.

```
x <- min(donner$age):max(donner$age)
plot(x,1/(1+exp(-(coef(mod3)[1]+coef(mod3)[3]*x))),ylim=c(0,1),
     type="l",col=4,lwd=2,xlab="Age",
     ylab="Estimated Probability of Survival", cex.lab=1.3)
lines(x,1/(1+exp(-(coef(mod3)[1]+coef(mod3)[3]*x +
coef(mod3)[2]))),col=2,lwd=2)
legend("topleft", legend=c("Females","Males"),
       col=c(4,2),lwd=2)
```
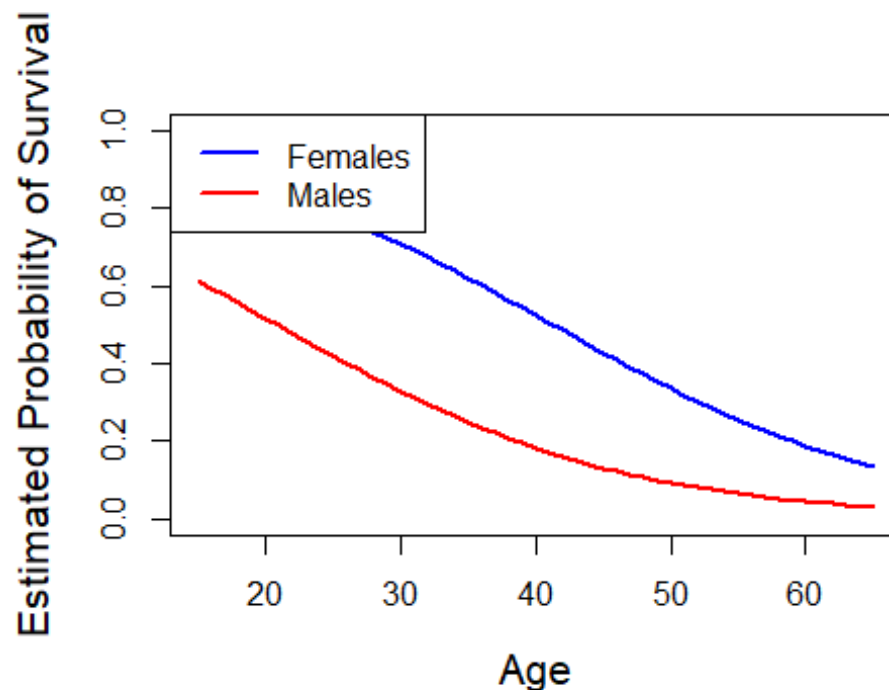


# 4)

The first thing to note is that the ANOVA model assumes normally distributed errors. Since the data is binomial count data and therefore not normal, it should be clear

that this assumption is violated and any inference based on this model will not be valid. Next, note that the data is unbalanced which will have an effect on R's estimate of variance terms, which will effect the standard error of the estimate and could lead to incorrect conclusions. Finally, since the data is binomial we know that it is not linear, using a standard linear model will not give a correct estimate.