Alyssa Semerdjian

Stat 510

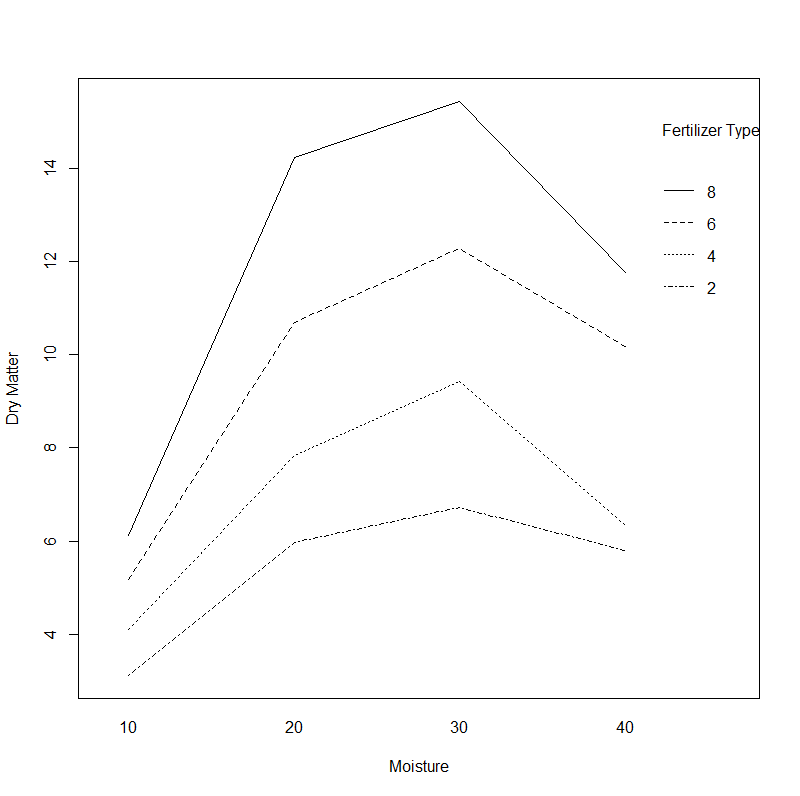
19 March 2018

Homework 8

1.

(a) The model lm(DryMatter~Moisture\*fertilizer) does not take into account that fertilizer type is nested within moisture treatments and therefore is not independent.

(b) (i) The lines on the interaction plot are not perfectly parallel, so I think that there is a significant interaction between fertilizer and moisture. Fertilizer 8 seems to be the best fertilizer. This fertilizer produces more dry matter regardless of the moisture treatment.



(ii) The t-value for most of the interaction terms in lmer(DryMatter~Moisture\*fertilizer+(1|Tray)) have t-values that does not fall between -2 and 2. This means that there is not a significant interaction.

> a=lmer(DryMatter~Moisture\*fertilizer+(1|Tray))

> summary(a)

Linear mixed model fit by REML ['lmerMod']

Formula: DryMatter ~ Moisture \* fertilizer + (1 | Tray)

REML criterion at convergence: 111.4

Scaled residuals:

Min 1Q Median 3Q Max

-1.42079 -0.57309 -0.09207 0.45819 2.11322

Random effects:

Groups Name Variance Std.Dev.

Tray (Intercept) 0.6636 0.8146

Residual 0.7521 0.8673

Number of obs: 48, groups: Tray, 12

Fixed effects:

Estimate Std. Error t value

(Intercept) 3.1220 0.6870 4.545

Moisture20 2.8491 0.9715 2.933

Moisture30 3.5931 0.9715 3.698

Moisture40 2.6737 0.9715 2.752

fertilizer4 0.9773 0.7081 1.380

fertilizer6 2.0605 0.7081 2.910

fertilizer8 3.0030 0.7081 4.241

Moisture20:fertilizer4 0.8787 1.0014 0.877

Moisture30:fertilizer4 1.7437 1.0014 1.741

Moisture40:fertilizer4 -0.4424 1.0014 -0.442

Moisture20:fertilizer6 2.6673 1.0014 2.663

Moisture30:fertilizer6 3.5035 1.0014 3.498

Moisture40:fertilizer6 2.3273 1.0014 2.324

Moisture20:fertilizer8 5.2618 1.0014 5.254

Moisture30:fertilizer8 5.7166 1.0014 5.708

Moisture40:fertilizer8 2.9676 1.0014 2.963

(c) The p-value for the anova comparison of the two models is 5.38e-6. Because the P-value is less than 0.05 we can reject the null hypothesis that the two models are the same, and therefore can conclude that the interaction term contributes significantly to the model. In short the long model is “better”.

> a=lmer(DryMatter~Moisture\*fertilizer+(1|Tray),REML=F)

> b=lmer(DryMatter~Moisture+fertilizer+(1|Tray),REML=F)

> anova(a,b)

Data: NULL

Models:

b: DryMatter ~ Moisture + fertilizer + (1 | Tray)

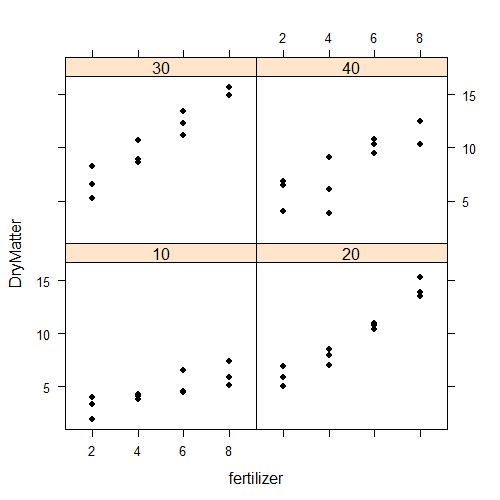
a: DryMatter ~ Moisture \* fertilizer + (1 | Tray)

Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)

b 9 180.04 196.88 -81.018 162.04

a 18 157.21 190.89 -60.605 121.21 40.826 9 5.38e-06 \*\*\*

(d) (i) Both fertilizer and moisture are significant. There is a positive relationship between the variables but the scatter plots comparing fertilizer to dry matter are slightly different for each moisture level.



(ii) Both fertilizer and moisture are significant. When we compare a full model with both factors to models where one of the factors has been removed we get p-vales less than 0.05, suggesting that the removed variable was contributing significantly to the model. This suggests that in both cases the longer model (with both terms) is better.

> a=lmer(DryMatter~Moisture+fertilizer+(1|Tray),REML=F)

> b=lmer(DryMatter~fertilizer+(1|Tray),REML=F)

> anova(a,b)

Data: NULL

Models:

b: DryMatter ~ fertilizer + (1 | Tray)

a: DryMatter ~ Moisture + fertilizer + (1 | Tray)

Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)

b 6 202.68 213.90 -95.338 190.68

a 9 180.04 196.88 -81.018 162.04 28.641 3 2.664e-06 \*\*\*

> a=lmer(DryMatter~Moisture+fertilizer+(1|Tray),REML=F)

> b=lmer(DryMatter~Moisture+(1|Tray),REML=F)

> anova(a,b)

Data: NULL

Models:

b: DryMatter ~ Moisture + (1 | Tray)

a: DryMatter ~ Moisture + fertilizer + (1 | Tray)

Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)

b 6 247.58 258.81 -117.790 235.58

a 9 180.04 196.88 -81.018 162.04 73.545 3 7.428e-16 \*\*\*

2.

(a) (i)

> tapply(strength,list(batch,cask),mean)

a b c

A 62.70 61.20 62.90

B 60.70 57.20 60.00

C 58.10 63.50 64.55

D 56.75 57.75 64.60

E 55.10 54.45 58.15

F 64.15 58.70 60.25

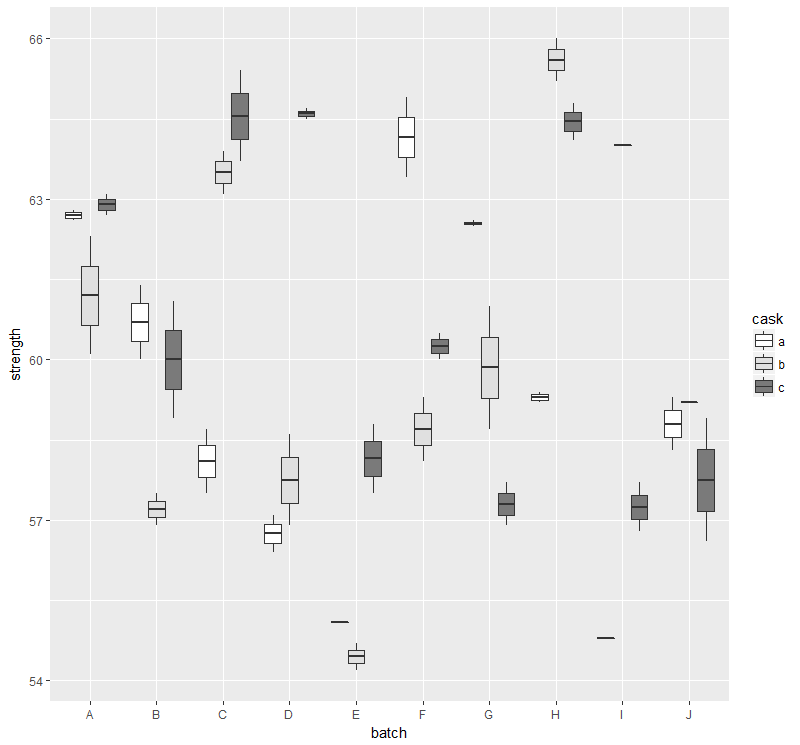
G 62.55 59.85 57.30

H 59.30 65.60 64.45

I 54.80 64.00 57.25

J 58.80 59.20 57.75

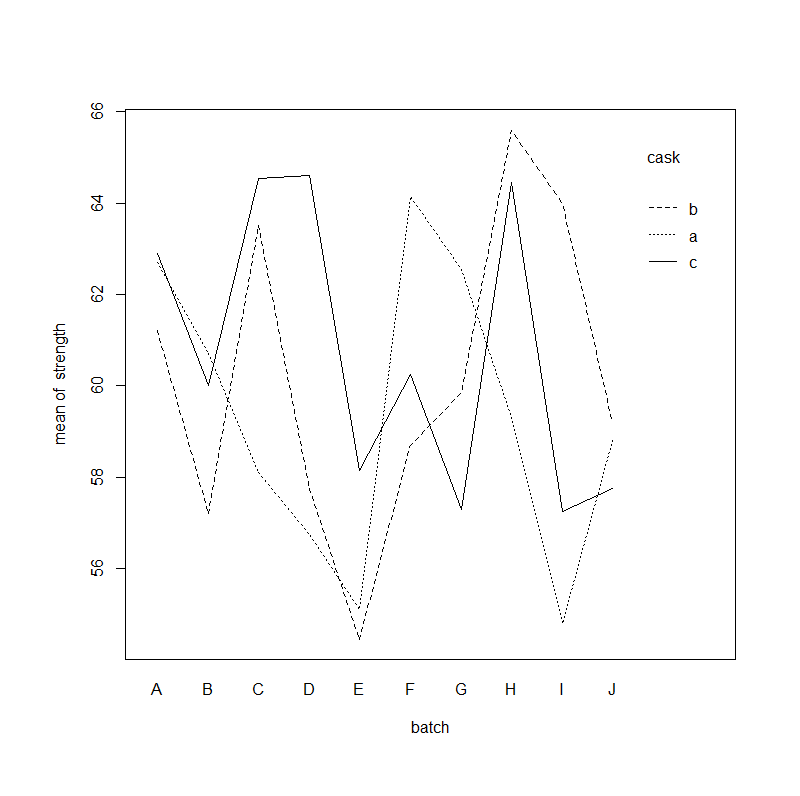
(ii)



(iii) It looks like there are significant differences between some of the batches but not others. There is also lot of variation within batches. Overall I’m guessing that batch is not a significant variable.

(iv), I think cask will be a more significant variable. There appear to be significant differences between casks within most of the batches.

(v) According to the crossed lines on the interaction plot there is a significant interaction between batch and cask.



(b) Model 1 [lmer(strength\_(1|batch)+(1|cask))] is incorrect because it does not take into account that casks are nested within batches. It combined all the “a” casks, “b” casks and “c” casks and considered them together even though they were from different batches.

(c) The estimated variance due to batch is 1.657

The estimated variance due to cask is 8.434

Tthe estimated variance due to pure error is 0.678

If you had a chance to redo the experiment, you'd like to collect more (circle one:

batches or **casks**).

(d) The p-value of the anova testing whether batch is a significant variable is 0.5234. This indicates that the batch variable is not contributing significantly to the model and that the shorter model is better.

> a=lmer(strength~(1|batch)+(1|batch:cask),REML=F)

> b=lmer(strength~(1|batch:cask),REML=F)

> anova(a,b)

Data: NULL

Models:

b: strength ~ (1 | batch:cask)

a: strength ~ (1 | batch) + (1 | batch:cask)

Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)

b 3 254.40 260.69 -124.2 248.40

a 4 255.99 264.37 -124.0 247.99 0.4072 1 0.5234

(e) The p-value of the anova testing whether cask is a significant variable is 1.474e-13. This indicates that the cask variable is contributing significantly to the model and that the longer model is better.

> a=lmer(strength~(1|batch)+(1|batch:cask),REML=F)

> b=lmer(strength~(1|batch),REML=F)

> anova(a,b)

Data: NULL

Models:

b: strength ~ (1 | batch)

a: strength ~ (1 | batch) + (1 | batch:cask)

Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)

b 3 308.60 314.88 -151.3 302.60

a 4 255.99 264.37 -124.0 247.99 54.605 1 1.474e-13 \*\*\*