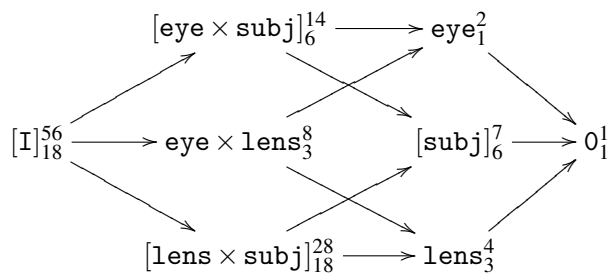


Besvarelse af opgave 5.2

1.

Faktorerne lens (niveauer: 6/6,6/18,6/36,6/60) og eye (niveauer: left,right) bør indgå i analyserne med systematisk effekt. Faktoren subj (niveauer: 1,2,3,4,5,6,7) og dermed også vekselvirkninger hermed bør indgå i modellen med tilfældig effekt.



2.

Statistisk model hørende til faktordiagrammet:

$$Y_i = \gamma(\text{eye} \times \text{lens}_i) + A(\text{subj}_i) + B(\text{eye} \times \text{subj}_i) + C(\text{lens} \times \text{subj}_i) + e_i,$$

hvor

- $\gamma(\text{left}, 6/6), \dots, \gamma(\text{right}, 6/60)$ er faste parametre hørende til $\text{eye} \times \text{lens}$
- $A(1), \dots, A(7)$ er uafh. $\sim N(0, \sigma_{\text{subj}}^2)$
- $B(\text{left}, 1), \dots, B(\text{right}, 7)$ er uafh. $\sim N(0, \sigma_{\text{eye} \times \text{subj}}^2)$
- $C(6/6, 1), \dots, C(6/60, 7)$ er uafh. $\sim N(0, \sigma_{\text{lens} \times \text{subj}}^2)$
- e_1, \dots, e_{56} er uafh. $\sim N(0, \sigma^2)$.

3.

Da forsøgsdesignet er balanceret, (-vi har netop en forsøgsenhed for hver kombination af lens, subj og eye) kan man lave F-test for varianskomponenterne svarende til $\text{eye} \times \text{subj}$ og $\text{lens} \times \text{subj}$ som i sædvanelige lineære modeller.

Test for: $\sigma_{\text{eye} \times \text{subj}}^2 = 0$

```

> model1=lm(timelag~subj*eye+subj*lens+eye*lens)
> model2a=lm(timelag~subj*lens+eye*lens)
> anova(model2a,model1)
  
```

Analysis of Variance Table

Model 1: timelag ~ subj * lens + eye * lens

Model 2: timelag ~ subj * eye + subj * lens + eye * lens

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	24	588.43				
2	18	231.00	6	357.43	4.6419	0.005109 **

Vi ser, at der er en signifikant effekt af $\text{eye} \times \text{subj}$ ($F=4.64, p=0.005$).

Test for: $\sigma_{\text{lens} \times \text{subj}}^2 = 0$

```
> model1=lm(timelag~subj*eye+subj*lens+eye*lens)
```

```
> model2b=lm(timelag~subj*eye+eye*lens)
```

```
> anova(model2b,model1)
```

Analysis of Variance Table

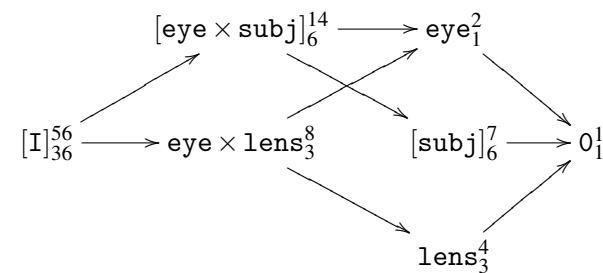
Model 1: timelag ~ subj * eye + eye * lens

Model 2: timelag ~ subj * eye + subj * lens + eye * lens

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	36	606.86				
2	18	231.00	18	375.86	1.6271	0.1554

Vi ser, at vi kan se bort fra effekten af $\text{lens} \times \text{subj}$ ($F=1.63, p=0.155$).

Faktordiagram hørende til den reducerede model



4.

Med udgangspunkt i modellen fra 3.

$$Y_i = \gamma(\text{eye} \times \text{lens}_i) + A(\text{subj}_i) + B(\text{eye} \times \text{subj}_i) + e_i,$$

reduceres den systematiske del af modellen.

```
> subje=eye:subj
```

```
> mod1=lme(timelag~eye*lens, random=~1|subj/subje, method="ML")
```

```
> mod2=lme(timelag~eye+lens, random=~1|subj/subje, method="ML")
```

```
> anova(mod2, mod1)
```

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
-------	----	-----	-----	--------	------	---------	---------

```

mod2      1  8 354.3246 370.5274 -169.1623
mod1      2 11 357.6031 379.8820 -167.8015 1 vs 2 2.721508 0.4366
> mod3=lme(timelag~lens, random=~1|subj/subjeye, method="ML")
> anova(mod3, mod2)
      Model df      AIC      BIC    logLik    Test   L.Ratio p-value
mod3      1  7 353.1798 367.3573 -169.5899
mod2      2  8 354.3246 370.5274 -169.1623 1 vs 2 0.8551892 0.3551
> mod4=lme(timelag~1, random=~1|subj/subjeye, method="ML")
> anova(mod4, mod3)
      Model df      AIC      BIC    logLik    Test   L.Ratio p-value
mod4      1  4 355.4422 363.5436 -173.7211
mod3      2  7 353.1798 367.3573 -169.5899 1 vs 2 8.262412 0.0409

```

Det ses, at vi kan fjerne vekselvirkningen $\text{eye} \times \text{lens}$ ($LR = 2.72, p = 0.43$) samt hovedvirkningen eye ($LR = 0.855, p = 0.355$), mens der er en signifikant effekt af lens ($LR = 8.262, p = 0.041$). Da p-værdien for testet for hovedeffekten af lens er meget tæt på signifikansniveauet (-her 5%), og da vi benytter os af det approksimative likelihood ratio test, kan man eventuelt forsøge at simulere sig frem til den eksakte p-værdi, som anført nedenfor

```

> sim<-simulate.lme(mod4, m2=mod3, nsim=5000, method="ML")
> lr.sim<-2*(sim$alt$ML-sim$null$ML)
> psim<-sum(lr.sim>8.262)/5000
> psim
[1] 0.0498

```

Vi vælger at stå fast på beslutningen om, at hovedeffekten af lens er signifikant, omend dette kan diskuteres. Det er meget muligt, at du med `simulate.lme` får en p-værdi lige over 5 %, og du er velkommen til at benytte modellen uden effekt af lens som slutmodel.

5.

Slutmodellen bliver

$$Y_i = \alpha(\text{lens}_i) + A(\text{subj}_i) + B(\text{eye} \times \text{subj}_i) + e_i,$$

hvor

- $\alpha(6/6), \dots, \alpha(6/60)$ er faste parametre hørende til lens
- $A(1), \dots, A(7)$ er uafh. $\sim N(0, \sigma_{\text{subj}}^2)$
- $B(\text{left}, 1), \dots, B(\text{right}, 7)$ er uafh. $\sim N(0, \sigma_{\text{eye} \times \text{subj}}^2)$
- e_1, \dots, e_{56} er uafh. $\sim N(0, \sigma^2)$.

og de tilhørende parameterestimer bliver (-se R-output nedenfor)

$$\hat{\sigma} = 4.075, \quad \hat{\sigma}_{\text{subj}} = 4.64, \quad \hat{\sigma}_{\text{eye} \times \text{subj}} = 3.205$$

```

      lens = 6/6   112.6  [108.1,117.2]
      lens = 6/18  113.4  [108.9,117.9]
      lens = 6/36  111.6  [107.1,116.2]
      lens = 6/60  115.9  [111.4,120.4].

> mod3ny=lme(timelag~lens-1,random=~1|subj/subjeye,method="REML")
> summary(mod3ny)
Linear mixed-effects model fit by REML
Data: NULL
      AIC      BIC    logLik
342.7098 356.3686 -164.3549

Random effects:
Formula: ~1 | subj
      (Intercept)
StdDev:    4.639616

      Formula: ~1 | subjeye %in% subj
      (Intercept) Residual
StdDev:    3.205242 4.074568

Fixed effects: timelag ~ lens - 1
      Value Std.Error DF   t-value p-value
lens6/18 113.4286   2.234914 39 50.75300     0
lens6/36 111.6429   2.234914 39 49.95399     0
lens6/6  112.6429   2.234914 39 50.40143     0
lens6/60 115.9286   2.234914 39 51.87161     0

> intervals(mod3ny)
Approximate 95% confidence intervals

Fixed effects:
      lower    est.    upper
lens6/18 108.9080 113.4286 117.9491
lens6/36 107.1223 111.6429 116.1634
lens6/6  108.1223 112.6429 117.1634
lens6/60 111.4080 115.9286 120.4491

```

6.

Startmodellen fra spørgsmål 2. indeholder tre tilfældige faktorer $\text{eye} \times \text{subj}$, $\text{lens} \times \text{subj}$ og subj . Da faktorerne ikke alle er nestede, (-nemlig $\text{eye} \times \text{subj}$ og $\text{lens} \times \text{subj}$) kan man ikke fitte modellen ved brug af lme. I stedet benyttes lmer som kræver lme4-pakken.

```

> library(lme4)
> model=lmer(timelag~eye*lens+(1|subj)+(1|subj:eye)+(1|subj:lens))
> summary(model)

```

Linear mixed-effects model fit by REML

Formula: timelag ~ eye * lens + (1 | subj) + (1 | subj:eye) + (1 | subj:lens)

AIC	BIC	logLik	MLdeviance	REMLdeviance
331.5703	353.8492	-154.7852	334.3711	309.5703

Random effects:

Groups	Name	Variance	Std.Dev.
subj:lens	(Intercept)	4.0238	2.0059
subj:eye	(Intercept)	11.6845	3.4183
subj	(Intercept)	20.2857	4.5040
Residual		12.8333	3.5824

number of obs: 56, groups: subj:lens, 28; subj:eye, 14; subj, 7

Varianskomponenterne estimeres til

$$\hat{\sigma} = 3.582, \quad \hat{\sigma}_{\text{subj}} = 4.504, \quad \hat{\sigma}_{\text{eye} \times \text{subj}} = 3.418, \quad \hat{\sigma}_{\text{lens} \times \text{subj}} = 2.006.$$

I overensstemmelse med spørgsmål 3. ses at estimatet for varianskomponenten hørende til $\text{lens} \times \text{subj}$ er klart mindst. Når man vil sammenligne varianskomponenterne er det faktisk mere rigtig at kvadrere estimaterne først. Den totale varians på de enkelte observationer er

$$\text{Var}Y_i = \sigma_{\text{eye} \times \text{subj}}^2 + \sigma_{\text{lens} \times \text{subj}}^2 + \sigma_{\text{subj}}^2 + \sigma^2$$

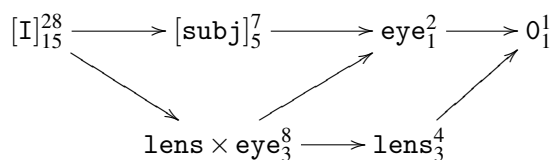
så den estimerede del af variansen, som stammer fra $\text{lens} \times \text{subj}$ udgør

$$\frac{\hat{\sigma}_{\text{lens} \times \text{subj}}^2}{\hat{\sigma}_{\text{eye} \times \text{subj}}^2 + \hat{\sigma}_{\text{lens} \times \text{subj}}^2 + \hat{\sigma}_{\text{subj}}^2 + \hat{\sigma}^2} = \frac{2.006^2}{3.418^2 + 2.006^2 + 4.504^2 + 3.582^2} = 0.082$$

dvs. omkring 8%.

7.

For det beskrevne forsøg er der kun 28 observationer, og vi konstaterer at $\text{subj} \times \text{lens}$ er lig med den identiske faktor I. Desuden bemærkes at subj er finere end eye . Den statistiske model bør således indeholde subj som tilfældig faktor, og vekselvirkning $\text{eye} \times \text{lens}$ samt tilhørende hovedvirkninger bør indgå med systematisk effekt. Faktordiagrammet ser ud som følger



Besvarelse af opgave 5.4

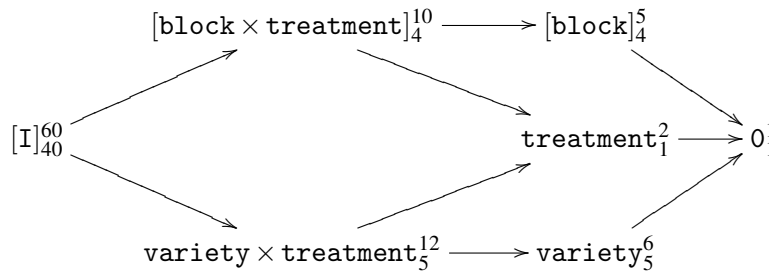
1.

I forsøget indgår 60 forsøgsenheder som hver især er beskrevet ved de tre faktorer **block**, **treatment** og **variety**. Behandlingsfaktoren, **treatment**, kan kun benyttes på større enheder (hel plots) hver bestående af 6 forsøgsenheder. Inden for hvert helplot randomiseres de 6 forskellige sorter (**variety**). Der er således tale om et split-plot forsøg. Som en ekstra krølle på

forsøgsdesignet er de 10 helpplots organiseret i 5 blokke (block) med 2 helpplots i hver. Ved udarbejdelsen af forsøgsplanen har man således også sørget for at randomisere de to behandlinger inden for de to helpplots på hver block.

2.

Faktoren block bør indgå i modellen med tilfældig effekt, mens treatment og variety indgår som systematiske faktorer. Som anført i opgaveformulering ses bort fra vekselvirkningen $\text{block} \times \text{variety}$, men de øvrige vekselvirkninger $\text{block} \times \text{treatment}$ og $\text{treatment} \times \text{variety}$ bør indgå i den statistiske model. Da block inddrages som tilfældige effekt skal vekselvirkningen hørende til helpplotfaktoren $\text{block} \times \text{treatment}$ også være tilfældig. Vi ender således op med følgende faktordiagram



Den statistiske model hørende til faktordiagrammet bliver

$$Y_i = \gamma(\text{variety} \times \text{treatment}_i) + A(\text{block} \times \text{treatment}_i) + B(\text{block}_i) + e_i,$$

hvor

- $B(1), \dots, B(5)$ er uafhængige og normalfordelte $N(0, \sigma_B^2)$
- $A(1, T), \dots, A(5, T), A(1, U), \dots, A(5, U)$ er uafhængige og normalfordelte $N(0, \sigma_A^2)$
- e_1, \dots, e_{60} er uafhængige og normalfordelte $N(0, \sigma^2)$

3.

I første omgang forsøger vi at fjerne vekselvirkningen $\text{variety} \times \text{treatment}$ og reducere udgangsmodellen til

$$Y_i = \alpha(\text{variety}_i) + \beta(\text{treatment}_i) + A(\text{block} \times \text{treatment}_i) + B(\text{block}_i) + e_i,$$

hvor

- $B(1), \dots, B(5)$ er uafhængige og normalfordelte $N(0, \sigma_B^2)$
- $A(1, T), \dots, A(5, T), A(1, U), \dots, A(5, U)$ er uafhængige og normalfordelte $N(0, \sigma_A^2)$
- e_1, \dots, e_{60} er uafhængige og normalfordelte $N(0, \sigma^2)$.

I følge R-udskriften er testet stærkt signifikant ($LR = 63.02, p < 0.0001$).

Selvom det ikke er en del af opgaven, kan man (som en ekstra øvelse) godt forsøge reducere den tilfældige del af modellen. Det viser sig, at man lige akkurat kan fjerne begge de to tilfældige effekter. Da vi ønsker at generalisere resultaterne til vilkårlige helpplots, vælger vi dog at afrapportere estimater og konfidensintervaller for modellen med tilfældige effekter.

4.

Estimaterne for varianskomponenterne bliver

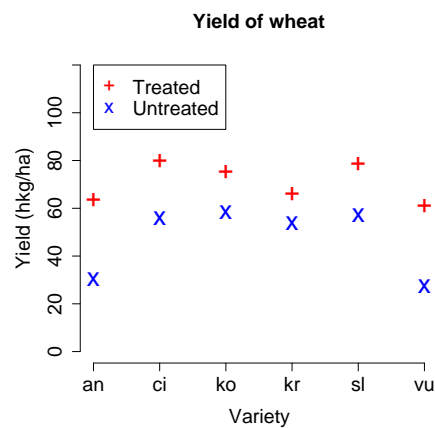
```

block      :  $\hat{\sigma}_B = 0.766$ 
block  $\times$  treatment :  $\hat{\sigma}_A = 1.227$ 
residual   :  $\hat{\sigma} = 3.041$ 

```

5.

Parameterestimaterne for de 12 kombinationer af variety og treatment fremgår af figuren nedenfor.



Med udgangspunkt i udskriften af modelA1 i R-udskriften i opgaveformuleringen viser følgende R-program, at behandlingen har en positiv effekt på udbyttet for alle niveauer af variety.

```

> library(gmodels)
> andiff<-c(1,0,0,0,0,0,-1,0,0,0,0,0)
> cidiff<-c(0,1,0,0,0,0,-1,0,0,0,0,0)
> kodiff<-c(0,0,1,0,0,0,-1,0,0,0,0,0)
> krdiff<-c(0,0,0,1,0,0,-1,0,0,0,0,0)
> sldiff<-c(0,0,0,0,1,0,-1,0,0,0,0,0)
> vudiff<-c(0,0,0,0,0,1,-1,0,0,0,0,0)
>
> diff<-rbind(andiff,cidiff,kodiff,krdiff,sldiff,vudiff)
> estimable(modelA1,diff)

```

	Estimate	Std. Error	t value	DF	Pr(> t)
andiff	32.88	2.073952	15.853789	39	0.000000e+00
cidiff	23.60	2.073952	11.379240	39	5.817569e-14
kodiff	16.54	2.073952	7.975111	39	1.021820e-09
krdiff	12.20	2.073952	5.882488	39	7.542024e-07
sldiff	21.14	2.073952	10.193099	39	1.484590e-12
vudiff	33.42	2.073952	16.114161	39	0.000000e+00

Den signifikante vekselvirkning mellem `variety` og `treatment` må således forklares med, at behandlingen virker bedre for visse af de 6 sorter i forsøget. Dette kan uddybes med supplerende anvendelser af `estimable`-funktionen i R.

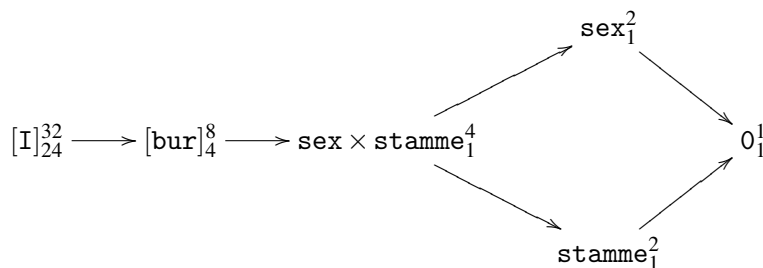
Besvarelse af opgave 5.5

1.

Det vigtige er at indse, at `bur` er finere end både `sex` og `stamme` og dermed også end produkt-faktoren `sex × stamme`. Udgangsmodellen bør være

$$Y_i = \gamma(\text{sex} \times \text{stamme}_i) + A(\text{bur}_i) + e_i,$$

hvor e_1, \dots, e_{32} er uafhængige og $N(0, \sigma^2)$ -fordelte og $A(1), \dots, A(8)$ er uafhængige og $N(0, \sigma_A^2)$ -fordelte. Faktordiagrammet bliver



2.

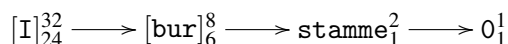
Vekselvirkningen mellem `sex` og `stamme` er ikke signifikant ($LR = 0.115, p = 0.734$), og der er en klart signifikant effekt af `stamme` ($LR = 8.77, p = 0.003$.) Det (approksimative) likelihood ratio test viser en svagt signifikant effekt af `sex` ($LR = 5.05, p = 0.025$). Når man benytter sig af approksimative test, bør man ikke være alt for skråsikker i forbindelse med p-værdier omkring signifikansniveauet på 5 %. En mulighed er at lave eksakte F-test som beskrevet ved torsdagsforelæsningen. Som alternativ kan man forsøge at simulere sig frem til p-værdien, som beskrevet i R-udskriften til opgavenformuleringen. I følge denne fås en p-værdi på $p = 0.0822$. På denne baggrund foreslås at man accepterer hypotesen om, at der ikke er nogen effekt af `sex`. (Prøv f.eks. at lave F-testet, som giver $F = 4.40, p = 0.09$.)

3.+4.

Slutmodel med tilhørende faktordiagram

$$Y_i = \beta(\text{stamme}_i) + A(\text{bur}_i) + e_i,$$

hvor e_1, \dots, e_{32} er uafhængige og $N(0, \sigma^2)$ -fordelte og $A(1), \dots, A(8)$ er uafhængige og $N(0, \sigma_A^2)$ -fordelte.



Vi genfitter modellen i R med `method="REML"` for at få parameterestimer under slutmodellen. For de systematiske parametre fås følgende estimer

$$\begin{aligned} \hat{\beta}(\text{normal}) &= 327.06 \quad [268.65, 385.48] \\ \hat{\beta}(\text{transgene}) - \hat{\beta}(\text{normal}) &= -100.94 \quad [-198.88, -2.99] \end{aligned}$$

For de tilfældige effekter fås estimaterne

$$\begin{aligned}\hat{\sigma}_A &= 50.61 \quad [24.85, 103.08] \\ \hat{\sigma} &= 50.70 \quad [38.21, 67.27].\end{aligned}$$

5.

Med parametrisering af slutmodellen som i spm. **3.-4.** ovenfor, svarer spørgsmål i opgaveformuleringen til hypotesen

$$\alpha(\text{normal}) = 1.2 \cdot \alpha(\text{transgene}) \text{ eller } \alpha(\text{normal}) - 1.2 \cdot \alpha(\text{transgene}) = 0.$$

Dette kan i R testes med følgende kode

```
> slutmodelny<-lme(kryds~stamme-1,random=~1|bur,method="REML")
> slutmodelny
Linear mixed-effects model fit by REML
  Data: NULL
Log-restricted-likelihood: -167.9375
Fixed: kryds ~ stamme - 1
      stamnormal stammetransgen
      327.0625      226.1250

Random effects:
Formula: ~1 | bur
      (Intercept) Residual
StdDev:    50.61441 50.69752

Number of Observations: 32
Number of Groups: 8
> diff20<-c(1,-1.2)
> estimable(slutmodelny,diff20,conf.int=0.95)
      Estimate Std. Error t value DF Pr(>|t|) Lower.CI Upper.CI
(1 -1.2)  55.7125    44.21166 1.260132  6 0.2544141 -52.46952 163.8945
```

p -værdien for testet bliver $p = 0.254$, så forsøget bekræfter således formodning om, at normale mus er 20 % mere frygtsomme end de transgene mus.

6.

Vi betragter her den additive model for tosidet variansanalyse

$$Y_i = \beta(\text{stamme}_i) + \gamma(\text{bur}_i) + e_i, \quad e_i \text{ uafh} \sim N(0, \sigma^2).$$

Estimatet for forskellen mellem transgene og normale mus bliver

$$\hat{\beta}(\text{transgene}) - \hat{\beta}(\text{normal}) = -108.25 \quad [-182.24, -34.26].$$

At der er forskel på estimatet fra **4.** og **6.** viser blot, at estimer for den samme effekt beregnes forskelligt afhængigt af, om man modellerer effekten af bur som tilfældig eller systematisk.

Konfidensintervallet i **6.** er smallest, fordi det her udtaler sig om den forventede forskel mellem transgene og normale mus, i et af de 8 bure fra forsøget. Fortolkningen af estimatet i **4.** er, at dette beskriver forskellen mellem de to typer mus, i et (tilfældigt) valgt bur, som ikke nødvendigvis er med i forsøget. Prisen for at kunne generalisere resultaterne til andre bure er, at vi må leve med et bredere konfidensinterval.