Eksamen i Statistisk Dataanalyse 2, 12. april 2012

Vejledende besvarelse

Opgave 1

1. Den statistiske model kan skrives som

$$\texttt{log(maxLA)}_i = \alpha(\texttt{race}_i) + \beta(\texttt{race}_i) \cdot \texttt{log(vgt)}_i + e_i,$$

hvor $e_1, \ldots, e_{82}$ er uafhængige $\sim N(0, \sigma^2)$. Modellen er fittet som `modA` i `R`-udskriften og parameterestimaterne bliver

$$\begin{aligned} \alpha(\texttt{Chihuahua}) &= -0.38 & \beta(\texttt{Chihuahua}) &= 0.72 \\ \alpha(\texttt{Dalmatiner}) &= 0.95 & \beta(\texttt{Dalmatiner}) &= 0.57 \\ \alpha(\texttt{Grand\_Danois}) &= 0.74 & \beta(\texttt{Grand\_Danois}) &= 0.67 \\ \alpha(\texttt{Gravhund}) &= -1.3 & \beta(\texttt{Gravhund}) &= 1.37 \end{aligned}$$

Estimatet for residualvariansen bliver $\hat{\sigma}^2 = 0.2676^2$.

2. Først konstateres, at `modA` kan reduceres til modellen (`modB` i `R`-udskrift)

$$\texttt{log(maxLA)}_i = \alpha(\texttt{race}_i) + \beta \cdot \texttt{log(vgt)}_i + e_i,$$

hvor hældningen ikke længere tillades at afhænge af `race` ($F = 1.981, p = 0.124$). Dernæst konstateres, at skæringen heller ikke lader til at afhænge af `race` ($F = 1.509, p = 0.219$) svarende til, at vi kan foretage en reduktion ned til modellen

$$\texttt{log(maxLA)}_i = \alpha + \beta \cdot \texttt{log(vgt)}_i + e_i.$$

Det kan desuden bemærkes, at det af testet givet ved kommandoen `anova(modE,modC)` fremgår, at hældningen, β, er signifikant forskellig fra nul.

Parameterestimaterne til beskrivelse af middelværdistrukturen bliver

$$\hat{\alpha} = -0.627[-0.767, -0.487](\text{skæring}) \quad \hat{\beta} = 1.015[0.966, 1.063](\text{hældning})$$

og variansestimatet bliver $\hat{\sigma}^2 = 0.076$.

3. Slutmodellen fra 2. siger, at `race` ikke har indflydelse på volumen af venstre forkammer, således at vi ender med den statistiske model

$$\texttt{log(maxLA)}_i = \alpha + \beta \cdot \texttt{log(vgt)}_i + e_i,$$

hvor $e_1, \ldots, e_{82}$ er uafhængige $\sim N(0, \sigma^2)$. Estimatet for *logaritmen* til det forventede volumen for en hund med vægt 35 *kg* kan fås ved brug af `estimable()` på slutmodellen `modC` svarende til *bestillingslisten* `estC4` nedenfor

```
> modC <- lm(log(maxLA) ~ log(vgt), data = data1)
> estC4 <- c(1, log(35))
> library(gmodels)

> est <- rbind(Dalmatiner.35 = estC4)
> estimable(modC, est, conf.int = 0.95)

               Estimate Std. Error  t value DF Pr(>|t|) Lower.CI Upper.CI
Dalmatiner.35 2.980812 0.03814226 78.14985 80        0 2.904906 3.056717
```

Estimatet med 95 %-konfidensinterval for det forventede volumen af venstre forkammer for en `Dalmatiner` på 35 kg bliver 19.704 [18.264,21.258] mL.

4. Slutmodellen udtrykker, at den forventede værdi af den naturlige logaritme til hjertevolumen er givet ved
$$\alpha + \beta \cdot \texttt{log}(\texttt{vgt}).$$
Ved at tage eksponentialfunktionen får man, at den forventede værdi af hjertevolumen (`maxLA`) kan estimeres ved
$$\exp(\alpha + \beta \cdot \texttt{log}(\texttt{vgt})) = \underbrace{\exp(\alpha)}_{=\gamma} \cdot \texttt{vgt}^{\beta}.$$
Da estimatet for β har et 95 %-konfidensinterval, som indeholder tallet 1, kan man argumentere for, at det er rimeligt at antage, at
$$\mathbb{E}\texttt{maxLA} = \lambda \cdot \texttt{vgt}.$$
Tilsvarende kan man få et bud på en rimelig værdi af γ ved at tage eksponentialfunktionen på øvre og nedre grænse for konfidensintervallet for estimatet for α. Dette interval bliver: [0.464,0.615].

Opgave 2

1. Den statistiske analyse bør tage udgangspunkt i en model med en tilfældig effekt af `octagon` og en systematisk effekt af trefaktorvekselvirkningen `temp` $\times$ `drought` $\times$ `co2`. Startmodellen (`mod0` i `R`-udskriften) bliver således
$$\texttt{length}_i = \delta(\texttt{temp} \times \texttt{drought} \times \texttt{co2}_i) + b(\texttt{octagon}_i) + e_i,$$
hvor $b(1), \ldots, b(12)$ er uafhængige $\sim N(0, \sigma_b^2)$ og $e_1, \ldots, e_{48}$ er uafhængige $\sim N(0, \sigma^2)$.

2. Der er forskellige muligheder for, i hvilken rækkefølge modelreduktionen kan foretages. Man kan med fordel støtte sig til et faktordiagrammet (-som faktisk netop er faktordiagrammet fra Opgave 3, spørgsmål 3). Uanset fremgangsmåden, så når man frem til en slutmodel, hvor der kun er en hovedeffekt af `temp`.

   Nedenfor skitseres en mulig testrækkefølge. Ved besvarelsen refereres til `R`-udskriften nedenfor, der angiver et forslag til, hvordan et `R`-program til løsning af opgaven kunne se ud.

Test for effekt af trefaktorvekselvirkningen svarende til modellen (`mod1` i `R`-udskrift):

$$\texttt{length}_i = \alpha(\texttt{temp} \times \texttt{drought}_i) + \beta(\texttt{temp} \times \texttt{co2}_i) + \gamma(\texttt{drought} \times \texttt{co2}_i) + b(\texttt{octagon}_i) + e_i$$

Vi godkender hypotesen om, at trefaktorvekselvirkningen kan fjernes: $LR = 0.520, p = 0.471$.

Dernæst vælger jeg at fjerne de parvise vekselvirkninger i følgende rækkefølge: $\texttt{temp} \times \texttt{drought}$ ($LR = 0.800, p = 0.371$), $\texttt{temp} \times \texttt{co2}$ ($LR = 0.022, p = 0.312$), $\texttt{drought} \times \texttt{co2}$ ($LR = 1.580, p = 0.209$). Undervejs benyttes følgende statistiske modeller:

$$\begin{aligned}
\texttt{length}_i &= \beta(\texttt{temp} \times \texttt{co2}_i) + \gamma(\texttt{drought} \times \texttt{co2}_i) + b(\texttt{octagon}_i) + e_i \quad (\texttt{mod2a}) \\
\texttt{length}_i &= \delta(\texttt{temp}_i) + \gamma(\texttt{drought} \times \texttt{co2}_i) + b(\texttt{octagon}_i) + e_i \quad (\texttt{mod3a}) \\
\texttt{length}_i &= \delta(\texttt{temp}_i) + \mu(\texttt{drought}_i) + \nu(\texttt{co2}_i) + b(\texttt{octagon}_i) + e_i \quad (\texttt{mod4}).
\end{aligned}$$

Dernæst konstateres, at hovedeffekten af `co2` kan fjernes ($LR = 0.055, p = 0.815$), svarende til at vi har den additive model:

$$\texttt{length}_i = \delta(\texttt{temp}_i) + \nu(\texttt{drought}_i) + b(\texttt{octagon}_i) + e_i \quad (\texttt{mod5b}).$$

Endelig konstateres, at hovedeffekten af `drought` kan fjernes ($LR = 0.850, p = 0.357$) således, at vi ender op med følgende slutmodel:

$$\texttt{length}_i = \delta(\texttt{temp}_i) + b(\texttt{octagon}_i) + e_i \quad (\texttt{mod6a}),$$

hvor $b(1), \ldots, b(12)$ er uafhængige $\sim N(0, \sigma_b^2)$ og $e_1, \ldots, e_{48}$ er uafhængige $\sim N(0, \sigma^2)$.

Det kan bemærkes, at likelihoodratio teststørrelsen for test af om man kan fjerne af `temp` er $LR = 7.453$ svarende til en p-værdi på 0.006 (-se evt. `anova(mod7,mod6a)`).

3. Slutmodellen indeholder en systematisk effekt af `temp` og en tilfældig effekt af `octagon`.

   Parameterestimaterne for middelværdistrukturen bliver

   $$\hat{\delta}(\texttt{temp} = 0) = 154.19 \quad \hat{\delta}(\texttt{temp} = 1) = 101.75$$

   og parameterestimaterne for varianskomponenterne bliver

   $$\hat{\sigma}_B^2 = 13.8^2 = 190.33 \quad (\texttt{octagon}) \quad \hat{\sigma}^2 = 62.33^2 = 3885.29 \quad \text{(residualvarians)}.$$

   Konklusion er, at øget temperatur (`temp=1`) giver en lavere rodlængde. Den totale varians på en måling er givet ved 4075.62, hvoraf residualvariansen udgør 95.3 %.

4. Da der er foretaget gentagne målinger på hvert plot, vil det være naturligt at tage udgangspunkt i en model, som tillader en seriel korrelation mellem målinger hørende til samme plot. Man kunne desuden argumentere for det fornuftige i, at inddrage en tilfældig effekt af `octagon` ud over den tilfældig effekt af `plot`.

   På baggrund af `R`-udskriften kan man ved at se på `AIC` for de fire modeller se at Diggle-modellen lader til at være et godt udgangspunkt for den statistiske analayse. Modellen (`model1`) kan opskrives som

   $$\texttt{lenght}_i = \gamma(\texttt{temp} \times \texttt{drought} \times \texttt{co2} \times \texttt{session}_i) + A(\texttt{plot}_i) + D_i + e_i,$$

   hvor $A(1), \ldots, A(12)$ er uafhængige $\sim N(0, \nu^2)$ og e_i samt D_i er beskrevet ved korrelationsstrukturen som svarer til Diggle-modellen i kompendiets kapitel 10.3. Der lader ikke til at være brug for en tilfældig effekt af `octagon`.

Eksempel på R-kode som kunne være brugt til løsning af opgave 2

Koden er meget omfattende, fordi den skal benyttes som udgangspunkt for rettelse af eksamensopgaver. Det er langt fra nødvendigt, at man kører alle kommandoerne for at lave en fuldstændig besvarelse af opgaven.

```
> ### Indlæsning af data
> data2<-read.table(file="roots.txt",header=T)
> ### Da alle faktorer kun har 2 niveauer (0,1) er det ikke
> ### strengt nødvendigt at huske at lave dem om til faktorer.
> data2$teFac<-factor(data2$temp)
> data2$drFac<-factor(data2$drought)
> data2$co2Fac<-factor(data2$co2)
> head(data2)

  octagon co2 drought temp  length teFac drFac co2Fac
1       1   0       0    0 194.386     0     0      0
2       1   0       1    0 156.266     0     1      0
3       1   0       1    1  48.881     1     1      0
4       1   0       0    1  64.156     1     0      0
5      10   1       1    0 137.201     0     1      1
6      10   1       1    1 172.646     1     1      1

> ### Fjernelse af trefaktorvekselvirkning
> library(nlme)
> mod0<-lme(length~teFac*drFac*co2Fac,random=~1|octagon,method="ML",data=data2)
> mod1<-lme(length~teFac*drFac+teFac*co2Fac+drFac*co2Fac
+           ,random=~1|octagon,method="ML",data=data2)
> anova(mod1,mod0)

     Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod1     1  9 513.2809 529.5409 -247.6405
mod0     2 10 514.7612 532.8278 -247.3806 1 vs 2 0.5197666  0.4709

> ### Fjernelse af første parvise vekselvirkning
> ### Tre muligheder for reduktionsrækkefølge
> mod2a<-lme(length~teFac*co2Fac+drFac*co2Fac,random=~1|octagon,method="ML",data=data2)
> mod2b<-lme(length~teFac*drFac+drFac*co2Fac,random=~1|octagon,method="ML",data=data2)
> mod2c<-lme(length~teFac*drFac+teFac*co2Fac,random=~1|octagon,method="ML",data=data2)
> anova(mod2a,mod1) ### fjerner temp*drought

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod2a     1  8 512.0814 526.5347 -248.0407
mod1      2  9 513.2809 529.5409 -247.6405 1 vs 2 0.800417   0.371

> anova(mod2b,mod1) ### fjerner temp*co2
```

```
      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod2b     1  8 512.2803 526.7336 -248.1402
mod1      2  9 513.2809 529.5409 -247.6405 1 vs 2 0.9993748  0.3175

> anova(mod2c,mod1) ### fjerner drought*co2

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod2c     1  8 512.8544 527.3077 -248.4272
mod1      2  9 513.2809 529.5409 -247.6405 1 vs 2 1.573424  0.2097

> ### Fjernelse af anden parvise vekselvirkning
> ### Forskellige muligheder for reduktionsrækkefølge som
> ### afhænger af, hvilken effekt som blev fjernet ovenfor
> mod3a<-lme(length~teFac+drFac*co2Fac,random=~1|octagon,method="ML",data=data2)
> mod3b<-lme(length~drFac+teFac*co2Fac,random=~1|octagon,method="ML",data=data2)
> mod3c<-lme(length~co2Fac+teFac*drFac,random=~1|octagon,method="ML",data=data2)
> anova(mod3a,mod2a) ### fjerner temp*co2

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod3a     1  7 511.1034 523.7500 -248.5517
mod2a     2  8 512.0814 526.5347 -248.0407 1 vs 2 1.022027   0.312

> anova(mod3a,mod2b) ### fjerner temp*drought

      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod3a     1  7 511.1034 523.7500 -248.5517
mod2b     2  8 512.2803 526.7336 -248.1402 1 vs 2 0.8230693  0.3643

> anova(mod3b,mod2a) ### fjerner drought*co2

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod3b     1  7 511.6800 524.3267 -248.8400
mod2a     2  8 512.0814 526.5347 -248.0407 1 vs 2 1.598686  0.2061

> anova(mod3b,mod2c) ### fjerner temp*drought

      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod3b     1  7 511.6800 524.3267 -248.8400
mod2c     2  8 512.8544 527.3077 -248.4272 1 vs 2 0.8256792  0.3635

> anova(mod3c,mod2b) ### fjerner drought*co2

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod3c     1  7 511.8364 524.4831 -248.9182
mod2b     2  8 512.2803 526.7336 -248.1402 1 vs 2 1.556129  0.2122
```

```
> anova(mod3c,mod2c) ### fjerner temp*co2

      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod3c     1  7 511.8364 524.4831 -248.9182
mod2c     2  8 512.8544 527.3077 -248.4272 1 vs 2 0.9820802  0.3217

> ### På dette tidspunkt er der mulighed for ENTEN af fjerne den sidste
> ### vekselvirkning ELLER fjerne den hovedvirkning, som ikke indgår i en
> ### vekselvirkning.
>
> ### Fjernelse af sidste vekselvirkning
> mod4<-lme(length~teFac+drFac+co2Fac,random=~1|octagon,method="ML",data=data2)
> anova(mod4,mod3a) ### fjerner drought*co2

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod4      1  6 510.6837 521.5237 -249.3419
mod3a     2  7 511.1034 523.7500 -248.5517 1 vs 2 1.580339  0.2087

> anova(mod4,mod3b) ### fjerner temp*co2

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod4      1  6 510.6837 521.5237 -249.3419
mod3b     2  7 511.6800 524.3267 -248.8400 1 vs 2 1.003680  0.3164

> anova(mod4,mod3c) ### fjerner temp*drought

      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod4      1  6 510.6837 521.5237 -249.3419
mod3c     2  7 511.8364 524.4831 -248.9182 1 vs 2 0.8472787  0.3573

> ### Alternativt: fjernelse af hovedeffekt
> mod4a<-lme(length~drFac*co2Fac,random=~1|octagon,method="ML",data=data2)
> mod4b<-lme(length~teFac*co2Fac,random=~1|octagon,method="ML",data=data2)
> mod4c<-lme(length~teFac*drFac,random=~1|octagon,method="ML",data=data2)
> anova(mod4a,mod3a) ### fjerner temp

      Model df      AIC      BIC    logLik   Test L.Ratio p-value
mod4a     1  6 516.7739 527.6139 -252.3870
mod3a     2  7 511.1034 523.7500 -248.5517 1 vs 2 7.67052  0.0056

> anova(mod4b,mod3b) ### fjerner drought

      Model df      AIC      BIC    logLik   Test L.Ratio p-value
mod4b     1  6 510.4944 521.3344 -249.2472
mod3b     2  7 511.6800 524.3267 -248.8400 1 vs 2  0.8144  0.3668
```

```
> anova(mod4c,mod3c) ### fjerner co2

      Model df      AIC      BIC    logLik   Test    L.Ratio p-value
mod4c     1  6 509.8830 520.7229 -248.9415
mod3c     2  7 511.8364 524.4831 -248.9182 1 vs 2 0.04652188  0.8292

> ### Med udgangspunkt i mod4 kan man nu fjerne hovedvirkninger i en
> ### passende rækkefølge.
> mod5a<-lme(length~drFac+co2Fac,random=~1|octagon,method="ML",data=data2)
> mod5b<-lme(length~teFac+co2Fac,random=~1|octagon,method="ML",data=data2)
> mod5c<-lme(length~teFac+drFac,random=~1|octagon,method="ML",data=data2)
> anova(mod5a,mod4) ### fjerner temp

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod5a     1  5 516.1391 525.1725 -253.0696
mod4      2  6 510.6837 521.5237 -249.3419 1 vs 2 7.455428  0.0063

> anova(mod5b,mod4) ### fjerner drought

      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod5b     1  5 509.5243 518.5576 -249.7622
mod4      2  6 510.6837 521.5237 -249.3419 1 vs 2 0.8406095  0.3592

> anova(mod5c,mod4) ### fjerner co2

      Model df      AIC      BIC    logLik   Test    L.Ratio p-value
mod5c     1  5 508.7382 517.7715 -249.3691
mod4      2  6 510.6837 521.5237 -249.3419 1 vs 2 0.05450119  0.8154

> ### Alternativt kan man med udgangspunkt i mod4a, mod4b eller mod4c
> ### fjerne den sidste vekselvirkning
> anova(mod5a,mod4a) ### fjerner drought*co2

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod5a     1  5 516.1391 525.1725 -253.0696
mod4a     2  6 516.7739 527.6139 -252.3870 1 vs 2 1.365246  0.2426

> anova(mod5b,mod4b) ### fjerner temp*co2

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod5b     1  5 509.5243 518.5576 -249.7622
mod4b     2  6 510.4944 521.3344 -249.2472 1 vs 2 1.029889  0.3102

> anova(mod5c,mod4c) ### fjerner temp*drought
```

```
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod5c     1  5 508.7382 517.7715 -249.3691
mod4c     2  6 509.8830 520.7229 -248.9415 1 vs 2 0.855258  0.3551

> ### Fjernelse af endnu en hovedvirkning med udgangspunkt i
> ### additiv model med to hovedvirkninger
> mod6a<-lme(length~teFac,random=~1|octagon,method="ML",data=data2)
> mod6b<-lme(length~drFac,random=~1|octagon,method="ML",data=data2)
> mod6c<-lme(length~co2Fac,random=~1|octagon,method="ML",data=data2)
> anova(mod6a,mod5b) ### fjerner co2

      Model df      AIC      BIC    logLik   Test    L.Ratio p-value
mod6a     1  4 507.5878 514.8145 -249.7939
mod5b     2  5 509.5243 518.5576 -249.7622 1 vs 2 0.06350001   0.801

> anova(mod6a,mod5c) ### fjerner drought

      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod6a     1  4 507.5878 514.8145 -249.7939
mod5c     2  5 508.7382 517.7715 -249.3691 1 vs 2 0.8496083  0.3567

> anova(mod6b,mod5a) ### fjerner co2

      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod6b     1  4 514.2081 521.4347 -253.1041
mod5a     2  5 516.1391 525.1725 -253.0696 1 vs 2 0.0689407  0.7929

> anova(mod6b,mod5c) ### fjerner temp

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod6b     1  4 514.2081 521.4347 -253.1041
mod5c     2  5 508.7382 517.7715 -249.3691 1 vs 2 7.469867  0.0063

> anova(mod6c,mod5a) ### fjerner drought

      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod6c     1  4 514.9618 522.1885 -253.4809
mod5a     2  5 516.1391 525.1725 -253.0696 1 vs 2 0.8226966  0.3644

> anova(mod6c,mod5b) ### fjerner temp

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod6c     1  4 514.9618 522.1885 -253.4809
mod5b     2  5 509.5243 518.5576 -249.7622 1 vs 2 7.437515  0.0064
```

```
> ### Fjernelse af sidste hovedvirkning
> mod7<-lme(length~1,random=~1|octagon,method="ML",data=data2)
> anova(mod7,mod6a) ### fjerner temp

      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod7      1  3 513.0412 518.4612 -253.5206
mod6a     2  4 507.5878 514.8145 -249.7939 1 vs 2 7.453362  0.0063

> anova(mod7,mod6b) ### fjerner drought

      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
mod7      1  3 513.0412 518.4612 -253.5206
mod6b     2  4 514.2081 521.4347 -253.1041 1 vs 2 0.8331031  0.3614

> anova(mod7,mod6c) ### fjerner co2

      Model df      AIC      BIC    logLik   Test    L.Ratio p-value
mod7      1  3 513.0412 518.4612 -253.5206
mod6c     2  4 514.9618 522.1885 -253.4809 1 vs 2 0.07934727  0.7782

> ### Genfitter slutmodellen
> mod6arefit<-lme(length~teFac-1,random=~1|octagon,method="REML",data=data2)
> mod6arefit

Linear mixed-effects model fit by REML
  Data: data2
  Log-restricted-likelihood: -242.7484
  Fixed: length ~ teFac - 1
  teFac0   teFac1
154.1885 101.7468

Random effects:
 Formula: ~1 | octagon
        (Intercept) Residual
StdDev:    13.79609 62.33209

Number of Observations: 45
Number of Groups: 12
```

Opgave 3

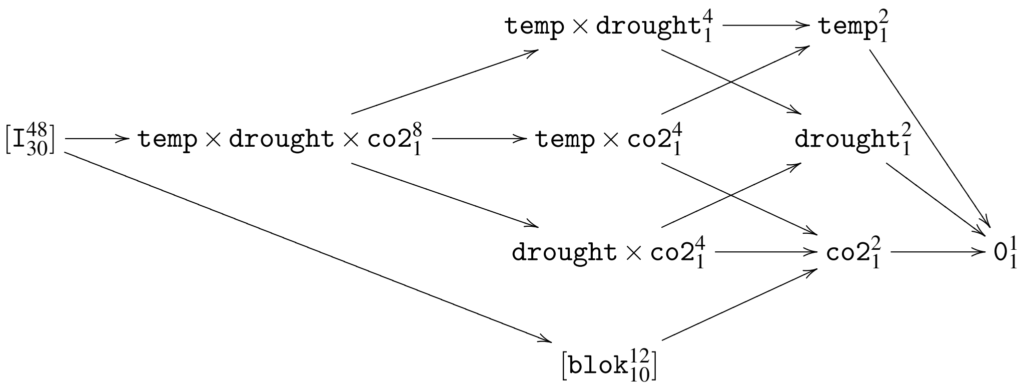
1. Der er tale om et 2^3-forsøg udført på to blokke af størrelse 4. Hvis 3-faktorvekselvirkningen `temp` $\times$ `drought` $\times$ `co2` skal konfunderes med blok, så skal de 8 behandlingskombinationer grupperes i to grupper i henhold til følgende skema

temp	drought	co2	temp + drought + co2	blok 1	blok 2
1	1	1	3	x	
1	1	2	4		x
1	2	1	4		x
1	2	2	5	x	
2	1	1	4		x
2	1	2	5	x	
2	2	1	5	x	
2	2	2	6		x

2. Det foreslåede blokforsøg skal have $v_T = 8$ behandlinger, $v_B = 12$ blokke og en blokstørrelse $r_B = 4$. Ifølge kompendiets Theorem 9.6 kræver dette, at hver behandling skal afprøves
$$r_T = \frac{r_B \cdot v_B}{v_T} = \frac{4 \cdot 12}{8} = 6$$
gange i forsøgsplanen. Endvidere skal hvert par af behandlinger forkomme
$$\lambda = \frac{r_T \cdot (r_B - 1)}{v_T - 1} = \frac{6 \cdot (4 - 1))}{8 - 1} = \frac{18}{7}$$
gange inden for samme blot i forsøgsplanen. Da λ *ikke* er et helt tal, kan vi konkludere, at det i det konkrete tilfælde ikke er muligt at lave et balanceret ufuldstændigt blokforsøg (BIBD).

3. Det foreslåede forsøg er et split-plot design med blokke som helplots, `co2` som helplot-faktor og vekselvirkningen `temp` $\times$ `drought` som delplotfaktor.

   Faktordiagrammet ser ud som følger



   Data fra forsøget bør analyseres med en model af formen
$$y_i = \alpha(\texttt{co2} \times \texttt{temp} \times \texttt{drought}_i) + A(\texttt{blok}_i) + e_i,$$
hvor $A(1), \ldots, A(12)$ er uafhængige $\sim N(0, \sigma_B^2)$ og $e_1, \ldots, e_{48}$ er uafhængige $\sim B(0, \sigma^2)$.

   Randomiseringen foretages i to trin: (i) først udvælges ved lodtrækning de 6 blokke, som skal modtage øget tilførsel af CO_2 (`co2`=1), (ii) dernæst fordeles de 4 kombinationer af `temp` $\times$ `drought` tilfældigt ud på de 4 forsøgsenheder inden for hver af de 12 blokke.