

# SD2 - løsning til case 6

*Anders Tolver*

*5 Oct 2017*

Dette dokument indeholder supplerende kommentarer til filen **casesol6.pdf** og vedrører løsningen af case 6 programsat torsdag i kursusuge 5.

## 1. Indlæsning af data

```
data <- read.table(file = "../data/case6.txt", header = T)
head(data)
```

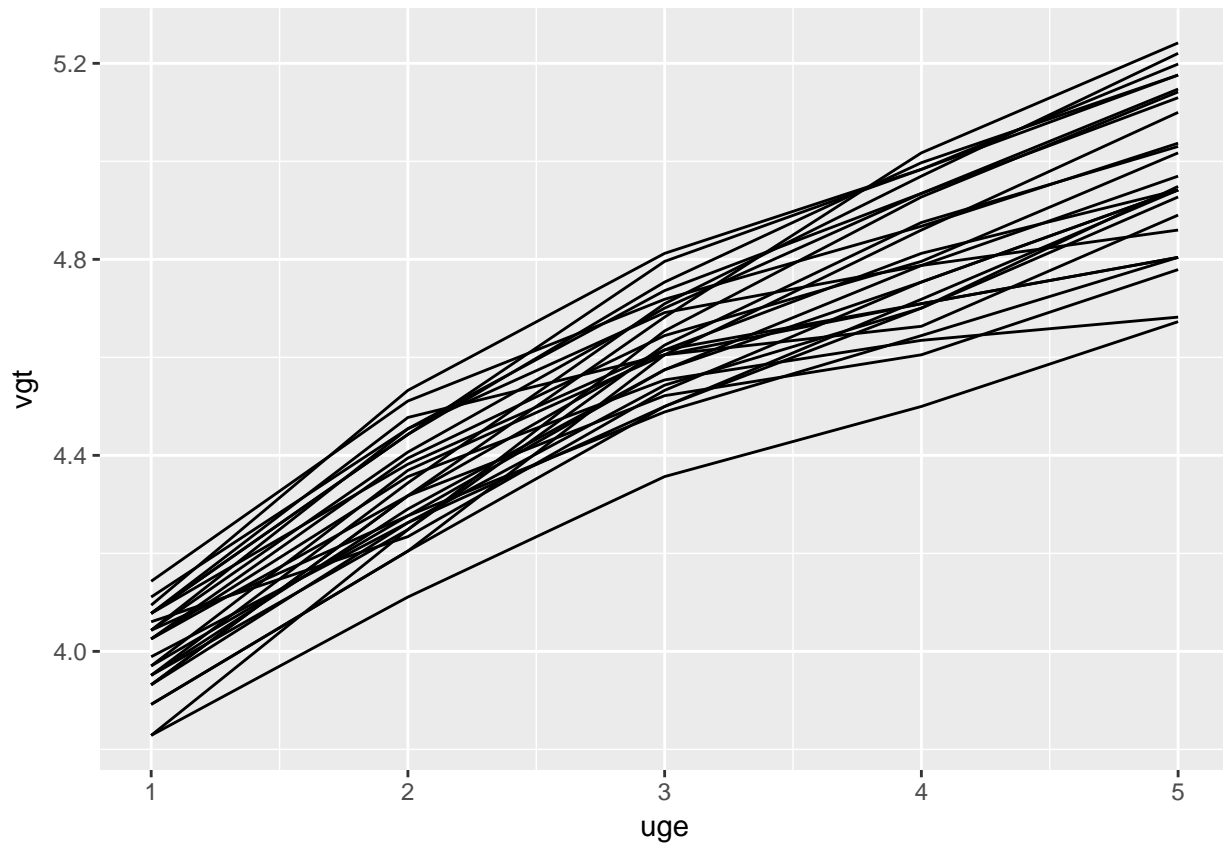
```
##   rotte uge    beh    vgt
## 1     1   1 Kontrol 4.0431
## 2     2   1 Kontrol 4.0943
## 3     3   1 Kontrol 3.9512
## 4     4   1 Kontrol 3.8918
## 5     5   1 Kontrol 4.0254
## 6     6   1 Kontrol 3.8286
```

```
data$rotte <- factor(data$rotte)
data$ugefac <- factor(data$uge)
```

## 2.

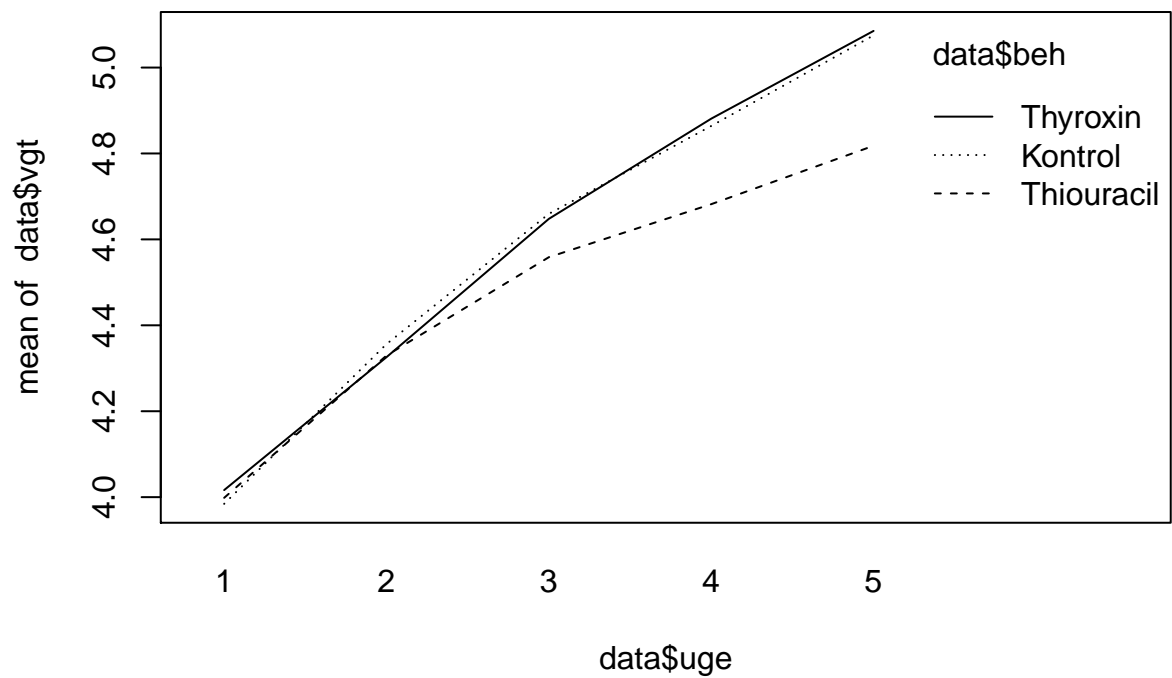
I stedet for at køre R-koden fra opgaveformuleringen kan følgende løsning vælges.

```
library(tidyverse)
ggplot(data = data) + geom_line(aes(x = uge, y = vgt, group = rotte))
```



3.

```
interaction.plot(data$uge, data$beh, data$vgt)
```



## 4.-5. Digglemodel

```
library(nlme)
model0 <- lme(vgt ~ beh * ugefac, random = ~ 1 | rotte, data = data
, corr = corGaus(form = ~ uge | rotte, nugget = T), method = "ML")
```

## 6.

Her er flere muligheder

- test om vekselvirkningen mellem **beh** og **ugefac** kan fjernes
- test om det er rimeligt lave **uge** indgå i modellen som en kovariat (svarende til at der er en lineær sammenhæng mellem **uge** og responsen **vgt**).

## 7. Sammenligning af kontrol- og thyroxinrotter

Vi benytter en fremgangsmåde, hvor vi konstruerer en *ny* behandlingsfaktor med kun 2 niveauer, hvor rotter som har modtaget **kontrol**- og **thyroxin**-behandling kommer i samme grupper. Forskellige løsningsmetoder er angivet.

Løsning som anført i *hint*

```
data$nybeh <- rep(1, 135)
data$nybeh[data$beh == "Thiouracil"] = 2
data$nybeh <- factor(data$nybeh)
table(data$beh, data$nybeh)
```

```
##
##           1  2
## Kontrol   50  0
## Thiouracil 0 50
## Thyroxin  35  0
```

Løsning, der benytter samme ide som i *hint*, men hvor niveauerne kaldes TRUE og FALSE

```
data$nybeh2 <- (data$beh == "Thiouracil")
table(data$beh, data$nybeh2)
```

```
##
##           FALSE TRUE
## Kontrol       50   0
## Thiouracil     0  50
## Thyroxin      35   0
```

Løsning, hvor niveauerne af behandlingsfaktoren ændres manuelt

```
data$nybeh3 <- factor(data$beh)
levels(data$nybeh3)
```

```
## [1] "Kontrol" "Thiouracil" "Thyroxin"
levels(data$nybeh3) <- c("KonThy", "Thiouracil", "KonThy")
table(data$beh, data$nybeh3)
```

```
##
##           KonThy Thiouracil
##   Kontrol      50          0
##   Thiouracil    0          50
##   Thyroxin     35          0
```

Test for om grupperne **Kontrol** og **Thyroxin** kan slås sammen

```
modelny <- lme(vgt ~ nybeh3 * ugefac, random = ~ 1 | rotte, data = data
, corr = corGaus(form = ~ uge | rotte, nugget = T), method = "ML")
anova(modelny, model0)
```

```
##           Model df          AIC          BIC   logLik   Test  L.Ratio p-value
## modelny       1 14 -380.6785 -340.0046 204.3392
## model0        2 19 -380.5067 -325.3065 209.2533 1 vs 2 9.828194 0.0803
```

## 8. Forskelle til hver af de enkelte uger

Med lidt snilde kan man få R til en anvende en parametrisering af modellen **modelny** fra spørgsmål 7., så man direkte kan aflæse forskelle mellem grupperne til hver af de 5 uger. Bemærk i øvrigt, at modellen genfittes med **method == "REML"**.

```
modelny_refit <- lme(vgt ~ ugefac + nybeh3 : ugefac - 1, random = ~ 1 | rotte, data = data
, corr = corGaus(form = ~ uge | rotte, nugget = T), method = "REML")
round(summary(modelny_refit)$tTable, digits = 4)
```

```
##           Value Std.Error DF   t-value p-value
## ugefac1      3.9976   0.0232 99 172.3652 0.0000
## ugefac2      4.3438   0.0232 99 187.2927 0.0000
## ugefac3      4.6547   0.0232 99 200.6963 0.0000
## ugefac4      4.8710   0.0232 99 210.0233 0.0000
## ugefac5      5.0792   0.0232 99 218.9991 0.0000
## ugefac1:nybeh3Thiouracil 0.0009   0.0381 99   0.0225 0.9821
## ugefac2:nybeh3Thiouracil -0.0142   0.0381 99  -0.3716 0.7110
## ugefac3:nybeh3Thiouracil -0.0961   0.0381 99  -2.5221 0.0133
## ugefac4:nybeh3Thiouracil -0.1890   0.0381 99  -4.9595 0.0000
## ugefac5:nybeh3Thiouracil -0.2612   0.0381 99  -6.8546 0.0000
```

Heraf aflæses direkte, at **Thiouracil**-rotterne først ligger signifikant under rotter fra de øvrige to behandlingsgrupper fra uge 3 (-se på p-værdier for det relevante test).

## 8.-9. Løsning ved brug af estimable()

Neden for benyttes **estimable** til at besvare de spørgsmål der ønskes besvaret under punkt 8. og 9. Her benyttes ikke en *smart* parametrisering af modellerne.

```
modelny_refit2 <- lme(vgt ~ nybeh3 * ugefac, random = ~ 1 | rotte, data = data
, corr = corGaus(form = ~ uge | rotte, nugget = T), method = "REML")
summary(modelny_refit2)
```

```
## Linear mixed-effects model fit by REML
## Data: data
##           AIC          BIC   logLik
##   -315.1069 -275.5105 171.5534
```

```

##
## Random effects:
## Formula: ~1 | rotte
##      (Intercept)  Residual
## StdDev:  0.05764853 0.07629465
##
## Correlation Structure: Gaussian spatial correlation
## Formula: ~uge | rotte
## Parameter estimate(s):
##      range      nugget
## 2.09802710 0.05500271
## Fixed effects: vgt ~ nybeh3 * ugefac
##
##              Value Std.Error DF   t-value p-value
## (Intercept)      3.997594 0.02319258 100 172.36522 0.0000
## nybeh3Thiouracil      0.000856 0.03810930 25   0.02246 0.9823
## ugefac2              0.346206 0.01300701 100  26.61686 0.0000
## ugefac3              0.657071 0.02059099 100  31.91059 0.0000
## ugefac4              0.873388 0.02451637 100  35.62469 0.0000
## ugefac5              1.081559 0.02584054 100  41.85511 0.0000
## nybeh3Thiouracil:ugefac2 -0.015016 0.02137270 100  -0.70257 0.4840
## nybeh3Thiouracil:ugefac3 -0.096971 0.03383445 100  -2.86603 0.0051
## nybeh3Thiouracil:ugefac4 -0.189858 0.04028451 100  -4.71293 0.0000
## nybeh3Thiouracil:ugefac5 -0.262079 0.04246034 100  -6.17232 0.0000
## Correlation:
##              (Intr) nybh3T ugefc2 ugefc3 ugefc4 ugefc5 ny3T:2
## nybeh3Thiouracil      -0.609
## ugefac2              -0.280  0.171
## ugefac3              -0.444  0.270  0.792
## ugefac4              -0.529  0.322  0.543  0.848
## ugefac5              -0.557  0.339  0.351  0.627  0.868
## nybeh3Thiouracil:ugefac2  0.171 -0.280 -0.609 -0.482 -0.330 -0.214
## nybeh3Thiouracil:ugefac3  0.270 -0.444 -0.482 -0.609 -0.516 -0.382  0.792
## nybeh3Thiouracil:ugefac4  0.322 -0.529 -0.330 -0.516 -0.609 -0.528  0.543
## nybeh3Thiouracil:ugefac5  0.339 -0.557 -0.214 -0.382 -0.528 -0.609  0.351
##              ny3T:3 ny3T:4
## nybeh3Thiouracil
## ugefac2
## ugefac3
## ugefac4
## ugefac5
## nybeh3Thiouracil:ugefac2
## nybeh3Thiouracil:ugefac3
## nybeh3Thiouracil:ugefac4  0.848
## nybeh3Thiouracil:ugefac5  0.627  0.868
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -1.77411499 -0.68094671  0.09841366  0.65444213  1.81956919
##
## Number of Observations: 135
## Number of Groups: 27

```

```

thi1 <- c(1, 1, 0, 0, 0, 0, 0, 0, 0, 0)
thi2 <- c(1, 1, 1, 0, 0, 0, 1, 0, 0, 0)
thi3 <- c(1, 1, 0, 1, 0, 0, 0, 1, 0, 0)
thi4 <- c(1, 1, 0, 0, 1, 0, 0, 0, 1, 0)
thi5 <- c(1, 1, 0, 0, 0, 1, 0, 0, 0, 1)
kon1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0)
kon2 <- c(1, 0, 1, 0, 0, 0, 0, 0, 0, 0)
kon3 <- c(1, 0, 0, 1, 0, 0, 0, 0, 0, 0)
kon4 <- c(1, 0, 0, 0, 1, 0, 0, 0, 0, 0)
kon5 <- c(1, 0, 0, 0, 0, 1, 0, 0, 0, 0)
diff1 <- kon1 - thi1
diff2 <- kon2 - thi2
diff3 <- kon3 - thi3
diff4 <- kon4 - thi4
diff5 <- kon5 - thi5
kon_incr1_5 <- kon5 - kon1
thi_incr1_5 <- thi5 - thi1
diff_incr1_5 <- kon_incr1_5 - thi_incr1_5

est <- rbind(kon1, kon2, kon3, kon4, kon5
             , thi1, thi2, thi3, thi4, thi5
             , diff1, diff2, diff3, diff4, diff5
             , kon_incr1_5, thi_incr1_5, diff_incr1_5)
library(gmodels)
round(estimable(modelny_refit2, est, conf.int = 0.95), digits = 4)

## Warning in estimable.default(modelny_refit2, est, conf.int = 0.95): Degrees
## of freedom vary among parameters used to construct linear contrast(s):
## 6, 7, 8, 9, 10, 12, 13, 14, 15. Using the smallest df among the set of
## parameters.

##           Estimate Std. Error  t value  DF Pr(>|t|) Lower.CI Upper.CI
## kon1          3.9976    0.0232 172.3652 100  0.0000   3.9516   4.0436
## kon2          4.3438    0.0232 187.2927 100  0.0000   4.2978   4.3898
## kon3          4.6547    0.0232 200.6963 100  0.0000   4.6087   4.7007
## kon4          4.8710    0.0232 210.0233 100  0.0000   4.8250   4.9170
## kon5          5.0792    0.0232 218.9991 100  0.0000   5.0331   5.1252
## thi1          3.9984    0.0302 132.2264  25  0.0000   3.9362   4.0607
## thi2          4.3296    0.0302 143.1787  25  0.0000   4.2674   4.3919
## thi3          4.5586    0.0302 150.7486  25  0.0000   4.4963   4.6208
## thi4          4.6820    0.0302 154.8303  25  0.0000   4.6197   4.7443
## thi5          4.8179    0.0302 159.3261  25  0.0000   4.7557   4.8802
## diff1         -0.0009    0.0381  -0.0225  25  0.9823  -0.0793   0.0776
## diff2          0.0142    0.0381   0.3716  25  0.7133  -0.0643   0.0926
## diff3          0.0961    0.0381   2.5221  25  0.0184   0.0176   0.1746
## diff4          0.1890    0.0381   4.9595  25  0.0000   0.1105   0.2675
## diff5          0.2612    0.0381   6.8546  25  0.0000   0.1827   0.3397
## kon_incr1_5    1.0816    0.0258  41.8551 100  0.0000   1.0303   1.1328
## thi_incr1_5    0.8195    0.0337  24.3227 100  0.0000   0.7526   0.8863
## diff_incr1_5   0.2621    0.0425   6.1723 100  0.0000   0.1778   0.3463

```

Konklusioner vedrørende forskellen mellem vægten for kontrolrotter og Thiouracil-rotter

- forskellen øges (stort set monotont) fra uge 1 til uge 5
- forskellene er signifikante fra uge 3 (-se på P-værdierne eller bemærk at 95 % - konfidensintervaller for

forskellene *ikke* indeholder null!)

Et par hovedkonklusioner omkring tilvæksten fra uge 1 til uge 5

- tilvæksten i kontrolgruppen estimeres til 1.0816 [95%-KI: 1.0303-1.1328]
- tilvæksten i Thiouracil-gruppen estimeres til 0.8195 [95%-KI: 0.7526-0.8863]
- tilvæksten er således 0.2621 højere i kontrolgruppen end i Thiouracil-gruppen [95%-KI: 0.1778-0.3463]

## 10.

Estimaterne for variansparametrene ses ved at lave et summary af slutmodellen (`summary(modelny_refit2)`). Se yderligere information på slide 24-25 fra dagens forelæsning.

## 11.

Her vises to forskellige konstruktioner, som begge laver et datasæt indeholdende tilvæksten fra uge 1 til uge 5 som *summary measure*.

```
data1 <- subset(data, uge == 1) ### udvælg målinger fra uge 1
data5 <- subset(data, uge == 5) ### udvælg målinger fra uge 5
data1$vgt5 <- data5$vgt ### tilføj måling fra uge 5 til det reducerede datasæt
data1$incr15 <- data1$vgt5 - data1$vgt ### konstruer tilvæksten fra uge 1 til 5
head(data1)
```

```
##   rotte uge   beh   vgt ugefac nybeh nybeh2 nybeh3   vgt5 incr15
## 1     1   1 Kontrol 4.0431     1     1 FALSE KonThy 5.1475 1.1044
## 2     2   1 Kontrol 4.0943     1     1 FALSE KonThy 5.1761 1.0818
## 3     3   1 Kontrol 3.9512     1     1 FALSE KonThy 5.2204 1.2692
## 4     4   1 Kontrol 3.8918     1     1 FALSE KonThy 5.0999 1.2081
## 5     5   1 Kontrol 4.0254     1     1 FALSE KonThy 5.0173 0.9919
## 6     6   1 Kontrol 3.8286     1     1 FALSE KonThy 5.0304 1.2018
```

Alternativ løsning (som bruger værktøjer fra **tidyverse** R-pakken).

```
head(data)
```

```
##   rotte uge   beh   vgt ugefac nybeh nybeh2 nybeh3
## 1     1   1 Kontrol 4.0431     1     1 FALSE KonThy
## 2     2   1 Kontrol 4.0943     1     1 FALSE KonThy
## 3     3   1 Kontrol 3.9512     1     1 FALSE KonThy
## 4     4   1 Kontrol 3.8918     1     1 FALSE KonThy
## 5     5   1 Kontrol 4.0254     1     1 FALSE KonThy
## 6     6   1 Kontrol 3.8286     1     1 FALSE KonThy
```

```
data_ny <- select(data, rotte, uge, beh, nybeh3, vgt)
data_ny <- spread(data_ny, key = uge, value = vgt)
names(data_ny)[4:8] <- paste("vgt", 1:5, sep = "")
data_ny <- mutate(data_ny, incr15 = vgt5 - vgt1)
```

```
## Warning: package 'bindrcpp' was built under R version 3.3.2
```

```
head(data_ny)
```

```
##   rotte   beh nybeh3   vgt1   vgt2   vgt3   vgt4   vgt5 incr15
## 1     1 Kontrol KonThy 4.0431 4.4543 4.7362 4.9345 5.1475 1.1044
```

```
## 2      2 Kontrol KonThy 4.0943 4.5326 4.8122 4.9836 5.1761 1.0818
## 3      3 Kontrol KonThy 3.9512 4.3438 4.7095 4.9698 5.2204 1.2692
## 4      4 Kontrol KonThy 3.8918 4.2047 4.6052 4.8598 5.0999 1.2081
## 5      5 Kontrol KonThy 4.0254 4.3944 4.6444 4.7958 5.0173 0.9919
## 6      6 Kontrol KonThy 3.8286 4.2485 4.6250 4.8752 5.0304 1.2018
```

## 12.

Vi laver en ensidet variansanalyse med henblik på at estimere tilvæksterne fra uge 1 til uge 5 og forskellen i tilvækster mellem kontrol-rotter og thiouracil-rotter.

Først inddrages behandlingsfaktoren med 3 niveauer (**beh**)

```
m1 <- lm(incr15 ~ beh, data = data1)
summary(m1)

##
## Call:
## lm(formula = incr15 ~ beh, data = data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19657 -0.08570  0.01393  0.08112  0.19032
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.09047    0.03523  30.951 < 2e-16 ***
## behThiouracil -0.27099    0.04983  -5.439 1.37e-05 ***
## behThyroxin   -0.02164    0.05490  -0.394  0.697
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1114 on 24 degrees of freedom
## Multiple R-squared:  0.5932, Adjusted R-squared:  0.5593
## F-statistic: 17.5 on 2 and 24 DF, p-value: 2.055e-05
```

Vi konkluderer bl.a. at forskellen i tilvækster for **Thyroxin** og **Kontrol** er - 0.02164 og at denne forskel *ikke* er signifikant (P=0.697).

En formelt test giver samme P-værdi

```
m2 <- lm(incr15 ~ nybeh3, data = data1)
anova(m2, m1)

## Analysis of Variance Table
##
## Model 1: incr15 ~ nybeh3
## Model 2: incr15 ~ beh
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      25 0.29984
## 2      24 0.29791  1 0.0019285 0.1554 0.6969
```

Endelig udskrives estimatorne ...

```
summary(m2)
```

```
##
```



```
## Call:
## lm(formula = incr15 ~ nybeh3, data = data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18766 -0.09041  0.01422  0.08112  0.19032
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.08156    0.02656  40.720 < 2e-16 ***
## nybeh3Thiouracil -0.26208    0.04364  -6.005 2.85e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1095 on 25 degrees of freedom
## Multiple R-squared:  0.5906, Adjusted R-squared:  0.5742
## F-statistic: 36.06 on 1 and 25 DF,  p-value: 2.85e-06

confint(m2)

##              2.5 %      97.5 %
## (Intercept)      1.0268551  1.1362626
## nybeh3Thiouracil -0.3519663 -0.1721914
```

Vi konkluderer at

- tilvæksten fra uge 1 til uge 5 for kontrolgruppen estimeres til 1.0816 [95%-KI: 1.0268-1.1363]
- tilvæksten fra uge 1 til uge 5 for Thiouracil-grupper er 0.2621 lavere [95%-KI: 0.1722-0.3520]

Diskussion / sammenligning med resulaterne fra Diggle-modellen:

Diggle-modellen inddrager alle målinger på en gang og fitter en samlet statistisk model. På baggrund af denne ene model er det muligt at udtrække estimater for mange interessant forskelle (mellem behandlinger til et givet tidspunkt eller mellem tilvækster fra uge 1 til uge 5). Resultater og konfidensintervaller er (næsten) identiske med resultaterne som fås ved at lave en analyse af et *summary measure* (her: tilvæksten fra uge 1 til uge 5). Sådan bør det naturligvis også være!