

Menarke i Nord-Norge

HLJ

4/3/2022

Bakgrunn

Definerer grupper og variabler

Menarkealder

- med og uten at "NA" for menarke-måned er omdefinert til =0

```
ff_uttrekk14062021_prno774000072 <- read.csv("~/Documents/ff_uttrekk14062021_
prno774000072/ff_uttrekk14062021_prno774000072.csv")
View(ff_uttrekk14062021_prno774000072)

## Warning in system2("/usr/bin/otool", c("-L", shQuote(DSO)), stdout = TRUE)
:
## running command ''/usr/bin/otool' -L '/Library/Frameworks/R.framework/Reso
urces/
## modules/R_de.so'' had status 1

data <- ff_uttrekk14062021_prno774000072 #uten menarke-mnd NA omdefinert til
0
Data <- ff_uttrekk14062021_prno774000072
Data$MENARCHE_AGE_MONTH_FF1[is.na(Data$MENARCHE_AGE_MONTH_FF1)] <- 6 #omdefin
ert
#menarke-mnd NA definert som lik 0

Data$menarke_alder <- (Data$MENARCHE_AGE_YEAR_FF1 + Data$MENARCHE_AGE_MONTH_F
F1/12)
#variabel for nøyaktig alder for menarke
```

Deskriptiv statistikk for populasjonen vår

Menarke

- hvor mange har startet å menstruere ved undersøkelsestidspunkt?

```
table(Data$SEX_FF1)
```

```
##
## 0 1
## 508 530
```

```

table(Data$MENSES_FF1) #3 har ikke startet å menstruere,
##
##    0    1
##   3 504

#mens 504 har startet å menstruere, 1 missing
table(Data$MENARCHE_FF1) #Men kun 495 har sagt at de har hatt menarche/
##
##    0    1
##   5 495

#første menstruasjon. 5 har svart at de olle har hatt menarke. Dvs 8 NA.

```

Kommentar:

504 har svart at de har menstruasjon, men kun 495 har svart at de har hatt menarke. Totalt hadde 507 svart på spm om menstruasjon, mens 500 svarte på spm om menarke. Har altså 8 missing/ NA. Greit å vite.

IOTF / aldersjustert vektklasse målt ved førskolealder

```

#Trenger kun jenter i settet nå
Data <- Data[ which(Data$SEX_FF1==0), ]

#Definerer de ulike aldersintervallene
library(data.table)
agebreaks <- c(0,4.5,5,5.5,6,6.5,7,7.5,8,10)
agelabels <- c("4.0-4.5", "4.5-4.9", "5.0-5.4", "5.5-5.9", "6.0-6.4",
              "6.5-6.9", "7.0-7.4", "7.5-7.9", "8.0+")
setDT(Data)[ , agegroups := cut(EXACT_AGE_PRECHOOLCONT_FF1,
                               breaks = agebreaks,
                               right = FALSE,
                               labels = agelabels)]

table(Data$agegroups) #Ser bra ut
##
## 4.0-4.5 4.5-4.9 5.0-5.4 5.5-5.9 6.0-6.4 6.5-6.9 7.0-7.4 7.5-7.9 8.0+
##      5      0      18      143      148      50      11      6      4

# BMI-cutoffs til 9 aldersgrupper

#Undervektig for alle aldre
Data[Data$BMI_PRECHOOLCONT_FF1 < 13.815 & Data$agegroups=="4.0-4.5" ,
      "Vektstatus"] <- "undervektig"
Data[Data$BMI_PRECHOOLCONT_FF1 < 13.685 & Data$agegroups=="4.5-4.9" ,
      "Vektstatus"] <- "undervektig"
Data[Data$BMI_PRECHOOLCONT_FF1 < 13.565 & Data$agegroups=="5.0-5.4" ,
      "Vektstatus"] <- "undervektig"
Data[Data$BMI_PRECHOOLCONT_FF1 < 13.46 & Data$agegroups=="5.5-5.9" ,

```

```

    "Vektstatus"] <- "undervektig"
Data[Data$BMI_PRECHOOLCONT_FF1 < 13.385 & Data$agegroups=="6.0-6.4" ,
    "Vektstatus"] <- "undervektig"
Data[Data$BMI_PRECHOOLCONT_FF1 < 13.35 & Data$agegroups=="6.5-6.9" ,
    "Vektstatus"] <- "undervektig"
Data[Data$BMI_PRECHOOLCONT_FF1 < 13.365 & Data$agegroups=="7.0-7.4" ,
    "Vektstatus"] <- "undervektig"
Data[Data$BMI_PRECHOOLCONT_FF1 < 13.44 & Data$agegroups=="7.5-7.9" ,
    "Vektstatus"] <- "undervektig"
Data[Data$BMI_PRECHOOLCONT_FF1 < 13.44 & Data$agegroups=="8.0+" ,
    "Vektstatus"] <- "undervektig"

#IOTF 2 = normal for alle aldre
Data[Data$BMI_PRECHOOLCONT_FF1 >= 13.815 & Data$BMI_PRECHOOLCONT_FF1 < 16.8
85
    & Data$agegroups=="4.0-4.5" , "Vektstatus"] <- "normalvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 13.685 & Data$BMI_PRECHOOLCONT_FF1 < 16.8
15
    & Data$agegroups=="4.5-4.9" , "Vektstatus"] <- "normalvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 13.565 & Data$BMI_PRECHOOLCONT_FF1 < 16.7
8
    & Data$agegroups=="5.0-5.4" , "Vektstatus"] <- "normalvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 13.46 & Data$BMI_PRECHOOLCONT_FF1 < 16.78
5
    & Data$agegroups=="5.5-5.9" , "Vektstatus"] <- "normalvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 13.385 & Data$BMI_PRECHOOLCONT_FF1 < 16.8
6
    & Data$agegroups=="6.0-6.4" , "Vektstatus"] <- "normalvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 13.35 & Data$BMI_PRECHOOLCONT_FF1 < 17
    & Data$agegroups=="6.5-6.9" , "Vektstatus"] <- "normalvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 13.365 & Data$BMI_PRECHOOLCONT_FF1 < 17.2
    & Data$agegroups=="7.0-7.4" , "Vektstatus"] <- "normalvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 13.44 & Data$BMI_PRECHOOLCONT_FF1 < 17.47
5
    & Data$agegroups=="7.5-7.9" , "Vektstatus"] <- "normalvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 13.44 & Data$BMI_PRECHOOLCONT_FF1 < 17.47
5
    & Data$agegroups=="8.0+" , "Vektstatus"] <- "normalvektig"

#IOTF 3 = overvektig for alle aldre
Data[Data$BMI_PRECHOOLCONT_FF1 >= 16.885 & Data$BMI_PRECHOOLCONT_FF1 < 19.1
4
    & Data$agegroups=="4.0-4.5" ,
    "Vektstatus"] <- "overvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 16.815 & Data$BMI_PRECHOOLCONT_FF1 < 19.1
4
    & Data$agegroups=="4.5-4.9" ,
    "Vektstatus"] <- "overvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 16.78 & Data$BMI_PRECHOOLCONT_FF1 < 19.2

```

```

    & Data$agegroups=="5.0-5.4" ,
    "Vektstatus"] <- "overvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 16.785 & Data$BMI_PRECHOOLCONT_FF1 < 19.3
6
    & Data$agegroups=="5.5-5.9" ,
    "Vektstatus"] <- "overvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 16.86 & Data$BMI_PRECHOOLCONT_FF1 < 19.61
    & Data$agegroups=="6.0-6.4" ,
    "Vektstatus"] <- "overvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 17 & Data$BMI_PRECHOOLCONT_FF1 < 19.96
    & Data$agegroups=="6.5-6.9" ,
    "Vektstatus"] <- "overvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 17.2 & Data$BMI_PRECHOOLCONT_FF1 < 20.39
    & Data$agegroups=="7.0-7.4" ,
    "Vektstatus"] <- "overvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 17.475 & Data$BMI_PRECHOOLCONT_FF1 < 20.8
9
    & Data$agegroups=="7.5-7.9" ,
    "Vektstatus"] <- "overvektig"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 17.475 & Data$BMI_PRECHOOLCONT_FF1 < 20.8
9
    & Data$agegroups=="8.0+" ,
    "Vektstatus"] <- "overvektig"

#IOTF 4 = fedme for alle aldre
Data[Data$BMI_PRECHOOLCONT_FF1 >= 19.14 & Data$agegroups=="4.0-4.5" ,
    "Vektstatus"] <- "fedme"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 19.14 & Data$agegroups=="4.5-4.9" ,
    "Vektstatus"] <- "fedme"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 19.2
    & Data$agegroups=="5.0-5.4" ,
    "Vektstatus"] <- "fedme"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 19.36
    & Data$agegroups=="5.5-5.9" ,
    "Vektstatus"] <- "fedme"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 19.61
    & Data$agegroups=="6.0-6.4" ,
    "Vektstatus"] <- "fedme"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 19.96
    & Data$agegroups=="6.5-6.9" ,
    "Vektstatus"] <- "fedme"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 20.39
    & Data$agegroups=="7.0-7.4" ,
    "Vektstatus"] <- "fedme"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 20.89
    & Data$agegroups=="7.5-7.9" ,
    "Vektstatus"] <- "fedme"
Data[Data$BMI_PRECHOOLCONT_FF1 >= 20.89
    & Data$agegroups=="8.0+" ,
    "Vektstatus"] <- "fedme"

```

```
table(Data$Vektstatus) #IOTF1=15, IOTF2=267, IOTF3=82, IOTF4=17
```

```
##  
##      fedme normalvektig  overvektig  undervektig  
##      17      267      82      15
```

```
#Dvs undervekt=15, normal=267, overvekt=82, fedme=17
```

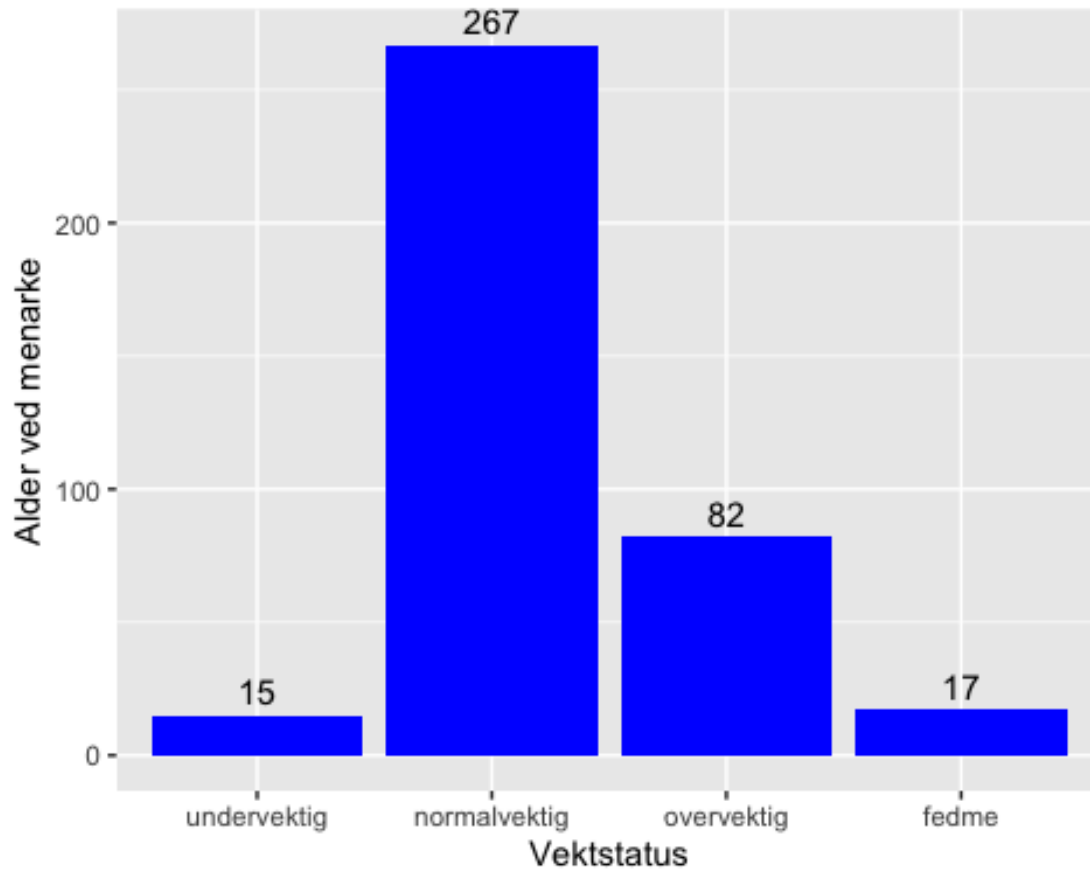
Søylediagram

```
library(ggplot2)  
library(dplyr)
```

```
iotf_barplot <- Data %>% group_by(Vektstatus)%>%summarise(counts=n())  
iotf_barplot
```

```
## # A tibble: 5 × 2  
##   Vektstatus counts  
##   <chr>      <int>  
## 1 fedme      17  
## 2 normalvektig 267  
## 3 overvektig   82  
## 4 undervektig  15  
## 5 <NA>       127
```

```
barplot_iotf <- ggplot(iotf_barplot, aes(x=Vektstatus, y=counts))+geom_bar(fill="blue", stat="identity") + geom_text(aes(label=counts),vjust=-0.5)  
barplot_iotf + scale_x_discrete(limits=c("undervektig", "normalvektig", "overvektig", "fedme")) + labs(y="Alder ved menarke")
```



Sosioøkonomisk status (SES)

Lager dummy-variabel for lav sosioøkonomisk status

```
Data$Low_SES <- ifelse(Data$FATHER_EDUCATION_FF1<4 &
                      Data$MOTHER_EDUCATION_FF1<4, 1, 0) #dummy for SES
#Her bestemt utifra om minst en av foreldrene har høyere utdanning enn vgs
#(vgs=3 i utdanningsskalaen)
table(Data$Low_SES) #ser bra ut, 247 av jentene regnes som Lav SES

##
##  0  1
## 247 247

#(ingen av foreldrene har mer enn vgs-utdanning)
```

Kommentar:

247 jenter har lav SES, mens 247 har høy SES basert på vår def, altså at minst en av foreldrene har mer utdanning enn vgs. (494 jenter har svart på SES - altså ytterligere 1 NA)

Etnisitet

```
summary(Data$ETHNICITY_NORWEGIAN_FF1) #0.964 har svart norsk, 8 NA
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##  0.000  1.000  1.000  0.964  1.000  1.000    8
```

```
summary(Data$ETHNICITY_SAMI_FF1) #0.032 har svart samisk, 8 NA
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##  0.000  0.000  0.000  0.032  0.000  1.000    8
```

```
summary(Data$ETHNICITY_KVENFINNISH_FF1) #0.02 har svart kvensk, 8 NA
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##  0.00  0.00  0.00  0.02  0.00  1.00    8
```

```
summary(Data$ETHNICITY_OTHER_FF1) #0.068 er annen etnisitet, 8 NA
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##  0.000  0.000  0.000  0.068  0.000  1.000    8
```

```
#Tydeligvis har noen markert mer enn en etnisitet...får ta det med i  
#diskusjonen
```

```
#Vi endrer ingen analyser av den grunn
```

```
table(Data$ETHNICITY_NORWEGIAN_FF1) #484 har sagt norsk etnisitet
```

```
##  
##  0  1  
## 18 482
```

```
table(Data$ETHNICITY_SAMI_FF1) #16 har sagt samisk etnisitet
```

```
##  
##  0  1  
## 484 16
```

```
table(Data$ETHNICITY_KVENFINNISH_FF1) #10 har sagt kvensk etnisitet
```

```
##  
##  0  1  
## 490 10
```

```
table(Data$ETHNICITY_OTHER_FF1) #34 har sagt annen etnisitet
```

```
##  
##  0  1  
## 466 34
```

```
#500 har svart på alle, dvs 7 missing
```

```
#Lager også dummy variabel for Ikke_norsk siden de aller fleste er norske
```

```
Data$Ikke_norsk <- ifelse(Data$ETHNICITY_NORWEGIAN_FF1==0,1,0)
```

```
table(Data$Ikke_norsk) #gjør modellen lett å forstå
```

```
##
##   0   1
## 482 18
```

Kommentar:

I spørreskjemaet kunne deltakere svare flere etnisiteter, og dermed beskriver vi fordelingen på to måter. Først så har 484 (96.8%) svart at de er av norsk etnisitet, 16 (3.2%) har svart samisk etnisitet, 10 (2%) har svart kvensk og 34 (6.8%) har svart at de er av annen etnisitet enn norsk, samisk og kvensk. 7 stykker svarte ikke på spørsmålet. Når en kun ser på antallet som ikke har svart på at de er av norsk etnisitet (men svarte på spørsmålet), så er antallet 18(3.6%).

Etnisitet, IOTF og SES

Ikke-norsk etnisitet; SES og IOTF ved førskolealder

```
iotf_ikke_norsk <- Data[ which(Data$Ikke_norsk==1), ]
table(iotf_ikke_norsk$Vektstatus) #kun 2 stykker tilgjengelig, begge to norma
lvektige

##
## normalvektig
##           2

table(iotf_ikke_norsk$Low_SES) #Av 18 stykker, har 11 Lav SES (61%)

##
##   0   1
##   7  11
```

Kommentar:

Kun 2 med ikke-norsk etnisitet hadde tilgjengelig vektstatus ved førskolealder, og begge var normalvektige (IOTF=2). Av de 18 som ikke svarte at de var norsk etnisitet, hadde 11 (61%) lav SES.

Norsk etnisitet; SES og Vektklasse ved førskolealder

```
iotf_norsk <- Data[ which(Data$Ikke_norsk==0), ]
table(iotf_norsk$Vektstatus) #15 undervektige, 261 normal, 98 overvektige. 37
1

##
##          fedme normalvektig   overvektig   undervektig
##          17           261           81           15

#totalt altså (Mye missing for vektstatus i førskolealder)
table(iotf_norsk$Low_SES) #Av 374 stykker, har 236 Lav SES (63%)
```



```
##  
## 0 1  
## 240 236  
  
236/476  
  
## [1] 0.4957983
```

Kommentar:

Av de 374 med norsk etnisitet, var 15 (4%) undervektige, 261 (70%) normalvektige, 81 (22%) overvektige, og 17 (4.5%) hadde fedme. Av de 476 med norsk etnisitet som hadde svart på spm om foreldres utdanning, hadde 236 (50%) lav SES.

Antallet med ikke-norsk etnisitet er for lavt til å gjøre seg opp noen konklusjoner rundt dette.

Fordeling menarkealder (KM-analyse)

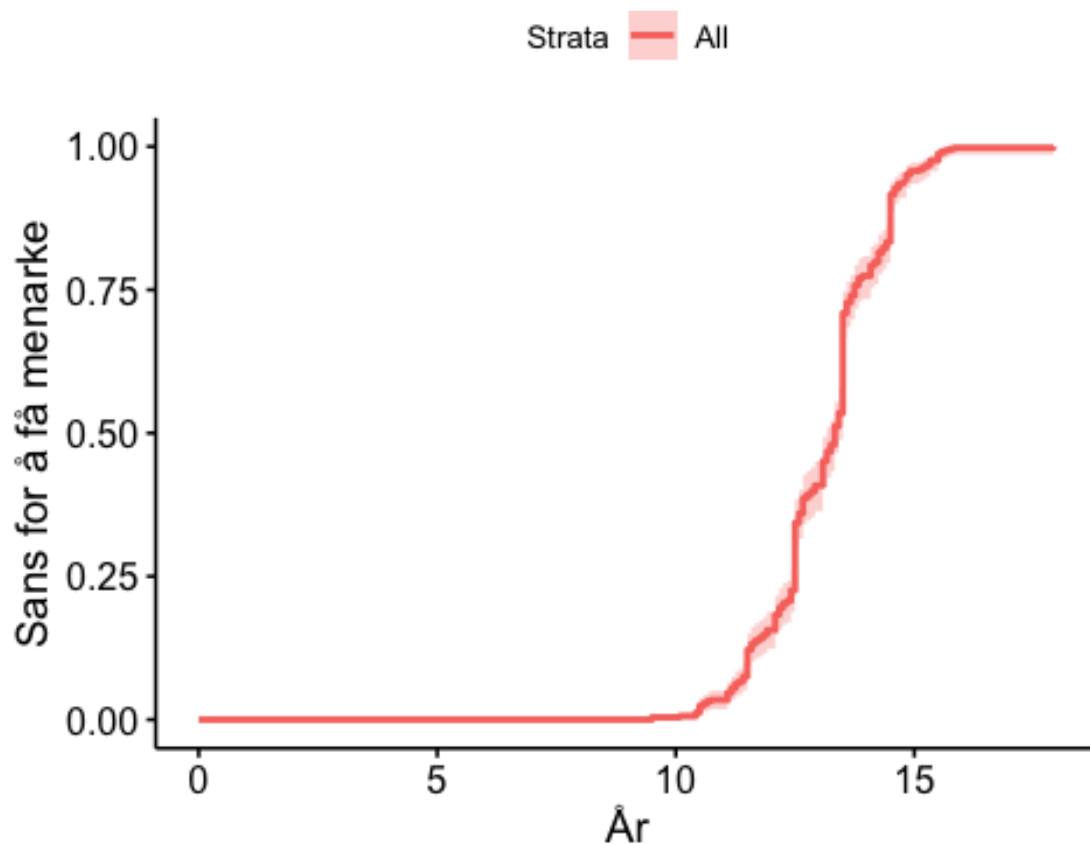
Bruker Kaplan-Meyer overlevelsesanalyse for å finne gj.snitt og median menarkealder. Bruker KM siden det ble brukt i Bratke et al. 2017, så resultatet kan sammenlignes.

Overlevelseskurver

- Sannsynlighet for å få menarke ved ulik alder (mnd-vis)

Menarke

```
library(survival) #trenger denne pakken for analysen  
#Definerte tidlige "menarke_alder" for eksakt alder  
  
library(survminer)  
  
f1 <- survfit(Surv(menarke_alder, MENARCHE_FF1) ~ 1, data = Data)  
names(f1)  
  
ggsurvplot(f1,  
           fun = "event",  
           xlab = "År",  
           ylab = "Sans for å få menarke")
```



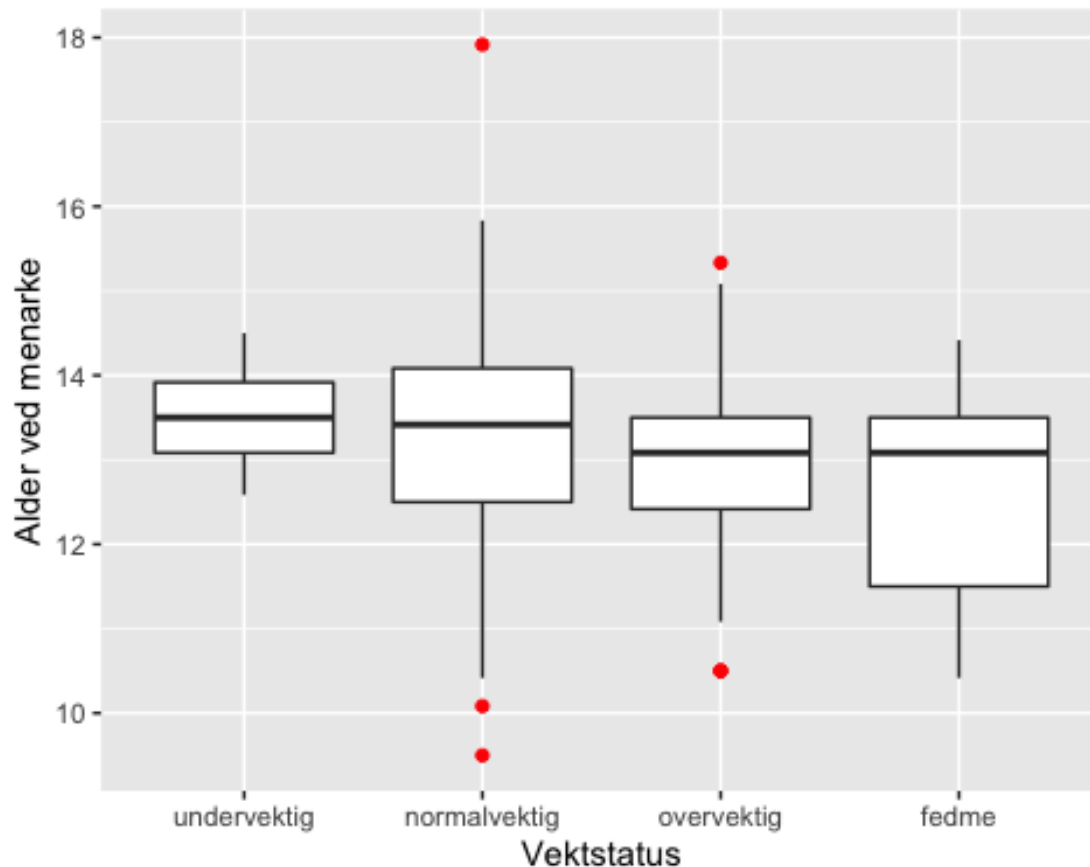
Kommentar:

Ser tydelig at f.o.m. rett før fylte 10 år starter noen å få menarke, og så er grafen brattest rundt 13-14 år, og etter 15 år har den flatet nesten helt ut. Den rosa skyggen viser konfidensintervallet.

Boksdigram

Splitter opp for hver vektklasse, SES og etnisitet:

```
Data$Vektstatus <- as.factor(Data$Vektstatus)
library(ggplot2)
boxp_1 <- ggplot(Data, aes(x=Vektstatus, y=menarke_alder))+geom_boxplot(outlier.colour = "red")
boxp_1 + scale_x_discrete(limits=c("undervektig", "normalvektig", "overvektig", "fedme")) + labs(y="Alder ved menarke")
```



Ønsker noe mer informativt enn boksdiagrammer, da en mister distribusjonen i hver gruppe her. Lager derfor fiolin-diagram i tillegg.

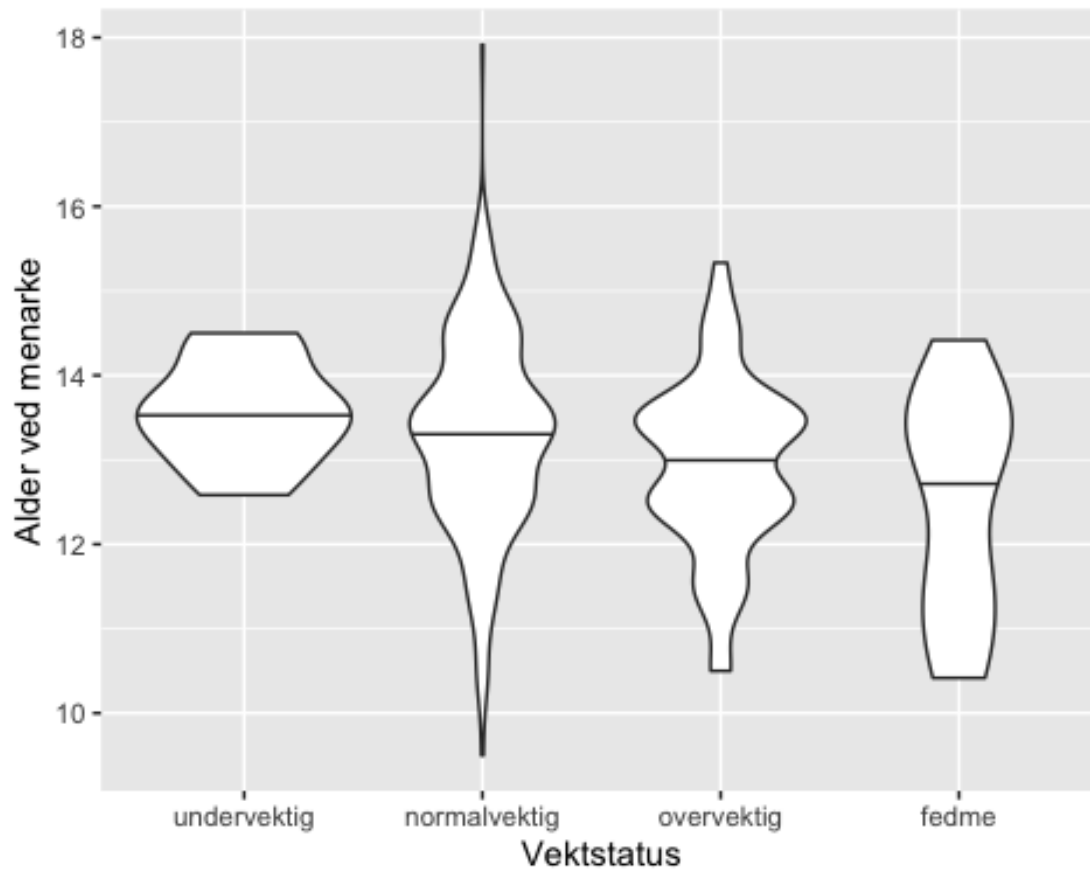
Fiolindiagram

Kommentar:

Streken markert inne i diagrammet viser median alder ved menarke, og fasongen på «fiolinen» viser aldersdistribusjonen i gruppen.

Menarkealder for ulike vektklasser

```
p <- ggplot(Data, aes(x=Vektstatus, y=menarke_alder))
p + geom_violin(draw_quantiles = c(0.5)) + scale_x_discrete(limits=c("undervektig", "normalvektig", "overvektig", "fedme")) + labs(y="Alder ved menarke")
## Warning: Removed 138 rows containing non-finite values (stat_ydensity).
```



Kommentar:

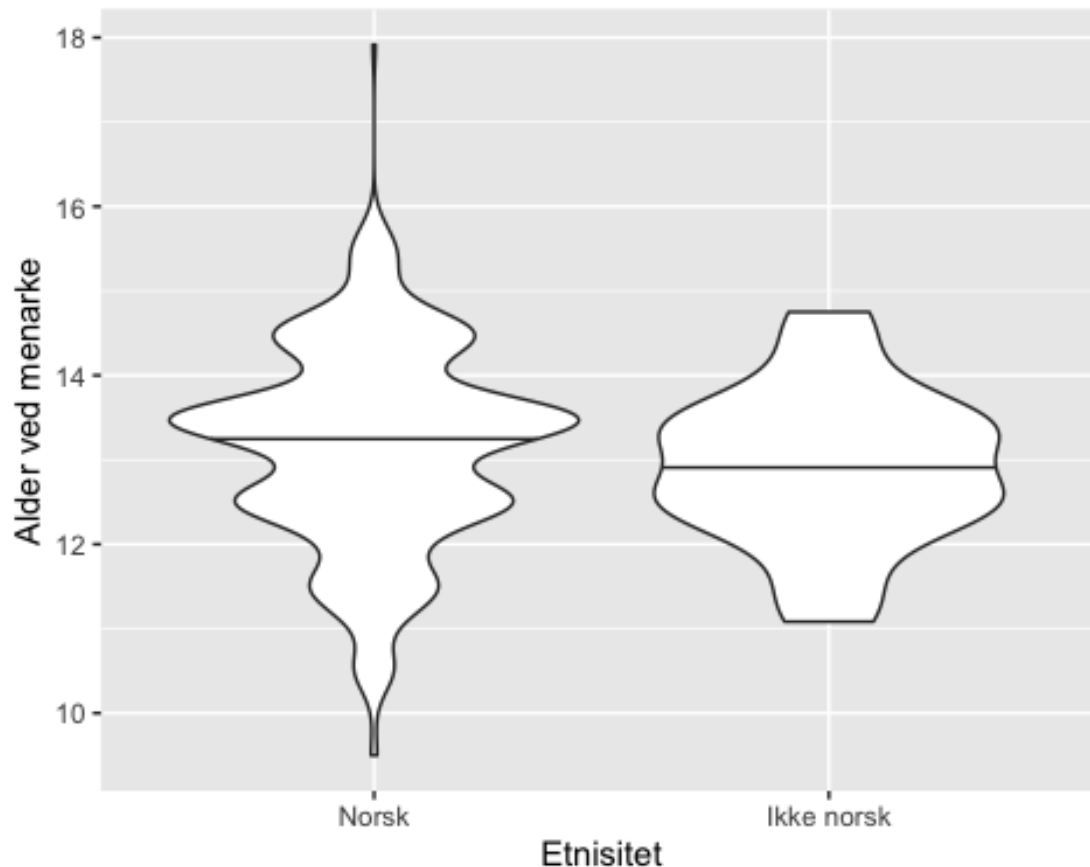
Ser at distribusjonen er litt ulik mellom normalvektige og overvektige, mens de små gruppene (undervektige og fedme) er litt mer vage. Ser dog ut som en svak trend av litt lavere menarkealder, jo høyere vektklasse en er (undervektig eldst, fedme yngst).

Menarkealder ift etnisitet

```
Data$Etnisitet_binær <- ifelse(Data$Ikke_norsk==1,"Ikke norsk", "Norsk")
table(Data$Etnisitet_binær)

##
## Ikke norsk      Norsk
##           18      482

q <- ggplot(Data, aes(x=Etnisitet_binær, y=menarke_alder))
q + geom_violin(draw_quantiles = c(0.5)) + labs(y="Alder ved menarke", x="Etnisitet") + scale_x_discrete(limits=c("Norsk", "Ikke norsk"))
```



Kommentar:

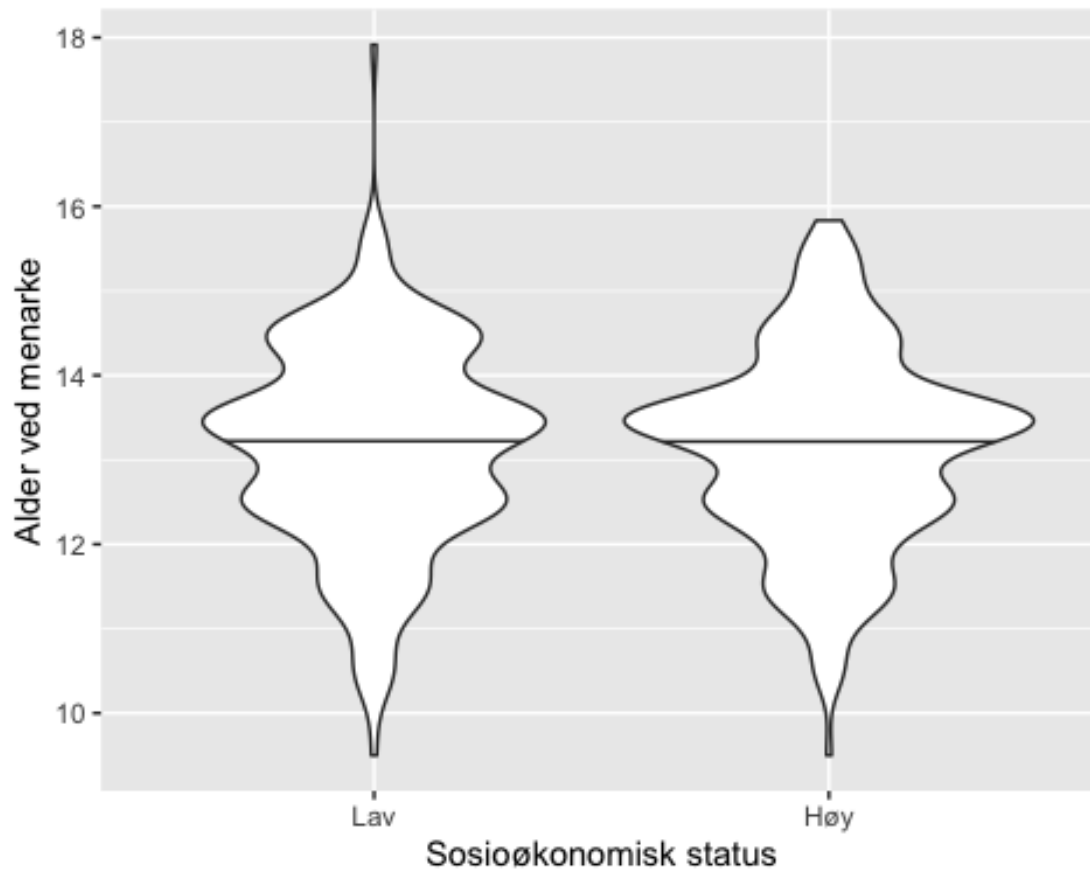
Her er det umulig å si om fordelingen egentlig er forskjellig, da det er så få med ikke-norsk etnisitet. Ser dog at de med norsk etnisitet har en (litt hakkede) ganske normal fordeling rundt 13-14 år ved menarke. Ser også en svak tendens til at de norske er litt eldre ved menarke – men vær obs på skalaen siden det er så mange flere i «norsk»-gruppen.

Menarkealder ved ulik SES

```
Data$SES <- ifelse(Data$Low_SES==1,"Lav", "Høy")
table(Data$SES)

##
## Høy Lav
## 247 247

w <- ggplot(Data, aes(x=SES, y=menarke_alder))
w + geom_violin(draw_quantiles = c(0.5)) + labs(y="Alder ved menarke", x="So
sioøkonomisk status") + scale_x_discrete(limits=c("Lav", "Høy"))
```



Kommentar:

Ser omtrent helt likt ut i høy og lav SES.

Log-rank / Mantel-Haenszel test

Tester om det er sig. ulik menarkealder i ulike grupper vha. Log-rank (Mantel-Haenszel) test for to eller flere grupper. Log-rank test er en ikke-parametrisk test, som ikke forutsetter antagelse om overlevelsesdistribusjonen (her: distribusjonen i alder ved menarke). Tolkes som at $p < 0.05$ viser at to grupper er like med 95% KI, mens $p > 0.05$ viser at en ikke kan kalle de like med 95% KI. En forventer her at menarkealder ikke kan sies å være lik, i ulike kategorier av våre uavhengige variabler.

```
library(survival)
#Ulike vektklasser
survdifff(Surv(menarke_alder, MENARCHE_FF1) ~ Vektstatus, Data, rho=0) #p=0.00
5

## Call:
## survdifff(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Vektstatus,
##           data = Data, rho = 0)
##
```

```
## n=370, 138 observations deleted due to missingness.
##
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Vektstatus=fedme      17      17    10.3    4.364    5.190
## Vektstatus=normalvektig 261     261   282.9    1.692    8.641
## Vektstatus=overvektig  77      77    60.1    4.763    6.756
## Vektstatus=undervektig 15      15    16.7    0.181    0.224
##
## Chisq= 13.1 on 3 degrees of freedom, p= 0.005
```

#-> De ulike vektklassene er statistisk signifikant ulike på p<0.05-nivå.

#SES

```
survdifff(Surv(menarke_alder, MENARCHE_FF1) ~ Low_SES, Data, rho=0)
```

```
## Call:
## survdifff(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Low_SES,
## data = Data, rho = 0)
##
## n=486, 22 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Low_SES=0 246      246    241    0.124    0.294
## Low_SES=1 240      240    245    0.121    0.294
##
## Chisq= 0.3 on 1 degrees of freedom, p= 0.6
```

#p=0.6 -> Kan ikke si at det er stat. sig. ulik menarkealder hos de med ulik SES.

#Etnisitet

```
survdifff(Surv(menarke_alder, MENARCHE_FF1) ~ Ikke_norsk, Data, rho=0)
```

```
## Call:
## survdifff(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Ikke_norsk,
## data = Data, rho = 0)
##
## n=492, 16 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Ikke_norsk=0 476      476   480.3    0.0384    1.86
## Ikke_norsk=1  16       16    11.7    1.5770    1.86
##
## Chisq= 1.9 on 1 degrees of freedom, p= 0.2
```

#p=0.2 -> Kan ikke si at det er stat. sig. ulik alder ved menarke mellom #de med norsk og de med ikke-norsk etnisitet.

Kommentar:

Log-rank-test viser at det er statistisk signifikant forskjell i menarkealder mellom de ulike vektkategoriene (vektstatus), men ikke i ulik SES eller etnisitet. Deltagerne er nesten delt 50/50 i hvilken SES de har, så en kan anta at svaret her er representativt. 96% av deltagerne har norsk etnisitet, så for denne variabelen er det vanskelig å si om det er reelt at menarkealder ikke varierer mellom ulike etnisiteter i populasjonen vår - eller om det bare er for få med ikke-norsk etnisitet til å konkludere noe.

CoxPH-modeller

Kommentar:

Lager coxph-modell for hver uavhengig variabel, og bygger deretter modell med utg.pkt. i vektstatus. Alle individuelle modeller med $p < 0.20$ blir tatt med i første multivariabel-modell.

Vektstatus deles opp i dummy-variabler, slik at en fjerner antagelsen om at det er lik forskjell i menarkealder mellom alle vektclassene (at menarkealder øker like mye om en går fra normal til overvektig, som fra overvektig til å ha fedme).

```
library(survival)

coxph_vekt <- coxph(Surv(menarke_alder, MENARCHE_FF1) ~ Vektstatus, data=Data
)
summary(coxph_vekt) #overvekt og fedme sig.:

## Call:
## coxph(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Vektstatus,
##       data = Data)
##
##      n= 370, number of events= 370
##      (138 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Vektstatusnormalvektig -0.6557    0.5191  0.2525 -2.597  0.00942 **
## Vektstatusovervektig   -0.2605    0.7707  0.2685 -0.970  0.33198
## Vektstatusundervektig  -0.6601    0.5168  0.3547 -1.861  0.06276 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Vektstatusnormalvektig  0.5191    1.927   0.3164   0.8515
## Vektstatusovervektig    0.7707    1.298   0.4553   1.3044
## Vektstatusundervektig   0.5168    1.935   0.2579   1.0358
##
## Concordance= 0.552 (se = 0.014 )
```



```

## Likelihood ratio test= 13  on 3 df,  p=0.005
## Wald test          = 14.2  on 3 df,  p=0.003
## Score (logrank) test = 14.5  on 3 df,  p=0.002

#Overvekt: HR1.5 (1.1-1.9), fedme: HR 1.9 (1.2-3.2)

#Lager dummy-variabler for vekt
Data$Overvekt <- ifelse(Data$Vektstatus=="overvektig",1,0)

Data$Undervekt <- ifelse(Data$Vektstatus=="undervektig",1,0)
Data$Fedme <- ifelse(Data$Vektstatus=="fedme",1,0)

coxph_Ovekt <- coxph(Surv(menarke_alder, MENARCHE_FF1) ~ Overvekt, data=Data)
summary(coxph_Ovekt) #p<0.05

## Call:
## coxph(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Overvekt,
##       data = Data)
##
## n= 370, number of events= 370
## (138 observations deleted due to missingness)
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## Overvekt 0.3631  1.4378  0.1293  2.807  0.005 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## Overvekt  1.438      0.6955  1.116  1.853
##
## Concordance= 0.532 (se = 0.012 )
## Likelihood ratio test= 7.35  on 1 df,  p=0.007
## Wald test          = 7.88  on 1 df,  p=0.005
## Score (logrank) test = 7.97  on 1 df,  p=0.005

coxph_Fedme <- coxph(Surv(menarke_alder, MENARCHE_FF1) ~ Fedme, data=Data)
summary(coxph_Fedme) #p<0.05

## Call:
## coxph(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Fedme, data = Data)
##
## n= 370, number of events= 370
## (138 observations deleted due to missingness)
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## Fedme 0.5719  1.7716  0.2499  2.288  0.0221 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## Fedme  1.772      0.5645  1.085  2.891

```

```

##
## Concordance= 0.513 (se = 0.007 )
## Likelihood ratio test= 4.44 on 1 df, p=0.04
## Wald test = 5.24 on 1 df, p=0.02
## Score (logrank) test = 5.38 on 1 df, p=0.02

coxph_Uvekt <- coxph(Surv(menarke_alder, MENARCHE_FF1) ~ Undervekt, data=Data
)
summary(coxph_Uvekt) #p=0.67 --> tas ikke med videre

## Call:
## coxph(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Undervekt,
## data = Data)
##
## n= 370, number of events= 370
## (138 observations deleted due to missingness)
##
## coef exp(coef) se(coef) z Pr(>|z|)
## Undervekt -0.1127 0.8934 0.2645 -0.426 0.67
##
## exp(coef) exp(-coef) lower .95 upper .95
## Undervekt 0.8934 1.119 0.532 1.5
##
## Concordance= 0.51 (se = 0.005 )
## Likelihood ratio test= 0.19 on 1 df, p=0.7
## Wald test = 0.18 on 1 df, p=0.7
## Score (logrank) test = 0.18 on 1 df, p=0.7

coxph_SES <- coxph(Surv(menarke_alder, MENARCHE_FF1) ~ Low_SES, data=Data)
summary(coxph_SES) #Lav SES exp(coef)=-0.98, exp(CI)=0.82-1.18, ikke sig. (p>
0.05)

## Call:
## coxph(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Low_SES,
## data = Data)
##
## n= 486, number of events= 486
## (22 observations deleted due to missingness)
##
## coef exp(coef) se(coef) z Pr(>|z|)
## Low_SES -0.01242 0.98765 0.09110 -0.136 0.892
##
## exp(coef) exp(-coef) lower .95 upper .95
## Low_SES 0.9877 1.012 0.8262 1.181
##
## Concordance= 0.504 (se = 0.014 )
## Likelihood ratio test= 0.02 on 1 df, p=0.9
## Wald test = 0.02 on 1 df, p=0.9
## Score (logrank) test = 0.02 on 1 df, p=0.9

#p=0.892

```

```

coxph_etnisitet <- coxph(Surv(menarke_alder, MENARCHE_FF1) ~ Ikke_norsk, data
=Data)
summary(coxph_etnisitet) #Ikke norsk exp(coef) // HR=1.14, exp(CI)=0.86-2.33,

## Call:
## coxph(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Ikke_norsk,
##       data = Data)
##
## n= 492, number of events= 492
## (16 observations deleted due to missingness)
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## Ikke_norsk 0.3493   1.4181  0.2548 1.371   0.17
##
##               exp(coef) exp(-coef) lower .95 upper .95
## Ikke_norsk    1.418    0.7052  0.8606    2.337
##
## Concordance= 0.505 (se = 0.004 )
## Likelihood ratio test= 1.69 on 1 df,  p=0.2
## Wald test               = 1.88 on 1 df,  p=0.2
## Score (logrank) test = 1.9 on 1 df,  p=0.2

#p=0.17

```

Kommentar:

Vektstatus **overvektig** og **fedme**, samt **etnisitet**, blir tatt med videre i byggingen av modellen.

Bygger modell

Kommentar:

Det gjøres baklengs og stegvis fjerning av variabler som ikke er statistisk sig. på $P < 10$ -nivå, også kjent som "backwise stepwise selection". Det regnes som et konservativt og trygt nivå å legge seg på i seleksjonsprosessen, slik at en ikke fjerner variabler ved å ha for høy terskel.

```

#Inkluderer alle de som hadde en p<0.20 på individuelle modeller
Coxph_1 <- coxph(Surv(menarke_alder, MENARCHE_FF1) ~ Overvekt+Fedme+Ikke_norsk, data=Data)
summary(Coxph_1)

## Call:
## coxph(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Overvekt +
##       Fedme + Ikke_norsk, data = Data)
##
## n= 370, number of events= 370
## (138 observations deleted due to missingness)

```

```

##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## Overvekt   0.3930   1.4814  0.1304  3.013 0.00258 **
## Fedme      0.6538   1.9228  0.2520  2.595 0.00947 **
## Ikke_norsk -0.3160   0.7290  0.7109 -0.445 0.65665
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## Overvekt      1.481      0.6750      1.147      1.913
## Fedme         1.923      0.5201      1.173      3.151
## Ikke_norsk    0.729      1.3717      0.181      2.937
##
## Concordance= 0.547 (se = 0.014 )
## Likelihood ratio test= 13.22 on 3 df,  p=0.004
## Wald test              = 14.38 on 3 df,  p=0.002
## Score (logrank) test = 14.69 on 3 df,  p=0.002

#Fjerner de som ikke tilfredsstillter et krav på p<0.10
#-->Fjerner etnisitet

Coxph_2 <- coxph(Surv(menarke_alder, MENARCHE_FF1) ~ Overvekt+Fedme, data=Data)
summary(Coxph_2) #Både overvekt og fedme er sig.

## Call:
## coxph(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Overvekt +
##       Fedme, data = Data)
##
## n= 370, number of events= 370
## (138 observations deleted due to missingness)
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## Overvekt 0.3955   1.4852  0.1303  3.035 0.00241 **
## Fedme    0.6560   1.9271  0.2520  2.604 0.00922 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## Overvekt      1.485      0.6733      1.150      1.917
## Fedme         1.927      0.5189      1.176      3.158
##
## Concordance= 0.547 (se = 0.014 )
## Likelihood ratio test= 13 on 2 df,  p=0.002
## Wald test              = 14.2 on 2 df,  p=8e-04
## Score (logrank) test = 14.5 on 2 df,  p=7e-04

#overvekt: HR 1.49 (1.15-1.92), fedme: HR 1.93 (1.18-3.16)

```

Endelig modell

```
Endelig_coxph <- Coxph_2
summary(Endelig_coxph) #Dersom overvektig (justert KMI <25,30>), så er

## Call:
## coxph(formula = Surv(menarke_alder, MENARCHE_FF1) ~ Overvekt +
##       Fedme, data = Data)
##
## n= 370, number of events= 370
## (138 observations deleted due to missingness)
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## Overvekt  0.3955   1.4852  0.1303  3.035  0.00241 **
## Fedme     0.6560   1.9271  0.2520  2.604  0.00922 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## Overvekt     1.485     0.6733     1.150     1.917
## Fedme        1.927     0.5189     1.176     3.158
##
## Concordance= 0.547 (se = 0.014 )
## Likelihood ratio test= 13 on 2 df,  p=0.002
## Wald test              = 14.2 on 2 df,  p=8e-04
## Score (logrank) test = 14.5 on 2 df,  p=7e-04

#HR 1.49 (1.15-1.92), sammenlignet med resten (normal- og undervektige).
#Dersom en har fedme, er HR 1.93 (1.18-3.16)
#n=370, concordance=0.55, Likelihood ratio=13 på 2 df, Wald=14.2, Logrank=14.5
```

Kommentar:

Modellen estimerer at jenter som er overvektige i førskolealder har HR=1.49 (1.15-1.92 95% KI), mens de med fedme i førskolealder har HR=1.93 (1.18-3.16 95% KI), for å få menarke **før** de som er normal- og undervektige i førskolealder. Modellen brukte n=370. Deltagere i FF1 om ikke hadde tilgjengelig vektstatus i førskolealder, ble ikke brukt til å bygge modellen. Modellen angir en concordance=0.547 (se0.014), likelihood ratio test=12 på 2 df, p=0.002 Wald-test = 14.2 på 2 df med p=8e-04, og score (logrank)-test=14.5 på 2 df med p=7e-04.

Tester antagelser for Cox PH regresjon

Kommentar:

Cox proposjonal Hazard modell har to antagelser; 1) overlevelseskurvene for ulike grupper har HR-funksjoner som er proposjonal over tid, og 2) Forholdet mellom log-hazard og hver kovariate er lineær (dette kan verifiseres med residualplot).

- 1) For å teste om proporsjonalitet er konstant over tid, brukes Schoenfeldts test (og plot). Testen bruker modell-residualer for å se om de er like over tid, dvs uavhengige av tid). Om denne testen viser sig. Sammenheng mellom tid og residualer, anses antagelsen om konstant proporsjonalitet som feil. På samme måte, tolker en plottet ved at ikke-tilfeldig mønster over tid anses det som bevis for at antagelsen er feil./ ikke møtt.
- 2) For å teste linearitet brukes *Martingale* residualer. Metoden går ut på å plote modell-residualer opp mot hver enkelt log av kontinuerlige uavhengige variabler. Da dette kun er aktuelt for kontinuerlige variabler, er det ikke aktuelt å gjennomføre her. Antagelsen anses som møtt.

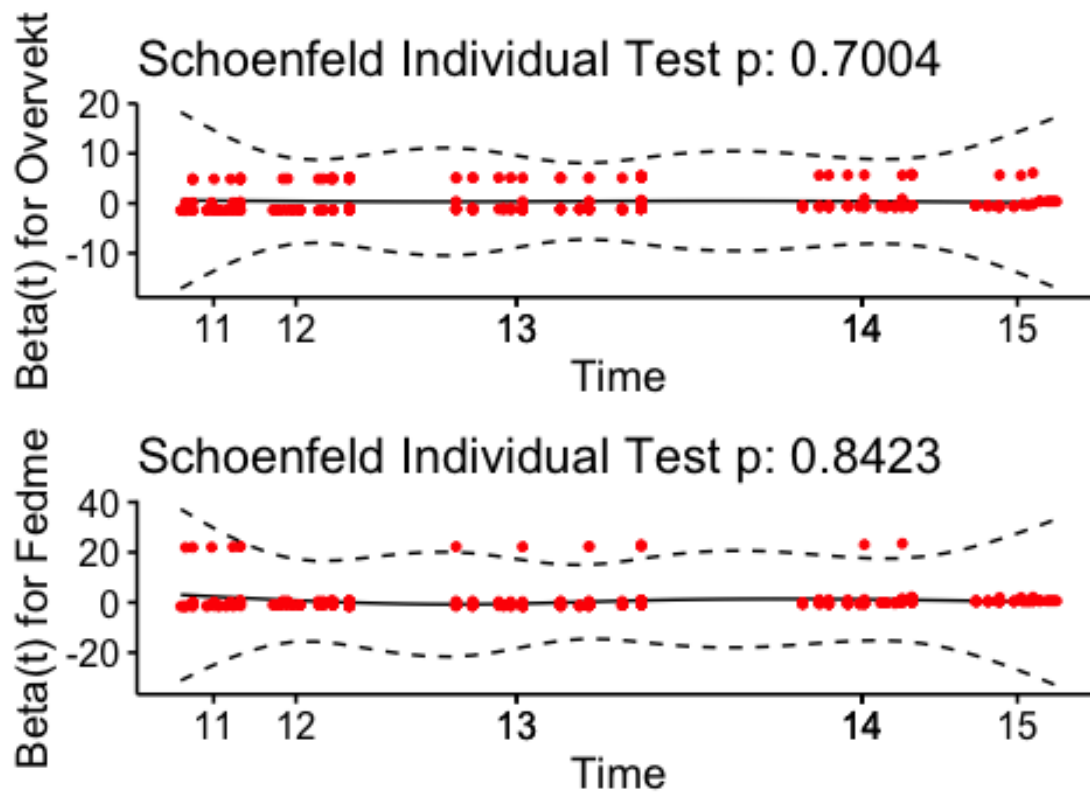
```
library(survival)
(ftest <- cox.zph(Endelig_coxph))#Ser at p>0.05 både globalt, og for hver enkelt

##           chisq df    p
## Overvekt  0.1481  1 0.70
## Fedme     0.0396  1 0.84
## GLOBAL    0.2051  2 0.90

#kovariabel, som betyr at antagelsen er møtt - og en kan bruke modellen.

library(survminer)
ggcoxzph(ftest)#Ser ekstremt konstant ut over tid -->
```

Global Schoenfeld Test p: 0.9025



#Antagelse anses også møtt grafisk.

Kommentar:

Ser på Schoenfeldts test at $p > 0.05$ både globalt, og for hver enkel uavhengige variabel, som betyr at antagelsen er møtt - og en kan bruke modellen. Dette vises også tydelig på den grafiske framstillingen av begge uavhengige variabler. Siden ingen av de uavhengige variablene er kontinuerlige, anses også antagelsen om linearitet som møtt (uten behov for å teste).