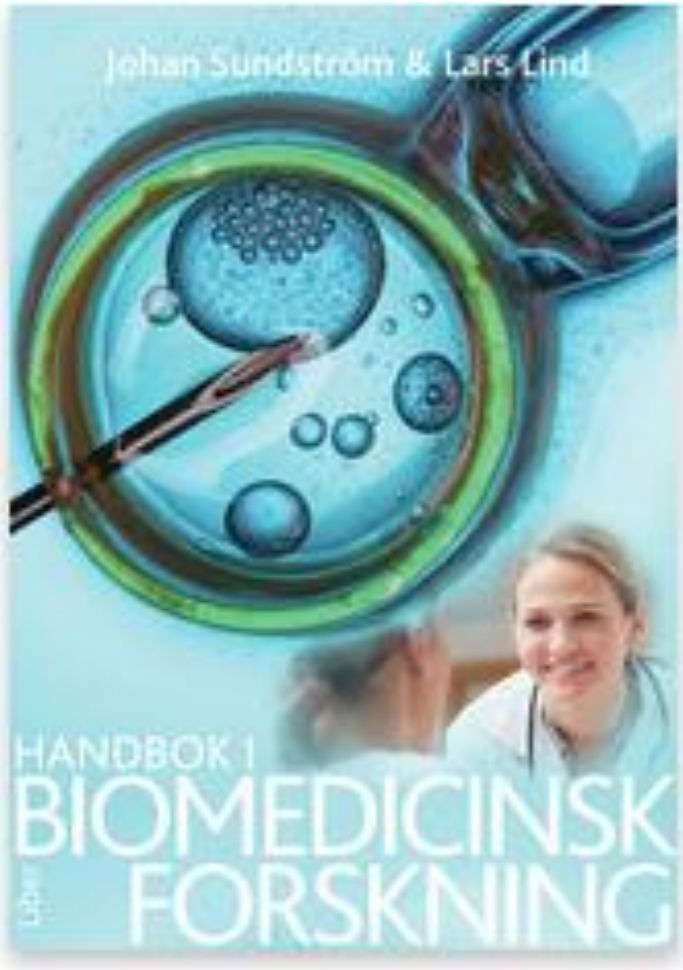
EXTRAMATERIAL



Extramaterial till Handbok i biomedicinsk forskning

Exempel 5.6–5.24

Exempel 5.6. Multipel imputering

I detta exempel vill vi ta reda på om utfallet fettmassa är relaterat till exponeringen triglycerider. Vi har identifierat 8 variabler som vi vill justera för, men vid närmare inspektion av data saknas det 150 (av cirka 1 000 observationer) värden på två av dessa variabler (”alco” och ”enerkcal”). Vi utför en multipel imputation med hjälp av en regressionsmodell där dessa två variabler förklaras av de övriga exponeringar som vi vill justera för. Vi bestämmer att vi vill ha 10 imputerade värden för varje saknat värde.

```
. mi set mlong

. mi register imputed alco enerkcal

. mi impute mvn alco enerkcal= sex lngd utbildning13 rkarenu motion ,
add(10) force


Multivariate imputation                 Imputations =       10
Multivariate normal regression                added =       10
Imputed: m=1 through m=10                   updated =        0

Prior: uniform                           Iterations =     1000
                                            burn-in =      100
                                            between =      100

               |              Observations per m
               |-----------------------------------------------
      Variable |   complete   incomplete   imputed |     total
---------------+-----------------------------------+----------
          alco |        861          155       149 |      1016
      enerkcal |        861          155       149 |      1016
-------------------------------------------------------------
```

Dessa kommandon leder till att vi nu har imputerat 149 av totalt 155 saknade värden för de båda variablerna. I dessa fall har 10 nya rader för varje saknat värde adderats till datasetet, som nu har drygt 2 500 rader i stället för 1 016.

I kommandot nedan tar ”mi estimate” hand om dessa extrarader i regressionsmodellen och gör ett medelvärde av resultaten i de imputerade datasetet. Resultatet ser ut som en vanlig regressionsmodell.

```
mi estimate:reg fett_total triglycerider sex lngd lean_total utbildning13 rkarenu
///
motion alco enerkcal

Multiple-imputation estimates                     Imputations     =         10
Linear regression                                 Number of obs   =        853
                                                  Average RVI     =     0.0235
                                                  Complete DF     =        843
DF adjustment:   Small sample                     DF:     min     =     417.67
                                                          avg     =     752.99
                                                          max     =     838.05
Model F test:       Equal FMI                     F(   9,  835.0) =      35.66
Within VCE type:          OLS                     Prob > F        =     0.0000

------------------------------------------------------------------------------
  fett_total |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
triglyceri~r |   1802.268   446.9521     4.03   0.000     924.9911    2679.545
         sex |   12802.37   965.1027    13.27   0.000     10908.07    14696.68
        lngd |  -93.53042   52.17065    -1.79   0.073    -195.9328    8.871916
  lean_total |   .6557334   .0525573    12.48   0.000     .5525726    .7588941
utbildning13 |  -500.1956   327.0741    -1.53   0.127    -1142.185    141.7936
     rkarenu |   -2030.15   890.2903    -2.28   0.023     -3777.63   -282.6694
      motion |  -965.2026   359.1317    -2.69   0.007    -1670.108   -260.2974
        alco |   44.99134   96.23237     0.47   0.640    -144.1688    234.1515
    enerkcal |  -2.807086   .6013229    -4.67   0.000    -3.988875   -1.625298
       _cons |   9744.315   8127.572     1.20   0.231    -6208.543    25697.17
```

Exempel 5.7. Rate ratio

Om vi har en kohort av personer med låg utbildning (enbart grundskola) respektive högre utbildning (minst gymnasieskola) samt räknar ut raterna och 95 % CI för respektive grupp, får vi följande resultat (tabell 1).

Tabell 1. Rate (och 95 % CI) I förhållande till utbildning.

	Rate (per 1 000 pyar)	95 % CI undre	95 % CI övre
Hög utbildning	12,1	11,1	13,2
Låg utbildning	24,9	22,4	27,7

För lågutbildade får vi ett rate ratio (RR) på 2,0, dvs. lågutbildade har fördubblad risk för att avlida under uppföljningstiden jämfört med högutbildade. Vi kan även se att de båda gruppernas 95 % CI inte överlappar varandra. Detta är inget formellt hypotestest, men man brukar kunna vara ganska säker på att skillnaderna i rate mellan grupperna inte bara har tillkommit av slumpen. Här skulle 95 av 100 stickprov i gruppen med hög utbildning ha en rate mellan 11,1 och 13,2, medan 95 av 100 stickprov i gruppen med låg utbildning skulle ligga inom dess CI, som ligger betydligt högre.

Det finns formella hypotestest som man kan göra för att se att skillnaden i RR mellan grupperna är signifikant. Det finns dels ”klassiska metoder”, dels regressionsmodeller, t.ex. Poisson-regression och Cox proportionella hasardmodell.

Bild 1. Rate och 95 % CI I fem åldersgrupper.

```
. tabrate all agep, e(y)
. table of cases (D), person-years (Y), and rates per 1000 person-years

------------------------------------------------------------------------
       agep  |     _D          _Y       _rate     Ci_low     Ci_high
-------------+----------------------------------------------------------
          1  |     46      9774.8       4.706      3.525       6.283
          2  |    111     15057.3       7.372      6.120       8.879
          3  |    160     12335.9      12.970     11.109      15.144
          4  |    250     10771.5      23.209     20.504      26.272
          5  |    276      6919.2      39.889     35.450      44.884

Cisq test for unequal rates = 453.41 (df, p = 0.000)
------------------------------------------------------------------------
```

Bild 1 visar mortaliteten i en grupp män där åldern delats in i kvintiler där ”agegr 5” är de äldsta och ”agegr 1” de yngsta personerna. Raten ökar med stigande ålder, och det finns knappt någon överlappning i 95 % CI mellan åldersgrupperna. Längst ned till höger i bilden ser vi att p-värdet är mycket lågt. Det säger att vi med gott samvete kan förkasta nollhypotesen, dvs. att det inte fanns någon skillnad i rate i olika åldersgrupper.

Med en likartad analys kan vi se att även p-värdet för eventuell skillnad mellan dem med hög respektive låg utbildning var mycket lågt ($p < 0.0001$).

Eftersom ålder var associerad till mortaliteten, och i denna kohort högre hos dem med låg utbildning än dem med hög utbildning, är ålder en potentiell förväxlingsfaktor med avseende på associationen mellan utbildning och mortalitet. Med dataprogrammets hjälp kan vi räkna ut RR för

associationen mellan utbildning och mortalitet i dessa fem åldersgrupper (stratifiering med avseende på ålder).

Bild 2. Rate ratio 95 % CI och p-värde för effekten av utbildning i fem åldersgrupper.

```
Mhrate all grade, e(y) by (agep)
Maximum likelihood estimate of the rate ratio
Comparing grade = 2 vs grade = 1 by agep

RR estimate, lower and upper 95% confidence limits,
and chi-squared test for RR = 1 (dgree of freedom)
------------------------------------------------------------------
   agep  |     RR    Lower      Upper     Chisq     p_value
---------------------------------------------------------------
      1  |  1.858    0.944      3.658      3.316      0.069
      2  |  1.195    0.750      1.906      0.563      0.453
      3  |  1.546    1.105      2.164      6.552      0.010
      4  |  1.359    1.055      1.751      5.696      0.017
      5  |  1.326    1.042      1.687      5.295      0.021
Mantel-Haenszel estimate controlling for: agegp
               RR    Lower      Upper     Chisq     p_value
---------------------------------------------------------------
            1.377    1.192      1.589     19.169       0.000
------------------------------------------------------------------
```

I bild 2 ser vi att RR varierar något mellan grupperna, men inte så mycket att vi behöver räkna med någon betydande interaktion mellan ålder och utbildningsgrad med avseende på mortalitet. Programmet använder sig även av metoden *Mantel–Haenszel* för att sammanväga dessa fem olika RR till ett enda (det som ses på nedersta raden, dvs. 1.377). Denna metod ger även ett 95 % CI för detta nya RR, som nu är korrigerat för effekten av åldersvariationen, samt ett p-värde. Vi ser då att p-värdet för associationen mellan utbildning och mortalitet, när vi nu har justerat för ålderseffekten, fortfarande är mycket lågt. Korrigeringen för ålderseffekten gjorde att RR för utbildningsgrad sjönk från 2.1 till 1.4, dvs. ålder var en förväxling med avseende på associationen mellan utbildningsgrad och mortalitet. Ålder var däremot bara en partiell förväxlingsfaktor då RR fortfarande var > 1.0 och höggradigt signifikant.

I stället för ålder delar vi nu in rökvanor i fem klasser där ”1” står för icke-rökare och ”5” för storrökare (bild 3). Vi prövar nu rökningens roll som förväxlingsfaktor i stället för ålder. Det visar sig då att det nya justerade RR bara har sjunkit till 1.881, dvs. att rökning bara är en svag förväxlingsfaktor med avseende på associationen mellan utbildning och mortalitet.

Bild 3. Rate ratio 95 % CI och p-värde för effekten av utbildning given i fem grupper med avseende på rökning.

```
Mhrate all grade, e(y) by (smok)
Maximum likelihood estimate of the rate ratio
Comparing grade = 2 vs grade = 1 by smok

RR estimate, lower and upper 95% confidence limits,
and chi-squared test for RR = 1 (dgree of freedom)
------------------------------------------------------------------
   smok  |     RR    Lower      Upper     Chisq     p_value
---------------------------------------------------------------
      1  |  2.248    1.482      3.722      14.019      0.000
      2  |  1.870    1.457      2.399      24.998      0.000
      3  |  1.659    1.241      2.219      11.927      0.000
      4  |  2.173    1.641      2.877      30.857      0.000
      5  |  1.557    1.046      2.318       4.830      0.028
```

```
Mantel-Haenszel estimate controlling for: agegp
              RR    Lower      Upper     Chisq     p_value
------------------------------------------------------------------
           1.881    1.637     2.161     81.938       0.000
------------------------------------------------------------------
```

En av de klassiska metodernas begränsningar är att man på ett effektivt sätt bara kan undersöka en förväxlingsfaktor i taget. En annan svaghet är att man inte kan få en uppskattning av RR, 95 % CI och p-värden för båda exponeringarna samtidigt – då måste man gå vidare med andra modeller.

Ett annat analysproblem uppkommer när vi arbetar med kohorter där individerna har inkluderats under en lång tid; en individ som t.ex. var 40 år på 1940-talet hade i allmänhet en kortare förväntad livslängd än en 40-åring som levde på 1980-talet. Vidare förändras en rate med åldern. Vi kan se individernas expositionstid, dvs. den tid de har varit med i studien, både längs en åldersaxel (bild 4) och en kalenderaxel (bild 5). Oftast delas båda dessa in i perioder av 5 eller 10 år.

Bild 4. Tre individers expositionstid (de horisontella linjerna under åldersskalan) i förhållande till deras ålder vid inklusion i studien. D = avliden, C = censurerad (levande vid eller missad i uppföljningen).

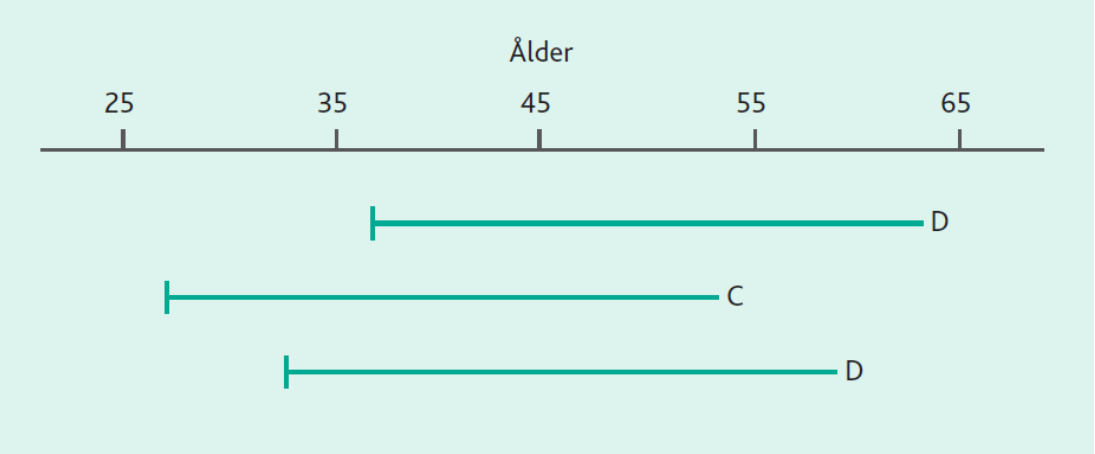
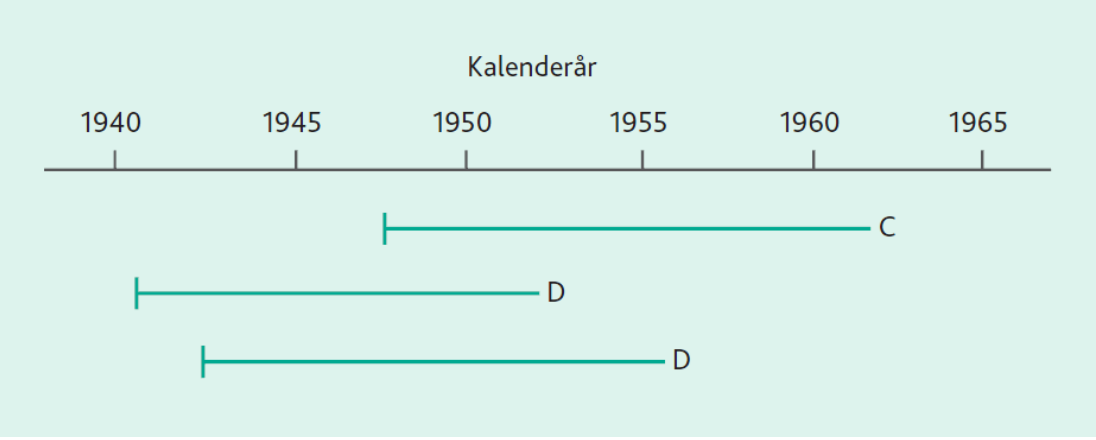
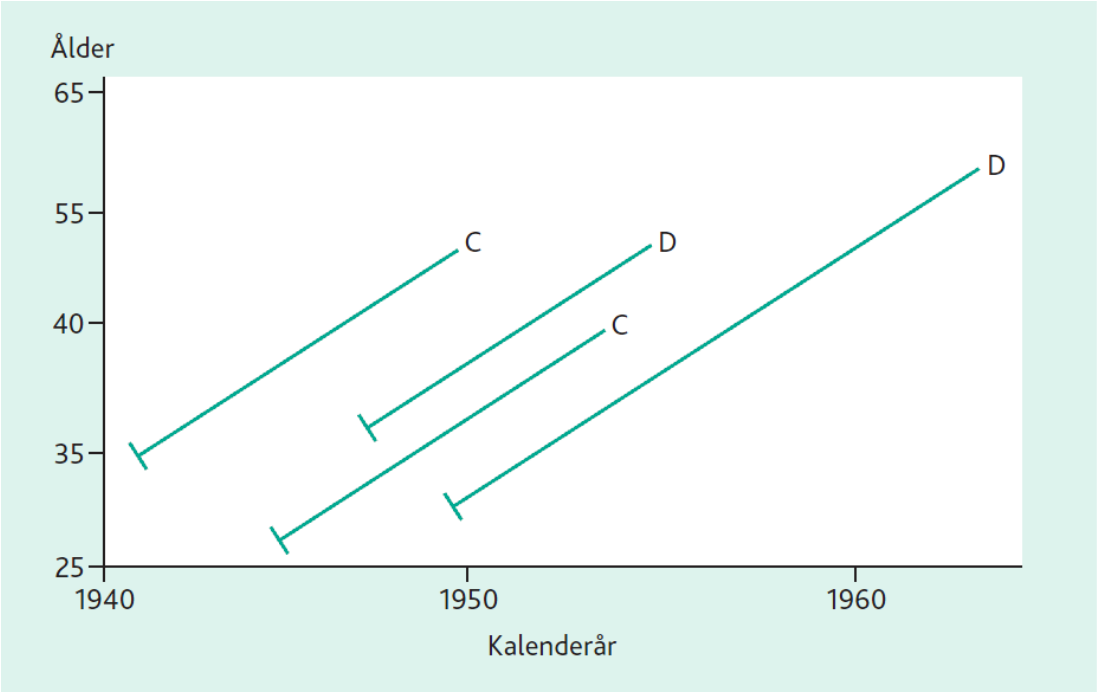


Bild 5. Tre individers expositionstid (de horisontella linjerna under åldersskalan) i förhållande till kalenderår vid inklusion i studien. D = avliden, C = censurerad (levande vid eller missad i uppföljningen).



Med hjälp av ett s.k. *Lexis-diagram* kan vi illustrera expositionstiden i förhållande till både åldern vid inklusion och till kalenderåret (bild 6).

Bild 6. Fyra individers expositionstid (de sneda parallella linjerna) i förhållande till kalenderår (x-axeln) och ålder vid inklusion i studien (y-axeln). D = avliden, C = censurerad (levande vid eller missad i uppföljningen).



På detta sätt kan vi beräkna både kalender- och åldersspecifika rater. Det går också att beräkna rate ratio för olika expositioner under olika kalender- eller åldersperioder för att se om de är lika.

Exempel 5.8. Faktoriell ANOVA med post hoc-test

I stället för att undersöka skillnaden i BMI mellan dem som tar/inte tar blodtrycksmedicin använder vi en exponering som består av tre grupper:
– inte medicin med normalt blodtryck (grupp 0)
– inte medicin men med förhöjt blodtryck (grupp 1)
– blodtrycksmedicin (grupp 2).

Vi gör då först en faktoriell ANOVA som visar att det är stor skillnad mellan de tre grupperna.

```
. table btvariabel, c( n bmi mean bmi sd bmi)

-----------------------------------------------
btvariabe |
l         |     N(bmi)    mean(bmi)      sd(bmi)
----------+------------------------------------
        0 |         310     25.67032     3.759436
        1 |         377     27.09947       4.3578
        2 |         316     28.27658     4.45190
```

```
. anova bmi btvariabel

                           Number of obs =    1003     R-squared     =  0.0567
                           Root MSE      = 4.21316     Adj R-squared =  0.0548

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  1065.97523     2  532.987617      30.03     0.0000
                         |
              btvariabel |  1065.97523     2  532.987617      30.03     0.0000
                         |
                Residual |  17750.7336  1000  17.7507336
              -----------+----------------------------------------------------
                   Total |  18816.7089  1002  18.7791506
```

Därefter gör vi ett post hoc-test av dessa tre grupper.

```
. oneway bmi  btvariabel, bon

                      Comparison of BMI by btvariabel
                                (Bonferroni)
Row Mean-|
Col Mean |          0          1
---------+----------------------
       1 |    1.42915
         |      0.000
         |
       2 |    2.60626    1.17711
         |      0.000      0.001
```

Om vi vill jämföra alla tre grupperna med varandra blir det kritiska p-värdet för vilket vi förkastar nollhypotesen < 0.0166 (Bonferroni-korrektionen 0,05/x, där x motsvarar antalet post hoc-test). Om hypotesen i stället är att BMI är förhöjt vid högt blodtryck är det ju bara jämförelserna mellan gruppen med normal blodtryck och de två grupperna med förhöjt blodtryck som är relevanta. Vi behöver då bara göra två post hoc-test. Det kritiska p-värdet blir då 0.0250.

Trots Bonferroni-korrektionen är p-värdena mycket låga (andra raden i varje jämförelse, första raden är skillnaden i medelvärden) och vi kan därför lugnt konstatera att de tre grupperna skiljer sig åt sinsemellan.

Exempel 5.9. Oparat t-test

I detta exempel har vi bara två grupper i den kategoriska exponerings-variabeln (blodtrycksmedicin/inte blodtryckmedicin). Då lämpar det sig bäst att använda oparat t-test.

```
. ttest bmi , by(blodtrycksmedicin)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     690     26.4642    .1581174    4.153406    26.15375    26.77465
       1 |     317    28.27697    .2496482     4.44486    27.78579    28.76815
---------+--------------------------------------------------------------------
combined |    1007    27.03486    .1363823    4.327855    26.76723    27.30248
---------+--------------------------------------------------------------------
    diff |           -1.812769    .2881798               -2.378272   -1.247266
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -6.2904
Ho: diff = 0                                     degrees of freedom =     1005

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Förutom att p-värdet är < 0.05 (mellersta p-värdet i sista raden) ser vi att 95 % CI för skillnaden mellan gruppen med respektive utan blodtrycks-medicin (diff) ligger mellan –2.378272 och –1.247266. Detta konfidens-intervall innehåller inte talet 0 och nollhypotesen kan därför förkastas.

Exempel 5.10 Tvåvägs variansanalys

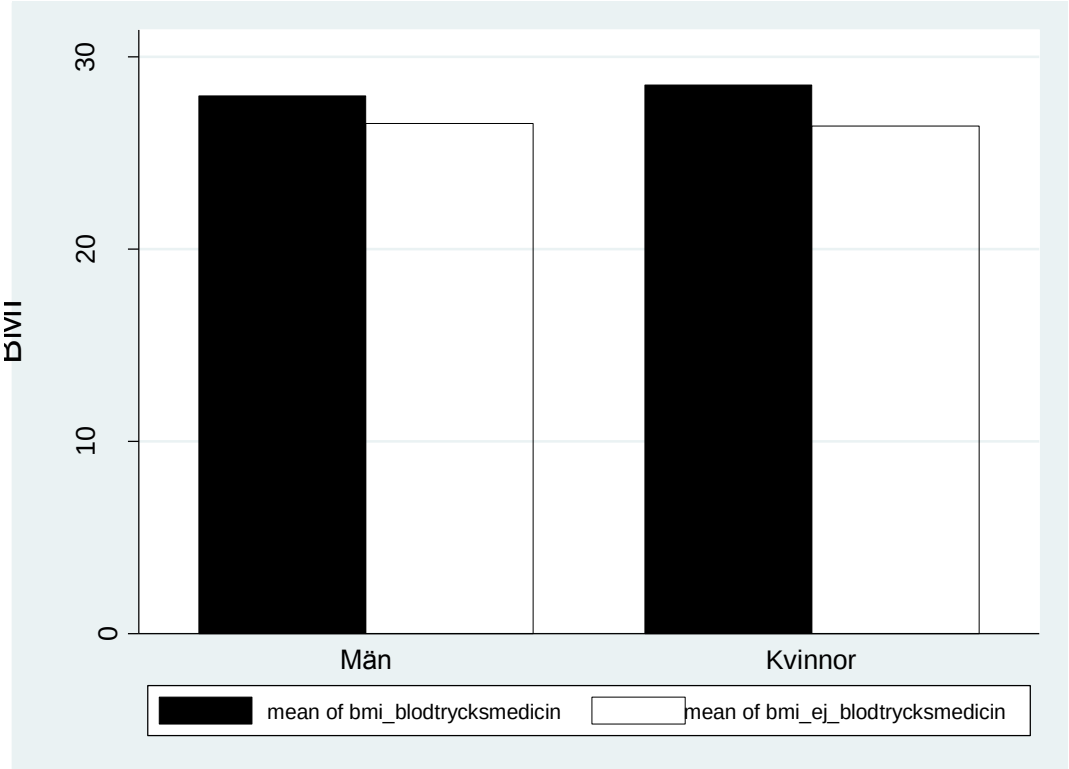
I en tvåvägs variansanalys (two-way ANOVA) ingår två kategoriska exponeringsvariabler.

```
. anova bmi blodtrycksmedicin##sex
```

```
                           Number of obs =    1007     R-squared     =  0.0398
                           Root MSE      = 4.24718     Adj R-squared =  0.0369

                  Source |  Partial SS    df       MS           F     Prob > F
-----------------------+----------------------------------------------------
                   Model |  750.013564     3  250.004521       13.86     0.0000
                         |
        blodtryck~n |  714.738463     1  714.738463       39.62     0.0000
                     sex |  12.7907162     1  12.7907162        0.71     0.4000
        blodtryck~n#sex |  33.8982581     1  33.8982581        1.88     0.1707
                         |
                Residual |  18092.6931  1003  18.0385773
-----------------------+----------------------------------------------------
                   Total |  18842.7066  1006  18.7303247
```

På raden ”blodtryck~n#sex ” ser vi interaktionstermen. Om detta p-värde är < 0.05 betyder det att det finns en signifikant könsinteraktion med avseende på skillnaden i BMI mellan dem som får respektive inte får blodtrycksbehandling. Eftersom p-värdet inte är lågt i detta fall, verkar det inte ha någon viktig könseffekt. Detta exemplifieras grafiskt i bild 7 som visar medelvärden i grupperna med respektive utan blodtrycksmedicin uppdelade på kön. Skillnaden mellan dem som tar/inte tar blodtrycksmedicin visar sig här vara ganska lika för män och kvinnor.

Bild 7. Medelvärden i grupperna med respektive utan blodtrycksmedicin uppdelade på kön.



Exempel 5.11. Faktoriell ANOVA

I detta exempel studerades 40 råttor som fick två olika interventioner i en s.k. faktoriell design. Den ena interventionen var injektioner av miljögiftet PCB, den andra var att äggstockarna togs bort så att östrogennivåerna minskade (OVX). Eftersom PCB anses utöva en del av sin giftiga verkan via östrogen-receptorer, var det intressant att studera interaktionen av dessa två inter-ventioner. I detta försök studerades effekten på benbildning, och i exemplet utvärderades effekten på slutvikten (slutvikt(g)).

I den faktoriella designen togs äggstockarna bort på hälften av råttorna, medan den andra hälften fick genomgå samma procedur, dvs. narkos och snitt i huden m.m. utan att äggstockarna togs bort. Denna ”placebo-operation” brukar kallas *sham-operation* eller *skenoperation*. Hälften av de råttor vilkas äggstockar opererades bort och hälften av dem som genomgick sham-operation fick sedan PCB-injektioner. Resten fick injektioner av det lösningsmedel som PCB var löst i, s.k. placeboinjektioner.

Detta resulterade i fyra olika grupper med lika många observationer (10) i varje (tabell 2).

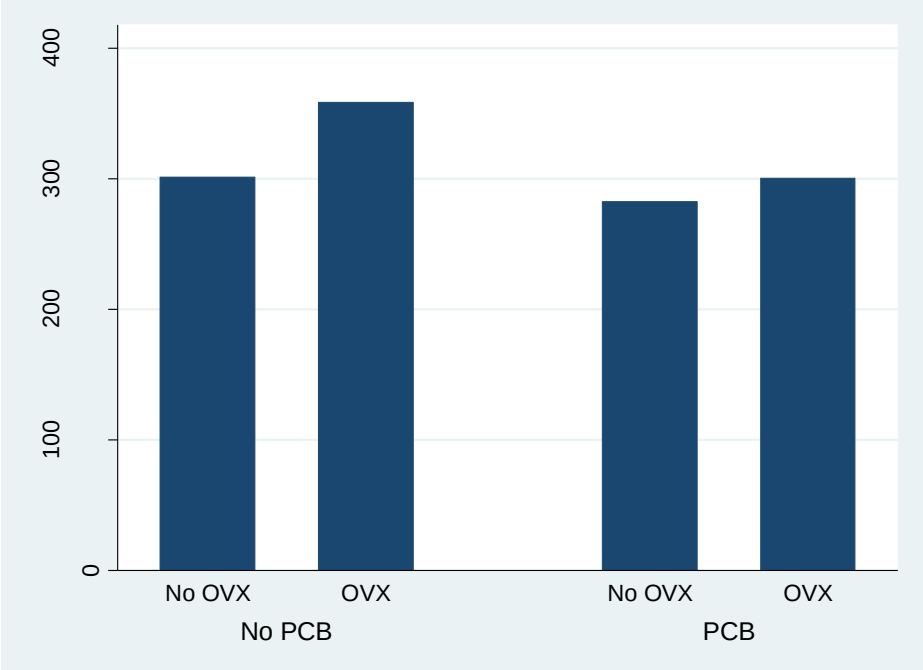
Tabell 2. Exempel på en 2 × 2 faktoriell design i en interventionsstudie.

OVX + PCB	OVX + placebo
Sham-operation + PCB	Sham-operation + placebo

Härmed kan vi både utvärdera effekten av att ta bort äggstockarna och av att ge PCB. Egentligen skulle vi bara behöva göra en envägs variansanalys med gruppvariabeln OVX/sham-operation som faktor. På samma sätt skulle vi kunna göra en envägs variansanalys för att ta reda på effekten av PCB/placebo, eftersom grupperna är identiska med avseende på den andra faktorn. Det blir dock mycket enklare att göra detta i en enda analys, en tvåvägs variansanalys, där vi dessutom får information om det finns någon samverkan mellan interventionerna, interaktionstermen.

I stapeldiagrammet (bild 8) ser vi att medelvärdet för de två PCB-staplarna ligger under medelvärdet för de två placebostaplarna (No PCB). Vi ser också att medelvärdet för de två OVX-staplarna är högre än hos de sham-opererade råttorna (No OVX).

Bild 8.



p-värdena i ANOVA-tabellen (nedan) visar ett mycket lågt p-värde både för raden ”ovx” och för raden ”pcb”. Dessutom är p-värdet för interaktions-termen ”ovx#pc” lågt.

```
. anova  slutviktg ovx##pcb
```

```
                          Number of obs =      38     R-squared     =  0.6922
                          Root MSE      = 19.1304     Adj R-squared =  0.6651

      Source |  Partial SS    df       MS           F     Prob > F
-------------+----------------------------------------------------
       Model |  27988.8947     3  9329.63158       25.49     0.0000
             |
         ovx |  13270.6118     1  13270.6118       36.26     0.0000
         pcb |  13914.3765     1  13914.3765       38.02     0.0000
     ovx#pcb |  3708.44706     1  3708.44706       10.13     0.0031
             |
    Residual |       12443    34  365.970588
-------------+----------------------------------------------------
       Total |  40431.8947    37  1092.75391
```

Resultatet av denna studie kan sammanfattas som så att PCB-exponeringen minskade vikten, medan OVX ökade vikten. Dessutom var OVX mindre uttalad för råttor som fått PCB. Tack vare denna analysmetod fick vi mer information om effekten av PCB med samma antal råttor än om vi hade gjort två separata studier, dvs. en med PCB/placebo till 20 råttor och en med OVX/sham-operation till 20 andra råttor.

Exempel 5.12. Variansanalys för upprepade mätningar

Med variansanalys för upprepade mätningar (ANOVA for repeated measurements) undersöker man om det finns en skillnad i medelvärden mellan olika tidpunkter. Liksom vid den faktoriella variansanalysen får man inget svar på mellan vilka specifika mättillfällen det kan föreligga en skillnad, utan det är ett test på den sammanvägda variansen av alla mätningar.

Nedan undersöker vi hur det systoliska blodtrycket utvecklas över 5 år hos en stor grupp 70-åringar. Blodtrycket mäts vid 70 respektive 75 års ålder.

```
. sum  SBP70 SBP75

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       SBP70 |       335    139.8925    19.32949         84        200
       SBP75 |       335    145.0597     19.2928         94        196

. anova SBP lpnr year, repeated(year)

                           Number of obs =     670     R-squared     =  0.8549
                           Root MSE      = 10.4977     Adj R-squared =  0.7093

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  216775.799   335  647.091936        5.87     0.0000
                         |
                    lpnr |  212303.618   334  635.639575        5.77     0.0000
                    year |   4472.1806     1   4472.1806      40.58     0.0000
                         |
                Residual |  36807.3194   334  110.201555
              -----------+----------------------------------------------------
                   Total |  253583.118   669  379.048009
```

Resultatet visar att det skiljer drygt 5 mm Hg mellan 70 år och 75 år. På raden för ”year” ses ett mycket lågt p-värde. Det betyder att vi kan förkasta noll-hypotesen att det inte är någon skillnad i blodtryck mellan 70 år och 75 år.

Exempel 5.13. Parat t-test

Nedan visas ett exempel på ett parat t-test (paired t-test) av huruvida blodtrycket (”SPB”) skiljer sig vid 70 respektive 75 års ålder hos samma individer.

```
. ttest SBP70 = SBP75

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   SBP70 |     335    139.8925    1.056083    19.32949    137.8151    141.9699
   SBP75 |     335    145.0597    1.054078     19.2928    142.9862    147.1332
---------+--------------------------------------------------------------------
    diff |     335   -5.167164    .8111225    14.84598   -6.762717   -3.571612
------------------------------------------------------------------------------
     mean(diff) = mean(SBP70 - SBP75)                              t =  -6.3704
 Ho: mean(diff) = 0                               degrees of freedom =      334

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Differensen 0 ligger här inte ligger inom konfidensintervallet (–6.762717 till –3.571612). Vi kan därför förkasta nollhypotesen att skillnaden mellan de olika mättillfällena var 0. p-värdet för detta ges av p-värdet i mitten på den nedersta raden.

Exempel 5.14. Variansanalys för upprepade mätningar med en faktor

Ovan konstaterade vi att blodtrycket (”SBP”) stiger med åldern. Frågan är nu om denna stegring är större hos dem som är överviktiga (BMI > 30) vid 70 års ålder än hos dem som inte är överviktiga.

```
. anova SBP fet / lpnr|fet year year#fet ,repeated(year)

                           Number of obs =     358     R-squared     =  0.8290
                           Root MSE      = 11.0877     Adj R-squared =  0.6551

                  Source |  Partial SS    df       MS           F     Prob > F
-----------------------+----------------------------------------------------
                   Model |  105491.268   180  586.062599        4.77     0.0000
                         |
                     fet |   865.11642     1   865.11642        1.50     0.2225
               lpnr|fet |  102178.694   177  577.280755
-----------------------+----------------------------------------------------
                    year |  880.415893     1  880.415893        7.16     0.0081
              year#fet |  36.7287419     1  36.7287419        0.30     0.5854
                         |
                Residual |  21760.0422   177  122.938092
-----------------------+----------------------------------------------------
                   Total |   127251.31   357  356.446247
```

Raden ”year#fet” indikerar interaktionen mellan blodtryckets utveckling med åren och fetma. p-värdet är högt vilket betyder att vi inte kan förkasta nollhypotesen att fetma vid 70 år inte påverkar blodtrycksutvecklingen mellan 70 och 75 år.

Exempel 5.15 Kovariansanalys (ANCOVA)

Vi fortsätter nu med ovanstående exempel där vi undersökte huruvida BMI skilde sig åt mellan blodtrycksmedicinerande/icke-blodtrycksmedicinerande personer. Vi undersökte även könets effekt på denna relation. Då även åldern kan påverka detta samband vill vi nu också undersöka denna effekt, varför ålder (”age”) tas med som kontinuerlig variabel i modellen. Vi börjar då med att göra en kovariansanalys.

```
. anova bmi blodtrycksmedicin sex c.age
```

```
                           Number of obs =    1007     R-squared     =  0.0389
                           Root MSE      = 4.24919     Adj R-squared =  0.0360

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  732.958396     3  244.319465      13.53     0.0000
                         |
              blodtryck~n |  723.189273     1  723.189273      40.05     0.0000
                     sex |  .116676093     1  .116676093       0.01     0.9359
                     age |  16.8430904     1  16.8430904       0.93     0.3344
                         |
                Residual |  18109.7482  1003  18.0555815
              -----------+----------------------------------------------------
                   Total |  18842.7066  1006  18.7303247
```

Det visar sig då att blodtrycksmedicin är relaterad till BMI oberoende av ålder och kön, något vi nu prövar i en multipel linjär regressionsmodell.

```
. reg bmi  blodtrycksmedicin sex age
```

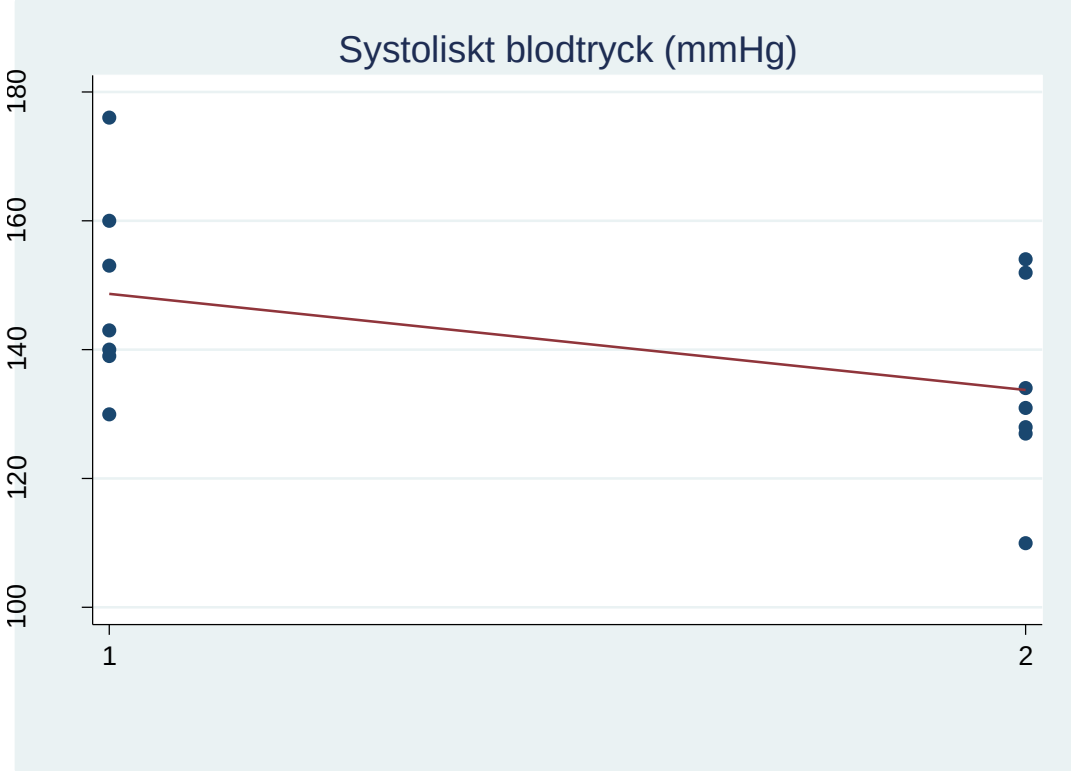
```
      Source |       SS       df       MS              Number of obs =    1007
-------------+------------------------------           F(  3,  1003) =   13.53
       Model |  732.958396     3  244.319465           Prob > F      =  0.0000
    Residual |  18109.7482  1003  18.0555815           R-squared     =  0.0389
-------------+------------------------------           Adj R-squared =  0.0360
       Total |  18842.7066  1006  18.7303247           Root MSE      =  4.2492

------------------------------------------------------------------------------
         bmi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
blodtrycks~n |   1.827085   .2886945     6.33   0.000     1.260571    2.393599
         sex |  -.0237347   .2952559    -0.08   0.936    -.6031249    .5556554
         age |   .0025298   .0026193     0.97   0.334    -.0026101    .0076696
       _cons |   26.33331   .2266148   116.20   0.000     25.88862    26.77801
------------------------------------------------------------------------------
```

p-värdena för dessa exponeringar (”blodtrycksmedicin”, kön (”sex”) och ålder ”age”) visar sig vara desamma i såväl kovariansanalysen som i den linjära regressionsmodellen. Till skillnad från kovariansanalysen ger regressionsmodellen inte kvadratmedelvärdet (”MS”) för varje ingående exponeringsvariabel, utan enbart för hela modellen och residualen.

Exempel 5.16. Linjär regression

Låt oss ta ett exempel som vi normalt inte använder oss av i praktiken – en regressionsanalys mellan ett utfall med en mängd olika värden och en exponering med bara två värden. Resultatet visas i ett regressionsdiagram (bild 9).

Bild 9. Regressionsdiagram över en beroende kontinuerlig variabel med flera olika värden och en oberoende variabel med bara två värden.

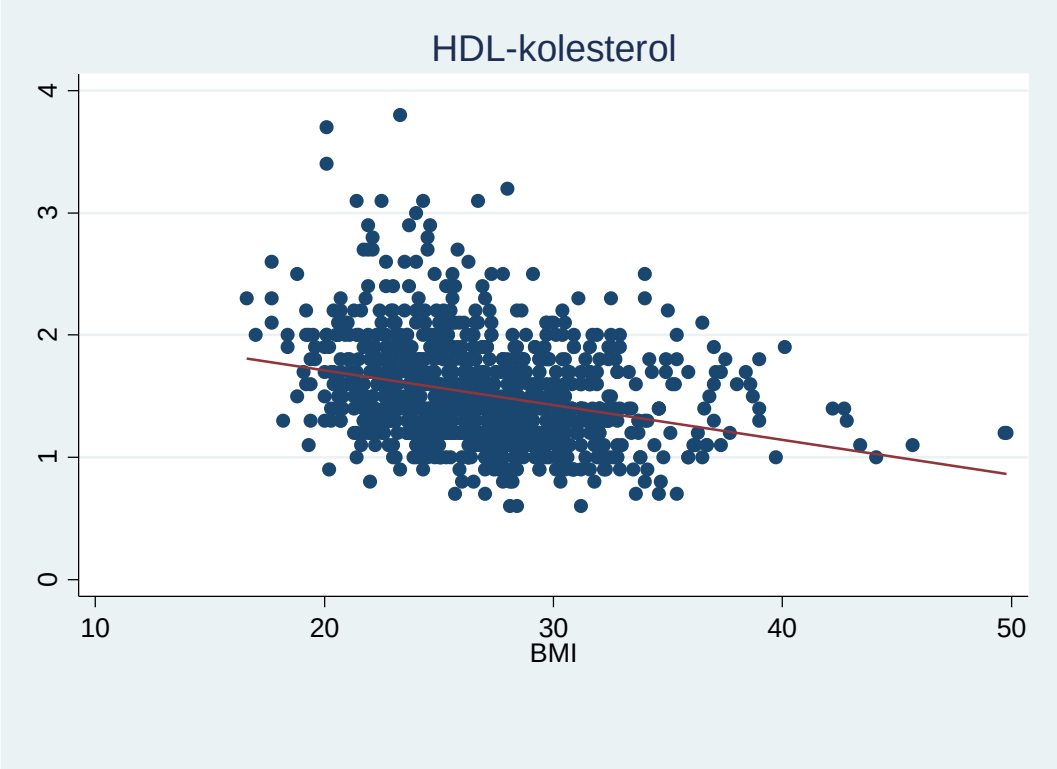


Här hamnar det systoliska blodtrycket på y-axeln (utfallet), medan kön är kodat som 1 eller 2 på x-axeln (exponeringen). I detta fall går regressionslinjen från medelvärdet för observationerna ovanför 1 (männen) till medelvärdet för observationerna ovanför 2 (kvinnorna).

Dessa medelvärden utgör summan av kvadraterna av avståndet mellan de enskilda observationerna. Medelvärdet är så litet som möjligt om man ger denna kvadrat (variansen) olika tecken beroende på om de ligger ovanför eller nedanför den tänkta regressionslinjen. I detta fall, med bara två grupper på x-axeln, kommer linjen av matematiska skäl därför att gå igenom de båda medelvärdena.

Nu behöver ju en relation inte alltid se ut som i exemplen i boken (t.ex. bild 5.17 och 5.18), utan en regressionslinje kan se ut som i bild 10 där HDL-kolesterol relaterats till BMI.

Bild 10. Exempel på en invers korrelation.



I detta fall lutar regressionslinjen åt motsatt håll jämfört med de tidigare exemplen. Där fanns det en direkt (positiv) relation mellan de två kontinuerliga variablerna, medan relationen i bild 10 är invers (negativ). Man kan annars tillämpa samma resonemang som tidigare när det gäller överlappning mellan varianserna, korrelationskoefficient, R^2 m.m.

Bild 11. Regressionsanalys.

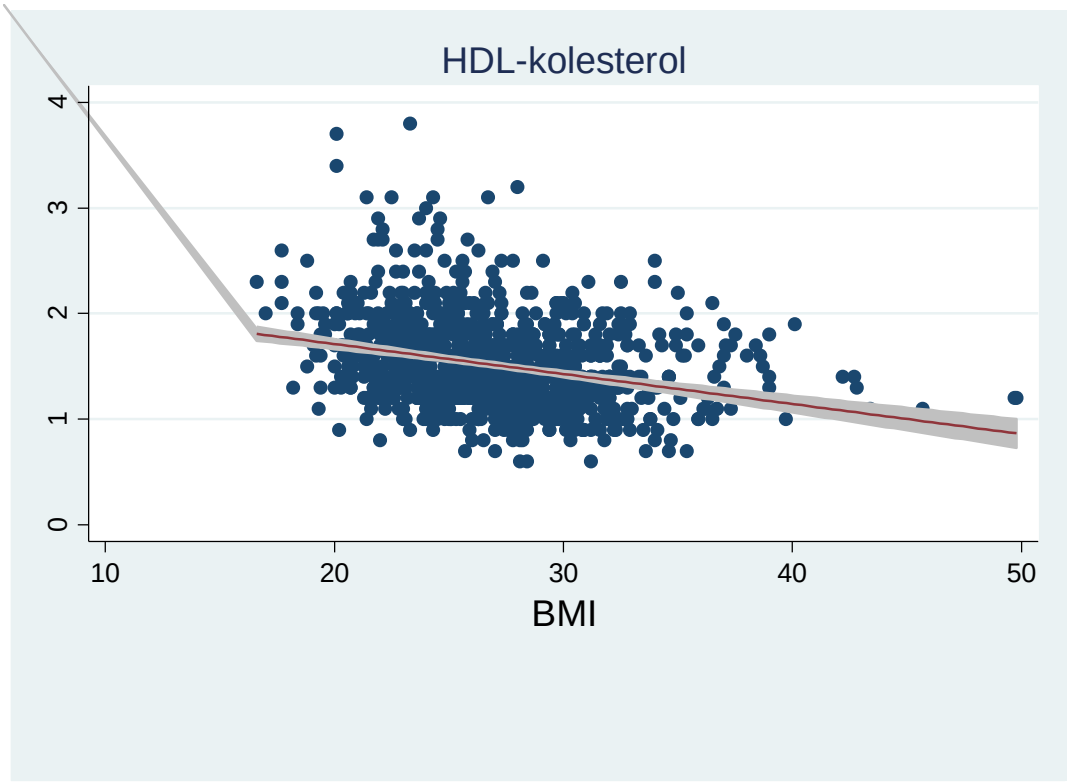
```
. reg hdl bmi

      Source |       SS       df       MS              Number of obs =    1013
-------------+------------------------------           F(  1,  1011) =   91.65
       Model |  15.3271454     1  15.3271454           Prob > F      =  0.0000
    Residual |  169.083058  1011  .167243381           R-squared     =  0.0831
-------------+------------------------------           Adj R-squared =  0.0822
       Total |  184.410203  1012  .182223521           Root MSE      =  .40895

------------------------------------------------------------------------------
         hdl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         bmi |  -.0284277   .0029695    -9.57   0.000    -.0342549   -.0226006
       _cons |   2.280447   .0813187    28.04   0.000     2.120874     2.44002
------------------------------------------------------------------------------
```

I den räta linjens ekvation (bild 11) finns det ett minustecken före lutningskoefficienten. Det innebär att för varje enhetsökning på x-axeln kommer y-värdet att minska med 0.0284277 enheter. I denna regressionsanalys (som i bild 11) får relationen mellan de två variablerna automatiskt en signifikansnivå . Man använder då det p-värde som står till höger på raden för exponeringen och *inte* det p-värde på raden över, som gäller för skärningspunkten (intercept). I exemplet ovan är p-värdet lågt (0.001), vilket gör att vi kan förkasta nollhypotesen att relationen mellan de två variablerna bara är ett fall av slumpen. I samma bild kan vi också se att

siffran 0 inte ingår i det 95 % CI för regressionskoefficienten (–0.0342549 till –0.0226006). Detta är ett annat sätt att visa att nollhypotesen kan förkastas.

Man kan även grafiskt visa 95 % CI för en regressionslinje. I bild 12 illustreras detta av den grå arean runt regressionslinjen. Hur vi än placerar regressionslinjen inom konfidensintervallet får vi varken en positiv relation eller en horisontell linje.

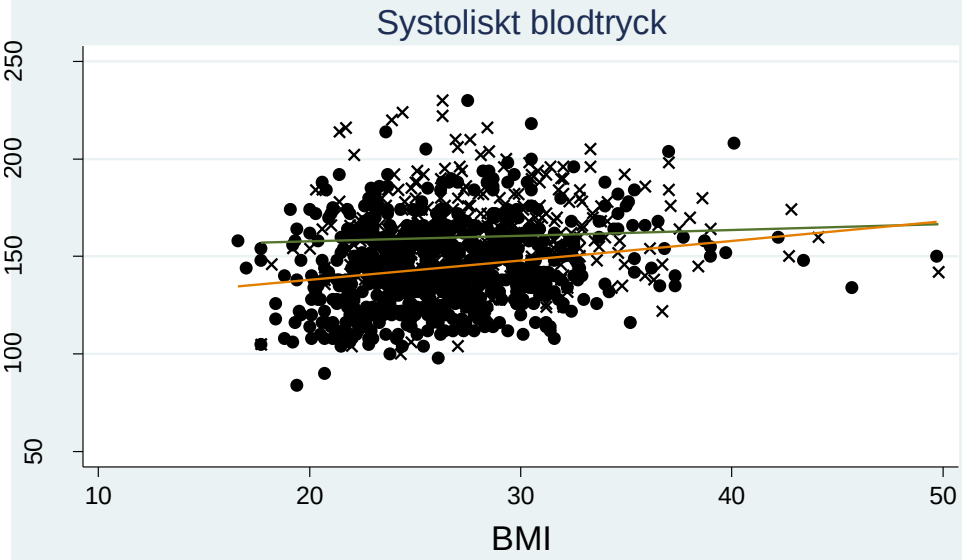
Bild 12.



Exempel 5.17. Att jämföra två regressionslinjer

I detta exempel studerar vi om relationen mellan blodtryck och BMI är likartad hos män och kvinnor. Precis som vid variansanalys inför vi då en interaktionsterm i regressionsmodellen.

Bild 13. Relationen mellan blodtryck och BMI. Kryssen motsvarar observationer för personer som tar blodtrycksmedicin, medan cirklarna representerar dem som inte tar någon medicin.



Regressionslinjen för personer utan blodtrycksmedicin är flackare än regressionslinjen för personer med blodtrycksmedicin.

```
. gen bmibtmed=blodtrycksmedicin*bmi

. reg SBP blodtrycksmedicin bmi bmibtmed
```

Source	SS	df	MS		
Model	70250.1851	3	23416.7284	Number of obs	1003
Residual	444477.789	999	444.922712	F(3, 999)	52.63
Total	514727.974	1002	513.700573	Prob > F	0.0000
				R-squared	0.1365
				Adj R-squared	0.1339
				Root MSE	21.093

SBP	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
blodtrycks~n	37.12588	9.235948	4.02	0.000	19.00179	55.24996
bmi	.9933639	.1937271	5.13	0.000	.6132053	1.373523
bmibtmed	-.796501	.3298427	-2.41	0.016	-1.443765	-.1492371
_cons	118.3265	5.187769	22.81	0.000	108.1463	128.5067

Först skapar vi produkten (= interaktionstermen, ”bmibtmed”) mellan blodtrycksmedicin och BMI. När vi sedan tar med denna term i regressionsmodellen ser vi att p-värdet blir 0.016, dvs. vi kan förkasta nollhypotesen att regressionslinjerna för relationen mellan SBP och BMI är lika för personer som tar respektive inte tar blodtrycksmedicin.

Exempel 5.18. Icke-linjär regression

Den icke-linjära ekvationen för regressionslinjen blir då y = 2.429911 x – 0.0141462 x 2 –3.129848.

```
. reg y x

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  1,     8) =   45.31
       Model |  10114.1941     1  10114.1941            Prob > F      =  0.0001
    Residual |  1785.80591     8  223.225739            R-squared     =  0.8499
-------------+------------------------------           Adj R-squared =  0.8312
       Total |       11900     9  1322.22222            Root MSE      =  14.941

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |    1.01051   .1501229     6.73   0.000     .6643258    1.356694
       _cons |   18.34192    9.26649     1.98   0.083    -3.026643    39.71049
------------------------------------------------------------------------------

. gen x2=x*x

. reg y x x2

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  2,     7) =  320.57
       Model |  11771.4783     2  5885.73916            Prob > F      =  0.0000
    Residual |  128.521683     7  18.3602405            R-squared     =  0.9892
-------------+------------------------------           Adj R-squared =  0.9861
       Total |       11900     9  1322.22222            Root MSE      =  4.2849

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   2.429911   .1554783    15.63   0.000     2.062263    2.797558
          x2 |  -.0141462    .001489    -9.50   0.000     -.017667   -.0106254
       _cons |  -3.129848   3.488582    -0.90   0.399    -11.37903    5.119337
------------------------------------------------------------------------------
```

Exempel 5.19. Multipel regression

Vi återgår till exemplet med relationen mellan blodtryck och BMI då även ålder och kön togs med som oberoende variabler i såväl en kovariansanalys som en regressionsmodell. I båda fallen visade sig exponeringen BMI vara relaterad till utfallet blodtryck, oberoende av effekterna av ålder och kön. En eventuell förväxlingseffekt av ålder och kön förklarar alltså inte fullständigt relationen mellan blodtryck och BMI.

I ett annat fall har vi funnit att både BMI och HDL-kolesterolnivåerna är relaterade till blodtrycket i separata enkla linjära regressionsmodeller (inverst för HDL). Om vi nu sätter in både BMI och HDL-kolesterol som oberoende variabler i en multipel modell visar det sig att enbart BMI är signifikant relaterat till blodtrycket. Med andra ord är BMI och HDL-kolesterol inte båda oberoende av varandra relaterat till blodtrycket.

```
. reg SBP bmi hdl

      Source |       SS       df       MS              Number of obs =    1009
-------------+------------------------------           F(  2,  1006) =   21.94
       Model |  21491.0065     2  10745.5032            Prob > F      =  0.0000
    Residual |  492796.135  1006  489.856993            R-squared     =  0.0418
-------------+------------------------------           Adj R-squared =  0.0399
       Total |  514287.142  1008  510.205498            Root MSE      =  22.133

------------------------------------------------------------------------------
         SBP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         bmi |   1.110137   .1680023     6.61   0.000     .7804615    1.439812
         hdl |   2.516823   1.706802     1.47   0.141    -.8324772    5.866123
       _cons |   115.8778   5.876796    19.72   0.000     104.3457      127.41
------------------------------------------------------------------------------
```

Exempel 5.20. Urval av variabler och stegvis regression

Ibland vill man pröva om en mängd olika variabler är relaterade till ett utfall. Det kan då vara lämpligt att som en första åtgärd utföra enkla regressioner mellan den beroende variabeln och de övriga var för sig. Ett annat sätt att skaffa sig en snabb uppfattning om vilka variabler som kan vara intressanta är att göra en korrelationsmatris. Då analyserar man korrelationerna för alla kombinationer mellan de olika variablerna (bild 14).

Bild 14. Korrelationsmatris.

```
. pwcorr lvmim27 SBP DBP bmi kn ldl hdl diabetes

             |  lvmim27      SBP      DBP      bmi       kn      ldl      hdl
-------------+---------------------------------------------------------------
     lvmim27 |   1.0000
         SBP |   0.3583   1.0000
         DBP |   0.2579   0.5963   1.0000
         bmi |   0.4764   0.1990   0.2517   1.0000
          kn |  -0.0751   0.1620  -0.0637   0.0102   1.0000
         ldl |  -0.1057   0.0873   0.0780  -0.0364   0.1607   1.0000
         hdl |  -0.2104  -0.0141  -0.0813  -0.2883   0.3533   0.0951   1.0000
    diabetes |   0.1999   0.0591  -0.0020   0.1533  -0.0549  -0.2227  -0.1398
```

I första kolumnen finns de sju andra variablerna som kan vara korrelerade med hjärtmuskelmassan (”lmvim2”7). Med en enkel korrelationsanalys får vi fram att BMI är den högsta korrelationskoefficienten följt av SBP m.fl. Med detta antal observationer vet vi att en korrelationskoefficient > 0.10 ger ett p-värde < 0.05, dvs. att alla de sju undersökta variablerna, förutom kön, är signifikant relaterade till hjärtvolymen.

Vi kan även leta efter kombinationer av variabler med hög korrelations-koefficient. Vi ser då att det bara är paret systoliskt och diastoliskt blodtryck som är nära relaterade (r = 0.60). Alla andra kombinationer har r < 0.50. Det är därför knappast av värde att inkludera både systoliskt och diastoliskt blod-tryck i en multipel regressionsmodell, något som vi tidigare har sett exempel på. Inom ett sådant par väljer man därför den variabel som är bäst relaterad till den tilltänkta beroende variabeln. Vi utför därefter en multipel regressionsmodell med alla de sex andra oberoende variablerna (bild 15).

Bild 15. Multipel regressionsmodell.

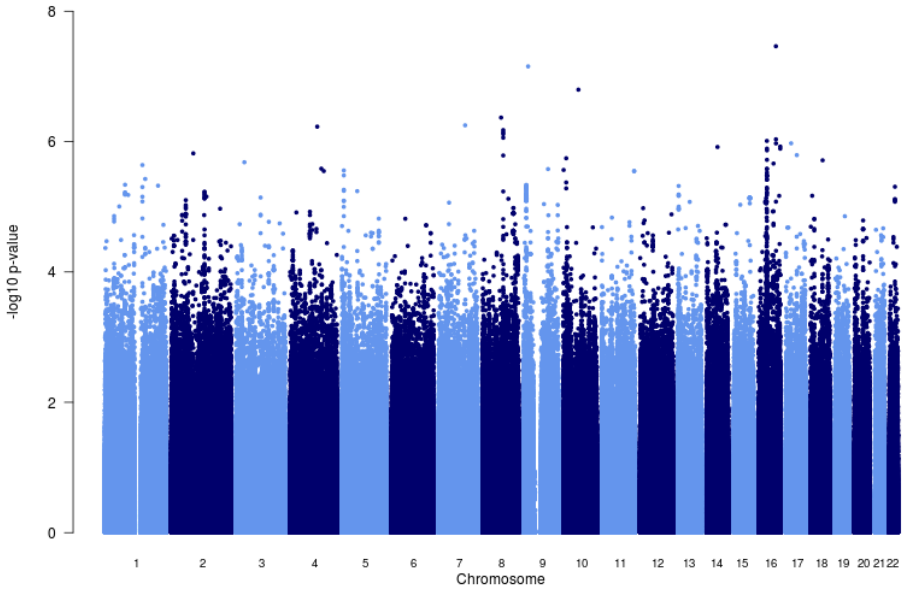
```
. reg lvmim27 SBP  bmi kn ldl hdl diabetes

      Source |       SS       df       MS              Number of obs =     908
-------------+------------------------------           F(  6,   901) =   78.10
       Model |  54114.7752     6   9019.1292           Prob > F      =  0.0000
    Residual |  104047.704   901  115.480249           R-squared     =  0.3421
-------------+------------------------------           Adj R-squared =  0.3378
       Total |  158162.479   907  174.379801           Root MSE      =  10.746

------------------------------------------------------------------------------
     lvmim27 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         SBP |   .1743452   .0163547    10.66   0.000     .1422474     .206443
         bmi |   1.273197   .0935793    13.61   0.000     1.089538    1.456856
          kn |  -2.360415   .7856236    -3.00   0.003     -3.90228   -.8185496
         ldl |  -1.263339   .4236189    -2.98   0.003    -2.094734   -.4319448
         hdl |  -1.150226   .9449237    -1.22   0.224    -3.004734    .7042814
    diabetes |   5.527854   1.339811     4.13   0.000      2.89834    8.157367
       _cons |  -10.37007   3.873624    -2.68   0.008    -17.97244   -2.767693
------------------------------------------------------------------------------
```

Här uppvisar alla sex variablerna, förutom HDL, ett p-värde < 0.05. HDL kommer därför egentligen bara att vara en belastning för modellen. Ett sätt att lösa detta problem är att i olika steg plocka bort variabler med p-värde > 0.05 tills det bara finns kvar variabler med p < 0.05. Detta brukar kallas *stegvis multipel regression*. Detta sätt att plocka bort en icke-signifikant variabel i taget brukar kallas *backward selection*. Man kan förstås lika gärna använda *forward selection*, dvs. plocka in en oberoende variabel i taget i modellen så länge den sist inkluderade får ett p < 0.05 i modellen. Det är viktigt att komma ihåg att stegvisa modeller bara väljer ut de statistiskt starkast kopplade variablerna i ett urval, inte de biologiskt mest relevanta variablerna för den sanna populationen.

Under senare år har en tillämpning av linjär regressionsanalys blivit vanlig, den s.k. genome-wide association study (GWAS). Detta är inte en jättestor multipel regression, utan många enkla regressioner eller regressioner med ett fåtal oberoende variabler. Denna typ av analyser används i stora populations- eller patientmaterial där man har analyserat genetiska variationer i en stor mängd baspar, s.k. enbaspolymorfier (single nucleotide polymorphism, SNP), spridda över hela genomet. Dessa analyser kan innehålla flera miljoner olika linjära regressionsmodeller där utfallet kan vara antingen en sjukdom eller en kontinuerlig variabel, och där varje SNP kodas som 0, 1 eller 2 beroende på om personen är homozygot (0 eller 2) eller heterozygot (1). Regressionskoefficienten påvisar då den additiva effekten av att ha den ena allelen jämfört med den andra allelen i detta baspar.

Resultatet av en sådan analys redovisas ofta som ett s.k. *Manhattan-diagram* (bild 16). Varje punkt representerar en SNP (en regressionsmodell); på y-axeln får vi då –10 log för p-värdet, medan x-axeln visar var i genomet denna SNP är lokaliserad. På grund av att extremt många test utförs sätts p-värdet för när man kan förkasta nollhypotesen för en enskild SNP extremt lågt, ofta $p < 5 * 10^{-8}$. En av prickarna i bild 16 nedan ligger just vid detta p-värde.

Bild 16. Manhattan-diagram.



Exempel 5.21. Logistisk regression

Exemplet ovan, som analyserades med chitvåtest, ser i en logistisk regressionsmodell ut så här:

```
. logistic blodtrycksmedicin fet

Logistic regression                                     Number of obs   =       1007
                                                        LR chi2(1)      =      32.69
                                                        Prob > chi2     =     0.0000
Log likelihood =  -610.8975                             Pseudo R2       =     0.0261

------------------------------------------------------------------------------
blodtrycks~n | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         fet |   2.460323   .3848303     5.76   0.000     1.810722     3.34297
------------------------------------------------------------------------------
```

Jämfört med chitvåtestet, som framför allt ger ett p-värd,e får vi nu även fram en oddskvot (odds ratio, OR) och dess konfidensintervall. Då talet 1 inte ingår i konfidensintervallet (1.810722 till 3.34297) kan nollhypotesen förkastas. Denna oddskvot säger oss att individer med högt BMI har nästan 2,5 gånger ökad risk för att behöva blodtrycksmedicin än individer med normalt BMI.

Vi försöker nu förklara utfallet blodtrycksmedicin med två nomonalvariabler (fetma, ”fet”, och kön ,”sex”,) och två kontinuerliga variabler (HDL-kolesterol och triglycerider). p-värdena visar att alla fyra variablerna, förutom kön, är oberoende av varandra relaterade till utfallet.

```
. logistic blodtrycksmedicin fet sex hdl triglycerider

Logistic regression                                     Number of obs   =       1003
                                                        LR chi2(4)      =      56.15
                                                        Prob > chi2     =     0.0000
Log likelihood = -596.87838                             Pseudo R2       =     0.0449

------------------------------------------------------------------------------
blodtrycks~n | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         fet |   2.109499   .3439964     4.58   0.000     1.532406    2.903922
         sex |   1.091649   .1672275     0.57   0.567      .808517     1.47393
         hdl |   .6029421   .1256408    -2.43   0.015     .4007768    .9070863
triglyceri~r |   1.429991    .174636     2.93   0.003     1.125593    1.816709
------------------------------------------------------------------------------
```

Exempel 5.22. Logistisk regression med undersökning av icke-linjära samband med kvartiler)

I exemplet ovan delade vi in triglycerider i kvintiler. I analysen nedan och på rad 2 jämförs den andra kvartilen mot den första, på rad 3 jämförs kvintilen 3 mot den första osv. Vi ser då att OR inte ökar/minskar på ett linjärt sätt från kvintil 2 till 4, dvs. i denna modell är variabeln triglycerider inte linjärt relaterad till blodtryckmedicinering.

```
. xtile q5triglycerider=triglycerider, nq(5)
. logistic blodtrycksmedicin fet sex hdl i.q5triglycerider

Logistic regression                                     Number of obs   =       1003
                                                        LR chi2(7)      =      64.40
                                                        Prob > chi2     =     0.0000
Log likelihood = -592.75298                             Pseudo R2       =     0.0515

------------------------------------------------------------------------------
blodtrycks~n | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         fet |    2.08759     .34126     4.50   0.000       1.5153    2.876019
         sex |   1.052932    .162835     0.33   0.739     .7776118    1.425732
         hdl |   .6784995    .145413    -1.81   0.070     .4457833    1.032702
             |
q5triglyce~r |
           2 |   1.219007   .3024881     0.80   0.425     .7495255    1.982558
           3 |   1.965718    .473678     2.80   0.005     1.225766    3.152353
           4 |    1.66105   .4107058     2.05   0.040     1.023099    2.696796
           5 |   2.418759    .607075     3.52   0.000     1.478948    3.955783
------------------------------------------------------------------------------
```

Exempel 5.23. Ordinal logistisk regression

I exemplet nedan har vi som utfall valt en blodtrycksvariabel som kodats på tre nivåer:
– 0 om man inte har högt blodtryck
– 1 om man har högt blodtryck men inte medicinerar
– 2 om man har blodtrycksmedicinering

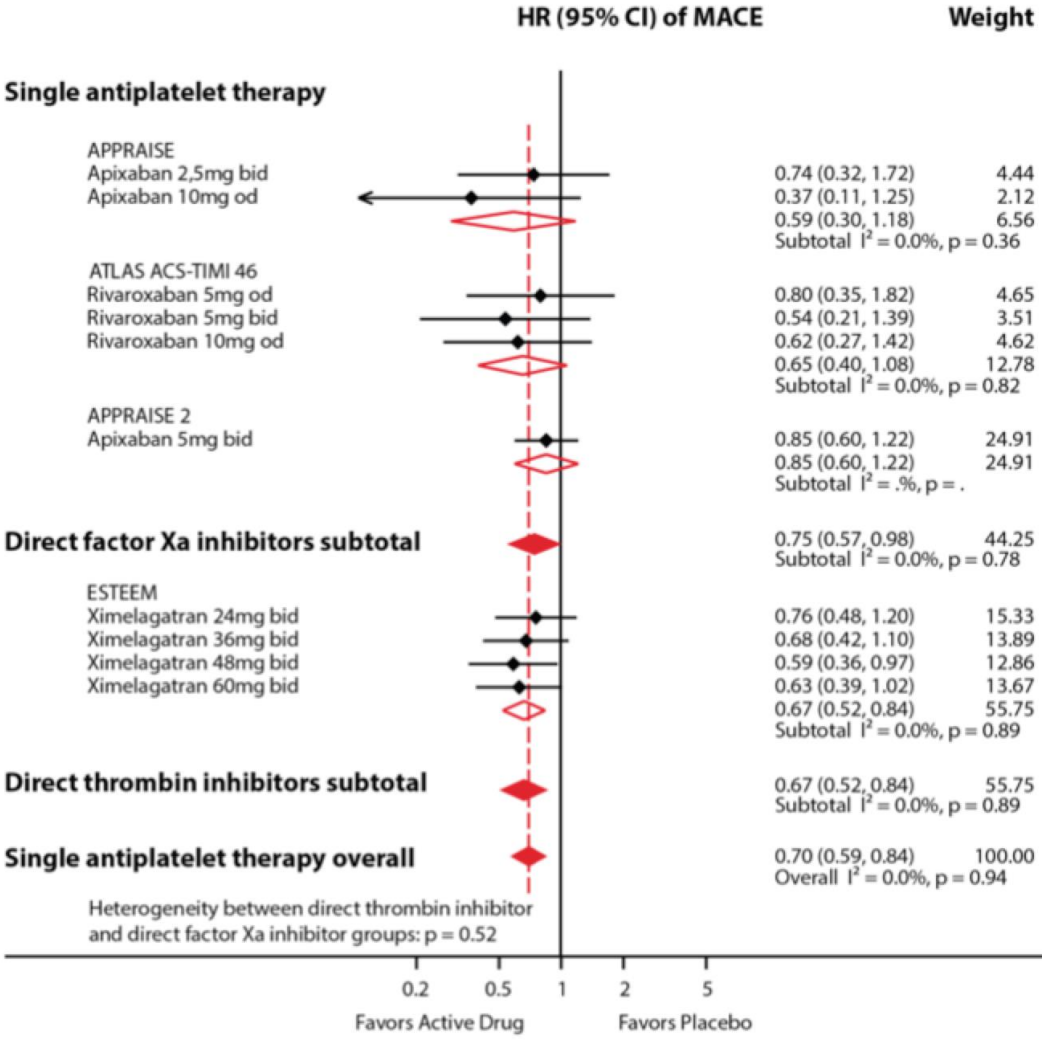
```
. ologit btvariabel fet sex hdl triglycerider, or

Ordered logistic regression                       Number of obs   =        999
                                                  LR chi2(4)      =      67.52
                                                  Prob > chi2     =     0.0000
Log likelihood = -1059.4289                       Pseudo R2       =     0.0309

------------------------------------------------------------------------------
  btvariabel | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         fet |   2.114692   .3159823     5.01   0.000     1.577824    2.834234
         sex |   1.234761   .1579212     1.65   0.099     .9609871     1.58653
         hdl |   .7461346   .1226832    -1.78   0.075     .5405779    1.029855
triglyceri~r |   1.520494    .169043     3.77   0.000     1.222788     1.89068
-------------+----------------------------------------------------------------
       /cut1 |  -.4986584   .3209837                     -1.127775    .1304581
       /cut2 |   1.176781   .3229756                      .5437603    1.809801
------------------------------------------------------------------------------
```

Vi ser då att OR för de fyra exponeringarna är tämligen lika jämför med exemplet då vi använde utfallet blodtrycksmedicin som en binär variabel. HDL är dock inte längre oberoende relaterat till denna nya mer komplexa variabel (p = 0.075).

Exempel 5.24. Metaanalys med blobbogram, metaregression och trattdiagram

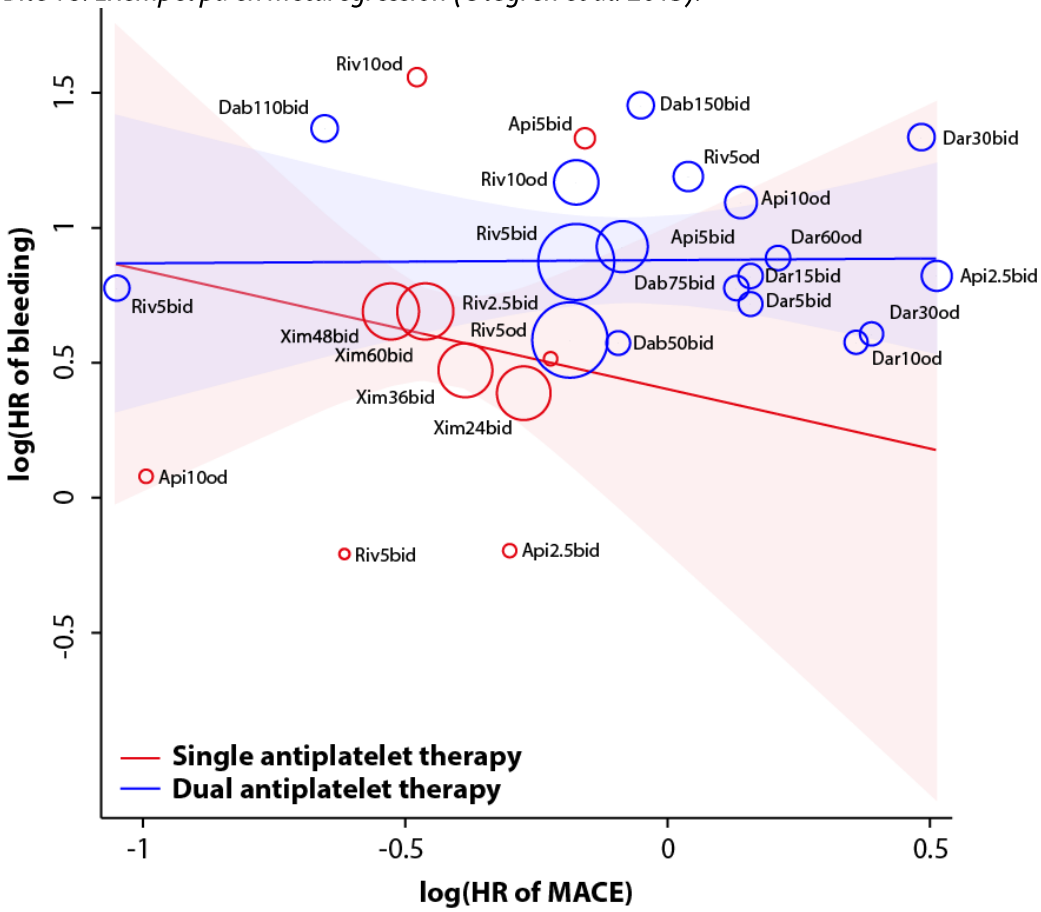
Bild 17 visar en metaanalys av effekterna av att ge ytterligare en blodförtunnande tablett (förutom acetylsalicylsyra) till personer som har haft en hjärtinfarkt. Det finns två familjer av nya blodförtunnare, faktor Xa-hämmare och trombinhämmare. Denna analys har gjorts med en modell med slumpmässiga effekter (random effects model), eftersom vi tänkte oss att den sanna effekten troligen varierar mellan olika studier. Eftersom I^2 är mycket lågt i varje analys är det lämpligt att presentera ett sammanvägt resultat i ett s.k. *blobbogram* (forest plot, bild 17) – den vanligaste grafiska framställningen av en metaanalys. Ibland representeras varje studies estimat av en kvadrat med en yta som motsvarar vikten i analysen, men i denna studie valde vi att i stället att ge vikten en kolumn vid sidan om. För varje studie finns även 95 % CI inritad. De sammanvägda resultaten i subgrupper respektive totalt har ritats som romber, vikas bredd är lika med 95 % CI. Observera att skalan på x-axeln är logaritmisk.

Bild 17. Exempel på ett blobbogram (Oldgren et al. 2013).



I bild 18 har vi använt metaregression för att studera kopplingen mellan två behandlingseffekter, en gynnsam och en skadlig. I denna studie sågs ingen tydlig koppling mellan den gynnsamma (skydd mot hjärthändelser) och den skadliga (blödningar) effekten.

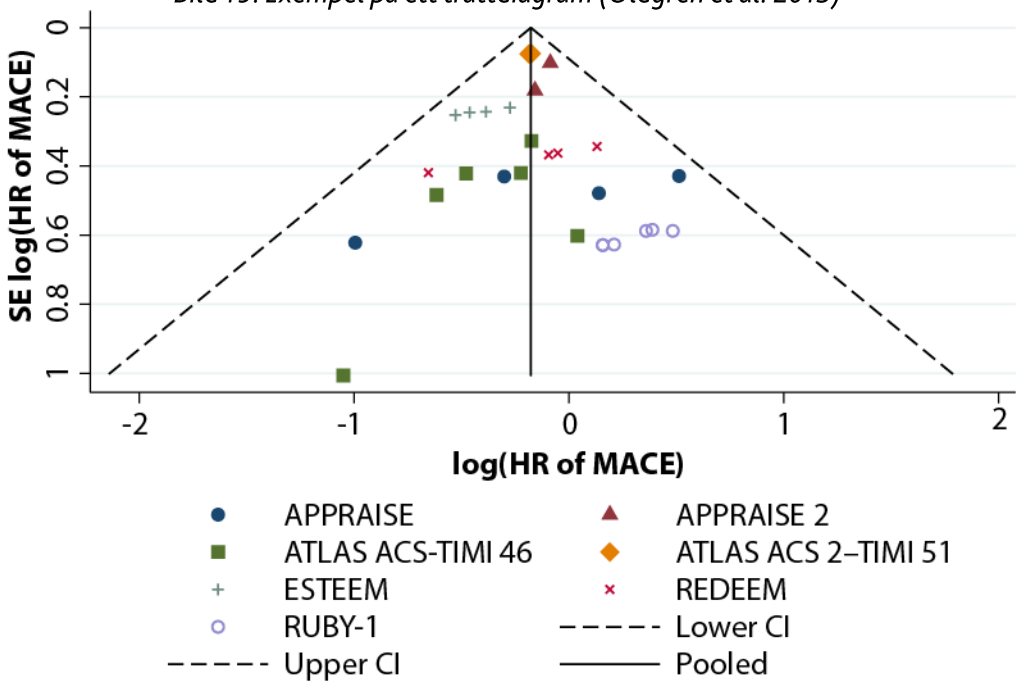
Bild 18. Exempel på en metaregression (Oldgren et al. 2013).



I bild 19 ses ett *trattdiagram* (funnel plot)med log hasardkvot på x-axeln som ett mått på studiens resultat, och standardfel log hasardkvot på y-axeln som ett mått på studiens precision (ju större studie, desto mindre standardfel på y-axeln). Denna graf ger inget stöd för någon betydande publikationsbias.

Små studier har en tendens att ge stor spridning i resultatet; det är därför som man ser desto större spridning ju högre standardfelet är (dvs. ju längre ner på y-axeln). Om det hade funnits publikationsbias hade det troligen saknats studier långt ner till höger (små studier som inte visar någon effekt av behandlingen). Det finns formella statistiska test för att kvantifiera detta fenomen, t.ex. Eggers test och Beggs test, men dessa behandlas inte här.

Bild 19. Exempel på ett trattdiagram (Oldgren et al. 2013)



Referenser

Oldgren, J., Wallentin, L., Alexander, J.H., James, S., Jönelid, B., Steg, G. & Sundström, J. et al. (2013). New oral anticoagulants in addition to single or dual antiplatelet therapy after an acute coronary syndrome: a systematic review and meta-analysis. *Eur Heart J*, 34–:1670–80.