Stata-solution: Exercises Day 1

PSY8003

Matthias Mittner

spring 2022

Exercise 2: Simple regression

Interpret the Intercept as the level where the predictor is zero, i.e., a person with no work experience earned about 198.8 thousand NOK. The slope is the increase with the predictor, i.e., for each year of work experience, someone will earn 0.88 thousand NOK more.

```
use "../data/loenn.dta"
regress loenn erfaring
regress loenn kvinne
* etc...
```

```
      Source |       SS           df       MS      Number of obs   =       464
-------------+----------------------------------   F(1, 462)       =      4.18
       Model |  50797.8678         1  50797.8678   Prob > F        =    0.0416
    Residual |  5620317.65       462  12165.1897   R-squared       =    0.0090
-------------+----------------------------------   Adj R-squared   =    0.0068
       Total |  5671115.52       463  12248.6296   Root MSE        =     110.3

------------------------------------------------------------------------------
       loenn | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    erfaring |   .8807123   .4309935     2.04   0.042     .0337617    1.727663
       _cons |   198.7606   9.268197    21.45   0.000     180.5475    216.9736
------------------------------------------------------------------------------

      Source |       SS           df       MS      Number of obs   =       471
-------------+----------------------------------   F(1, 469)       =     18.93
       Model |  223139.819         1  223139.819   Prob > F        =    0.0000
    Residual |  5527759.06       469  11786.2667   R-squared       =    0.0388
```

```
-------------+----------------------------------   Adj R-squared   =    0.0368
       Total |  5750898.88       470  12235.9551   Root MSE        =    108.56

------------------------------------------------------------------------------
       loenn | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      kvinne |  -43.83181    10.0737    -4.35   0.000    -63.62699   -24.03664
       _cons |    234.136   6.694378    34.98   0.000     220.9814    247.2907
------------------------------------------------------------------------------
```

Exercise 3: Multiple regression

When interpreting, it is important to include the "keeping constant" qualifier for all included variables. I.e., For each year in work experience, someone would earn 2400 NOK more **controlling for gender and education**.

```
use "../data/loenn.dta"
regress loenn erfaring kvinne utdann

regress loenn kvinne alder fagfor gift
margins, at(alder=(40) gift=(1) fagfor=(1) kvinne=(0))
```

```
      Source |       SS           df       MS      Number of obs   =       455
-------------+----------------------------------   F(3, 451)       =     44.75
       Model |  1281049.52         3  427016.508   Prob > F        =    0.0000
    Residual |  4303262.68       451   9541.6024   R-squared       =    0.2294
-------------+----------------------------------   Adj R-squared   =    0.2243
       Total |   5584312.2       454  12300.2471   Root MSE        =    97.681

------------------------------------------------------------------------------
       loenn | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    erfaring |   2.417205   .4131308     5.85   0.000     1.605304    3.229105
      kvinne |  -50.11816   9.241881    -5.42   0.000    -68.28065   -31.95566
      utdann |   19.08477   1.872227    10.19   0.000      15.4054    22.76414
       _cons |  -58.81691   28.71604    -2.05   0.041    -115.2508   -2.383058
------------------------------------------------------------------------------

      Source |       SS           df       MS      Number of obs   =       471
-------------+----------------------------------   F(4, 466)       =     10.35
       Model |  469247.606         4  117311.901   Prob > F        =    0.0000
```

```
    Residual |  5281651.27       466  11334.0156   R-squared       =    0.0816
-------------+----------------------------------   Adj R-squared   =    0.0737
       Total |  5750898.88       470  12235.9551   Root MSE        =    106.46

------------------------------------------------------------------------------
       loenn | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      kvinne |   -43.7214   10.04341    -4.35   0.000    -63.45738   -23.98541
       alder |    1.79559    .453446     3.96   0.000     .9045381    2.686642
      fagfor |   23.02044   12.62738     1.82   0.069    -1.793211    47.83409
        gift |  -2.455176   10.95962    -0.22   0.823    -23.99158    19.08123
       _cons |   164.5541   17.47449     9.42   0.000     130.2155    198.8926
------------------------------------------------------------------------------


Adjusted predictions                                      Number of obs = 471
Model VCE: OLS

Expression: Linear prediction, predict()
At: kvinne =  0
    alder  = 40
    fagfor =  1
    gift   =  1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   256.9429   11.75805    21.85   0.000     233.8376    280.0483
------------------------------------------------------------------------------
```

Exercise 4: Brain weight and total sleep across species

The two "weight" variables are highly correlated (collinear). Individually, they are both related to the total sleep. However, when inputting them simultaneously, the effect disappears.

```
use "../data/total_sleep.dta"
correlate
regress totalsleep brainwt
regress totalsleep bodywt
```

```
(species ignored because string variable)
(obs=42)

             |  brainwt   bodywt totals~p
-------------+---------------------------
     brainwt |   1.0000
      bodywt |   0.9558   1.0000
  totalsleep |  -0.3371  -0.3428   1.0000

      Source |       SS           df       MS      Number of obs   =        42
-------------+----------------------------------   F(1, 40)        =      5.13
       Model |  103.320849         1  103.320849   Prob > F        =    0.0290
    Residual |  806.001996        40  20.1500499   R-squared       =    0.1136
-------------+----------------------------------   Adj R-squared   =    0.0915
       Total |  909.322845        41   22.178606   Root MSE        =    4.4889

------------------------------------------------------------------------------
  totalsleep | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     brainwt |  -.0021661   .0009566    -2.26   0.029    -.0040995   -.0002328
       _cons |   11.11656   .7235503    15.36   0.000     9.654209    12.57891
------------------------------------------------------------------------------

      Source |       SS           df       MS      Number of obs   =        42
-------------+----------------------------------   F(1, 40)        =      5.33
       Model |  106.879467         1  106.879467   Prob > F        =    0.0262
    Residual |  802.443378        40  20.0610845   R-squared       =    0.1175
-------------+----------------------------------   Adj R-squared   =    0.0955
       Total |  909.322845        41   22.178606   Root MSE        =     4.479

------------------------------------------------------------------------------
  totalsleep | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      bodywt |  -.0040155   .0017397    -2.31   0.026    -.0075315   -.0004995
       _cons |   11.04768   .7130249    15.49   0.000     9.606599    12.48875
------------------------------------------------------------------------------
```

Exercise 5: Random data

When running enough of the regressions, some will be significant (approximately 5% when using the alpha-level of 0.05). The correlation matrix has some pretty large correlations. This is because of the low number of observations (N=20).

```
use "../data/random_data.dta"
regress x1 x43 x21
pwcorr x1 x2 x3 x4 x5 x6 x7 x8 x9 x10, sig star(0.05)
```

```
      Source |       SS           df       MS      Number of obs   =        20
-------------+----------------------------------   F(2, 17)        =      1.81
       Model |  5.25624358         2  2.62812179   Prob > F        =    0.1939
    Residual |  24.6906185        17  1.45238932   R-squared       =    0.1755
-------------+----------------------------------   Adj R-squared   =    0.0785
       Total |   29.946862        19  1.57615063   Root MSE        =    1.2052

------------------------------------------------------------------------------
          x1 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         x43 |   .4971569   .2718395     1.83   0.085    -.0763742    1.070688
         x21 |   .3040078   .2632187     1.15   0.264    -.2513351    .8593507
       _cons |  -.0543406   .2773098    -0.20   0.847    -.6394132    .5307319
------------------------------------------------------------------------------

             |       x1       x2       x3       x4       x5       x6       x7
-------------+---------------------------------------------------------------
          x1 |   1.0000
             |
             |
          x2 |   0.1170   1.0000
             |   0.6234
             |
          x3 |  -0.0762   0.2533   1.0000
             |   0.7494   0.2812
             |
          x4 |   0.0482  -0.2952   0.0128   1.0000
             |   0.8401   0.2063   0.9571
             |
          x5 |  -0.3391  -0.4572* -0.4202  -0.1140   1.0000
             |   0.1436   0.0427   0.0651   0.6322
             |
          x6 |  -0.0941  -0.0638  -0.0186   0.1202   0.2956   1.0000
             |   0.6932   0.7894   0.9378   0.6136   0.2057
             |
          x7 |  -0.1527   0.0902  -0.1023  -0.2309   0.1321   0.0611   1.0000
             |   0.5204   0.7053   0.6679   0.3273   0.5789   0.7979
```

```
             |
          x8 |   0.0304  -0.3157   0.1056   0.1504  -0.0900   0.1400  -0.5311*
             |   0.8987   0.1751   0.6578   0.5267   0.7058   0.5562   0.0160
             |
          x9 |   0.0191   0.1300   0.2776   0.2956   0.0020   0.1334  -0.0816
             |   0.9362   0.5848   0.2360   0.2058   0.9934   0.5750   0.7325
             |
         x10 |   0.0095   0.0285  -0.0426   0.2529   0.1370   0.0693   0.3339
             |   0.9684   0.9052   0.8585   0.2820   0.5646   0.7716   0.1502
             |

             |       x8       x9      x10
-------------+---------------------------
          x8 |   1.0000
             |
             |
          x9 |   0.1846   1.0000
             |   0.4360
             |
         x10 |  -0.1022   0.2040   1.0000
             |   0.6682   0.3882
             |
```