Sebastian E. Baumeister, Carsten-Oliver Schmidt, Till Ittermann, Henry Völzke

Version 4 - 14.02.2012

SHIP/KEF-Richtlinien zur Durchführung qualitätsgesicherter Auswertungen

1. Für Auswertungsprojekte ist im Regelfall ein bewilligter Datennutzungsantrag erforderlich.

2. Vor Beginn der Analysen ist ein Analyseplan zu erstellen. Dieser ist zumindest zwischen Erst-, Zweit- und Letztautor abzustimmen.

3. Variablendefinitionen und statistische Analysen sind vollständig in Programmen (z.B. Stata-Do-Files und Log-Files) zu dokumentieren. Diese Programme sind vor Ersteinreichung des Manuskripts an den Leiter des Publikationskomitees zu schicken. (siehe Appendix mit Beispielanalysen)

4. Die Log-Files, die alle Variablendefinitionen und statistischen Analysen beinhalten, sind durch eine zweite Person zu kontrollieren. Hierzu erfolgt insbesondere ein Abgleich mit den Tabellen des Manuskripts.

5. Die Analysen sind sachgerecht durchzuführen, dies umfasst:

    - die Auswertungen mit adäquaten statistischen Verfahren und

    - die Überprüfung der Annahmen der verwendeten Verfahren [1,2,3]. Bei erheblicher Verletzung der Annahmen sind die Analysen entsprechend anzupassen. Die Überprüfung der Annahmen ist ebenfalls im Do- und Log-File zu dokumentieren.

Appendix

Allgemeine Hinweise

In der Regel stehen die Ergebnisse von Regressionsmodellen in einem Auswertungsprojekt im Vordergrund. Bei der Schätzung von Regressionsmodellen ist sicherzustellen, dass die zugrundeliegenden Annahmen erfüllt sind [1,2,3]. Substantielle Verletzungen dieser Annahmen können zu Verzerrungen der Schätzer (Koeffizienten, P-Werte, Konfidenzintervalle, etc.) führen.

Wichtige Annahmen im Rahmen klassischer Regressionsmodellen umfassen:

a. Normalverteilung der Residuen
b. Linearität
c. Homoskedaszität
d. Einflussreiche Beobachtungen
e. Unabhängigkeit der Residuen
f. Missing completely at random (MCAR) Annahme bei fehlenden Werten und Ausfällen

Hinweise zur Kovariablen-Auswahl

Die Berücksichtigung und Auswahl von Kovariablen in einem Regressionsmodell kann unterschiedliche Ziele verfolgen. Grundsätzlich ist zwischen Konfounder- und Prädiktionsmodellen zu differenzieren [1,2,3]. Im Konfoundermodell steht der Zusammenhang zwischen Exposure und Outcome im Vordergrund. Die Auswahl zu adjustierender Konfounder erfolgt theoriegeleitet (z.B. mittels kausaler Grafen) oder anhand empirischer Kriterien (z.B. 15%-Change-in-Coefficient) [4]. Bei der Entwicklung von Prädiktionsmodellen [3] kommen die theoriegeleitete und empirisch-automatisierte Selektion (wie Stepwise-Verfahren) zur Anwendung [3].

Quellen:

1. Vittinghoff et al 2005. Regression Methods in Biostatistics.
2. Harrell 2001. Regression Modeling Strategies.
3. Steyerberg 2010. Clinical Prediction Models.
4. Rothman, Grennland, Lash 2008. Modern Epidemiology.

Beispielsyntax und Log-File für die Durchführung und Überprüfung einer Regressionsanalyse

Das folgende Beispiel dient zur Illustration von Do- und Log-Dateien. Analysiert werden Korrelate des Geburtsgewichts mittels linearer Regression.

1. Log-File zur Variablendefinition

```
**************************************************************************************
. ** Sebastian Baumeister
. ** Titel of project: Maternal correlates of birth weight
. ** Pfad:  E:\Arbeit\Greifswald\ship_statistik\Richtlinie zur Durchführung von Auswertungen\
**************************************************************************************

glo bw "E:\Arbeit\Greifswald\ship_statistik\Richtlinie zur Durchführung von
Auswertungen\data\"
```

```
. *Load original dataset and save under new name
. ********************
. use http://www.stata-press.com/data/r10/lbw3.dta, clear
(Hosmer & Lemeshow data)

save "$bw\bw_smoking.dta", replace
file E:\Arbeit\Greifswald\ship_statistik\Richtlinie zur Durchführung von
Auswertungen\data\\bw_smoking.dta saved

codebook, c
Variable   Obs Unique      Mean  Min   Max  Label
-------------------------------------------------------------------------------------------------
-------------------------------------------------
id         189    189  121.0794    4   226  identification code
low        189      2  .3121693    0     1  birth weight<2500g
age        189     24   23.2381   14    45  age of mother
lwt        189     76  129.8201   80   250  weight at last menstrual period
race       189      3  1.846561    1     3  race
smoke      189      2  .3915344    0     1  smoked during pregnancy
ptl        189      4  .1957672    0     3  premature labor history (count)
ht         189      2  .0634921    0     1  has history of hypertension
ui         189      2  .1481481    0     1  presence, uterine irritability
ftv        189      6  .7936508    0     6  number of visits to physician during 1st trimester
bwt        189    133  2944.286  709  4990  birth weight (grams)
white      189      2  .5079365    0     1  race==white
black      189      2  .1375661    0     1  race==black
other      189      2  .3544974    0     1  race==other
ptd        189      2  .1587302    0     1
lwd        189      2  .2222222    0     1
* Define new outcome variables
. ************
Low birth weight
recode bwt (min/2500=1) (2500.001/max=0), gen(lowbwt)
(189 differences between bwt and lowbwt)
label variable lowbwt "birth weight"
label define lowbwt 1 "1,birth weight<2500g" 0 "0,birth weight>=2500"
label value lowbwt lowbwt

. *Covariables
. ************
. * Age groups
. recode age (min/19=1) (20/23=2) (24/26=3) (27/max=4) (.=.), gen(age4)
(189 differences between age and age4)

. save, replace
file E:\Arbeit\Greifswald\ship_statistik\Richtlinie zur Durchführung von
Auswertungen\data\\bw_smoking.dta saved

. log close
```

Der 2. Log-File unten veranschaulicht das Vorgehen bei der Durchführung und Überprüfung der Annahmen der linearen („kleinste Quadrate“) Regression.

2. Log-File zur Auswertung

```
.
**************************************************************************************************
**************************************************************************************************
. ** Sebastian Baumeister
. ** Titel of project: Maternal correlates of birth weight
. ** Pfad:  E:\Arbeit\Greifswald\ship_statistik\Richtlinie zur Durchführung von Auswertungen\
**************************************************************************************************
**************************************************************************************************
glo bw "E:\Arbeit\Greifswald\ship_statistik\Richtlinie zur Durchführung von
Auswertungen\data\"

. use "$bw\bw_smoking.dta", clear
(Hosmer & Lemeshow data)

des, sh

Contains data from E:\Arbeit\Greifswald\ship_statistik\Richtlinie zur Durchführung von
Auswertungen\data\\bw_smoking.dta
  obs:           189                          Hosmer & Lemeshow data
```

```
 vars:            18                          10 Jan 2012 14:15
 size:         5,292
Sorted by:

. * Analytical sample
. ***********************
. glo out bwt

. glo cov  age lwt race smoke ptl ht ui ftv

. tabmiss $out $cov
    Variable |     Obs       Missings   Feq.Missings    NonMiss   Feq.NonMiss
-------------+-----------------------------------------------------------------
         bwt |     189           0            0            189          100
         age |     189           0            0            189          100
         lwt |     189           0            0            189          100
        race |     189           0            0            189          100
       smoke |     189           0            0            189          100
         ptl |     189           0            0            189          100
          ht |     189           0            0            189          100
          ui |     189           0            0            189          100
         ftv |     189           0            0            189          100

. egen nm=rowmiss($out $cov)

. fre nm

nm
-------------------------------------------------------
          |      Freq.    Percent      Valid       Cum.
----------+--------------------------------------------
Valid   0 |        189     100.00     100.00     100.00
-------------------------------------------------------

. recode nm (0=0) (1/max=1),gen(miss)
(0 differences between nm and miss)

. * No missing values on any variable included in the (regression analyses)
.
. * Table 1: Characteritics of the mothers
. ************************
. tabstat  age lwt  ptl  ftv if miss==0, s(n median p25 p75 mean sd) c(s)

    variable |         N       p50       p25       p75      mean        sd
-------------+------------------------------------------------------------
         age |       189        23        19        26   23.2381  5.298678
         lwt |       189       121       110       140  129.8201  30.57515
         ptl |       189         0         0         0  .1957672  .4933419
         ftv |       189         0         0         1  .7936508  1.059286
----------------------------------------------------------------------------

. fre race smoke ht ui if miss==0

race -- race
-------------------------------------------------------------
                |      Freq.    Percent      Valid       Cum.
----------------+--------------------------------------------
Valid   1 white |         96      50.79      50.79      50.79
        2 black |         26      13.76      13.76      64.55
        3 other |         67      35.45      35.45     100.00
        Total   |        189     100.00     100.00
-------------------------------------------------------------

smoke -- smoked during pregnancy
-------------------------------------------------------------
              |      Freq.    Percent      Valid       Cum.
--------------+--------------------------------------------
Valid   0     |        115      60.85      60.85      60.85
        1     |         74      39.15      39.15     100.00
        Total |        189     100.00     100.00
-------------------------------------------------------------

ht -- has history of hypertension
-------------------------------------------------------------
              |      Freq.    Percent      Valid       Cum.
--------------+----------------------------------------------
```

```
Valid   0     |        177      93.65      93.65      93.65
        1     |         12       6.35       6.35     100.00
        Total |        189     100.00     100.00
-------------------------------------------------------------

ui -- presence, uterine irritability
-------------------------------------------------------------
              |      Freq.    Percent      Valid       Cum.
--------------+----------------------------------------------
Valid   0     |        161      85.19      85.19      85.19
        1     |         28      14.81      14.81     100.00
        Total |        189     100.00     100.00
-------------------------------------------------------------

.
. * Table 2: Correlates of birth weight
. ************************
. reg bwt age i.race i.smoke i.ht lwt  ptl  ftv if miss==0

      Source |       SS       df       MS              Number of obs =     189
-------------+------------------------------           F(  8,   180) =    5.09
       Model |  18427359.2     8   2303419.9           Prob > F      =  0.0000
    Residual |  81487939.3   180  452710.774           R-squared     =  0.1844
-------------+------------------------------           Adj R-squared =  0.1482
       Total |  99915298.6   188  531464.354           Root MSE      =  672.84

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -1.400947   9.932138    -0.14   0.888    -20.99935    18.19745
             |
        race |
          2  |  -488.0906   155.1772    -3.15   0.002     -794.291   -181.8901
          3  |   -372.048   118.6461    -3.14   0.002    -606.1641    -137.932
             |
     1.smoke |  -362.5319    110.121    -3.29   0.001     -579.826   -145.2379
        1.ht |  -524.4655   208.4684    -2.52   0.013    -935.8216   -113.1093
         lwt |   4.801987   1.791585     2.68   0.008     1.266777    8.337197
         ptl |  -127.4346    103.163    -1.24   0.218     -330.999    76.12981
         ftv |  -9.086507   48.06139    -0.19   0.850    -103.9227     85.7497
       _cons |   2759.884   320.2368     8.62   0.000     2127.982    3391.785
------------------------------------------------------------------------------

.
. * Test of linear regression (OLS) assumptions
. *************************
. * Normality of residuals: looks fine
. *****
. predict r, res

. kdensity r, norm

. qnorm r

. pnorm r

. * Homoscedasticity: looks fine
. *****
. estat hettest age i.race i.smoke i.ht lwt  ptl  ftv, iid

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: age i.race i.smoke i.ht lwt ptl ftv

         chi2(8)      =     7.45
         Prob > chi2  =   0.4887

. rvfplot, yline(0)

. * Linearity
. *****
. * Age - looks nonlinear, modeled using restricted cubic splines
. twoway (scatter bwt age) (lfit bwt age)  (lowess bwt age) (fpfit bwt age)

. acprplot age, lowess lsopts(bwidth(1))
```

```
. centile age if miss==0, c(5 35 65 95)

                                                      -- Binom. Interp. --
    Variable |     Obs  Percentile      Centile        [95% Conf. Interval]
-------------+-------------------------------------------------------------
         age |     189          5            16              15          17
             |                 35            20              20          21
             |                 65            25              24          26
             |                 95          32.5              31          35

. mkspline2 ag_=age, cubic knots(16 20 25 32.5) dis

             |     knot1      knot2      knot3      knot4
-------------+--------------------------------------------
         age |        16         20         25       32.5

. * Weight at last menstrual period - looks linear
. twoway (scatter bwt lwt) (lfit bwt lwt)  (lowess bwt lwt) /*(fpfit bwt lwt)*/

. * premature labor history
. twoway (scatter bwt ptl) (lfit bwt ptl)  (lowess bwt ptl) /*(fpfit bwt lwt)*/

. fre ptl

ptl -- premature labor history (count)
------------------------------------------------------------
              |      Freq.    Percent      Valid       Cum.
--------------+---------------------------------------------
Valid   0     |        159      84.13      84.13      84.13
        1     |         24      12.70      12.70      96.83
        2     |          5       2.65       2.65      99.47
        3     |          1       0.53       0.53     100.00
        Total |        189     100.00     100.00
------------------------------------------------------------

. *only few subjects with premature labor history >1 therefore recode with new variables
values 0 and 1+
. recode ptl (0=0 "0,zero") (1/3=1 "1,1+"), gen(plt_2c)
(6 differences between ptl and plt_2c)

. * Collinearity - not an issue
. *****
. vif

    Variable |       VIF       1/VIF
-------------+----------------------
         age |      1.15    0.869446
        race |
          2  |      1.19    0.838428
          3  |      1.34    0.743603
     1.smoke |      1.21    0.829112
        1.ht |      1.08    0.926932
         lwt |      1.25    0.802511
         ptl |      1.08    0.929649
         ftv |      1.08    0.929060
-------------+----------------------
    Mean VIF |      1.17

. collin bwt age race smoke ht lwt  plt_2c  ftv if miss==0
(obs=189)

  Collinearity Diagnostics

                       SQRT                  R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
       bwt      1.22    1.10    0.8211      0.1789
       age      1.13    1.06    0.8848      0.1152
      race      1.30    1.14    0.7684      0.2316
     smoke      1.28    1.13    0.7816      0.2184
        ht      1.12    1.06    0.8955      0.1045
       lwt      1.20    1.09    0.8364      0.1636
    plt_2c      1.11    1.05    0.9010      0.0990
       ftv      1.07    1.04    0.9304      0.0696
----------------------------------------------------
  Mean VIF      1.18
```

```
                        Cond
       Eigenval         Index
--------------------------------
    1     5.7714          1.0000
    2     0.9465          2.4693
    3     0.8763          2.5664
    4     0.5870          3.1356
    5     0.5637          3.1997
    6     0.1628          5.9545
    7     0.0438         11.4792
    8     0.0373         12.4466
    9     0.0112         22.6979
--------------------------------
 Condition Number        22.6979
 Eigenvalues & Cond Index computed from scaled raw sscp (w/ intercept)
 Det(correlation matrix)    0.5271

.
. * Omitted variables / exogeneity -> not suspected
. *****
. linktest

      Source |       SS       df       MS              Number of obs =     189
-------------+------------------------------           F(  2,   186) =   21.23
       Model |  18566610.8     2  9283305.38           Prob > F      =  0.0000
    Residual |  81348687.8   186  437358.537           R-squared     =  0.1858
-------------+------------------------------           Adj R-squared =  0.1771
       Total |  99915298.6   188  531464.354           Root MSE      =  661.33

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   2.244969   2.211734     1.02   0.311    -2.118339    6.608278
      _hatsq |  -.0002112   .0003743    -0.56   0.573    -.0009497    .0005272
       _cons |  -1813.978   3246.971    -0.56   0.577    -8219.603    4591.646
------------------------------------------------------------------------------

. estat ovtest

Ramsey RESET test using powers of the fitted values of bwt
       Ho:  model has no omitted variables
                 F(3, 177) =      1.49
                  Prob > F =      0.2184

.
. * Influential observations
. *****
. *Cookd distance > 4/N
. reg bwt ag_* i.race i.smoke i.ht lwt  ptl  ftv if miss==0

      Source |       SS       df       MS              Number of obs =     189
-------------+------------------------------           F( 10,   178) =    4.64
       Model |  20672624.4    10  2067262.44           Prob > F      =  0.0000
    Residual |  79242674.1   178  445183.563           R-squared     =  0.2069
-------------+------------------------------           Adj R-squared =  0.1623
       Total |  99915298.6   188  531464.354           Root MSE      =  667.22

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ag_1 |   16.37202   49.33889     0.33   0.740     -80.9924    113.7364
        ag_2 |  -232.9313   218.3098    -1.07   0.287    -663.7397     197.877
        ag_3 |   612.1989   476.4075     1.29   0.200    -327.9347    1552.332
             |
        race |
          2  |  -495.2098   154.1368    -3.21   0.002    -799.3804   -191.0392
          3  |  -353.5734   117.9436    -3.00   0.003     -586.321   -120.8258
             |
     1.smoke |  -369.0548   109.3287    -3.38   0.001     -584.802   -153.3076
        1.ht |  -511.5395   207.5715    -2.46   0.015    -921.1572   -101.9219
         lwt |   4.648354     1.7802     2.61   0.010     1.135341    8.161367
         ptl |  -104.6919   102.8709    -1.02   0.310    -307.6954    98.31169
         ftv |  -10.29435   47.67845    -0.22   0.829    -104.3821    83.79341
       _cons |   2535.593   906.5436     2.80   0.006     746.6368    4324.548
------------------------------------------------------------------------------
```

```
. predict d, cooksd

. li bwt age race smoke ht lwt  plt_2c  ftv if d>4/189

     +-------------------------------------------------------+
     |  bwt   age    race   smoke   ht   lwt   plt_2c   ftv |
     |-------------------------------------------------------|
 23. | 2836    36   white       0    0   202   0,zero     1 |
 65. | 3260    26   other       1    0   133     1,1+     0 |
 94. | 3637    25   white       1    0    95     1,1+     0 |
102. | 3756    19   white       1    1   184   0,zero     0 |
106. | 3790    25   black       0    1   241   0,zero     0 |
     |-------------------------------------------------------|
110. | 3860    16   black       0    0   170   0,zero     4 |
130. | 4990    45   white       0    0   123   0,zero     1 |
131. |  709    28   other       1    0   120     1,1+     0 |
132. | 1021    29   white       0    0   130   0,zero     2 |
133. | 1135    34   black       1    1   187   0,zero     0 |
     |-------------------------------------------------------|
134. | 1330    25   other       0    1   105     1,1+     0 |
147. | 1928    21   black       0    0   200   0,zero     2 |
     +-------------------------------------------------------+

. *dfits > 2*sqrt(k/N)
. predict dfits, dfits

. quiet scalar thresh=2*sqrt((e(df_m)+1)/e(N))

. di "dfits threshold=" %6.3f thresh
dfits threshold= 0.482

. li dfits bwt age race smoke ht lwt  plt_2c  ftv if abs(dfits)>2*thresh & e(sample)

     +-----------------------------------------------------------------------+
     |     dfits    bwt   age    race   smoke   ht   lwt   plt_2c   ftv |
     |-----------------------------------------------------------------------|
 94. |  1.032447   3637    25   white       1    0    95     1,1+     0 |
130. |  1.464016   4990    45   white       0    0   123   0,zero     1 |
133. | -.9895521   1135    34   black       1    1   187   0,zero     0 |
     +-----------------------------------------------------------------------+

. reg bwt ag_* i.race i.smoke i.ht lwt  ptl  ftv if miss==0 & d<4/189

      Source |       SS       df       MS              Number of obs =     177
-------------+------------------------------           F( 10,   166) =    4.90
       Model |  17613347.8    10  1761334.78           Prob > F      =  0.0000
    Residual |  59651728.5   166  359347.762           R-squared     =  0.2280
-------------+------------------------------           Adj R-squared =  0.1815
       Total |  77265076.3   176  439006.115           Root MSE      =  599.46

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ag_1 |   22.39731   46.39013     0.48   0.630    -69.19341     113.988
        ag_2 |  -215.2098   216.1807    -1.00   0.321    -642.0279    211.6083
        ag_3 |   569.4761   486.5547     1.17   0.244     -391.157    1530.109
             |
        race |
          2  |  -441.4639   146.7486    -3.01   0.003    -731.1982   -151.7296
          3  |  -320.2998   109.1899    -2.93   0.004    -535.8797     -104.72
             |
     1.smoke |  -366.6639   101.9441    -3.60   0.000     -567.938   -165.3898
        1.ht |  -512.8817   221.7429    -2.31   0.022    -950.6815   -75.08184
         lwt |   4.619319   1.707276     2.71   0.008     1.248546    7.990092
         ptl |  -225.7582   106.8228    -2.11   0.036    -436.6646   -14.85177
         ftv |  -7.792252    45.2459    -0.17   0.863    -97.12384    81.53934
       _cons |   2396.499   843.6123     2.84   0.005     730.9063    4062.091
------------------------------------------------------------------------------

. reg bwt ag_* i.race i.smoke i.ht lwt  ptl  ftv if miss==0 & abs(dfits)<2*thresh

      Source |       SS       df       MS              Number of obs =     186
-------------+------------------------------           F( 10,   175) =    4.14
       Model |  17586801.3    10  1758680.13           Prob > F      =  0.0000
    Residual |    74385541   175  425060.234           R-squared     =  0.1912
```

```
-------------+------------------------------           Adj R-squared =  0.1450
       Total |  91972342.3   185  497147.796           Root MSE      =  651.97

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ag_1 |   5.318512   49.27923     0.11   0.914    -91.93959    102.5766
        ag_2 |  -149.2398   226.7889    -0.66   0.511    -596.8333    298.3537
        ag_3 |   405.2366   507.4508     0.80   0.426    -596.2747    1406.748
             |
        race |
          2  |  -424.4026   153.8329    -2.76   0.006    -728.0091   -120.7961
          3  |   -310.511   116.1151    -2.67   0.008    -539.6773   -81.34473
             |
     1.smoke |  -324.2243    108.012    -3.00   0.003    -537.3981   -111.0505
        1.ht |  -403.0304   210.9367    -1.91   0.058    -819.3377    13.27692
         lwt |   5.147605   1.756653     2.93   0.004     1.680652    8.614557
         ptl |  -208.9705   109.7989    -1.90   0.059    -425.6709    7.729915
         ftv |   -9.28851   46.96304    -0.20   0.843    -101.9754    83.39833
       _cons |   2615.481   896.9764     2.92   0.004     845.1968    4385.764
------------------------------------------------------------------------------

. *Some influential observation that distort estimates
. *Robust and quantile regression
. rreg bwt ag_* i.race i.smoke i.ht lwt  ptl  ftv

Robust regression                                      Number of obs =     189
                                                       F( 10,   178) =    4.44
                                                       Prob > F      =  0.0000

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ag_1 |   14.28029   51.93272     0.27   0.784    -88.20275    116.7633
        ag_2 |   -206.659   229.7867    -0.90   0.370    -660.1157    246.7977
        ag_3 |   555.4032   501.4531     1.11   0.270    -434.1548    1544.961
             |
        race |
          2  |  -485.2457     162.24    -2.99   0.003    -805.4071   -165.0843
          3  |  -350.0935   124.1441    -2.82   0.005    -595.0771   -105.1099
             |
     1.smoke |  -379.7665   115.0763    -3.30   0.001    -606.8559   -152.6771
        1.ht |  -503.1772   218.4839    -2.30   0.022    -934.3291    -72.0252
         lwt |   4.742754   1.873788     2.53   0.012     1.045056    8.440453
         ptl |   -133.221   108.2791    -1.23   0.220    -346.8968    80.45477
         ftv |   -8.67221     50.185    -0.17   0.863    -107.7063     90.3619
       _cons |   2564.002   954.2022     2.69   0.008     680.9976    4447.006
------------------------------------------------------------------------------

. xi: qreg bwt ag_* i.race i.smoke i.ht lwt  ptl  ftv
i.race            _Irace_1-3          (naturally coded; _Irace_1 omitted)
i.smoke           _Ismoke_0-1         (naturally coded; _Ismoke_0 omitted)
i.ht              _Iht_0-1            (naturally coded; _Iht_0 omitted)
Iteration  1:  WLS sum of weighted deviations =  98922.856

Median regression                                      Number of obs =       189
  Raw sum of deviations   111569 (about 2977)
  Min sum of deviations 98122.23                       Pseudo R2     =    0.1205

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ag_1 |   45.50179   92.97873     0.49   0.625    -137.9807    228.9842
        ag_2 |  -272.5861   411.4234    -0.66   0.508    -1084.481     539.309
        ag_3 |   663.8722   898.2262     0.74   0.461     -1108.67    2436.415
    _Irace_2 |  -412.5939   279.1732    -1.48   0.141    -963.5091    138.3212
    _Irace_3 |  -411.2986   219.5242    -1.87   0.063    -844.5035    21.90631
   _Ismoke_1 |  -492.6311   203.2619    -2.42   0.016    -893.7442   -91.51801
      _Iht_1 |  -459.1765   361.2127    -1.27   0.205    -1171.987    253.6339
         lwt |   5.314155   3.313364     1.60   0.111    -1.224375    11.85268
         ptl |  -172.1293   191.6626    -0.90   0.370    -550.3526    206.0939
         ftv |  -18.20978   88.36572    -0.21   0.837     -192.589    156.1694
       _cons |   1977.775   1705.814     1.16   0.248    -1388.446    5343.996
------------------------------------------------------------------------------

.
```

```
.
.
.
. *Final robust regression model
. xi: rreg bwt ag_* i.race i.smoke i.ht lwt  i.plt_2c  ftv if miss==0
i.race            _Irace_1-3          (naturally coded; _Irace_1 omitted)
i.smoke           _Ismoke_0-1         (naturally coded; _Ismoke_0 omitted)
i.ht              _Iht_0-1            (naturally coded; _Iht_0 omitted)
i.plt_2c          _Iplt_2c_0-1        (naturally coded; _Iplt_2c_0 omitted)

Robust regression                                     Number of obs =     189
                                                      F( 10,   178) =    4.79
                                                      Prob > F      =  0.0000

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ag_1 |   14.10818   51.27275     0.28   0.784    -87.07248    115.2888
        ag_2 |  -199.0313   227.0517    -0.88   0.382    -647.0908    249.0282
        ag_3 |   539.2802   495.4929     1.09   0.278     -438.516    1517.076
    _Irace_2 |  -469.5862   160.6347    -2.92   0.004    -786.5796   -152.5928
    _Irace_3 |  -335.1016   123.1111    -2.72   0.007    -578.0467   -92.15643
   _Ismoke_1 |  -358.5469   114.0825    -3.14   0.002    -583.6751   -133.4187
      _Iht_1 |  -487.7192    215.899    -2.26   0.025    -913.7703   -61.66819
         lwt |   4.594307   1.848273     2.49   0.014     .9469598    8.241655
  _Iplt_2c_1 |  -285.7212   144.2481    -1.98   0.049    -570.3777    -1.06463
         ftv |  -7.178501   49.54356    -0.14   0.885    -104.9468    90.58981
       _cons |    2579.96   942.5491     2.74   0.007     719.9511    4439.968
------------------------------------------------------------------------------

. testparm ag_*

 ( 1)  ag_1 = 0
 ( 2)  ag_2 = 0
 ( 3)  ag_3 = 0

       F(  3,   178) =    1.52
            Prob > F =    0.2110

. log close
```