Working with Stata

Advanced modelling topics

April 13, 2017

Nicola Orsini

Biostatistics Team

Department of Public Health Sciences

Karolinska Institutet

Outline

- Testing hypothesis and compare multivariable models
- Interaction analysis
- Flexible modelling of quantitative predictors
- Extension of logistic

Likelihood Ratio Test

The likelihood ratio test compares the log-likelihoods (or deviances) of two nested models fitted on the same data

$$LRT = 2 * \text{abs}(\log L_R - \log L_U)$$

where $\log L_R$ is the log-likelihood of the restricted model and $\log L_R$ is the log-likelihood of the unrestricted model.

The two models are nested because are estimated on the same number of subjects but different number of parameters in the model.

It means that you can obtain the restricted model from the full model by putting constraints on the parameters you want to test.

Let's perform a likelihood ratio test for the association between race duration (modeled using indicator variables 1 "<=3:30"; 2 "3:30-4:30"; and 3 ">3:30") and the risk of hyponatremia. The lowest category is used as referent or comparison group.

Unrestricted model

$$\log(\text{odds}|\text{runtimehc}) = \beta_0 + \beta_1 \text{I}(\text{runtimehc} = 2) + \beta_2 \text{I}(\text{runtimehc} = 3)$$

Restricted model

$$\log(\text{odds}|\text{runtimehc}) = \beta_0$$

The null hypothesis is

$$H_0: \beta_1 = \beta_2 = 0$$

. xi: logit nas135 i.runtimehc

```
Logistic regression                               Number of obs     =        477
                                                  LR chi2(2)        =      26.55
                                                  Prob > chi2       =     0.0000
Log likelihood = -167.17359                       Pseudo R2         =     0.0736

------------------------------------------------------------------------------
       nas135 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
_Iruntimehc_2 |    1.309275   .4311212     3.04   0.002     .4642931    2.154257
_Iruntimehc_3 |    1.895113   .4149536     4.57   0.000     1.081819    2.708407
        _cons |   -3.124565   .3612402    -8.65   0.000    -3.832583   -2.416547
------------------------------------------------------------------------------
```

. xi: logit nas135 if runtimehc != .

```
Logistic regression                               Number of obs     =        477
                                                  LR chi2(0)        =       0.00
                                                  Prob > chi2       =          .
Log likelihood = -180.44774                       Pseudo R2         =     0.0000

------------------------------------------------------------------------------
      nas135 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   -1.938742   .1380752   -14.04   0.000    -2.209364   -1.668119
------------------------------------------------------------------------------
```

The likelihood ratio test is given by

$$LRT = 2 * \text{abs}(-180.44774 - -167.17359\ ) = 26.55$$

The greater is the change in the log likelihood because of the inclusion of the predictors (race duration categorized) and the greater is the evidence against the null hypothesis (no association between race duration and hyponatremia risk).

Under the null hypothesis (assuming the restricted model is true), the likelihood ratio test follows a χ^2 (chi-square) distribution with degrees of freedom equal to the difference in the number of parameters (2 in our example) of the two models being compared.

The likelihood ratio test and the p-value can be calculated using a hand calculator with probability functions.

```
. di 2*abs(-180.44774--167.17359 )
26.5483

. di chi2tail(2, 2*abs(-180.44774--167.17359 ))
1.718e-06
```

The *P*-value associated with the test is $\Pr[\chi^2(2) > 26.5] < 0.05$.

We reject the null hypothesis that β_1 and β_2 are equal to zero.

We found a significant association between race duration (categorized in 3 intervals and modeled using indicators) and the risk of hyponatremia ($p<0.001$).

In Stata one can easily store and subsequently compare likelihoods of different nested models using a command to perform a likelihood ratio test.

```
logit nas135 i.runtimehc
est store u

logit nas135 if runtimehc != .
est store r
```

```
. lrtest r u

Likelihood-ratio test                                 LR chi2(2)  =     26.55
(Assumption: r nested in u)                           Prob > chi2 =    0.0000
```

Wald-type test

A Wald-type test is equal to the ratio of the maximum likelihood estimate of the parameter to an estimate of its standard error.

$$W = \frac{\widehat{\beta_1}}{SE(\widehat{\beta_1})}$$

Under the null hypothesis $H_0: \beta_1 = 0$ and assuming a large sample, this ratio follows a standard normal distribution z.

A two-tailed P-value is obtained by calculating $P(|z| > W)$.

The squared Wald-test ($W^2 = z^2$) follows a Chi-Square distribution with 1 degree of freedom.

A two-tailed *P*-value is obtained by calculating $P(\chi_1^2 > W^2)$.

Example. The estimated log odds ratio of hyponatremia comparing a race duration of 3:30-4:00 vs <3:30 was $\widehat{\beta_1} = 1.309275$ with $SE(\widehat{\beta_1}) = .4311212$.

$$W = \frac{1.309275}{.4311212} = 3.04$$

$$P(|z| > 3.04) = 0.002$$

```
. display normal(-3.04)*2
```

$$W^2 = \left(\frac{1.309275}{.4311212}\right)^2 = 9.22$$

Then a *P*-value is obtained as

$$P(\chi_1^2 > 9.22)=0.002$$

```
. di chi2tail(1, (1.309275/.4311212)^2)
```

At 5% level, we reject the null hypothesis that the population log odds ratio of hyponatremia comparing a race duration of 3:30-4:00 vs <3:30 is equal to zero.

In our example of race duration modelled with 2 indicator variables, one might be interested in a single *P*-value testing overall whether race duration is significantly associated with hyponatremia risk.

$$H_0: \beta_1 = \beta_2 = 0$$

We use the multivariate analogue of the Wald-test which, under the null, follows a χ^2 distribution with 2 degrees of freedom.
So a two-sided *P*-value is obtained $P(\chi_2^2 > W^2)$.

```
. testparm i.runtimehc
 ( 1)  [nas135]2.runtimehc = 0
 ( 2)  [nas135]3.runtimehc = 0

           chi2(  2) =   21.04
         Prob > chi2 =    0.0000
```

Information criteria

The AIC is a popular measure for comparing maximum likelihood models.

AIC is defined as

$$AIC = -2*\log(\text{likelihood}) + 2k$$

where

k = number of parameters estimated
N = number of observations

AIC is a measure that combine fit and complexity.

Fit is measured negatively by -2*log(likelihood); the larger the value, the worse the fit.

Complexity is measured positively, by 2 times k (AIC).

Given two models fit on the same data, the model with the smaller value of the information criterion is considered to be better.

Multivariable logistic regression

The population probability (and its logit transformation) depends (conditionally) on a linear combination of *k* covariates or predictors.

$$(p|x_1, \ldots, x_k) = \text{invlogit}(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

$$\log(\text{odds}|x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

where x represents any type of predictor and the corresponding β is the conditional log odds ratio.

Let's suppose that in the hyponatremia study the main exposure of interest is gender.

$$\log(\text{odds}|\text{female}) = \beta_0 + \beta_1 \text{female}$$

Women (22%) were about 3 times more likely to develop hyponatremia compared to men (8%).

The crude or unadjusted (marginal) odds ratio was 3.4 (95% CI = 2 to 6).

This association may be partially or totally explain by other characteristics of the marathon runners (weight change, body mass index, running time).

Women may be more likely to gain weight and weight change is associated with higher risk of hyponatremia.

Therefore, one might expect that part of the association between gender and hyponatremia to be partly explained by weight change.

The association between gender and hyponatremia risk further adjusted for (conditionally on) weight change can be estimated by

$$\log(\text{odds}|\text{female}, \text{wtdiff}) = \beta_0 + \beta_1 \text{female} + \beta_2 \text{wtdiff}$$

The (adjusted) conditional odds ratio for female is likely to be different from the marginal (unadjusted) odds ratio (non-collapsibility property).

```
. logit nas135 female wtdiff, or

Logistic regression                                     Number of obs   =        455
                                                        LR chi2(2)      =      62.33
                                                        Prob > chi2     =     0.0000
Log likelihood =  -140.5062                             Pseudo R2       =     0.1815

------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   2.385677   .7417703     2.80   0.005     1.297032    4.388063
      wtdiff |   2.018988   .2298481     6.17   0.000     1.615215    2.523698
       _cons |   .1001656   .0235007    -9.81   0.000     .0632428     .158645
------------------------------------------------------------------------------
```

Compared to men and adjusting for weight change, women had 2.4 fold increase odds of hyponatremia (95% CI = 1.3 to 4.4).

“Adjusting for” or “controlling for” means that for any given value of weight change the odds of hyponatremia among women is 2.4 times larger the one among men.

Every 1 kg increase in weight the gender-adjusted odds of hyponatremia increased by 2-fold (95% CI = 1.6 to 2.5).

“Adjusting for” or “controlling for” means that for either men or women the odds of hyponatremia doubles.

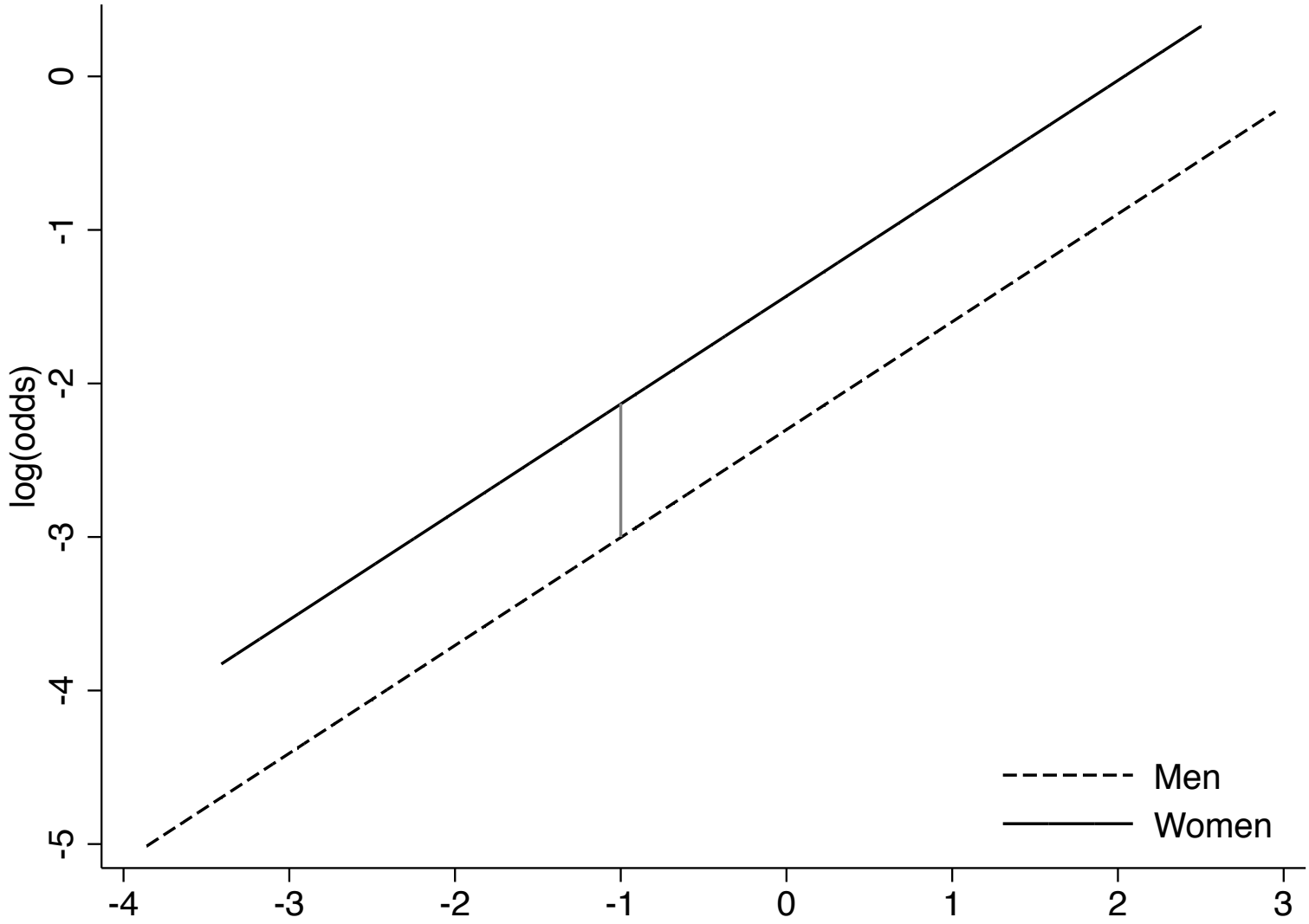
What is the population odds ratio of hyponatremia comparing women vs men among those increasing 1 kg?

$$\log(\text{odds}|\text{female} = 1, \text{wtdiff} = 1) - \log(\text{odds}|\text{female} = 0, \text{wtdiff} = 1)$$
$$= \beta_1 \text{female} + \beta_2(1 - 1) = \beta_1$$

What is the population odds ratio of hyponatremia comparing women vs men among those lost 2 kg?

$$\log(\text{odds}|\text{female} = 1, \text{wtdiff} = -2)$$
$$- \log(\text{odds}|\text{female} = 0, \text{wtdiff} = -2)$$
$$= \beta_1 \text{female} + \beta_2(-2 + 2) = \beta_1$$

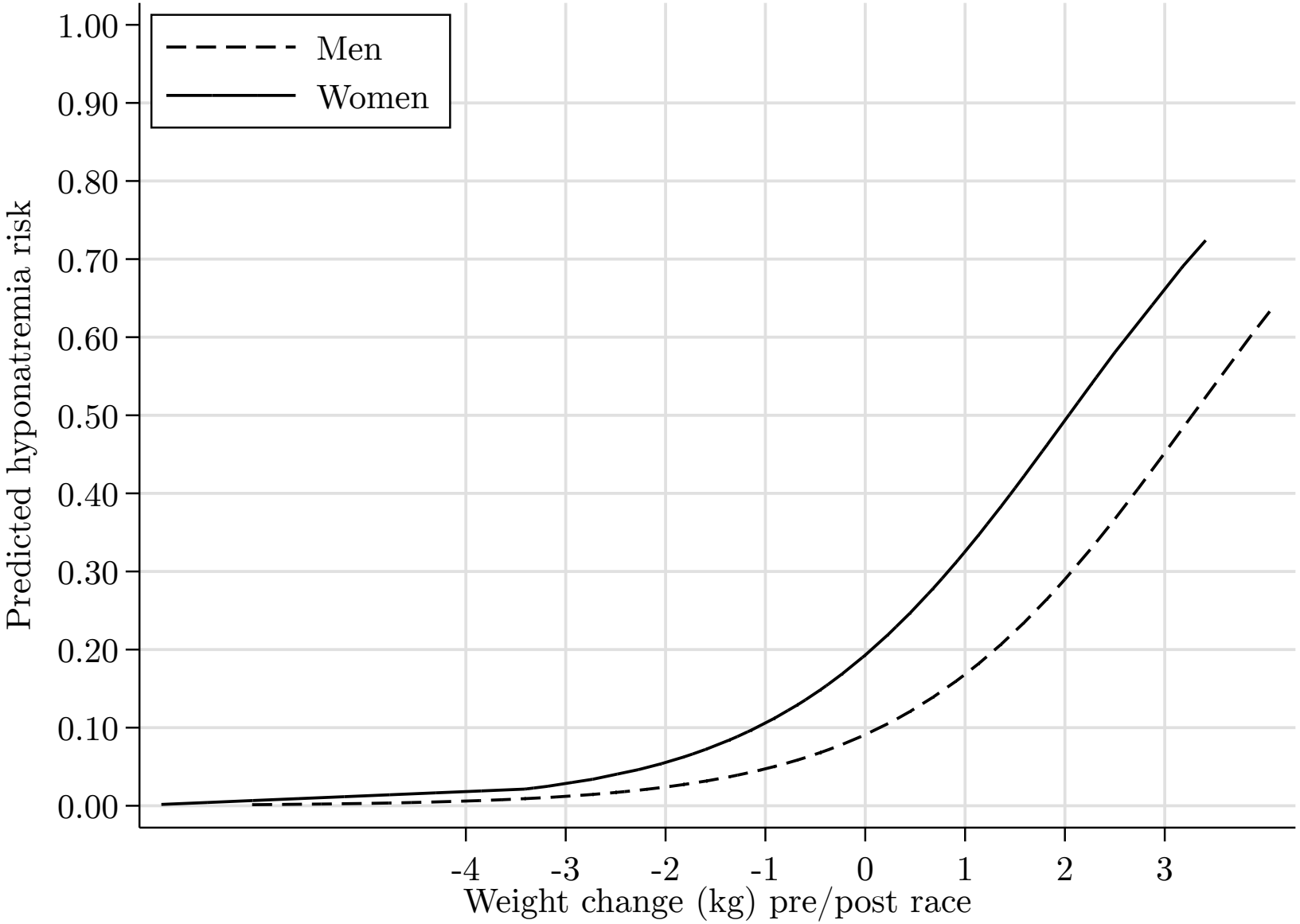
Graphical presentation of the multivariable model.



The predicted population hyponatremia risk is obtained by recalling the mathematical relationship between odds and risk.

$$(\text{risk}|\text{female}, \text{wtdiff}) = \frac{e^{(\beta_0+\beta_1\text{female}+\beta_2\text{wtdiff})}}{1 + e^{(\beta_0+\beta_1\text{female}+\beta_2\text{wtdiff})}}$$

$$(\text{risk}|\text{female}, \text{wtdiff}) = \text{invlogit}(\beta_0 + \beta_1\text{female} + \beta_2\text{wtdiff})$$

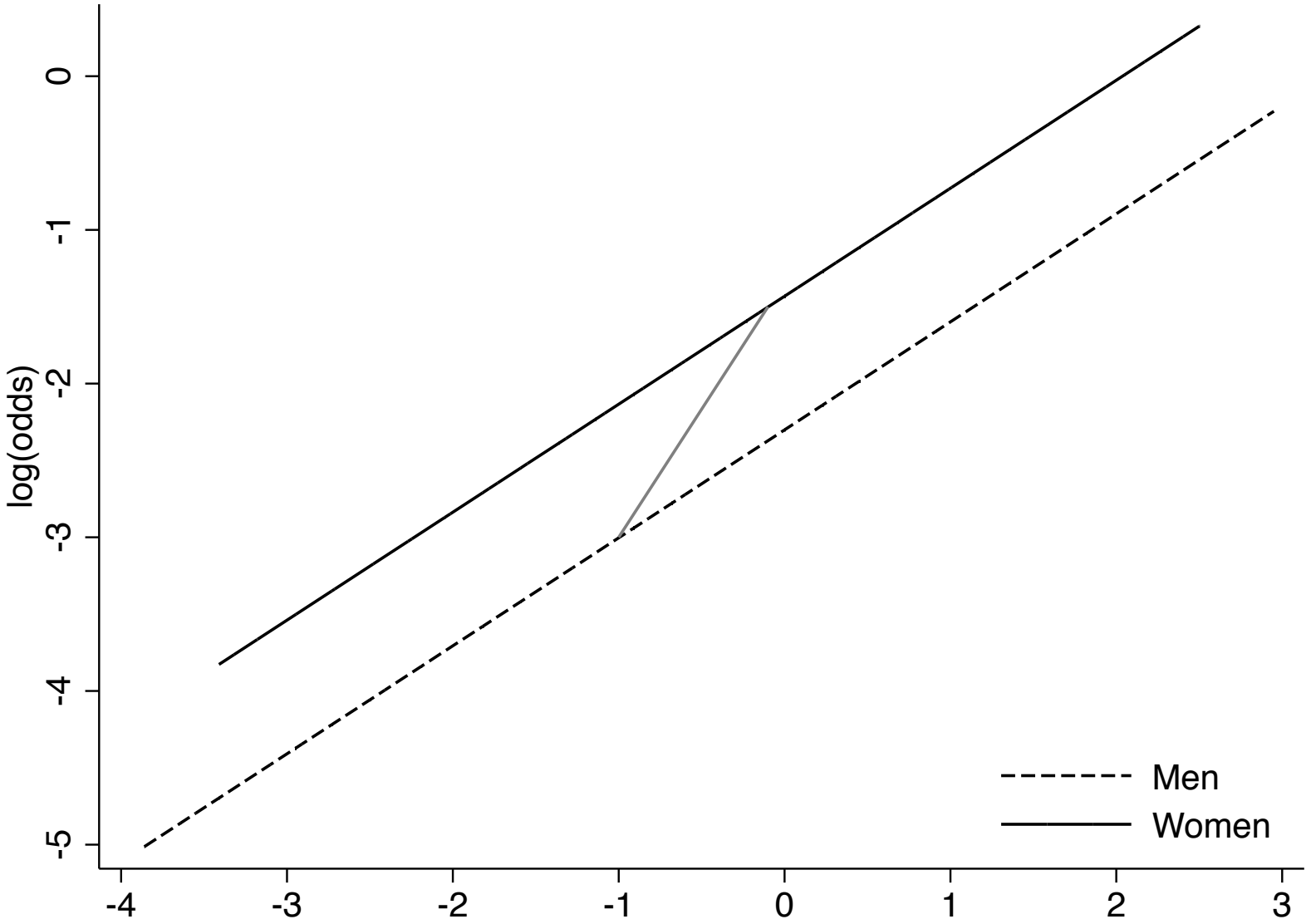

Men
Women
Predicted hyponatremia risk
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-4
-3
-2
-1
0
1
2
3
Weight change (kg) pre/post race

What is the odds ratio of hyponatremia comparing women who lost 0.1 kg vs men who lost 1 kg?

$$\log(\text{odds}|\text{female} = 1, \text{wtdiff} = -0.1) - \log(\text{odds}|\text{female} = 0, \text{wtdiff} = -1) = \beta_1(1-0) + \beta_2(-0.1+1) = \beta_1 + \beta_2 0.9$$

$$\text{OR} = \exp(\beta_1 + \beta_2 0.9) = 4.5$$

```
lincom _b[female]+_b[wtdiff]*.9, eform
```


log(odds)
0
-1
-2
-3
-4
-5
-4
-3
-2
-1
0
1
2
3
Men
Women

Statistical interaction

The question is: is the effect of the predictor *x* on the response *y* varying according to another factor *z*?

You could try to answer this question by doing stratified analysis or including an interaction term between *x* and *z* in the regression model.

Testing that the coefficient/s of the interaction term/s is/are equal to zero can help us to answer the question.

In epidemiological language, *z* is called effect modifier of the relation between *x* and *y*.

We can specify a model taking into account the possible interaction between the variable x and z

$$\log(\text{odds}|x,z) = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz$$

We can specify the above model in two equivalent ways

$$\log(\text{odds}|x,z) = \beta_0 + (\beta_1 + \beta_3 z)x + \beta_2 z$$

$$\log(\text{odds}|x,z) = \beta_0 + \beta_1 x + (\beta_2 + \beta_3 x)z$$

The effect of x on y depends on z via β_3

The effect of z on y depends on x via β_3

We test the hypothesis of no (multiplicative) interaction between x and z

$$H_0 : \beta_3 = 0$$

If we can't reject the null hypothesis we can remove the interaction term from the model.

Example: Does the association between weight change and hyponatremia risk vary with gender? Or viceversa, does the association between gender and hyponatremia vary according to weight change?

Binary X and Quantitative Z

The question is: is the effect of the predictor *x* on the response *y* varying according to a quantitative factor *z*?

1. We first create the interaction between weight change and female

```
. gen inter = wtdiff*female
```

2. We fit a model with weight change, female, and their interaction. We are assuming that weight change is linearly associated with the log odds of hyponatremia among men and women.

$$\log(\text{odds}) = \beta_0 + \beta_1 \text{wtdiff} + \beta_2 \text{female} + \beta_3 \text{inter}$$

```
. logistic nas135 wtdiff female inter

Logistic regression                                     Number of obs   =        455
                                                        LR chi2(3)      =      62.36
                                                        Prob > chi2     =     0.0000
Log likelihood = -140.49355                             Pseudo R2       =     0.1816

------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      wtdiff |   1.987425   .2986368     4.57   0.000     1.480425    2.668056
      female |   2.351299   .7620755     2.64   0.008     1.245737    4.438021
       inter |    1.03724   .2386386     0.16   0.874     .6607582     1.62823
       _cons |   .1007185   .0237613    -9.73   0.000     .0634301    .1599275
------------------------------------------------------------------------------
```

The *p*-value for interaction is large (p=0.874). Therefore there is no evidence of (multiplicative) interaction between gender and weight gain in predicting risk of hyponatremia.

Interpretation

$\exp(\beta_0)$ Among men who did not change weight, the odds of hyponatremia were 10 cases every 100 non-cases.

$\exp(\beta_1)$ Among men, every 1 kg increment in weight was associated with a 2 fold increase odds of hyponatremia.

$\exp(\beta_2)$ Compared to men, women had 2.35 fold increase odds of hyponatremia among those runners who did not change weight.

There are other odds ratios that can be estimated from the model.

Question 1. What is odds ratio of hyponatremia associated with every 1 kg increase in weight among women?

$$\exp(\beta_1 + \beta_3 1)$$

```
. lincom _b[wtdiff]+_b[inter]*1
```

nas135	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
(1)	2.061436	.3591489	4.15	0.000	1.465113	2.90047

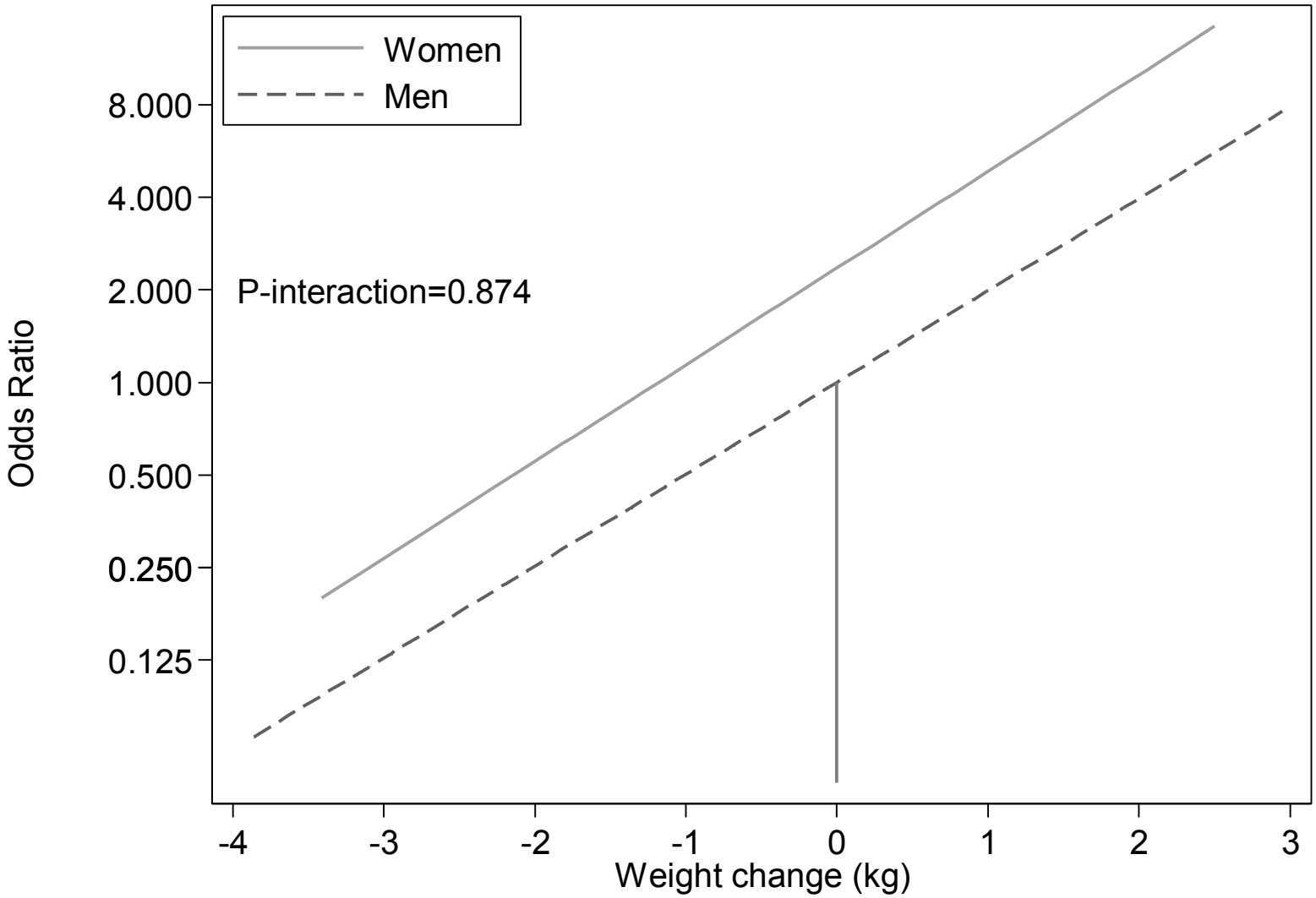
Among women, the odds of hyponatremia doubles (OR=2.1, 95% CI = 1.5-2.9) for every one kilogram increase in weight.

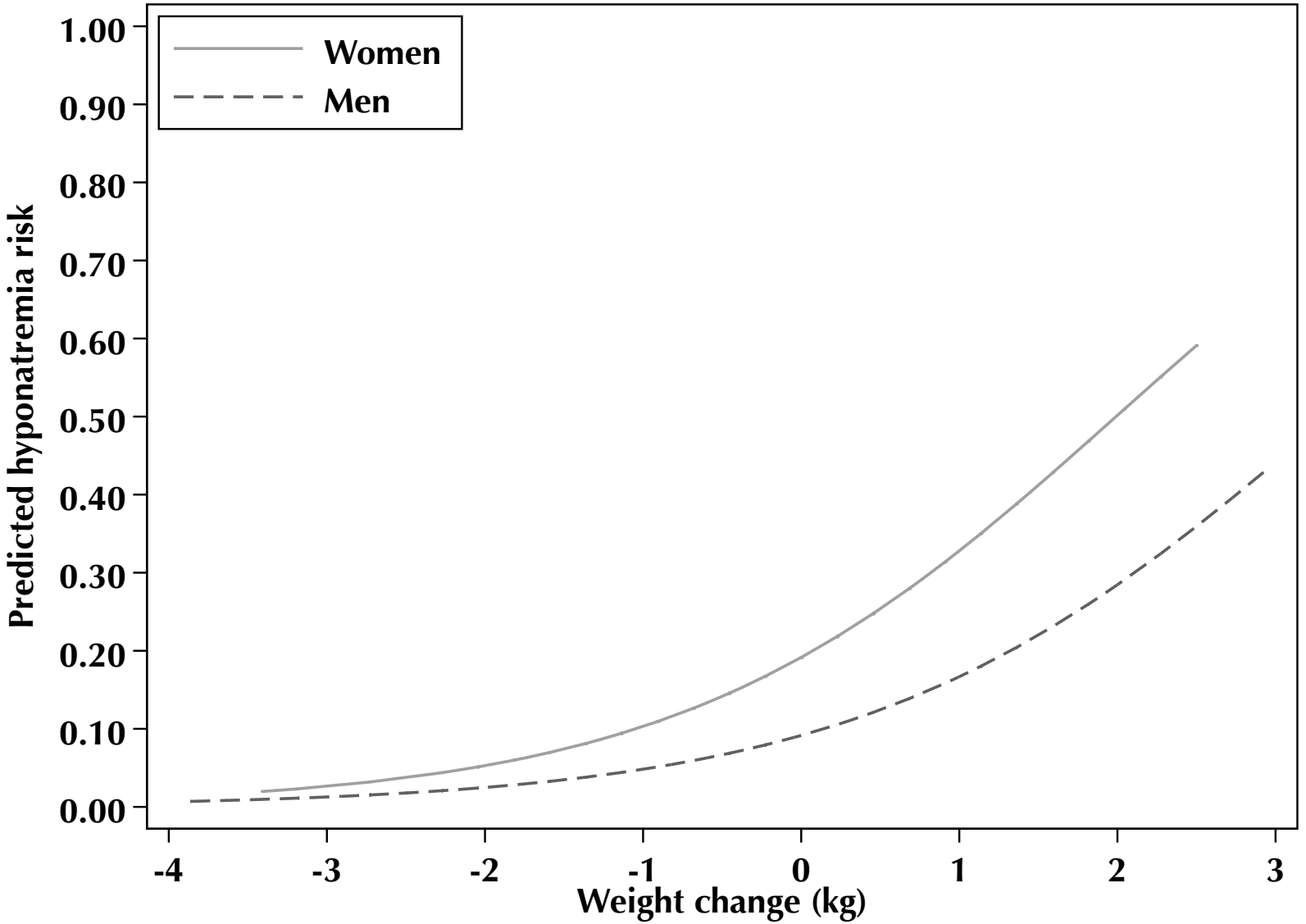
Question 2: What is the odds ratio of hyponatremia comparing women vs men among those runners who gained 3 kg during the race?

$$(\beta_2 + \beta_3 3)$$

```
. lincom _b[female]+_b[inter]*3

( 1)  [nas135]female + 3*[nas135]inter = 0

------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   2.623888   1.771526     1.43   0.153     .6986412    9.854544
------------------------------------------------------------------------------
```

Compared to men, women had 2.6 fold increase odds of hyponatremia, although not significant (95% CI = 0.7 to 10).


Women
Men
P-interaction=0.874
Odds Ratio
8.000
4.000
2.000
1.000
0.500
0.250
0.125
-4
-3
-2
-1
0
1
2
3
Weight change (kg)


Women
Men
Predicted hyponatremia risk
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-4
-3
-2
-1
0
1
2
3
Weight change (kg)

Binary *X* and binary *Z*

Suppose we have two binary predictors of hyponatremia risk: gain weight (post-race > pre-race weigth) and gender.

```
. gen gain = wtdiff > 0 if wtdiff !=.
. codebook gain

                  type:  numeric (float)
                 label:  gw

                 range:  [0,1]                        units:  1
         unique values:  2                    missing .:  33/488

            tabulation:  Freq.   Numeric  Label
                           320         0  Post<=Pre
                           135         1  Post>Pre
                            33
```

2. We fit a model with weight gain, female, and their product

$$\log(\text{odds}) = \beta_0 + \beta_1 \text{gain} + \beta_2 \text{female} + \beta_3 \text{inter}$$

```
. generate inter = gain*female
. logistic nas135 gain female inter

Logistic regression                               Number of obs     =        455
                                                  LR chi2(3)        =      53.31
                                                  Prob > chi2       =     0.0000
Log likelihood = -145.01767                       Pseudo R2         =     0.1553

------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        gain |   11.02976   5.492052     4.82   0.000     4.156494    29.26881
      female |   5.690763   2.902987     3.41   0.001     2.093887    15.46634
       inter |   .2964859   .1895923    -1.90   0.057     .0846624    1.038287
       _cons |   .0275229   .0113898    -8.68   0.000     .0122305    .0619364
------------------------------------------------------------------------------
```

There is no evidence of (multiplicative) interaction (p=0.057).

Interpretation

$\exp(\beta_0)$ Among men who lost weight, the odds of hyponatremia were 3 cases every 100 non-cases.

$\exp(\beta_1)$ Among men, those runners who gained weight had 11-fold increase odds of hyponatremia compared to those who lost weight (95% CI = 4- 29).

$\exp(\beta_2)$ Among those who lost weight, women had 5.7 fold increase odds of hyponatremia than men (95% CI = 2-15).

$\exp(\beta_1 + \beta_3)$ Among women, those runners who gained weight had 3.3-fold increased odds of hyponatremia compared to those who lost weight (95% CI = 1.5-7.2).

```
. lincom gain + inter, eform

------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   3.270169    1.31207     2.95   0.003     1.489525    7.179472
------------------------------------------------------------------------------
```

$\exp(\beta_2 + \beta_3)$ Among those who gained weight, women had 1.7 higher odds of hyponatremia than men (95% CI = 0.8-3.6).

```
. lincom female + inter, eform

------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.687231    .650602     1.36   0.175     .7924071    3.592533
------------------------------------------------------------------------------
```

Compared to men who lost weight, women who gained weight had 19 fold increased odds of hyponatremia (95% CI = 7 - 49).

```
. lincom gain + female + inter, eform

------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   18.60976    9.17865     5.93   0.000     7.078048    48.92917
------------------------------------------------------------------------------
```

Logistic regression model with a quadratic term

The quadratic model for a quantitative exposure x is

$$\log(\text{odds}|x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

The linear response model is nested in the quadratic model.

A p-value for linearity is obtained by testing the coefficient β_2 equal to zero.

If the p-value is small (saying < 0.05), there is departure from linearity that needs care and attention. Otherwise, the simpler linear model fits adequately the data.

We first generate a new variable containing weight change to the power of 2 (wtdiff squared).

```
. gen wtdiffsq = wtdiff^2
```

Then we fit the quadratic regression model

```
. logistic nas135 wtdiff wtdiffsq
```

nas135	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
wtdiff	2.141649	.2756308	5.92	0.000	1.664172	2.75612
wtdiffsq	.964032	.0588865	-0.60	0.549	.8552578	1.08664
_cons	.1613851	.0300918	-9.78	0.000	.1119821	.2325832

Question 1. Is weight change predicting the odds of hyponatremia?

We test simultaneously the two coefficients equal to zero

```
. testparm wtdiff wtdiffsq

 ( 1)  [nas135]wtdiff = 0
 ( 2)  [nas135]wtdiffsq = 0

           chi2(  2) =   41.49
         Prob > chi2 =    0.0000
```

The p-value is small, so the answer is yes.

Question 2. Is a quadratic model for weight change predicting odds of hyponatremia better compared to a simpler linear-response model?

We test the coefficient of the squared exposure equal to zero

```
. testparm wtdiffsq

 ( 1)  [nas135]wtdiffsq = 0

           chi2(  1) =    0.36
         Prob > chi2 =    0.5487
```

The p-value is large, so the answer is no.

There is no evidence of departure from linearity (*P*-value = 0.55).

Question 3. What is the odds ratio of hyponatremia comparing those who increased 2 kg as compared to those who did not change weight?

To put it more generally, the predicted mean responses for any two values of x of a quadratic model are

$$\log(\text{odds}|x = x_1) = \beta_0 + \beta_1 x_1 + \beta_1 x_1^2$$

$$\log(\text{odds}|x = x_2) = \beta_0 + \beta_1 x_2 + \beta_2 x_2^2$$

The log odds ratio is given by

$$\log(\text{odds}|x = x_1) - \log(\text{odds}|x = x_2) =$$

$$\beta_1(x_1 - x_2) + \beta_2(x_1^2 - x_2^2)$$

The odds ratio is given by

$$\text{OR} = \frac{\text{odds}|x = x_1}{\text{odds}|x = x_2} = \exp(\beta_1(x_1 - x_2) + \beta_2(x_1^2 - x_2^2))$$

Once again, you can easily estimate the above quantity with the postestimation commands **lincom** or **predictnl**.

Example, using the post-estimation **lincom** command.

```
. lincom _b[wtdiff]*(2-0) + _b[wtdiffsq]*(4-0)

 ( 1)  2*[nas135]wtdiff + 4*[nas135]wtdiffsq = 0

------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   3.961524   1.015595     5.37   0.000     2.396869    6.547571
------------------------------------------------------------------------------
```

Compare to those runners who did no change weight, those runners who increased 2 kg had a 4 fold increase odds of hyponatremia.

One can tabulate differences in mean responses for a list of specific values of the exposure.

Question 4. How to plot the odds ratios with 95% confidence intervals as function of the exposure using a chosen exposure value as reference?

To create a plot we need to store the numbers we are interested in as variables. Once again, we can use the post-estimation command predictnl

```
predictnl logorsq = _b[wtdiff]*(wtdiff-0)+ ///
                    _b[wtdiffsq]*(wtdiffsq-0)

gen orsq = exp(logorsq)
```

We now overlay the three approaches used so far.

```
tw (line orl orsq lbl ubl wtdiff, sort ///
    lp(l dash dot dot) ) ///
(pci 0.02 0 1 0) ///
if inrange(wtdiff, -4,3) , ///
text(8 -3 "P<0.001") ///
xlabel(-4(1)3) scheme(s1mono) legend(off) ///
ylabel(.125 0.25 0.25 0.5 1 2 4 8 ///
, angle(horiz) format(%4.3fc))  ///
ytitle("Odds Ratio") ///
xtitle("Weight change (kg)") yscale(log)
```

Let's compare 4 logistic regression models of increasing complexity

```
egen nmissing = rowmiss(nas135 wtdiff bmi runtime)
keep if nmissing==0

logit nas135 wtdiff
estimate store m1

logit nas135 wtdiff runtimeh
estimate store m2

logit nas135 wtdiff runtimeh  bmi bmisq
estimate store m3

logit nas135 wtdiff runtimeh  bmi bmisq female
estimate store m4
```

```
. estimates table m1 m2 m3 m4, stat(N ll aic) star eform

-------------------------------------------------------------
    Variable |      m1          m2          m3          m4
-------------+-----------------------------------------------
      wtdiff |    2.05***      2.01***     2.08***     2.06***
    runtimeh |                 2.64***     2.63***     2.33**
         bmi |                             0.11***     0.12***
       bmisq |                             1.05***     1.05***
      female |                                         1.70
       _cons |    0.15***      0.00***    1.8e+09**   4.6e+08**
-------------+-----------------------------------------------
           N |       446         446         446         446
          ll |   -142.24      -132.99     -126.16     -125.34
         aic |    288.48       271.98      262.32      262.69
-------------------------------------------------------------
                     legend: * p<0.05; ** p<0.01; *** p<0.001
```

The logistic model associated with a better fit of the data is **m3** with the lowest AIC.

The final multivariable model fitted for the risk of hyponatremia is

$$\log(\text{odds}|\text{wtdiff}, \text{runtime}, \text{bmi}) = \beta_0 + \beta_1 \text{wtdiff} + \beta_2 \text{runtime} + \beta_3 \text{bmi} + \beta_4 \text{bmi}^2$$

```
. logit nas135 wtdiff runtime bmi bmisq

Logistic regression                               Number of obs   =        446
                                                  LR chi2(4)      =      84.73
                                                  Prob > chi2     =     0.0000
Log likelihood = -126.16244                       Pseudo R2       =     0.2514

------------------------------------------------------------------------------
      nas135 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      wtdiff |   .7327378   .1231096     5.95   0.000     .4914474    .9740282
     runtime |   .0161168   .0043701     3.69   0.000     .0075516     .024682
         bmi |  -2.229458   .6222846    -3.58   0.000    -3.449113   -1.009802
       bmisq |    .044857     .01293     3.47   0.001     .0195148    .0701993
       _cons |   21.33502   7.480436     2.85   0.004     6.673632     35.9964
------------------------------------------------------------------------------
```

Overall, the predictors included in multivariable logistic regression are significantly associated with the risk of hyponatremia.

The p-value associated with the likelihood ratio test for the null hypothesis that all the regression coefficients of the predictors are jointly equal to zero is small (LRT=84.73, p<0.001).

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

For any given set of predictors one can estimate the log odds, odds, and probability of hyponatremia using the equation

$$\log(odds) = 21.3 + 0.73\ \text{wtdiff} + 0.16\ \text{runtime} - 2.2\ \text{bmi} + 0.04\ \text{bmi}^2$$

Example 1. What is the predicted risk of hyponatremia for persons who increase 3 kg, complete the marathon in 4 hours, and have a BMI of 19 kg/m^2?

$$\log(odds) = 21.3 + 0.73\times3 + 0.16\times240 - 2.2\times19 + 0.04\times(19\times19)$$

$$\log(odds) = 1.8$$

$$\text{odds} = \exp(1.8) = 6$$

$$\text{p} = \exp(1.8)/(1 + \exp(1.8)) = 0.9$$

Odds of hyponatremia are 6 cases for every 1non-case.

Risk of hyponatremia is 90%.

```
. scalar logodds = _b[_cons] + ///
_b[wtdiff]*3+_b[runtime]*240 + ///
_b[bmi]*18 + _b[bmisq]*324

. display logodds
1.8046985

. scalar odds = exp(logodds)

. scalar risk = invlogit(logodds)

. display odds
6.0781387

. display risk
.85871992
```

Example 2. What is the predicted risk of hyponatremia for persons who loose 1 kg, complete the marathon in 3.5 hours, and have a BMI of 25 kg/m^2?

$$\log(odds) = 21.3 + 0.73\times -1 + 0.16\times 210 - 2.2\times 25 + 0.04\times 625$$

$$\log(odds) = -3.7$$

$$\text{odds} = \exp(-3.7) = 0.02$$

$$\text{p} = \exp(-3.7)/(1 + \exp(-3.7)) = 0.02$$

Odds of hyponatremia are 2 cases for every 100 non-cases.

Risk of hyponatremia is 2%, 2 cases every 100 cases.

```
. scalar logodds = _b[_cons] + ///
 _b[wtdiff]*-1+_b[runtime]*210 + ///
 _b[bmi]*25 + _b[bmisq]*625

. display logodds
-3.7139972

. scalar odds = exp(logodds)

. scalar risk = invlogit(logodds)

. display odds
.02437988

. display risk
.02379964
```

The predicted risk of the outcome depends on the linear combination of the *k* predictors.

$$\log(odds|x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

$$p = odds/(1 + odds)$$

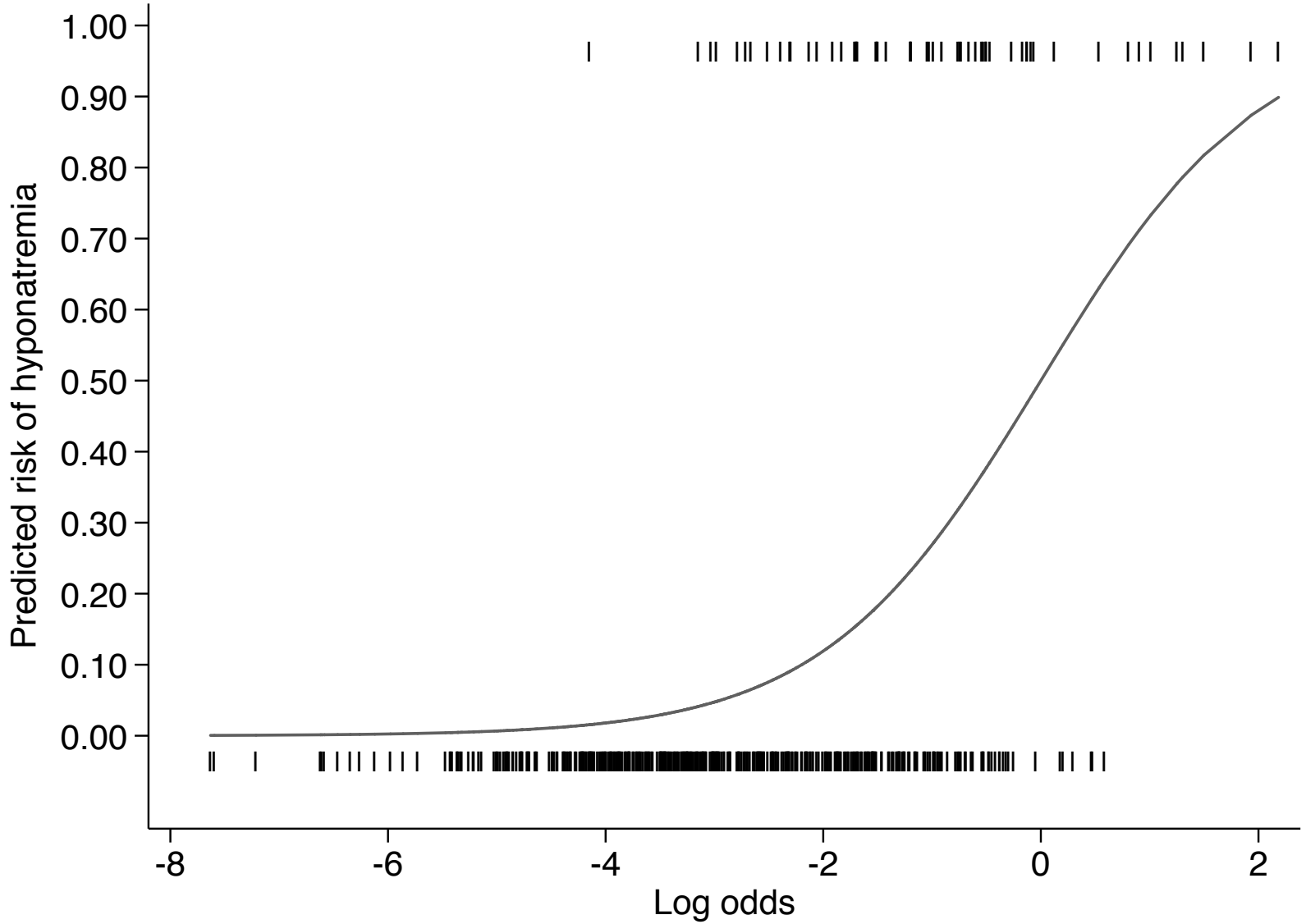
$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}$$

The greater is the logit or log odds value and the higher is the odds and risk of the outcome.

```
. logit nas135 wtdiff runtime bmi bmisq

Logistic regression                                     Number of obs   =        446
                                                        LR chi2(4)      =      84.73
                                                        Prob > chi2     =     0.0000
Log likelihood = -126.16244                             Pseudo R2       =     0.2514

------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      wtdiff |   2.080769   .2561627     5.95   0.000      1.63468    2.648592
     runtime |   1.016247   .0044411     3.69   0.000      1.00758    1.024989
         bmi |   .1075867   .0669496    -3.58   0.000     .0317738     .364291
       bmisq |   1.045878   .0135232     3.47   0.001     1.019706    1.072722
       _cons |   1.84e+09   1.38e+10     2.85   0.004     791.2645    4.30e+15
------------------------------------------------------------------------------

* Get variables containing log odds and risks

. predict logodds, xb

. predict p, pr
```

We now plot the predicted risk of hyponatremia as function of the values of the log odds (or logit score).

```
gen pipe = "|"

twoway ///
  (line p logodds, sort) ///
  (scatter nas135 logodds, ms(none) ///
           mlabpos(6) mlabel(pipe)) , ///
 ylabel(0(.1)1, angle(horiz) format(%3.2fc)) ///
 yscale(range(-.1 1)) ///
 plotregion(style(none)) legend(off) ///
 ytitle("Predicted risk of hyponatremia") ///
 xtitle("Log odds")
```


Predicted risk of hyponatremia
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-8
-6
-4
-2
0
2
Log odds

```
. lincom wtdiff*1, eform cformat(%2.1fc)

------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |        2.1        0.3     5.95   0.000           1.6          2.6
------------------------------------------------------------------------------

. lincom runtime*30, eform cformat(%2.1fc)
------------------------------------------------------------------------------
      nas135 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |        1.6        0.2     3.69   0.000           1.3          2.1
------------------------------------------------------------------------------

. testparm bmi bmisq

 ( 1)  [nas135]bmi = 0
 ( 2)  [nas135]bmisq = 0

           chi2(  2) =   14.18
         Prob > chi2 =    0.0008
```

Summary of the findings

A 1-kg increase in weight conferred an odds ratio of 2.0 (95% CI=1.6-2.6).

A 30-minute increase in running time conferred an odds ratio of 1.6 (95% CI=1.3-2.1).

Body-mass-index extremes were also associated with hyponatremia. An overall Wald-test suggests that body-mass index is significantly associated with hyponatremia ($p<0.001$).

Graphing multivariable adjusted odds ratios

Figure 2 of the NEJM paper shows adjusted odds ratios as function of the continuous predictors arising from the final multivariable logistic regression model.

$$\log(odds) = \beta_0 + \beta_1 \text{wtdiff} + \beta_2 \text{runtime} + \beta_3 \text{bmi} + \beta_4 \text{bmi}^2$$

$$\text{OR}_{\text{wtdiff}} = \exp[\beta_1(\text{wtdiff} - 0)]$$

$$\text{OR}_{\text{runtime}} = \exp[\beta_2(\text{runtime} - 210)]$$

$$\text{OR}_{\text{bmi}} = \exp[\beta_3(\text{bmi} - 22.5) + \beta_4(\text{bmi}^2 - 506.25)]$$

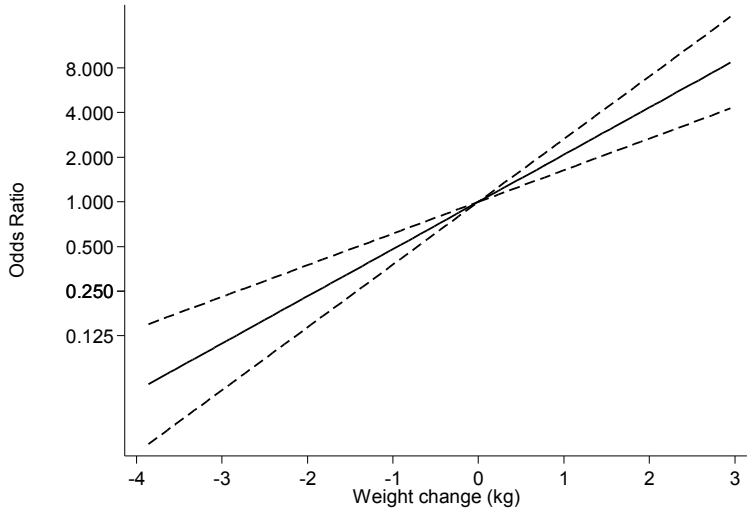
One can easily display or save the adjusted odds ratios (with a 95% CI) with postestimation commands like **lincom** or **predictnl**.

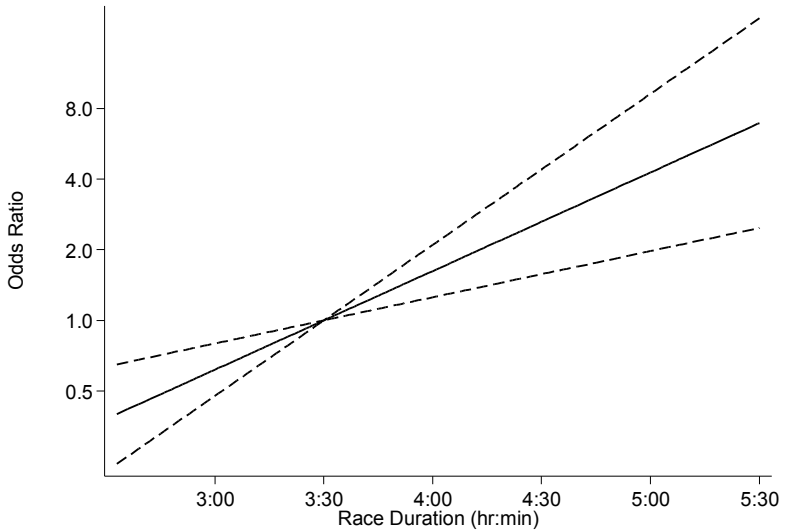
The post-estimation command **xblc** can simplify tabulating and plotting adjusted associations regardless of how the continuous predictors are modeled. The paper also shows how to fit different kinds of spline models (more flexible than the quadratic model

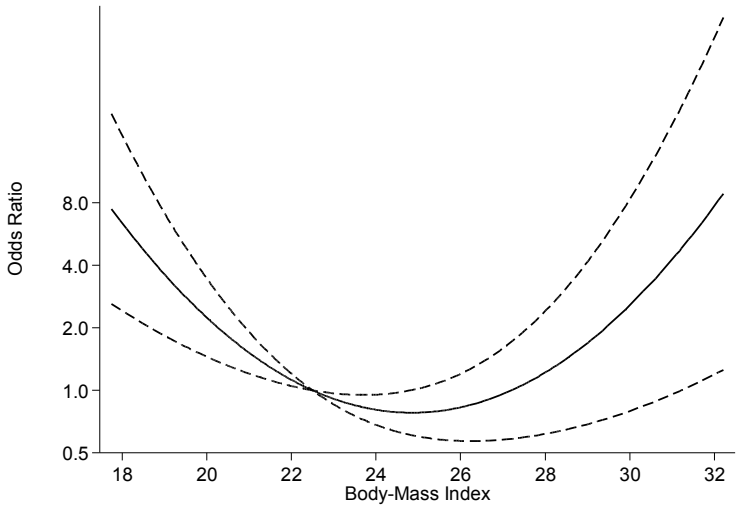
Orsini N., Greenland S. A procedure to tabulate and plot results after flexible modeling of a quantitative covariate. *Stata Journal*. 2011. 11, Number 1, pp. 1–29.

Figure 2

```
qui levelsof wtdiff  if inrange(wtdiff, -4,3)
qui xblc wtdiff  , cov(wtdiff ) ref(0) at(`r(levels)') eform ///
yscale(log) line ///
ylabel(.125 0.25 0.25 0.5 1 2 4 8 ///
, angle(horiz) format(%4.3fc))  ///
ytitle(Odds Ratio)  ///
xtitle("Weight change (kg)") ///
name(g1, replace)

qui levelsof runtimeh if inrange(runtimeh, 2.5, 5.5)
qui xblc runtimeh , cov(runtimeh) ref(3.5) at(`r(levels)') eform ///
yscale(log) line ///
ylabel(.5 1 2 4 8) ytitle(Odds Ratio) ///
xlabel(3 "3:00" 3.5 "3:30" 4 "4:00" 4.5 "4:30" 5 "5:00" 5.5 "5:30") ///
xtitle("Race Duration (hr:min)") ///
name(g2, replace)

qui levelsof bmi if inrange(bmi, 17.5, 33)
qui xblc bmi bmisq  , cov(bmi) ref(22.5) at(`r(levels)') ///
eform yscale(log) line ///
ylabel(.5 1 2 4 8) ///
xtitle("Body-Mass Index") ///
ytitle(Odds Ratio)  ///
name(g3, replace)
```


Odds Ratio
8.000
4.000
2.000
1.000
0.500
0.250
0.125
-4
-3
-2
-1
0
1
2
3
Weight change (kg)


Odds Ratio
8.0
4.0
2.0
1.0
0.5
3:00
3:30
4:00
4:30
5:00
5:30
Race Duration (hr:min)


Odds Ratio
8.0
4.0
2.0
1.0
0.5
18
20
22
24
26
28
30
32
Body-Mass Index