A command for Laplace regression

Nicola Orsini

Joint work with Matteo Bottai

4th Nordic and Baltic Stata Users Group meeting 2011

11/11/11

Outline

- Laplace regression model with censored data
- Example 1 – Randomized Clinical Trial
- Laplace regression model with no censored data
- Example 2 – Observational study

Laplace regression with censored data

- Let T_i be a sample of a continuous variable (time to an event) with $i = 1, \ldots, n$
- T_i may be censored (study end, lost to follow-up)
- $y_i = \min(T_i, C_i)$
- $\delta_i = \mathrm{I}(T_i < C_i)$

Let x_i be a k-dimensional vectors of covariates

- $y_i = x'_i\beta_p + u_i$
- $P(u_i |\, x_i \leq 0) = p$ with $p \in (0, 1)$
- $P(\, y_i \leq x'_i\beta_p | x_i) = p$
- $x'_i\beta_p$ is the p -quantile of the conditional distribution of y_i given x_i

Asymmetric Laplace Distribution

We assume that conditionally on covariates the response variable follows an asymmetric Laplace distribution with probability density function

- $f(y_i | x_i) = \exp[(I(\mathrm{B}) - p)z)]\, p(1\text{-}p)/\sigma_\mathrm{p}$

and cumulative distribution function

- $F(y_i | x_i) = \exp[(I(\mathrm{B}) - p)z)](p\text{-}\, I(\mathrm{A})) + I(\mathrm{A})$

where $I(\mathrm{B}) = (y_i \leq x'_i\beta_p)$, $I(\mathrm{A}) = (y_i > x'_i\beta_p)$, and $\mathrm{z} = (y_i - x'_i\beta_p)/\sigma_\mathrm{p}$

Log-Likelihood

The contribution of the i-th observation to the log-likelihood is

- $l(\beta_p, \sigma_p | y_i, x_i, \delta_i) = \sum_{i=1}^{n} [\delta_i \log f(y_i | x_i) + (1 - \delta_i) \log(1 - F(y_i | x_i))]$

The likelihood estimating equations (first derivatives in β_p and σ_p equal to zero) do not have a closed form solution.

Estimation algorithm

- Maximize the log-likelihood function directly
- Continuity and concavity of the log-likelihood
- Maximum exists and is global
- Gradient search maximization algorithm
- Standard errors based on bootstrap samples

Example 1

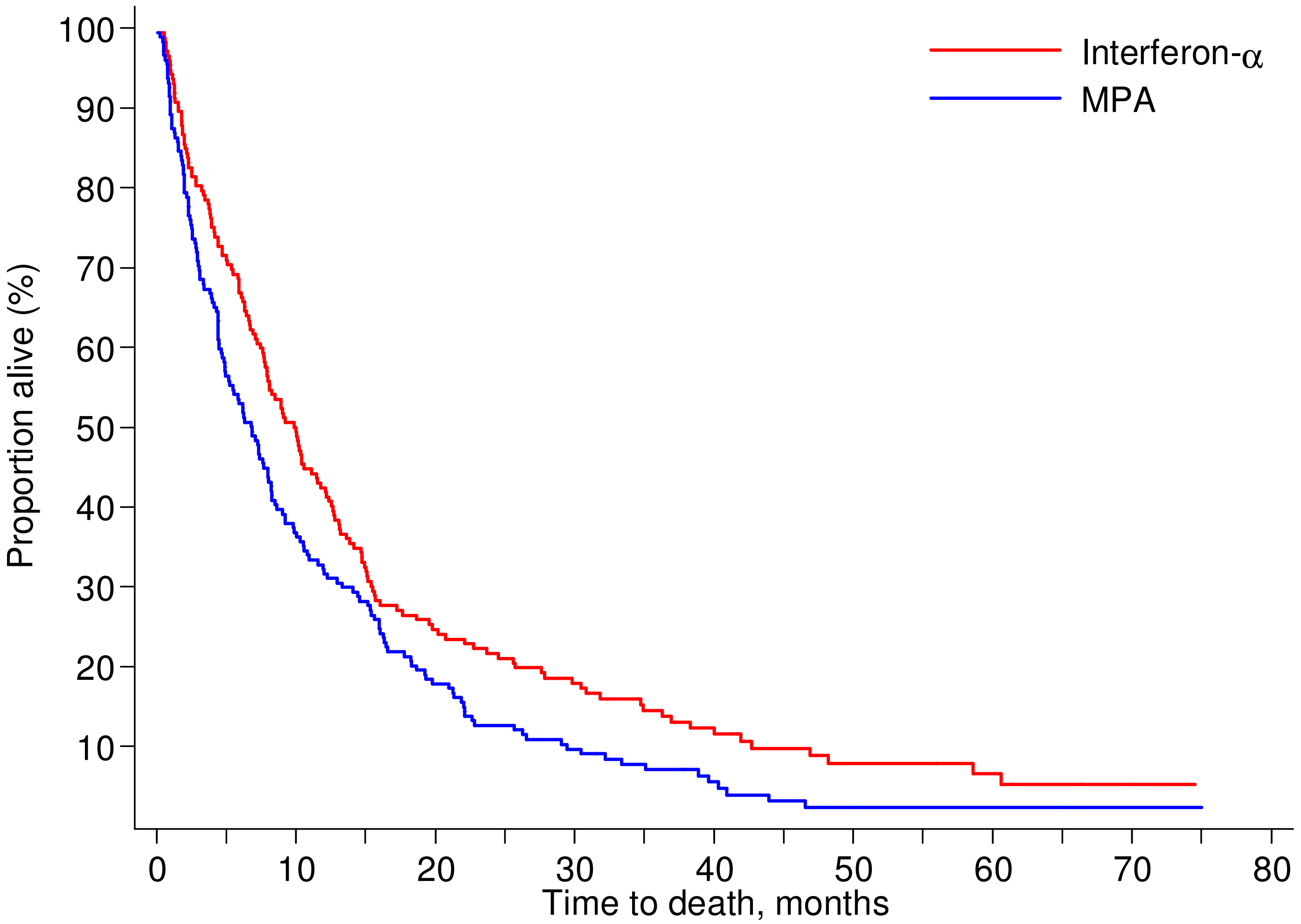
Lancet 1999; **353:** 14–17

ARTICLES

Interferon-α and survival in metastatic renal carcinoma: early results of a randomised controlled trial

*Medical Research Council Renal Cancer Collaborators**

Kaplan-Meier



```
. laplace month trt , failure(cens) q(10 50 90) reps(500)
```

```
Simultaneous Laplace regression                          Number of obs =        347
  Optimization: Gradient Search
  bootstrap(500) SEs                               q10 Log likelihood =     -1.5581
                                                   q50 Log likelihood =     -1.9044
                                                   q90 Log likelihood =     -2.6071
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
       month |      Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
q10          |
         trt |        0.6         0.3     1.84   0.066          -0.0         1.2
       _cons |        1.0         0.2     5.67   0.000           0.6         1.3
-------------+----------------------------------------------------------------
q50          |
         trt |        3.1         1.2     2.50   0.013           0.7         5.5
       _cons |        6.8         0.8     8.54   0.000           5.3         8.4
-------------+----------------------------------------------------------------
q90          |
         trt |       11.7         5.9     1.99   0.047           0.1        23.2
       _cons |       29.5         4.2     6.98   0.000          21.2        37.8
------------------------------------------------------------------------------
```

- Ninety percent of patients assigned to interferon-α therapy have at least two more months to live

10th percentile of survival is 2 month

- Half of the patients assigned to interferon-α therapy live longer than ten months

Median survival is 10 months

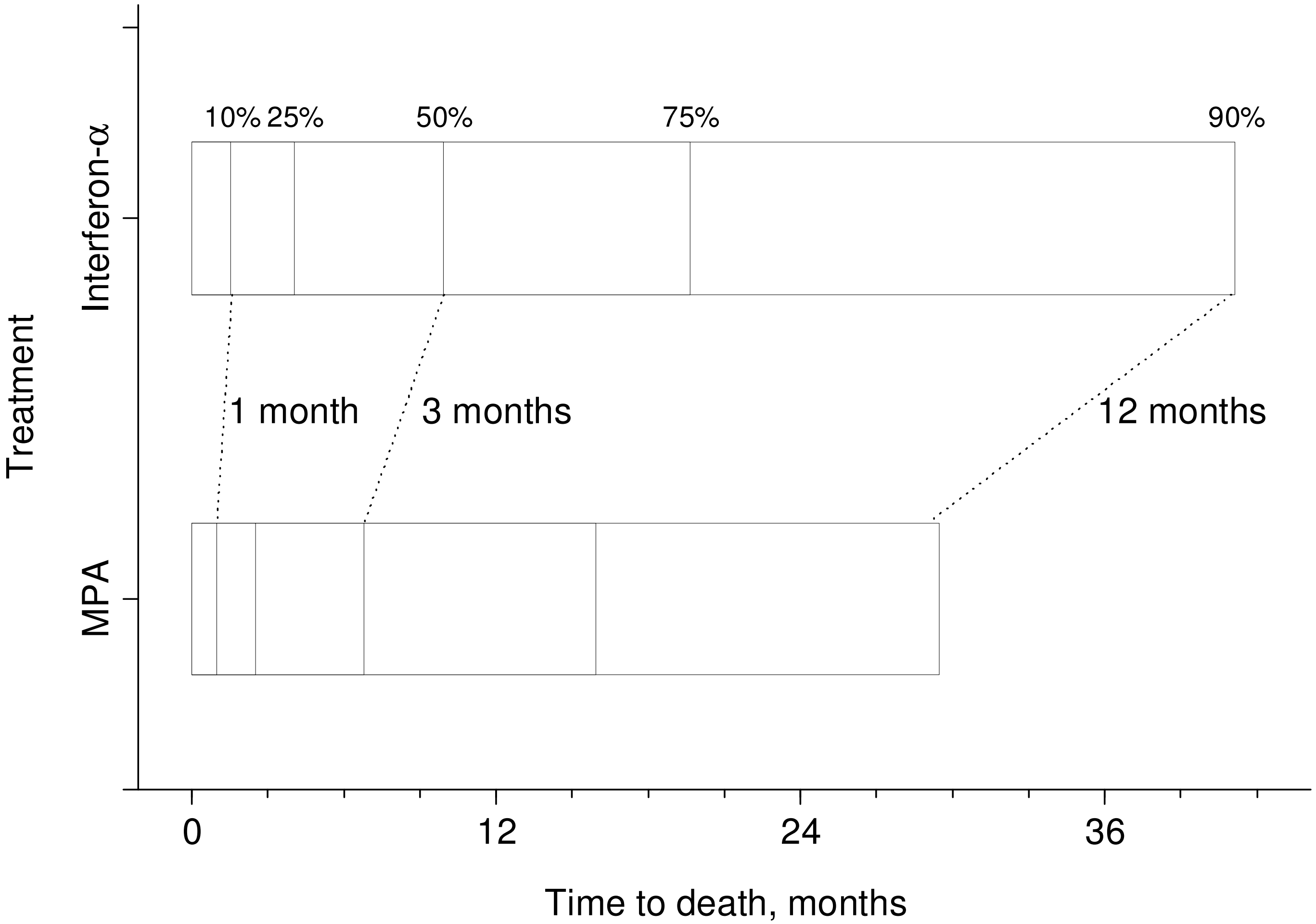
- Ten percent of the patients assigned to interferon-α therapy live longer than 41 months

90th percentile of survival is 41 months

Tabulate percentiles

Time to death, months	Interferon-α	MPA	Percentile Difference (95% CI)	*P*-value
10th Percentile	1.6	1.0	0.6 (-0.0, 1.2)	0.066
Median	9.9	6.8	3.1 (0.7, 5.5)	0.013
90th Percentile	41.2	29.5	11.7 (0.1, 23.2)	0.047

Plot percentiles



Testing across percentiles

```
. test [q10]trt = [q50]trt

 ( 1)  [q10]trt - [q50]trt = 0

       F(  1,   345) =    4.51
            Prob > F =    0.0343
```

The treatment effect on the 10th (1 month) and 50th (3 months) survival percentiles are significantly different.

Flexible modeling of predictors

We can investigate the change in survival percentiles according to a quantitative covariate using flexible tools (fractional polynomials, splines).

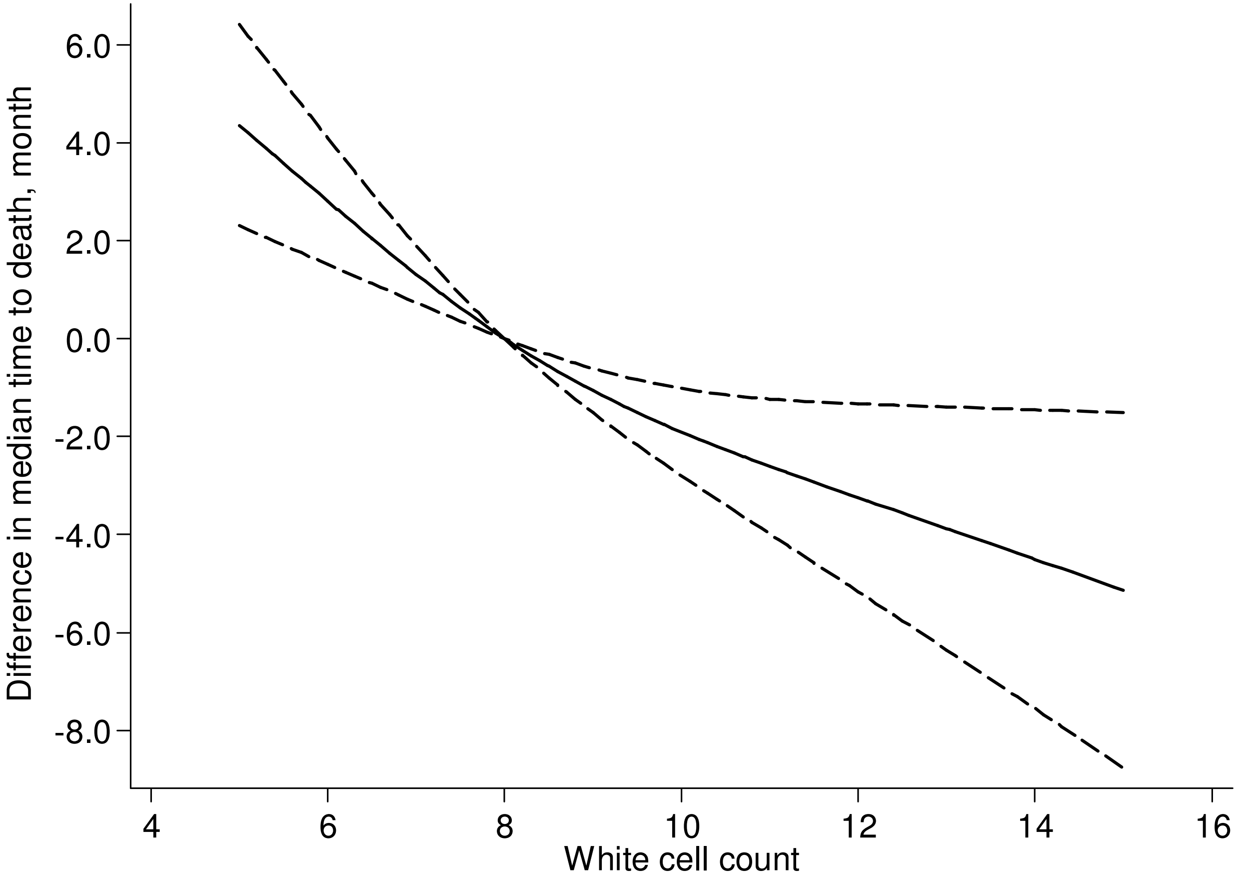
Let's consider white cell count (`wcc`) as predictor of the median survival (or any other percentile).

Laplace with restricted cubic splines

```
. mkspline wccs = wcc, nk(3) cubic

. laplace month trt wccs1 wccs2, failure(cens)

Laplace regression                              Number of obs =        343
  Optimization: Gradient Search
  bootstrap(500) SEs                         q50 Log likelihood =    -1.8687
------------------------------------------------------------------------------
             |  Observed   Bootstrap                        Normal-based
       month |     Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
q50          |
         trt |       2.3         1.1     2.20   0.029          0.2         4.4
       wccs1 |      -1.6         0.4    -3.94   0.000         -2.3        -0.8
       wccs2 |       0.8         0.4     1.79   0.075         -0.1         1.7
       _cons |      19.1         3.3     5.78   0.000         12.6        25.6
------------------------------------------------------------------------------
```

Some indication of departure from linearity for white cell count (p=0.075).

```
. levelsof wcc if inrange(wcc, 5, 15)

. xblc wccs1 wccs2 , cov(wcc) at(`r(levels)') line ///
   ytitle("Difference in median time to death, month") ///
   xtitle("White cell count") eq(q50) ref(8)
```



Interactions

British Journal of Cancer (2004) 90, 794–799

Is treatment with interferon-α effective in all patients with metastatic renal carcinoma? A new approach to the investigation of interactions

P Royston[*,1], W Sauerbrei[2] and A Ritchie[3]

Laplace with restricted cubic splines and interactions

```
. gen inter1 = wccs1*trt
. gen inter2 = wccs2*trt
```

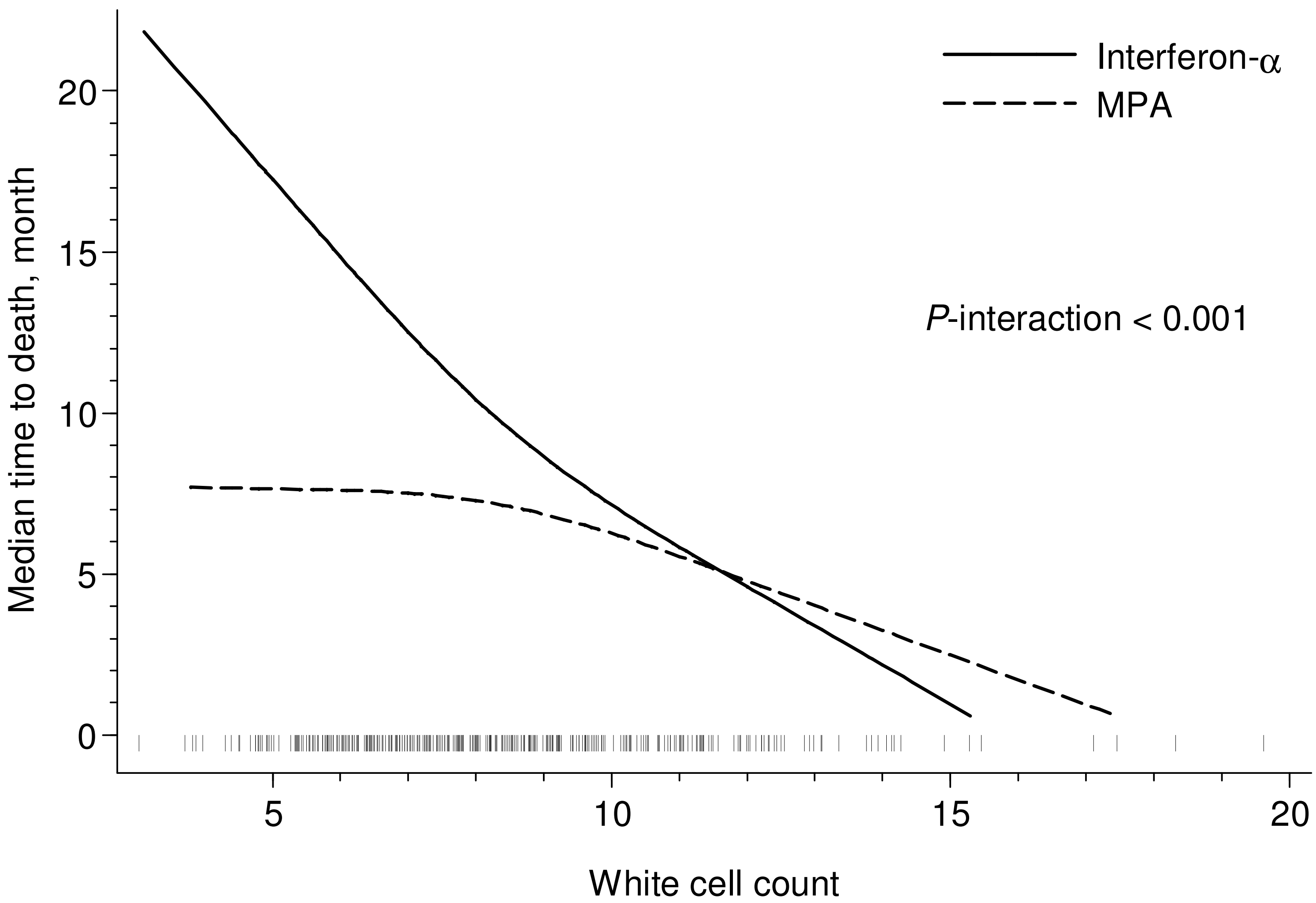
```
laplace month trt wccs1 wccs2 inter1 inter2 , failure(cens)
```

```
Laplace regression                                  Number of obs =       343
  Optimization: Gradient Search
  bootstrap(500) SEs                            q50 Log likelihood =    -1.8647
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
       month |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
q50          |
         trt |       21.4         1.2    18.07   0.000          19.1       23.8
       wccs1 |       -0.0         0.3    -0.13   0.900          -0.7        0.6
       wccs2 |       -0.6         0.3    -1.78   0.075          -1.3        0.1
      inter1 |       -2.4         0.1   -27.28   0.000          -2.5       -2.2
      inter2 |        1.6         0.3     4.77   0.000           1.0        2.3
       _cons |        7.8         2.5     3.10   0.002           2.9       12.8
------------------------------------------------------------------------------
```

Is the treatment effect dependent of white cell counts?

```
. testparm  inter1 inter2

 ( 1)   [q50]inter1 = 0
 ( 2)   [q50]inter2 = 0

       F(  2,   337) =  373.12
            Prob > F =    0.0000
```

Yes ($p<0.001$), based on a joint test that the two coefficients for interaction are equal to zero.


Interferon-α
MPA
P-interaction < 0.001
Median time to death, month
White cell count
0
5
10
15
20
5
10
15
20

Laplace regression with no censored data

- If there is no censoring ($\delta_i = 1$ for all observations)
- Laplace likelihood estimator is simplified
- Laplace regression is equivalent to quantile regression

```
laplace = sqreg
```

Example 2

The NEW ENGLAND JOURNAL *of* MEDICINE

ORIGINAL ARTICLE

Hyponatremia among Runners
in the Boston Marathon

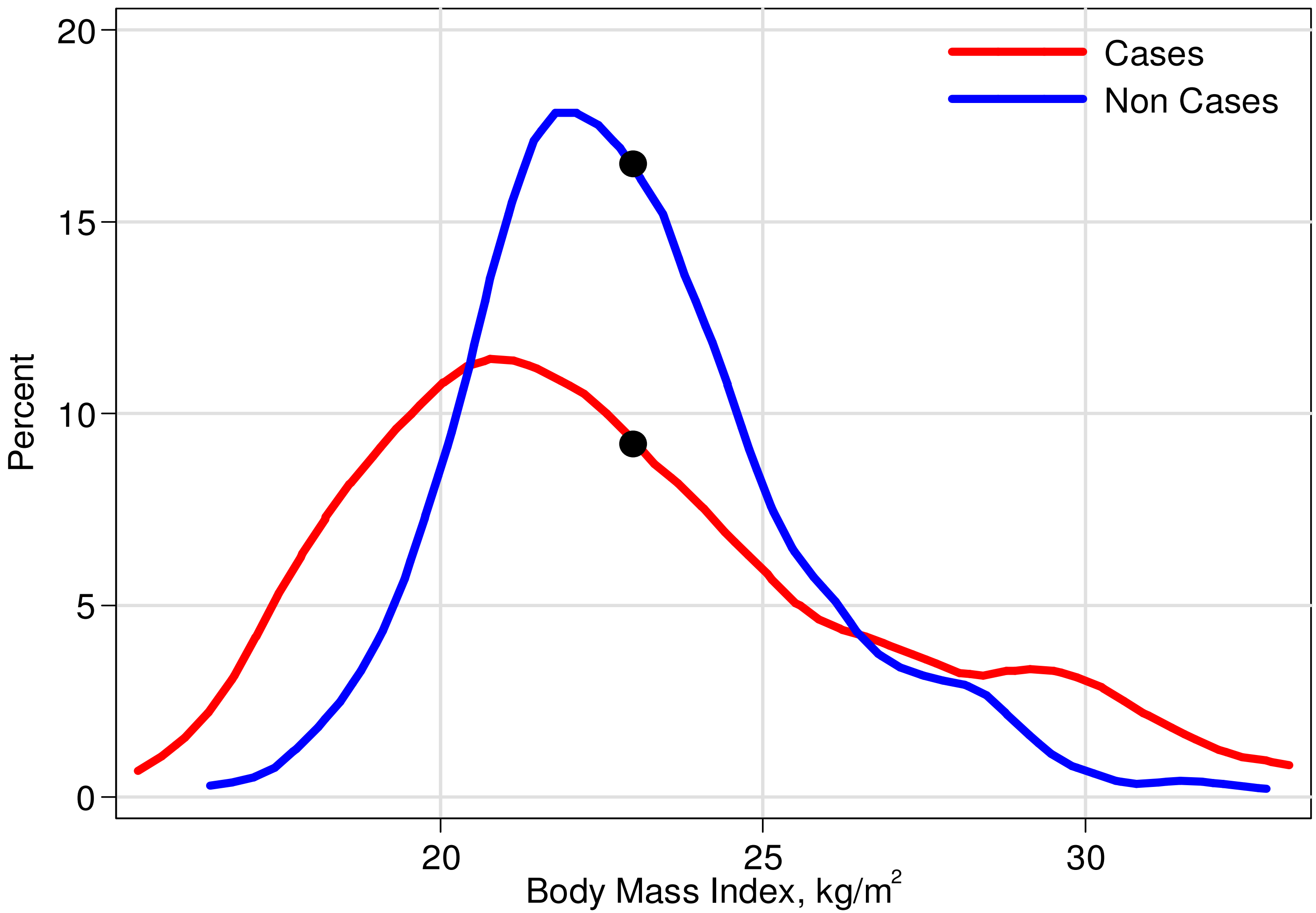
N ENGL J MED 352;15 WWW.NEJM.ORG APRIL 14, 2005

Table 2. **Univariate and Multivariate Predictors of Hyponatremia.***

Variable	Univariate Predictors		
	Hyponatremia (N=62)	No Hyponatremia (N=426)	P Value†
Demographic characteristics			
Body-mass index	22.8±3.7	23.0±2.5	0.68
Category of body-mass index			0.01
<20 (%)	25	8	—
20–25 (%)	54	73	—
>25 (%)	21	19	—

† For the univariate analysis, all continuous variables were analyzed with the use of t-tests.

Histogram



Cases
Non Cases
Percent
20
15
10
5
0
20
25
30
Body Mass Index, kg/m2

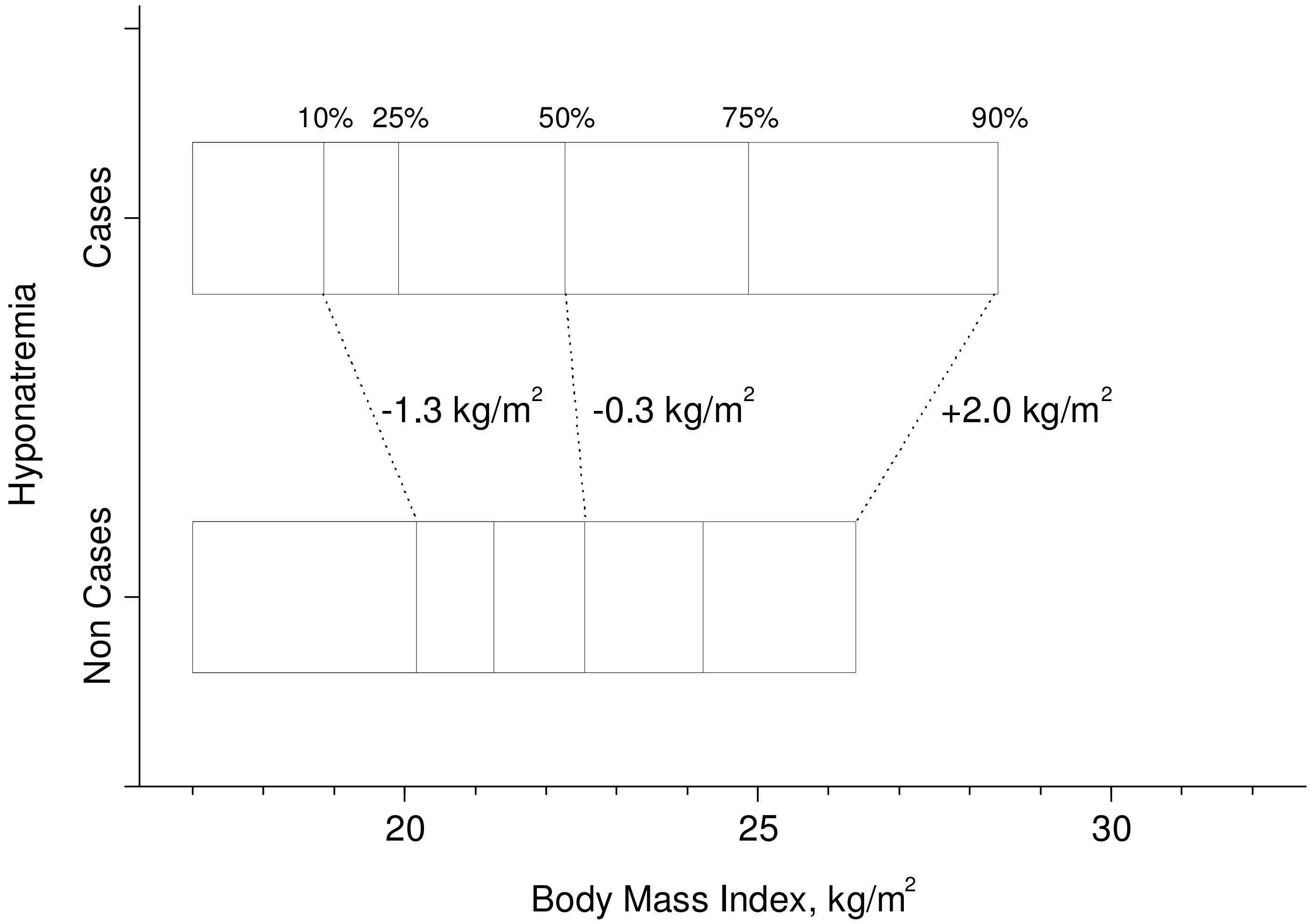
```
. sqreg bmi nas135, reps(500)    // 8 seconds

Simultaneous quantile regression                   Number of obs =       465
  bootstrap(500) SEs                               .50 Pseudo R2 =    0.0014

------------------------------------------------------------------------------
             |              Bootstrap
         bmi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
q50          |
      nas135 |      -0.27        0.61    -0.44   0.660        -1.48        0.94
       _cons |      22.56        0.13   173.33   0.000        22.30       22.82
------------------------------------------------------------------------------

. laplace bmi nas135, reps(500) // 2 seconds

Laplace regression                                 Number of obs =       465
  Optimization: Gradient Search
  bootstrap(500) SEs                         q50 Log likelihood =     0.6640
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
         bmi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
q50          |
      nas135 |      -0.27        0.60    -0.46   0.648        -1.45        0.91
       _cons |      22.56        0.12   188.38   0.000        22.32       22.80
------------------------------------------------------------------------------
```

```
. laplace bmi nas135, q(10 50 90)

Simultaneous Laplace regression                     Number of obs =        465
  Optimization: Gradient Search
  bootstrap(500) SEs                              q10 Log likelihood =     0.2533
                                                  q50 Log likelihood =     0.6640
                                                  q90 Log likelihood =     0.3569
------------------------------------------------------------------------------
             |  Observed   Bootstrap                         Normal-based
         bmi |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
q10          |
      nas135 |     -1.33        0.42   -3.15   0.002        -2.15       -0.50
       _cons |     20.18        0.15  130.82   0.000        19.88       20.48
-------------+----------------------------------------------------------------
q50          |
      nas135 |     -0.27        0.60   -0.46   0.645        -1.44        0.89
       _cons |     22.56        0.12  182.96   0.000        22.32       22.80
-------------+----------------------------------------------------------------
q90          |
      nas135 |      2.02        0.98    2.05   0.040         0.09        3.95
       _cons |     26.39        0.34   77.32   0.000        25.72       27.06
------------------------------------------------------------------------------
```

Tabulate percentiles

BMI, kg/m²	Hyponatremia (N=62)	No Hyponatremia (N=426)	Percentile Difference (95% CI)	*P*-value
10th Percentile	18.9	20.2	-1.3 (-2.2, -0.5)	0.005
Median	22.3	22.6	-0.3 (-1.4, 0.9)	0.645
90th Percentile	28.4	26.4	2.0 (0.1, 4.0)	0.040

Plot percentiles



Other features

- **`laplace`** can model heteroschedasticity (error term dependent on covariates) using the option **`sigma(`***`varlist`***`)`**

- **`laplace`** can take on any kind of weights

Summary

- **`laplace`** is a parametric model that estimates quantiles of a continuous response variable conditionally on covariates

- **`laplace`** is another way of analyzing censored data

References

Bottai, M. and Zhang, J. (2010), Laplace regression with censored data. *Biometrical Journal*, 52: 487–503.

Orsini N., Bottai, M. (2011) Laplace regression. *Stata Journal*, in preparation.