Back-Transformation in Nomograms: A Note for 432 Class 14

Thomas E. Love

2020-03-03

```
library(here); library(janitor); library(rms)
library(tidyverse)
```

A Sample Data Set

Consider the `lind.Rds` data available on our site, which was also used in Quiz 1.

The `lind` data describe 970 of the subjects in an observational study of adults receiving an initial Percutaneous Coronary Intervention (PCI) at Ohio Heart Health, Christ Hospital, Cincinnati in 1997 and followed for at least 6 months by the staff of the Lindner Center. The patients thought to be more severely diseased were assigned to treatment with abciximab (an expensive, high-molecular-weight IIb/IIIa cascade blocker); while the rest of the patients received usual-care-alone with their initial PCI. Data elements we'll use today are:

Variable	Description
`ptid`	subject ID (assigned by Dr. Love for this Quiz)
`cardbill`	Cardiac related costs incurred within 6 months of patient's initial PCI; numeric value in 1998 dollars
`abcix`	Treatment indicator: 0 means usual PCI care alone; 1 means usual PCI care augmented by treatment with abciximab.
`stent`	Coronary stent deployment, with 1 meaning YES and 0 meaning NO.
`acutemi`	Acute myocardial infarction in the previous 7 days, with 1 meaning YES and 0 meaning NO.
`ejecfrac`	Left ventricular ejection fraction; numeric value from 0 percent to 90 percent.
`ves1proc`	Number of vessels involved in the patient's initial PCI procedure; integer from 0 to 5.
`diabetic`	Diabetes mellitus diagnosis, with 1 meaning YES and 0 meaning NO.

```
lind <- readRDS(here("data/lind.Rds"))

summary(lind)
```

```
     ptid              cardbill          abcix            stent      
 Length:970         Min.   :  2216   Min.   :0.0000   Min.   :0.0000  
 Class :character   1st Qu.: 10172   1st Qu.:0.0000   1st Qu.:0.0000  
 Mode  :character   Median : 12395   Median :1.0000   Median :1.0000  
                    Mean   : 15496   Mean   :0.7082   Mean   :0.6691  
                    3rd Qu.: 16597   3rd Qu.:1.0000   3rd Qu.:1.0000  
                    Max.   :178534   Max.   :1.0000   Max.   :1.0000  
    acutemi          ejecfrac        ves1proc        diabetic      
 Min.   :0.0000   Min.   : 0.00   Min.   :0.000   Min.   :0.0000  
```

```
 1st Qu.:0.0000   1st Qu.:45.00   1st Qu.:1.000   1st Qu.:0.0000
 Median :0.0000   Median :55.00   Median :1.000   Median :0.0000
 Mean   :0.1412   Mean   :51.18   Mean   :1.385   Mean   :0.2186
 3rd Qu.:0.0000   3rd Qu.:56.00   3rd Qu.:2.000   3rd Qu.:0.0000
 Max.   :1.0000   Max.   :90.00   Max.   :5.000   Max.   :1.0000
```

Model 1: A Linear Regression on a Transformed Outcome

Suppose we plan to fit a model for `cardbill` using `stent`, `ejecfrac` and `ves1proc`.

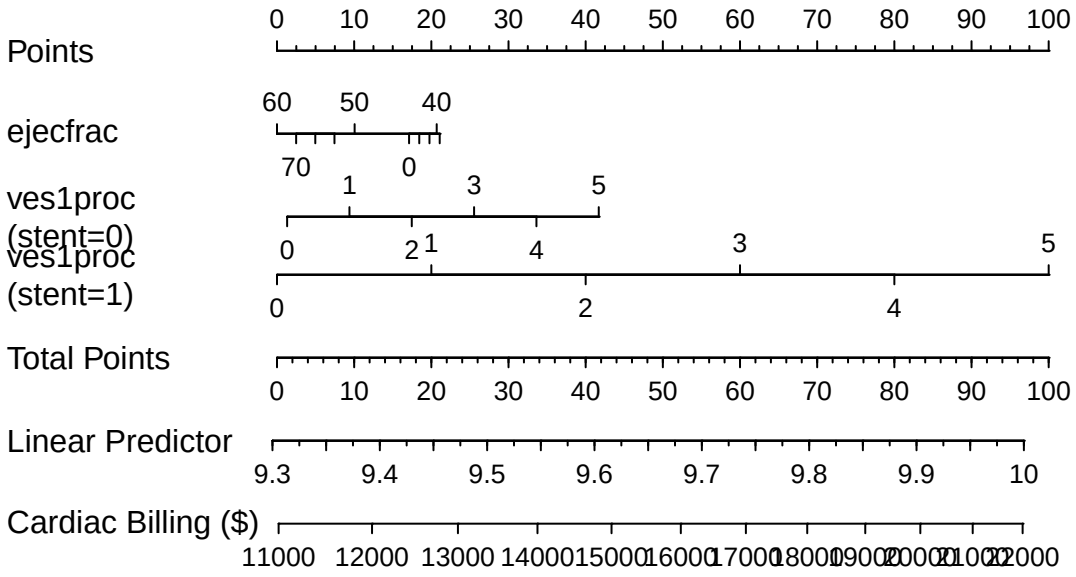
The purpose of this little example is to suggest strategies for building a nomogram for the models, under several different potential strategies for transforming the outcome. First, I'll create the data set to include several potential transformations of the `cardbill` data. The transformations we're considering are the the natural log after 1 has been added to `cardbill` (which would be appropriate if there were any subjects with `cardbill` = 0, since all the values in the outcome need to be positive to use our available power transformations), the base-10 log, the square root and the inverse.

```
lind1 <- lind %>%
  mutate(log_costs = log(cardbill + 1),
         log10_costs = log10(cardbill),
         sqrt_costs = sqrt(cardbill),
         inv_costs = 1/cardbill)
```

The model for the natural log of costs

Here's the nomogram showing costs in dollars as the last line.

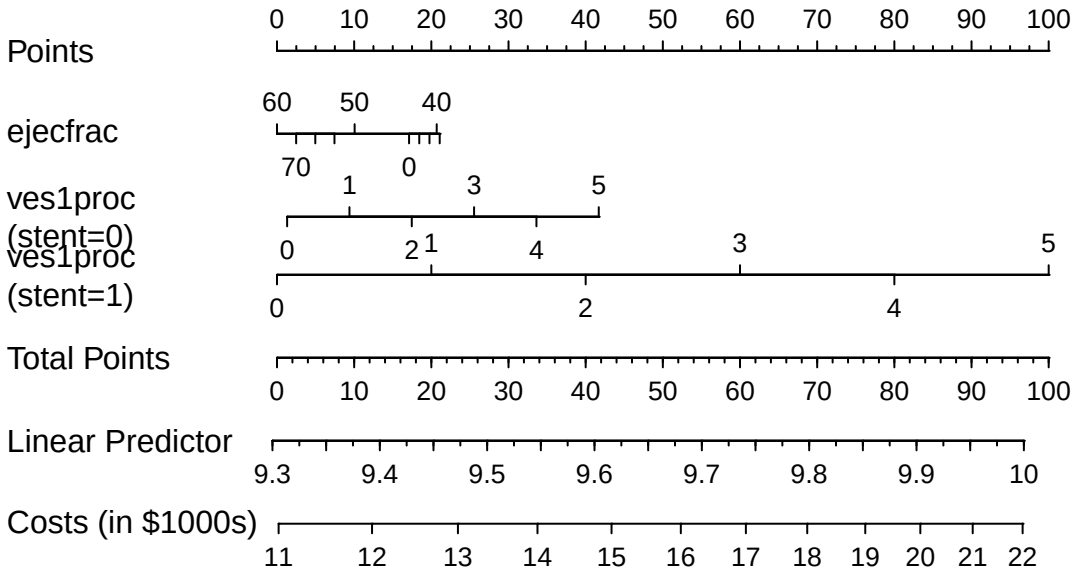
```
d <- datadist(lind1)
options(datadist = "d")

m1a <- ols(log_costs ~ ves1proc * stent + rcs(ejecfrac, 4),
           data = lind1, x = TRUE, y = TRUE)

plot(nomogram(m1a,
              fun = list(function(x) exp(x) - 1),
              funlabel = "Cardiac Billing ($)"))
```



That doesn't look so good.

An alternative would be to divide the costs by 1000 at the end and show the nomogram with costs in $1000s.

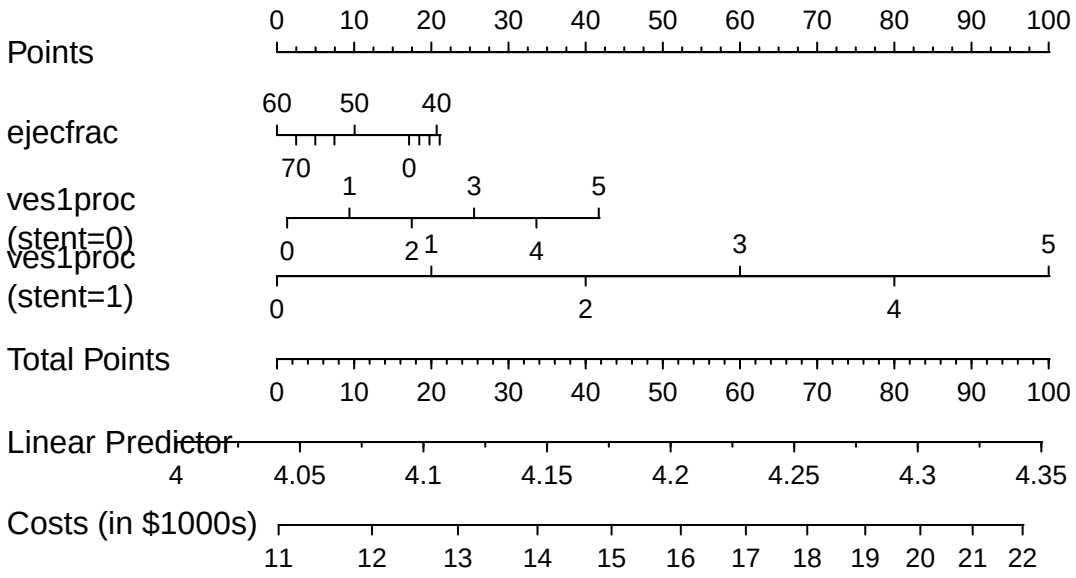
```
plot(nomogram(m1a,
              fun = list(function(x) (exp(x) - 1)/1000),
              funlabel = "Costs (in $1000s)"))
```



The model for the base-10 log of costs

Here's the nomogram showing costs in thousands of dollars as the last line.

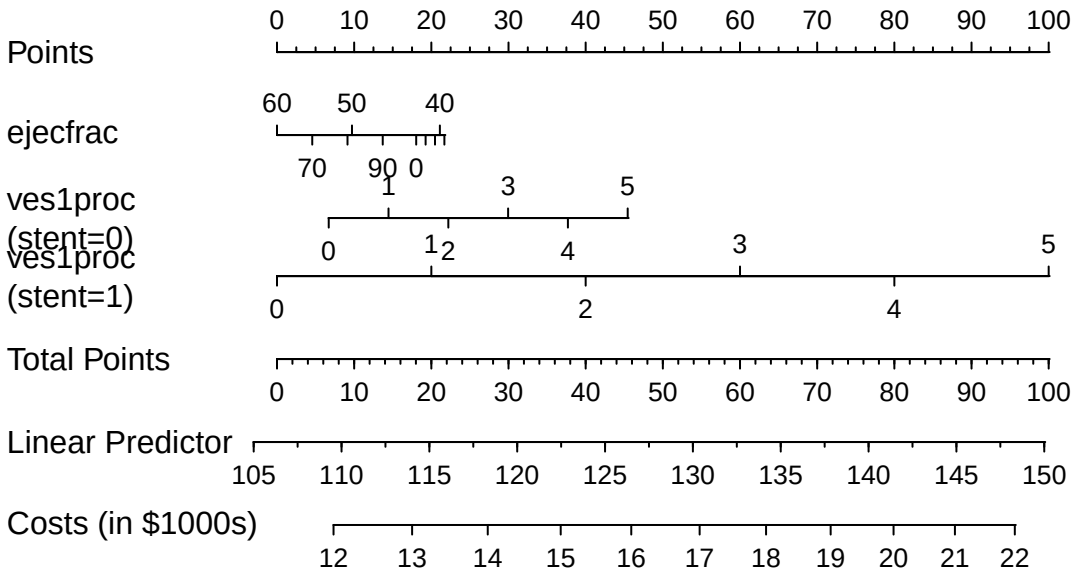
```
d <- datadist(lind1)
options(datadist = "d")

m1b <- ols(log10_costs ~ ves1proc * stent + rcs(ejecfrac, 4),
           data = lind1, x = TRUE, y = TRUE)

plot(nomogram(m1b,
              fun = list(function(x) (10^x / 1000)),
              funlabel = "Costs (in $1000s)"))
```



The model for the square root of costs

Here's the nomogram showing costs in thousands of dollars as the last line.

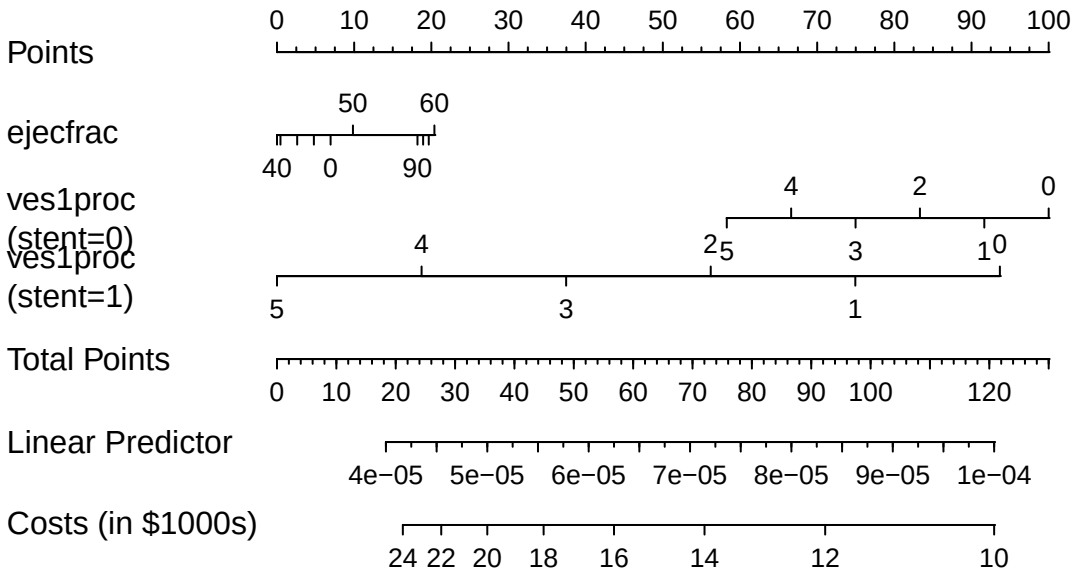
```
d <- datadist(lind1)
options(datadist = "d")

m1c <- ols(sqrt_costs ~ ves1proc * stent + rcs(ejecfrac, 4),
           data = lind1, x = TRUE, y = TRUE)

plot(nomogram(m1c,
              fun = list(function(x) x^2 / 1000),
              funlabel = "Costs (in $1000s)"))
```



The model for the inverse of costs

Here's the nomogram showing costs in thousands of dollars as the last line. Note the change in direction.

```
d <- datadist(lind1)
options(datadist = "d")

m1d <- ols(inv_costs ~ ves1proc * stent + rcs(ejecfrac, 4),
           data = lind1, x = TRUE, y = TRUE)

plot(nomogram(m1d,
              fun = list(function(x) (1/x) / 1000),
              funlabel = "Costs (in $1000s)"))
```



A Logistic Regression Model

Let's predict `abcix` status based on the same predictors. First, we make sure that `abcix` is numeric and takes the values 1 and 0.

```
lind %>% tabyl(abcix)
```

```
 abcix   n   percent
     0 283 0.2917526
     1 687 0.7082474
```

```
d <- datadist(lind1)
options(datadist = "d")

m2a <- lrm(abcix ~ ves1proc * stent + rcs(ejecfrac, 4),
           data = lind1, x = TRUE, y = TRUE)

plot(nomogram(m2a,
              fun = plogis,
              funlabel = "Prob(abcix = 1)"))
```