STA304

Neil Montgomery
2016-03-14

ratio, regression, difference estimation

recap

If one has a sample $\{(y_1, x_1), \ldots, (y_n, x_n)\}$ the following may be of interest:

1. To estimate the population ratio $R = \tau_y \big/ \tau_x$ using $\hat{R} = r = \bar{y} \big/ \bar{x}$.
2. To enable estimation of τ_y when N is unknown.
3. To enable improved estimation of τ_y or μ_y when y and x are correlated.

Formula summary:

r	$\hat{\tau}_y$	$\hat{\mu}_y$	$\hat{V}(r)$
$\bar{y} \big/ \bar{x}$	$r\tau_x$	$r\mu_x$	s_r^2

The error bounds are based on this new object:

$$s_r^2 = \frac{\sum_{i=1}^{n} (y_i - r x_i)^2}{n - 1}$$

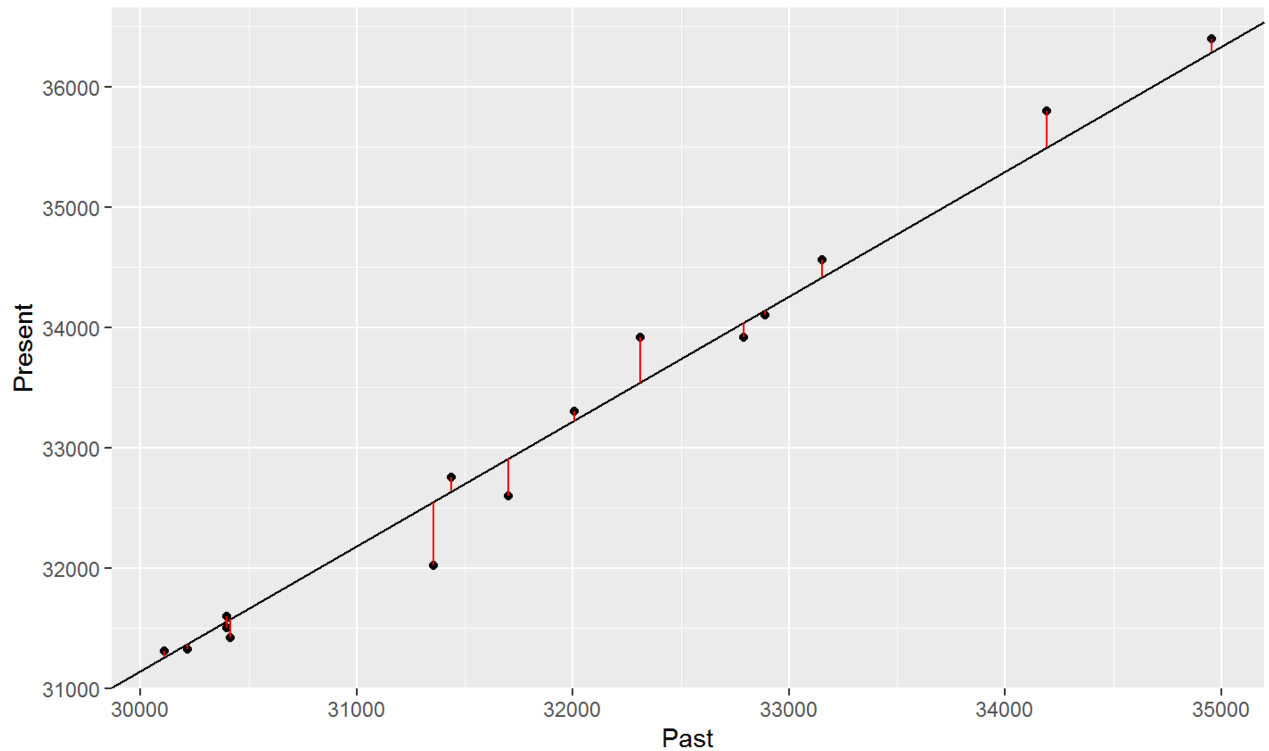
$r_i = y_i - rx_i$ is a "residual"

Reconsider the teacher salary example with $r = 1.0380283$. First 5 lines shown:

```
##   Teacher  Past Present Predicted = r*Past        r_i
## 1       1 30400   31500           31556.06  -56.06028
## 2       2 31700   32600           32905.50 -305.49706
## 3       3 32792   33920           34039.02 -119.02397
## 4       4 34956   36400           36285.32  114.68280
## 5       5 31355   32020           32547.38 -527.37730
```

Think of rx_i as the "predicted" value for y_i. Then $y_i - rx_i$ is in some sense a prediction error, or "residual". And s_r^2 is just the "average" of the squared residuals.

"residuals" plotted



example of improved estimation

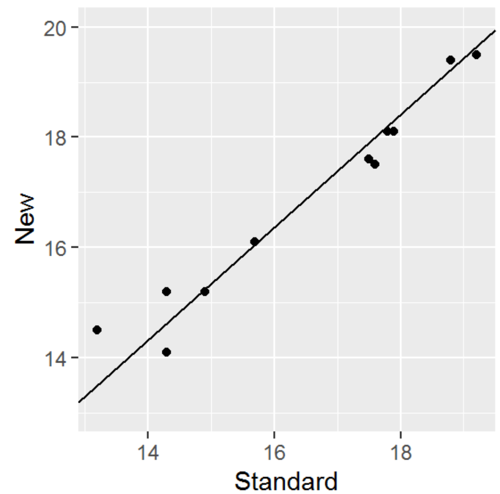
We have estimated a ratio and estimated τ_y using ratio techniques because there was no other option. In this example we'll estimate μ_y using ratio techniques simply to take advantage of the information contained in the x variable.

Consider question 6.6. "Rats doing mazes while on drugs". They have $N = 763 who completed the maze on the standard drug in an average of $\mu_x = 17.2$ seconds.

A random sample of 11 rats are given a new drug. Their old times x_i were known from before and they complete the maze while on the new drug in time y_i .

The task is to estimate the average maze time μ_y for the new drug.

these are your rats on drugs



The estimated ratio is $r = 1.0226269$. The mean estimate is$$_y =

$s_r^2 = 0.2049424$.