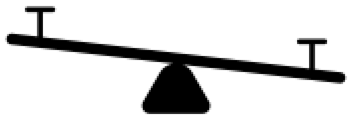
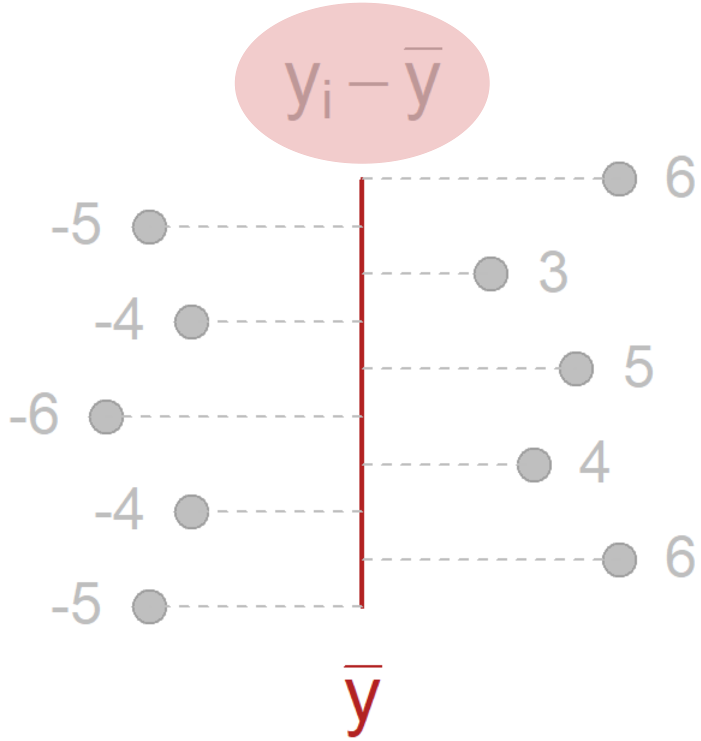
DISTRIBUTION OF THE SAMPLE

vs

THE SAMPLING DISTRIBUTION

$$var(y) = \frac{\sum (y_i - \bar{y})^2}{(n-1)}$$



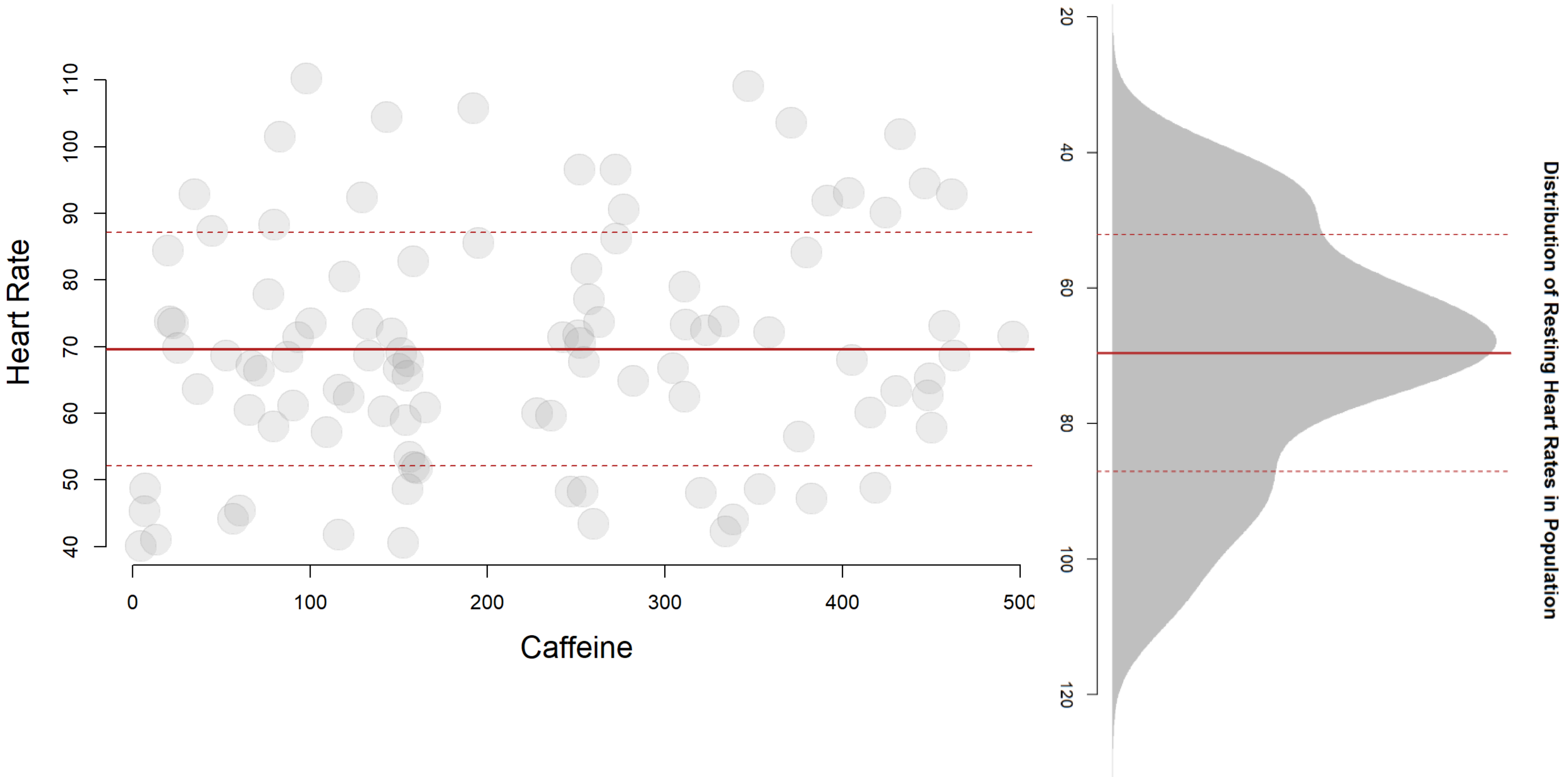


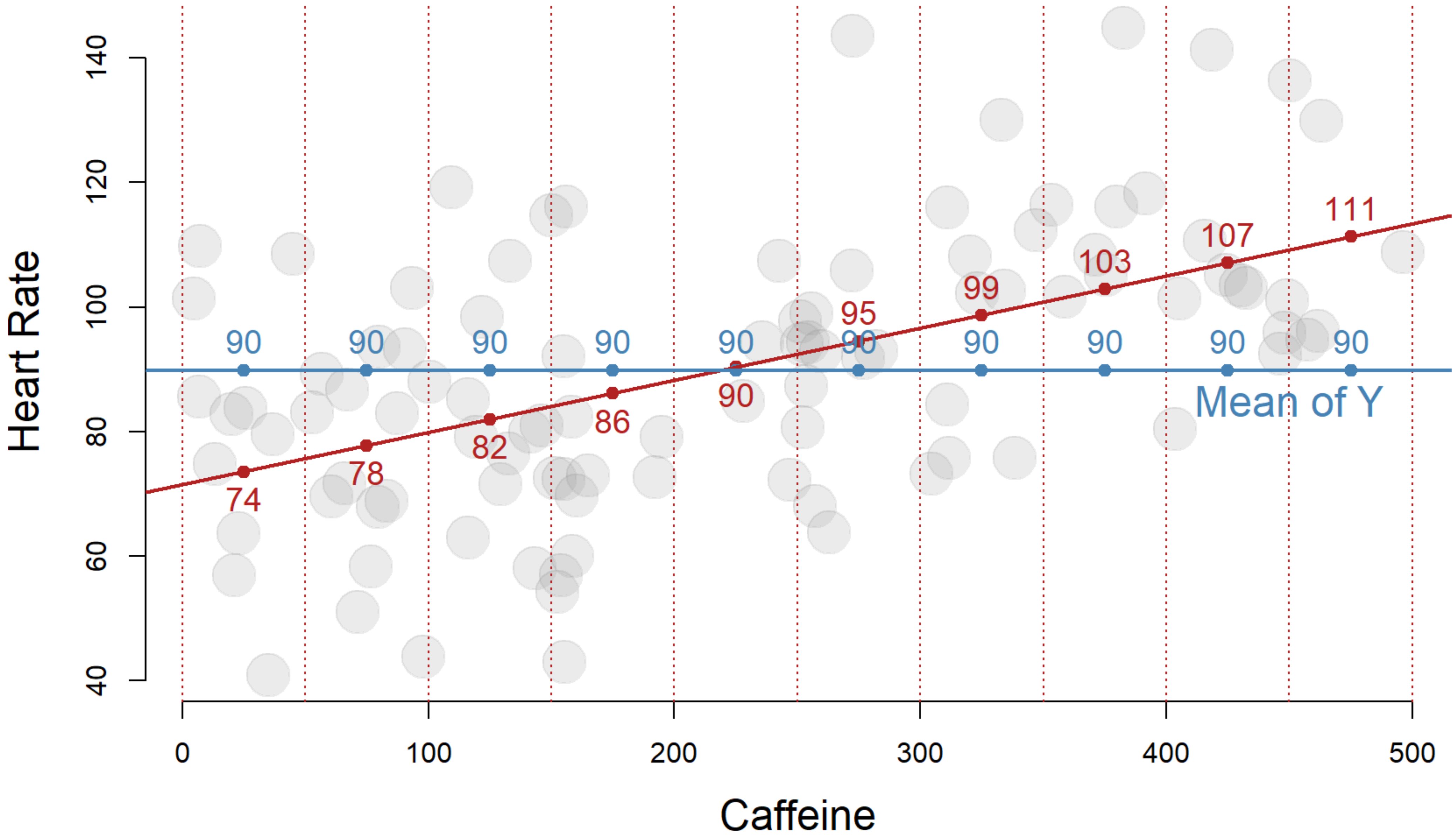
"DEVIATIONS"
are the centerpiece of every variance and the currency of regression models

https://watts-college.github.io//paf-510-template/lectures/caffeine.html

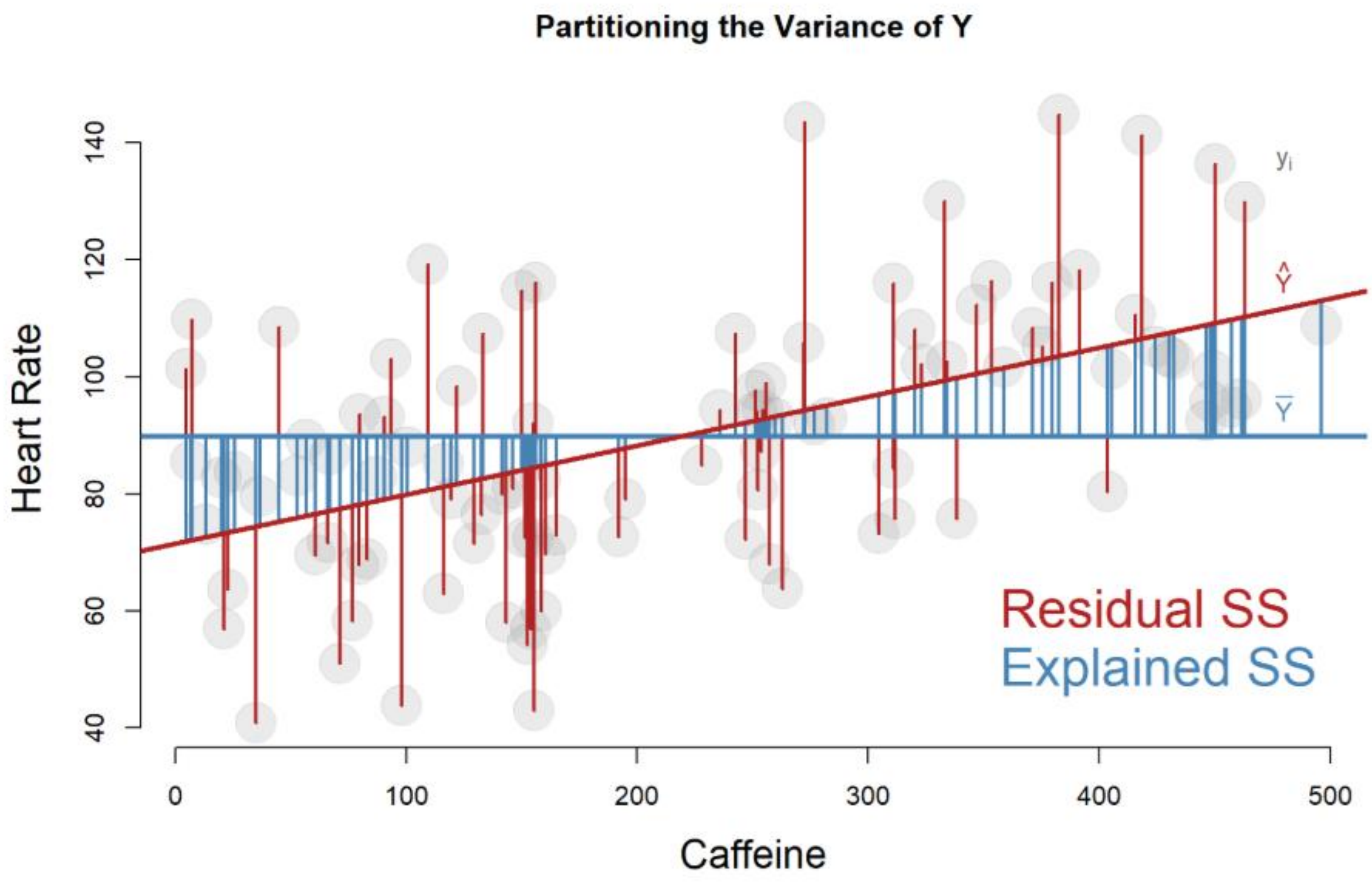
THIS IS SAMPLE VARIANCE

THIS IS SAMPLE VARIANCE



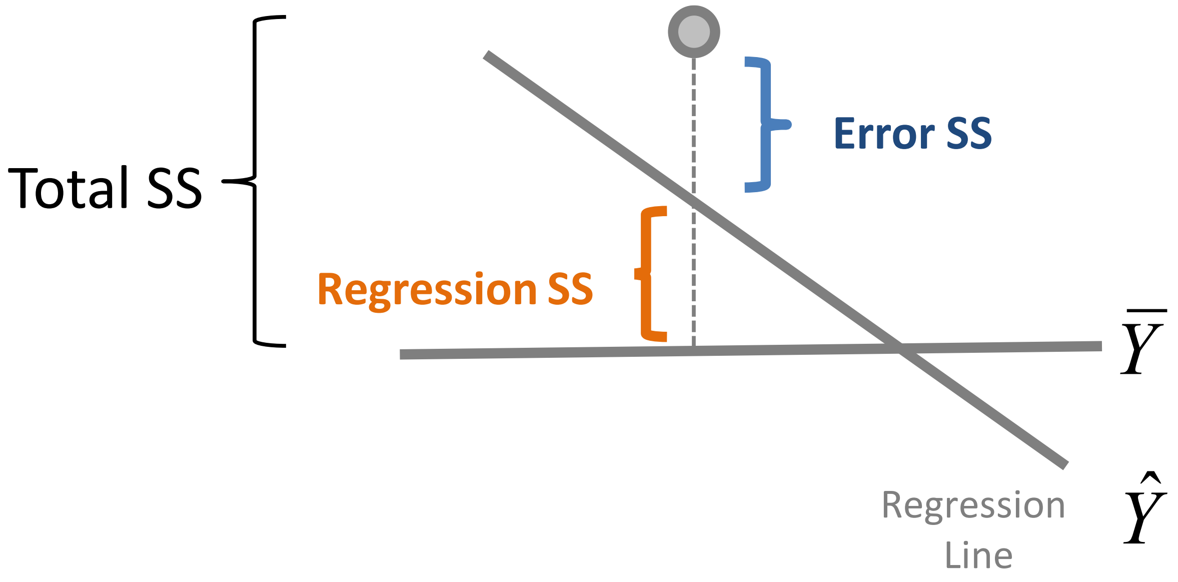


Heart Rate
Caffeine
Mean of Y
140
120
100
80
60
40
0
100
200
300
400
500
90
90
90
90
90
90
90
90
90
90
74
78
82
86
90
95
99
103
107
111



Partitioning the Variance of Y
Heart Rate
Caffeine
0
100
200
300
400
500
40
60
80
100
120
140
y_i
$\hat{Y}$
$\bar{Y}$
Residual SS
Explained SS

PARTITIONING THE VARIANCE OF Y



TSS becomes the variance of Y

$$TotalSS = \sum (y_i - \bar{y})^2$$

$$RegressionSS = \sum (\hat{y}_i - \bar{y})^2$$

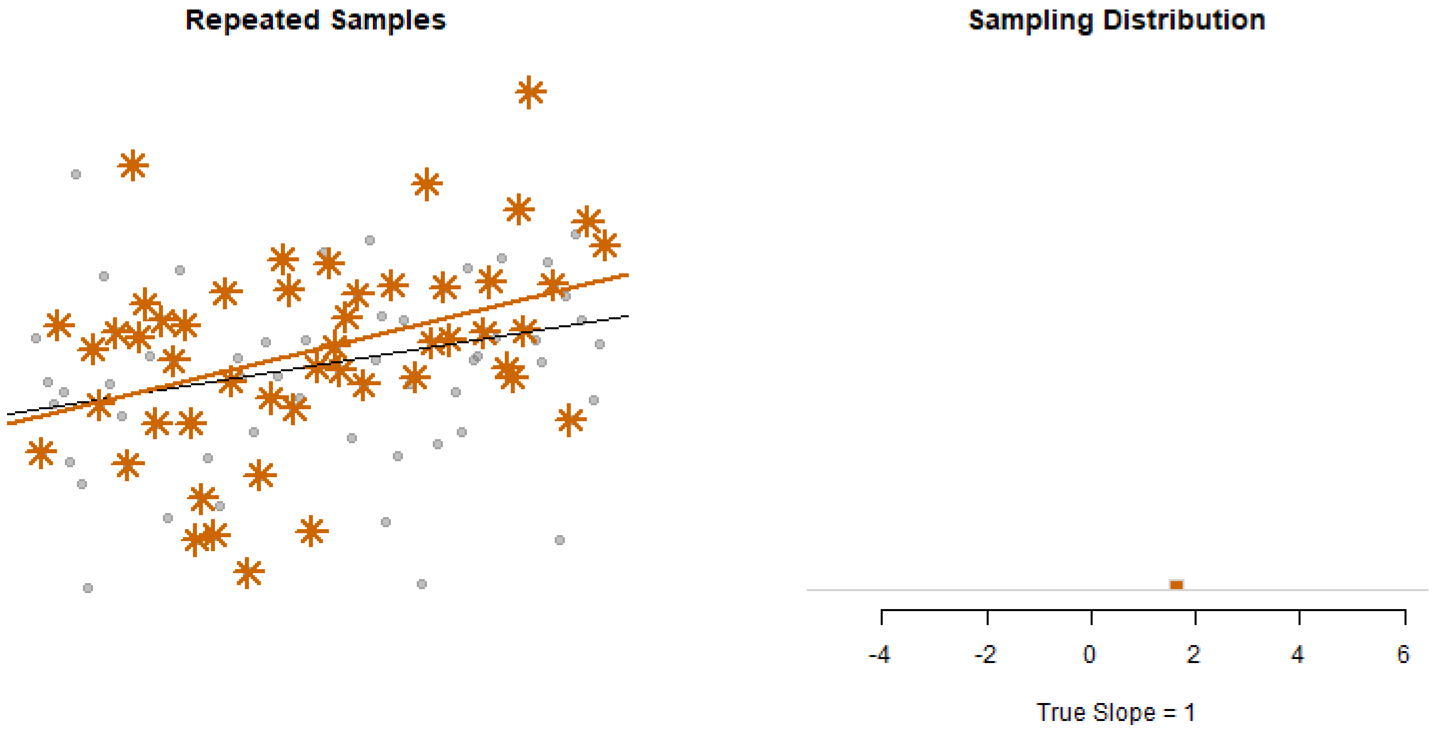
$$ErrorSS = \sum (y_i - \hat{y}_i)^2$$

$$TSS = RSS + ESS$$

ESS becomes the variance of the slope

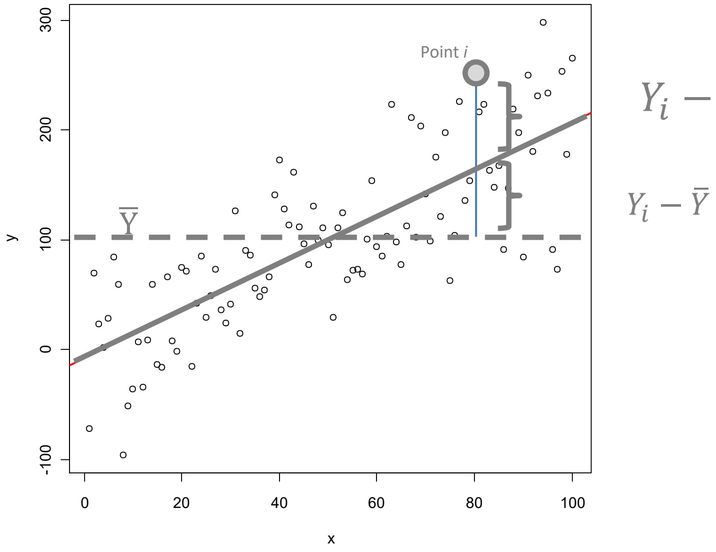
THE CENTRAL LIMIT THEOREM

CENTRAL LIMIT THEOREM:
THE SAMPLE STATISTIC (THE SLOPE) HAS A PREDICTABLE VARIANCE (THE STANDARD ERROR)



https://github.com/lecy/regression-simulations

STANDARD ERROR OF THE SLOPE



$$SSE = \sum e_i^2$$

Sum of Squared Error Terms

$$\hat{\sigma}_{\varepsilon}^2 = \frac{SSE}{n-2}$$

Variance of the residual

$$SE_{b_1} = \sqrt{\frac{\hat{\sigma}_{\varepsilon}^2}{\sum (x_i - \bar{x})^2}}$$

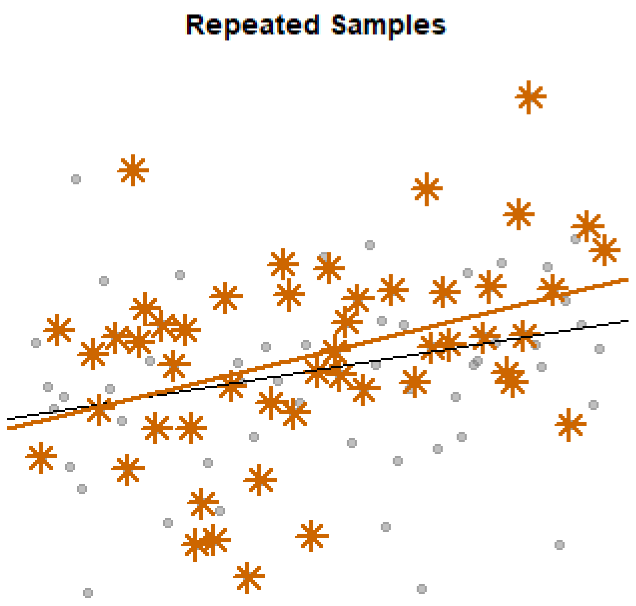
Standard error of the slope

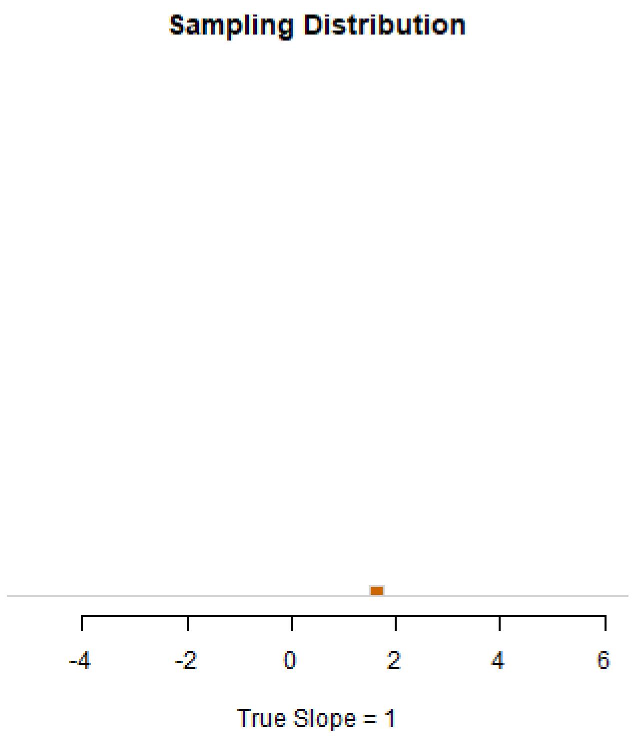
Standard error: on average, how far off is our estimate (best guess) of the slope from the true slope?

The standard error of the slope is one of the most important concepts in regression because it determines the size of the confidence interval and thus the statistical significance of our study. We want standard errors to be as small as possible. We see here that the size of the standard error will be directly related to the amount of unexplained variance we have in our model, the residual *e*. The important thing to note is the unexplained portion shows up in the numerator of the standard error. As a result, the size of the standard error will be proportional to the amount of unexplained variance (plus a couple of other considerations to be covered later).

$$\textbf{Standard Error of the Slope} \approx \frac{\text{residual}}{\text{sample size} \cdot \text{variance X}}$$

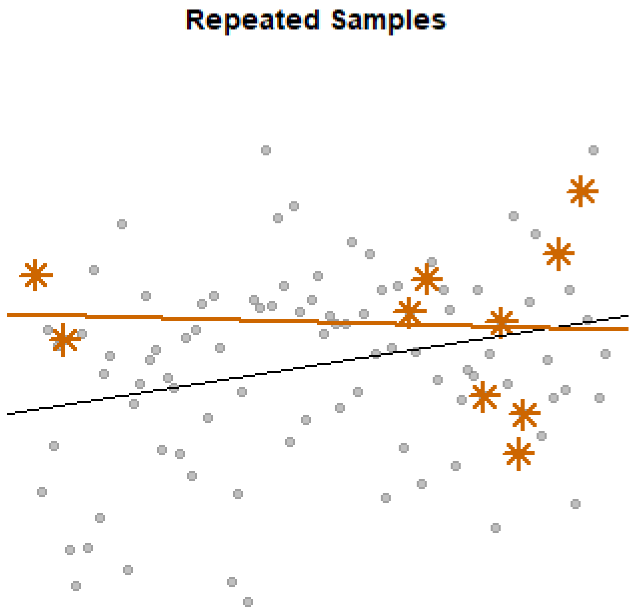
THIS IS SAMPLING VARIANCE

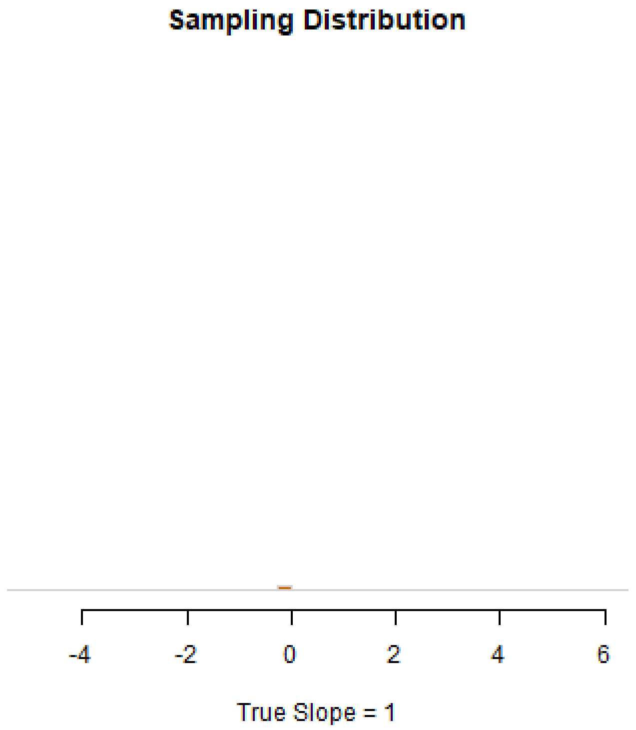




https://github.com/lecy/regression-simulations

THIS IS SAMPLING VARIANCE

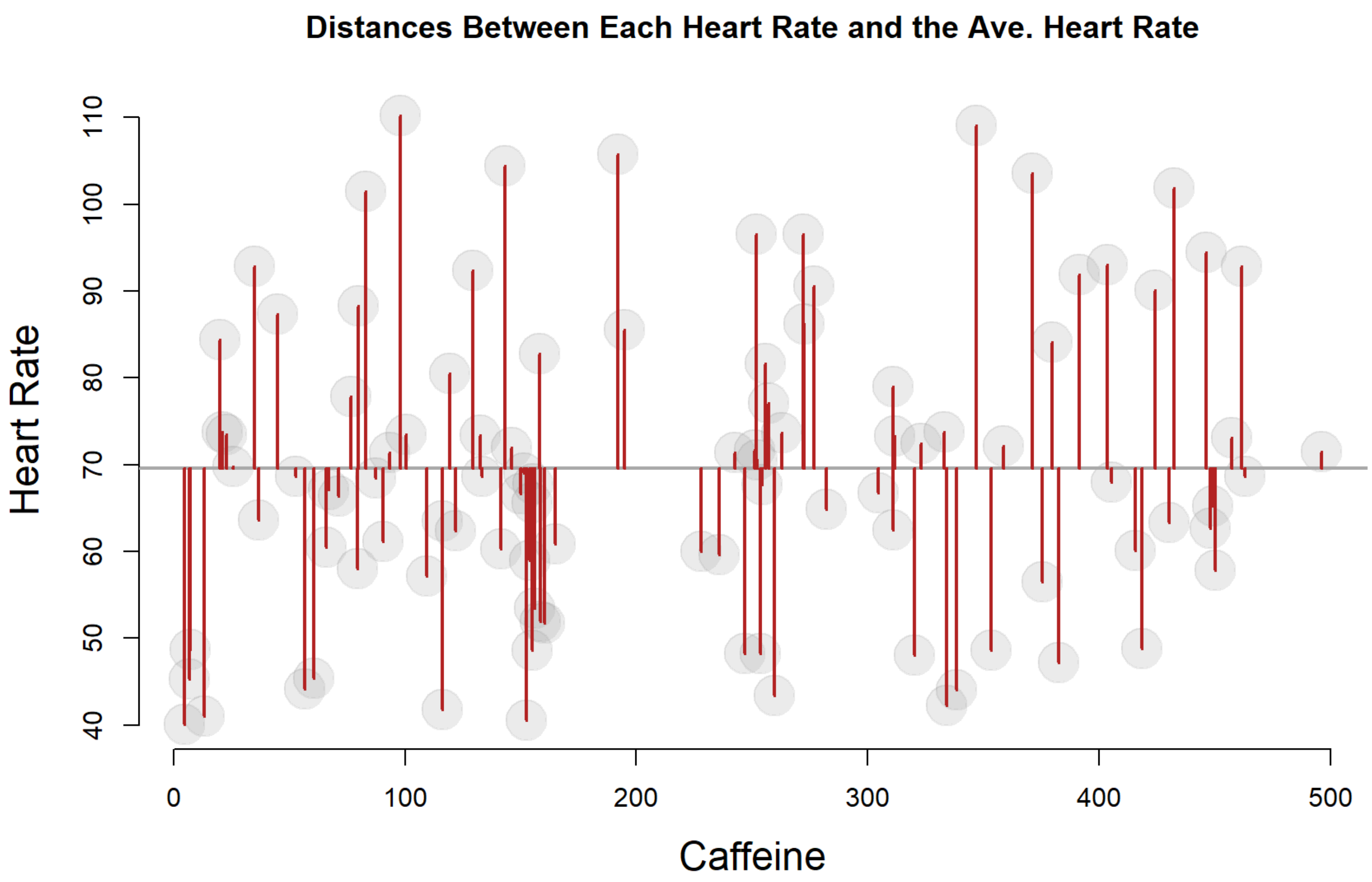




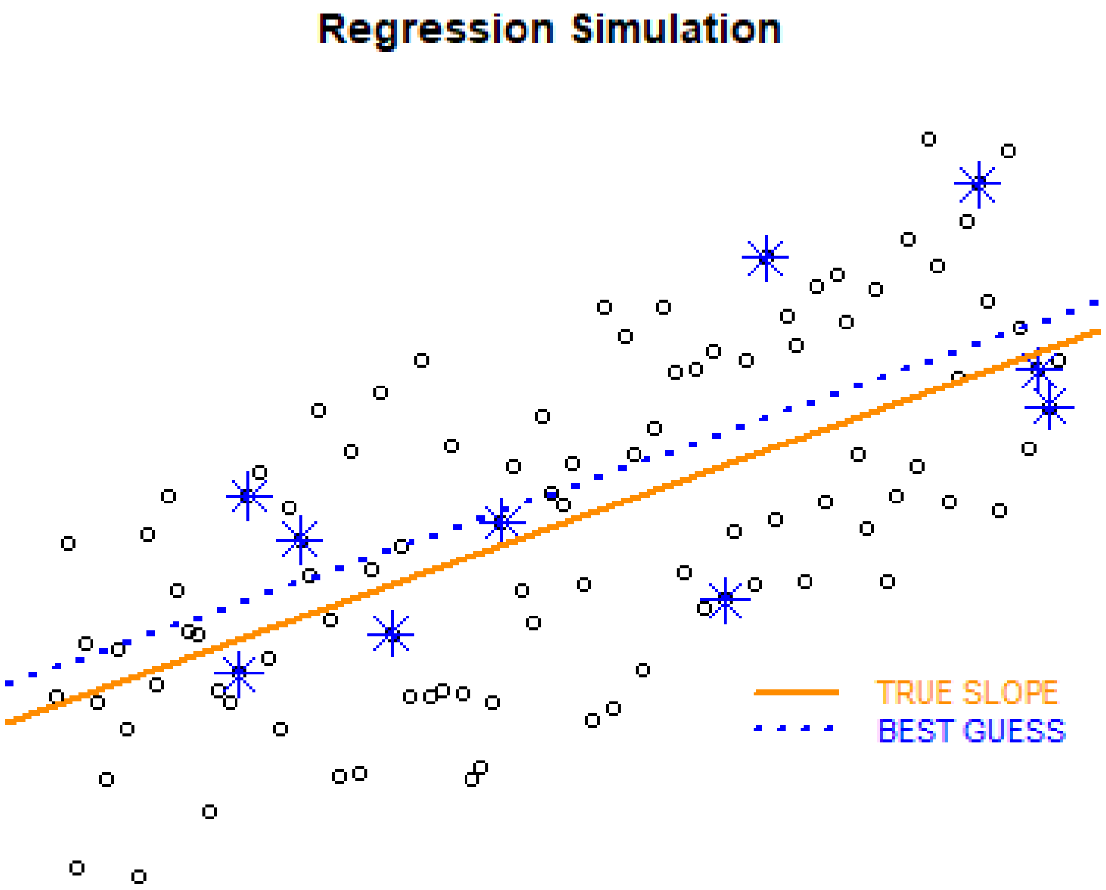
https://github.com/lecy/regression-simulations

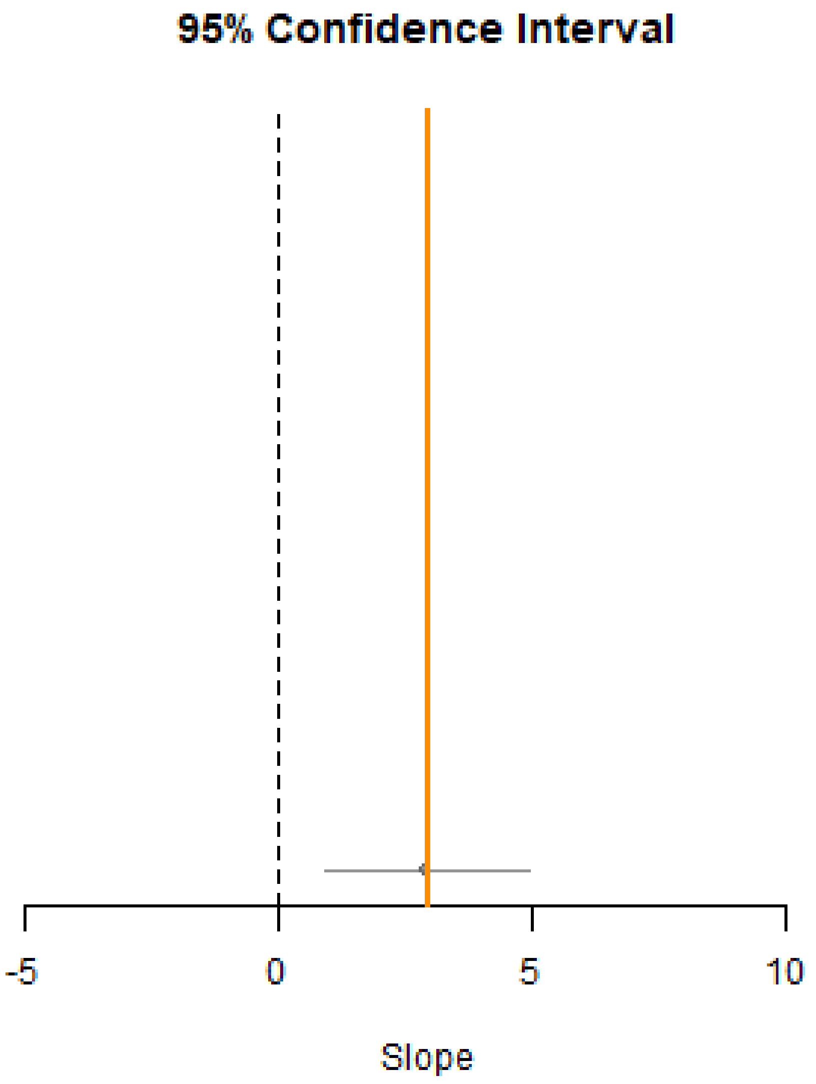
THIS IS SAMPLE VARIANCE

THIS IS SAMPLE VARIANCE

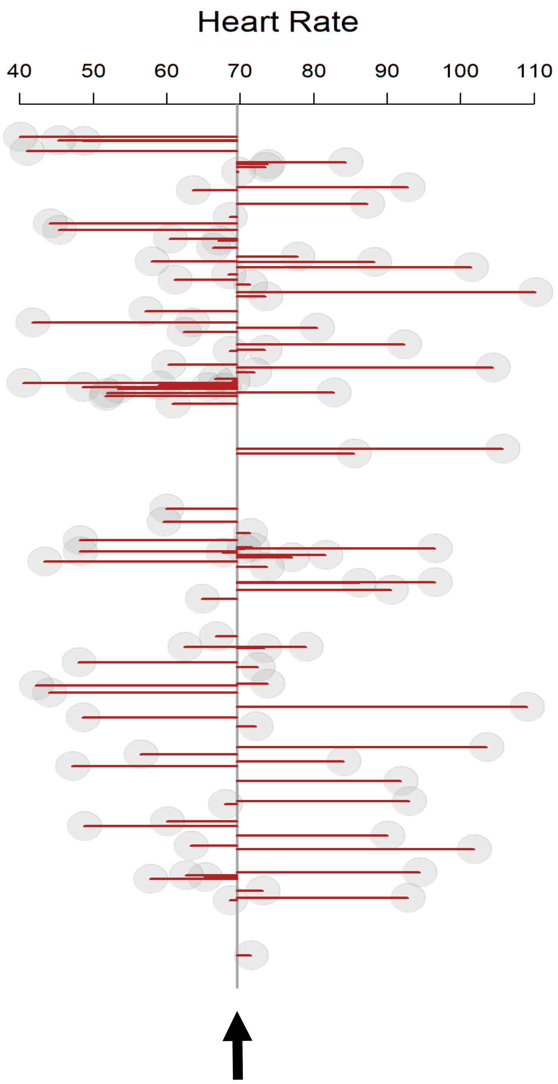


THIS IS SAMPLING VARIANCE



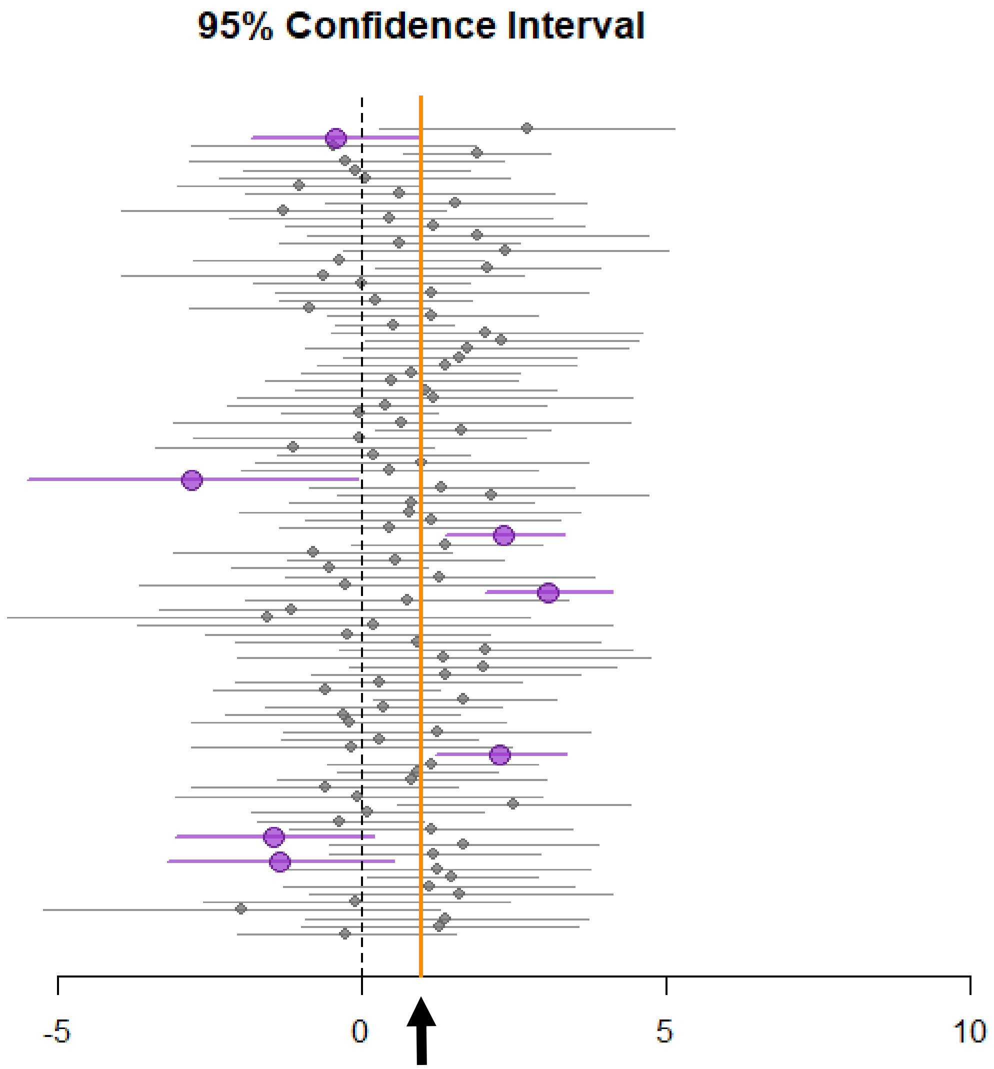


SAMPLE Distribution



Center of the Data

SAMPLING Distribution



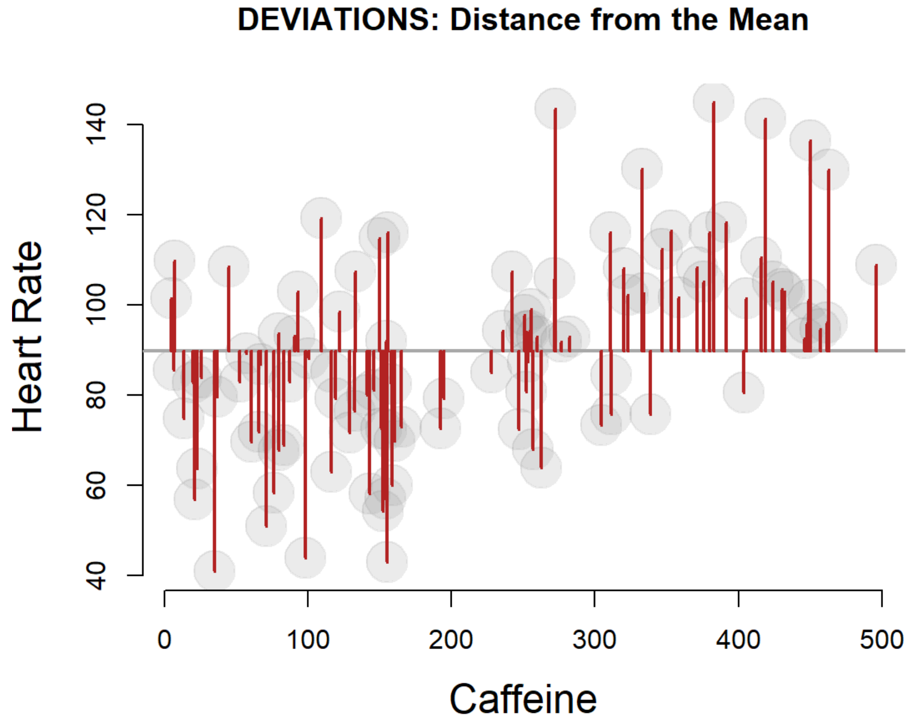
"True" Slope

variance:

Measured as Standard DEVIATIONS

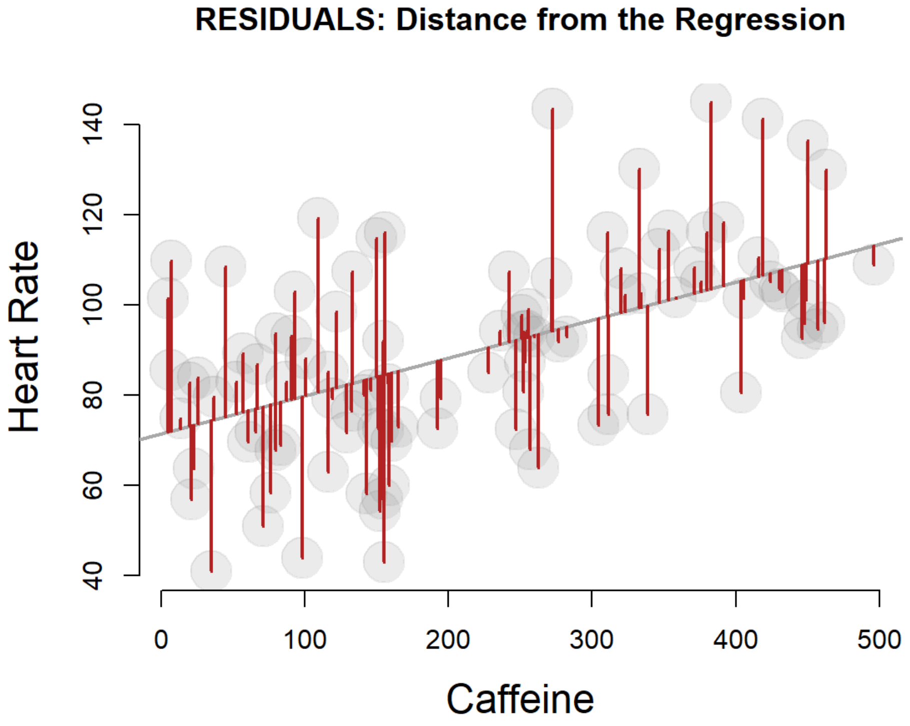
Measured as Standard ERRORS

TSS BECOMES VARIANCE OF Y



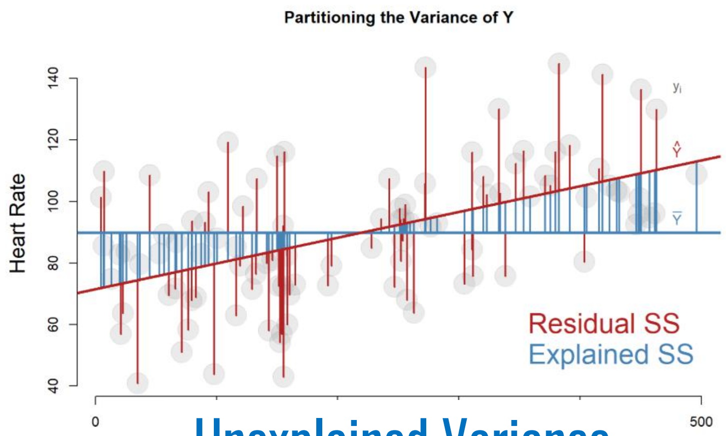
Standard deviation: on average, how far is the **MEAN** from each data point

ESS BECOMES VARIANCE OF THE SLOPE



One interpretation of the SD of the residual: on average, how far is the **CONDITIONAL MEAN** from each data point

More Importantly:



Unexplained Variance

Error SS
same as
the Residual SS

Standard
ERROR

Confidence
Interval

p-value

SAMPLING Distribution