

Practice Problems on Least Squares

1. We might model Barack Obama's vote share in 2012 as a function of the percent favorable to the Affordable Care Act. After all, the ACA was Barack Obama's signature legislative achievement. We might suppose the following model:

$$y = \alpha + \beta x + \epsilon, \tag{1}$$

where y represents Obama's vote share and x represents the percent favorable to the ACA. What parts of this model are known variables? What parts are parameters that we'd like to estimate?

2. Use the `lm()` function to estimate the unknowns using least squares. Notice that the `lm()` function has a `data` argument, so we can tell it where to find the variables and not need to use the awkward `$` operator.
3. Think really hard about what you expect the value of β to be? A one point increase in the percent favorable to the ACA should lead to about a X point increase in Obama's vote share? Should β be positive or negative? About what value?
4. Print out the coefficients using the `coef()` function. Does the estimated value of α make sense? What about the estimate of β ? Interpret $\hat{\alpha}$ and $\hat{\beta}$.
5. Print out the estimated errors (i.e., the residuals) using the `residuals()` function. These are harder to interpret—just be aware they exist and how to calculate them.
6. Produce a scatterplot of the percent favorable to the ACA and Obama's vote share. Add the least squares fit to the plot. Add a red, dashed line to the plot with an intercept of zero and a slope of one. Hint: It's `abline(a = 0, b = 1)`. Why do you think the line fit with least squares doesn't match the red, dashed line (i.e., the one I'd expect in theory).