A major issue with antiretroviral drugs is the mutation of the virus’ genes. Because of its high rate of replication (10^9 to 10^10 virus per person per day) and error-prone polymerase^1, HIV can easily develop mutations that alter susceptibility to antiretroviral drugs. The emergence of resistance to one or more antiretroviral drugs is one of the more common reasons for therapeutic failure in the treatment of HIV.

In the following paper^2, a sample of in vitro^3 HIV viruses were grown and exposed to a particular antiretroviral therapy. The susceptibility of the virus to treatment and the number of genetic mutations of each virus were recorded.

Anytime I ask for ‘test set prediction error’ for a method, use the following

\[
X_0 = \text{hiv.test}\$x \\
Y_0 = \text{hiv.test}\$y
\]

\[
\hat{Y} = \text{prediction of method on } X_0
\]

\[
\text{print(} \text{mean}((\hat{Y} - Y_0)^2) \text{)}
\]

Let’s look at comparing the lasso with CV, elastic net, refitted lasso, and scaled sparse regression.

1. Find the test set prediction error for lasso with CV. Compare the CV estimate of the risk to the test set prediction.

2. Find the test set prediction error for elastic net with

---

^1 An enzyme that ‘stitches’ back together DNA or RNA after replication

^2 The entire paper is on the website. Try to see what you can get out of it.

^3 Latin for in glass, sometimes known colloquially as a test tube
3. Find the test set prediction error for refitted lasso
4. Find the test set prediction error for SSR

(a) $\alpha = .5$
(b) $\alpha = .9$