**Reading Yifan Ma Thesis**

Talks about Predictive Mean squared Error

should i do package or data size research $ questino

**Literature review:**

Stochastic predictor variables are random variables.B

Kerridge

* “Stochastic predictor variables”
* What Is “variance of error term in the regression”
* A number and a symbol

  Description automatically generated with medium confidence
* Widely used approximation

Narula says

* A math problem with numbers and symbols

  Description automatically generated
* Talks about taking subsets of data

Sawyer

* A math equations with numbers

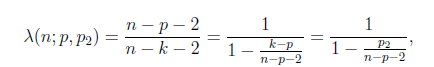
  Description automatically generated with medium confidence
* Approximate distribution of mean absolute error (MAE)
* Better than normal distribution

**Analysis:**

* Following Multivariate normal disitrubiton
* 
* A y value, and the multiple z values are multivariate?
* With the specific distributions, of mean, and covariance (multiple variance)
* <https://en.wikipedia.org/wiki/Multivariate_normal_distribution>
* What des this mean A black and white text

  Description automatically generated with medium confidence

Inflation Equation

* 

Efficient Sample size

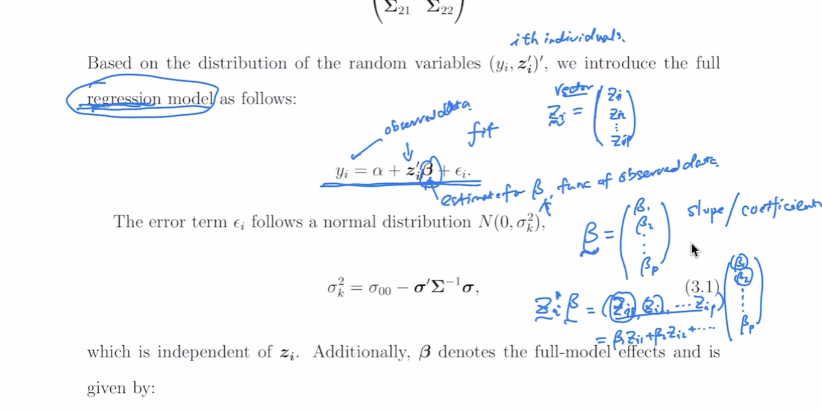
Notes:

(yi, zi)

Expectation, E() is what the difference is expectd.

Alpha and beta is used for prediction model.

Estimate for beta uses least square estimation

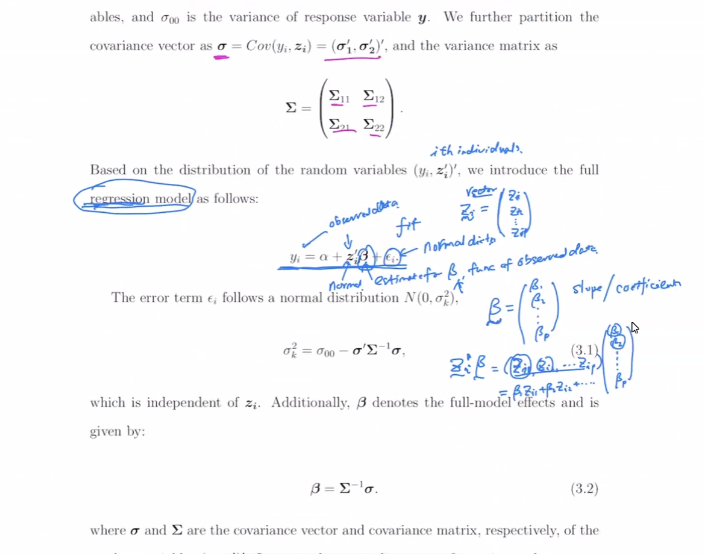


Error follows normal distribution

‘ means transpose

Bold means vector

Covariance is needed to understand how related different factors are.



Splitting up data into parts.

Inverse of matrix, is like reciprocal

Partiionaing can help with identifying differing amounts of predictors.

Understand how prediction if only know amount, not wihih

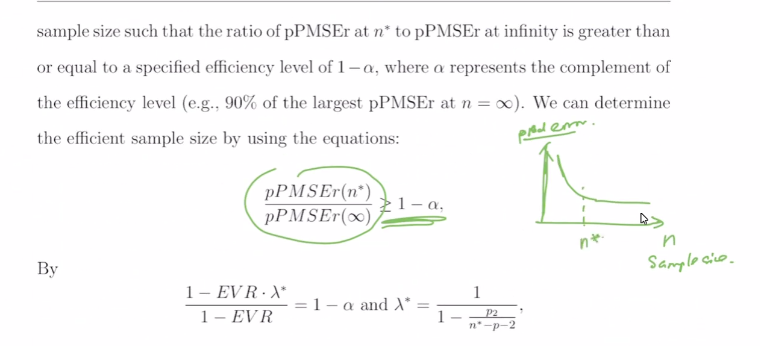
Percentage to represent prediction accuracy.

PMSEr (prediction mSE reduction) can identify relationships between nsmaller and bigger mdoels.

n-how mayn indivuaiu,s

p-how many predictors total

k-how many predictors used.



N\* is a good enough amount where 1-a.

Alpha Is some value as a big value near the ratio., how close infitie predict error to n\* predict error is.

Effect size, how deterministic. A paper with text and equations

Description automatically generated

Measure how storng predictors are.

Can replace regression model with any model, like neural network.

Check sample\_size\_predi\_simulation.rmd file

Check yifanma code.

Check based on simulation, what results to get.

Get the idea and then look into code.

Get a clear view into the idea.

PMSE improvement document.

Summaries made on work.??? (no need)

Focus on computing part.