# Pre-requisites

* + is invertible
  + , ( is full rank) then is injective:
* Linearity of : , ,
* Covariance:
* ,
* Transposition: , ,
  + Symmetric invertible matrix is symmetric.
  + is positive symmetric (symmetric with positive eigenvalues).
* Dot product: , , ,
* Gradient: , in general, if is symmetric.
* Trace of a matrix is defined by .
  + Linearity:
  + . Hence, if is diagonalizable, the trace is the sum of the eigenvalues.
  + If is an orthogonal projector, .
* Normal distribution:
  + , independent
  + Confidence interval for with known variance:
* Chi-squared distribution:
  + ,
* T-Student distribution:
  + ,
  + Confidence interval for with unknown variance:
  + Confidence interval for the regression coefficients :
  + Confidence interval for the predicted values :
  + Confidence interval for the predicted values :
* Eigenvalues: is invertible if and only if its eigenvalues are nonzero.
  + If denotes the set of eigenvalues of , then
* Singular Value Decomposition (SVD): , orthogonal, and diagonal such that .
  + The eigenvectors of are the columns of .
  + The eigenvectors of are the columns of .
  + Singular values in are on the diagonal component and are the square roots of eigenvalues, arranged in descending order.
* Convexity: and symmetric positive is convex.
* An orthogonal projector on , a subspace of : , , .
  + Hat matrix: is an orthogonal projector onto the column space of
* eigenvalue of eigenvector:
  + The eigenvalues of an idempotent matrix () are either or
    - Number of eigenvalues equal to is then
* Orthogonal matrix:
* Similar matrices and : there exists an orthogonal matrix such that , they share the same eigenvalues
* Diagonalizable matrix : there exists an orthogonal matrix , such that is diagonal, and its elements being are the eigen values of
* Quantile function:

# Synthèse

## Ordinary Least Square

* Gram matrix:
* Orthogonal projector on :
* The OLS estimator always exists, and the associated prediction is given by . It is either:
  + *uniquely defined* the Gram matrix is invertible, which is equivalent to
  + *non-unique*, with an infinite number of solutions. This happens if and only if
    - , where is a particular solution
    - The traditionally considered solution is
      * Moore-Penrose inverse: For a positive semi-definite symmetric matrix with eigenvectors and corresponding eigenvalues ,
  + , and
* Determination coefficient , because of the orthogonality between and , and between and
  + , implying that is one OLS estimator.

## Statistical Model

### Fixed-design model

* , iid
* Matrix notations : each row corresponds to a sample or .
  + We handle the intercept by either centering the vectors or by fixing the first coordinate of each sample .
* Bias:
  + Unbiased if
* Quadratic risk:
* Prediction risk:
* Linear estimator: , depends only on
* Under the fixed design model:
* Empirical variance:
  + Unbiased:

### Gaussian model

* Hat matrix
* Cochran lemma
  + and are independent
* is independent of
* Central Limit Theorem (CLT): sequence of iid random variables with the same mean and the same standard deviation , by defining :
  + Sufficiently large:

## Hypothesis testing

* Level
* Errors:
  + Type 1: to reject whereas is true
  + Type 2: not to reject whereas is false
* Test of no effect:

## Ridge Regression

* When is not full rank, one can add L2 regularization to make the problem solvable:
  + - Reduce bias
    - Reduce variance

## Least Absolute Shrinkage and Selection Operator (LASSO) Regression

* If we know that only certain coordinates of the samples are useful for predicting , we can perform variable selection. One simple way is to use L1 regularization, which forces most coordinates of to be zero: