# Sobel Test and Bootstrap Test

## 1. Sobel Test

### 1.1. Theory

The Sobel test is a statistical test used to analyze mediation effects. Mediation analysis is used to examine the relationship between an independent variable (X) and a dependent variable (Y) by introducing a mediator (M) to see if M mediates the effect of X on Y. The Sobel test checks the statistical significance of the indirect effect of X on Y through M.

### 1.2. Formula

The Sobel test evaluates the significance of the mediation effect using the following formula:  
  
Z = (a \* b) / sqrt(b^2 \* sa^2 + a^2 \* sb^2)  
  
Where:  
- a: The regression coefficient from X to M (i.e., the coefficient a in M = aX + e1).  
- b: The regression coefficient from M to Y (i.e., the coefficient b in Y = bM + c'X + e2).  
- sa: The standard error of coefficient a.  
- sb: The standard error of coefficient b.  
  
If the Z value exceeds the critical value at a given significance level (e.g., α=0.05), the mediation effect is considered significant.

### 1.3. Real-world Example

Example: In a psychology study, researchers want to investigate the relationship between stress (X) and health (Y). They hypothesize that stress affects health through depression (M).  
  
1. Independent Variable (X): Stress level.  
2. Mediator (M): Level of depression.  
3. Dependent Variable (Y): Health status.  
  
The process:  
- First, analyze the relationship between stress (X) and depression (M) to obtain the coefficient a.  
- Second, analyze the relationship between depression (M) and health (Y) to obtain the coefficient b.  
- Finally, substitute these coefficients into the Sobel test formula to check the significance of the mediation effect.  
  
If the Sobel test is significant, it can be concluded that depression partially mediates the effect of stress on health.

## 2. Bootstrap Test

### 2.1. Theory

The Bootstrap test is a method that uses resampling to estimate the uncertainty of a statistical estimate. Bootstrap involves repeatedly drawing samples from the original data and estimating the statistic of interest (e.g., mean, variance) from each sample. It is particularly useful when the sample size is small or when there is uncertainty about the distribution.

### 2.2. Formula and Process

The basic process of the Bootstrap test is as follows:  
  
1. Draw repeated resamples from the original data.  
- From the given original data, draw n resamples multiple times (e.g., 1000 times).  
- Each resample is drawn with replacement from the original data (i.e., the same data point can be selected multiple times).  
  
2. Calculate the statistic from each resample.  
- For each resample, calculate the statistic of interest (e.g., mean, variance).  
  
3. Generate the distribution of the statistic.  
- Create a distribution of the statistic using the calculated statistics from all resamples.  
  
4. Compare the original data statistic with the resample distribution.  
- Compare the statistic calculated from the original data with the distribution generated from the resamples to calculate the p-value or set confidence intervals.

### 2.3. Real-world Example

Example: Estimating the confidence interval of the mean using the Bootstrap method.  
  
1. Data: Suppose we have exam scores of students (e.g., [70, 75, 80, 85, 90]).  
2. Resampling: Draw 1000 resamples of size 5 from this data with replacement.  
3. Calculate Statistic: Calculate the mean of each resample.  
4. Generate Distribution: Create a distribution of the means from the resamples.  
5. Confidence Interval: Estimate the 95% confidence interval of the mean using the 2.5% and 97.5% percentiles of this distribution.  
  
The Bootstrap method allows reliable estimates of statistics such as the mean, variance, etc., even with small sample sizes. It is particularly useful when there are no strong assumptions about the data distribution, making it flexible for non-normal or complex data.