# Kapitel 1

* Kvantitativ data: mätbar data
* Kategorisk data: ej mätbar - kategorier

## Kategorisk data

#### Kategorisk data, typer

* Nominal: Ex On = 1 / Off = 0. Dvs numrena i sig har ingen betydelse
* Ordinal: (ordnad?) Finns rank mellan kategorierna. Ex låg - mellan - hög

## Kvantitativ data

#### Olika skalor

* Intervall skala: Kan inte meningsfullt jämföra datapunkter med division/multiplikation (addition/subtraktion funkar). Ex temperatur 20 grader vs 40 grader.
* Ratio skala: Kan jämföra data med division/multiplikation. Ex priser. 40 kronor är dubbels så dyrt som 20 kronor.

## Population vs Sample

* Population: The entire group.
* Sample: The group that you have data on. Subset of population.  
  The sample needs to be random if we are to infer any conclusions upon the underlying population.

## Descriptive vs Inferential

* Descriptive statistics: Describes the sample.
* Inferential statistics: Uses sample to draw conclusions of the population.

Sample statistics are used as estimators of the population. We have to quantify bias and variance. Methods for this can be: - Parametric (asssumed shape, ex Gaussian) - Non-parametric (no assumed shape)

## Resampling of sample (Normally used in machine learning)

#### The following methods exist

* Simple random sample. Uses random number generator to choose subset.
* Stratified random sample. Used when specific groups need to be represented proportionally.
* bootstrap sample: Worst method. Used when not enough data points exists for the 2 methods above.