TMA 4255 Applied Statistics

Statistics is the science that concerns the collection, organization, analysis, interpretation and presentation of data.

TMA 4240 Statistics

Probability calculus

Distributions

Inference { Estimation
Confidence interval
Hypothesis testing

As an applied statistician one also needs ~~to~~:

Make models

Collect data

Analyse data

Interpret data.

Usefulness of statistics

Process controlled by operators. (Production of nickel plates for nickel hydrogen batteries).

1. Make random adjustment based on subjective judgement.

2. Registrer the weight of plates, look at earlier registrations and control the process according to these.

Fundamental ideas in the field of Statistics

Opens up for variation in the data
Accept risk and uncertainty.

Central Concepts

Population	Random sample
The set of all possible observations of interest to the problem we are studying Parameters known $\mu, \quad \sigma^2$	$X_1, \ldots, X_n$ random sample if they are independent and identically distributed Statistics: $\bar{X}$, $S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$

Data are normally of two types

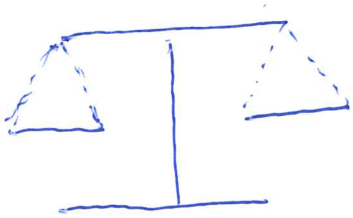
Obtained from planned experiments (causal relationships)

Obtained from observational studies (Correlated relations

An example on why planning collection of the data is important.

A bowl-weight with measurement error σ: Want to decide the weight of 2 bodies A and B (μ_A and μ_B)

Method 1: $\hat{A} = Y_1 \qquad Var(\hat{A}) = \sigma^2$
$\hat{B} = Y_2 \qquad Var(\hat{B}) = \sigma^2$



Method 2. A and B are weighed together

First A+B: Y_3 where $E[Y_3] = \mu_A + \mu_B$

Then A−B: Y_4 where $E[Y_4] = \mu_A - \mu_B$

We get $A^* = \frac{Y_3 + Y_4}{2}$, $B^* = \frac{Y_3 - Y_4}{2}$

$$Var[A^*] = \frac{1}{4}(Var(Y_3) + Var(Y_4)) = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2}$$

$$Var[B^*] = \frac{1}{4}(Var(Y_3) + Var(Y_4)) = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2}$$

An idustrial example

Obtaining a good yild from a process.

Method A: $X_1, \ldots, X_{10}$, sample $E[X_i] = \mu_A$, $i = 1, \ldots, 10$

Method B: $Y_1, \ldots, Y_{10}$, $E[Y_i] = \mu_B$, $i = 1, \ldots, 10$

Interesting to know if $\mu_B > \mu_A$.

How should we start:

Descriptive statistics will provide some information like $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$, sample median and sample quartiles.

Boxplot may tell oss about outliers and how symmetric the data are aroud their central measures ($\bar{x}$, and sample med

Normal-plot may ~~inform~~ be a good check for if ~~it~~ ~~the~~ ~~assumption~~ normal-distribution can be ~~as~~ assumed.