

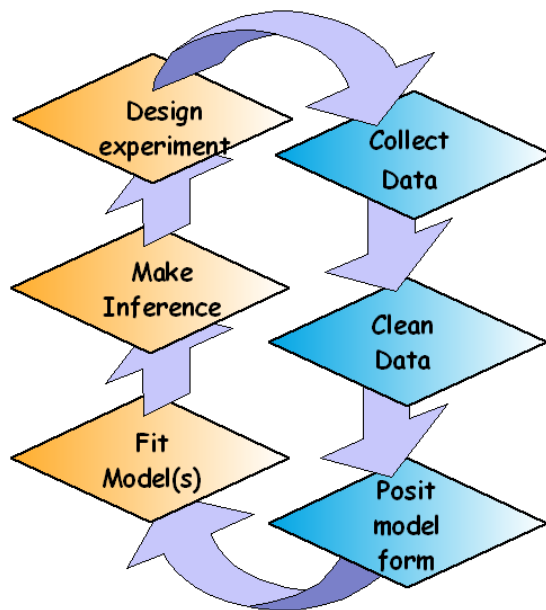
Assumption: Data Generating Mechanism (or model)
→ Data (sample)

Probability Given complete knowledge about data generating mechanism, make assertions about the data.

Statistics Given the data, make inference about the data generating mechanism.

In practice, we move back and forth between data and model:

Model Creation Process



17.2 Data description

Descriptive statistics: Numerical summaries, graphical summaries

Graphical summaries:

- Dot Plot
- Histogram
- Stem and leaf plot
- Box Plot (later)
- Probit (or quantile-quantile) plot (later)

Example: Exam 3

Summary Statistics:

Measures of central tendency:

- sample mean
- sample median
- mode

Definition: if our sample is x_1, x_2, \dots, x_n then

$$\bar{x} = \text{sample mean} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Example: Exam 3

Another useful notation:

$$\sum_{k=M}^N f(k) =_{def} f(M) + f(M+1) + \cdots + f(N)$$

Examples:

$$\sum_{k=2}^5 k^2 = ?$$

$$\sum_{k=0}^3 7 = ?$$

$$\sum_{a=1}^5 \frac{1}{2^a} = ?$$

$$\sum_{k=M}^N Z_k =_{def} Z_M + Z_{M+1} + \cdots + Z_N$$

In this notation,

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$$

Other useful statistics:

MIN, MAX

First Quartile $Q_1 =$ the $(\frac{n}{4})^{th}$ smallest element of the sample

Third Quartile $Q_3 =$ the $(\frac{3n}{4})^{th}$ smallest element of the sample

(example)

Measures of spread (or dispersion):

- range=MAX-MIN
- inter-quartile range=IQR= $Q_3 - Q_1$
- sample variance
- sample standard deviation

$$s^2 = \text{sample variance} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$s = \text{sample standard deviation} = \sqrt{s^2}$$