Data Collection → Theory based
Identifying distribution with data
Parameter estimation
Goodness of fit test

Do from Jerry Banks

Input Modelling → stop Theory 10M

Verification & Validation Theory

$$\mathrm{cov}(x, y) = \frac{1}{n-1} \sum_{1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$\hat{\rho} = \frac{\mathrm{cov}(x, y)}{\sigma_1 \sigma_2}$$

$(6.5 - 6.14)(103 - 101.8)$
$(4.3 - 6.14)(83 - 101.8)$
$(6.9 - 6.14)$
$(6.6 - 6.14)$

$\sum(x - \bar{x})(y - \bar{y})$

$\sum xy - \sum y\bar{x} - \sum x\bar{y} + \bar{x}\bar{y}$

stock broker has data of buy & sell order in seconds

find correlation & covariance & how to use it to model EAR process

- 1.95
- 1.58
- 1.28
- 1.04
- 0.84
- 0.68
- 11.98
- 9.71
- 12.62
- 10.22

$\longrightarrow$

$$\bar{X} = \frac{\sum x_i}{n} = 5.19$$

lag 1 correlation $\hat{\phi}$

$$\hat{\phi} = \rho = \frac{\operatorname{Cov}(x_t, x_{t+1})}{\hat{\sigma}^2}$$

$$\operatorname{Cov}(x_t, x_{t+1}) = \frac{\sum x_t \times x_{t-1} - (n-1)\bar{x}^2}{n-1}$$

$$383.86 - 8 \times (5.19)^2$$

$$= 15.71$$

$$\sigma^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1}$$

$$\hat{\phi} = \frac{15.71}{26.92} = 0.58$$

$$\lambda = \frac{1}{\mu} = 0.19$$

NGO records $\rightarrow$

Apply x^2 to test if it is poisson

1. AR
2. FAR
3. Poisson fit

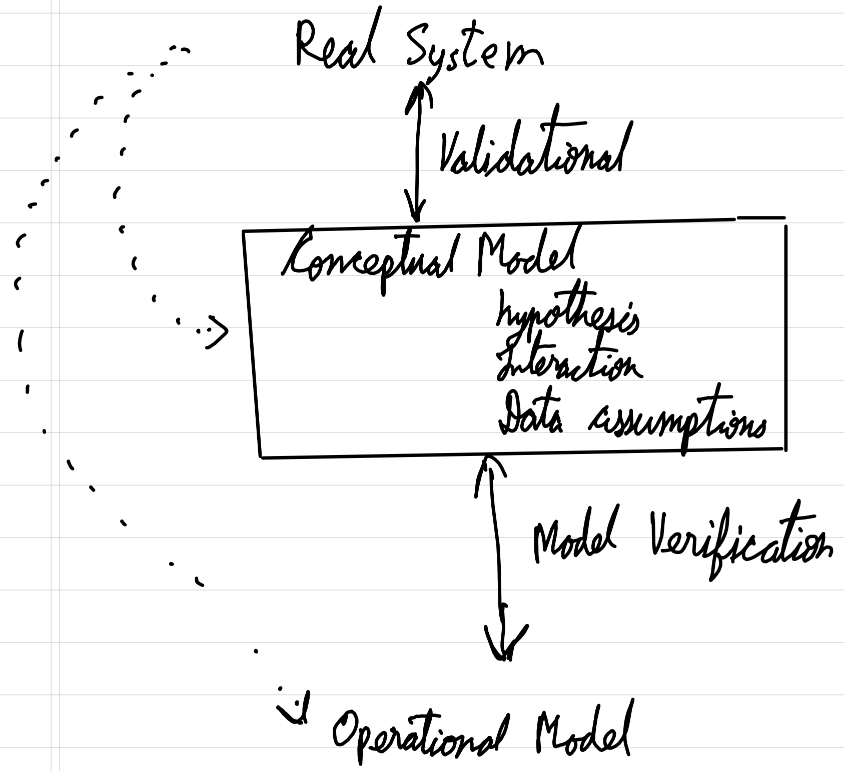
} Numerical

Verification → No coding errors
Check for implementation errors

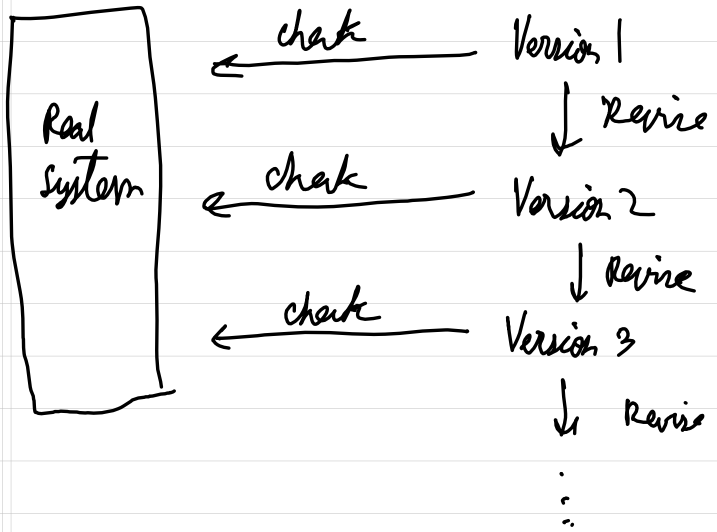
Validation → Model assumptions are correct
Accurate representation of real model

Verification : Building model correctly

Validation : Building correct model.



Calibration → Update the model, tweak
iterative process



Techniques for Verification & Validation

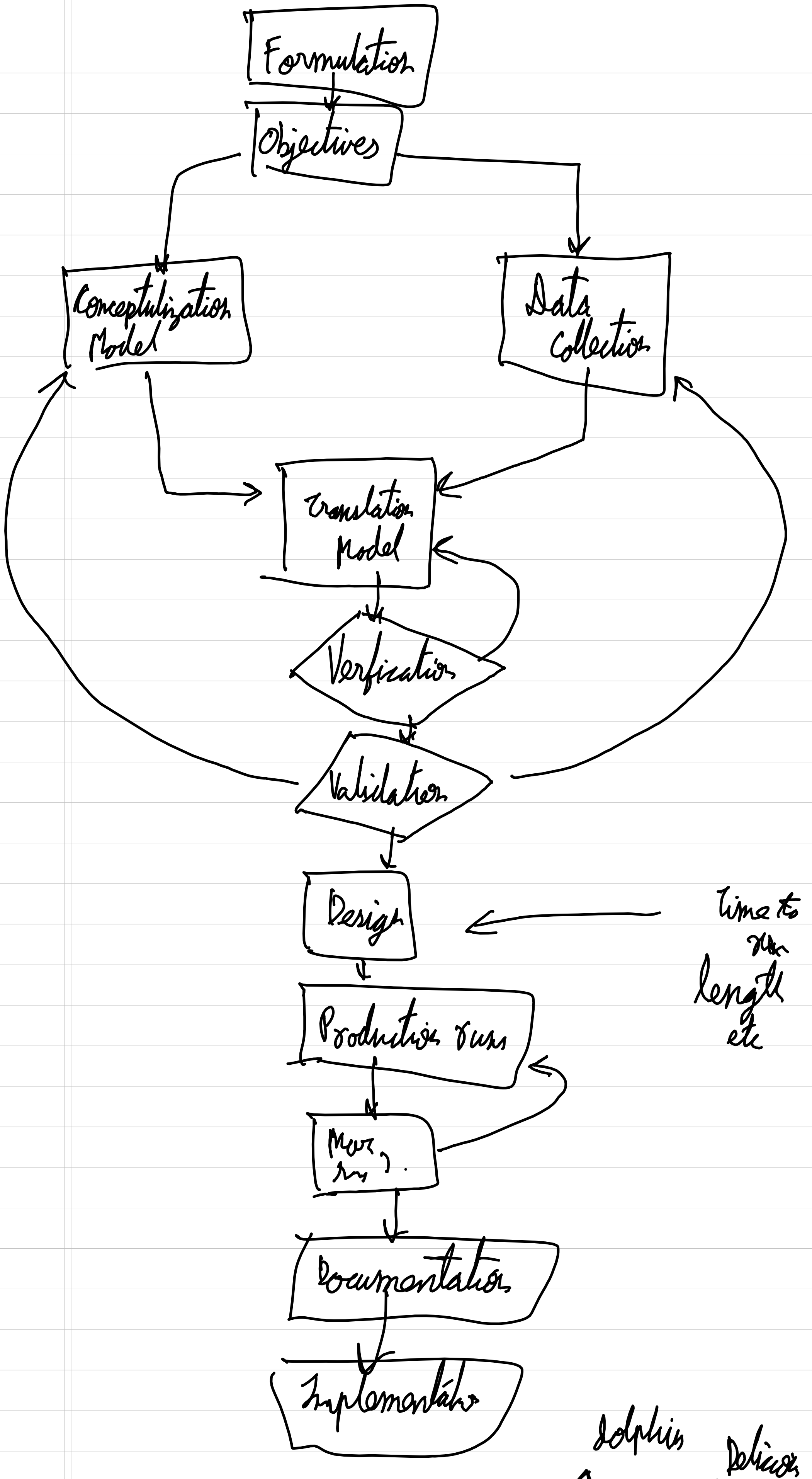
* Verification →

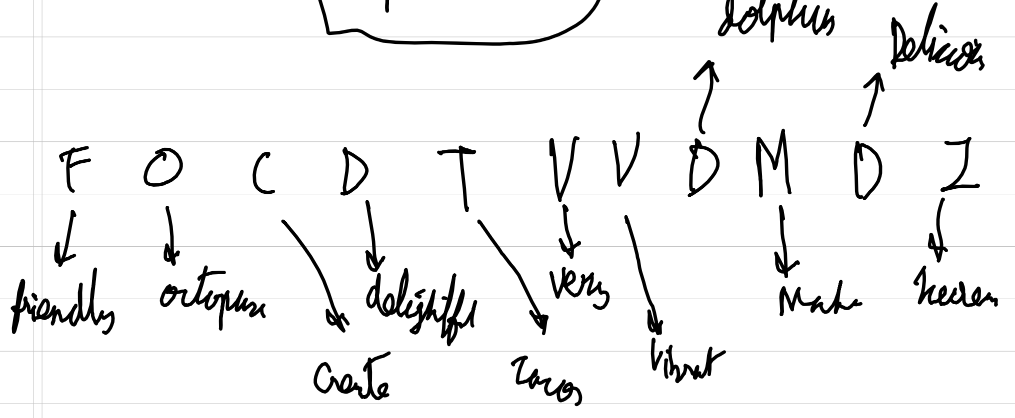
1. Code review by someone else
2. Debugger
3. Manual testing
4. Reasonability analysis → high face value
5. Sensitivity Analysis → check changes in the output when one or more parameters are changed.

6. Historical data

De Co Re Se da Te

* Validation → 1. High face validity (Compare with reality)
  2. Interview experts eg bus
  3. Subjective tests


Formulation
Objectives
Conceptulization Model
Data Collection
Translation Model
Verfication
Validation
Design
Time to run length etc
Production runs
More runs?
Documentation
Implementation


F O C D T V V D M D I
friendly
octopus
Create
delightful
Tacos
very
vibrant
dolphins
Make
Delicious

Steps for input Modelling

$$\begin{array}{ll} \text{Data Collection} & \\ \downarrow & \\ \text{Identifying distribution} & \\ \downarrow & \\ \text{Parameter estimation} & \text{mean, s.D. etc} \\ \downarrow & \\ \text{Goodness of fit test} & \chi^2,\ \text{k.s. etc} \end{array}$$

Data Collection

Garbage in Garbage Out

Data mislead → Inaccurately collected
Inaccurately Analyzed
Not representative of environment

Plan ahead – observe for unusual circumstances

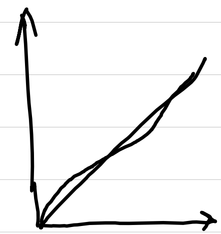
Identifying Distribution

① Histogram $\rightarrow$ Interval size $= \sqrt{s}$

② Selecting family of distributions $\rightarrow$ when to use which distribution

③ Q.Q. plots $\rightarrow$ Evaluate distribution fit

Must be straight line



Quantile Quantile Plots

Plot quantiles of a sample distribution against the theoritical distribution

QQ plot is a useful tool for evaluating distribution fit.

Helps to determine if a dataset follows any particular type of probability distribution

Uniform : All outcomes are equally likely

Binomial : Number of successes in n independant trials

Geometric : Number of trials to achieve k successes

Poission : Number of independant events occur in fixed time

Normal : Models distribution in Bell curve - sum of component processes

lognormal : Product of component process

Exponential : Time between independant events for a memoryless process

Gamma : Nonnegative random variables

Beta : Bounded random variables

Erlang : Sum of several exponentially distributed processes

Weibull : Time to failure

Triangular : Only minimum, most likely & maximum values known

Empirical : Resamples from the actual data collected when no theoritical distribution appropriate

Input output validation using Turing Test

step 1 Collect real reports of a system in a familiar format

step 2 Generate fake reports using simulation in same format

step 3 Give reports to engineers & managers

step 4 If they can't tell the difference that means the simulation is good

step 5 If they tell difference then learn from data.

AR $\quad u = \bar{x}$

$$\sigma_\varepsilon^2 = \sigma^2(1-\hat{\phi}^2)$$

lag 1 correlation $\quad \hat{\phi} = \frac{Cov(x_t, x_{t+1})}{\sigma^2}$

EAR $\quad \lambda = 1/\bar{x}$

$$\hat{\phi} = \frac{Cov(x_t, x_{t+1})}{\sigma^2}$$

$$Cov(x_t, x_{t+1}) = \frac{\sum (x_t - \bar{x})(x_{t-1} - \bar{x})}{N-1}$$

total N

$$= \frac{\left(\sum_{1}^{N-1} x_t x_{t+1} - (n-1)\bar{x}^2\right)}{N-1}$$

$$\sigma^2 = \frac{\sum (x_t - \bar{x})^2}{N}$$