Planeamiento Avanzado en Cadenas de Aprovisionamiento

centrum.pucp.edu.pe

centrum PUCP ESCUELA PARA LOS BUENOS NEGOCIOS

UNIDAD 2: SUPPLY CHAIN FORECASTING

Sesión 5

Regresión Lineal Múltiple

Multiple Regression

Multiple Regression allows us to:

- Use several variables at once to explain the variation in a continuous dependent variable.
- Isolate the unique effect of one variable on the continuous dependent variable while taking into consideration that other variables are affecting it too.
- Write a mathematical equation that tells us the overall effects of several variables together and the unique effects of each on a continuous dependent variable.

The Multiple Regression Model

Multiple linear regression is very similar to simple linear regression except that the dependent variable Y is described by k independent variables $X_1, \ldots, X_k$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

Multiple Linear Regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

- Intercept is the same
- Slope b_i is the change in Y given a unit change n X_i while holding all other variables constant (more on this later)
- SST, SSE, SSR, and R^2 are the same
- s_e is the same except now s_e = sqrt(SSE / (n-k-1))
- Slope coefficient C.I.s are the same
- p-values (one for each X_i) are the same

Example: Housing Data

We have data on 15 randomly selected house sales from last year:

price	size	age	lotsize
89.5	20.0	5	4.1
79.9	14.8	10	
83.1	20.5	8	6.3
56.9	12.5	7	5.1
66.6	18.0	8	4.2
82.5	14.3	12	8.6
126.3	27.5	1	4.9
79.3	16.5	10	6.2
119.9	24.3	2	
87.6	20.2	8	5.1
112.6	22.0	7	6.3
120.8	19.0	11	12.9
78.5	12.3	16	9.6
74.3	14.0	12	5.7
74.8	16.7	13	4.8

size in 100 sq-feet

age in years

How does selling price relate to the three variables?

```
> fit=lm(price~size+age+lotsize)
> summary(fit)

Call:
lm(formula = price ~ size + age + lotsize)

Residuals:
     Min       1Q   Median       3Q      Max
-14.3848  -1.7477   0.5549   4.0566   8.6598

Coefficients:
             Estimate Std. Error t value   Pr(>|t|)
(Intercept) -16.0580    19.0710  -0.842   0.417712
size          4.1462     0.7512   5.520   0.000181  ***
age          -0.2361     0.8812  -0.268   0.793730
lotsize       4.8309     0.9011   5.361   0.000230  ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01   '*' 0.05  '.' 0.1 ' ' 1

Residual standard error: 6.894 on 11 degrees of freedom
Multiple R-squared:  0.9161,    Adjusted R-squared:  0.8932
F-statistic: 40.03 on 3 and 11 DF,  p-value: 3.278e-06
```

S_e

Planeamiento Avanzado de Cadenas de Aprovisionamiento

Interpretation

The relationship between house size and price is measured by b_1 = 4.146. This indicates that in this model, for each additional 100 square feet, the price of the house increases (on average) by $4,146 (assuming that the other independent variables are fixed).

The coefficient b_2 = -.236 specifies that for each additional year in the age of the house, the price decreases by an average of $236 (as long as the values of the other independent variables do not change).

The coefficient b_3 = 4.831 means that for each additional 1000 sq-feet if lot size, the price increases by an average of $4831 (assuming that house size and age remain the same).

This Held Fixed Concept

- In a multiple regression model, the interpretation of a parameter is entirely dependent upon the model in which the parameter appears.

- If you have the "wrong" sign, you may not be thinking clearly about the "held fixed" meaning of the parameters (it can be confusing).

Adjusted R-squared

It can be shown that every time you add a new X variable to a multiple regression the error sum of squares (SSE) goes down (math fact).

Since,

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} = 1 - \frac{SSE}{SST}$$

this means that every time you add an X variable, R^2 goes up.

The adjusted R-squared is designed to build in an automatic penalty for adding an X.

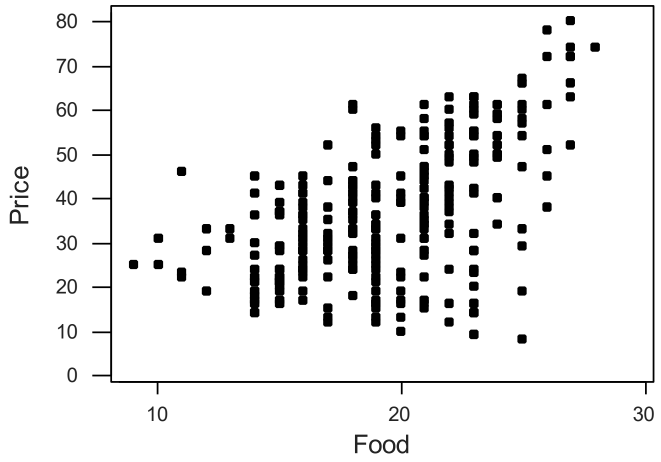
$$R_a^2 = 1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n} e_i^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\frac{1}{n-k-1}SSE}{\frac{1}{n-1}SST}$$

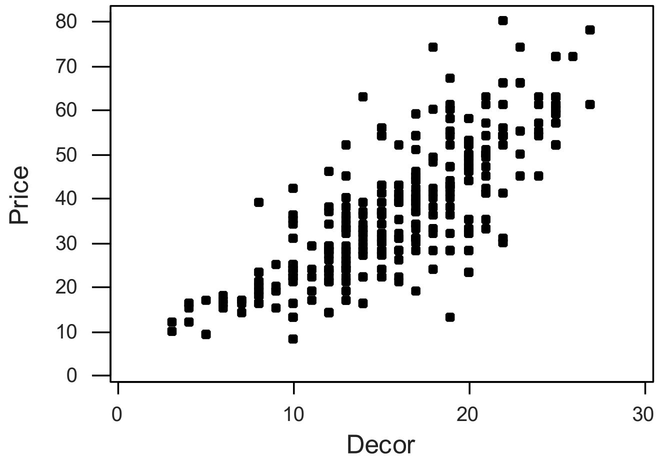
I find the “penalty” artificial.

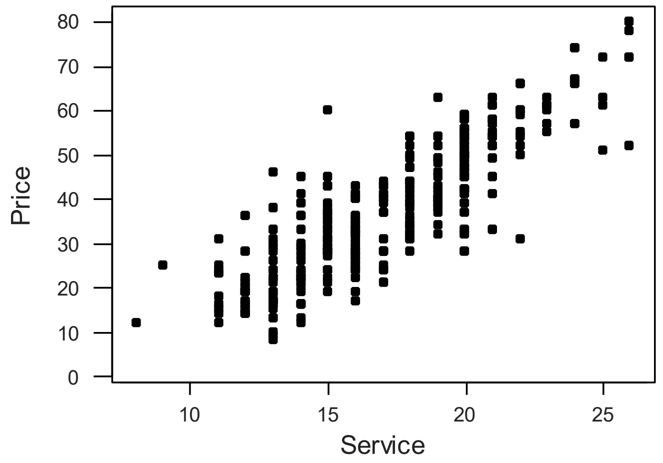
Another Example

- Consider Zagat food ratings for Manhattan.
- We have data on price of meal, and ratings for food quality, décor and service.

Zagat data : Relationship between price and other variables at the same time!







First,

- Regress price on food quality:

```
> fit=lm(price~food)
> summary(fit)

Call:
lm(formula = price ~    food)

Residuals:
    Min       1Q Median      3Q     Max
-23.49   -8.31   -1.85    7.11   42.59

Coefficients:
               Estimate Std. Error t    value Pr(>|t|)
(Intercept)      -3.871       10.047    -0.39   0.70046
food              1.640        0.436     3.76   0.00022 ***
---
Signif. codes:     0 '***' 0.001 '**'    0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12 on 193 degrees of freedom
Multiple R-squared:  0.0683,    Adjusted R-squared:  0.0635
F-statistic: 14.2 on 1 and 193 DF,  p-value: 0.000223
```

Planeamiento Avanzado de Cadenas de Aprovisionamiento

Now a multiple regression

```
> fit=lm(price~food+decor+service)
> summary(fit)

Call:
lm(formula = price ~          food + decor + service)

Residuals:
    Min       1Q Median              3Q     Max
-20.50   -5.90   -0.38            4.78   47.76

Coefficients:
              Estimate  Std. Error t  value    Pr(>|t|)
(Intercept) -28.3326      8.2553   -3.43    0.00073 ***
food          -0.0401      0.3993   -0.10     0.92016
decor          0.7471      0.2702    2.76     0.00626 **
service        2.5097      0.4495    5.58  0.00000008 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.7 on 191 degrees of freedom
Multiple R-squared:  0.424,     Adjusted R-squared:  0.415
F-statistic: 46.8 on 3 and 191 DF,  p-value: <0.0000000000000002
```

Compare: What Happened?

```
Call:
lm(formula = price ~       food)

Residuals:
      Min       1Q Median       3Q      Max
   -23.49    -8.31  -1.85     7.11    42.59

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    -3.871     10.047   -0.39  0.70046
food            1.640      0.436    3.76  0.00022 ***
---

Call:
lm(formula = price ~ food + decor + service)


Coefficients:
             Estimate  Std. Error t  value    Pr(>|t|)
(Intercept)  -28.3326      8.2553   -3.43     0.00073 ***
food          -0.0401      0.3993   -0.10     0.92016
decor          0.7471      0.2702    2.76     0.00626 **
service        2.5097      0.4495    5.58  0.00000008 ***
```

Capacity Analysis

Variability Views

Variability:

- Any departure from uniformity
- Random versus controllable variation

Randomness:

- Essential reality?
- Artifact of incomplete knowledge?
- Management implications: robustness is key

Variability

Definition: Variability is anything that causes the system to depart from regular, predictable behavior.

Sources of Variability:

- setups
- machine failures
- materials shortages
- yield loss
- rework
- operator unavailability
- workpace variation
- differential skill levels
- engineering change orders
- customer orders
- product differentiation
- material handling

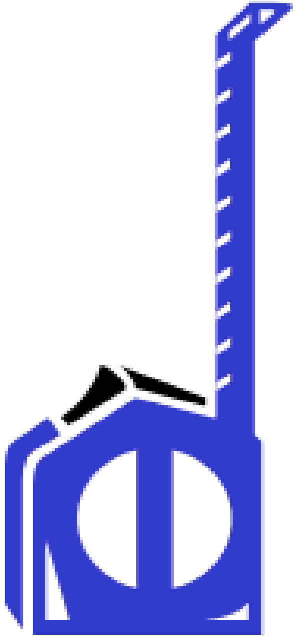
Measuring Process Variability

$$t_e = \text{mean process time of a job}$$

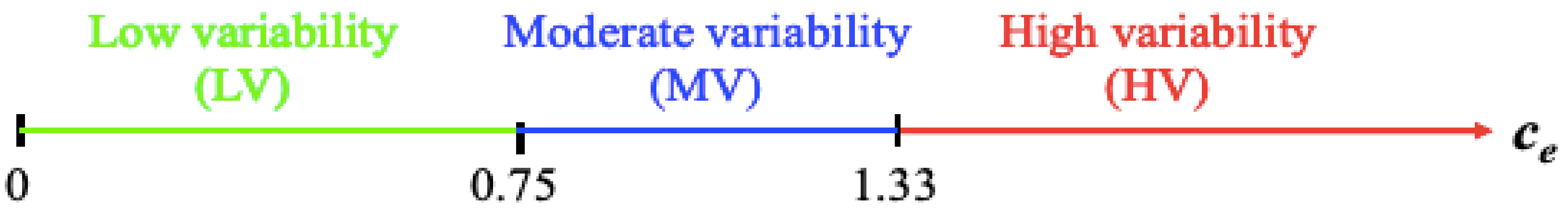
$$\sigma_e = \text{standard deviation of process time}$$

$$c_e = \frac{\sigma_e}{t_e} = \text{coefficient of variation, CV}$$

Note: we often use the "squared coefficient of variation" (SCV), c_e^2



Variability Classes in Factory Physics®



Effective Process Times:

- ***actual*** process times are generally LV
- ***effective*** process times include setups, failure outages, etc.
- HV, LV, and MV are all possible in effective process times

Relation to Performance Cases: For balanced systems

- MV – Practical Worst Case
- LV – between Best Case and Practical Worst Case
- HV – between Practical Worst Case and Worst Case

Trial	Machine 1	Machine 2	Machine 3
1	22	5	5
2	25	6	6
3	23	5	5
4	26	35	35
5	24	7	7
6	28	45	45
7	21	6	6
8	30	6	6
9	24	5	5
10	28	4	4
11	27	7	7
12	25	50	500
13	24	6	6
14	23	6	6
15	22	5	5
t_e	25.1	13.2	43.2
s_e	2.5	15.9	127.0
c_e	0.1	1.2	2.9
Class	LV	MV	HV

Natural Variability

Definition: variability without explicitly analyzed cause

Sources:

- operator pace
- material fluctuations
- product type (if not explicitly considered)
- product quality

Observation: natural process variability is usually in the LV category.

Definitions

$$t_0 = \text{base process time}$$

$$c_0 = \text{base process time coefficient of variability}$$

$$r_0 = \frac{1}{t_0} = \text{base capacity (rate, e.g., parts/hr)}$$

$$m_f = \text{mean time to failure}$$

$$m_r = \text{mean time to repair}$$

$$c_r = \text{coefficent of variability of repair times } (\sigma_r / m_r)$$

Planeamiento Avanzado de Cadenas de Aprovisionamiento

Down Time – Mean Effects (cont.)

Availability: Fraction of time machine is up

$$A = \frac{m_f}{m_f + m_r}$$

Effective Processing Time and Rate:

$$r_e = A r_0$$

$$t_e = t_0 / A$$

Two systems analysis

MX1	TZ00
$t_0 = 15$ min	$t_0 = 15$ min
$\sigma_0 = 3.35$ min	$\sigma_0 = 3.35$ min
$c_0 = \sigma_0 / t_0 = 3.35/15 = 0.05$	$c_0 = \sigma_0 / t_0 = 3.35/15 = 0.05$
$m_f = 12.4$ hrs (744 min)	$m_f = 1.9$ hrs (114 min)
$m_r = 4.133$ hrs (248 min)	$m_r = 0.633$ hrs (38 min)
$c_r = 1.0$	$c_r = 1.0$

Availability:

$$A = \frac{m_f}{m_f + m_r} = \frac{744}{744+248} = 0.75$$

$$A = \frac{m_f}{m_f + m_r} = \frac{114}{114+38} = 0.75$$

No difference between machines in terms of availability.

Variability effects

Effective Variability:

$$t_e = t_0 / A$$

$$\sigma_e^2 = \left(\frac{\sigma_0}{A}\right)^2 + \frac{(m_r^2 + \sigma_r^2)(1-A)t_0}{Am_r}$$

$$c_e^2 = \frac{\sigma_e^2}{t_e^2} = c_0^2 + (1 + c_r^2)A(1-A)\frac{m_r}{t_0}$$

Planeamiento Avanzado de Cadenas de Aprovisionamiento

Two systems analysis

MX1 | **TZ00**

MX1

$$t_e = \frac{t_0}{A} = \frac{15}{0.75} = 20 \text{ min}$$

$$c_e^2 = c_0^2 + (1 + c_r^2)A(1-A)\frac{m_r}{t_0} =$$

$$0.05 + (1+1)0.75(1-0.75)\frac{248}{15} =$$

$$6.25 \quad \text{high variability}$$

TZ00

$$t_e = \frac{t_0}{A} = \frac{15}{0.75} = 20 \text{ min}$$

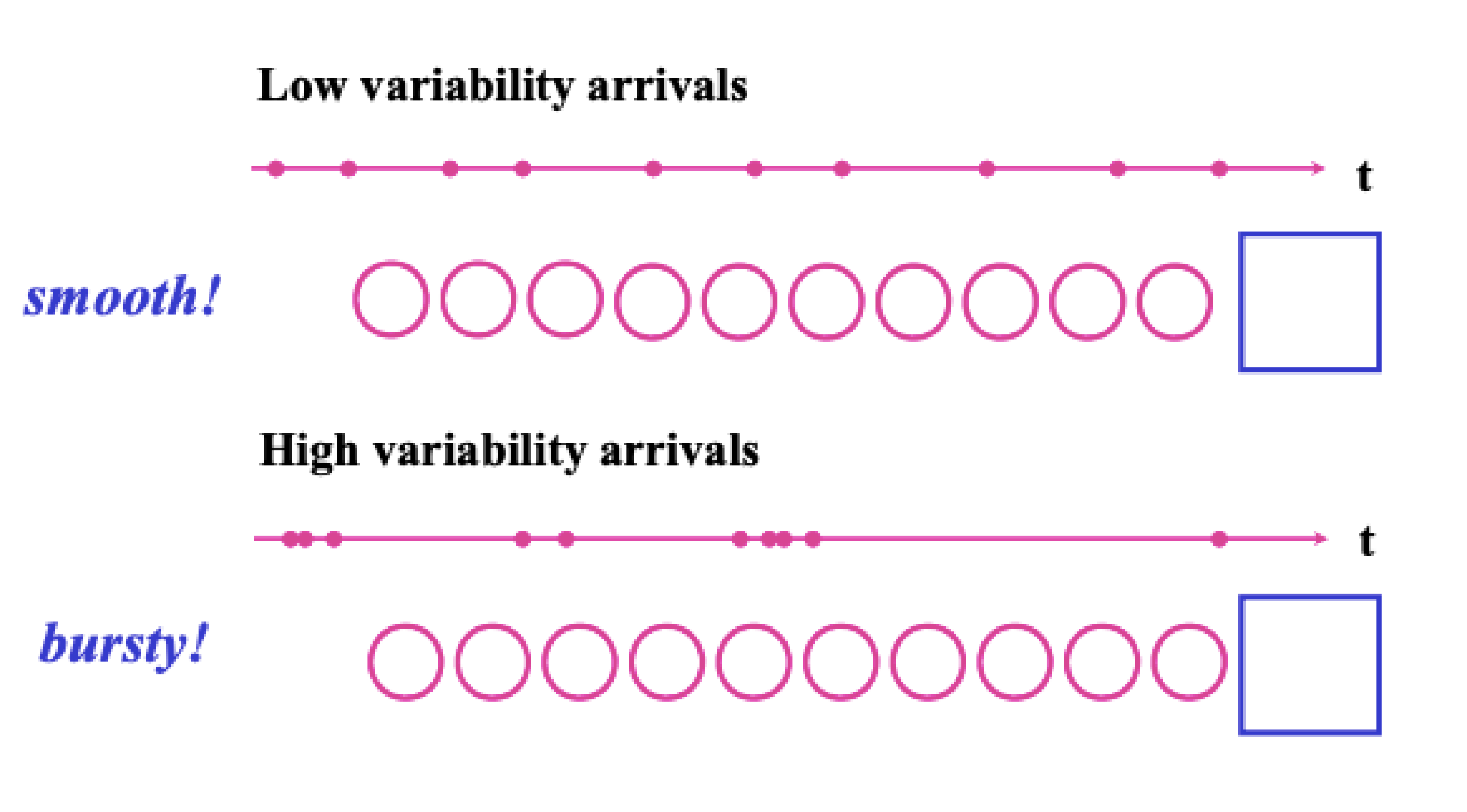
$$c_e^2 = c_0^2 + (1 + c_r^2)A(1-A)\frac{m_r}{t_0} =$$

$$0.05 + (1+1)0.75(1-0.75)\frac{38}{15} =$$

$$1.0 \quad \text{moderate variability}$$

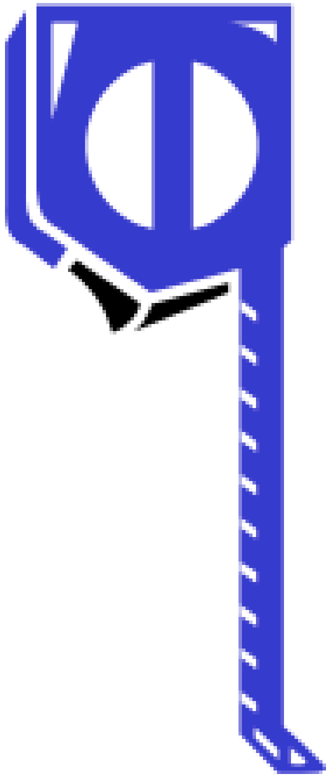
Planeamiento Avanzado de Cadenas de Aprovisionamiento

Two Flow Variability



Measuring Flow Variability

$$t_a = \text{mean time between arrivals}$$

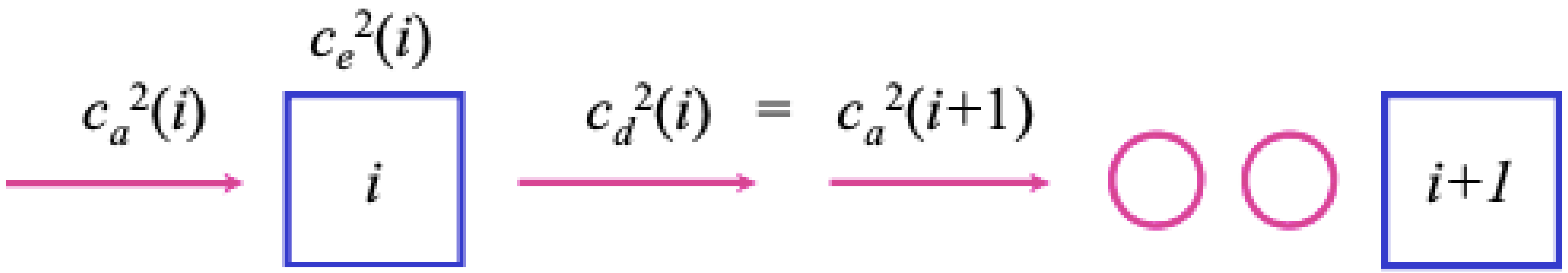


$$r_a = \frac{1}{t_a} = \text{arrival rate}$$

$$\sigma_a = \text{standard deviation of time between arrivals}$$

$$c_a = \frac{\sigma_a}{t_a} = \text{coefficient of variation of interarrival times}$$

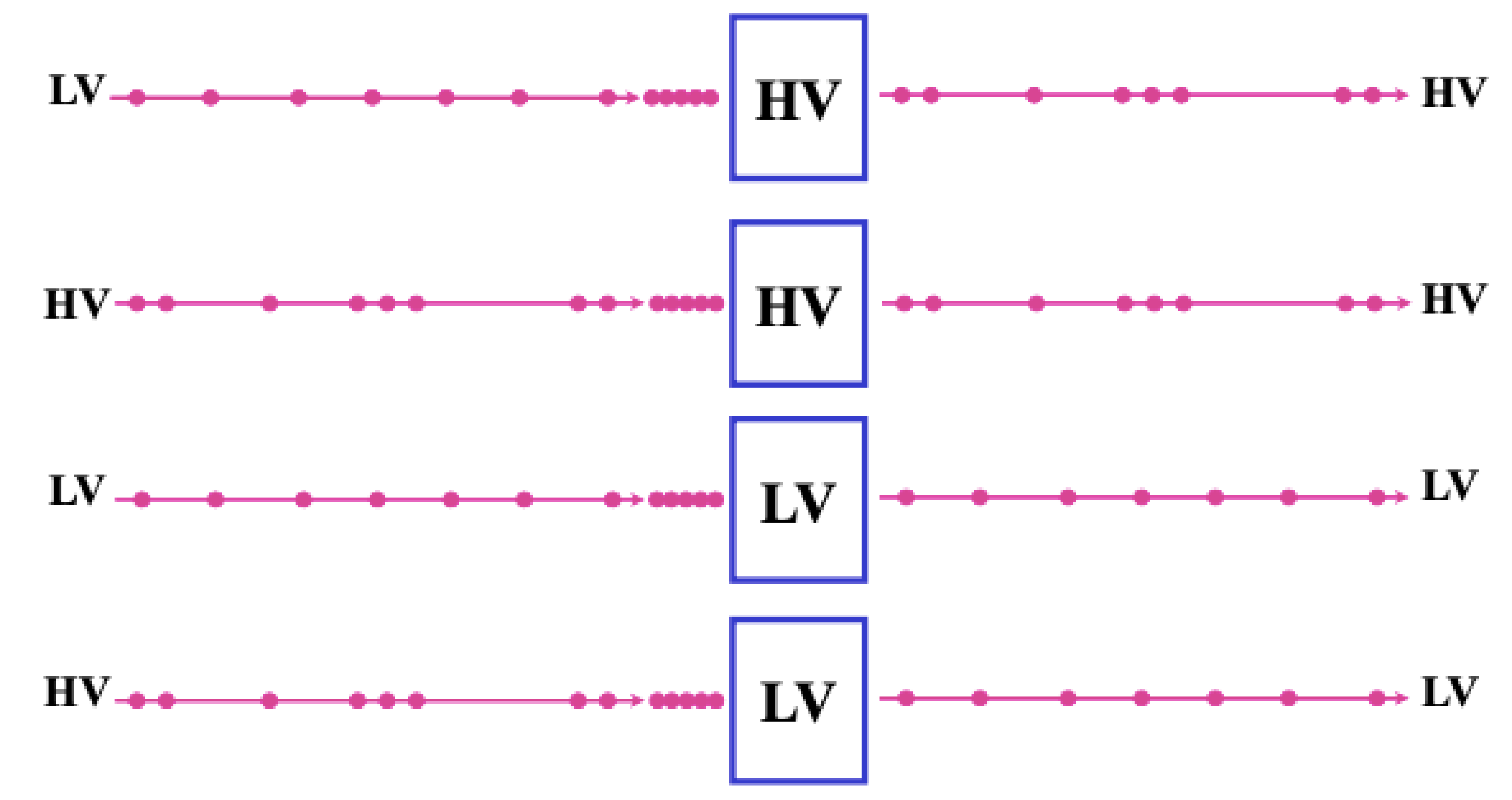
Propagation of Variability



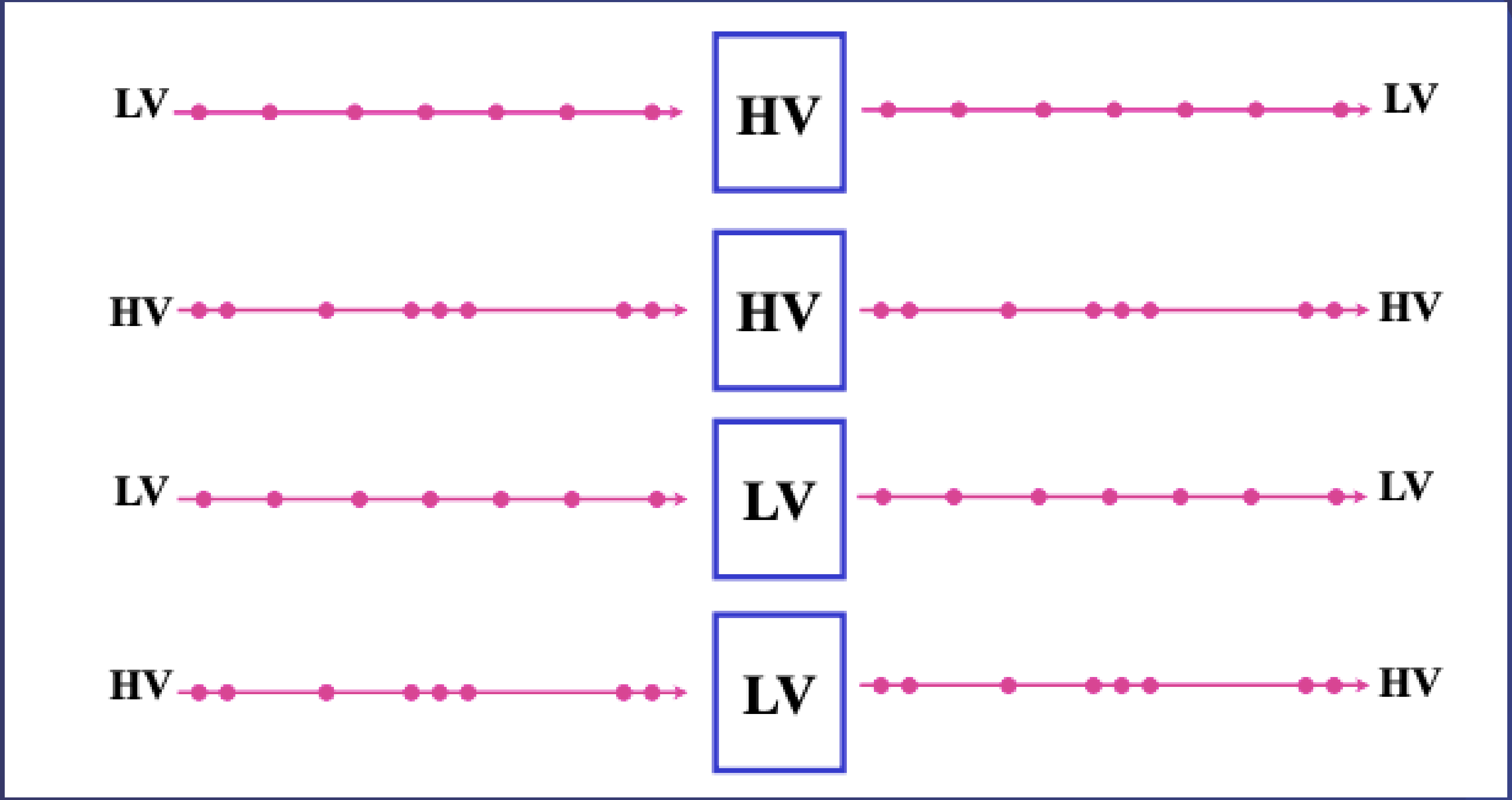
$$\boxed{c_d^2 = u^2 c_e^2 + (1-u^2) c_a^2}$$

where u is the station utilization given by $u = r_a t_e$

Propagation of Variability – High Utilization Station



Propagation of Variability – Low Utilization Station



Spectrum of Location Decisions

- Transportation services link locations into an integrated logistical system.
- Selection of individual locations represents competitive and cost-related logistical decisions.
  - Manufacturing plant locations may require several years to fully deploy
  - Warehouses can be arranged to use only during specified times
  - Retail locations are influenced by marketing and competitive conditions

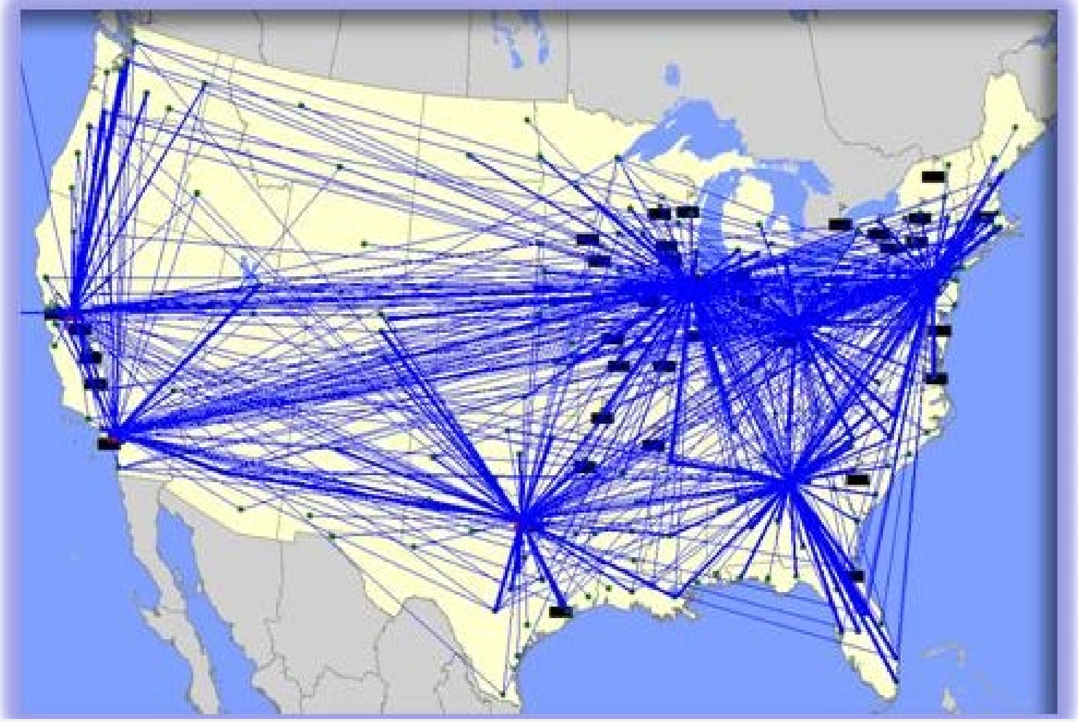
Local Presence: An Obsolete Paradigm

Local presence paradigm

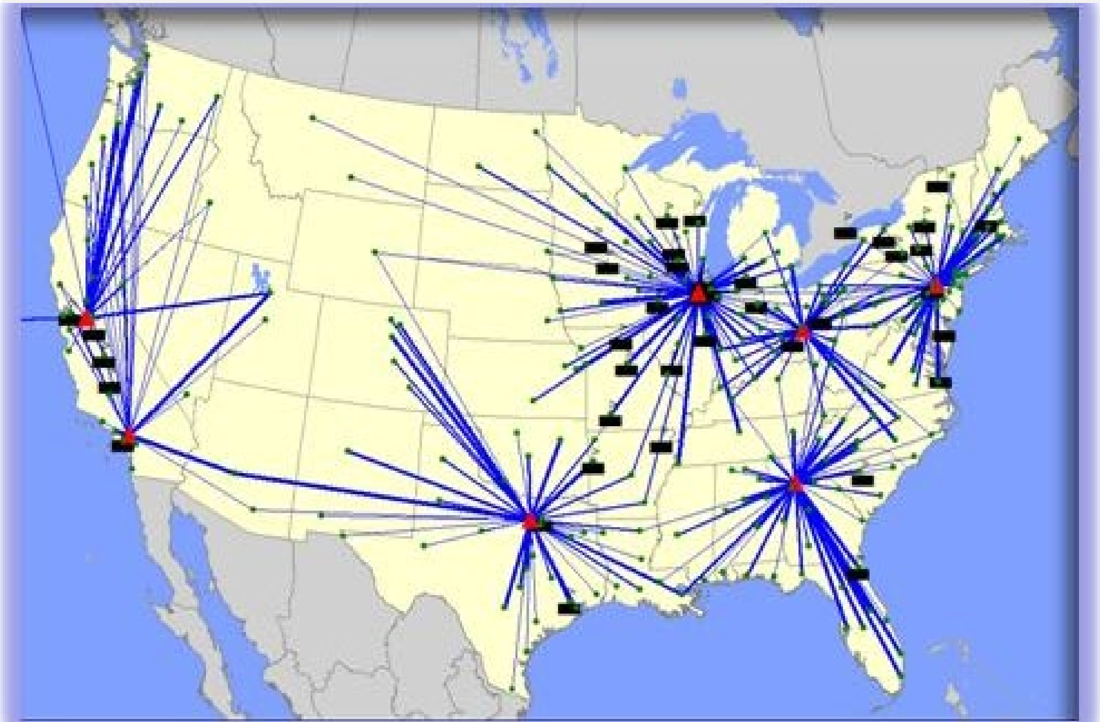
- Transportation services started out erratic with few choices.
- Customers felt that inventory within the local market area was needed to provide consistent delivery.

Contemporary view

- Transportation services have expanded.
- Shipment arrival times are dependable and consistent.
- Information technology.
  - Provides faster access to customer requirements
  - Enables tracking of transport vehicles



Network Modeling Steps



Planeamiento Avanzado de Cadenas de Aprovisionamiento

Strategic Importance of Network Design

- Critical variables in network design:
  - Changing Customer Service Requirements
  - Shifting Locations of Customer and/or Supply Markets
  - Change in Corporate Ownership
  - Cost Pressures
  - Competitive Capabilities
  - Corporate Organizational Change

Supply Chain Logistics Management, First Edition. Bowersox, Closs, and Cooper.

High-level Modeling Steps

Supply Chain Strategy
- Review project goals & objectives
- Describe current network strategy
- Analyze logistics processes & enablers
- Define customer needs
- Determine logistics cost drivers
- Define products & markets
- Develop conceptual model

Gather Data
- Develop data requirements
- Submit data request
- Receive & organize data

Analyze Data
- Data cleaning
- Statistical analysis
- Product flow summary
- Assumption generation

Baseline Network
- Product flow modeled
- Transportation profile developed
- Modeled baseline / operating model
- Assumptions documented
- Calibrate model

Analyze & Optimize Network
- Develop scenarios to be modeled
- Clean-sheet approach
- Model with sensitivity analysis
- Determine optimal locations
- Model output analysis
- Sensitivity trade-off analysis
- Overall network strategy
- Validation

Recommend
- Transportation cost impact
- Overall cost analysis
- Service analysis
- Implementation impact assessment
- Network design recommendation

Modeling Approaches: Optimization Models

- Based on precise mathematical procedures guaranteed to find the "best" solution from among a number of feasible solutions.
- One approach is Linear Programming (LP).
  - Useful in linking facilities in a network.
  - Defines optimum distribution patterns.
  - Modern computers facilitate LP modeling.

Supply Chain Logistics Management, First Edition. Bowersox, Closs, and Cooper.

Modeling Approaches: Simulation Models

- Based on developing a model of a real system and conducting experiments with this model.
- In location theory, a firm can test the effect of various locations on costs and profitability.
- Does not guarantee an optimum solution but evaluates through the iterative process.
- Simulations are either static or dynamic depending upon how whether they incorporate data from each run into the next run.

Supply Chain Logistics Management, First Edition. Bowersox, Closs, and Cooper.

Modeling Approaches: Heuristic Models

- Based upon developing a model that can provide a good approximation to the least-cost location in a complex decision problem.
- Can reduce a problem to a manageable size.
- This approach can be as sophisticated as mathematical optimization approaches.

THANK YOU!!

Section for questions

Carlos Mariño, Ph.D.

Director de Investigación
Profesor

(511) 626 7100 ext 7200

cmarino@pucp.pe

www.centrum.pucp.edu.pe