Institute of Business Forecasting & Planning

Welcome

Advanced Demand Planning & Forecasting: Applications

Facilitator Name

Sara Brumbaugh



(About me)

ibf.org/demand-planning-and-forecasting-awards/winners/2015

About Us | Membership | Journal | Research | S&OP Self-Assessment

Institute of Business Forecasting & Planning

Events | Train

2026 winners | 2025 winners | 2024 winners | 2023 winn

2017 winners | 2016 winners | 2015 winners | 2014 winn



IN BUSINESS FORECASTING & PLANNING

Sara Brumbaugh

BIOARRAY Therapeutics www.bioarray.us

Gene expression profiles accompanying phenotypic changes during non-malignant breast epithelial cells acini formation to explain MRI phenotypes

Marcia V. Fournier Ph.D.[1], Alan Derr[1], Alex Margulis Ph.D.[1], Kevin Reid[1], Sara Brumbaugh[2], Richard Anderson[3], W Fraser Symmans[4], ACRIN 6657 Trial investigators, Laura Esserman[5], and Nola Hylton[5].

Background

Magnetic resonance imaging (MRI) captures the three dimensional organization of tumors n the breast, defined as imaging phenotypes in the I-SPY 1 trial (CALGB 150007/150012, ACRIN 6657).

We developed a gene set based on the breast epithelial cells organization in three dimensional cultures. We investigated whether these organizational genes correspond to the imaging phenotypes.

MRI phenotypes have been shown to correspond to pathological complete response (pCR) to neoadjuvant chemotherapy, and are used to predict the ability to achieve breast conservation treatment (Mukhtar, et al, Ann Surg Oncol, 20: 3823–3830, 2013).

We hypothesized that the molecular profile accompanying phenotypic changes occurring during the organization process of non-malignant

Cell

Resource

Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain

AJP Reports

A Companion to the American Journal of Perinatology

Authors

Arpiar Saunders, Evan Z. Macosko, Alec Wysoker, ..., Sara Brumbaugh, David Kulp, Steven A. McCarroll

Correspondence

asaunders@genetics.med.harvard.edu (A.S.), emacosko@broadinstitute.org (E.Z.M.), nccarroll@genetics.med.harvard.edu (S.A.M.)

In Brief

Sampling across multiple brain regions identifies hundreds of transcriptionally distinct groups of cells and reveals large-scale features of brain organization and neuronal diversity.

▸ AJP Rep. 2020 Mar 16;10(1):e93–e100. doi: 10.1055/s-0040-1705141

Electronic Fetal Monitoring Credentialing Examination: The First 4000

Mark W Tomlinson [1,✉], Sara A Brumbaugh [2], Marin O'Keeffe [1], Richard L Berkowitz [3], Mary D'Alton [3], Michael Nageotte [4]; on behalf of the Perinatal Quality Foundation

▸ Author information ▸ Article notes ▸ Copyright and License information

PMCID: PMC7075713 PMID: 32190412

Abstract

Objective Recognized variability in fetal heart rate interpretation led the Perinatal Quality Foundation (PQF) to develop a credentialing exam. We report an evaluation of the 1st 4000 plus PQF Fetal Monitoring Credentialing (FMC) exams.

FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Forecasting

the best job in an organization

- Opportunities for skills
- Data access
- Working relationships
- Answerable questions



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Given a wealth of data…



How do I predict a metric?
(sales, market return, or weight loss)

How do I predict a yes/no event?
(response to a promotion, probability of attrition)

How do I know what will drive or predict a result?
(… all of the above!)

How do I know which products need advanced forecasting tools?

How can I construct market segments?
(customers or patients or portfolios)

What products are purchased together?
(or events occur together)

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Building on Day 1…

I. **Applying what you've learned**: a bit more background

II. Begin from **Exploratory Data Analysis**

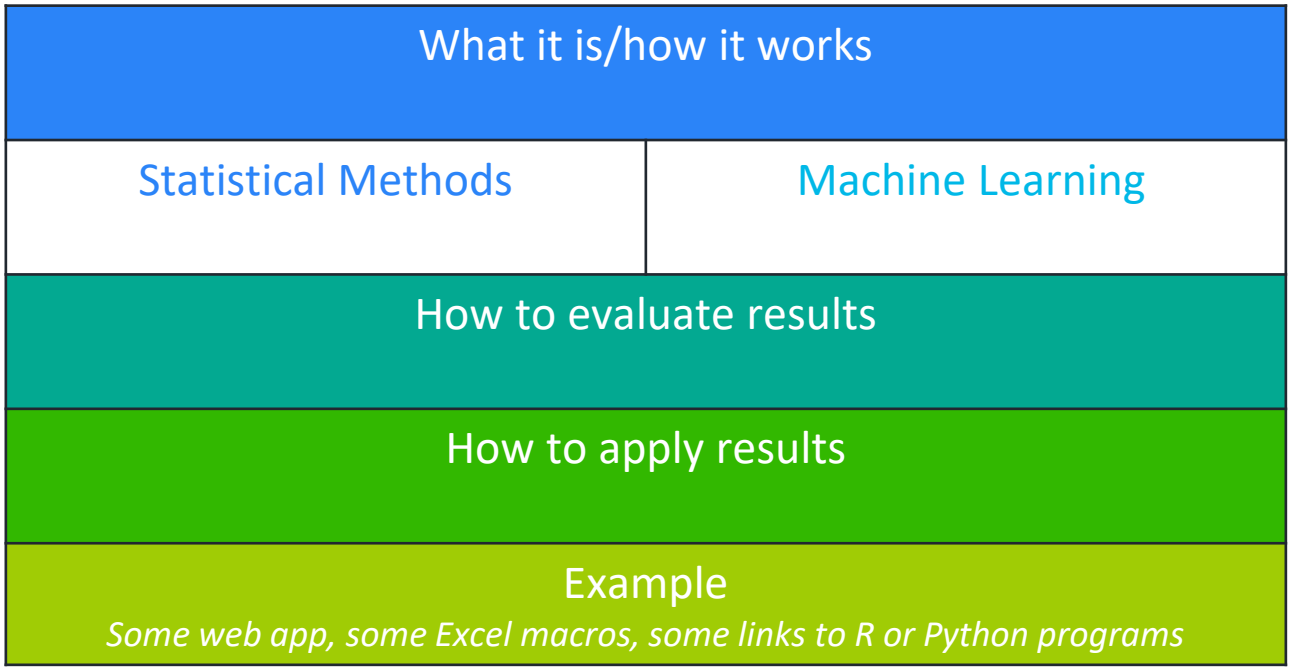
III. **Applied Descriptive Analytics**
   A. Rules-based Segmentation
   B. Cluster Analysis
   C. Principal Components Analysis (PCA)

IV. **Applied Predictive Analytics:** automated variable selection in linear regression
   A. Stepwise Regression
   B. All Possible Subsets Regression
   C. Lasso (Machine Learning)

v. Putting it all together

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

When we apply each technique we'll address:



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Applying what you've learned

A bit more background



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Predictive

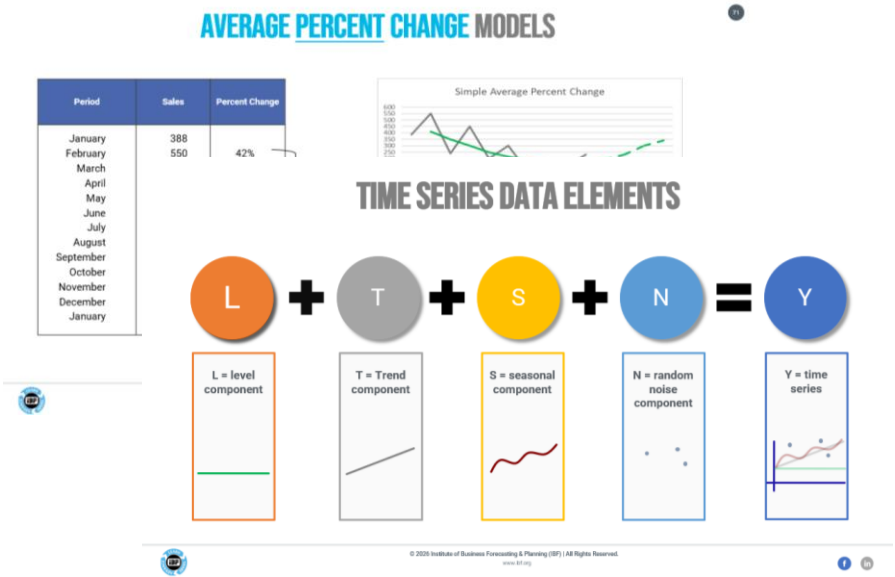
Context

Source: Wilson, Eric (IBF, 2026): ACPF Advanced Demand Planning & Forecasting

Univariate vs. Multivariate Predictive Models

Univariate

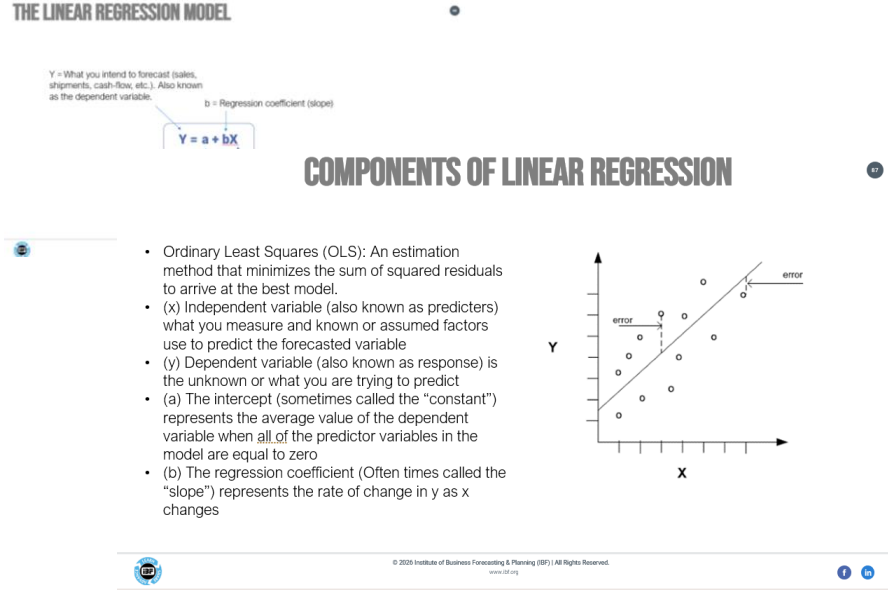
Time Series Models



How does demand track itself through time?

Multivariate

Regression Models



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Cause-and-effect models in forecasting workflows

- Many workflows for short-term forecasting rely on univariate time-series models
- When longer-term forecasts are required, cause-and-effect models may generate more reliable forecasts
- Linear regression models can be explored efficiently with variable selection algorithms
- Segmentation (rules-based or cluster analysis) can identify groups (of products or customers) that require special attention or improve groupings for top-down forecasting
- Machine learning models (e.g., neural networks, decision trees) can efficiently address nonlinear and interactive relationships between causal variables (predictors) and the result (demand forecast)

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS



The Forecasting Trap: Why Segmentation Is Your Way Out

Charles Chase
Executive Advisor/Thought Leader/Author/Speaker

October 7, 2025

The forecasting trap refers to the common mistake of applying a single forecasting model or approach across all products, regardless of their demand behavior, lifecycle stage, or data availability. It's the assumption that one method can fit all, which often leads to poor accuracy, wasted resources, and misguided decisions. Forecasting every product the same way is a common mistake that leads to inefficiency and inaccuracy. Products differ widely, some have long, stable demand histories, while others are new, volatile, or exhibit intermittent patterns.

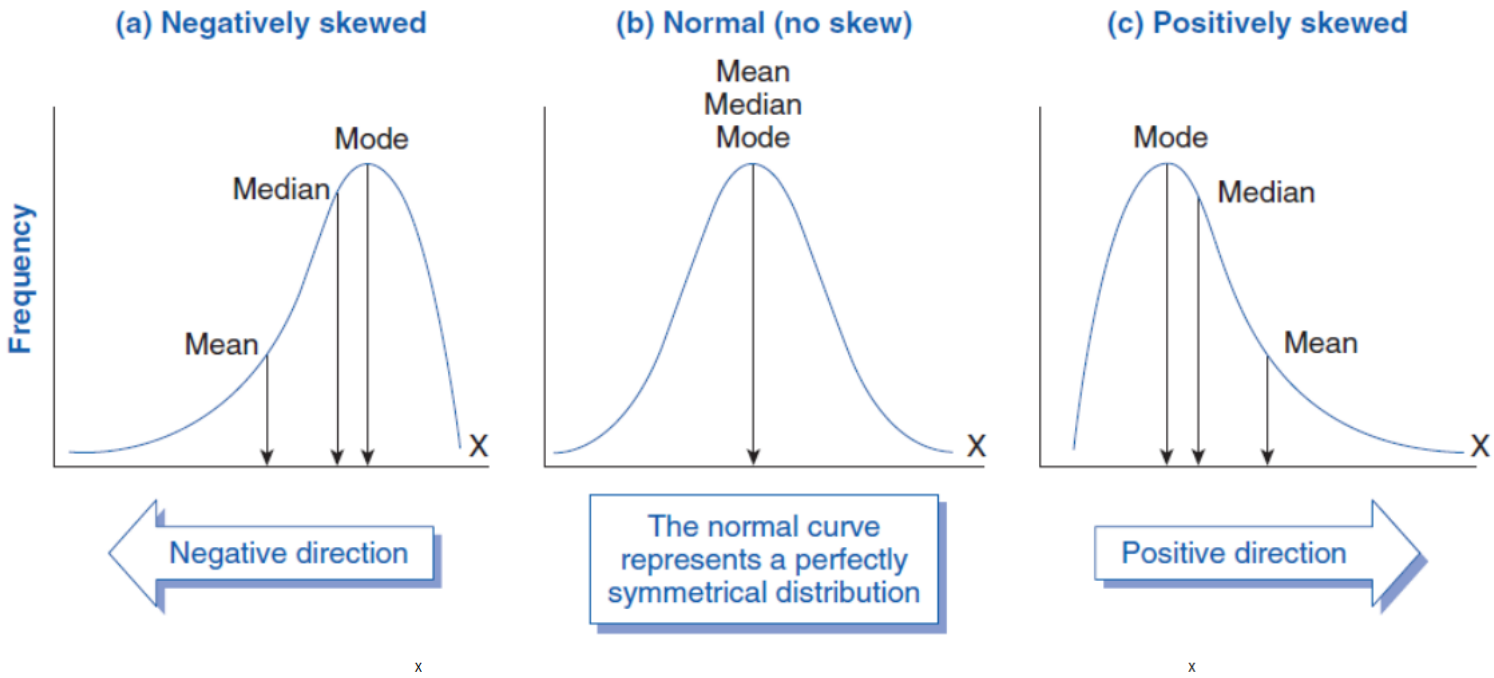
Should all products be forecasted in the same way?

"Products differ widely, some have long, stable demand histories, while others are new, volatile, or exhibit intermittent patterns...

Segmenting a product portfolio for demand forecasting is essential because it enables businesses to tailor their predictions to the unique characteristics of each product group."

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Distributions

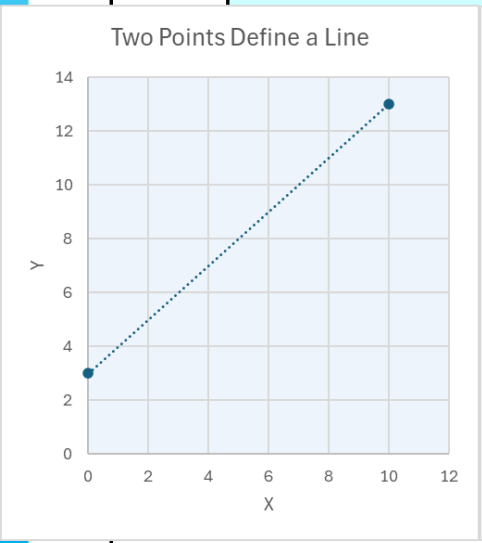


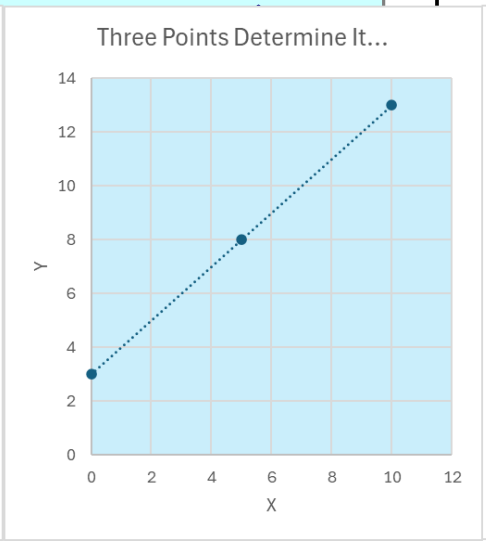
https://www.biologyforlife.com/skew.html

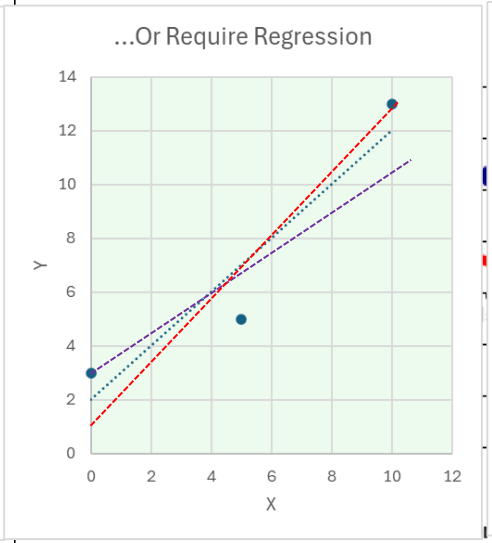
Errata: distributed copy was slightly off in boundaries for +/-σ and IQR

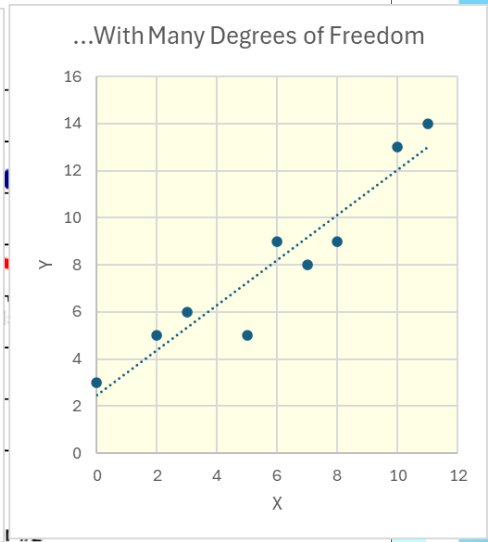
Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Regression Refresher



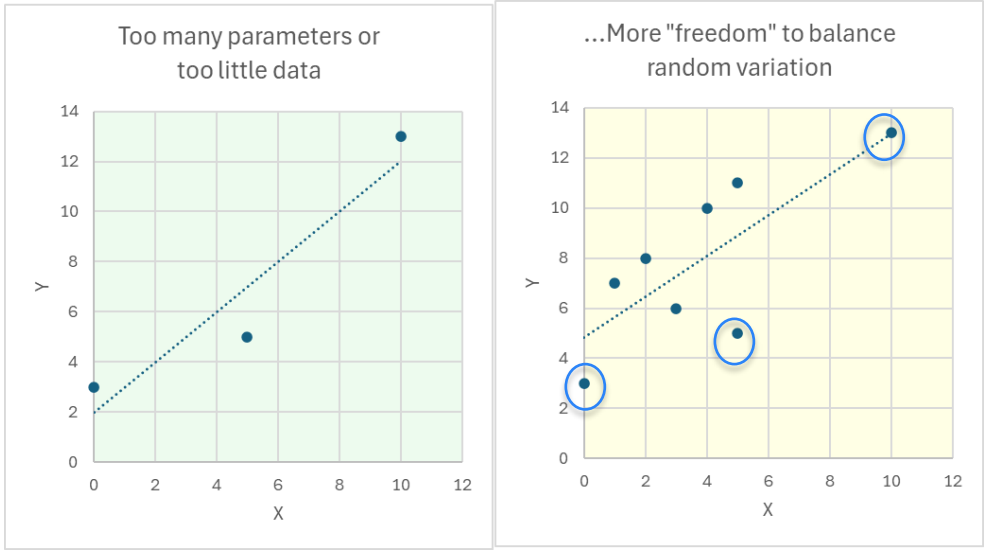






Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

What is "overfitting"



- **Mistaking random** variation **for systematic** variation
- Basing your **model** on that **mistaken "system"**
- Usually happens with **too many parameters** (relative to observations) — or—in ML, **too much "learning"**

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

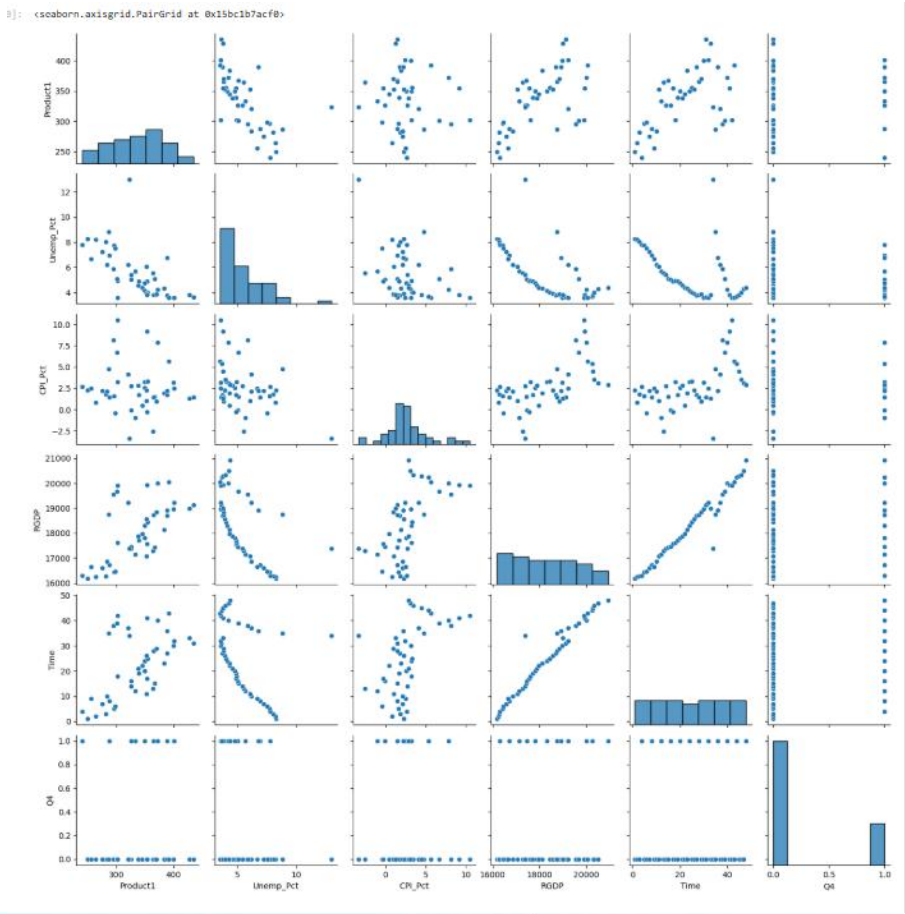
Begin from Exploratory Data Analysis (EDA)

- Conventional techniques
- LLM Contribution (ChatGPT)



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Conventional techniques include scatter plots and distributions



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Note: Use of Online LLM's

- Proprietary data should not be exposed, such as by entry or uploading, into an online tool unless company management has authorized such exposure, having made provisions to assure confidentiality of data and their analyses

- Even if de-identified, non-proprietary data are to be entered online, company policies may prohibit use of online tools, even if the IT department has not blocked access to them

In other words, if you wish to enter data online into websites like https://sbthesis.com/shiny_IBF/ or https://chatgpt.com/ , please assure that your company approves

Here's what ChatGPT advises (next slides):

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

ChatGPT advises…
(1 of 3)

New chat
Search chats
Images
Apps
Deep research
Codex
Projects

GPTs
ChatGPT Classic
Explore GPTs

Recents
Data Analysis Without LLMs
Inventory Metrics in SCP
Shiny Error Detail Options
Scatter Plots in R
Novice Guide to Forecasting
Proxy Voting Authorization ...

Sara Brumbaugh
Ceres Analytics LLC

Invite team members

If I want to analyze data but my employer hasn't authorized entry into an LLM tool like ChatGPT, what should I do?

If your employer hasn't authorized the use of an LLM tool like ChatGPT for analyzing data, the safest and most professional approach is to **avoid using it with company data until you have approval**. Many organizations restrict these tools due to **data privacy, confidentiality, regulatory, and IP risks**. Here are practical options you can consider:

1️⃣ Follow your company's policy first

- Check your **AI / data security / acceptable use policies**.
- If the policy says external AI tools are not allowed for company data, **don't upload or paste internal data** into them.
- Violating this can create **security risks and compliance issues** for the company.

2️⃣ Ask for approval or a sanctioned tool

You could ask your manager, IT, or security team:

- Whether **LLM tools are allowed for data analysis**
- Whether they have **an approved AI platform** (many companies deploy enterprise versions such as internal LLMs or secure APIs)
- Whether you can use **ChatGPT Enterprise / Azure OpenAI / internal tools**

Often companies **do allow AI tools but only through approved environments**.

3️⃣ Use traditional approved tools

Ask anything

ChatGPT can make mistakes. OpenAI doesn't use Ceres Analytics LLC workspace data to train its models.

ChatGPT advises…
(2 of 3)

New chat
Search chats
Images
Apps
Deep research
Codex
Projects

GPTs
ChatGPT Classic
Explore GPTs

Recents
Data Analysis Without LLMs
Inventory Metrics in SCP
Shiny Error Detail Options
Scatter Plots in R
Novice Guide to Forecasting
Proxy Voting Authorization ...

Sara Brumbaugh
Ceres Analytics LLC

Invite team members

3 Use traditional approved tools

If AI tools aren't approved yet, rely on tools typically allowed:

- **Python (pandas, numpy, matplotlib)**
- **R**
- **Excel / Google Sheets**
- **SQL**
- **BI tools** (Tableau, Power BI, Looker)

These can handle most analysis workflows safely inside company systems.

4 Use LLMs only for general guidance (no company data)

If policies allow it, you can still use LLMs to:

- Learn **analysis techniques**
- Ask about **coding patterns**
- Get help with **generic examples**

Example (safe):

> "How do I group by a column and calculate a moving average in pandas?"

But **don't paste real company data, queries, or proprietary schemas.**

5 Propose a safe AI workflow

If AI would meaningfully help your work, suggest a compliant solution:

- Enterprise LLM access
- A **private company-hosted model**
- **Data anonymization pipelines**
- Using **synthetic or sampled data**

Ask anything

ChatGPT can make mistakes. OpenAI doesn't use Ceres Analytics LLC workspace data to train its models.

New chat
Search chats
Images
Apps
Deep research
Codex
Projects

GPTs
ChatGPT Classic
Explore GPTs

Recents
Data Analysis Without LLMs
Inventory Metrics in SCP
Shiny Error Detail Options
Scatter Plots in R
Novice Guide to Forecasting
Proxy Voting Authorization ...

Sara Brumbaugh
Ceres Analytics LLC

Invite team members

Propose a safe AI workflow

If AI would meaningfully help your work, suggest a compliant solution:

- Enterprise LLM access
- A **private company-hosted model**
- **Data anonymization pipelines**
- Using **synthetic or sampled data**

Framing it as a **productivity improvement with proper safeguards** can help approval.

✅ **Simple rule:**

- **No approval → don't use external LLMs with company data.**
- **Use approved tools or request authorization.**

If you'd like, I can also explain:

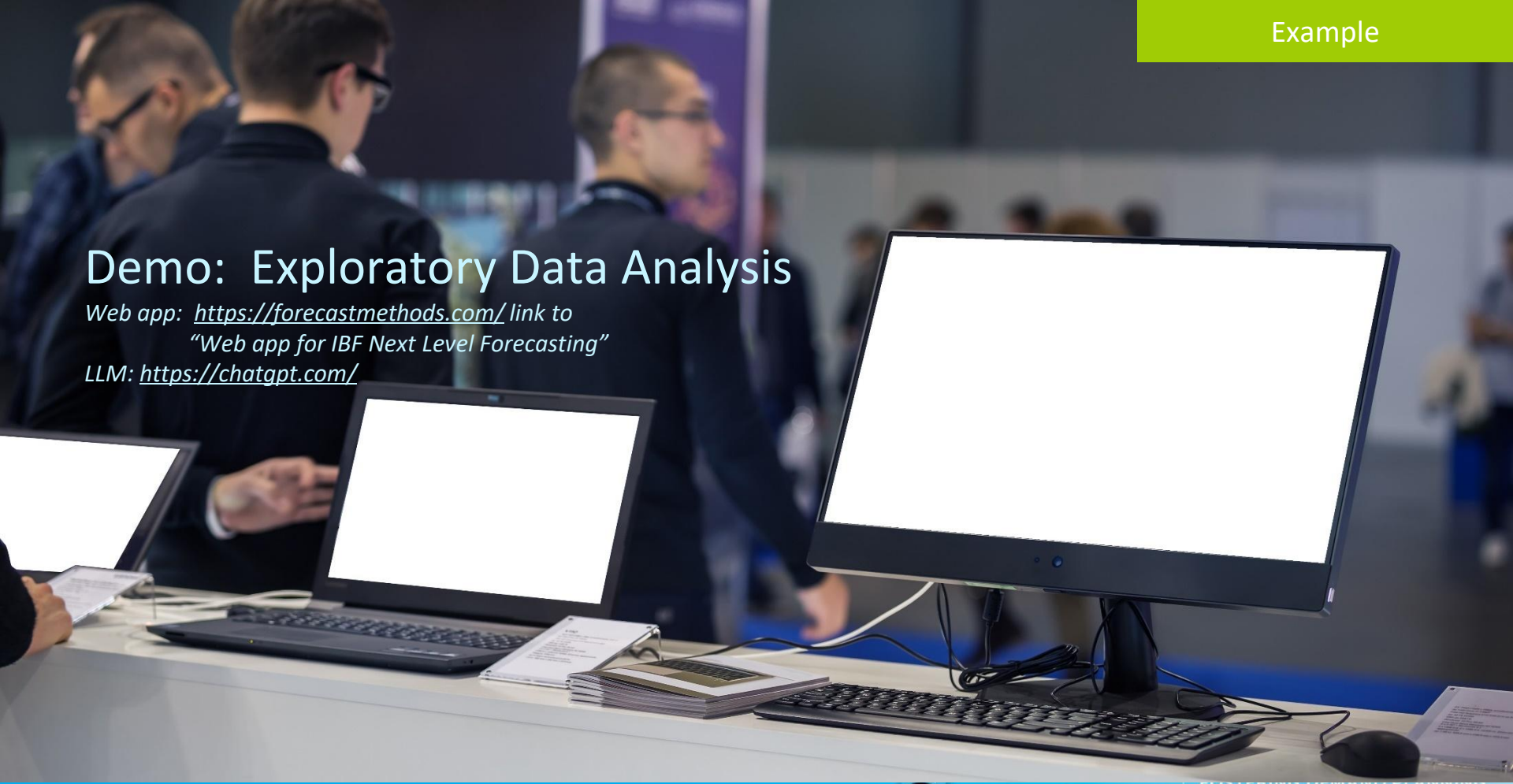
- **How companies safely deploy LLMs for internal data analysis**
- **What security teams worry about with ChatGPT**
- **Ways analysts use LLMs without exposing sensitive data.**

Ask anything

ChatGPT can make mistakes. OpenAI doesn't use Ceres Analytics LLC workspace data to train its models.

Demo: Exploratory Data Analysis

Web app: https://forecastmethods.com/ link to
"Web app for IBF Next Level Forecasting"
LLM: https://chatgpt.com/



https://forecastmethods.com/

Forecast Methods

forecastmethods.com

Forecast Methods for Workshops and Training

Developed by Ceres Analytics

Welcome!

Jupyter Python Notebooks for IBF Workshop

Web app for IBF Next Level Forecasting

Downloads to run neural networks in Excel:

- neuralNet_cfTiberius.xlsm
- HowToRun_neuralNet_cfTiberius.pptx
- neuralNet_cfTiberius_betaFC.xlsm

Downloads to run rules-based segmentation in Excel

- Excel file with macro
- Excel file without macro
- VBA code for macro
- Excel how-to

Links to reference cards ("cheat sheets") for Python and Markdown (used in Jupyter):

- DataCamp Python Cheat Sheet (PDF)
- DataCamp Python Pandas Cheat Sheet (PDF)
- DataCamp Python NumPy Cheat Sheet (PDF)
- MarkdownGuide.org Cheat Sheet

Download IBF Scottsdale script--Decision Tree with sine pattern

- sineBasedTree.py

Scripts used in IBF Next Level Forecasting

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

SEGUE…

For those who wish to code

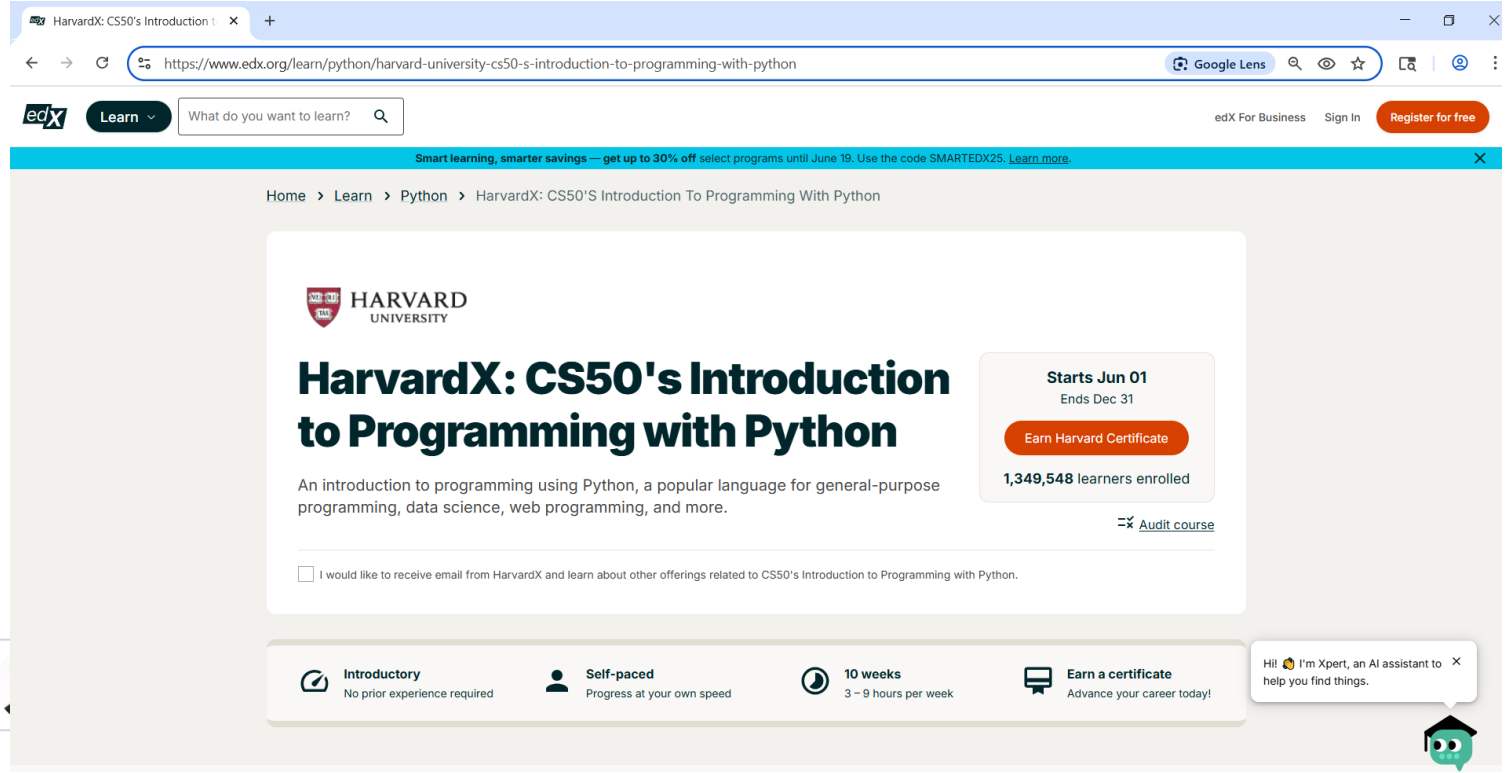


Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

If you want to learn to code

in Python, free Harvard course through edX platform

https://www.edx.org/learn/python/harvard-university-cs50-s-introduction-to-programming-with-python



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

OR begin with Python step by step with the IBF
(April 2027, Scottsdale AZ)

Forecast Methods

forecastmethods.com

Forecast Methods for Workshops and Training

Developed by Ceres Analytics

Welcome!

Jupyter Python Notebooks for IBF Workshop

Web app for IBF Next Level Forecasting

Downloads to run neural networks in Excel:

- neuralNet_cfTiberius.xlsm
- HowToRun_neuralNet_cfTiberius.pptx
- neuralNet_cfTiberius_betaFC.xlsm

Downloads to run rules-based segmentation in Excel

- Excel file with macro
- Excel file without macro
- VBA code for macro
- Excel how-to

Links to reference cards ("cheat sheets") for Python and Markdown (used in Jupyter):

- DataCamp Python Cheat Sheet (PDF)
- DataCamp Python Pandas Cheat Sheet (PDF)
- DataCamp Python NumPy Cheat Sheet (PDF)
- MarkdownGuide.org Cheat Sheet

Download IBF Scottsdale script--Decision Tree with sine pattern

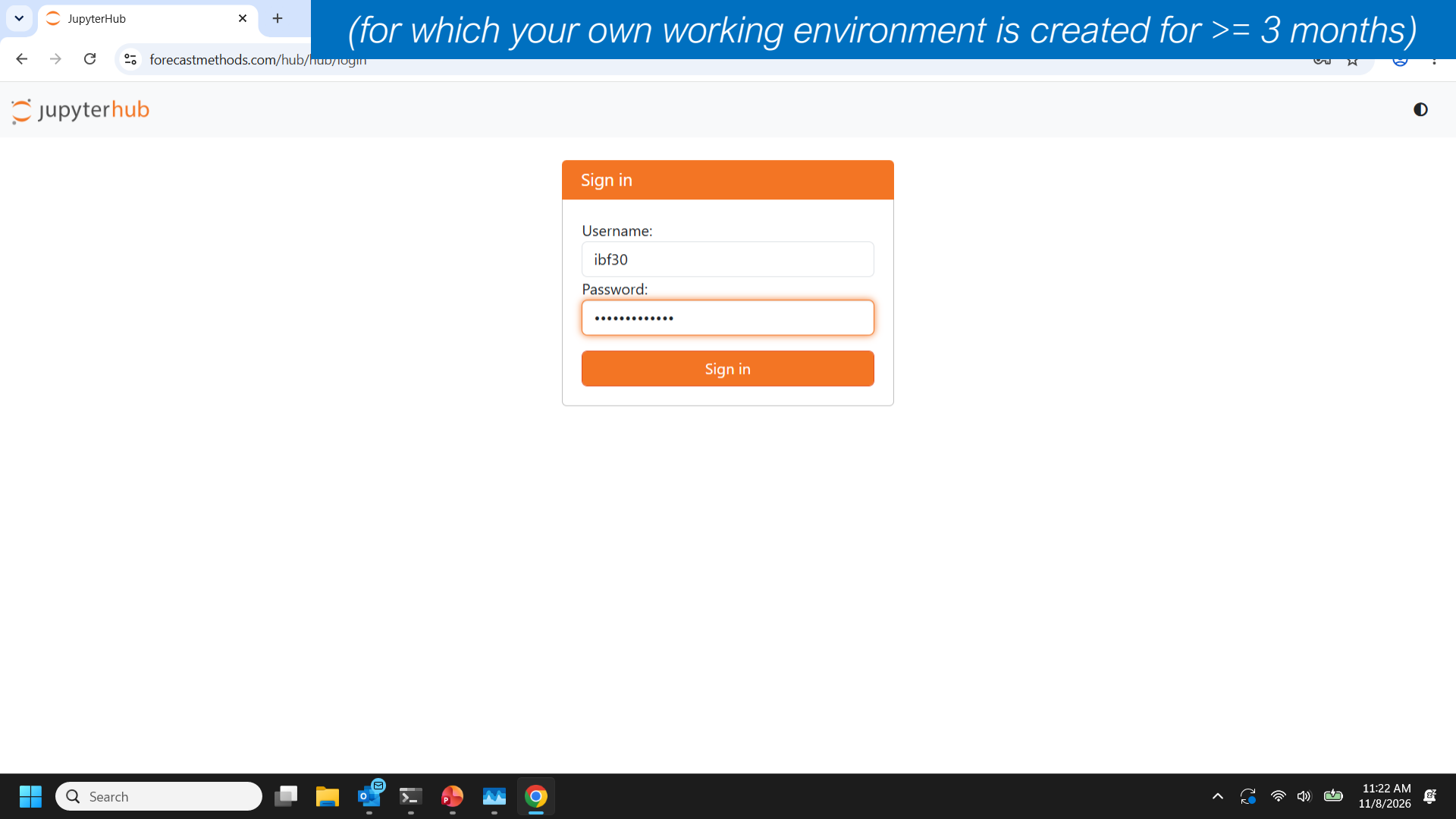
- sineBasedTree.py

Scripts used in IBF Next Level Forecasting

Search

11:13 AM
11/8/2026

(for which your own working environment is created for >= 3 months)



We start with one-line basics…

1_Functions.... (2) - JupyterLab

forecastmethods.com/hub/user/lbl50/lab/tree/_basics/1_Functions.ipynb

File Edit View Run Kernel Tabs Settings Help

/ _basics /

Name	Modified
0_YesYouCanTypeCode.ip...	6mo ago
1_Functions.ipynb	now
2_Comments_and_types.i...	6mo ago
3_Operators_and_Logic.i...	6mo ago
4_Collections_and_Loops....	6mo ago
5_Libraries_and_DataInpu...	6mo ago

Launcher | 0_YesYouCanTypeCode.ipynb | 1_Functions.ipynb

Markdown | Notebook | Python 3 (ipykernel)

A. The classic first program, "Hello, world"

Type this in the grey box below:

> *print("Hello, world")*

Then, to run it, either hit the triangular button above,
or press Shift+Enter ¶

[2]:
```
print("Hello, world")
```
```
Hello, world
```

B. Add two numbers

Type this in the grey box below,
then run it (triangular button above or Shift+Enter):

> *1 + 1*

[1]:
```
1+1
```
[1]: 2

C. Add two numbers, but assign the result to a variable

Type this in the grey box below,
then run it (triangular button above or Shift+Enter):

BTW, if you want to stay in this cell, to change numbers and re-run, press Control+Enter instead of Shift+Enter

> *x = 1 + 1*

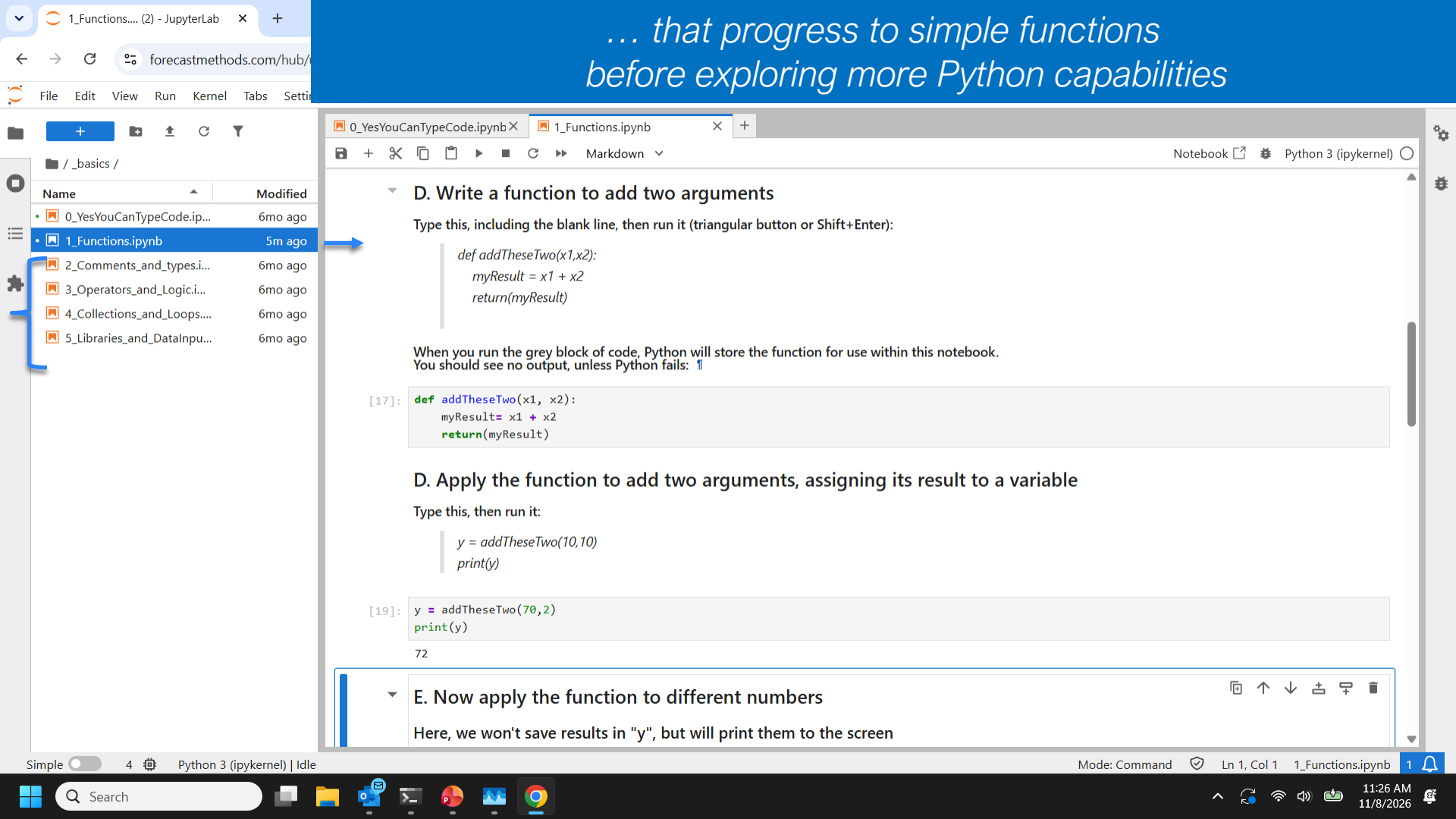
Simple 4 Python 3 (ipykernel) | Idle

Mode: Command Ln 1, Col 1 1_Functions.ipynb 1

Search

11:19 AM
11/8/2026

… that progress to simple functions before exploring more Python capabilities



we show how to import and use Python libraries for EDA before building an AI/ML neural network

Step1_ReadDa... (7) - JupyterLab

forecastmethods.com/hub/

File Edit View Run Kernel Tabs Settings

/

Name	Modified
_basics	50s ago
data	6mo ago
sarasFunctions	6mo ago
ForecastData_IBF_Produc...	6mo ago
model_fc_plot1.png	6mo ago
myProject_IBF_Product1_...	6mo ago
myProject_IBF_Product1_...	6mo ago
prep_Textbook_crimeDat...	6mo ago
prep_Textbook_forecastD...	6mo ago
Step1_ReadData_and_ED...	6mo ago
Step2_Data_Prep.ipynb	6mo ago
Step3_BuildNeuralNet.ipy...	6mo ago
Untitled.ipynb	6mo ago

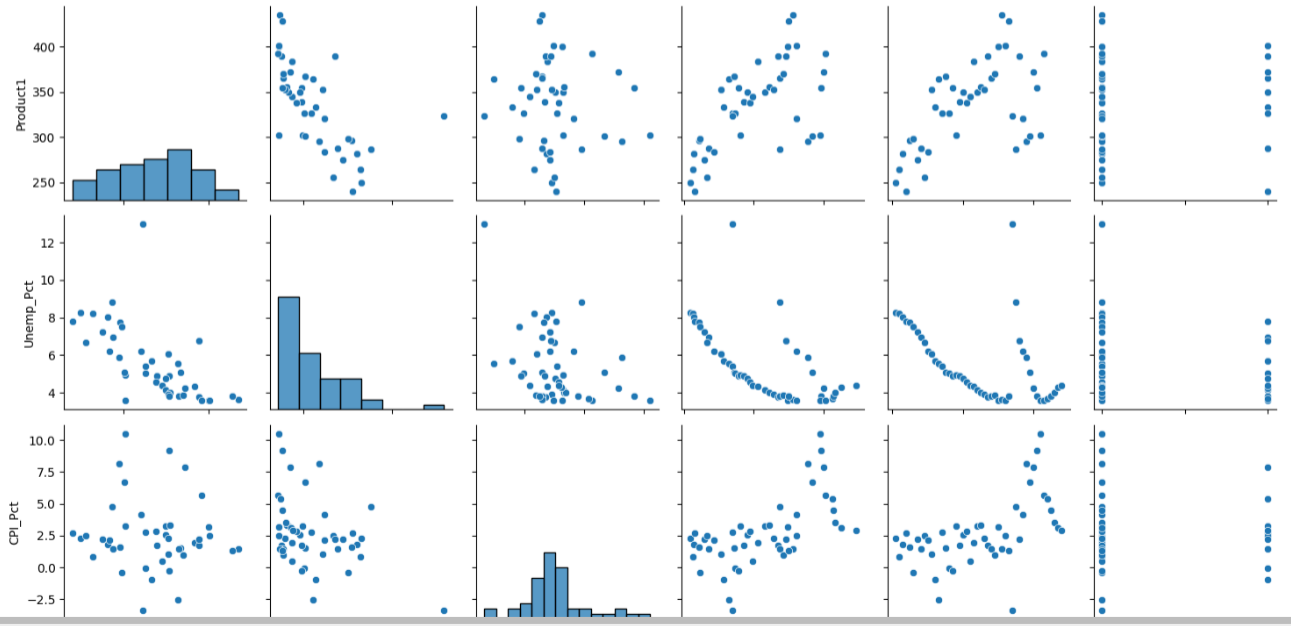
0_YesYouCanTypeC | 1_Functions.ipynb | 2_Comments_and_t | 3_Operators_and_L | 4_Collections_and_I | 5_Libraries_and_Dat | Step1_ReadData_an

Markdown | Notebook | Python 3 (ipykernel)

C. Pair plots, including dependent variable

[17]:
```
sb.pairplot(revData[[myDepVarName] + myIndepVarNames])
```

[17]: <seaborn.axisgrid.PairGrid at 0x7f904c286810>



Simple 9 Python 3 (ipykernel) | Idle

Mode: Command Ln 1, Col 1 Step1_ReadData_and_EDA.ipynb 1

Search

11:31 AM
11/8/2026

… then we build a neural network

Step3_BuildN... (8) - JupyterLab

forecastmethods.com/hub/user/ibf50/lab/tree/Step3_BuildNeuralNet.ipynb

File Edit View Run Kernel Tabs Settings Help

0_YesYouCanTyp | 1_Functions.ipyn | 2_Comments_an | 3_Operators_an | 4_Collections_an | 5_Libraries_and_ | Step1_ReadData | Step3_BuildNeu

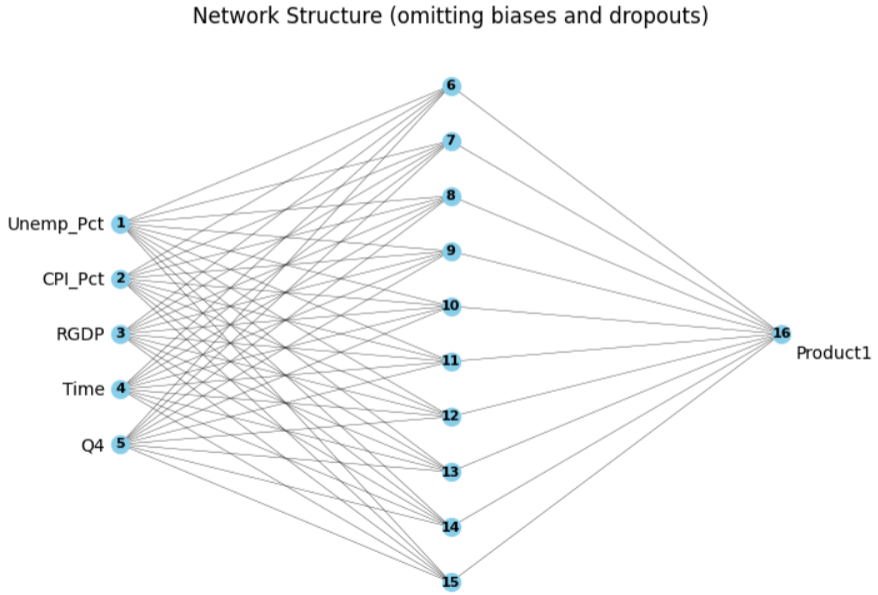
Markdown | Notebook | Python 3 (ipykernel)

/

Name	Modified
_basics	50s ago
data	6mo ago
sarasFunctions	6mo ago
ForecastData_IBF_Produc...	6mo ago
model_fc_plot1.png	6mo ago
myProject_IBF_Product1_...	6mo ago
myProject_IBF_Product1_...	6mo ago
prep_Textbook_crimeDat...	6mo ago
prep_Textbook_forecastD...	6mo ago
Step1_ReadData_and_ED...	6mo ago
Step2_Data_Prep.ipynb	6mo ago
Step3_BuildNeuralNet.ipy...	6mo ago
Untitled.ipynb	6mo ago

C. See network structure

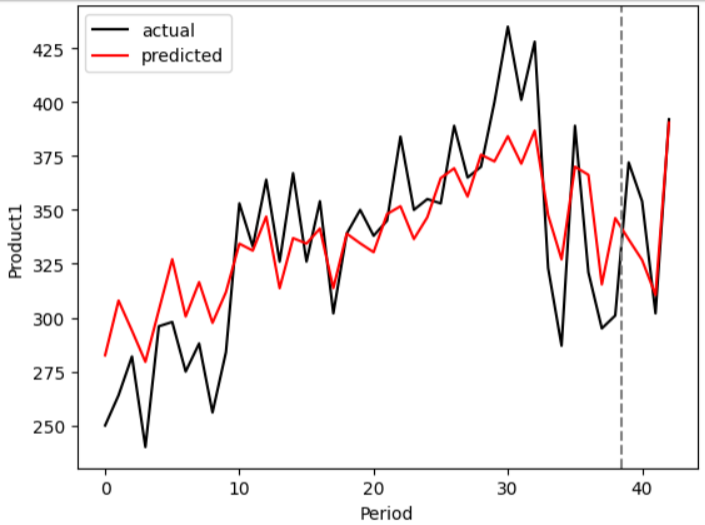
[7]:
```
drawNetworkGraph(listForGraph=myListForGraph,model=myModel,myDepVarName=myDepVarName,myIndepVarNames=myIndepVarNames,picFileName=None)
```



Simple 9 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 Step3_BuildNeuralNet.ipynb 1

Search 11:33 AM 11/8/2026

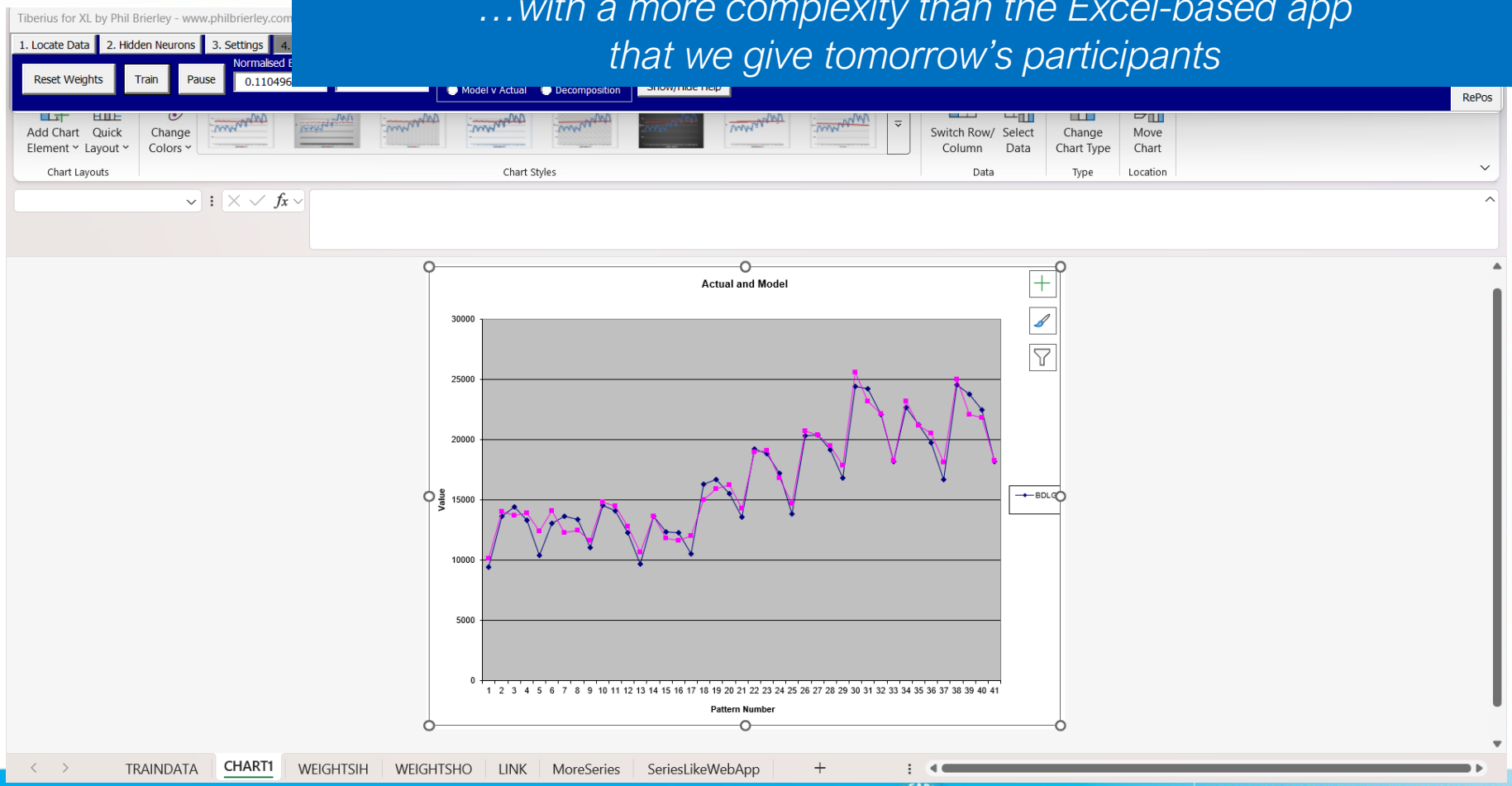
… and test its ability to forecast



YearQtr	Unemp_Pct	CPI_Pct	RGDP	Time	Q4	Product1	Product1_Pred	Error
2021Q4	4.23	7.9122	20006.2	40	1	372.0	336.284515	-35.715485
2022Q1	3.80	9.2009	19924.1	41	0	354.0	326.735229	-27.264771
2022Q2	3.60	10.5309	19895.3	42	0	302.0	310.511566	8.511566
2022Q3	3.57	5.6856	20039.4	43	0	392.0	390.454651	-1.545349

```
=== Metrics on test dataset, calculated like ChatGPT script ===
MAE : 18.2593
RMSE: 22.8790
R^2 : 0.5314
```

…with a more complexity than the Excel-based app that we give tomorrow's participants



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Applied Descriptive Analytics

- Rules-based Segmentation
- Cluster Analysis
- PCA



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Rules-based Segmentation

- *Hands on with Syntetos-Boylan mechanism*



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Rules-based Segmentation

(Syntetos-Boylan)

- **What it is/how it works**
  - Straightforward arithmetic classification of series as Intermittent, Lumpy, Smooth or Erratic
  - Statistics vs. Machine Learning—Statistics
- **How to evaluate your results**
  - Read the labels and compare time-series plots to understand them better
- **How to apply your results**
  - Apply similar forecasting methodologies within each class, e.g.:
  - Smooth may provide good univariate time-series forecasts
  - Intermittent may require methods that account for seasonality
  - Lumpy or Erratic may require cause-and-effect or machine learning methods like decision trees)
- **Examples**
  - Excel workbook

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Why would I construct product segments?

https://www.linkedin.com/pulse/forecasting-trap-why-segmentation-your-way-out-charles-chase-71vde/



The Forecasting Trap: Why Segmentation Is Your Way Out



Charles Chase
Executive Advisor/Thought Leader/Author/Speaker

October 7, 2025

The forecasting trap refers to the common mistake of applying a single forecasting model or approach across all products, regardless of their demand behavior, lifecycle stage, or data availability. It's the assumption that one method can fit all, which often leads to poor accuracy, wasted resources, and misguided decisions. Forecasting every product the same way is a common mistake that leads to inefficiency and inaccuracy. Products differ widely, some have long, stable demand histories, while others are new, volatile, or exhibit intermittent patterns.

Why would I construct product segments?

https://www.linkedin.com/pulse/forecasting-trap-why-segmentation-your-way-out-charles-chase-71vde/

Differing Historical patterns -> Differing Forecast patterns

Segmenting a product portfol
because it enables business
characteristics of each produ
seasonality, price sensitivity,
products, whether by categ
product maturity, companie
that reflect the nuances of e
size-fits-all forecasting, whic
mismanagement.

Type of forecast required

Segmentation also enhances respons
example, fast-moving consumer goo
frequency forecasts, while industrial e
term projections based on macroecor
segments, businesses can allocate res
supply chains, and reduce waste. It al
planning, as each segment can be str
different market conditions.

Strategic Importance

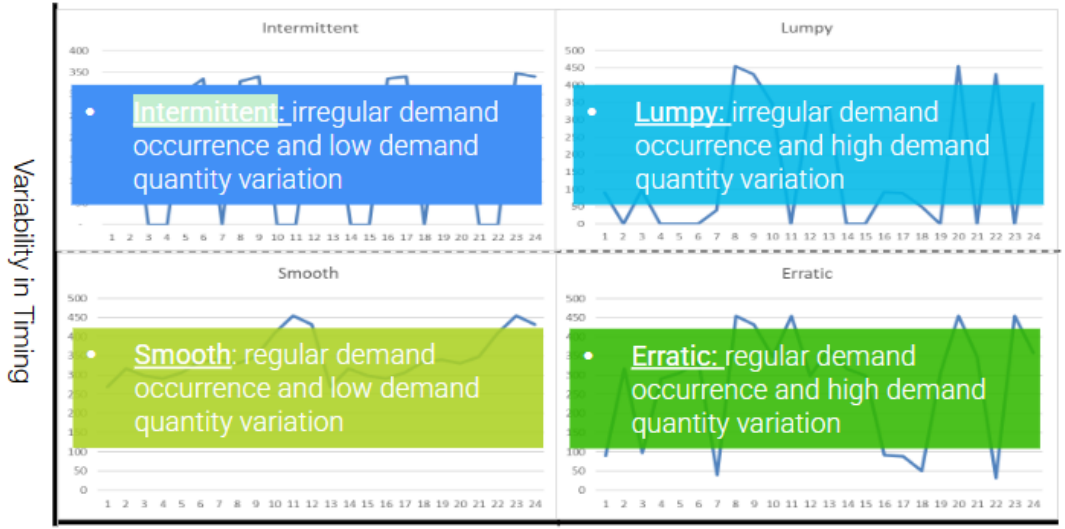
Moreover, segmentation supports strategic decision-making. It helps identify which product lines are driving growth, which are declining, and which require innovation or repositioning. This insight is invaluable for portfolio optimization, pricing strategies, and promotional planning. It also facilitates collaboration across departments, marketing, sales, and operations planning, by aligning forecasts with segment-specific goals.

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Source: Wilson, Eric (IBF, 2026): ACPF Advanced Demand Planning & Forecasting

DEMAND TYPES

Forecastability reveals the extent to which an SKU can be forecasted (average demand interval / average variation) and provides the crucial context for our forecast accuracy.



© 2026 Institute of Business Forecasting & Planning (IBF) | All Rights Reserved.
www.ibf.org

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Rules-based Segmentation, Simplified:

Demand Pattern Classification Framework based on ADI and CV²

Metric	Measures	Computed as
Average Demand Interval (ADI)	How frequently demand occurs	$\frac{Number\ of\ time\ periods}{Number\ of\ periods\ with\ demand}$
Squared Coefficient of Variation (CV^2)	Demand variability	$\left(\frac{\sigma}{\mu}\right)^2$ *NOTE: σ = standard deviation; μ = average*

Sources:

- *Wilson, Eric (IBF, 2026): ACPF Advanced Demand Planning & Forecasting, ~ slide 44*
- *Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. Journal of the operational research society, 56(5), 495-503.*

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Rules-based Segmentation, Complex:

Demand Pattern Classification Framework based on ADI and CV²

i.e.,
$\sigma >= .7 * \mu$

		Demand Variability (CV^2)	
		Low $CV^2 < 0.49$	High $CV^2 \geq 0.49$
Demand Frequency (ADI)	Frequent ADI < 1.32	SMOOTH Stable volume Frequent demand	ERRATIC Variable volume Frequent demand
	Infrequent (ADI ≥ 1.32)	ITTENT Stable volume Gaps between orders	LUMPY Variable volume Gaps between orders

i.e.,
> ¼ of time periods have non-zero demand

Source: Syntetos, A Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the operational researc* 503.

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS



Demo: Demand Pattern Classification (Syntetos-Boylan)

Excel file with macro: https://ceresanalytics.com/R_for_IBF_NextLevel/makeProductSegments_withMacro.xlsm
Excel file without macro: https://ceresanalytics.com/R_for_IBF_NextLevel/makeProductSegments.xls
VBA code for macro: https://ceresanalytics.com/R_for_IBF_NextLevel/productSegments.bas
Excel how-to: https://ceresanalytics.com/R_for_IBF_NextLevel/HowToRun_productSegments.pptx

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Cluster Analysis

- *Hands on with Ward's Hierarchical Method*



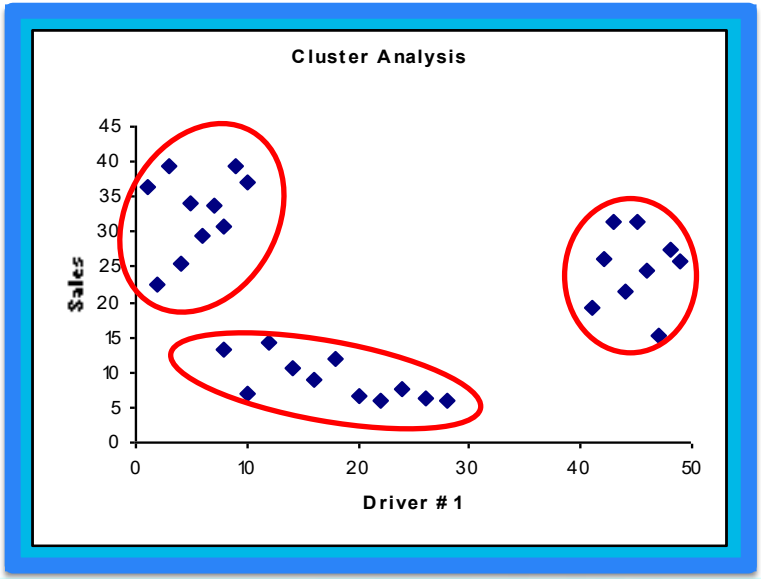
Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

How can I construct market segments?

Examples

Statistics: Cluster Analysis

Machine Learning: Cluster Analysis



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Cluster Analysis

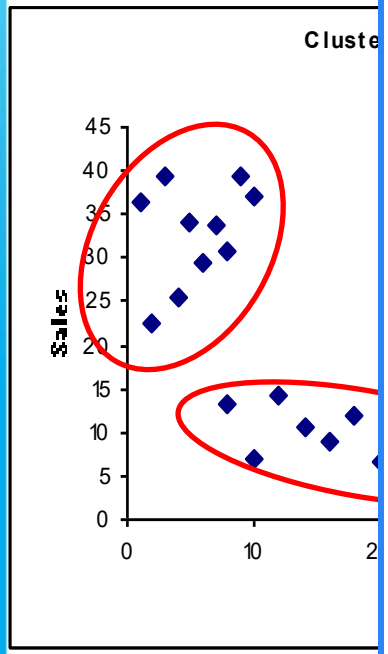
- **What it is/how it works**
  - Descriptive technique that—through either statistics or machine learning—finds groups in data for which observations are as similar as possible, while groups are as different as possible
  - Statistics vs. Machine Learning—different approaches to same task
- **How to evaluate your results**
  - Key measures, visualization (color the clusters)
- **How to apply your results**
  - Name them and learn from their profiles
  - Aggregate by segment to create class-level history and forecasts
- **Examples**
  - Web app; Scripts in Python and R

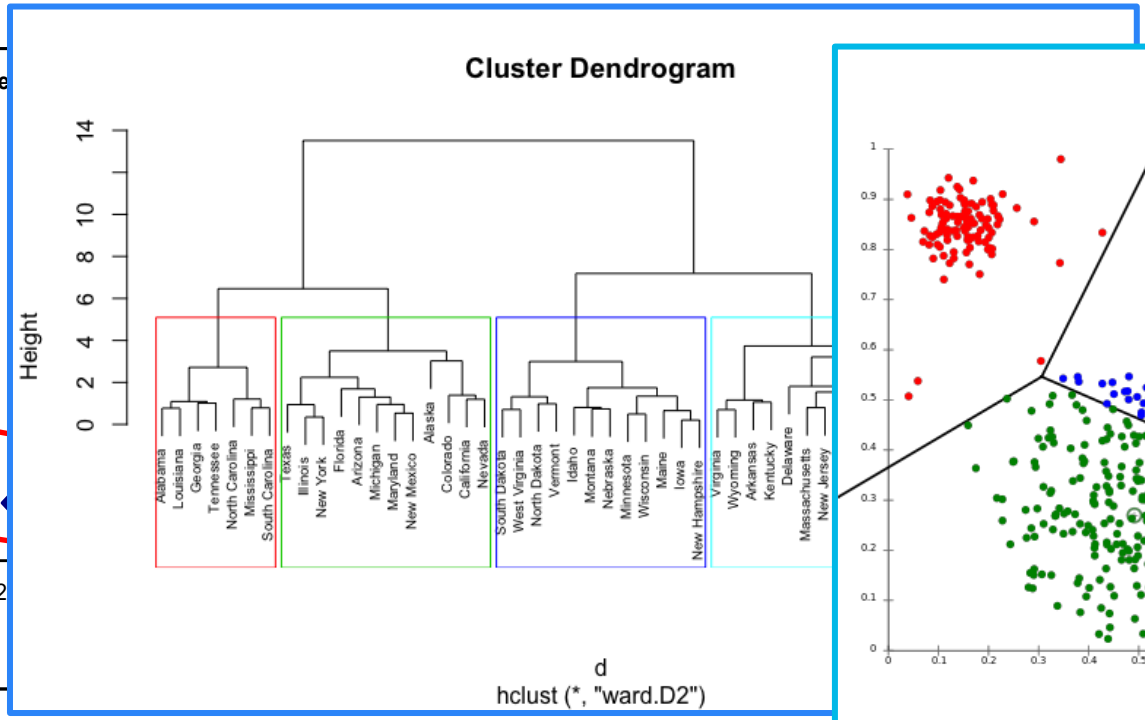
Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

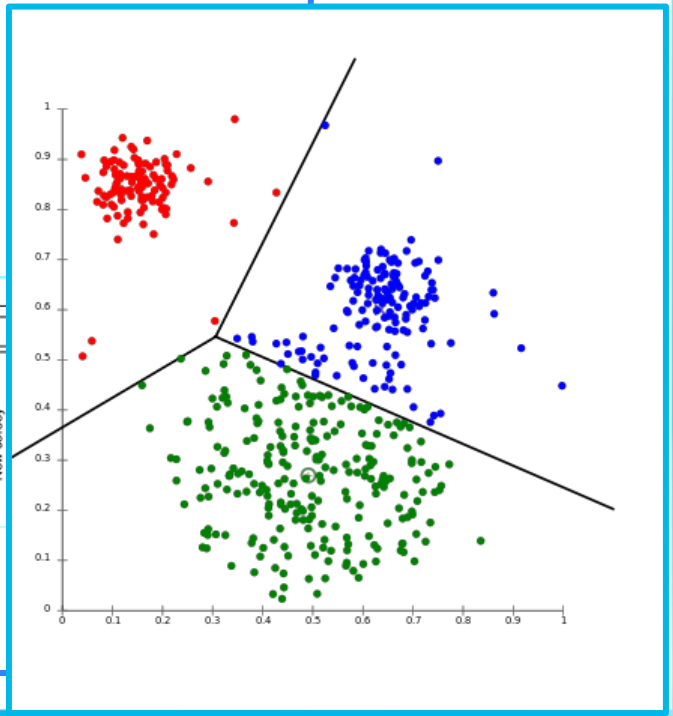
Cluster Analysis

A Generalization: Hierarchical vs. Partitioning Methods

https://stats.stackexchange.com/questions/146339/what-does-cluster-size-mean-in-context-of-k-means





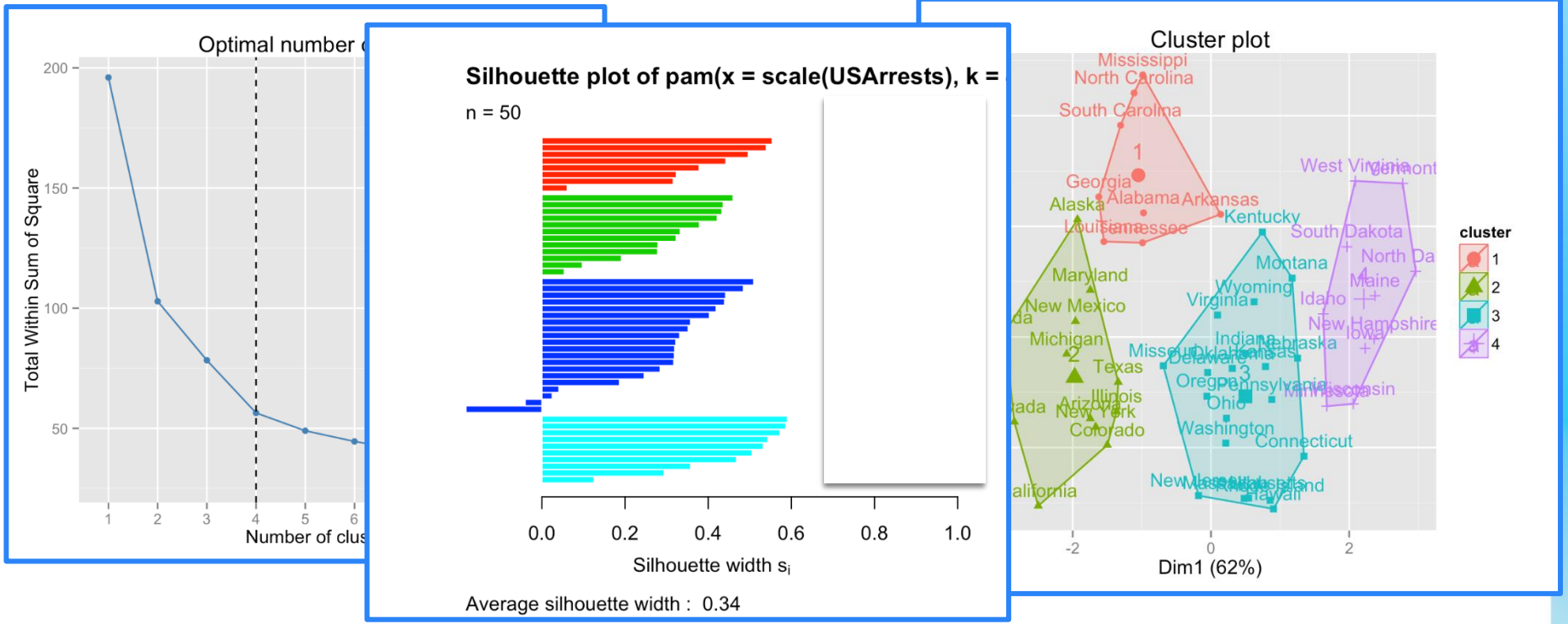


https://uc-r.github.io/hc_clustering

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Cluster Analysis

Evaluation of Results



Source (with silhouettes modified): http://www.sthda.com/english/wiki/print.php?id=236

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Cluster Analysis

Process

Concept	Considerations
Identify the purpose of the cluster analysis	• To construct new aggregates for forecasting? • To learn about/isolate group(s) from customer base? • To learn about/isolate group(s) from products?
Identify what will be clustered	• Identify **customer** segments? • Segregate **products** by key sales characteristics? • Characterize unprofitable **time** periods?
Assemble data	• One **row** per observation (e.g., customer, product) • **Columns** are measures that will define groups (e.g., RFM, inventory and profit metrics)

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

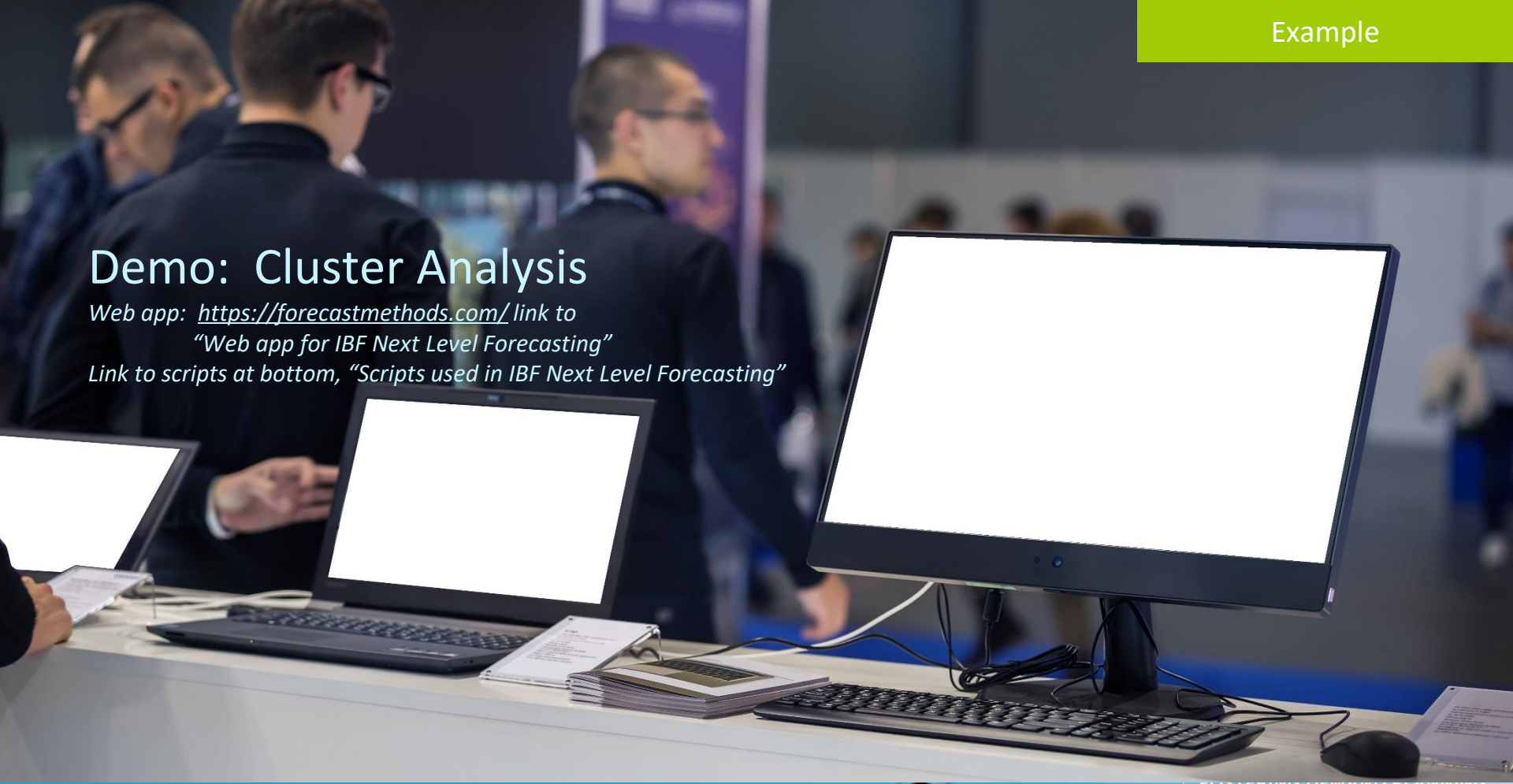
Cluster Analysis

e.g., Aggregate to class-level by assigned cluster #



Customer	Mon_AvgSales	Freq_NumPurchPastYear	Recency_TimeSinceLastPchs	OldGroupNum	OldGroup	NewCluster	NewClusterName
10000001	41	52	41	1	SouthEast	1	ThirstyThrifties
10000002	106	6	202	1	SouthEast	2	SippingSnobs
10000003	105	6	189	1	SouthEast	2	SippingSnobs
10000004	43	53	92	1	SouthEast	1	ThirstyThrifties
10000005	71	26	207	1	SouthEast	3	ImpressiveImbibers
10000006	69	26	137	1	SouthEast	3	ImpressiveImbibers
10000007	72	27	249	1	SouthEast	3	ImpressiveImbibers
10000008	70	26	174	2	NorthEast	3	ImpressiveImbibers
10000009	43	53	120	2	NorthEast	1	ThirstyThrifties
10000010	49	56	337	2	NorthEast	1	ThirstyThrifties
10000011	46	54	208	2	NorthEast	1	ThirstyThrifties
10000012	44	53	134	2	NorthEast	1	ThirstyThrifties
10000013	41	52	20	2	NorthEast	1	ThirstyThrifties
10000014	45	54	186	2	NorthEast	1	ThirstyThrifties
10000015	67	25	58	2	NorthEast	3	ImpressiveImbibers
10000016	68	25	124	2	NorthEast	3	ImpressiveImbibers
10000017	69	25	133	2	NorthEast	3	ImpressiveImbibers
10000018	71	26	224	2	NorthEast	3	ImpressiveImbibers
10000019	100	4	2	2	NorthEast	2	SippingSnobs
10000020	108	7	303	2	NorthEast	2	SippingSnobs
10000021	67	25	83	2	NorthEast	3	ImpressiveImbibers

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS



Demo: Cluster Analysis

Web app: https://forecastmethods.com/ link to
"Web app for IBF Next Level Forecasting"
Link to scripts at bottom, "Scripts used in IBF Next Level Forecasting"

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Principal Components Analysis

Data redundancy and key, underlying trends



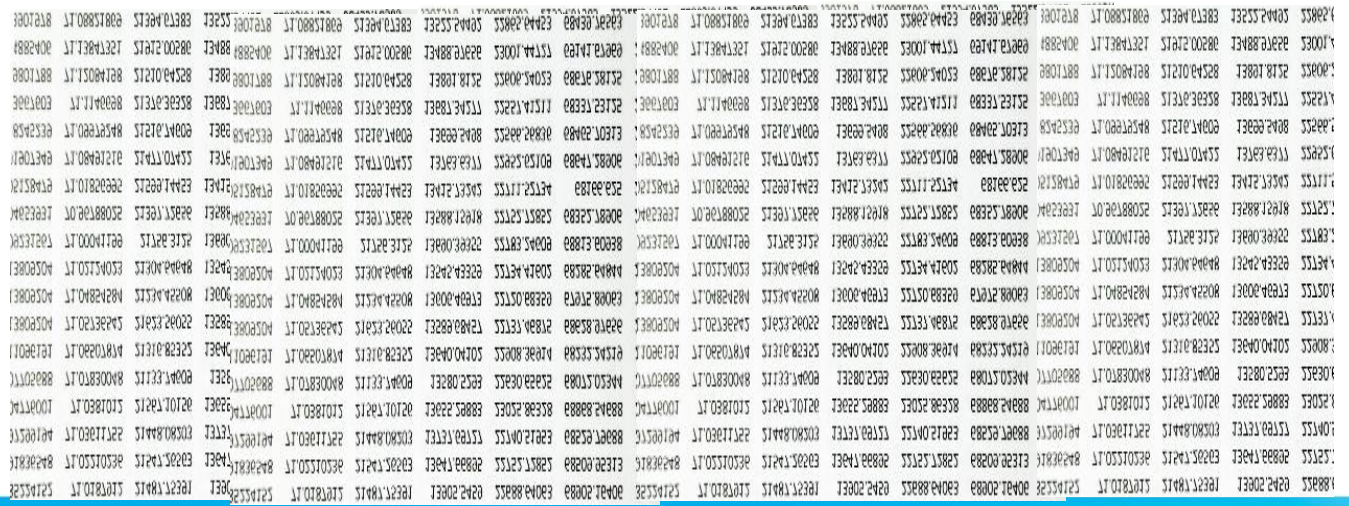
Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

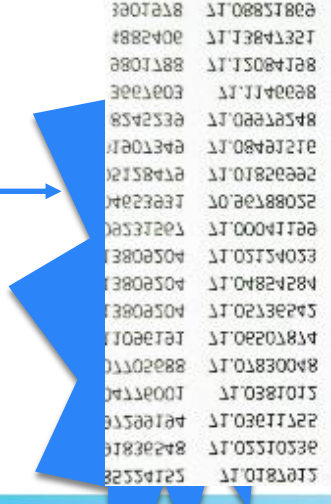
What key trends underlie all these data?

Examples

Statistics: Principal Components Analysis (PCA), Independent Components Analysis (ICA), Multi-Dimensional Scaling

Machine Learning: (t-distributed Stochastic Neighbor Embedding) (t-SNE), Uniform Manifold Approximation and Projection(UMAP)





Institute of Business Forecasting & Planning

FOSTERI DEMAND PLANNING, FORECA NG, S&OP & ANALYTI S FOR 40+ YEARS

Principal Components

- **What it is/how it works**
  - An algorithm by which, after a pool of eligible X (explanatory) variables is identified, they are weighted and combined into new explanatory variables ("Z's")
  - Statistics vs. Machine Learning—although a number of sources will classify PCA as machine learning, the PC's are computed directly through an algorithm known as "finding the root of a matrix", leading to an "eigen decomposition"
- **How to evaluate your results**
  - "Knee of curve": cumulative eigenvalues
- **How to apply your results**
  - Use top (or top few) component(s) as composite variables
  - Note which variables contribute most heavily to each component
- **Example**
  - Excel VBA script

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Principal Components

the process

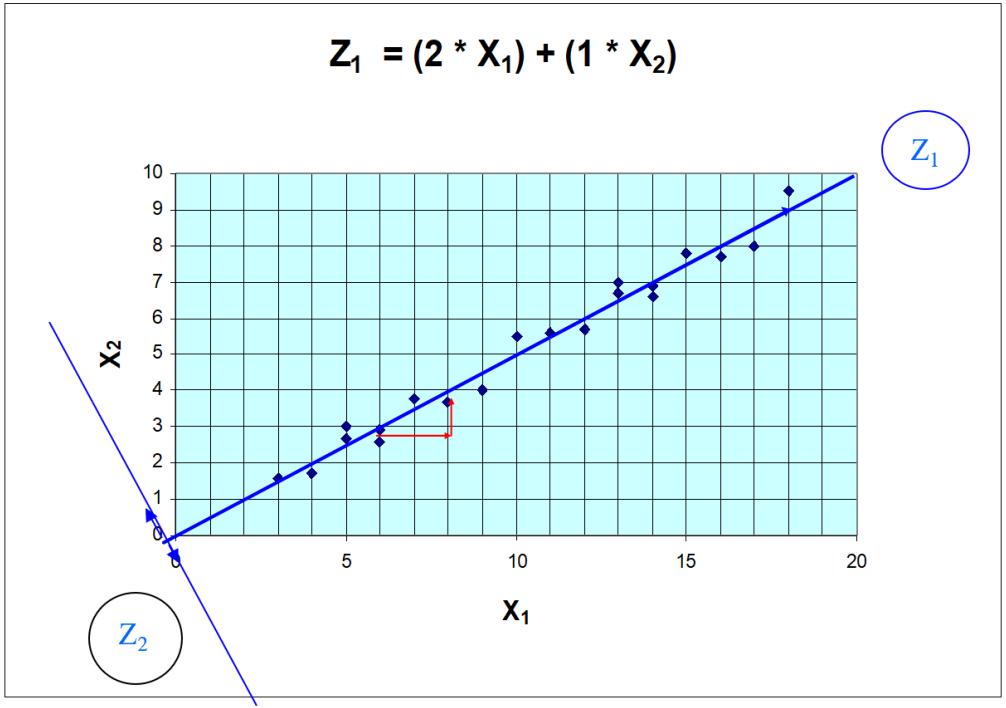
- The **computer constructs the first Z**--**the weighted combination** of X's which **captures the maximum amount of variation** within the sample
- **Logic** (why you use them to forecast): **variation in the X's** must help to **explain variation in Y**
- *after all, If Y didn't vary, forecasting would be easy: your forecast would be a constant value, such as Y's average*
- *but there's no guarantee…*
- **Subsequent combinations**, each capturing maximum movement, while **remaining uncorrelated** with other new Z's, are created
- Usually the first **three to five** new Z's capture most of the sample variation

https://uc-r.github.io/hc_clustering

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Principal Components

- In this simple example based on **two predictors** (X_1 and X_2), **Z_1 captures most of the variation** in the two X's, along with their **correlation**
- In more advanced applications, **factor rotation** re-expresses the **loading of X_1 and X_2** on the new **axes**



https://uc-r.github.io/hc_clustering

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

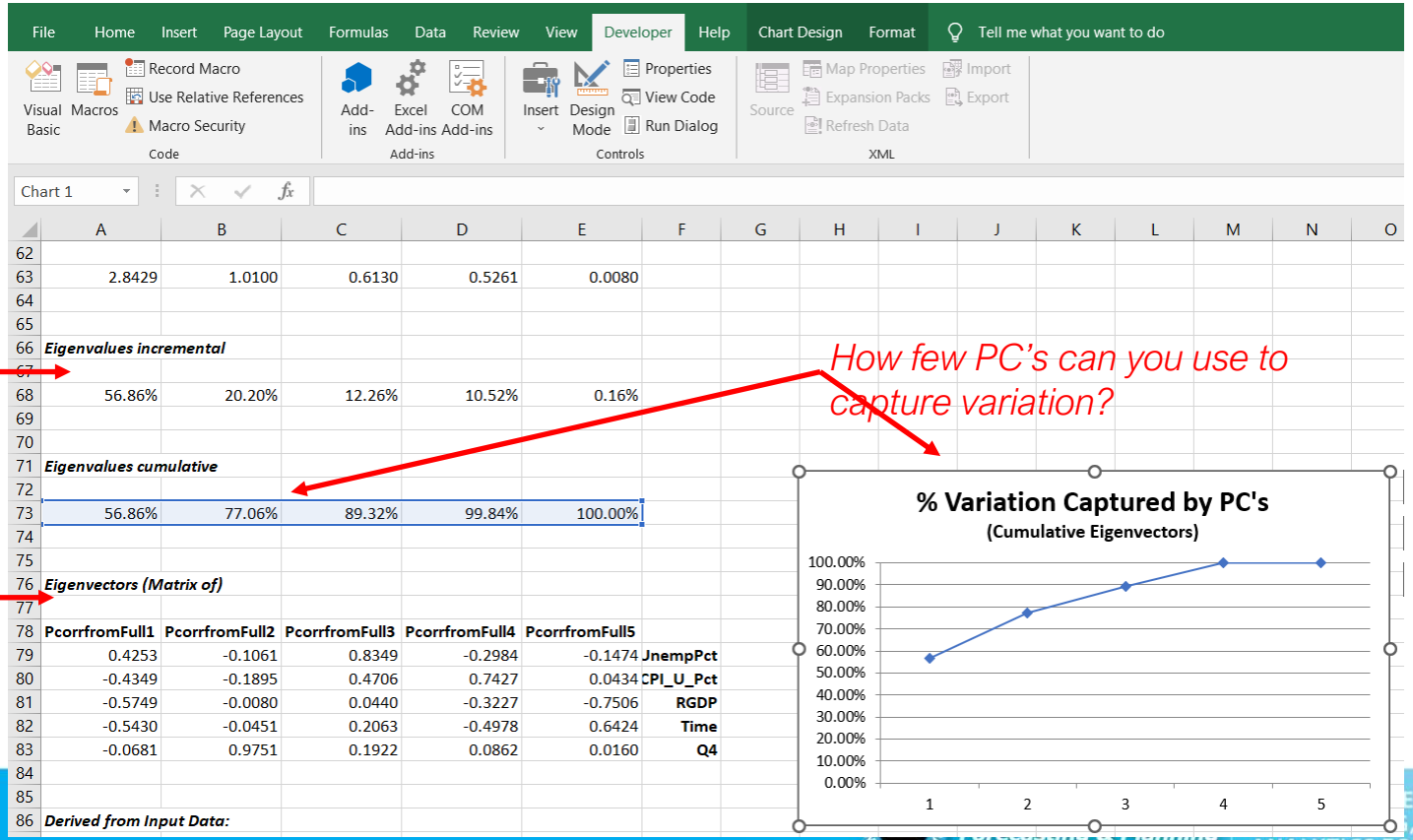
Principal Components

Evaluation of Results

How much variation is "packed" into each new Z

How are original X's weighted in new Z's?

How few PC's can you use to capture variation?



EMAND PLANNING, , S&OP & ANALYTICS FOR 40+ YEARS

Principal Components

Use one or more PC's ("Z's") in place of explanatory variables (X's]

The **construction of composite variables**, through PCA, is most useful when:

- there are **too many explanatory variables** and too few observations
- a number of explanatory variables are **similar in concept**
- **multicollinearity** confounds use of separate predictors
- You're going to be **data mining** and would like to know how much redundancy is in your data

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

How to apply it

Principal Components

Use one or more PC's ("Z's") in place of explanatory variables (X's]

	A	B	C	D	E	F	G	H
1	Year	Quarter	YearQuart	TIME	WASA	EMPL	QTR_4	BLDG
2	1	1	Y01Q1	1	1193.3	87973	0	9392
3	1	2	Y01Q2	2	1217.3	90021	0	13637
4	1	3	Y01Q3	3	1254.4	90120	0	14392
5	1	4	Y01Q4	4	1284.7	91180	1	13301
6	2	1	Y02Q1	5	1319.8	89832	0	10411
7	2	2	Y02Q2	6	1335.1	90668	0	13057
8	2	3	Y02Q3	7	1360.1	89850	0	13621
9	2	4	Y02Q4	8	1411.6	91276	1	13354
10	3	1	Y03Q1	9	1451.7	89964	0	11056
11	3	2	Y03Q2	10	1478.1	91566	0	14559
12	3	3	Y03Q3	11	1512.6	91405	0	14067
13	3	4	Y03Q4	12	1530.6	91691	1	12286
14	4	1	Y04Q1	13	1542.7	89341	0	9658
15	4	2	Y04Q2	14	1563.9	90310	0	13659
16	4	3	Y04Q3	15	1579.8	89324	0	12351
17	4	4	Y04Q4	16	1586	89436	1	12307
18	5	1	Y05Q1	17	1610.7	88815	0	10508
19	5	2	Y05Q2	18	1643.5	89920	0	16287
20	5	3	Y05Q3	19	1671.3	90480	0	16697
21	5	4	Y05Q4	20	1715.4	92367	1	15523
22	6	1	Y06Q1	21	1755.8	91637	0	13540
23	6	2	Y06Q2	22	1793.1	94256	0	19190
24	6	3	Y06Q3	23	1819.5	95021	0	18827
25	6	4	Y06Q4	24	1847.7	96547	1	17215
26	7	1	Y07Q1	25	1882.7	95450	0	13848
27	7	2	Y07Q2	26	1939.8	97657	0	20319
28	7	3	Y07Q3	27	1976	97985	0	20403
29	7	4	Y07Q4	28	2012.8	99383	1	19179

Prin Comp demo CORR | Prin Comp demo VCOV | EDA_timeSeries



	A	B	C	D
1	Pcorrdemo1	Pcorrdemo2	Pcorrdemo3	Pcorrdemo4
2	-2.6069	-0.3871	0.3173	0.0238
3	-2.3283	-0.3921	0.5308	-0.0421
4	-2.2157	-0.4021	0.4760	-0.0220
5	-1.8817	1.8829	0.4325	-0.0187
6	-2.0497	-0.4216	0.3068	0.0157
7	-1.8993	-0.4273	0.3636	-0.0301
8	-1.8919	-0.4370	0.1935	-0.0057
9	-1.4916	1.8455	0.1862	0.0284
10	-1.6487	-0.4597	0.0623	0.0715
11	-1.4091	-0.4657	0.2148	0.0241
12	-1.3250	-0.4758	0.1268	0.0471
13	-1.0829	1.8097	-0.0128	0.0499
14	-1.3803	-0.4930	-0.2586	0.0697
15	-1.2085	-0.4993	-0.1876	0.0315
16	-1.2306	-0.5080	-0.3747	0.0430
17	-1.0226	1.7789	-0.5305	0.0279
18	-1.1368	-0.5229	-0.5528	0.0193
19	-0.9349	-0.5307	-0.4705	-0.0001
20	-0.7922	-0.5386	-0.4581	-0.0127
21	-0.3591	1.7458	-0.3994	-0.0075
22	-0.4604	-0.5585	-0.4457	0.0183
23	-0.1079	-0.5645	-0.1638	-0.0389
24	0.0522	-0.5719	-0.1232	-0.0605
25	0.4274	1.7142	-0.1032	-0.0757
26	0.2831	-0.5899	-0.1953	-0.0492
27	0.6260	-0.5994	0.0197	-0.0545
28	0.7590	-0.6089	-0.0040	-0.0433
29	1.1348	1.6757	-0.0062	-0.0375

Prin Comp demo CORR | Prin Comp demo VC

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS



Demo: Principal Components

https://ceresanalytics.com/R_for_IBF_NextLevel/PrincipalComponents.bas
https://ceresanalytics.com/R_for_IBF_NextLevel/HowToRunPCA_inExcel.pptx
https://ceresanalytics.com/R_for_IBF_NextLevel/EDA_timeSeries.txt

Institute of Business Forecasting & Planning
FORECASTING, S&OP &
ANALYTICS FOR 40+ YEARS

How much variation is packed into each component?

How much variation is packed into first "n" components?

How is each "X" variable weighted in each new "Z" component?

- *E.g., first PC weights Time, WASA, and Empl roughly equally*

EDA_timeSeries_withMacro - Excel

Chart Tools

File Home Insert Page Layout Formulas Data Review View Developer Help Chart Design Format Tell me what you want to do

Visual Basic | Macros | Record Macro | Use Relative References | Macro Security — Code

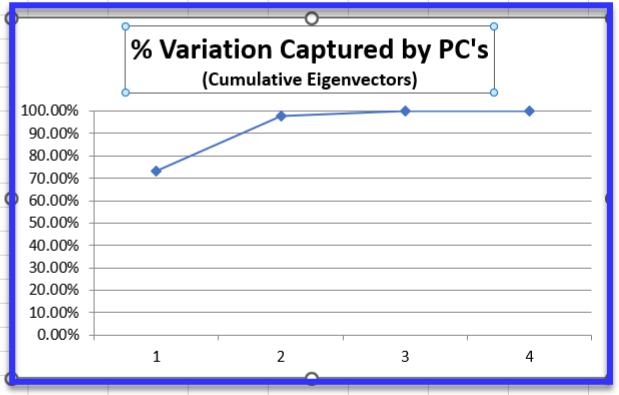
Add-ins | Excel Add-ins | COM Add-ins — Add-ins

Insert | Design Mode | Properties | View Code | Run Dialog — Controls

Source | Map Properties | Expansion Packs | Refresh Data | Import | Export — XML

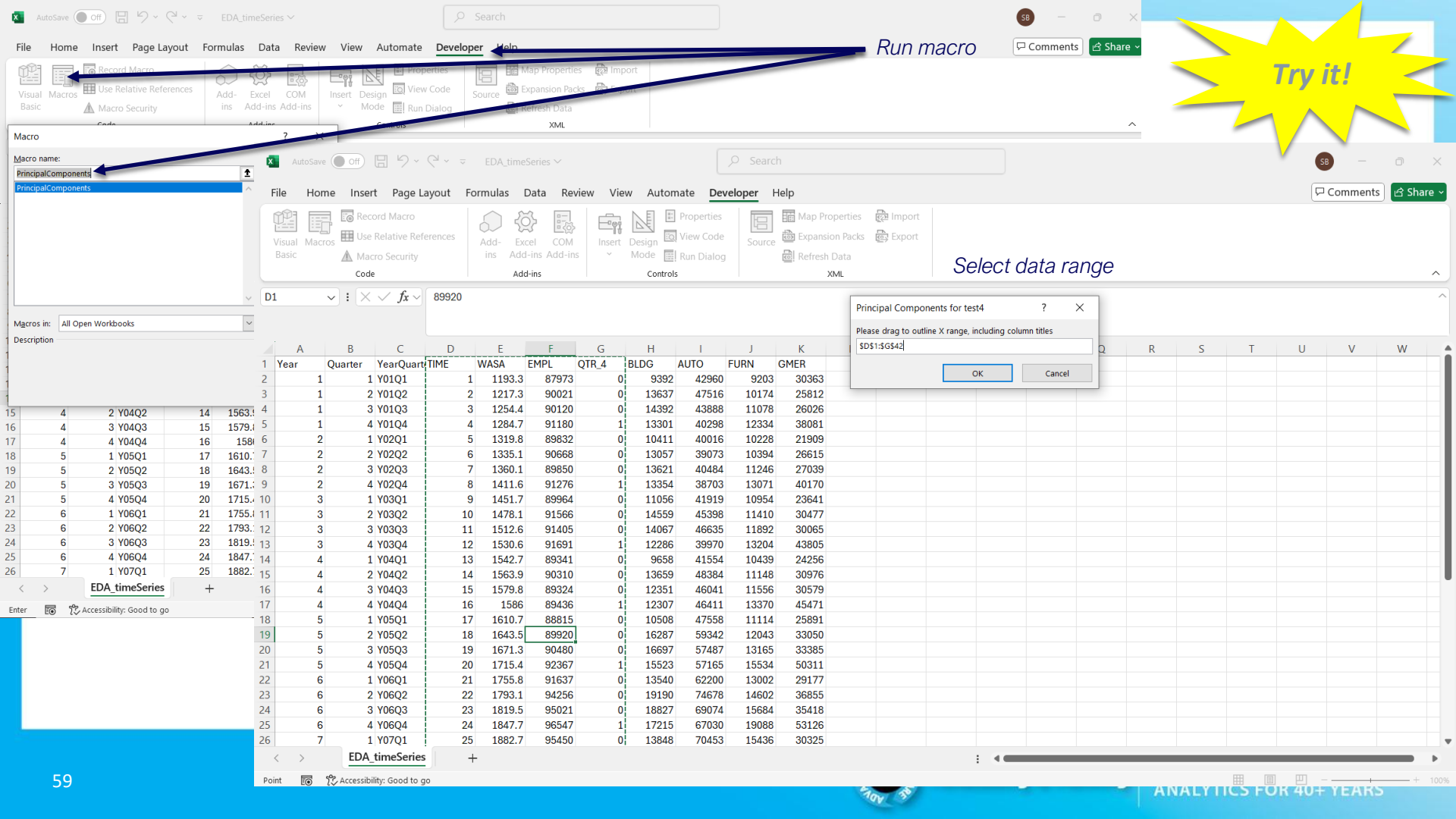
Chart 1

	A	B	C	D
54				
55	2.9229	0.9955	0.0792	0.0024
56				
57				
58	***Eigenvalues incremental***			
59				
60	73.07%	24.89%	1.98%	0.06%
61				
62				
63	***Eigenvalues cumulative***			
64				
65	73.07%	97.96%	99.94%	100.00%
66				
67				
68	***Eigenvectors (Matrix of)***			
69				
70	**Pcorrdemo1**	**Pcorrdemo2**	**Pcorrdemo3**	**Pcorrdemo4**	
71	0.5767	-0.0573	-0.5487	-0.6026	**TIME**
72	0.5823	-0.0569	-0.2346	0.7763	**WASA**
73	0.5698	0.0097	0.8007	-0.1847	**EMPL**
74	0.0609	0.9967	-0.0527	0.0115	**QTR_4**
75				
76				
77	***Derived from Input Data:***			
78				
79	**TIME**	**WASA**	**EMPL**	**QTR_4**
80	1	1193.3	87973	0
81	2	1217.3	90021	0
82	3	1254.4	90120	0



Prin Comp demo CORR | Prin Comp demo VCOV | EDA_timeSeries

Institute of Business Forecasting & Planning
FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS


Run macro
Try it!
Select data range
Macro
Macro name:
PrincipalComponents
Macros in:
All Open Workbooks
Description
Principal Components for test4
Please drag to outline X range, including column titles
$D$1:$G$42
OK
Cancel
EDA_timeSeries

AutoSave Off | EDA_timeSeries | Search

File Home Insert Page Layout Formulas Data Review View Automate Developer Help

Visual Basic | Macros | Record Macro | Use Relative References | Macro Security (Code)
Add-ins | Excel Add-ins | COM Add-ins (Add-ins)
Insert | Design Mode | Properties | View Code | Run Dialog (Controls)
Source | Map Properties | Expansion Packs | Refresh Data | Import | Export (XML)

Try it!

V66

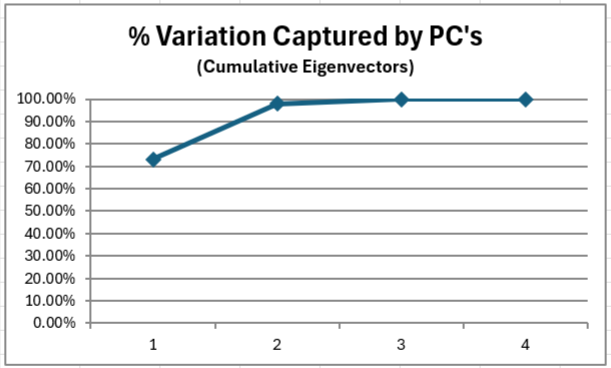
Almost all variation from 4 predictive variables can be captured in 2 composites

Eigenvalues incremental

73.07%	24.89%	1.98%	0.06%

Eigenvalues cumulative

73.07%	97.96%	99.94%	100.00%



Eigenvectors (Matrix of)

Pcorrtest41	Pcorrtest42	Pcorrtest43	Pcorrtest44	
0.5767	-0.0573	-0.5487	-0.6026	**TIME**
0.5823	-0.0569	-0.2346	0.7763	**WASA**
0.5698	0.0097	0.8007	-0.1847	**EMPL**
0.0609	0.9967	-0.0527	0.0115	**QTR_4**

Correlated predictors have nearly-equal weights in primary composite

Derived from Input Data:

TIME	WASA	EMPL	QTR_4
1	1193.3	87973	0
2	1217.3	90021	0

The uncorrelated predictor "stands alone" with nearly full weight (~1.0) while the correlated predictors' weights approach 0

Prin Comp test4 CORR | Prin Comp test4 VCOV | EDA_timeSeries

Ready | Accessibility: Investigate | 100%

Forecasting & Planning
ANALYTICS FOR 40+ YEARS

Applied Predictive Analytics

- Advanced Regression Concepts
- Automated Linear Regression: Selection of Causal Drivers



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Advanced Regression Concepts

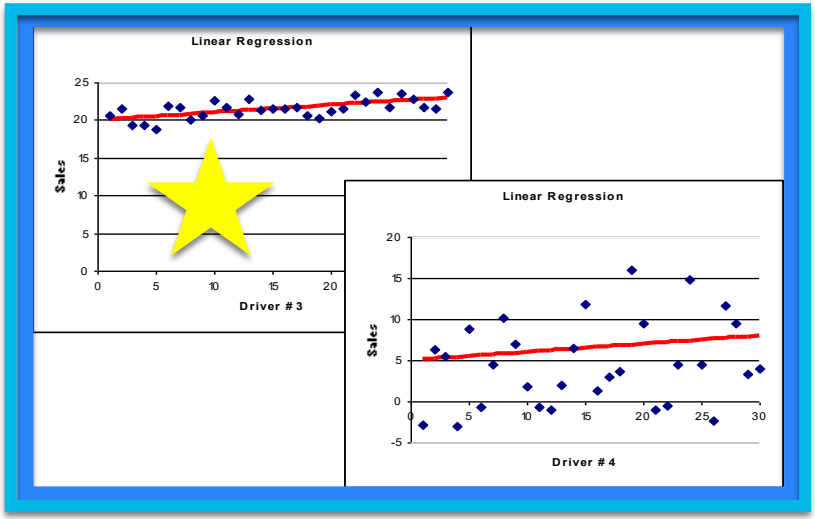


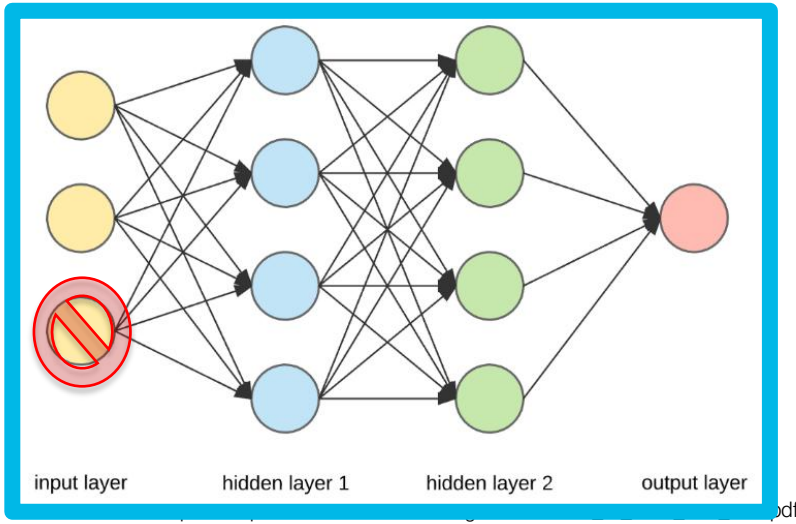
Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

How do I know what will drive a result?

Examples

Statistics: Automated variable selection (stepwise, all possible subsets)
Machine Learning: Filter Methods, Wrapper Methods, Embedded Methods





Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

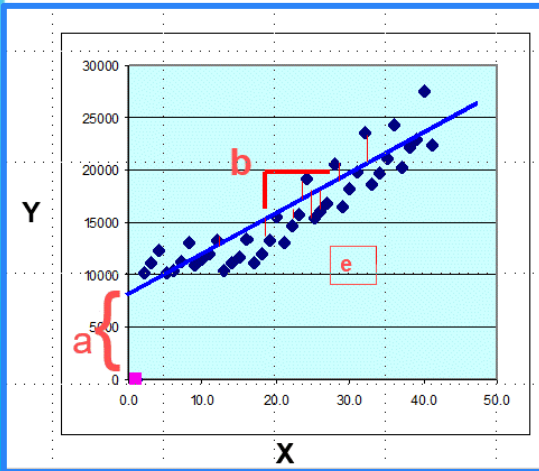
Linear Regression Analysis

- **What it is/how it works**
  - Predictive technique that articulates a linear relationship between predictors (independent variables) and the level of an outcome (dependent variable)
  - Statistics vs. Machine Learning—variable selection algorithms, optimization algorithms
- **How to evaluate your results**
  - R^2, Adjusted R^2, t-statistics, outliers, RMSE, ex-post forecasts
- **How to apply your results**
  - Extrapolate with equation
- **Example**
  - R (today), Excel

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Linear Regression

Estimate an equation to predict Y from one or more X's

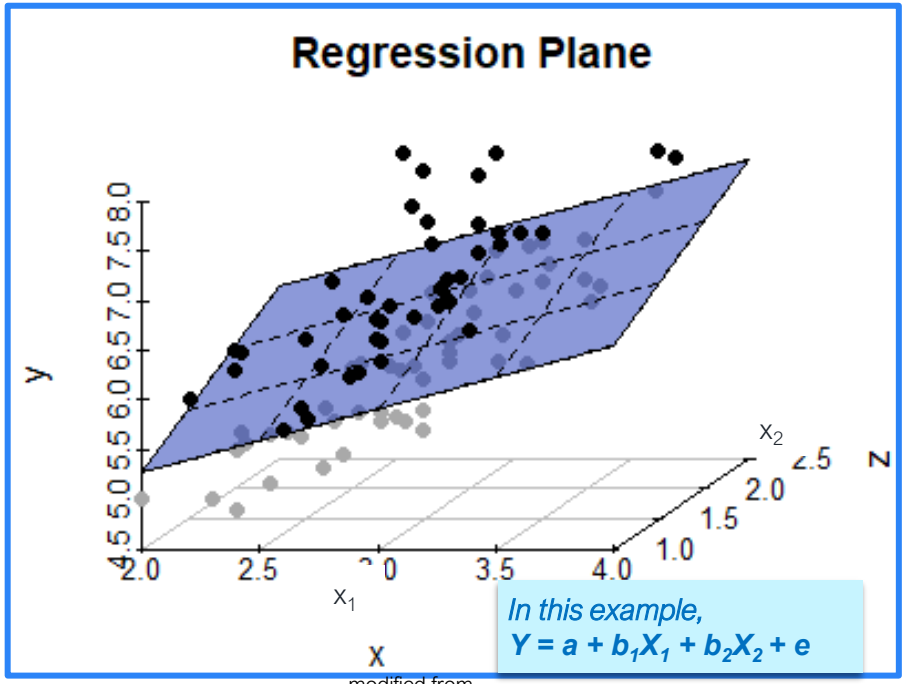


You'll forecast by the equation,

$Y = a + bx$

In this example,

$Y = 8000 + 40X$



In this example,

$Y = a + b_1X_1 + b_2X_2 + e$

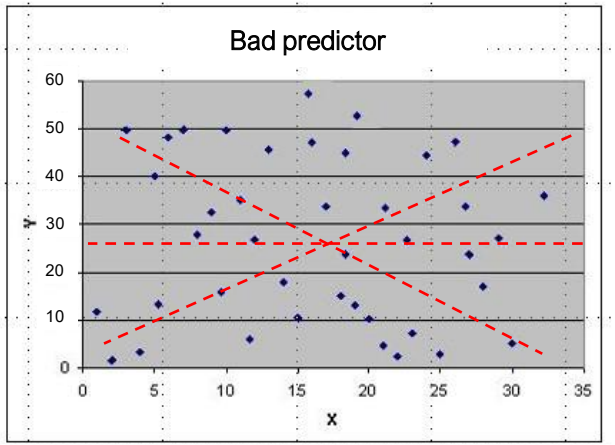
modified from
https://stackoverflow.com/questions/47344850/scatterplot3d-regression-plane-with-residuals

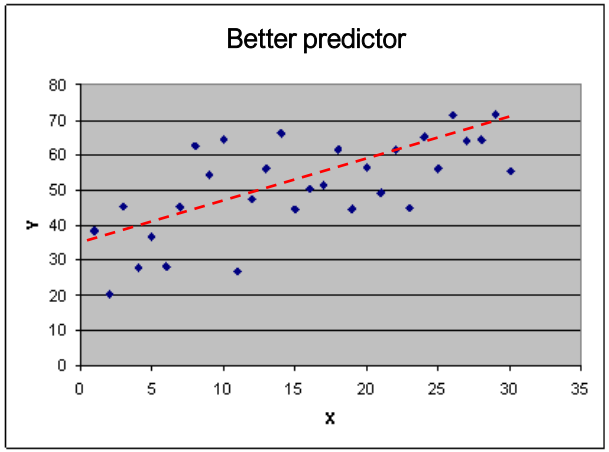
Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

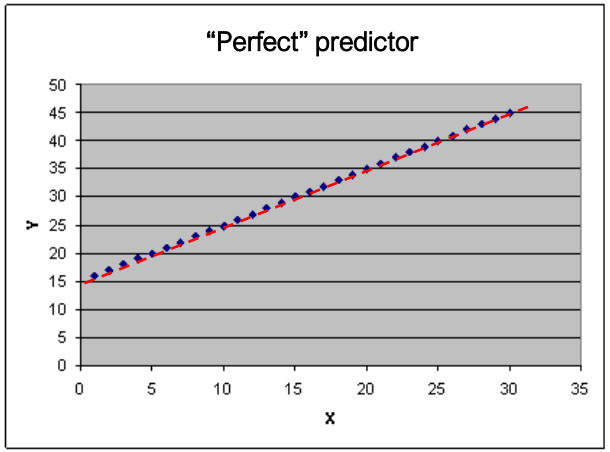
Automated Variable Selection

Helps to construct models when you're unsure which predictors drive a result

Techniques such as Stepwise, All Possible Subsets, and Lasso (Machine Learning) Regression select the set of predictors which best explains what you're trying to forecast







Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Calculation of Regression Coefficients

straightforward application of matrix algebra

Least squares estimates in matrix notation

Here's the punchline: the $(k+1) \times 1$ vector containing the estimates of the $(k+1)$ parameters of the regression function can be shown to equal:

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = (X'X)^{-1}X'Y$$

where:

- $(X'X)^{-1}$ is the **inverse** of the $X'X$ matrix, and
- X' is the **transpose** of the X matrix.

Source: https://online.stat.psu.edu/stat462/node/132/

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Linear Regression

A few key measures

Concept	Measures
Overall Fit	• R^2 and Adjusted R^2 • F –Statistic • RMSE
Each predictor's fit	• Coefficients (+ or -) and their t-statistics • Error patterns
Other (depends on software)	• Standard error of the equation • Durbin-Watson (error tracks itself through time) • More

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

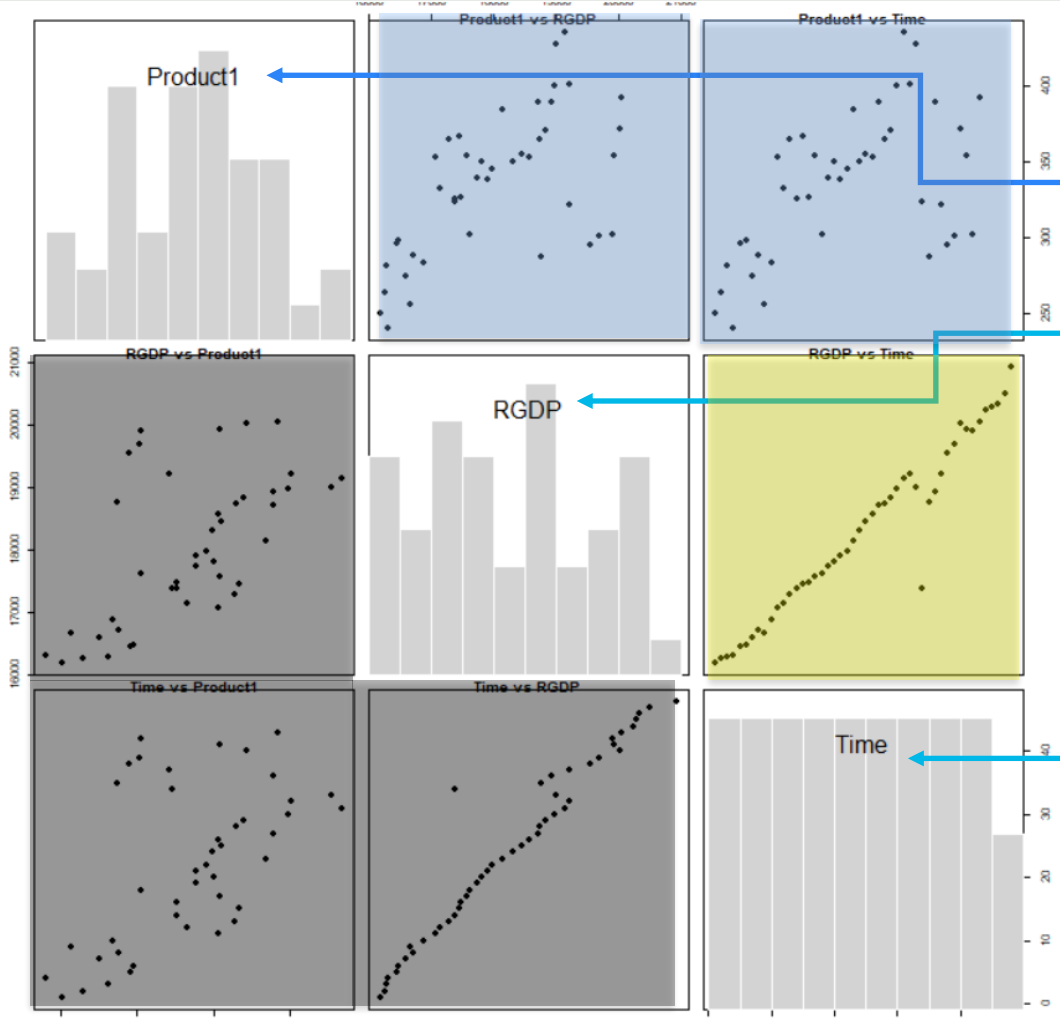
Linear Regression

Three big words

Word	Meaning	Impact
Multicollinearity	2 or more explanatory (X) variables are highly correlated	OLS regression algorithm can't attribute Y to each, unambiguously
Heteroskedasticity	Variance increases with scales	Regression's assumption of Independent, **Identically Distributed** errors is violated, suggesting that regression coefficients are invalid
Autocorrelation	The regression error tracks itself through time	Regression's assumption of **Independent,** Identically Distributed errors is violated, suggesting that regression coefficients are invalid

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Multicollinearity
What It Looks Like



"pairplots" for Dependent Variable and 2 Predictors

Strong correlation between RGDP and Time is evidenced by near-linearity

The correlation between these two predictors (X's) is stronger than either's with the dependent variable (Y)!

Multicollinearity
What It Looks Like

	A	B	C	D	E
1		***Product1***	*Unemp_Pct*	*CPI_Pct*	***RGDP***
2	Product1	1			
3	Unemp_Pct	-0.6445	1		
4	CPI_Pct	-0.0715	-0.3556	1	
5	**RGDP**	0.6067	-0.6044	0.6372	1
6	**Time**	0.5553	-0.4124	0.5608	0.9647
7					
8	***Single variable r -> R^2 for Product1***				
9	***RGDP***	*0.3680*			
10	***Time***	*0.3083*			

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.6067
R Square	0.3680
Adjusted R Square	0.35263
Standard Error	38.3545
Observations	43

SUMMARY OUTPUT

ssion Statistics	
	0.5553
	0.3083
Square	0.2915
rror	40.1254
ns	43

	df	*SS*	*MS*	*F*	*Significance F*
	1	29427.7208	29427.72	18.2775367	0.00011113
	41	66011.9892	1610.049		
	42	95439.7101			

	Coefficient	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	ov
	288.48	12.4548	23.1622	3.6186E-25	263.3273	313.6330	2
	2.10806	0.4931	4.2752	0.0001	1.1123	3.1039	1

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.6172
R Square	0.3810
Adjusted R Square	0.3500
Standard Error	38.4316
Observations	43

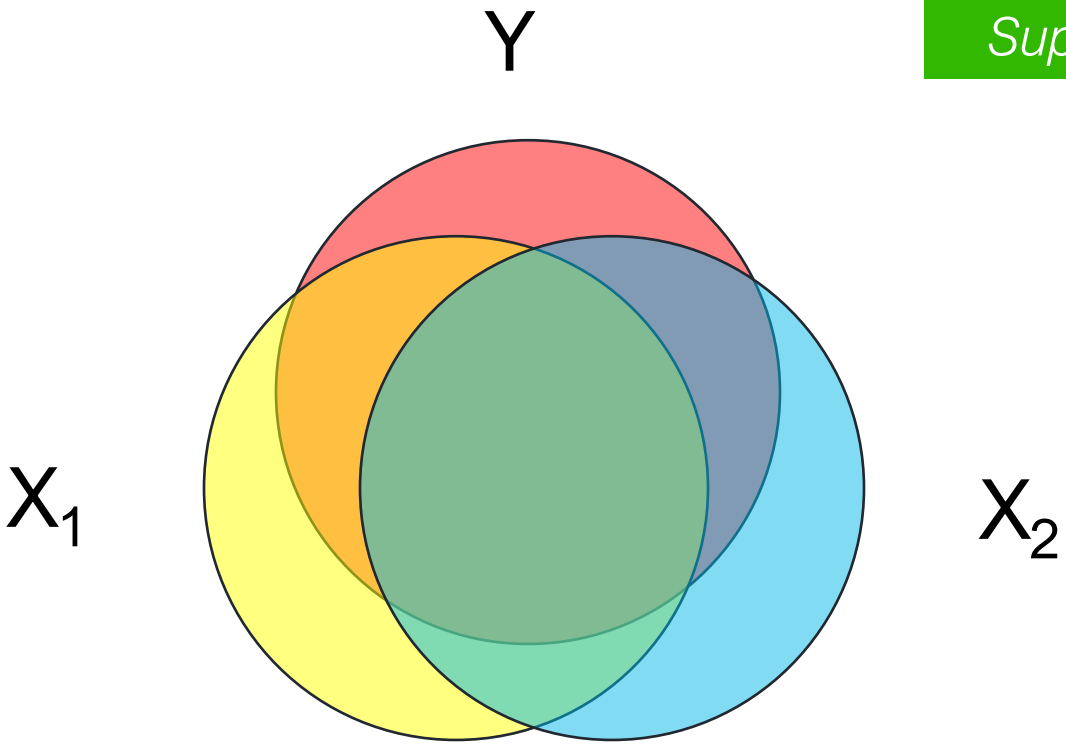
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	36360.1912	18180.1	12.3088989	6.8259E-05
Residual	40	59079.5189	1476.988		
Total	42	95439.7101			

	Coefficient	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-375.09	306.5216	-1.2237	0.22822442	-994.5943	244.4123
RGDP	0.04142	0.0191	2.1665	0.0363	0.0028	0.0801
Time	-1.6389	1.7928	-0.9141	0.3661	-5.2623	1.9846

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

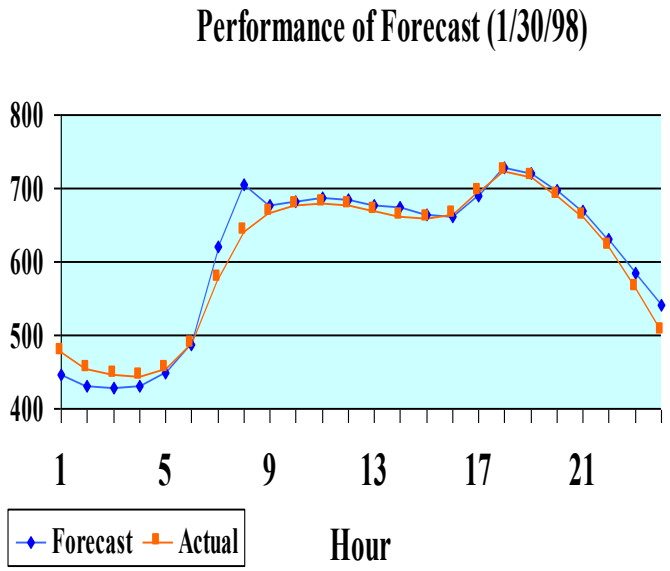
Multicollinearity
Supplemental Note

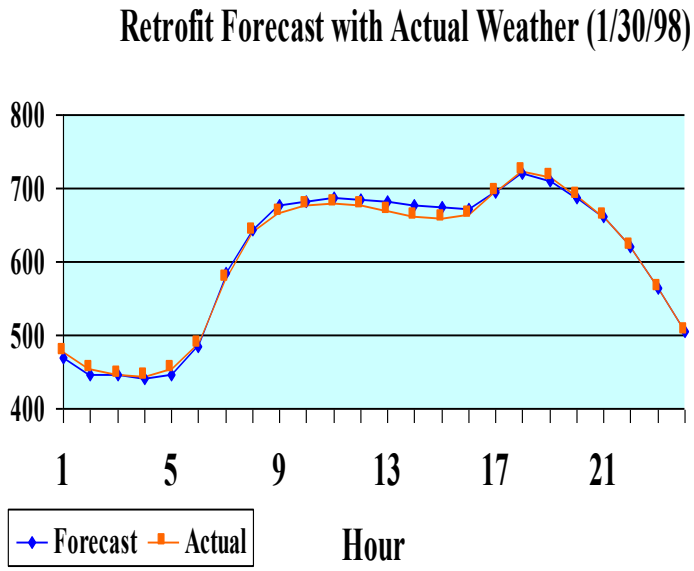


Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Some folks say it doesn't matter, but look what happened to me!

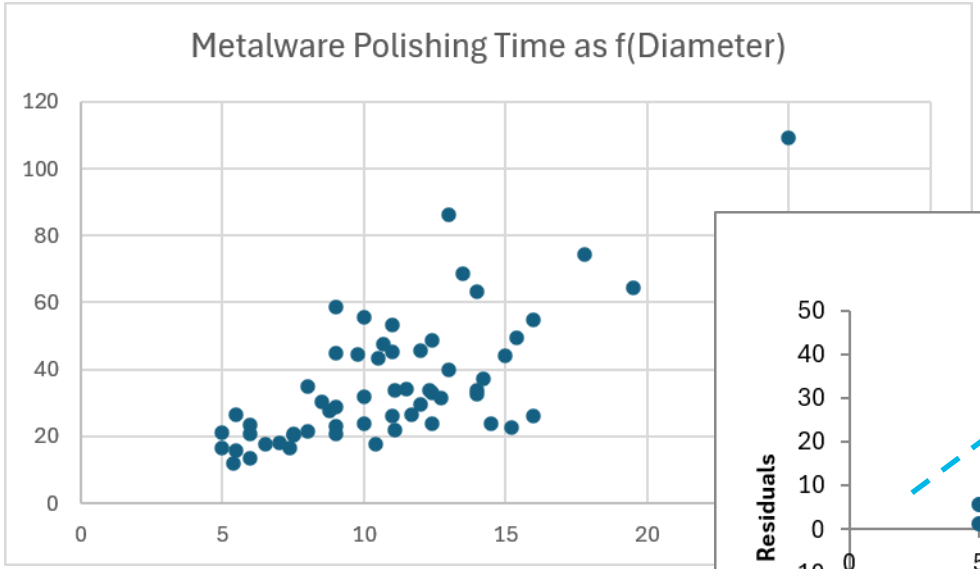
Sudden divergence in collinear explanatory variables

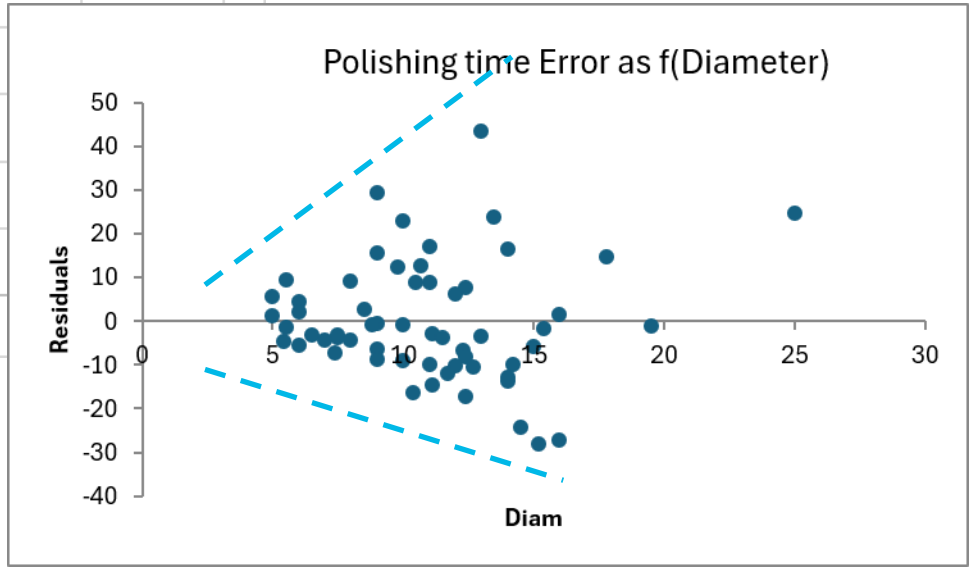




Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Heteroskedasticity
What It Looks Like

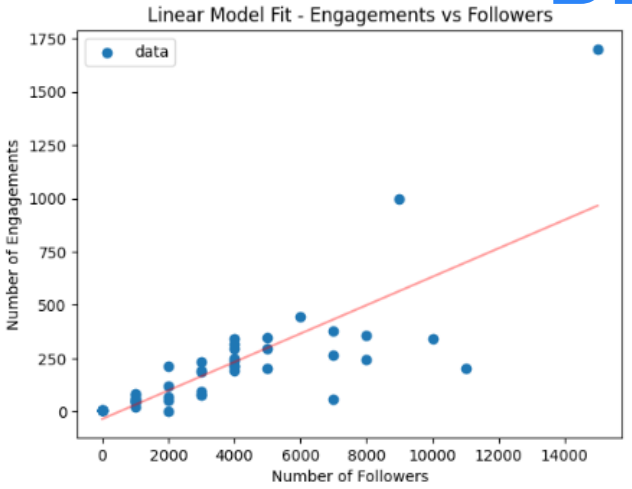


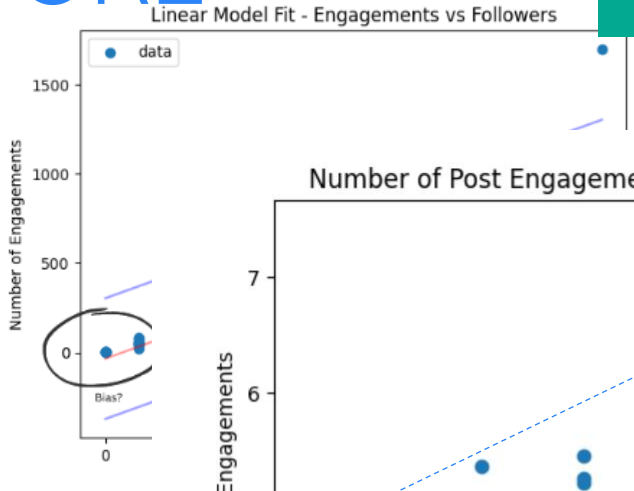


Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

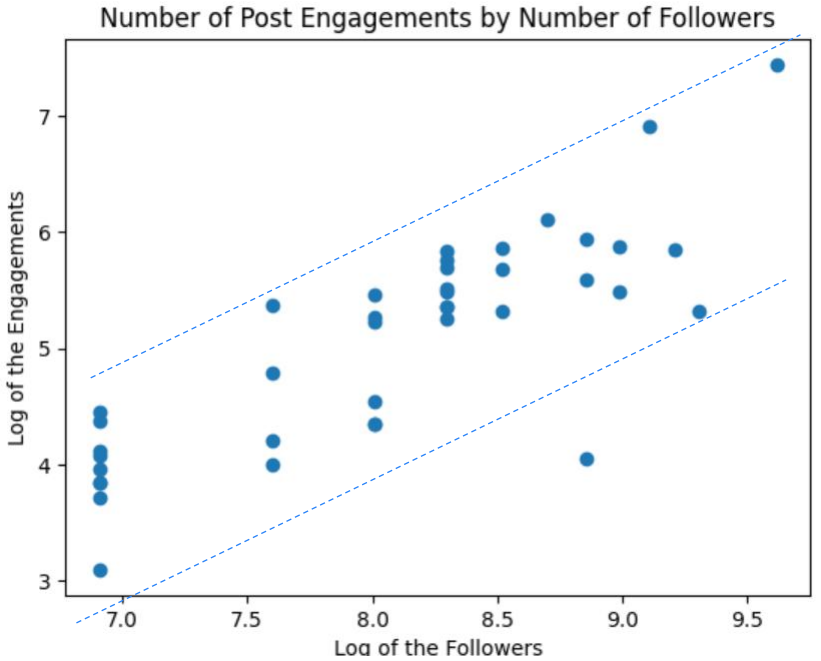
Heteroskedasticity
What It Looks Like, v2

BEFORE





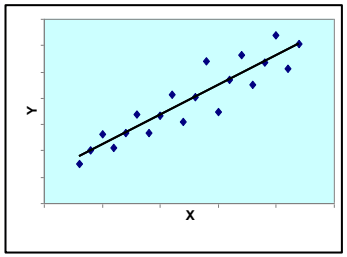
AFTER



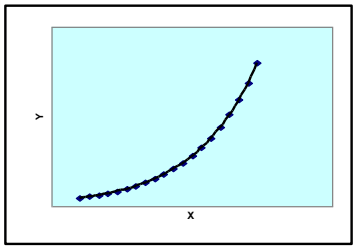
Source: https://aaron-pickering.com/2023/01/25/working-with-heteroscedastic-data/

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

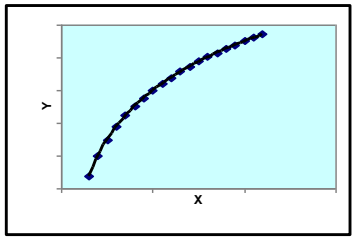
BTW
Logarithms



For every unit increase in X (____), there is a corresponding *n*-unit increase in Y (____)



As X (____) increases, its positive impact on Y (____) accelerates

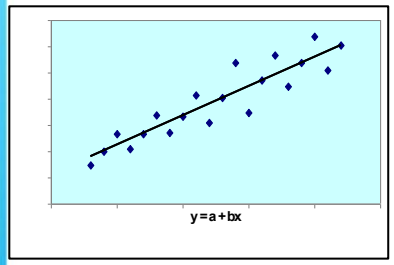


As X (_____) increases, its positive impact on Y (____) trails off

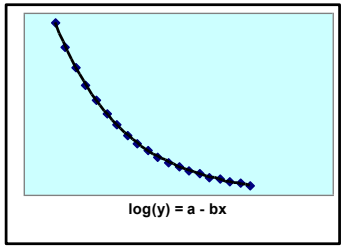
Non-linearities: Emphasis on Meaning

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

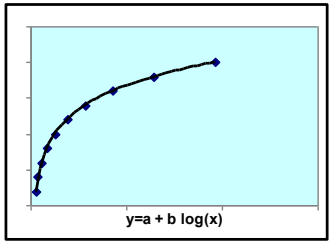
BTW
Logarithms



For every unit increase in X (____), there is a corresponding *n*-unit increase in Y (____)



For every unit increase in X (____), there is a corresponding *n*-percent decrease in Y (____)



For every 1-percent increase in X (____), there is a corresponding *n*-unit increase in Y (____)

Logarithms: The Concept of Proportionate Variation

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

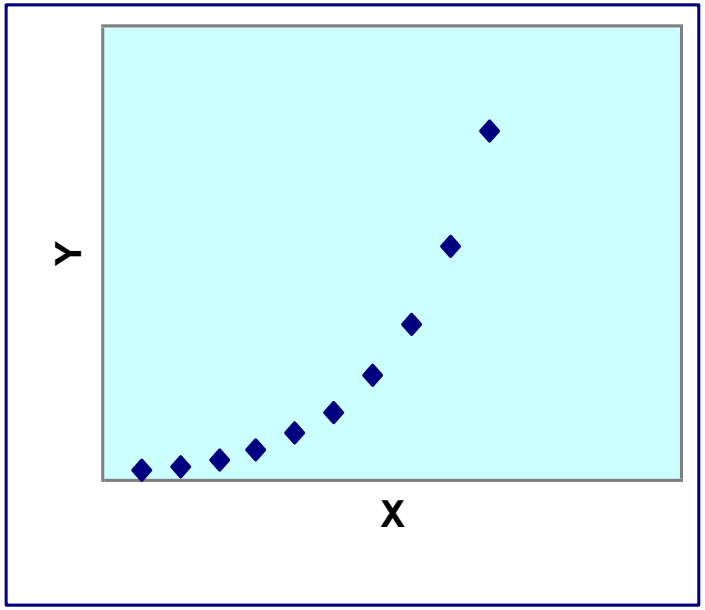
BTW
Logarithms

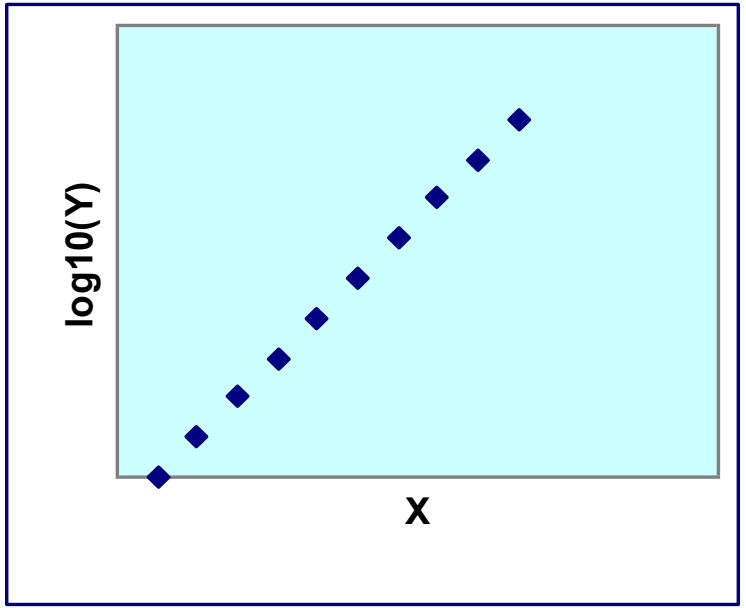
	X	Y			=LOG10(y)
	1	1			0
---> + 1	2	1.5	<----+ 50% (* 1.5)		0.17609126
|	3	2.25	|		0.35218252
|	4	3.375	|		0.52827378
|	5	5.0625	|		0.70436504
|	6	7.59375	|		0.8804563
|	7	11.390625	|		1.05654755
|	8	17.085938	|		1.23263881
|	9	25.628906	|		1.40873007
V	10	38.443359	V		1.58482133

For the Cynic:
An Exercise

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

And Now
the Cynic's Graphs ...

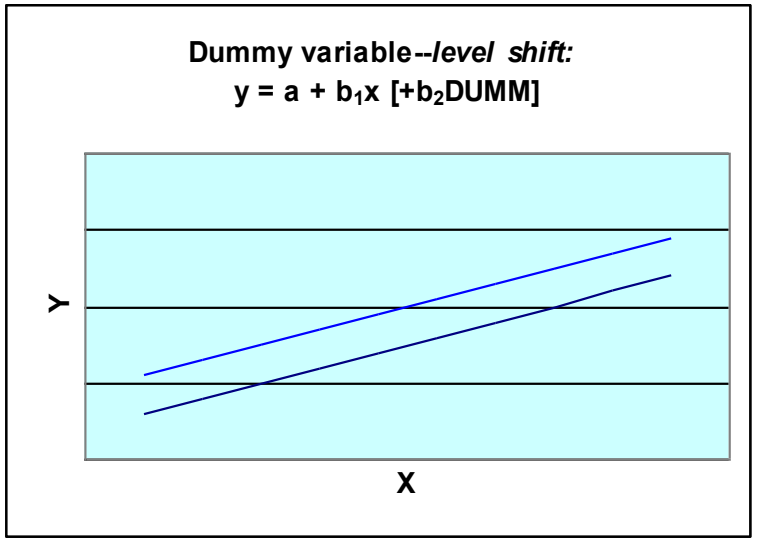


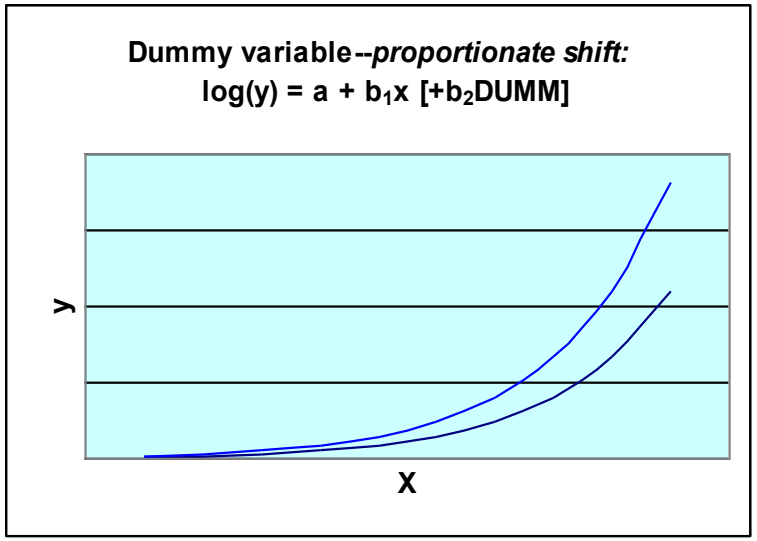


Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

BTW
Logarithms

Effect of
(0,1) Dummy Variable





Shifts $y=a + b_1x + b_2DUMM$

Rotates $\log(y)=a + b_1x + b_2DUMM$

e.g. $Y = 4 + 2X + [4 * 0]$ or $Y = 4 + 2X + [3 * 1]$

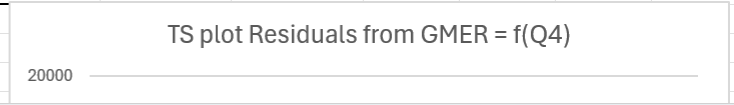
Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Autocorrelation

What it (can) look like

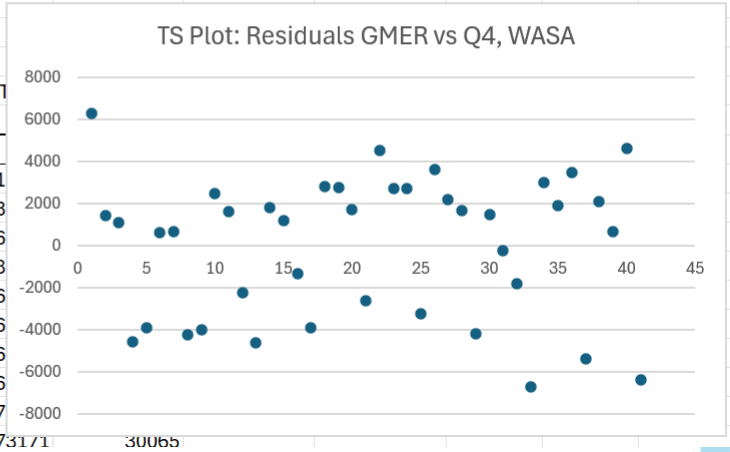
	A	B	C	D	E	F	G	H	I
62		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
63	Intercept	32292.51613	1155.552395	27.94552308	2.092E-27	29955.19079	34629.84147	29955.19079	34629.84147
64	QTR_4	17945.68387	2339.815267	7.669701161	2.6263E-09	13212.96078	22678.40697	13212.96078	22678.40697
65									
66	**DW**	**0.5373**		corr (e w/lag e)	**0.73040827**				
67									
68	RESIDUAL OUTPUT					PROBABILITY OUTPUT			
69									
70	*Observation*	*Predicted GMER*	*Residuals*	*Standard Residuals*		*Percentile*			
71	1	24061.08351	6301.916486	1.875861347		1.21951			
72	2	24391.44023	1420.559768	0.422851234		3.65853			
73	3	24902.11666	1123.883343	0.334540981		6.09756			
74	4	42667.11228	-4586.112277	-1.365126112		8.53658			
75	5	25802.33871	-3893.338713	-1.158911518		10.9756			
76	6	26012.94112	602.0588798	0.179211988		13.4146			
77	7	26357.0627	681.937299	0.202989015		15.8536			
78	8	44413.87342	-4243.873421	-1.263253508		18.2926			
79	9	27617.92417	-3976.924173	-1.183792002		20.7317			
80	10	27981.31656	2495.683438	0.742878155		23.17073171			
81									

BEFORE



AFTER

	A	B	C	D	E	F	G	H	I	J	K	L
62		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*			
63	Intercept	7635.47222	2567.881657	2.973451755	0.005090995	2437.06758	12833.87686	2437.06758	12833.87686			
64	QTR_4	17347.92026	1254.946099	13.82363775	2.09945E-16	14807.41471	19888.42582	14807.41471	19888.42582			
65	WASA	13.76486323	1.391246908	9.893903916	4.58839E-12	10.94843111	16.58129535	10.94843111	16.58129535			
66												
67	DW	2.1647		corr (e w/lag e)	-0.08121218							
68												
69	RESIDUAL OUTPUT					PROBABILITY OUT						
70												
71	*Observation*	*Predicted GMER*	*Residuals*	*Standard Residuals*		*Percentile*						
72	1	24061.08351	6301.916486	1.875861347		1.21951						
73	2	24391.44023	1420.559768	0.422851234		3.65853						
74	3	24902.11666	1123.883343	0.334540981		6.09756						
75	4	42667.11228	-4586.112277	-1.365126112		8.53658						
76	5	25802.33871	-3893.338713	-1.158911518		10.9756						
77	6	26012.94112	602.0588798	0.179211988		13.4146						
78	7	26357.0627	681.937299	0.202989015		15.8536						
79	8	44413.87342	-4243.873421	-1.263253508		18.2926						
80	9	27617.92417	-3976.924173	-1.183792002		20.7317						
81	10	27981.31656	2495.683438	0.742878155		23.17073171						



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

A Few Ways to Fix Them

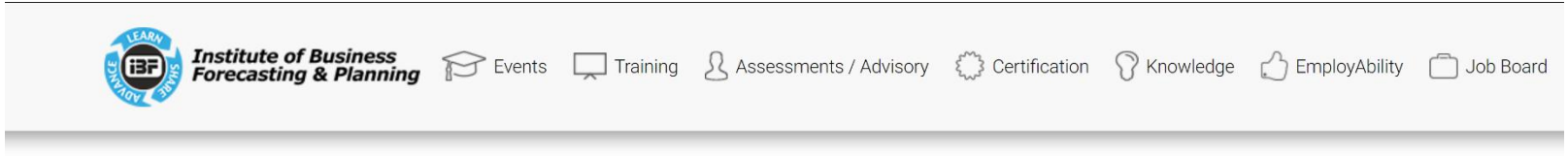
Multicollinearity, Heteroskedasticity, and Autocorrelation

Word	Fix(es)	
Multicollinearity	• Choose between X's • Combine X's (PCA) • Change method (Ridge Regression)	
Heteroskedasticity	• Arithmetic transformations (logarithms if percentage changes make sense)	
Autocorrelation	• Change functional form, e.g. differences: forecast ΔY_t, then construct $Y_{t+1} = Y_t, + \Delta y_{t+1}$ • Add missing predictors	Account for curvature, trend, seasonality, or omitted predictors that could explain the residual pattern
	• Forecast the error and add it on?	

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Predictive Model Performance

Mean Average Percent Error (MAPE)



Home › Knowledge › Glossary › MAPE (Mean Absolute Percentage Error)

MAPE (Mean Absolute Percentage Error)

What is MAPE? It is a simple average of absolute percentage errors. The MAPE calculation is as follows:

$$\text{MAPE} = \frac{\sum \frac{|A-F|}{A} \times 100}{N}$$

Here A= Actual, F= Forecast, N= Number of observations, and the vertical bars stand for absolute values.

For intermittent demand (many Actuals=0), consider Mean Absolute Deviation (MAD)= $\frac{\sum |A-F|}{N}$*; see* https://robjhyndman.com/papers/foresight.pdf

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Linear Regression

Forecast by applying the equation

#	Step	Details
1	Assemble data…	…for what you wish to forecast (dependent variable) and the predictors that you think are associated
2	Separate “ex post” holdouts	• “training” and “test” data • Be wary how a model fits vs. how it forecasts (MAPE, MAD) • Historic random variation can be mistaken for “the real thing” • Hence, a predictive model should be tested on data that did not create it
3	Fit the model with “training” and “test”	• see what it did with those *ex post* holdouts!
4	Re-fit …	… with training and test data combined
5	Forecast (extrapolate)	Given future estimates of predictors

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Stepwise Regression

An algorithm by which, after a pool of eligible X (explanatory) variables is identified:

- the computer selects first the variable which explains the most variation in Y
- next the variable which explains the most remaining variation
- and so forth until no further X's can significantly contribute to the explanation of Y

NOTE: sequential selection of predictors imparts bias

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

All Possible Subsets Regression

An algorithm by which, after a pool of eligible X (explanatory) variables is identified, summary measures are provided for all possible combinations

- Best 1-variable models, 2-variable models, 3-variable models, etc.
- R^2 goes up with each new variable; watch Adjusted R^2
- Beware the high level of summary measures, which can hide basic flaws such as insignificant t-statistics, gross multicollinearity, etc.

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Lasso Regression

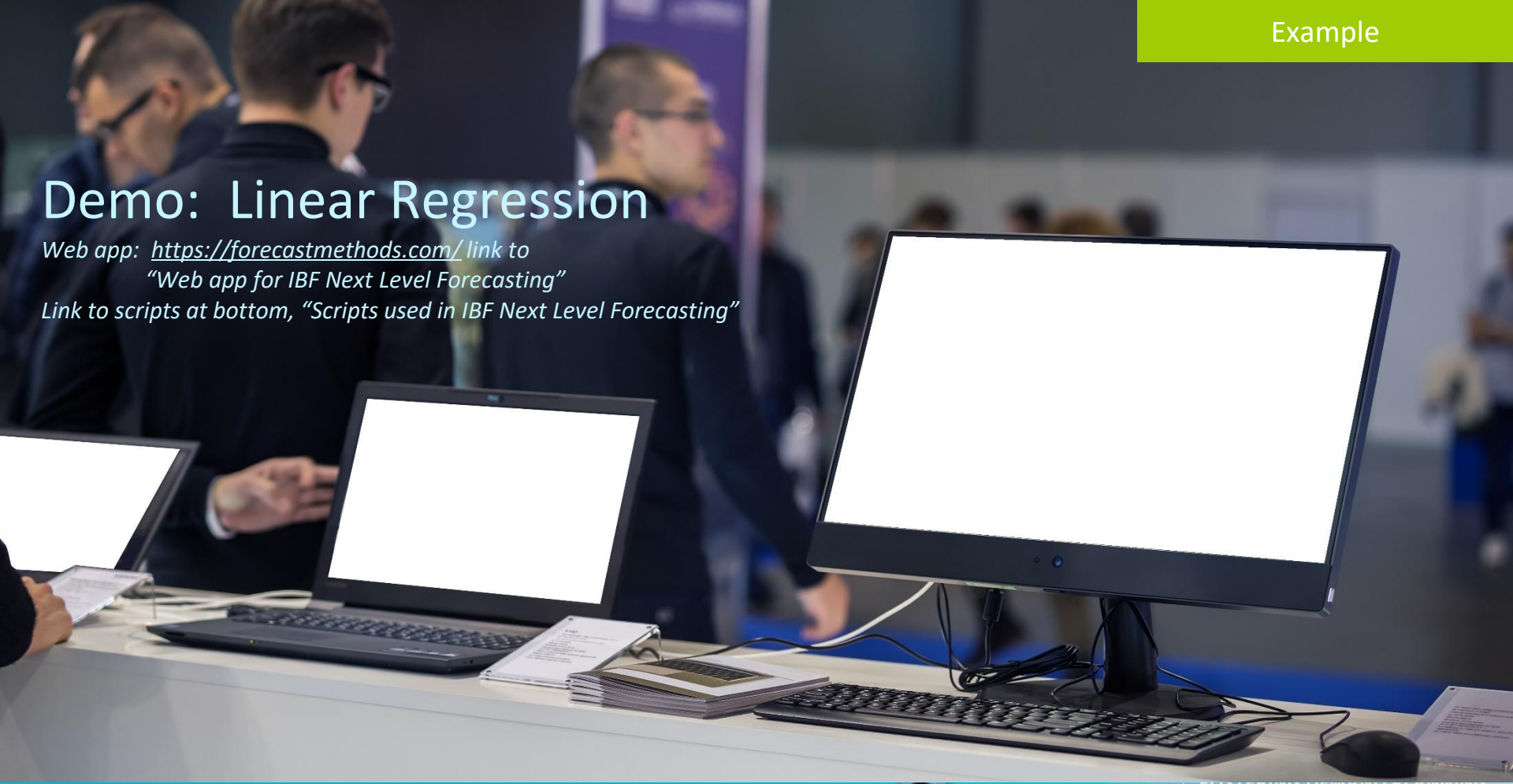
An algorithm by which, after a pool of eligible X (explanatory) variables is identified:

- a regression model is composed using all the X's
  - The model is **trained** on a fraction of the data set
  - The remainder is held out for **testing** (cross-validation, out-of-sample prediction)
- the model's **coefficients (weights)** are iteratively **shrunk toward zero**
  - **When decreasing** a variable's **influence** (weight) successively **improves accuracy**, that suggests the **variable was contributing more noise than value**
  - Accuracy is measured using the test fraction of the data
- the analyst usually chooses from 2 **final models**:
  - That which **minimizes test error**
  - The simplest version of the model **within 1 standard deviation of minimum** test error

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Demo: Linear Regression

Web app: https://forecastmethods.com/ link to
"Web app for IBF Next Level Forecasting"
Link to scripts at bottom, "Scripts used in IBF Next Level Forecasting"



Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Putting it all together

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS

Putting it all together

- The **ACPF exam** provides an opportunity and credential to demonstrate the depth of your forecasting knowledge
- **Applying concepts from the ACPF exam** can improve productivity in forecasting, along with accuracy of forecasts
  - **Product segmentation** can: (1) identify products that are important but difficult to forecast; (2) create homogenous groups for top-down forecasting
  - **Market segmentation** can identify customer groups, so that you can "get to know" your customer base from a high level
  - **Regression analysis** can implement "cause and effect" forecasting, which is useful not only for longer-term projections, but also scenario planning
  - Interestingly, the **most complex multiple regression models may be the easiest to build**, supported by automatic variable selection algorithms like stepwise, all possible subsets, and lasso regression

Institute of Business Forecasting & Planning
FOSTERING DEMAND PLANNING, FORECASTING, S&OP & ANALYTICS FOR 40+ YEARS