

オリエンテーション
デミング先生とデータサイエンス

“Data are for Prediction and Action”

from Deming (1942) On a Classification of the Problems of
Statistical Inference, *JASA* 37, 173-185.

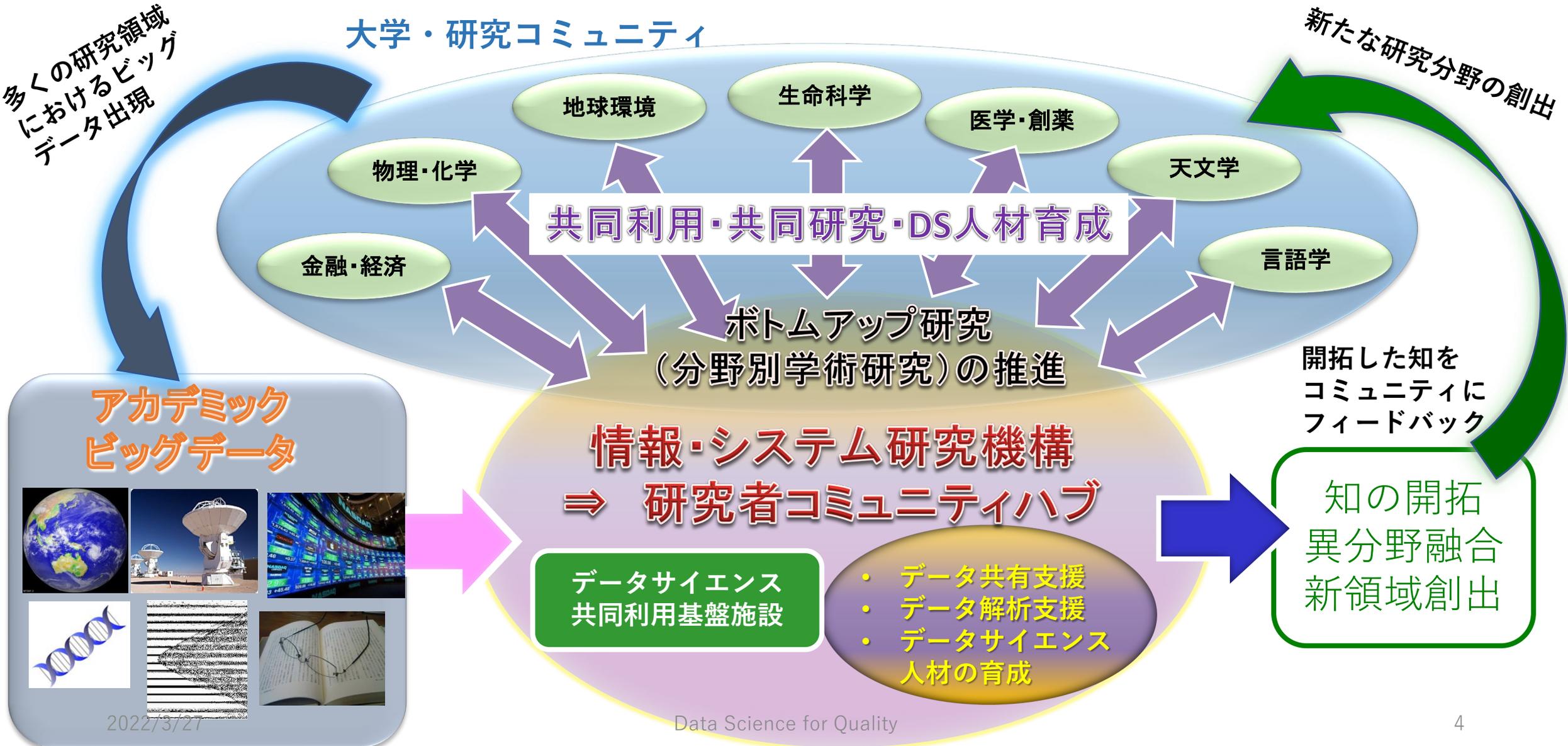
大学共同利用機関法人
情報・システム研究機構
統計数理研究所 椿 広計

前座の務め

- 自己組織紹介
- 職業としての統計家
- 統計家としてのデミング先生の貢献
- 日本的品質管理の父として
- 海外のデータサイエンス人づくりの比較的健全な発展
- おわりに
 - データサイエンス時代だからこそデミング・石川哲学の再興

自己組織紹介

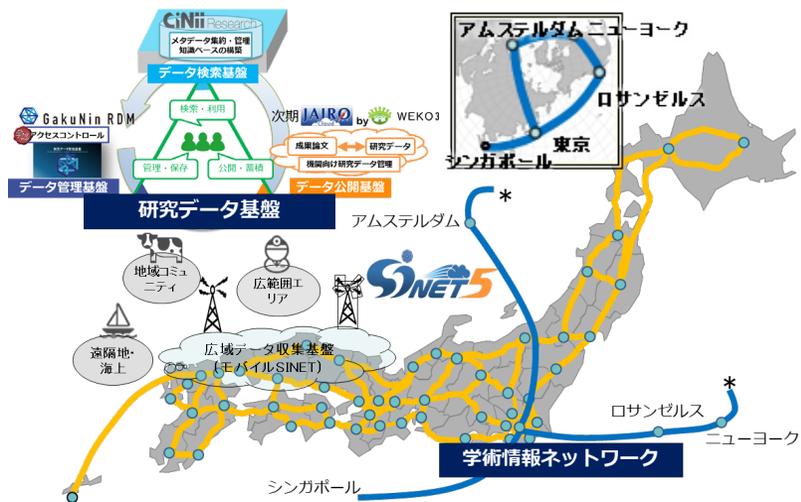
データサイエンス推進の概念



2000(平成12)年設置

(※前身の「学術情報センター」は(1986年4月設置)

情報研は、「情報学」という新しい分野で、私たちの生活をどのように豊かにするかを研究する国内唯一の学術総合研究所です。また、超高速ネットワークや学術コンテンツ等を提供し、日本の研究・教育活動に貢献しています。



◆学術情報ネットワーク (SINET)

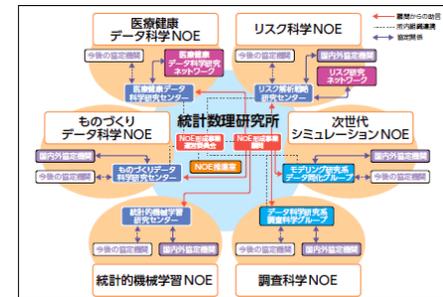
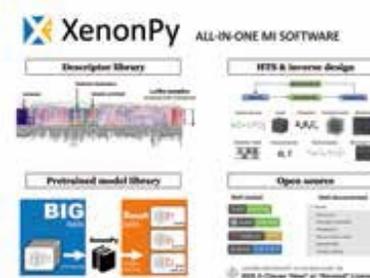
- 全県を100Gbpsで網目状に接続する国内回線、欧米を地球一周100Gbpsで接続する国際回線、セキュアな広域モバイル収集環境の運営
- オープンサイエンスの実現に向けた研究データ管理・公開・検索基盤の整備・運用
- 学術情報の検索・公開、クラウド利活用促進、学術認証等のための学術情報基盤の整備・運用

◆大学間連携によるセキュリティ体制の基盤構築

1944(昭和19)年6月設置

情報社会と言われる現代では、私たちの生活の中には膨大なデータが飛び交っています。統数研ではデータから本質的な情報を抽出し、将来を予測することにより、人間が賢くデータ共生する社会の実現を目指し、先進的な研究を日々行なっています。

マテリアル・インフォマティクスソフトウェア



NOE (Network Of Excellence) 形成によるネットワーク型共同研究の推進

◆ NOE形成による異分野融合研究, 新研究分野の創成

- ◆ものづくりデータ科学 (データ駆動型物質・材料探索)
- ◆データ同化・機械学習 (データ同化や機械学習を用いた予測・解析等)
- ◆リスク科学 (環境, 資源, 金融, 地震予測等)
- ◆医療健康データ科学 (医学・健康科学領域における先進的なデータサイエンスの研究・教育の推進)
- ◆調査科学 (日本人の国民性, 国際比較, 住民意識調査等) など

◆統計思考力育成

各種研究活動の担い手となるデータサイエンティスト等統計思考を備えた人材の育成

◆先進的科学技术計算資源の提供

⇒世界最高水準の統計計算環境等

統計科学スーパーコンピュータシステム



職業としての統計家

Deming先生の名刺”Statistician”

小浦孝三先生談

アメリカ合衆国労働統計局雇用統計2019/05

<https://www.bls.gov/ooh/math/mathematicians-and-statisticians.htm>などより作成

- 米国政府職業小分類867職種中：数理科学職「統計家」第5位、
計算機科学職「情報セキュリティアナリスト」第10位の**労働統計局成長率予測**
 - 「第1位は風力タービンサービス技術者（7000名：62%成長）
 - 2018年から職業分類に追加されたデータサイエンティストも数理科学職

アメリカ合衆国標準職業分類	分類コード	雇用者数	10年成長率	トップ産業雇用者数
統計家	15-2041	42,700	34.6%	科学研究開発, 6190
データサイエンティスト/その他	15-2051/99	33,200	30.9%	計算機システム設計, 9100
オペレーションズリサーチアナリスト	15-2031	105,100	24.8%	企業マネジメント, 9930
アクチュアリー	15-2011	27,700	17.6%	保険会社, 11,140
数理科学（中分類）	15-20XX	211,700	26.5%	
情報セキュリティアナリスト	15-1212	131,000	31.2%	計算機システム設計, 36,280
計算機科学（中分類）	15-12XX	4,633,400	11.5%	

米国労働統計局が示す社会における 「統計家(15-2041)」の役割

- 特定の質問や問題に答えるために必要なデータを決定
- ビジネス、工学、科学、その他の分野で実用的問題を解決に、数学理論と数理技術を適用
- データ収集のための調査・アンケート（設問の設計、ターゲット母集団からの適切な標本の決定、サンプリング、意思決定に必要なサンプルサイズの決定）、実験計画
- データ分析のための数理モデル・統計モデルの開発
- データを解釈し、専門家および専門家ではない者に分析をレポート
- 統計ソフトウェアを利用したデータを分析し、トレンドや関係性を明らかにし、データの妥当性や限界も検討
- レポート（ビジュアルライゼーション、集計表）を作成して、ビジネス等の意思決定を支援
 - **42,700名中：全米146統計学専攻の修士統計等取得者が65%、博士取得者が20%**

Top 10 Famous Statisticians

<https://onlinedegrees.mtu.edu/news/top-10-famous-statisticians>

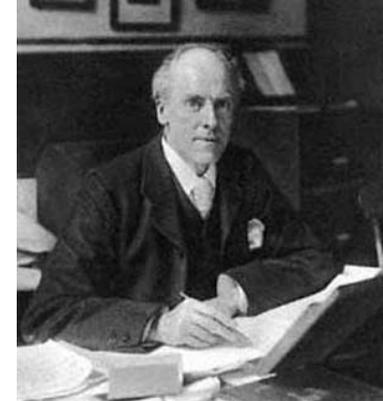
- Friedlich Gauss (1777~1855): 基礎数理：誤差論、最小二乗法、彗星軌道予測
- Florence Nightingale(1820~1910): 医学統計(医療の改善活動)：データの可視化、統計グラフ開発
 - Oxford大学応用統計学科構想提案 (Applied Statisticsという用語の創生者)
- Karl Pearson(1857~1936)：科学研究への統計適用 (注：科学の文法)：相関論、適合度検定
 - 1911 Univ. College London 応用統計学科創設(世界初の統計学科)
- W. S. Gosset (ペンネーム：Student,1876~1937): 生産への統計適用 (ギネスビールの改善活動)：小標本論
- R. A. Fisher (1890~1962)：近代統計学の創生：推論、実験計画法 (ロザムステッド農事試験場)
- Gertrude Cox(1900~1978): 高等統計教育：科学のパートナー、女性活躍
- **Edward Deming (1900~1993)：産業マネジメント：品質マネジメント**
- J. Tukey(1915~2000): 探索的データ解析、ロバスト統計、S言語の開発指導
 - Tukey (1962) The Future of Data Analysis, AMS 33(1), 1-67. 「だから2012はData Science 50周年記念」
- G. Box (1919~2013): 化学への統計適用(ICI)：工業実験計画、変数変換、時系列解析、応用ベイズ解析
- Janet Norwood (1923~2015): データマネジメント(労働統計局)：客観的・科学的データシステム、CPI, 失業率推定

数学は「科学の言語」統計は「科学の**文法**」

K. Pearson, *The Grammar of Science*, 1892.

何でも科学になる：**プロセスに基づく科学の定義**

Man Gives a Law to Nature



https://en.wikipedia.org/wiki/Karl_Pearson

- 科学の適用範囲 (Scope)

- 科学を特徴づけるのは対象ではなく，その**プロセス**である。
- あらゆる知的方法を用いて**真実を確かめること**

- **科学的プロセス：知識マネジメントプロセス**：認識の進化

- **分解：Analysis**; 事実の周到な分類と事実間の関連性と順序の観察
- **総合：Synthesis**; 創造的想像に支援された科学的法則の付与
- **妥当性検証：Validation**; “自己批判と全ての人が等しく妥当性を認めるか否かを検証”

- **科学的プロセス支援技法：ピアソンの統計的方法開発**

- 視覚に訴える：ヒストグラム，散布図
- 指標化：標準偏差，相関係数
- 妥当性検証：モデルの適合度検定

統計家としての Deming先生の功績



https://upload.wikimedia.org/wikipedia/commons/7/73/W._Edwards_Deming.jpg

American Statistical Association Deming Lecturer Award since 1995

- Deming先生の貢献の範囲と重要性について統計科学コミュニティの意識を高めるために創設
 - ASAが考えるDeming先生の主要な貢献
 - サンプルングへの貢献
 - 標本調査に関する最初の専門書の一つを1950年に発刊
 - 調査推計の質に影響を与える複数要因の検討を行ったパイオニア
 - 標本調査の質に関するアイデアを品質改善に一般化
 - ビジネスと産業界への統計的貢献
 - データから学ぶデータアナリスト、統計的方法開発者、思想家として、輸送業界における主要なデータアナリストとなったこと
 - それらを一般化してビジネスや産業界の品質向上のための統計的方法と統計的思考の利活用を強力に推進
 - マネジメントへの貢献
 - マネジメント理論（知識システムの真似人念と）を開発・普及し、日本と米国の企業パフォーマンスの向上に貢献
- 最近の受賞者
 - Techno-metrics 関係者(1988にG. Boxと共に来日、田口玄一、竹内啓、久米均らと懇談)
 - 2012: Jeff Wu (第1回赤池メモリアルレクチャー賞受賞者)
 - 2013: Vijay Nair (国際統計協会元会長)
 - その他の統計的品質管理・信頼性工学関係者
 - 2015: W. Meeker, 2019: Nicholas Fisher, 2020 Dennis Lin.
 - 公的統計関係者
 - 2017: F. Scheuren (元ASA会長、統計等のデータの質マネジメントで活躍)

Deming先生の統計理論分野の貢献

- 標本調査論実務のパイオニア
 - 確率抽出 (Probability Sampling)
 - Deming配分
 - 必要な推定精度を確保するためのコスト最適化標本配分
 - Deming, W. E.(1950) *Some Theory of Sampling*. John Wiley & Sons, Inc. New York. (Dover Publications edition, 1966).
- Deming Regression
 - 変数誤差モデル (説明変数に誤差のある回帰分析)
 - Deming, W. E. (1943) *Statistical adjustment of data*. Wiley, NY (Dover Publications edition, 1985).
 - 統計用語 : P 値 (Fisherの有意確率に代わって)

物理学者・統計家としてのDeming先生

<https://deming.org/timeline/>

- 1928: Yale大学Ph.D in Mathematical Physics
 - **Deming(1928) Equipotential Surface Electrons as an Explanation of the Packing Effect, *Physical Review*, 31(3), 453~465.**
- 1927~1939: 米国農務省窒素固定研究所数理物理職
 - **1927: ベル研究所物理職Shewhartと出会う**
 - Deming(1930) On the determination of the parameters in an empirical formula, *Proc. Phys. Soc.* 42 97
 - 1935~農務省大学院(GSDA)で数学・統計学の講義開始
 - 1935: Univ. College London でFisher, E.S. Pearson, Neymanと統計研究
 - Deming and Birge(1934) On the Statistical Theory of Error, *Reviews of Modern Physics*, 6, 119-161. 実証的物理学研究の基礎
 - **1937: ShewhartのSQC講義: 農務省大学院講義に招聘**
 - Shewhart(1939)Statistical Method from the Viewpoint of Quality Control.編集出版
 - このほかにも、Neyman, Fisher, Wishart, Cochranを招聘

管理を科学に:Shewhartの品質管理学創生

Shewhart(1939) *Statistical Method from Viewpoint of Quality Control*, Graduate School of the Department of Agriculture.

• Shewhartの管理の3成分

- Plan: 達成されるべき目標を規定する行為
- Do: 規定された目標を達成しようとする行為
- See: 目標が達成されたかどうかを判断する行為
- 「科学の文法」から派生した「実証科学のプロセス」
 - 仮説提示⇒データの採取（実験、調査）⇒仮説検定
- Shewhart流品質管理：**CAPD**：Cは検査ではなく：工程改善トリガーの発見
 - 問題（価値）の発見は外れ値の探索から
- 問題点（価値）の抽出技法
 - 期待と実際の乖離（残差）を改善のチャンスと考える
 - **米国Shewhart:管理図法の提唱⇒工程管理・改善重視の世界観⇒日本が引き継ぐ**
 - 検査を無くすことが品質管理の目的：西堀栄三郎
 - 検査のような合格が目的ではない⇒Out of Controlが出たら改善のチャンス
 - 1950年代～：品質は工程で作りこむ⇒1970年代～；品質は設計で作りこむ
 - 1980年代；顧客志向：当たり前品質から魅力的品質へ（Kano model）

- 1939~1946: 米国センサス局統計的標本調査アドバイザー
 - 1940年国勢調査に標本調査の概念導入：精度改善とコスト削減
 - Stephan, Deming and Hansen(1940) The Sampling Procedures of the 1940 Population Census, *JASA* 35, 615~630,
 - **統計編成（集計）に品質管理技法導入：世界初のWhite-Collar品質マネジメント**
 - 1943: 戦時下の技術者への応用統計教育
 - **1943：Stanford Statistics Training Program: 2000名にShewhart Cycle指導**
- 1946~1993: ニューヨーク大学大学院MBA統計学教授
 - 1946: ギリシャ選挙監視、インド政府標本調査指導
 - 1947~1952: 国連統計委員会標本調査分科会
 - Mahalanobis委員長, Fisher, Yatesらと共に
 - 国連統計委員会第一号レポート
 - 1947: GHQ/SCAP：ライス統計使節団で訪日：日本の官庁統計支援
 - 疲弊した家計の調査、1951国勢調査の準備支援
 - 統計数理研究所で標本調査論講義
 - 統計数理研究所員の海外論文誌投稿紹介支援（小川潤次郎先生との往復書簡）
 - 統計局図書室にDeming先生の残した資料（井上統計局長調べ）
 - 1947: 日本統計学会名誉会員(第1号)
 - 1954~1955 メキシコセンサス局、中央銀行等指導
 - 1959~1962：トルコ統計局指導

参考

- Deming and Stephan(1940) On a least squares adjustment of a sampled frequency table when the expected marginal totals are known, *AMS* 11(4), 427-444.引用件数2079
- Deming and Stephan (1941). On the interpretation of censuses as samples. *JASA* 36, 45-49.
- Deming (1944) On errors in surveys, *Amer. Sociological Review* 9(4),359-369.
- Sekar and Deming (1949). On a method of estimating birth and death rates and the extent of registration. *JASA* 44, 101-115.
- Deming(1953) On the distinction between enumerative and analytic surveys. *JASA* 48, 244-255.

日本的品質管理の父として

日本の品質管理活動から世界へ

- 1950: 日本科学技術連盟招聘
 - 歴史的講義（統計的品質管理の講義）：東京・九州
 - 箱根(経営者のための講義)
 - Deming Wheel: Investigate, Survey, Design, Sale Manufacture
 - <https://curiouscat.com/management/deming/deming-1950-japan-speech-mt-hakone>
 - 日本の技術者、数学者、統計学者
 - 講義録出版（1951:Deming賞創設, 第1回増山元三郎）
- 1955: ASQC Shewhart Medal
- 1980年代Deming先生の再発見：日本の高度経済成長を支えた
- 1986～ コロンビア大学Distinguished Professor

科学的ビジネスの文法：認識科学の文法から設計科学の文法への発展

デミング・石川のマネジメント・プロセスモデル + 支援技法の習熟

Japanese PDCA：日常管理(Control)と問題解決行動(Improvement)の合体

Nair, 2015 ISI会長講演「統計学の産業界への最大の貢献はDeming石川の統計哲学」

今、**自分はどこに立っているかを知る**

検査 & 統計的意思決定
統計的プロセス管理

Do

Planの

着実な実施

Shewhart(1939)

PDSサイクル

Deming + 石川の

PDC(S)Aサイクル, 1951

Plan

人・設備・

予算・情報

提供

マネジメントサイエンス技法

Check

あるべき姿と
実際とのずれ

What, Who,
When, Where,

How

GAP (異常検知)分析、Shewhart管理図

問題提起

解くべき
価値ある

問題・課題の
発見

問題解決型QCストーリー

改善の標準シナリオ

小松製作所, 1960

仮説提示

どう解くか何を
どう調べるかか

特性要因図

連関図

要求品質展開

質的調査計画
量的調査計画
実験計画

情報収集

現場や社会の調査情報

対策案比較検証実験情報

分析

要因の分析

原因と結果

因果モデリング：検証的解析
層別分析・回帰分析・
時系列解析

Action

=

対策立案

問題解決方針

原因に基づく対策
+ 効果確認

大原則：
適切な方法を適切な
プロセスに
埋め込む

石川馨先生をはじめとする第一世代の先生方 日本的品質管理の確立と世界への発信

• QC7つ道具の実証的提唱と現場展開

- 適切なデータをとる
 - チェックシート
- どこが重要問題なのかを知る
 - パレート図
 - 管理図
 - ヒストグラム
- 問題の原因候補（要因）に関する仮説を網羅する
 - 特性要因図(Ishikawa Diagram)
- 原因を事実に基づいて突き止める
 - グラフ・層別
- 効果の確認
 - ヒストグラム、パレート図等

• 問題解決型QCストーリー

- 第2世代開発方法論の世界展開
 - JSQC発足後のものも多い
 - 問題の攻め方：Akao's QFD
 - 価値の発見原理：Kano Model
 - 経営の因果モデル化：方針管理（コマツの旗管理）
⇒ Balanced Scorecard
 - 課題達成型QCストーリー
 - 工程能力指数：デンスー
 - 棟近雅彦先生、永田靖先生
 - 実験計画法の活用と改革：Taguchi Method
 - 多変量解析の活用
 - 奥野忠一先生、芳賀敏郎先生、吉沢正先生
- 第3世代管理技術
 - 人間系の誤りとFool Proof（中條武志先生）
 - Failure Mode（鈴木和幸先生）
 - Taguchi Methodと統計の再融合
 - 宮川雅己先生、永田靖先生

日本の応用統計の貢献：要因の分類とその技術分析戦略への影響

椿(2019) 田口玄一の統計学に対する貢献, 横幹、13(2)、110-116.

- 田口玄一，実験計画法(上)，(下)，丸善，(1957, 1958)からの発展
- 産業界の要因解析の目標
⇒ × 予測 ○ 制御因子最適化(Action)と技術知識価値獲得

• 変動要因の分類

- 制御因子：技術を最適化するために調整する因子
 - 意図的にその水準（変数値）を制御可能
 - 要因効果を推定すべき因子
- 標示因子：制御因子の最適化に影響を与える可能性のある要因
 - 観測可能で，意図的に実験で観測
 - その水準を制御することは不可能な因子
- AI時代でも重要な区別：ただの入力変数ではなりえない

最適行動決定を踏まえた技術分析戦略 価値最適化の場合分けを知る

- Actionのための統計的要因解析：交互作用(Interaction)解析
 - 統計的品質管理戦略
 - 標示因子と制御因子の交互作用を明らかにし標示因子の水準毎に制御因子を最適化
 - Tukey(交互作用の特異値分解), 広津千尋(交互作用のパターン分類), 田口玄一 (SN比解析) 等
 - 制御因子間の交互作用存在
 - 交互作用の少ない制御因子の合成・交互作用の消去技法
- 田口玄一のコトづくり：ロバストパラメータ設計
 - ロバストな制御因子最適化：
 - 標示因子の一部を制御因子と敵対するノイズと見なす直積実験
 - 制御因子 vs ノイズ因子のゲームの落としどころ
 - Waldの決定理論のもう一つの重要な特色：ミニマックス推論
 - 統計家(意思決定家)とそれを邪魔するNatureのゲームとして
 - 期待損失最小化決定理論を整備
 - 直積実験→調合誤差因子の理念(誤差因子の最適化)
 - AIの敵対的学習(GAN)の先駆的業績

George Box: 1986/06

Bell研若手研究者(V. Nair, Jeff Wuら) と共に
田口玄一博士との討論とトヨタを視察

R.B. Reich, 1984

- アメリカが10年間で50%以上シェアを落とした製造業リスト
 - Automobiles, Cameras, Stereo Components
 - Medical Equipment, Color Television Sets
 - Hand Tools, Radial Tires
 - Electric Motors, Food Processors
 - Microwave Ovens, Athletic Equipment
 - Computer Chips, Industrial Robots
 - Electron Microscopes, Machine Tools
- **高品質低コスト製品に敗退**

Box and Bisgaard, 1987

- The Scientific Context of Quality Improvement
- **"They Do It and We Don't"**
 - 小原、光藤、椿(2018)
 - 横幹連合コンファレンス: QCサークル
 - JSQC推薦: 至宝コレクションへの登録
 - 1988: 製造業事業所無作為標本調査
 - 6.6万事業所
 - 75万QCサークル, **550万名**
- 当時: 品質管理誌、QCC誌:
 - 問題解決ストーリー宝庫
 - 直交表実験・多変量解析の事例

海外のデータサイエンス 人づくりの比較的健全な発展

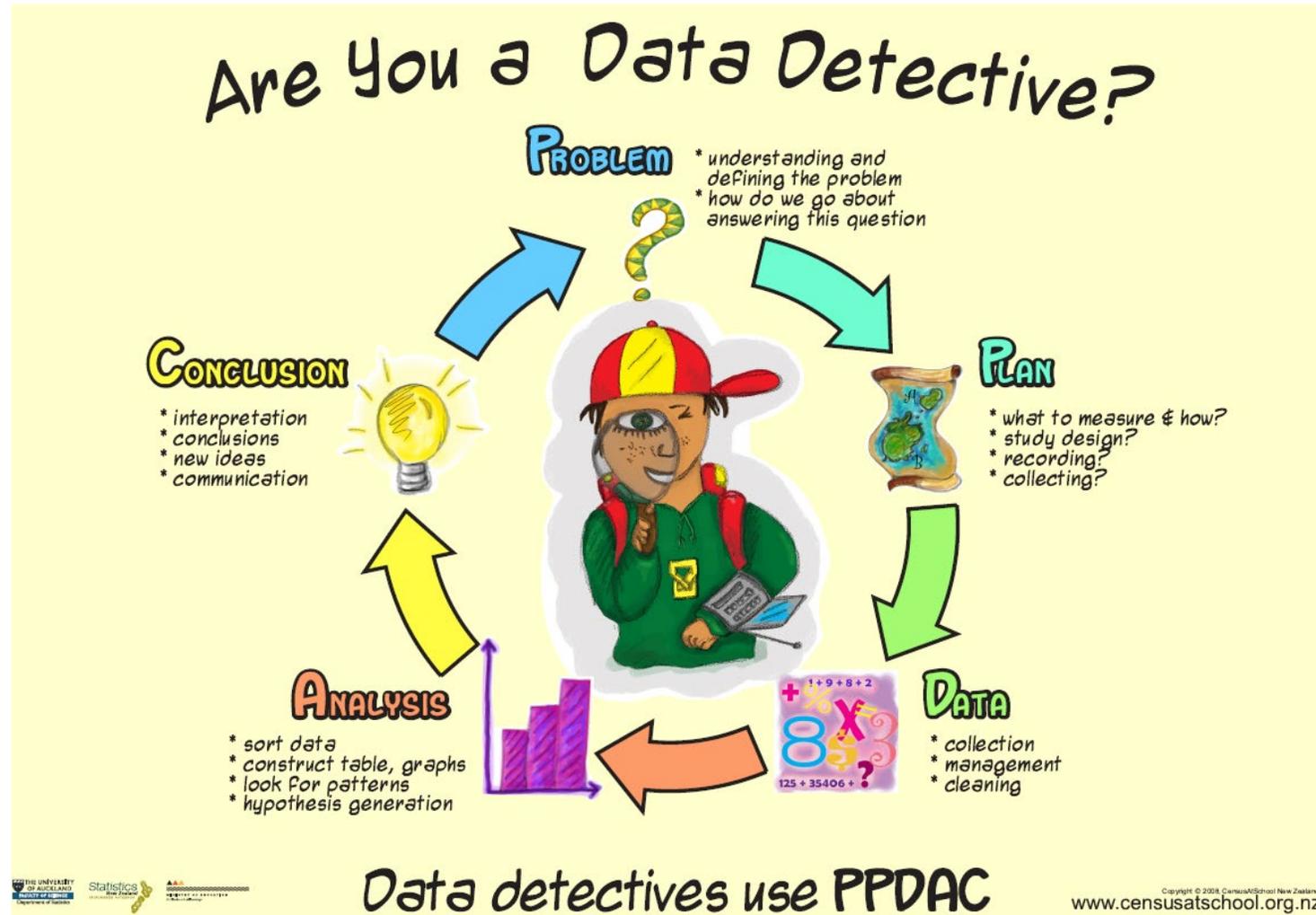
日本モノ作りの絶頂期にあって

- 1980 NBC “I Japan can, why can’t we?”
- 1989 MIT Commission on Industrial Productivity
 - “Made in America”：アメリカ再生のための米日欧産業比較
 - アメリカ復活の必要条件：将来投資⇒人的資源など
- **国家的対応**
 - **企業改善活動の組織化：Six Sigma活動 DMAIC**
 - **企業褒章活動：マルコムボルドリッジ国家品質賞**
 - **初中等教育：SCANSレポート⇒統計を学校教育への展開**
- 1992 A SCANS Report for America 2000
 - Secretary’s Commission on Achieved Necessary Skills, U.S. Department of Labor
 - 日独の労働者の70%が持っている力量を米国の労働者が持っているか
 - 日本の奇跡に学べ

各国の初中等統計教育の変貌: Census At School **New Zealand**

初中等教育での科学的問題解決教育：PPDACサイクル

<https://new.censusatschool.org.nz/wp-content/uploads/2012/11/data-detective1.pdf>



科学技術 総合戦略

科学技術イノベーション総合戦略 (2015年6月 閣議決定)

- 科学技術基本計画の中長期方針の下、各年度の重点項目を明確化
- 成長戦略の一環として、毎年策定、閣議決定

- 第5期科学技術基本計画の始動に向けた3つの政策分野
⇒ **大変革時代の挑戦**、地方創生、2020東京オリンピック・パラリンピック

重点的取組

- ・ 「**超スマート社会**」の実現に向けた**共通基盤技術**や**人材の強化**
(IoT、**ビッグデータ解析**、**数理科学**、**AI**、サイバーセキュリティ、センサ、ロボット、素材、ナノテクノロジー等)

「重点的課題」に、**強烈な問題意識**が示された

第1章 大変革時代における挑戦 2. 重点的課題

「我が国では欧米等と比較し、データ分析のスキルを有する人材や**統計科学を専攻する人材が極めて少なく、・・・危機的な状況にある。**」

米国DS教育の健全性と 日本の人工知能に対する誤解

- AIは画期的な予測技術：No
 - 予測に用いる統計的機械学習は回帰分析・層別分析の一種！
 - 回帰予測とは何か？
 - 自動層別とは何か？
 - **最適意思決定とは何か**
 - **機械学習が採用している幾つかの原理とは何か？**
- ビッグデータからNowcastや行動の可視化は完璧になる：Yes
- 既存ビッグデータからの将来予測や行動選択は完璧になる：No
 - どんなモデルを用いても将来の不確かさを超えた予測は不可能
- 行動とその結果の評価モデル（損失関数）が存在する場合
 - 自律的実験データを採取（シミュレーション）分析することで
 - 人工知能が自律的にPDCAサイクルを回すことで
 - 行動選択は既存の最適行動を上回るパフォーマンスとなる；Yes
 - 実験計画で**新たなデータを創らない限りは、革新的技術は生まれない！**

専門職修士課程（産業界への人材輩出）そのカリキュラム

DS修士課程基礎コース

2.5(統計学) : 2(計算機科学) : 0.5(応用AI)

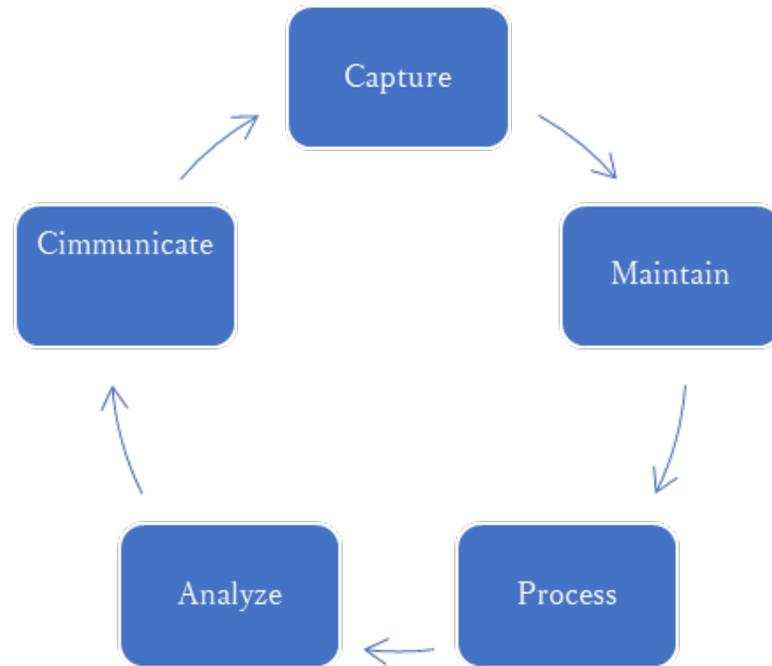
DS修士課程発展コース

2: 1: 2: 2(コミュニケーション・倫理・マネジメント)

科目名	キーワード
DSプログラミング入門	オブジェクト指向、Pythonによるデータ分析、モジュール等の開発、Jupyter Notebookによるプレゼン、GitとGithub
研究計画とデータや分析への適用	研究計画、研究の問いかけの定式化、データと意思決定、認知バイアスの理解、説得とアクションのためのデータ、データと固有知識の統合、データで物を語る
DSのための統計学	研究計画、様々な量的研究方法と統計的分析技法、記述統計、推測統計、サンプリング、実験計画、仮説検定、最小二乗法と回帰分析、ロジスティック回帰、Rによる実社会データ分析実践
データ工学の基礎	データセットの保存、管理、処理システムの基礎、分析によるソリューションのアーキテクチャー、分散データ処理、リレーショナルデータベース、グラフデータベース、ストリーミング分析、クラウドコンピューティング
応用機械学習	実験計画、機械学習アルゴリズム、特徴量エンジニアリング、予測か説明か、ネットワーク解析、協調フィルタリング、

科目名	キーワード
実験と因果推論	実験計画に基づく因果関係の推論。無作為化、統計分析と結果の伝達、データのクリーニング、データマイニングと探索
データの背景：人間と価値	倫理と法的枠組み、政策分析、DSライフサイクル全体での倫理問題、データの収集やタイプなどへの倫理・政策的制約、ケーススタディ
クラウドとエッジでの深層学習	画像やビデオなど大規模データの深層学習などの処理環境の実践、クラウド、分散ストレージ、イーサリアムブロックチェーン、Keras, IBM Watson, 利用可能なApacheのソフトウェア
質的回答・時系列・パネルデータのための統計的方法	断面的データと時系列データの可視化、確率と数理統計の重要概念、古典的回帰分析、変数変換、モデルの規定、因果推論、操作変数法、時系列モデル (ARIMA, GARCH, VAR モデル)、統計的予測、時系列データを伴う回帰分析
大規模機械学習	単一あるいは複数計算機上での機械学習アルゴリズムの実装、AWS, テラバイト級データの作業における問題点、ペタバイト級データの機械学習パイプライン、アルゴリズムの設計 (決定樹、グラフ処理、最急降下法)、並列処理
深層学習を用いた自然言語処理	言語現象の紹介と機械学習による分析、情報抽出、機械翻訳、センチメント分析、要約
データの可視化	探索的データ分析、効果的な文書コミュニケーション、効果的なデータの可視化プレゼンテーション、人間の知覚のデザイン

米国データサイエンス大学院ランキング1位： UC Berkley: オンライン大学院：質保証機関認証



データサイエンスのライフサイクル
これができるエキスパートを育成
専門職育成過ぎる嫌い

- **Capture**

- データ取得、データ入力、信号入力とデータの抽出

- **Maintain**

- データウェアハウスの作成、分析可能なデータの編成（クレンジング）、データ処理、データアーキテクチャー

- **Process**

- データマイニング、クラスタリング/分類、データモデリング、データの要約

- **Analyze**

- 探索的分析と検証的分析、予測分析、回帰分析、テキストマイニング、質的分析

- **Communicate**

- レポートニング、可視化、ビジネスインテリジェンス、意思決定

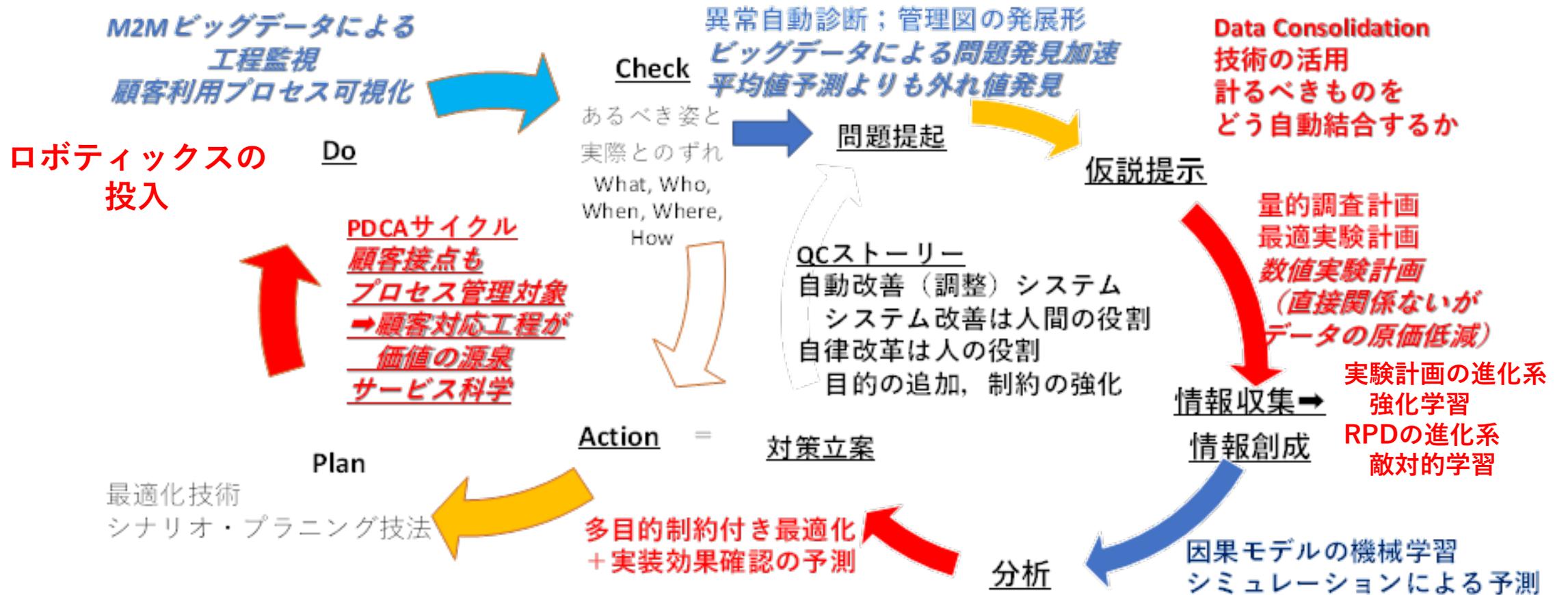
Berkleyの企業幹部向けDSリテラシーコース 10週間70時間オンライン2850\$:

- 確率的意思決定
 - データ分析の前に離散・連続データの比較等DSの背後にある基礎概念等
- 標本データの生成
 - 用語、抽出方法の種類、標本データの質、第1種の過誤と第2種の過誤等
- 仮説検定
 - 仮説検定、信頼区間、実験の基本原則、4M（動機、方法、メカニズム、メッセージ）モデルによる問題の定式化
- 標本データからの情報推論
 - 直線パターンや曲線パターンの探索とデータに線形モデルを当てはめる様々な方法の理解と産業界での応用
- 基本的回帰モデル
 - より精緻なビジネス意思決定の理論の基幹となる単回帰分析の使い方、
その利用上の注意とビジネスの意思決定がどう改善するのか
- 発展的回帰モデル:重回帰分析の様々な利用方法
- 予測と機械学習
 - 機械学習の基礎原理と様々な応用を分かりやすく解説、教師付き学習と教師無学習、時系列回帰予測等
- **A/BテストとのDSチームの効果的構築**
 - **組織のデータ主導型文化構築、データサイエンスとの効果的連携戦略と陥りやすい問題**

おわりに データサイエンス時代だからこそ Deming-石川の統計哲学を再興

本来の日本産業界の強みはマネジメントとソリューションの結合体
ただのDS専門職ではなく、Functional Manager

データ駆動型時代でも骨格は維持 フェイズ毎の適用技術が自動化しつつあるだけ 今、自分はこの立ち位置で人工知能を活用しているかを知る



Data Science for Quality?!

The ultimate purpose of taking data is action.

Every empirical state of science is a prediction.

Deming(1942) On a Classification of the Problems of Statistical Inference, JASA 37

石川馨先生「日本にはアクションが必要だ」

本日が様々な方々の活動や見解を共有する場となることを期待します