

学術と社会の問題解決を支える 統計エキスパートの育成に向けて

統計数理研究所長 椿 広計

2024/09/09

第4回統計エキスパート育成に向けたワークショップ

天文学から社会物理学： 社会問題解決のための応用統計学生成

L. A. Quetelet (1798~1874)

- ブリュッセルに天文台建設活動：1828設立
- 1835: 社会物理学（統計学の社会科学への適用）
 - **社会的要素を変える原因の影響**
 - **犯罪学（犯罪の社会的要因）** ・ **自殺率**
 - **Gauss分布と平均人、BMI(Body Mass Index)の提唱**
 - **F, Nightingale(1820~1910)：Applied Statistics 提唱への絶大な影響！**
 - 「社会学」勃興（**Durkheim：社会的事実**）に影響を与えたのか？
- 1853 第1回国際統計会議議長
- データに基づく問題解決
 - **Snow：1854, 「この水道の水を飲むとコレラになる」**
 - **Koch: 1884 コレラ病原体としてのコレラ菌発見**

統計家≡統計エキスパート≠統計学者 「米国労働統計局」が公式に示す職務記述

- 15-2021数学家と15-2041統計家
 - 特定の問題や疑問に答えるために
どのようなデータが必要であるかを判断
 - ビジネス、工学、科学、その他の分野の
実際の問題解決に数学理論や技法を適用
 - データを収集するために調査、実験等を計画
 - データ分析のための
数学的・統計的モデルを開発
 - データを解釈し、分析結果を専門家や
非専門家に説明
 - 統計ソフトウェアを使用してデータを分析し、
ビジネスにおける
意思決定を支援する可視化を作成
- 15-2051:データサイエンティスト
 - どのようなデータが利用可能で、
プロジェクトに有用かを判断
 - データの収集、分類、分析
 - アルゴリズムとモデルの作成、検証、
検査、および更新
 - データ可視化ソフトウェアを使用して
調査結果を提示
 - データ分析に基づくステークホルダー
へのビジネス提案

米国職業分類15-2041

15:計算機・数学職 15-20XX 数学職

米国労働統計局Occupational Outlook Handbook

<https://www.bls.gov/ooh/>

2023/09改訂

職名	職業小分類コード	2022就労者数	2032までの増加率(%)	主要な学歴	主要就職先
保険数理士	15-2011	30,000	23 (12位)	学士	金融保険79.8%
数学家	15-2021	2,300	2	修士	中央政府53.9%,
OR分析家 Operations Research	15-2031	109,900	23(14位)	学士	専門科学技術サービス25.4% 金融保険24.0%
統計家	15-2041	33,300	32 職種第4位	修士	科学研究開発サービス17.9% 中央政府15.9%、大学11.6%
データサイエンティスト	15-2051	168,900	35 職種第3位	学士	専門科学技術サービス32.8% 金融保険17.5% 企業マネジメント8.5%

全職業での10年増加率は3%：情報セキュリティアナリスト 15-1212 168,900の増加率；32%（職種第5位）

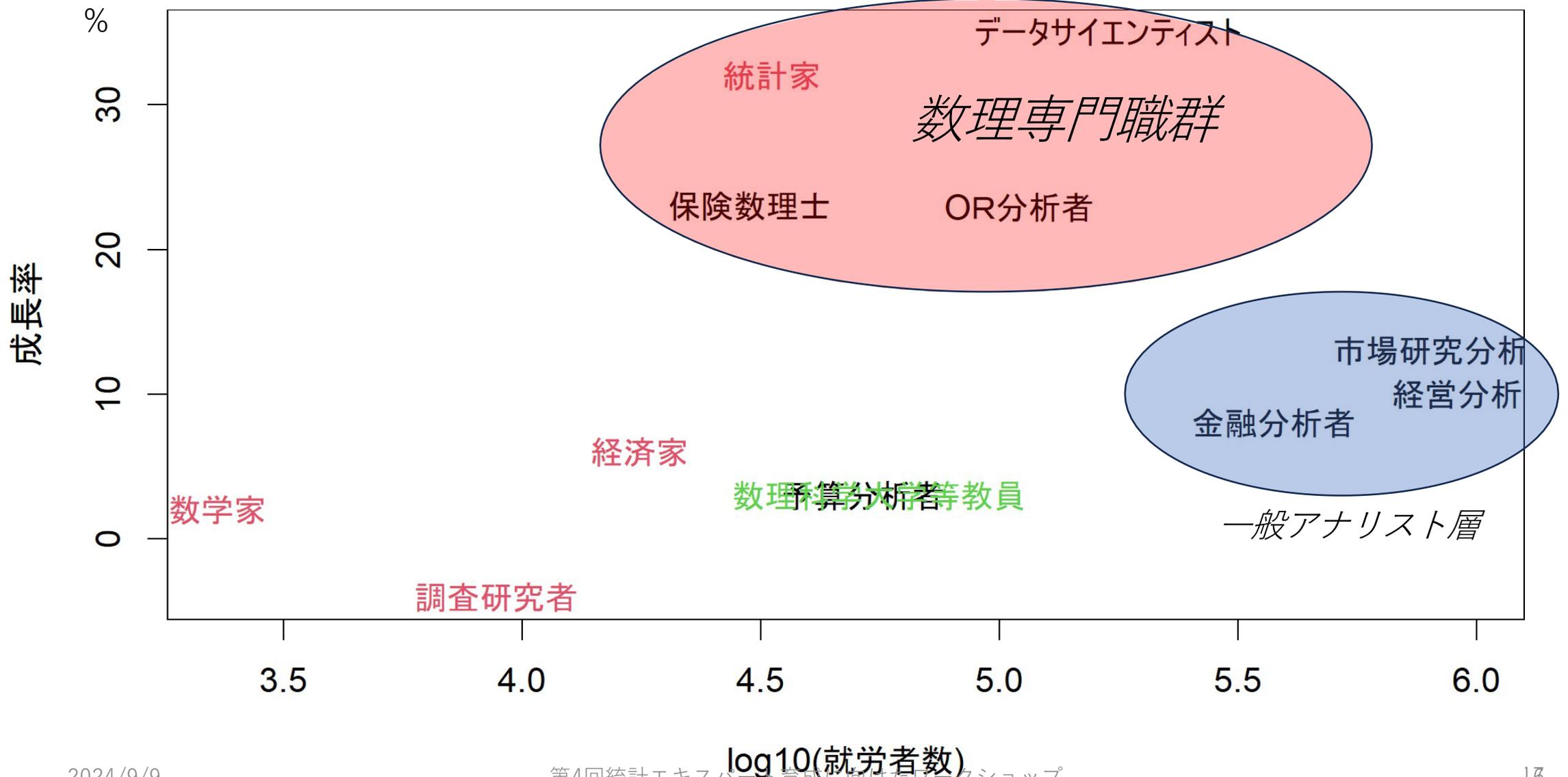
米国職業分類での労働統計局10年成長率予測：トップ20職種
 with the highest projected percent change of employment between 2022-32.
<https://www.bls.gov/ooh/fastest-growing.htm>

職種	10年成長率	年収中央値
Wind turbine service technicians	45%	\$57,320 per year
Nurse practitioners	45%	\$121,610 per year
Data scientists	35%	\$103,500 per year
Statisticians	32%	\$98,920 per year
Information security analysts	32%	\$112,000 per year
Medical and health services managers	28%	\$104,830 per year
Epidemiologists	27%	\$78,520 per year
Physician assistants	27%	\$126,010 per year
Physical therapist assistants	26%	\$62,770 per year
Software developers	26%	\$127,260 per year

Occupational therapy assistants	24%	\$64,250 per year
Actuaries	23%	\$113,990 per year
Computer and information research scientists	23%	\$136,620 per year
Operations research analysts	23%	\$85,720 per year
Solar photovoltaic installers	22%	\$45,230 per year
Home health and personal care aides	22%	\$30,180 per year
Taxi drivers	21%	\$30,670 per year
Personal care and service workers, all other	21%	\$34,670 per year
Veterinary technologists and technicians	21%	\$38,240 per year
Veterinary assistants and laboratory animal caretakers	20%	\$34,740 per year

2022年現在の米国就労者数と10年成長率

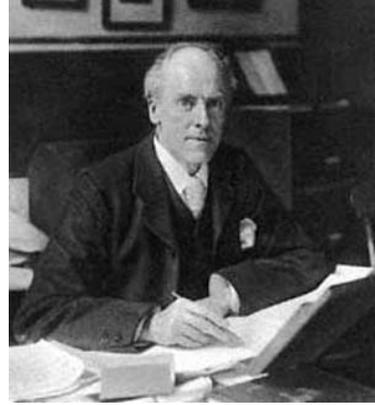
米国労働統計局情報より椿作図



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
科学的問題解決プロセスの提唱と支援ツールとしての統計的方法



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
 - 可視化：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
- 1911 University College Londonに世界初の応用統計学科設立

マネジメントを統計科学に:Shewhart(1891~1967)

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

- 「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
- **一般化された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←仕様
 - Do: 規定された目標を達成しようとする行為←生産
 - See: 目標が達成されたかどうかを判断する行為←検査
- Shewhart流品質管理：CAPD_o：
 - Cは検査ではなく：工程改善のトリガーの発見
 - 時系列解析（Shewhart 管理図）
 - データに基づく問題（価値）の発見は「外れ値の探索(異常検知)」から



https://en.wikipedia.org/wiki/W._Edwards_Deming

石川馨 (1915-1989)



https://www.juse.or.jp/qc_circle/case/

産業界の統計的問題解決 Japanese PDCAサイクル

- 1950: デミング博士来日 (国勢調査指導)
 - 統計的品質管理講義
 - デミング・ホイール
 - 設計→製造→市場投入→市場研究
- 1951: 日本科学技術連盟 (JUSE)
 - 戦後品質管理活動第一世代の講義
 - 石川馨(東大工学部)、朝香鐵一(東大工学部)、水野滋(東工大)
 - 計画→実施→チェック→アクション
 - Plan→Do→Check→Actionと英訳
 - デミング博士の講義を基に日本の産学品質管理第一世代が創成したサイクル
 - Deming先生はPDS(Study)Aを主張

産業界のKAIZEN活動を支える科学的基盤 Deming-IshikawaのPDCA + 問題解決標準シナリオ

- **データ科学的問題解決の標準シナリオ：(問題解決型) QCストーリーの創生と普及**
- 1961 小松製作所改善活動：当初はプレゼンテーションの標準
 - (問題の抽出) テーマの選定と取り上げた理由
 - 現状把握 (データの収集) と目標の設定
 - 要因分析 (データの収集による原因の追究)
 - 原因に対する対策の立案と実施
 - (データによる) 効果の確認
 - 歯止め (標準化と水平展開)
- 1960年代後半：QCプロジェクトの進め方として確立
- 1986:Six Sigma活動：モトローラ：改善活動標準化開始
 - DMAICへの影響：Define, Measure, Analysis, Improve, Control：ISO 13052-1: 2011
 - 英国Royal Statistical Societyと中国が主導：産業界の問題解決を行う統計家の要員認証規格
 - ISO18404:2015 プロセス改善における定量的方法—シックスシグマとリーンの実装関連する主要な担当者とその組織能力
- 2000: 欧米初中等数学教育への影響
 - 1990 米国労働省SCANS Report：日本の労働者の「自立的問題解決能力」の学校教育での習得
 - PPDACサイクル (Smith 1943) 導入 ¥ Problem, Plan, Data, Analysis, Conclusion

問題解決の標準シナリオを支える 初等的統計技術：小学校からの問題解決

- 1961～QCストーリーの各フェイズを支える統計的ツール
 - 現場のデータ収集：チェックシート
 - 重点志向による問題の絞り込み：パレート図
 - 問題の発見と効果の検証：管理図／グラフ
 - 問題の発見と効果の検証：ヒストグラム
 - **要因の定性的検討：特性要因図（石川ダイアグラム）**
 - 定性的；仮説形成ツール
 - 要因の探索：散布図
 - 要因の探索：層別
- PPDACサイクルは、日本では小学校5年算数の教科書に反映

社会に活かす統計の考え方

デジタル社会に求められる
データへのリテラシー(素養)

データに基づき
社会問題を解決するプロセス

**STEAMを貫く
問題解決プロセスによる
地域・社会・世界への貢献**

身近な
問題

日常生活のグラフ表示

小学校

地域の
問題

記述統計的方法

中学校

社会の
問題

データと確率、推測統計の基礎

高校

世界の
問題

数理・データサイエンス
人工知能・ビッグデータ

大学

統計は不確かさ・バラツキ・偶然等を
対象に研究します。

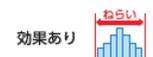
不確かな現象に対する意思決定の改善

コンピュータで乱数発生
偶然を短時間で大量に発生
シミュレーションによる効果の確認

**STEAMを貫く
問題解決すごろく**

GOAL
問題解決
大成功!
ルール化しよう
振り返りをしよう

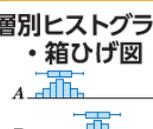
問題解決(対策)エリア
原因への対策立案とその効果の確認
および対策の実施・ルール化

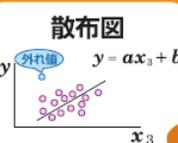
効果を確認めよう
効果あり 
効果なし(別の問題が出た) 
効果があった → 1マス進む
効果がなかった → 4マス戻る

原因を追及しよう
原因がわかった → 1マス進む
原因がわからなかった → 2マス戻る

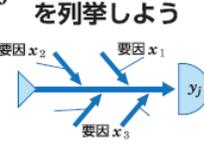
原因追及エリア
要因(原因の候補)と
結果の関連性の把握

棒グラフ・地図グラフ  小学校

層別ヒストグラム・箱ひげ図  中学校

散布図 $y = ax_3 + b$  高校

質的データ $x_2 = \{A, B\}$
量的データ

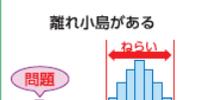
絞り込んだ問題 y_j の要因(原因の候補)を列挙しよう
要因 x_2 x_1 x_3 

START
テーマを
決めよう

サイコロの目は偶然だけど、
その背後には数理があるんだ。

データを探って
棒グラフ・
ヒストグラムを
描いてみよう

問題発見エリア
平均やバラツキが
予想と異なる気づき

問題を発見しよう
平均値が偏っている 
ばらつきが大きい 
離れ小島がある 

問題を絞り込もう
平均値が偏っている
→ 偏りの問題 y_1
ばらつきが大きい
→ ばらつきの問題 y_2
離れ小島がある
→ ルール不遵守の問題 y_3
絞り込みに成功 → 1マス進む
失敗 → 2マス戻る

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

Deming・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」

統計家は今、自分がどこに立っているかを知るべき

検査 & 統計的意思決定
統計的过程管理

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

特性要因図

連関図

要求品質展開

仮説提示

どう解くか何を
どう調べるかか

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

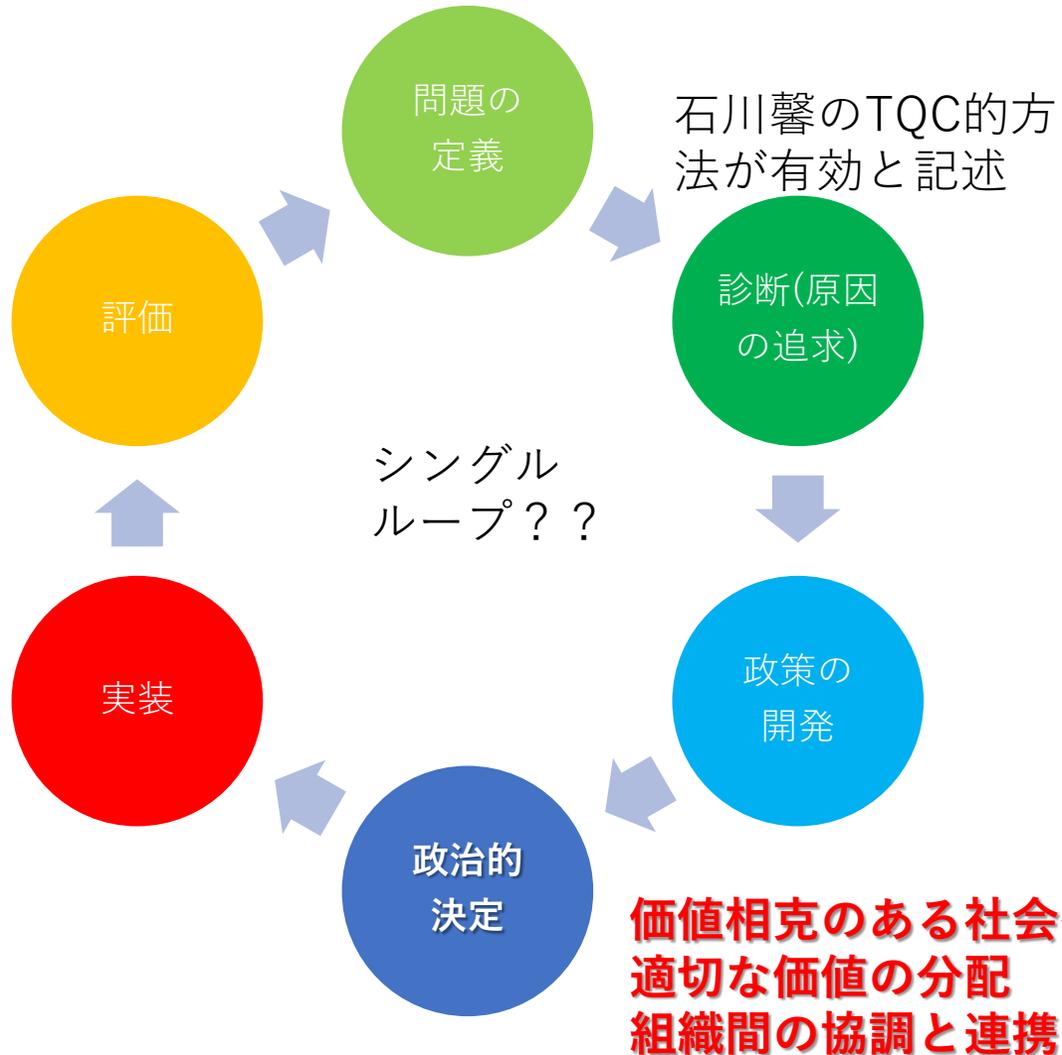
因果モデリング：検証的解析
層別分析・回帰分析・
時系列解析

Action = 対策立案
対策実装 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

企業活動マネジメントサイクルと 社会政策マネジメントサイクルの微妙な違い



• 政策のサイクル

- Hsiao, M.R.W., Berman, P. and Reich, M. :
• *Getting Health Reform Right: A Guide to Improving Performance and Equity*, Oxford University Press, (2008) 中村安秀訳:
事前ガイド医療改革をどう実現すべきか, 日本経済新聞社, (2010)
- シングルループモデルには不満

AI時代でも統計的マネジメント枠組みを維持！ フェイズ毎の適用技術が進化(自動化)するだけ

統計家はどの立ち位置で人工知能を活用しているかを知る



そういう統計家を日本は育てたい
それでは本日の産官学WS本格開始

“Data are for Prediction and Action”

The ultimate purpose of taking data is action.

Every empirical state of science is a prediction.

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