

# データサイエンスの基礎と応用

早稲田大学

創造理工学部 経営システム工学科

永田 靖

# 私のTalkの内容

椿先生の論説に啓発されて、私がどういう活動につなげていったかを具体的に紹介

1. ロジット変換
2. 科学の文法
3. 多重比較法
4. グラフィカルモデリング
5. タグチメソッド

# 1. ロジット変換(スライド:2枚)

1992年 日科技連ベーシックコース(BC)のテキストの全面改定  
久米均先生から「計数値データの解析」の執筆を永田(35歳)に依頼

表 ベーシックコースのカリキュラム

序論	技術者・スタッフにとってのTQM	基礎	データのとり方・まとめ方	品質管理技法	管理図
TQM実践のための要素技術	問題解決法 顧客ニーズの解析と管理 設計開発における品質管理 トラブル予測と未然防止 プロセスの設計と管理 購入と外注の管理 品質保証体系の構築 経営管理システムの構築と運営	統計	統計的方法の基礎 検定と推定		抜取検査 信頼性工学 官能評価と感性品質 新QC七つ道具
		高度な統計手法	回帰分析 分散分析 実験計画法 サンプリング 計数値データの解析 ノンパラメトリック法 ばらつきの推定と管理 タグチメソッド 多変量解析法	演習等	問題解決ケーススタディ 工程解析ケーススタディ QFDケーススタディ 回帰分析ケーススタディ 実験計画法ケーススタディ QCゲーム 演習 班別研究発表会

品質管理 実施法

統計的方法

講義:6時間×30日=180時間

その1年前, 日本規格協会のQSセミナーの新テキストで, 椿先生がロジット変換, ロジスティック回帰分析を記述. ⇒古典的な解析に一石を投じられた.

椿先生のテキストと「欧米で認められていない古い内容を取り込むことはできない」という主張に刺激されて、私は次の内容のテキストを作成した。

## BC 第5章「計数値データの解析」の内容

1. 母不良率に関する検定と推定
2. 母不良率の違いに関する検定と推定
3. 母欠点数に関する検定と推定
4. 母欠点数の違いに関する検定と推定
5. 適合度の検定
6. 分割表による検定
7. 母不良率に関する回帰分析
8. 母不良率に関する分散分析

従来の直接近似＋ロジット変換近似

従来の直接近似＋対数変換近似

従来の方法

ロジット変換による方法

二項確率紙は付録へ、0-1データの分散分析は取りやめ

➡ 年配の講師の方々から反発

➡ 飯塚先生に守ってもらった、感謝

## 2. 「科学の文法」(Grammar of Science) (スライド: 3枚)

ある時期, 椿先生は「科学の文法」について論説されていた.

「ピアソンは、『科学の文法』の中で, 科学的方法の基本は, 合意を形成するための科学的な議論をしやすいように, 事実をわかりやすく分類し, 相関関係を観察し, 法則を発見し, 検証することだとしています. 」(2011 ゑれきてる)

→ 私は, これを授業でのネタにしようと考えた.

大学での授業や社会人セミナーでのスライド

### 検定結果の表現

$$1. H_0 : \mu = \mu_0 \quad (\mu_0 = 6.0)$$

$$\alpha = 0.05$$

$$H_1 : \mu \neq \mu_0$$

$$R : |t_0| \geq t(4, 0.05) = 2.776$$

### 2. $|t_0| \geq t(\phi, 0.05)$ となる場合

(正-a) 検定統計量は棄却域に入る

(正-b) (有意水準5%で)有意である

(正-c)  $H_0$  を棄却する

(正-d)  $\mu$  は  $\mu_0$  と異なるといえる

⇒

検定統計量は棄却域に入るので, (有意水準5%で)有意である.

$H_0$  を棄却し,  $\mu$  は  $\mu_0$  と異なるといえる.

### 3. $|t_0| < t(\phi, 0.05)$ となる場合

(正 - e) 検定統計量は棄却域に入らない

(正 - f) (有意水準5%で)有意でない

(正 - g)  $H_0$  を棄却できない

(正 - d)  $\mu$  は  $\mu_0$  と異なるとはいえない

⇒

検定統計量は棄却域に入らないので,

(有意水準5%で)有意でない.

$H_0$  を棄却できず,  $\mu$  は  $\mu_0$  と異なるとはいえない.

(誤 - a) 無意である

(誤 - b)  $H_1$  を棄却する

(誤 - c)  $\mu$  は  $\mu_0$  に等しいといえる

よく間違う漢字:

○有意, ×優位, ×有為;

○直交, ×直行

○仮説, ×仮設;

○群内・群間, ×郡内・郡間

なぜ、「記号」や「言葉づかい」にうるさくいうのか？

**統計学は「あらゆる科学の言語」だから.**

言語には「文法」がある. それをおろそかにすると,  
「相手に伝わらない」「誤解される」

**She are a girl.**

**There is three pen.**

**意味は伝わるけれど, これ  
でよいか？**

**「科学の文法」The Grammar of Science**  
**(カール・ピアソン著)**

**カール・ピアソン(1857-1936):**  
**近代統計学を作った統計学者**

### 3. 多重比較法(スライド:5枚)

1995年頃, 吉村功先生, 大橋靖男先生, 医薬安全性研究会の皆さんから, 多重比較法の解説書の作成のお薦めが 私と吉田道弘氏 にあった.

当時は, わが国において多重比較法に関する成書はなかった.

椿先生や広津先生の書籍・論説・論文, 欧米の多重比較法の成書を勉強



1997年11月に出版

当時: 吉田道弘氏は武田薬品工業に勤務

現在: 岡山大学医学部教授

その後, 多重比較法は様々な方面から注目を集め, 発展している.

現在, わが国でも, 多くの専門書や解説書, 多くのページ数を多重比較法に割り当てた書籍が刊行されている.

## 多重比較法の特徴を一言で言うと・・・

下手な鉄砲でも数撃てば当たるので、  
下手な鉄砲なら数撃っても当たらないようにする  
( $\alpha$ を厳しくする)

多重比較法における $\alpha$  (有意水準) って何？

その結果、上手な鉄砲でも当たりにくくなる  
(検出力が下がる)

多重比較法における検出力って何？

たとえ話・・・(佐久間(1987)より, やや改題)

あるレストランの地下にはたくさんのワインが貯蔵

20本に1本の割合(5%)で酸っぱくなった不良品が混入

★客がきて1本のワインを注文・・・

ソムリエは20回に1回謝罪

★パーティでは6本単位で出す・・・

ソムリエは, パーティあたり26%, およそ4回に1度の割合で謝罪

◇ 1本がまともなワインである確率  $0.95$

◇ 2本がまともなワインである確率  $0.95^2$

少なくとも1本が不良品となる確率  $1 - 0.95^2 = 0.0975$

◇ 6本がまともなワインである確率  $0.95^6$

少なくとも1本が不良品となる確率  $1 - 0.95^6 = 0.2649$

★パーティでは6本単位で出す...

ソムリエが、パーティあたり**5%**の割合で謝罪するには、貯蔵ワインの不良率  $P$  はいくらであるべきか？

◇ 1本がまともなワインである確率  $1 - P$

◇ 6本がまともなワインである確率  $(1 - P)^6$

少なくとも1本が不良品となる確率  $1 - (1 - P)^6 = 0.05$

これより,  $P = 1 - (1 - 0.05)^{1/6} = 0.00851$  ( $> 0.00833 = 0.05 / 6$ )

(注)一般に次の不等式が成り立つ

$$P = 1 - (1 - \alpha)^{1/k} > \alpha / k$$

薬効については:

$H_0$ : 薬効がない vs  $H_1$ : 薬効がある  $\Rightarrow$  有意であれば薬は認可される

毒性については:

$H_0$ : 毒性がない vs  $H_1$ : 毒性がある  $\Rightarrow$  有意でなければ毒性なしと言える?

★検定で有意でないときに $H_0$ の主張は無理

(そのためには、意味のある効果量を決め、それ以上の効果があるときには高い検出力で有意になるようにサンプルサイズを設定しておく必要がある)

★多重比較で有意でないときに $H_0$ の主張はもっと無理

(必要なサンプルサイズは非常に大きくなってしまう)

椿先生は、「多重性」の問題について、「水準間比較の多重性の問題」「多項目検定の多重性の問題」「多時点比較の多重性の問題」「多種検定適用の多重性の問題」「解析対象(サブグループ解析)の多重性の問題」「分割表における検定の区切り直し検定の問題」「中間解析の問題」に分類して解説された。

椿先生は、さらに、毒性の問題において「多重比較を用いて有意でないとき、毒性がない」とみなすことに異議を唱えられていた。

## 4. グラフィカルモデリング(スライド:1枚)

1994年 日本品質管理学会, 椿先生が「テクノメトリックス研究会」を設立

初代主査: 宮川雅巳先生(東京工業大学)

現在も続いています

回帰分析の正しい理解, 偏相関係数の理解を最初の研究テーマ

椿先生が「グラフィカルモデリング」を教示してくださった.

→ 研究会の主要テーマ: グラフィカルモデリング

→ 芳賀敏郎先生, 廣野元久氏が解析ソフトを作成

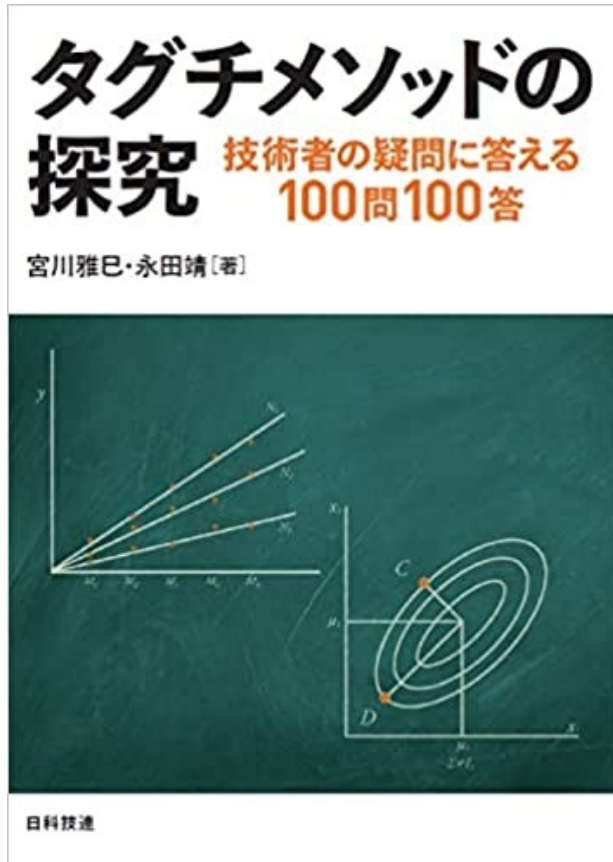
→ 宮川雅巳先生が『グラフィカルモデリング』(朝倉書店)を上梓

まえがきの一節:

「統計的因果分析に関心のあった筆者にグラフィカルモデリングの存在を教えてくれたのは椿広計先生(慶應義塾大学)であった。」

→ 永田が2代目主査のとき, 『グラフィカルモデリング実際』  
(日本品質管理学会テクノメトリックス研究会編, 日科技連出版)を上梓 13/14

## 5. タグチメソッド(スライド: 1枚)



2022年1月発刊

あとがき

...

私はタグチメソッドをこの研究会で学びました。宮川雅巳氏と椿広計氏から薫陶を受けました。お二人は田口玄一先生を熱烈に信奉されていました。

...

2021年10月  
永田 靖

終

ご清聴ありがとうございました