Guiding Readers to the Focus and Context of Industrial Statistical Reports

Ken Wakita (Tokyo Tech) Kohei Arimoto (TDB – Teikoku Databank)

2019/6/24

TDB Reports

- TDB: Japan-based business intelligence corporation founded in 1900.
- TDB Reports Daily PDF report on Japanese companies. Target readers are Japanese business leaders. Topics ranges:
 - Company capital investment incentives
 - Changes in industrial structure
 - Latest financial trends
 - Market surveys
 - Employment trends, ...

TDB

特別企画:全国・女性社長分析(2019年)

新任女性社長の就任経緯をみると、最も割合が高いのは「同族承継」(41.8%)で、以下「創業 者」(40.5%)、「内部昇格」(9.8%)の順。男性社長と比較すると、「同族承継」の割合が突出し て高いほか、「創業者」の割合も高かった。他方、「内部昇格」の割合は男性社長に比べて約半分 となったほか、「出向・分社化」(2.8%) は約4分の1程度にとどまった。



他方、最も低かったのは「岐阜県」(5.2%)となり、全国(7.9%)と比較して2.7ポイント低い水準となった。以下、「長野県」(5.9%)、「滋賀県」(6.0%)、「愛知県」(6.1%)、「静岡県」 (6.3%)の順。下位には特に中部地方の県が多く目立った。



Report structure

- (Text, Illustration), (Text, Illustration), ...
- Data reference
 Number → value
 representation in the
 illustration
- Statistical claim with "female leadership was the lowest in the construction industry"



Pattern & Formulation

- Simple data reference: "The ratio of female CEOs in 2019 in real estate (16.7%)"
- **Compound data reference**: "increase of 7.5% over the past 30 years"

- Statistical claims
 - X is highest in Y \Rightarrow Y. idxmax \equiv X
 - X is high in ... $\Rightarrow \min X > X$. mean ambiguity resolution
 - change / no change $D[Y] > D[Y'] \quad D[Y] \equiv D[Y']$ claims on temporal trend

• **generally** increased universal property $\Rightarrow \forall x \in \text{Industries}$

Focus & Context

- Focus: a set of data values under consideration in the text
- Context: a set of data or data distribution that is being compared to the focus values.
- **Example**: The "real estate" industry also had the highest increase of 7.5% over the past 30 years.
 - Focus: Data[2018, "Real Estate"], Data[1988, "Real Estate"]
 - Context: D[2018,x], D[1988,x] for all x in industries $\Rightarrow (D[2018] - D[1988])$. idxmax \equiv Real Estate

The Aim

Linking text and illustration aiming at:

inducing reader attention from report text content to the associated data illustrations (tables and charts). We hope to evoke a more self-motivated, critical, and explorative reading attitude by encouraging readers to:

- 1. locate illustrated data items corresponding to the text;
- 2. verify "claims" described in the text through statistic testing; understand the statistical standards that lead to factual descriptions in the text;
- 3. learn to discover undocumented insights on their own.

Ongoing & future work

- Analyzing more reports to test the generality of the proposed method
- Automatic text synthesis (Latif & Beck 2019)
- Task-based evaluation

- Poster #35
- more on this presentation at: <u>http://j.mp/tab-reports</u> (talk slide, demo, poster, and paper)



(Latif & Beck 2019) Interactive map reports summarizing bivariate geographic data, Visual informatics, 10.1016/j.visinf.2019.03.004.