

「データ利活用サポートデスク」の挑戦

1 「データ利活用サポートデスク」とは

県統計分析課では、庁内におけるEBPMに基づく施策展開を推進するため、庁内各部局等からリサーチデザインやデータの利活用等に関する相談を受付する窓口として、「データ利活用サポートデスク」という相談対応サービスを令和6年6月にスタートしました。

2 相談対応

このデスクは庁内の職員がチャットツールのボットを通じて、データ利用に関する相談を統計分析課が受け付けする仕組みとなっています。受付した質問は統計分析課職員が対応し、自ら調べて回答することでデータリテラシーや分析スキルの向上に取り組んでいます。

回答に当たっては、専門的な知識が必要な場合に備え、弘前大学人文社会科学部に所属する経済学の教授陣や株式会社あおもり創生パートナーズの専門家に加え、県外の大学等の大学教授などと連携して相談対応に当たっています。



(サポートデスク担当の坂本主事)



(チャットツールボットの画面例)

3 相談内容

相談内容としては、「こういったデータがあるかどうか」というデータの存在や、ダウンロード元、EXCELを使った簡単なクロス集計やグラフによる見える化といった基本的なことから、サンプルサイズ的设计、調査票的设计、回帰分析、産業連関表を用いた経済効果の分析、主成分分析、テキストマイニング、GIS(地理情報システム)を使った見える化など、幅広いデータ利活用に関する相談に対応しています。

4 今回紹介するケーススタディ

本稿では、1～4の事例について紹介していきます。

<ケーススタディ一覧>

ケーススタディ 1 議事録の見える化

～AIと統計ソフトを組み合わせたワードクラウドの作成～

ケーススタディ 2 人口シミュレーション

～合計特殊出生率2.0に道筋をつけるシミュレーション～

ケーススタディ 3 質的データの解析

～高校生の意識調査における回答同士の関連性を明らかにする～

ケーススタディ 4 効果的な調査票設計

～マメコバチの飼養管理等に関するアンケート調査票のブラッシュアップ～

ケーススタディ 1 議事録の見える化

～A I と統計ソフトを組み合わせたワードクラウドの作成～

自治体の多くの部署で議事録が作成され、公開されています。長いもので2時間程度の会議の議事録を一つ一つ確認していくことは大切なことですが、発言のポイントを短時間で押さえるには、「見える化」に適した、「ワードクラウド」が有効です。今回は、会議で使用することを想定したワードクラウド化のケースについて紹介します。

(1) 相談内容

青森県における少子化の要因分析や政策立案、政策効果の検証等を行い、合計特殊出生率2以上に道筋をつけるこども・子育て「青森モデル」を確立することを目的に、青森県こども未来県民会議を設置しました。本会議は令和5年度中に3回会議を開催しました。

令和6年度の第1回会議で、令和5年度中の3回の会議で出た意見を「見える化」し、更なる具体的な議論につなげるために、議事録をワードクラウド化してほしい。

(2) 相談対応

まずは、議事録を県のホームページからダウンロードしました。

<青森県こども未来県民会議>

<https://www.pref.aomori.lg.jp/soshiki/kodomo/kodomo/kodomomirai-kenminkaigi.html>

まずは、この約2万字の議事録の中から、知事、事務局、司会の発言を除いた会議委員の発言のみを抽出します。そして、各委員の発言を一つ一つ分割します。その分割した1つ1つの発言を①出会い・結婚、②妊娠・出産、③子育て費用の軽減、④子育てしやすい環境づくり(ソフト面)、⑤育てしやすい環境づくり(ハード面)、⑥仕事と家庭の両立、⑦女性の定着・還流の7つの分野に割り振っていきました。割り振り方は、自治体向け ChatGPT 活用サービス「LogoAI アシスタント」を用いて、ある発言が①～⑦のいずれに当たるかを判断してもらいます。

「①出会い・結婚」と回答が返ってくるものもあれば、「①出会い・結婚」と「③子育て費用の軽減」と複数の分野であるという回答も返ってきます。複数分野の回答が返ってきた場合には、その発言はそれぞれの分野の発言として取り扱いました。このように振り分けて、7つの分野に分かれた会議委員のみの発言議事録を作成し、この議事録を使って、統計ソフトR(アール)¹により、日本語形態素解析を行うRパッケージ「RMecab(アール・メカブ)²」を用いて分析しました。まずは、Rに取り込んだ議事録を動詞、名詞、形容詞、副詞、助詞等の形態素に分け、そ

¹ 統計ソフトRは、変数宣言などをする必要がなく、コーディングがシンプルな言語であり、WEB上でたくさんのライブラリーが作られている。

² RMecabは、Rでテキストマイニングを実行する際に必要なパッケージ及びライブラリーのことです。

ケーススタディ2 人口シミュレーション

～合計特殊出生率2.0に道筋をつけるシミュレーション～

多くの自治体において、人口シミュレーションを2015年の「まち・ひと・しごと創生総合戦略」の策定及び2020年の改定に合わせて策定しています。今回は、特定の条件を設定した場合の人口シミュレーションについて紹介します。

(1) 相談内容

こども・子育て「青森モデル」において合計特殊出生率2.0をめざす上で、合計特殊出生率だけで人口を維持するとなれば、どれくらいの合計特殊出生率の上昇が必要なのかシミュレーションしてほしい。

(2) 相談対応

シミュレーションに当たっては、いくつかの条件設定が必要です。条件設定に当たっては、内閣官房デジタル田園都市国家構想実現会議事務局・地方創生推進室(現：内閣官房新しい地方経済・生活環境創生本部事務局・内閣府地方創生推進事務局)が発行している「地方版総合戦略の策定等に向けた人口動向分析・将来人口推計の手引き」を基に行いました。

<地方版総合戦略の策定等に向けた人口動向分析・将来人口推計の手引き
(令和6年6月版)>

https://www.chisou.go.jp/sousei/about/chiouban/pdf/tebiki_20240624.pdf

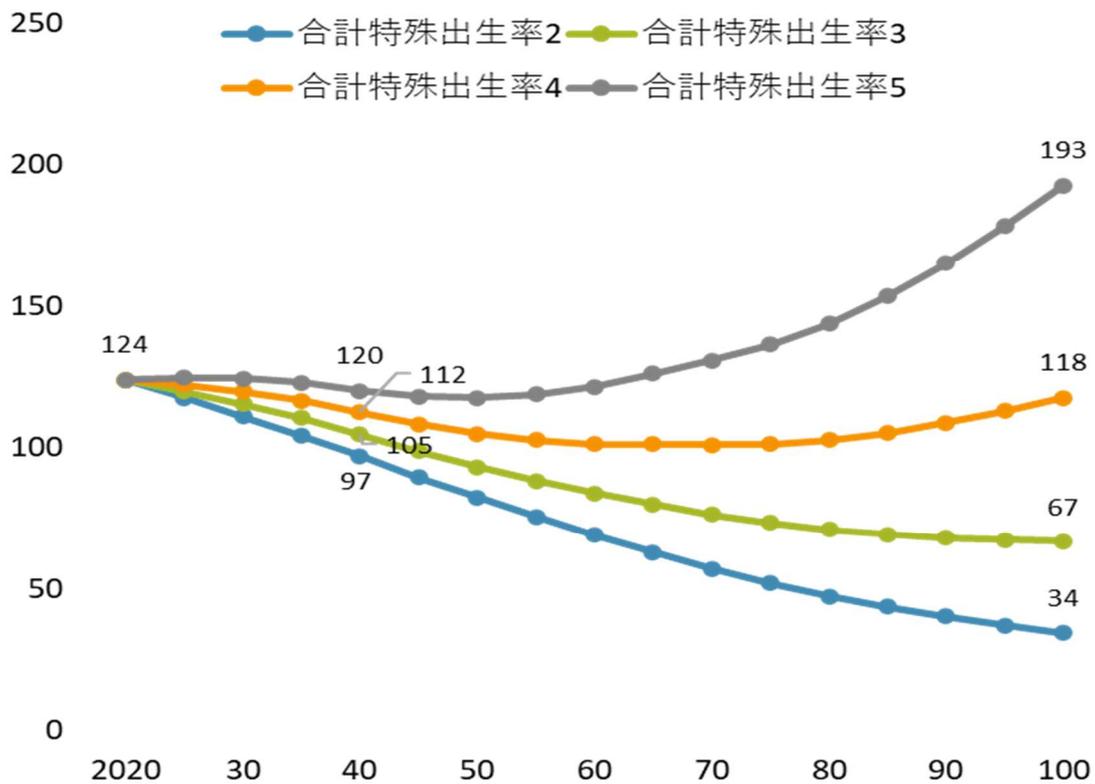
人口推計に当たっては、「人口変動の三要素」と呼ばれる出生・死亡・移動(転入・転出)を仮定していく必要があります。今回は合計特殊出生率だけを変動させていくという相談でしたから、死亡や移動については国立社会保障・人口問題研究所が2023(令和5)年に公表した日本の地域別将来推計人口(2023(令和5)年推計)における最終的な推計結果と整合的な、将来の生残率、純移動率を設定したうえで、「青森県基本計画『青森新時代』の架け橋」2040年時点の合計特殊出生率を2.0、3.0、4.0、5.0と1ポイントずつ変更していきました。また、2020(令和2)年から2040(令和22)年の間の合計特殊出生率については、線形補完により数値を補完しています。

(3) 分析結果

2040年時点で合計特殊出生率2.0だった場合には、2100年には34万人まで減少します。また、合計特殊出生率3.0であっても、現在の人口の半数程度に当たる67万人となります。合計特殊出生率4.0まで上げて、現在の人口とほぼ同等の規模を維持できます。さらに、合計特殊出生率4.0まで上げると、193万人と現在の人口よりも多くなります。

このように、合計特殊出生率の改善だけで現在の人口規模を維持・増加させていくのは非常に困難であることがわかりました。このことから、人口の維持・増加に向けては、合計特殊出生率の向上による出生数の増加のみではなく、若い世代の人口移動による増加と両面に取り組む必要があることがわかりました。

<人口シミュレーション>



ケーススタディ 3 質的データの解析

～高校生の意識調査における回答同士の関連性を明らかにする～

県では、中南地域の高校生に、地域の職業・企業を知ってもらう取組を長年にわたって続けています。今回は、その取組で実施したアンケート結果を用いて、解答項目の特徴を統計的手法で「見える化」したケースについて紹介します。

(1) 相談内容

管内の高校生に対して地域の職業・企業を知ってもらうため、「企業と高校生の交流会」を実施している。交流会の実施に当たり、参加する高校2年生を対象に事前に別添のアンケートを実施し、約88%(672名)の回答を得ました。その中で、「県内就職・県外就職・県内進学・県外進学」を複数選択可としており、本交流会のファシリテーターをしていただく先生から、生徒が何と何の選択肢で悩んでいるのか見えるようにしたらいいとのアドバイスがあり、その提示方法を示してほしい。

(2) 相談対応

アンケートの質問内容は以下のとおりです。

- ・現在考えている進路について教えてください。[県内就職・県外就職]
- ・就職について考えていることを選んでください。(大学進学後も含む、複数選択可)
[県内にも魅力的な企業がある]など。

これらについて多重対応分析(Correspondence-analysis)を使って、回答項目の特徴を「見える化」しました。この分析方法は、具体的には、選んだ項目を1、選んでない項目を0と置き、その対応関係についてみていくものです。分析に当たっては統計解析ソフト stata³を使用しました。

(3) 分析結果

多重対応分析は、その距離が近ければ同じ生徒の悩み同士の近さを指しています。例えば、以下の6つは同じ回答群とみなします。

回答群1

- A1 「県内にも魅力的な企業がある」
- A2 「地域に根付いた企業で自分を試したい」
- A4 「地域の仕事をもっと知りたい」
- A7 「大学や就職で一度県外に出て、県内に戻ってくるのもアリかも」
- A8 「就職できるか不安」
- A9 「やりたい仕事が県内に無い」

³ Stata とは、データ操作、統計、グラフ化、解析結果の出力など、研究に必要な機能を包括的に提供する統合統計ソフトウェアのことです。

この結果の一つの解釈としては、

- ・ 県内について好意的な見方を持っている(A 1、A 2、A 4)
- ・ しかし、現実的な不安や問題点も感じている(A 8、A 9)
- ・ 将来的なキャリアの一環として県外に出ることも視野に入れているが、最終的には県内に戻ることを考えている(A 7)

また、以下の3つが同じ回答群となります。

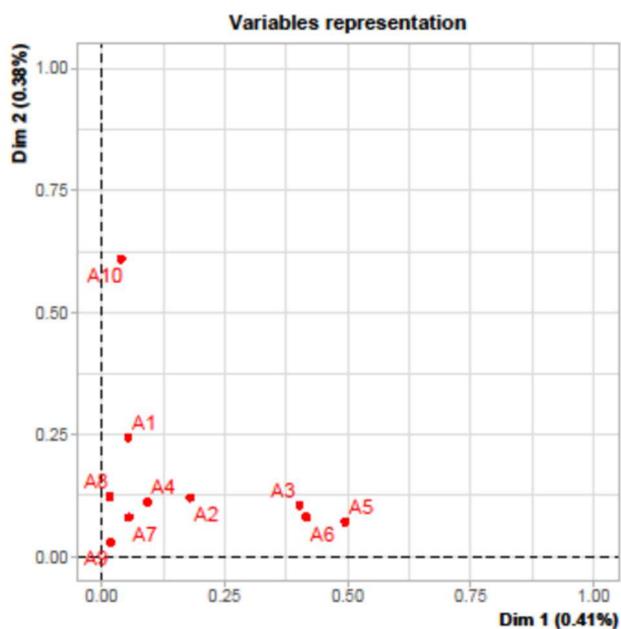
回答群2

- A 3 「将来的に県内に残りたい」
- A 5 「県外に出てみたい」
- A 6 「県外の興味ある職種で働きたい」

この結果の一つの解釈としては、

- ・ 最終的な目標として県内に残ることを望んでいる(A 3)
- ・ しかし、現時点では県外でのキャリアチャンスを積極的に探している(A 5、A 6)。
- ・ これらから、「将来的には県内に定住する意向があるが、今は県外で興味のある職種に就きたい」という前向きな姿勢が伺えます。

<多重対応分析結果>



A1	県内にも魅力的な企業がある
A2	地域に根付いた企業で自分を試したい
A3	将来的に県内に残りたい
A4	地域の仕事をもっと知りたい
A5	県外に出てみたい
A6	県外の興味のある職種で働きたい
A7	大学や就職で一度県外に出て、県内に戻ってくるのもアリかも
A8	就職できるか不安
A9	やりたい仕事がない
A10	まだ何も考えていない

ケーススタディ 4 効果的な調査票設計

～マメコバチの飼養管理等に関するアンケート調査票のブラッシュアップ～

生産量日本一を誇る本県のりんご。その安定的な結実を確保するために欠かせない、受粉を手助けする昆虫「マメコバチ」は、本県のりんご栽培にとって非常に重要な役割を果たしています。

ところが今春、マメコバチの活動量が低下してしまいました。今後の収量減が懸念されたことから、県でその活動量低下の原因を探るための調査を行うこととなり、アンケート調査票の設計をサポートしたケースを紹介します。

(1) 相談内容

今春のマメコバチの活動量が低下した原因を探るため、りんご農家を対象とするアンケート調査票を作成したが、質問項目や選択肢の内容について専門的な視点からアドバイスが欲しい。

(2) 相談対応

今回の相談では、アンケート調査票の設計や集計といった、データ分析の専門家とのオンライン相談会を開催し、質問項目を1つずつ確認しながら、得たいデータをよりの確に収集できるよう、質問や選択肢の文言を丁寧に見直しました。

このアンケートで明らかにしたいのは、マメコバチの飛翔状況の良い農家とそうでない農家で、マメコバチの管理方法等がどのように異なっているか(マメコバチは各農家が巣箱で管理し、開花期に放飼している)という点であったため、その差を客観的に把握することを念頭に置いた質問や選択肢としました。

例えば、マメコバチの飛翔状況に影響する要因については、構造的な要因も影響する可能性があることから、農家の基礎情報など、より幅広く質問した方が良いとのアドバイスを踏まえ、当初の仮説となっていた巣箱の管理方法に加えて、園地周辺の環境や農家の年齢・経営状況など、他にも考えられる要因を洗い出しながら、それらを尋ねるための質問項目を新たに設けました。

また、マメコバチの活動状況を答える選択肢は、当初「良い・平年並・悪い」の3択でしたが、回帰分析等が可能となるよう、活動状況の程度をより細かく分解かつ数値化し、「100～80%、80～60%、60～40%、40～20%、20～0%」の5択に増やしました。

このように、調査目的を明らかにし、考えられる要因を事前に仮説立てたうえで、質問項目を設定する「リサーチデザイン」の考え方にに基づき質問票を作成することで、より有用なデータを得られるアンケート調査とすることが可能となります。

(3) サポートの結果

今回のサポートにより、ブラッシュアップが図られた調査項目を一部抜粋して紹介します。

① 構造的要因を分析するための質問項目の追加

1 マメコバチの巣箱の周辺環境について		回答欄
1	園地の面積を教えてください（単位：a）。複数の園地を管理されている場合は、最も活動が活発だった園地についてお答えください（以下同じ）。	
2	園地の近く（園地の端から50m以内）にりんご以外の樹種はありますか。 ① あり ② なし ③ わからない	
3	園地の近く（園地の端から50m以内）に山林や雑木林はありますか。 ① あり ② なし	

5 回答者（園主）について		回答欄
18	年齢を教えてください。 ① 20～39歳 ② 40～59歳 ③ 60～79歳 ④ 80歳以上	
19	主な経営方を教えてください。 ① 法人経営 ② 家族経営 ③ その他（ ）	
20	外部雇用している労働力があるかお知らせください。 ① ある ② ない	
21	園地の管理に携わっている職員（又は家族）は何人ですか。 ① 1～2人 ② 3～4人 ③ 5～10人 ④ 10人以上	
22	経営面積（りんご）を教えてください。 ① 5ha以上 ② 5～2ha程度 ③ 2～1ha程度 ④ 1ha未満	

② 回帰分析等を念頭に置いた選択肢の細分化・数値化

【改良前】

マメコバチの活動状況について		回答欄
1	令和6年の開花期におけるマメコバチの活動（飛翔）状況をお知らせください。 ① 良い ② 平年並 ③ 悪い	
2	令和5年の開花期におけるマメコバチの活動（飛翔）状況をお知らせください。 ① 良い ② 平年並 ③ 悪い	

【改良後】

2 マメコバチの活動状況について		回答欄
4	令和6年（本年）の開花期におけるマメコバチの活動（飛翔）状況が前年と比べてどの程度であったかお知らせください。 ① 100～80% ② 80～60% ③ 60～40% ④ 40～20% ⑤ 20～0%	
5	令和5年（前年）の開花期におけるマメコバチの活動（飛翔）状況が前々年と比べてどの程度であったかお知らせください。 ① 100～80% ② 80～60% ③ 60～40% ④ 40～20% ⑤ 20～0%	