**問 1 事前に送付した PageRank に関する原著論文について、以下の問いに答えよ。 当該論文が発表された 1998 年当時のインターネット社会では、PageRank という新しい評価指標 は画期的なものであった。一方、現在の複雑化したインターネット社会において、PageRank が依然として通用する部分と通用しない部分について、具体的な PageRank のメカニズムの原理を現在のインターネット技術に対比させながら、800 字以内で論じよ**。

＜自分用メモ＞

* 小論文に必要な要素：背景・導入→問題提起→問題・原因分析→解決策
* 背景：PageRankはGoogle創業者のPageとBrinが開発した、全世界のウェブページのハイパーリンク構造を利用することでその重要度を決定するアルゴリズムである。PageRank 値は、ウェブ上をランダムに遷移するランダムサーファーが無限回の遷移を繰り返した後に、各 Webページへどれだけ遷移したかを示す存在確率と読み替える事ができる。よって、PageRank 値を算出することは、ランダムサーファーの遷移確率に関する行列の最大固有値 １に対応する固有ベクトルを導出することを意味する。PageRankは当時停滞していた検索エンジンのランキング技法のブレイクスルーとなり、現在も十分通用しているGoogle社の検索エンジンの基盤技術である。
* 問題提起：一方で、現在の複雑化したインターネット社会においてPageRankがそのままでは通用できない問題も存在する。特に、検索エンジンによるウェブ利用者への影響力は拡大した結果、サイト側がウェブマーケティングのために検索結果をコントロールするSEO分析・対策が登場した。また、検索結果を誘導する有害サイトの存在は健全なインターネット社会の発展に向けて大きな課題となっている。
* 問題・原因分析：SEO対策に関するサイト側の典型的対応は、キーワードの埋め込みから検索ニーズに合うコンテンツ制作、内部リンク構造の最適化、良い外部リンクの獲得、そしてページスピードや携帯電話での閲覧に関するページエクスペリエンスの最適化がある。これらの対応により、サイト製作者の意向に従う形でPageRankを意図的に向上させることができる。
* 解決策：以上に対して、PageRankに関する改良も考えられている。パーソナル化検索は課題論文でも指摘されたが、ページ外要素であるリンクによるランキングのスパムを信頼できるサイトへの推移確率を意図的に上げることで排除に成功した。今後もGoogle社を中心に検索結果のパーソナル化による恣意性の排除が導入されていくと考えられる。また、先述のSEO対策のいくつかの手段（ページエクスペリエンスの最適化等）はインターネット社会の発展を阻害させない内容とも考えられる。

参考：<https://orsj.org/wp-content/corsj/or57-6/or57_6_308.pdf>

<http://www.kentmiyajima.com/document/pagerank.pdf>

<https://www.willgate.co.jp/promonista/pagerank/>

＜回答＞

PageRankはGoogle創業者のPageとBrinが開発した、全世界のウェブページのハイパーリンク構造を利用することでその重要度を決定するアルゴリズムである。PageRank 値は、ウェブ上をランダムに遷移するランダムサーファーが無限回の遷移を繰り返した後に、各 Webページへどれだけ遷移したかを示す存在確率と読み替える事ができる。よって、PageRank 値を算出することは、ランダムサーファーの遷移確率に関する行列の最大固有値 １に対応する固有ベクトルを導出することを意味する。PageRankは当時停滞していた検索エンジンのランキング技法のブレイクスルーとなり、現在も十分通用しているGoogle社の検索エンジンの基盤技術である。

一方で、現在の複雑化したインターネット社会においてPageRankがそのままでは通用できない問題も存在する。特に、検索エンジンによるウェブ利用者への影響力は拡大した結果、サイト側がウェブマーケティングのために検索結果をコントロールするSEO分析・対策が登場したことで検索結果の公平性を阻害している。また、検索結果を誘導する有害サイトの存在は健全なインターネット社会の発展に向けて大きな課題となっている。

SEO対策に関するサイト側の典型的対応は、キーワードの埋め込みから検索ニーズに合うコンテンツ制作、内部リンク構造の最適化、良い外部リンクの獲得、そしてページスピードや携帯電話での閲覧に関するページエクスペリエンスの最適化がある。これらの対応により、サイト製作者の意向に従う形でPageRankを意図的に向上させることができる。

以上に対して、PageRankに関する改良も考えられている。パーソナル化検索は課題論文でも指摘されたが、ページ外要素であるリンクによるランキングのスパムを信頼できるサイトへの推移確率を意図的に上げることで排除に成功した。Google社を中心に検索結果のパーソナル化による恣意性の排除策も導入されてきたことから、今後も検索結果の民主化に向けた技術的解決は期待されると考えられる。（794字）

メモ：

・PageRank が依然として通用する部分と通用しない部分、はSEO対策に関する検索結果の民主化を採用した（他のネタ：プライバシーの問題等）。全体として現在主流の検索エンジンであるGoogleがPageRankを発展的に改良しながら使用しているため、回答時には「原著論文発表時点の」PageRankと読み替えた（PageRank自体は今も通用している）。

・現在のインターネット技術に対比、はSEO対策に関する説明で回答。よりテクニカルな話題であれば検索エンジンにおける機械学習やテキストマイニングに関する対比となると思われるが、広義解釈してSEO対策もインターネット技術の範疇に入ると考えた。

・最初の背景で原理（論文を読んで準備しておく）を説明しておくことで、論文理解している点を言及。あとは小論文作成の流れにのっとって論述展開した。

・ネット情報を引用しながら回答作成したので、本番では自分の知識で回答できるよう背景までよく理解しておく。

**問 2 次の「Mandated Bacillus Calmette-Guérin (BCG) vaccination predicts flattened curves for the spread ofCOVID-19」と題する論文(M.Bergetal.,ScienceAdvances,Vol.6,No.32,2020)の抄録 (Abstract) を読んで以下の問いに答えよ。**

**Problem 2 Read the following abstract of a paper entitled “Mandated Bacillus Calmette-Guérin (BCG) vaccination predicts flattened curves for the spread of COVID-19” (M. Berg et al., Science Advances, Vol.6, No. 32, 2020), and answer the question.**

**Abstract:**

**Bacillus Calmette-Guérin (BCG) vaccination may reduce the risk of a range of infectious diseases, and if so, it could protect against coronavirus disease 2019 (COVID-19). Here, we compared countries that mandated BCG vaccination until at least 2000 with countries that did not. To minimize any systematic effects of reporting biases, we analyzed the rate of the day-by-day increase in both confirmed cases (134 countries) and deaths (135 countries) in the first 30-day period of country-wise outbreaks. The 30-day window was adjusted to begin at the country-wise onset of the pandemic. Linear mixed models revealed a significant effect of mandated BCG policies on the growth rate of both cases and deaths after controlling for median age, gross domestic product per capita, population density, population size, net migration rate, and various cultural dimensions (e.g., individualism). Our analysis suggests that mandated BCG vaccination can be effective in the fight against COVID-19.**

**この論文では、COVID-19 の感染者数、死者数の増加率と BCG ワクチン強制接種との間に相関があるという結果が示されている。この結果が正しいとしても、BCG ワクチンが COVID-19 の予防に効果があるとは科学的に主張できない理由を述べよ。さらに、その効果を科学的に立証するためにはどんな研究や調査が必要か、実施の可能性も含めて、合計 800 字以内で論じよ。**

＜自分用メモ＞

* 小論文に必要な要素：背景・導入→問題提起→問題・原因分析→解決策
* 背景：2020年初から進展したCOVID-19のパンデミックは世界の経済・人々の生活を一変させた。早くから既存・新規を含めて様々なワクチンが検討されたのは記憶に新しい。本論文では、世界130カ国以上の国別にBCGワクチン(以下、BCG)の強制摂取の有無で分けながら、感染者数増加率(日)、死亡率(30日ドア)から比較している。また、比較前に年齢やGDP・人口密度や人口等の６条件をコントロールしている。その結果、BCG接種がCOVID-19に対して効果ある可能性を示唆していると述べている。
* 問題提起：しかし、本結果を持ってBCGがCOVID-19の予防に効果があるとは言えない。
* 問題：その理由は、第一に母集団抽出の課題である。ワクチン摂取データが揃っている国は相対的に保健衛生環境が高い。また、人種や気候条件、COVID-19以外の疾病や衛生環境に関する因子がコントロールに含まれていない。第二に、ワクチン摂取者と未摂取者の比較できていない点がある。国ごとにワクチン摂取者と未摂取者が存在するが、本論文ではワクチン接種率や摂取者の人口構成は比較出来ていない。第三に、時間の比較ができてない点がある。COVID-19の感染拡大やその対応策の迅速性は各国ごとに異なるため、コントロール要件として必要となる。
* 解決策：上記を踏まえて、今後の科学的検証で必要と考えられるのは、第一に１カ国内でのコントロール条件を揃えたBCG摂取者/被接種者に関するランダム化比較試験を実施すべきである。第二に感染状況を踏まえて時期を変えて複数回以上の試験を実施すべきである。最後に、国ごとの試験結果を十分集めて比較することで、BCG の効果を測定すべきである。

回答：

2020年初から進展したCOVID-19のパンデミックは世界の経済・人々の生活を一変させた。早くから既存・新規ワクチンが検討されたのは記憶に新しい。本論文では、世界130カ国以上の国別にBCGワクチン(以下、BCG)の強制摂取の有無で分けながら、感染者数増加率(日)、死亡率(30日ドア)から比較した。比較前に年齢やGDP・人口密度や人口等の６条件を調整している。その結果、BCG接種がCOVID-19に対して効果ある可能性を示唆している。

しかし、本結果が正しいとしても、その効果が科学的に十分検証できていない。その理由は、第一にサンプル抽出の課題がある。今回の条件同一化は不十分である。ワクチン摂取データが揃っている国は相対的に保健衛生環境が高いと考えられるし、国別の人種・遺伝要因や気候条件、COVID-19以外の疾病や衛生環境に関する因子がコントロールには含まれていない。第二に、ワクチン摂取者と未摂取者の比較ができていない点がある。国ごとにワクチン摂取者と未摂取者が存在するが、本論文ではワクチン接種率や摂取者の人口構成が比較できていない。よって、比較対象が国とBCG摂取者・非摂取者で異なるために検証が不十分である。第三に、時間比較ができてない点がある。COVID-19の感染拡大や対応策の迅速性は各国ごとに異なる。

上記を踏まえて、今後の科学的検証には統計的因果推論により手法改良する必要がある。第一に１カ国内での条件を揃えたBCG摂取者/被接種者のランダム化比較をすべきである。第二に、回帰不連続デザインのためパンデミック前後の有意な差を確認すべきである。COVID-19による外因として生活環境や衛生環境変化の条件を揃える必要がある。第三に、COVID-19以降のコントロール試験は時期を変えながら複数回実施すべきである。最後に、国ごとの試験結果を十分集めることにより、全世界で検証すべきである。以上の対応により、十分な同一条件下でのサンプル抽出と比較検証ができれば効果を検証できる。（787字）

メモ：

* 抄録だけだと本文内容はわからないが、否定的な質問文からその内容を類推して回答。
* データサイエンス観点での処理データの前処理・実験デザインを統計的因果推論の観点から回答させたいと思われるので、用語を使いながら回答した（用語のうろ覚えに寄る定義間違いなどは十分避ける）。
* データ処理の課題や実験デザインについては一般論で回答（本来なら本文読んでない状況なのでこれくらいで良い気がする）。突っ込んだ議論は制限時間からは難しそう。
* 全体的に統計処理に関する背景知識が必要そうで（知識を持っている者が試験時に圧倒的に有利）、2021年からは事前送付論文２報の出題形式に変更したかも。