Analysis of Prostate-Specific Antigen (PSA) Test Prevalence and Demographic Factors.

Introduction

This project presents a comprehensive time-series analysis of the prevalence rates of Prostate-Specific Antigen (PSA) testing across various demographic groups in the United States using data sourced from the Behavioural Risk Factor Surveillance System (BRFSS). Spanning from 2012 to 2020, the dataset encompasses around 12,000 entries with 27 columns, reflecting data points from 50 states.

While existing studies predominantly emphasize relationships between behavioural or demographic factors and prostate cancer testing, this analysis uniquely explores potential seasonal patterns in PSA test prevalence, systematically broken down by demographic segments.

Objectives

- To identify temporal trends in PSA testing prevalence.
- To investigate variations in PSA testing prevalence among demographic groups including age, gender, education level, household income, and race/ethnicity.
- To forecast future prevalence rates based on historical trends.
- To utilize clustering techniques to uncover inherent patterns and categorize prevalence groups effectively.

Analysis Methodology

The analytical methods include:

- **Exploratory Data Analysis (EDA)**:
    - Visualization of prevalence trends over years, segmented by whether PSA tests were administered within the past two years.
    - Detailed breakdowns across demographic segments such as gender, race/ethnicity, household income, education level, and age groups.
- **Confidence Interval Analysis**:
    - Investigating relationships between prevalence rates and associated confidence intervals (CI widths) across demographic groups.
- **Forecasting**:
    - Employing time-series forecasting models to project future trends in PSA testing prevalence.

- **Clustering Analysis**:
    - Application of K-means clustering technique, determining the optimal number of clusters via the elbow and silhouette methods, to classify states or regions by similarity in PSA testing prevalence patterns.

Key Findings and Insights

1. PSA Testing Prevalence Trends (2012-2020):

- **Overall Trends**:
    - A clear upward trend (from ~55% in 2012 to ~72% in 2020) in the percentage of people **not having** a PSA test within the last two years.
    - A complementary downward trend (from ~45% in 2012 to ~28% in 2020) in the percentage of people **having had** a PSA test within the last two years.
    - These inverse trends may suggest shifting public health practices, recommendations, or perceptions toward PSA screening.
- **Demographic Trends**:
    - Higher prevalence rates of not having PSA tests observed across all demographic groups over the years, reflecting the overall declining testing rates.
    - A consistent disparity exists across demographic categories, with variations prominently seen in race/ethnicity, education levels, and income brackets.

2. Detailed Demographic Breakdowns:

- **Gender**:
    - PSA testing specifically relevant to males, showing clear consistent upward trends in rates of non-testing (~55% in 2012 to ~72% in 2020).
- **Race/Ethnicity**:
    - Variations were notable; certain groups (e.g., Hispanic and Multiracial populations) showed steeper increases in non-testing prevalence.
    - Clear disparities emerged across racial groups, reflecting potential differences in healthcare accessibility and education regarding prostate cancer screening.
- **Household Income**:
    - Individuals from lower-income households exhibited higher prevalence rates of non-testing, highlighting possible economic barriers to regular PSA screening.
- **Education Attainment**:
    - Higher educational attainment correlated with lower non-testing prevalence, suggesting education-driven awareness and better access to healthcare.
- **Age Groups**:

    - Older age groups (65+) were consistently less likely to skip testing compared to younger age groups, likely due to greater awareness or higher perceived risk.

3. Prevalence vs. Confidence Interval Width:

- Scatter plots showed the relationship between the prevalence rates and their confidence intervals (CI widths).
- Wider CI values frequently accompanied lower prevalence rates, reflecting less data certainty in groups with lower PSA test participation.
- This correlation indicates higher variability and uncertainty in demographic segments that are less likely to engage in PSA screening practices.

4. Forecasting PSA Test Prevalence:

- The predictive analysis forecasts continued increasing prevalence of non-testing beyond 2020, indicating a potential decline in PSA testing practices.
- Conversely, declining trends are forecasted for PSA testing participation, further emphasizing shifts in health behaviours and screening recommendations or practices.

5. Clustering Analysis:

- **Elbow Method**: Identified optimal cluster count between 3 and 4 for both groups (tested and not tested within the last two years).
- **Silhouette Score**: Confirmed optimal cluster count at approximately k=3 for PSA non-testing prevalence and k=2 or 3 for PSA testing prevalence.
- These clusters reveal distinct segments based on similarities in demographic prevalence trends and could support targeted public health interventions.

Conclusions and Recommendations

The analysis demonstrates significant temporal and demographic differences in PSA testing prevalence. The identified trends and demographic variations underscore the need for targeted public health strategies to address gaps in prostate cancer awareness and healthcare accessibility.

Recommendations:

- **Enhanced Educational Programs**: Tailored outreach for demographic groups demonstrating higher non-testing prevalence (e.g., lower-income, less educated populations).
- **Targeted Screening Initiatives**: Implement focused screening programs for demographics showing sharp declines in testing prevalence.
- **Policy Adjustments**: Public health policies should consider demographic-specific factors revealed by clustering analysis to optimize resource allocation.

Limitations

- The dataset reflects self-reported data with inherent recall bias and accuracy limitations.

- Temporal analyses constrained by annual data collection may obscure finer seasonal patterns.
- Unmeasured external factors, including healthcare policy changes and public health recommendations, may influence observed trends.

Future Work

- Conduct finer-grained temporal analysis (monthly/quarterly) to validate seasonal patterns explicitly.
- Extend clustering analyses with additional demographic and behavioural variables to enhance segmentation precision.
- Explore qualitative research methods to complement quantitative findings, uncovering barriers and facilitators influencing PSA testing decisions.

References

[1] Li, J., Hall, I.J. & Zhao, G. **Prostate cancer screening decision-making in three states: 2013 behavioral risk factor surveillance system analysis.** *Cancer Causes Control* **28**, 235–240 (2017). https://doi.org/10.1007/s10552-017-0860-8

[2] Li J, Thompson TD, Richards TB, Steele CB. **Racial and Ethnic Differences in Health Behaviors and Preventive Health Services Among Prostate Cancer Survivors in the United States**. Prev Chronic Dis. 2016 Jul 21;13:E95. doi: 10.5888/pcd13.160148. PMID: 27442995; PMCID: PMC4956476.

[3] Garg, V., Raisch, D., Selig, J. *et al.* **Health disparities in clinical practice patterns for prostate cancer screening by geographic regions in the United States: a multilevel modeling analysis**. *Prostate Cancer Prostatic Dis* **16**, 193–203 (2013). https://doi.org/10.1038/pcan.2013.3

[4] Chang-Rong Chen, Logan Briggs, Mara Koelker, Benjamin V. Stone, Khalid Alkhatib, Muhieddine Labban, Alberto Briganti, Francesco Montorsi, Giorgio Gandaglia, Quoc-Dien Trinh, **The association between behavioral habits and physical health status in prostate cancer patients: a large US national health-related survey**, Prostate International, Volume 12, Issue 4, 2024, Pages 207-212, ISSN 2287-8882, https://doi.org/10.1016/j.prnil.2024.08.001.

[5] K.R. Fontaine, M. Heo, D.B. Allison, **Obesity and prostate cancer screening in the USA**, Public Health, Volume 119, Issue 8, 2005, Pages 694-698, ISSN 0033-3506, https://doi.org/10.1016/j.puhe.2004.09.002.