STA440 Final Project

Matthew O'Donnell

2024-12-01

1. Introduction

2. Methodology

2.1 Gender-Specific Name Usage Over Time

For each character in the dataset, a gender score is provided, which is computed as follows:

$$g_i = \frac{N_{\text{male}} - N_{\text{female}}}{N_{\text{male}} + N_{\text{female}}}$$

This gender score for a character g_i can range from -1 (completely feminine) to 0 (gender-neutral) to 1 (completely masculine). The data contains a parts-per-million metric for each character during each of six birthing cohorts/generations. A average of gender specificity of a cohort (absolute value of gender score), weighted by parts-per-million in that cohort, could be used to see changes in gender-specificity over time.

Changes in Gender–Specific Name Usage Over Time
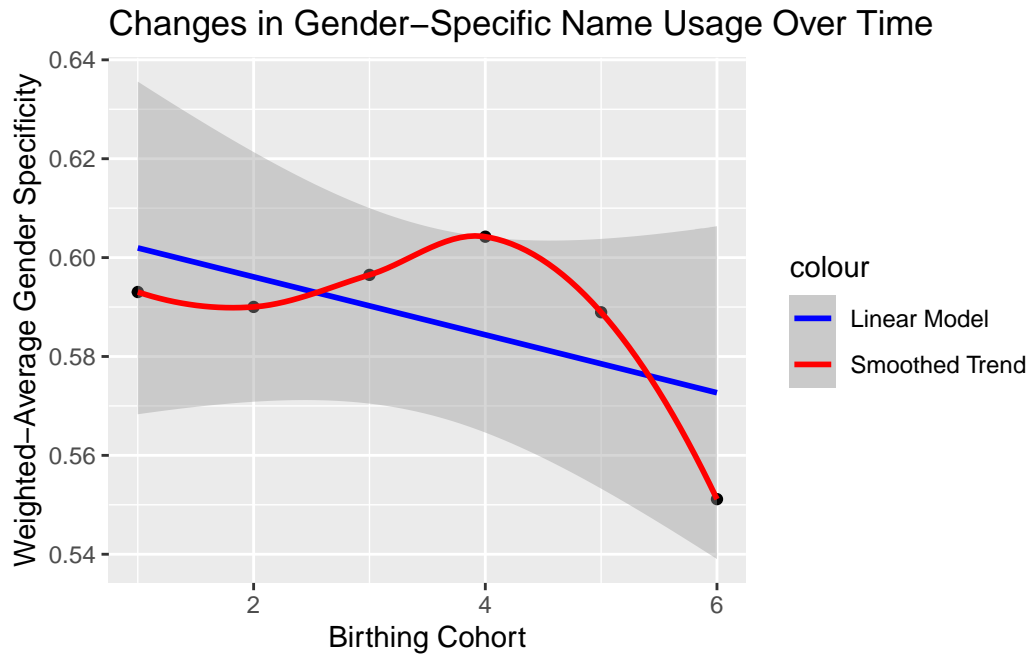
2.2 Subjective Name Characteristics By Gender

3. Results

3.1 Anova Results

Coefficients Table

Predictor	Slope (Estimate)	p-value
Intercept	0.593	< 0.001
Time Period : 1960-1969	-0.003	0.683
Time Period :1970-1979	0.003	0.642
Time Period : 1980-1989	0.011	0.144
Time Period : 1990-1999	-0.004	0.597
Time Period : 2000-2008	-0.042	< 0.001

ANOVA Results

Independent Variable	F-statistic	P-value
Time Period	10.792	< 0.001

3.2 WLS Output

Valence

Decade	Estimated Coefficient $(\hat{\beta}_1)$	P-value
1930–1959	-0.053	< 0.001
1960–1969	-0.026	0.0436
1970–1979	-0.009	0.494
1980–1989	0.001	0.907
1990–1999	0.015	0.251
2000–2008	0.027	0.0394

Warmth

Decade	Estimated Coefficient $(\hat{\beta}_1)$	P-value
1930–1959	-0.054	< 0.001
1960–1969	-0.038	0.0007
1970–1979	-0.032	0.0035
1980–1989	-0.026	0.0132
1990–1999	-0.022	0.0359
2000–2008	-0.037	0.0010

Competence

Decade	Estimated Coefficient $(\hat{\beta}_1)$	P-value
1930–1959	0.088	< 0.001
1960–1969	0.134	< 0.001
1970–1979	0.172	< 0.001
1980–1989	0.202	< 0.001
1990–1999	0.204	< 0.001
2000–2008	0.206	< 0.001

4. Discussion