
Left-over Men and Women in Post-One-Child-Policy China

Name Yilun Dai

Faculty Advisor: Kazuo Yamaguchi

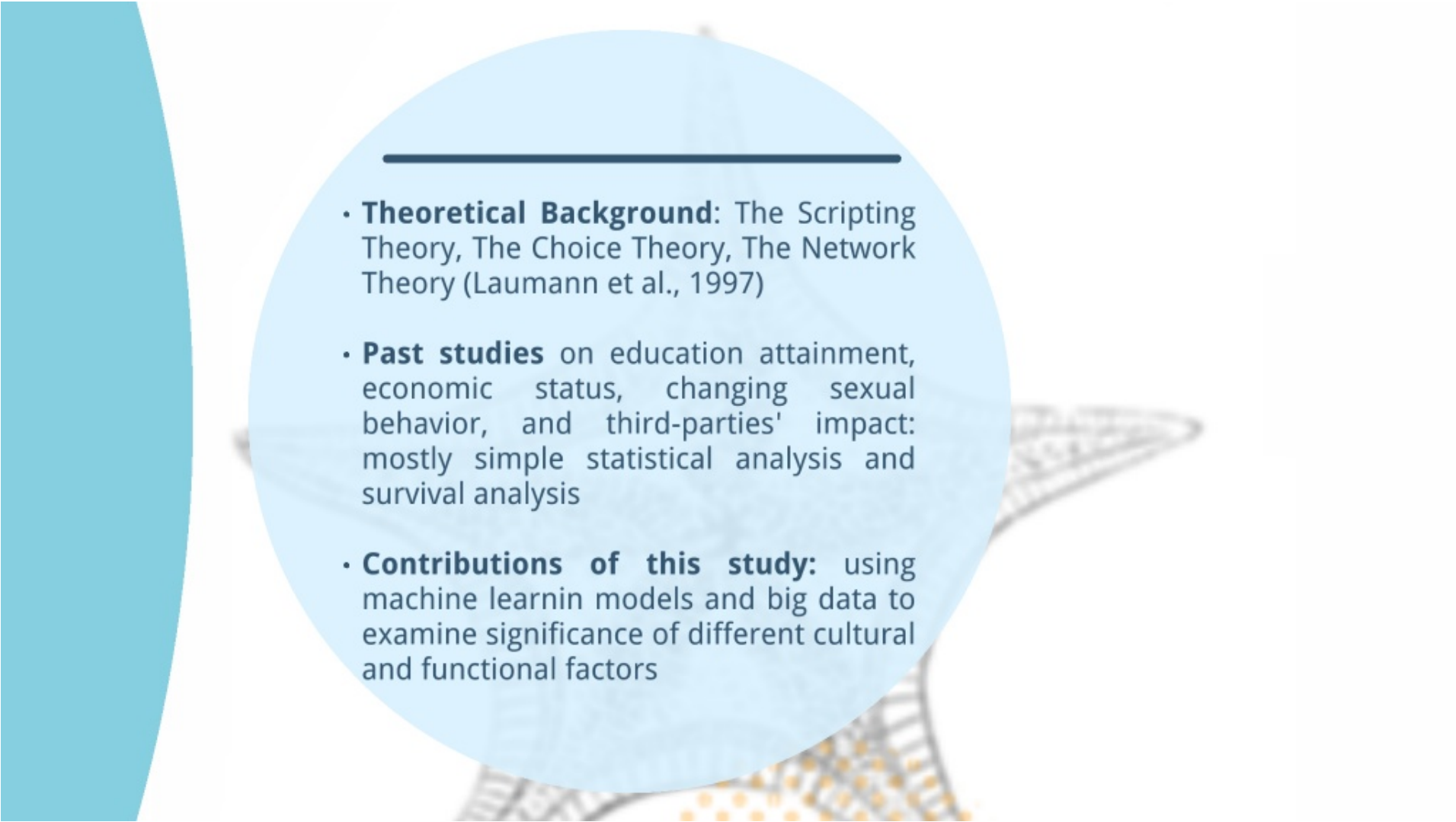


Introduction

This study will explore functional and cultural factors that cause Chinese people to postpone their age of first marriage after the change in birth control policy.



Literatures & Contributions

- 
- **Theoretical Background:** The Scripting Theory, The Choice Theory, The Network Theory (Laumann et al., 1997)
 - **Past studies** on education attainment, economic status, changing sexual behavior, and third-parties' impact: mostly simple statistical analysis and survival analysis
 - **Contributions of this study:** using machine learning models and big data to examine significance of different cultural and functional factors



Data

Data:

- Adults' personal, educational, economic and marital data (Chinese Family Panel Studies)
- Housing Price (National Bureau of Statistics of China)
- Zhihu

Variables:

- First marriage entrance age (Y)
- Years of education, way of knowing the partner, income level, weekly working time, housing price, premarital cohabitation, time spent on the internet (X)

Methods

Methods

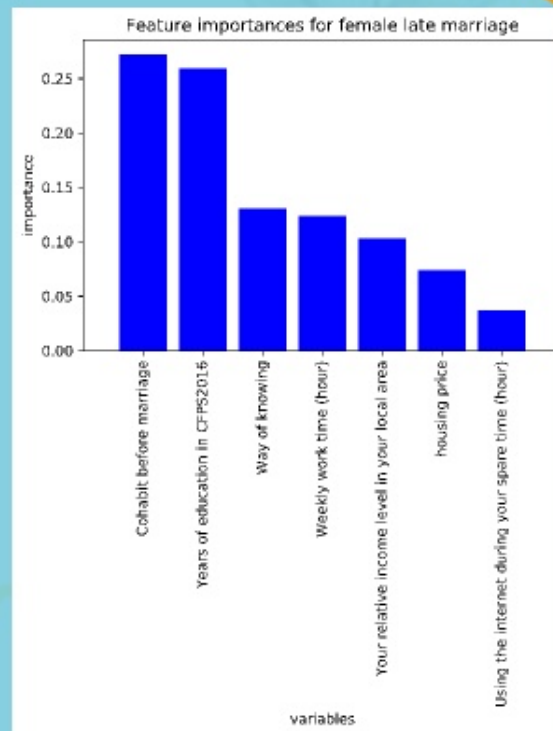
Machine Learning:

- Feature importance
- Regressor vs Classifier based on prediction power
- Random Forest vs SVM based on prediction power

Text Analysis:

- Crawling top answers to 3 Zhihu questions
- Jieba for extracting Chinese words
- Word Cloud for High-frequency words

Feature Importance



Model
Comparison

Random Forest vs SVC

Table 1: Performance comparison between Random Forest Classifier and SVC

	Female Late Marriage		Male Late Marriage	
	Random Forest Classifier	SVC	Random Forest Classifier	SVC
Accuracy	0.6669	0.4972	0.6616	0.5872
Informedness	0.3798	-0.0055	0.3231	0.1744
Sensitivity	0.4138	0.0344	0.5517	0.5172
Specificity	0.92	0.96	0.7714	0.6571

- Classifier > Regressor
- Random Forest > SVM

Zhihu Questions

1. "Are there a great many young people who no longer want to get married? Why?"
2. "What are left-over women insisting on?"
3. "What are left-over men insisting on?"

Patterns of answers to Question1, min 100 likes



Gender-Specific

Women vs Men

Patterns of answers to Question2



Patterns of answers to Question3



Conclusion

- Better accuracy if treated as a classification problem; Random Forest performs better than SVM
- Pre-marital Cohabitation is an important predictor
- Child-rearing is a big concern
- Gender differences in cultural scripting



Thank you!