Analysing Factors Contributing to Type II Diabetes Risk Status

Sean Mulherin and Ramiro Lobo

STATS 411 Multivariate Statistical Analysis
Dr. Maria Cha
Winter 2025

UCLA College | Physical Sciences Statistics & Data Science

Introduction

Background

- Diabetes is a chronic disease that describes the body's inability to properly regulate insulin levels
- Our bodies use insulin to break down glucose obtained from the foods/drinks we consume
- There are two main types
  - Type I: not preventable - typically inherited genetically or triggered via viral infection
  - Type II: preventable - typically a result of lifestyle habits

Introduction

Motivation

- Abnormal glucose levels can lead to many health issues, including:
  - Heart Disease
  - Kidney Disease
  - Cataracts
  - Nerve Damage
- As of 2021, 38 million people have diabetes - 11.6% of population (American Diabetes Association)
- Diabetes is the 8th leading cause of death in the U.S.
- Total costs span $412.9 billion (Emily et al., 2022)

Introduction

Objective

- The goal is to help prevent the onset of type II diabetes.

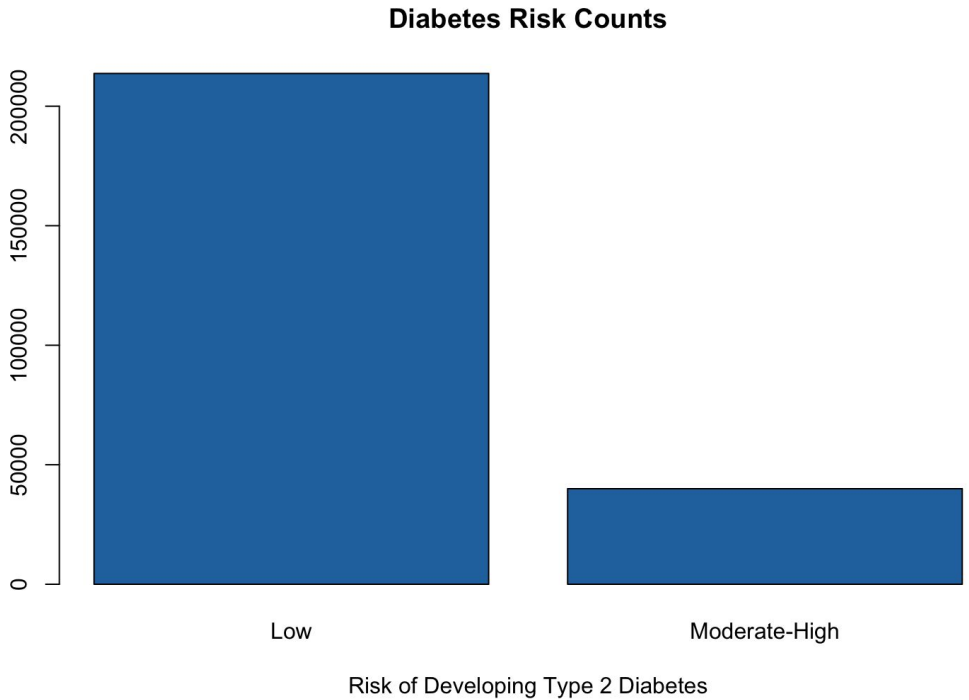
- To accomplish this, we leverage rigorous statistical analysis to identify factors that affect the development of type II diabetes, including:
  - Logistic Regression Analysis
  - Principal Component Analysis
  - Factor Analysis

Data

- Every year, the Center for Disease Control (CDC) conducts a behavioral survey of Americans
  - Behavioral Risk Factor Support Survey (BRFSS) collects health-related information via telephone surveys of over 400,000 Americans from each state
- Predictors
  - 21 variables of class numeric, ordinal, and categorical
  - **Biological factors**: blood pressure, cholesterol, BMI
  - **Lifestyle factors**: age, sex, smoker, diet, exercise, alcohol consumption
  - **Social factors**: income, education, mental health
- Response
  - Binary: low or medium-high risk of developing type II diabetes
- 253,680 total observations

Data

EDA


Diabetes Risk Counts
0
50000
100000
150000
200000
Low
Moderate-High
Risk of Developing Type 2 Diabetes

Data

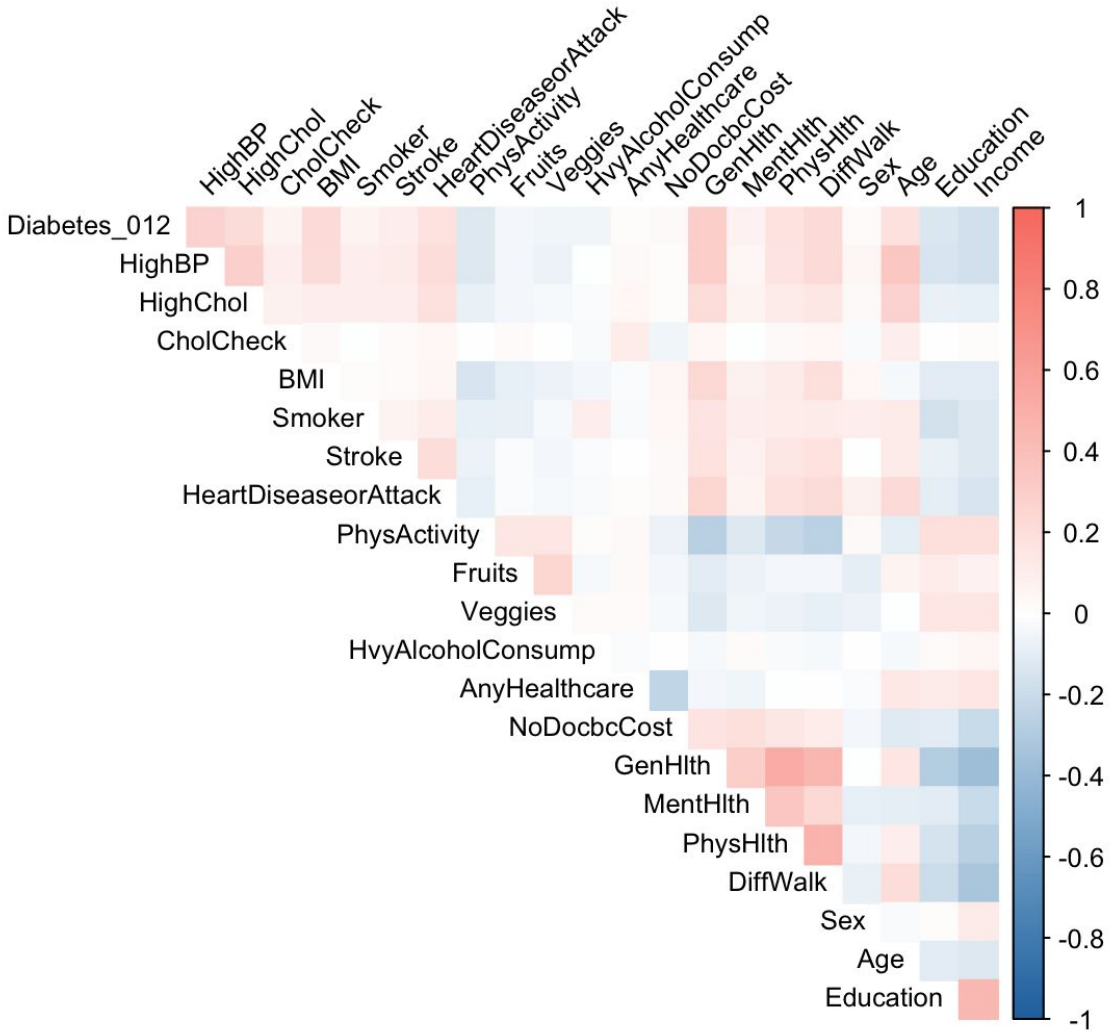
EDA

Highest Positive Correlations with Diabetes Risk Status (> 0.2)

- GenHlth
- HighBP
- BMI
- DiffWalk
- HighChol

Highest Negative Correlations with Diabetes Risk Status (< -0.12)

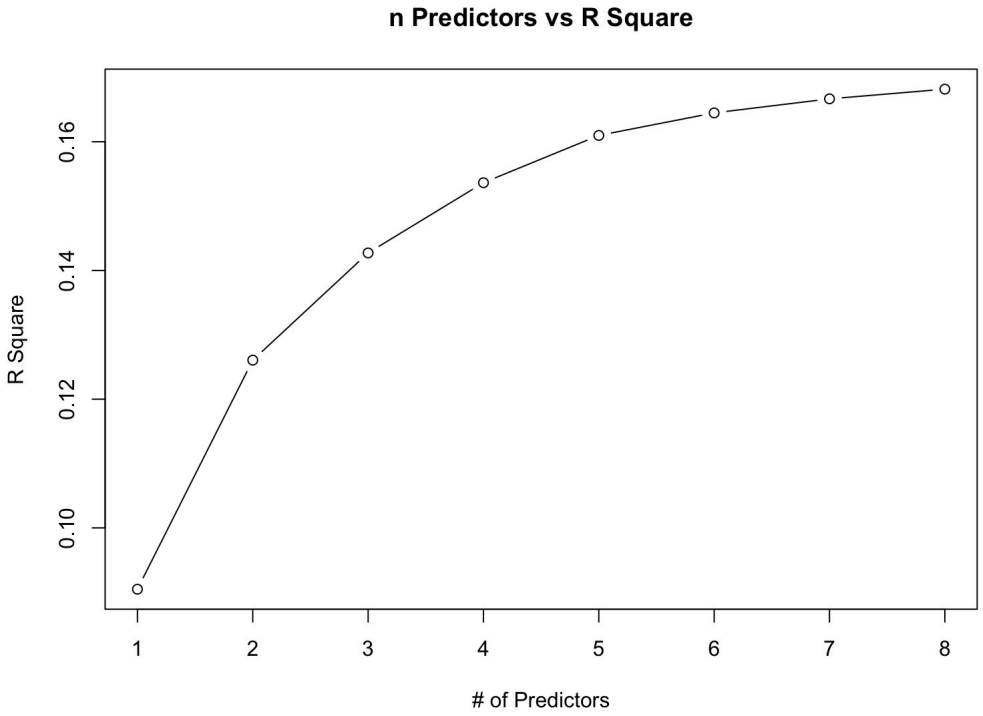
- Income
- Education
- PhysActivity



Methodology - Regression

Notable Factor Coefficients

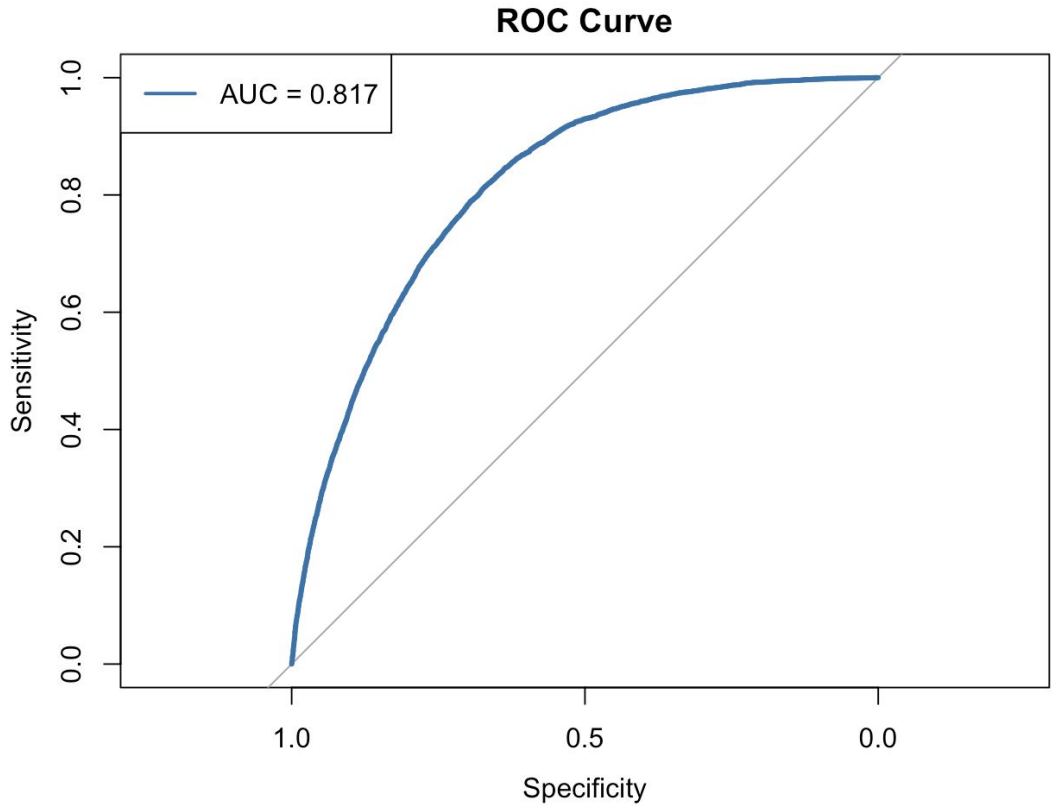
- CholCheck = 1.21
- HighBP = 0.71
- HvyAlcoholConsump = 0.66
- HighChol = 0.60
- GenHlth = 0.51
- Sex = 0.24



Methodology - Regression Cont.

Utilized the Youden Index to find the optimal cut-off threshold. (Ruopp et. al, 2008)

Optimal Threshold = 0.133



Methodology - Regression Cont.

We want to confidently capture medium-high risk patients, so we prioritize Recall

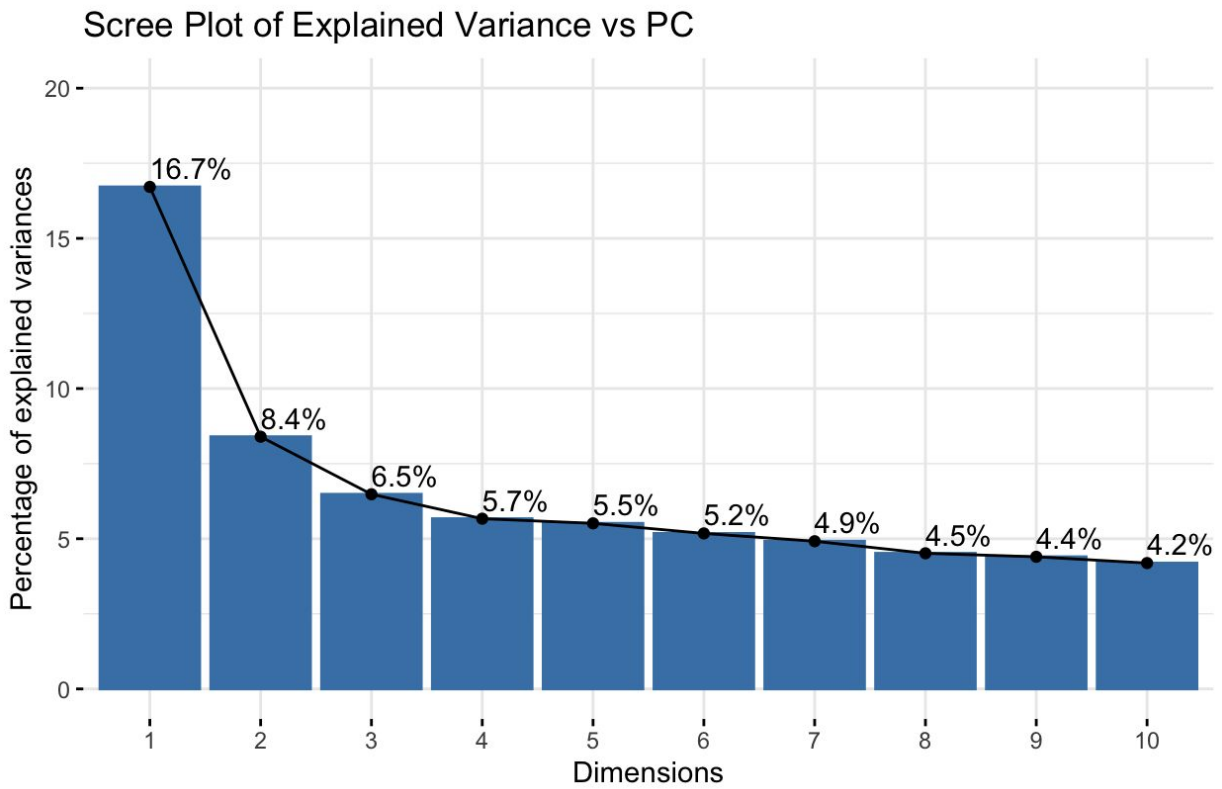
- Accuracy = 69.5%
- Precision = 31.8%
- Recall = 81.2%
- F1 Score = 45.7%

		Observed	
		Low Risk	**Medium-High Risk**
Predicted	**Low Risk**	28745	1508
	Medium-High Risk	13969	6514

Methodology - PCA

Top 10 PC's explain 66% of variance

Top 14 PC's explain 80% of variance

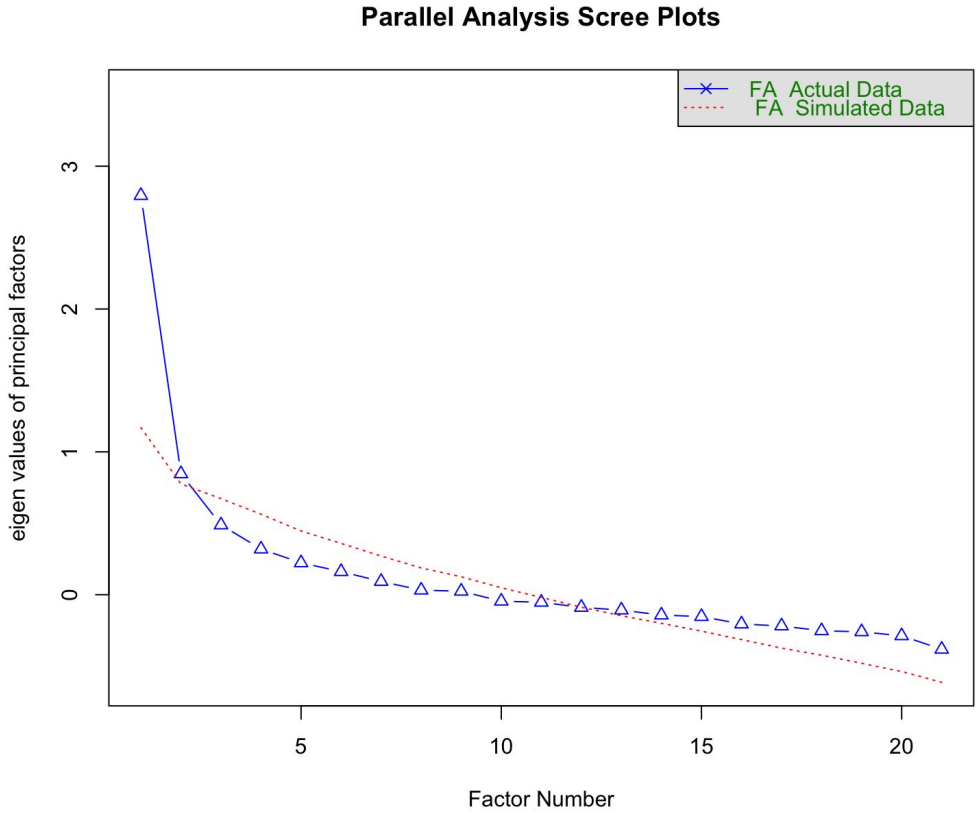


Methodology - Factor Analysis

Two factors are identified as the ideal amount to explain the variance.

Factor 1 is likely to represent the general health of a person.

Factor 2 is likely to represent biological factors, predominantly cholesterol health.



Conclusion

- Biological Factors:
  - High Cholesterol has the strongest adverse effect
  - High Blood Pressure has the second strongest adverse effect
  - Males are more likely to develop type II diabetes than females
  - As Age increases, the risk increases

- Lifestyle Factors:
  - Heavy Alcohol Consumption has the strongest adverse effect
    - Males > 14 drinks/week
    - Females > 7 drinks/week
  - Marginal negative effects:
    - Income, Education, Physical Activity, Diet, and Smoking

Conclusion

American Diabetes Association. (Nov, 2023). *Diabetes facts and statistics*. Diabetes.org. Retrieved March 1, 2025, from https://diabetes.org/about-diabetes/statistics/about-diabetes

Emily D. Parker, Janice Lin, Troy Mahoney, Nwanneamaka Ume, Grace Yang, Robert A. Gabbay, Nuha A. ElSayed, Raveendhara R. Bannuru; Economic Costs of Diabetes in the U.S. in 2022. Diabetes Care 2 January 2024; 47 (1): 26–43. https://doi.org/10.2337/dci23-0085

Centers for Disease Control and Prevention. (2025, February). 2023 BRFSS survey data and documentation. https://www.cdc.gov/brfss/annual_data/annual_2023.html

Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. Biom J. 2008 Jun;50(3):419-30. doi: 10.1002/bimj.200710415. PMID: 18435502; PMCID: PMC2515362.