40.016: The Analytics Edge

Week 3 Lecture 2

Health analytics

The Framingham heart study

Term 5, 2022

Franklin D. Roosevelt

- President of the United States: 1933 - 1945
- Longest serving president
- Lead country through the great depression
- Commander-in-Chief during World War II
- Died while president, April 12, 1945

FDR's blood pressure and the origin of FHS

- Before presidency (1932): BP 140/100
  - Healthy BP is around than 120/80
  - Today 140/100 is considered high BP
- 1935-1941: gradual rise from 136/78 to 188/105
  - Today this is called Hypersensitive crisis, requires immediate emergency care
  - FDR's physician said: "A moderate degree of arteriosclerosis, although no more than normal for a man of his age"
- Two months before demise: BP 260/150
- Day of death: BP 300/190

- Today, our belief is that normal BP (systolic/diastolic) is in the range of 90/60 to 120/80.

Harry Truman who was vice president under FDR, became president and signed into law the National Heart Act. This law allocated a US $500000 seed grant for a 20 year epidemiological study of heart disease.

Study of Coronary Disease

- High blood pressure was thought of as essential hypertension.
- After World War II ended, it was observed a third of deaths in the US was of heart disease, up four times from 1900.
- Plan: track a large cohort of initially healthy patients over time.
- City of Framingham, MA selected as site for study
  - Appropriate size
  - Stable population
  - Cooperative doctors and residents
- 1948: beginning of Framingham Heart Study

Framingham Heart Study

- 1948: 5209 men and women aged 30-62 enrolled. The first round of extensive physical examination and lifestyle interviews were conducted with participants returning every two years for examinations and tests.
- 1971: The second generation of 5124 original participant's children and spouses were enrolled into the program.
- 1994: An Omni cohort consisting of 507 men and women of African-American, Hispanic and Asian origins (residents of Framingham) were enrolled to reflect the increasing diversity of residents in the area.
- 2002: Third generation of participants were enrolled.
- 2003: A second group of Omni cohort was enrolled.

Most of the knowledge concerning heart diseases such as the effect of diet, exercise and blood pressure that we know today is based on this longitudinal study.

Coronary Heart Disease

- We predict 10-year risk of CHD influenced by:
  Prediction of coronary heart disease using risk factor categories (1998)
  by Wilson, Agostino, Levy, Belanger, Silbershatz, Kannel.
  This paper has been cited over 11000 times as of September 2021.
- It introduced the Framingham Risk Score.
- CHD is a disease of the blood vessels supplying the heart.
- Heart disease has been the leading cause of death worldwide since 1921.
- 7.3 million people died from CHD in 2008.

	TOTAL DEATHS BY BROAD CAUSES							
								Per Cent
Year	Tuberculosis	Other Communicable Diseases	Neoplasms	Cardiovascular Diseases	External Causes of deaths	Disease of Early Infancy	Other Causes of Deaths	Total
1950	12.0	32.5	2.8	6.3	4.0	7.2	35.3	100.0
1955	9.1	21.7	6.5	8.5	4.9	11.1	38.3	100.0
1960	6.3	18.7	10.4	10.6	5.0	11.2	37.8	100.0
1965	6.2	12.6	13.9	13.7	6.5	9.9	37.2	100.0
1970	4.2	12.7	15.1	27.0	7.9	5.9	27.1	100.0
1975	3.7	11.7	18.5	29.4	4.1	3.5	29.0	100.0
1980	1.8	11.4	21.0	34.4	7.2	3.3	20.9	100.0
1985	1.3	11.6	22.0	34.8	8.1	2.5	19.7	100.0
1990	0.8	10.3	23.9	37.1	7.3	2.2	18.4	100.0
1995	0.8	14.3	25.2	35.7	7.1	1.3	15.6	100.0
2000	0.6	13.9	27.0	36.6	7.2	0.8	13.8	100.0
2005	0.4	17.9	26.5	33.3	6.3	0.7	15.1	100.0
2006	0.4	15.3	28.8	33.2	6.3	0.7	15.4	100.0
2007	0.5	16.1	27.7	34.0	6.0	0.5	15.2	100.0
2008	0.5	15.3	29.3	33.6	5.8	0.6	14.9	100.0
2009	0.4	16.8	29.3	32.8	5.7	0.6	14.3	100.0
2010	0.4	17.2	28.8	33.0	5.5	0.5	14.5	100.0
2011	0.4	17.7	30.4	31.7	4.6	0.5	14.6	100.0
2012	0.4	18.0	30.6	31.1	5.6	0.5	13.9	100.0
2013	0.3	19.7	30.9	30.4	4.9	0.5	13.3	100.0
2014	0.3	20.0	29.9	30.9	4.7	0.5	13.8	100.0
2015	0.2	20.4	30.1	30.7	4.5	0.5	13.6	100.0
2016	0.2	20.4	29.9	30.5	4.4	0.5	14.0	100.0
2017	0.2	21.3	29.8	31.3	4.0	0.4	13.0	100.0
2018	0.1	21.8	29.1	30.2	4.3	0.5	14.1	100.0

Death by broad causes in Singapore. *Courtesy: MOH*

Coronary Heart Disease: data and analysis

- The paper develops clinical decision rules to predict 10 year CHD using **risk factors**: Blood Pressure, total cholesterol, LDL cholesterol in a white middle aged population.
- The coinage “risk factor” was done by William Kannell and Roy Dawber of FHS.
- The data set:
  - Source: `http://biolincc.nhlbi.nih.gov/home/`
  - Anonymized data
  - 4434 participants
  - Three examination periods (6 years apart) from 1956-1968.
  - Participants followed for 24 years.

Coronary Heart Disease: risk factors

- **Demographic risk factors:** *gender, age, education*.
- **Behavioral risk factors:** *Current smoker, Number of cigarettes per day*.
- **Medical history risk factors:** *BP meds, stroke history, hypertension, diabetes*.
- **Risk factors from first examination:** *Total cholesterol, sys/dia BP, BMI, heart rate, glucose level*

How we analyze

- Randomly split data into *training* and *test* sets.
- Create a logistic regression model on the training set to predict whether a patient experiences a CHD within the next ten year.
- Validate on the test set.

Findings

- Model never predicts 10-year CHD risk above 50%.
- AUC = 0.757 (differentiates between low-risk and high-risk cases).
- Significant variables: smoking, cholesterol, blood pressure, glucose.

- Population cohort mostly white middle class.

Intervention strategies and impacts

- The results from the FHS has been validated with external validation methods by generalizing to different populations such as African-Americans, Asians.
- The advantage of such an approach is that it can be used to develop intervention strategies for example drugs to lower blood pressure, drugs for lower cholesterol. The effects of these on reduced chances of coronary heart diseases can be tested by doing clinical trials.
- The market for diuretics (to reduce blood pressure) and statins (to reduce cholesterol levels) are now in billions of dollars.
- The FHS has also led to an increase in clinical decision rules in many areas of healthcare that predict clinical outcomes using patient data and results. These models are unbiased, unemotional and can assist new physicians with little experience to make decisions.

Point system

- While this approach predicts the risk of patients getting a heart disease, it is not particularly easy for patients and physicians to uses.
- A points system is often implemented to make these results more interpretable and usable in practice.
- We can develop such a points system from the logistic regression results as follows.
  - Assign the lowest risk to a person who is in the low age category, is female, has a systolic blood pressure of lower than 120 (from clinically meaningful states), smokes zero cigarettes per day and has a smaller value for the glucose level.
  - Based on this by appropriately normalizing and scaling the results, we can generate a points system that is particularly useful in telling patients on how improving on a certain aspect can help decrease the risk of CHD.

Points table

Variables	Category	Reference	Base difference	Logit units	Points
Age	30-39	34.5	0	0	0
	40-49	44.5	10	0.59	2
	50-59	54.5	20	1.18	4
	60-69	64.5	30	1.77	6
	70-79	74.5	40	2.36	8
Sex	Male	1	-1	0.54	2
	Female	2	0	0	0
Systolic blood pressure	< 120	104.5	-25	-0.4	-1
	120-139	129.5	0	0	0
	140-159	149.5	20	0.32	1
	≥ 160	174.5	45	0.72	2

Points system: divide by 0.059×5 and round off to nearest integer. Negative points (bonus) allowed.

Age	Points
20 – 34	- 9
35 – 39	- 4
40 – 44	0
45 – 49	3
50 – 54	6
55 – 59	8
60 – 64	10
65 – 69	11
70 – 74	12
75 – 79	13

Smoker	Points				
	Age 20 – 39	Age 40 – 49	Age 50 – 59	Age 60 – 69	Age 70 – 79
No	0	0	0	0	0
Yes	8	5	3	1	0

Total cholesterol mmol/L (mg/dL)	Points				
	Age 20 – 39	Age 40 – 49	Age 50 – 59	Age 60 – 69	Age 70 – 79
< 4.1 (160)	0	0	0	0	0
4.1 - 5.1 (160 - 199)	4	3	2	1	0
5.2 - 6.1 (200 - 239)	7	5	3	1	0
6.2 - 7.2 (240 - 279)	9	6	4	2	1
≥ 7.3 (280)	11	8	5	3	1

HDL cholesterol mmol/L (mg/dL)	Points
≥ 1.6 (60)	-1
1.3 - 1.5 (50 - 59)	0
1.0 - 1.2 (40 - 49)	1
< 1.0 (40)	2

Systolic BP (mmHg)	Points	
	If untreated	If treated
< 120	0	0
120 - 129	0	1
130 - 139	1	2
140 - 159	1	2
≥ 160	2	3

Allocate points based on person's age, total and HDL cholesterol levels, smoking status and systolic blood pressure as indicated in the tables to the left.

Check the total points against Table 1 to estimate a person's ten-year CHD risk.

For example, if you are a 45-year-old Chinese male who smokes every day with a total cholesterol of 7.5 mmol/L, a HDL cholesterol of 1.1 mmol/L and a systolic BP of 135 mmHg, then your total score is >20. You are estimated to have a 'high' risk of heart attack or coronary death within the next ten years.

This would mean that more than 20 out of 100 persons in your risk category would experience a heart attack or coronary death within the next ten years.

Table 1. Estimation of ten-year CHD risk for men in Singapore

Total points	Ten-Year Risk (%)		
	Chinese	Malay	Indian
-1	< 1	< 1	1
0	< 1	< 1	1
1	< 1	1	1
2	1	1	1
3	1	1	2
4	1	1	2
5	1	1	3
6	1	2	3
7	2	2	4
8	2	3	5
9	3	4	7
10	4	5	9
11	5	6	11
12	6	8	14
13	8	11	18
14	11	13	> 20
15	13	17	> 20
16	17	> 20	> 20
≥17	> 20	> 20	> 20

Age	Points
20 - 34	- 7
35 - 39	- 3
40 - 44	0
45 - 49	3
50 - 54	6
55 - 59	8
60 - 64	10
65 - 69	12
70 - 74	14
75 - 79	16

Smoker	Points				
	Age 20 – 39	Age 40 – 49	Age 50 – 59	Age 60 – 69	Age 70 – 79
No	0	0	0	0	0
Yes	9	7	4	2	1

Total cholesterol mmol/L (mg/dL)	Points				
	Age 20 – 39	Age 40 – 49	Age 50 – 59	Age 60 – 69	Age 70 – 79
< 4.1 (160)	0	0	0	0	0
4.1 - 5.1 (160 - 199)	4	3	2	1	1
5.2 - 6.1 (200 - 239)	8	6	4	2	1
6.2 - 7.2 (240 - 279)	11	8	5	3	2
≥ 7.3 (280)	13	10	6	4	2

HDL cholesterol mmol/L (mg/dL)	Points
≥ 1.6 (60)	-1
1.3 - 1.5 (50 - 59)	0
1.0 - 1.2 (40 - 49)	1
< 1.0 (40)	2

Systolic BP (mmHg)	Points	
	If untreated	If treated
< 120	0	0
120 - 129	1	3
130 - 139	2	4
140 - 159	3	5
≥ 160	4	6

Allocate points based on person's age, total and HDL cholesterol levels, smoking status and systolic blood pressure as indicated in the tables to the left.

Check the total points against Table 2 to estimate a person's ten-year CHD risk.

For example, if you are a 40-year-old Chinese non-smoker female with a total cholesterol of < 4.1mmol/L, a HDL cholesterol of 1.3 mmol/L and a systolic BP of <120 mmHg, then your total score is 0. You are estimated to have a 'low' risk of heart attack or coronary death within the next ten years.

This would mean that less than one out of 100 persons in your risk category would experience a heart attack or coronary death within the next ten years.

Table 2. Estimation of ten-year CHD risk for women in Singapore

Total points	Ten-Year Risk (%)		
	Chinese	Malay	Indian
5	< 1	< 1	1
6	< 1	< 1	1
7	< 1	1	1
8	< 1	1	1
9	1	1	2
10	1	1	2
11	1	2	3
12	1	2	3
13	1	3	4
14	2	4	6
15	3	5	7
16	3	6	10
17	4	8	12
18	5	10	16
19	7	13	20
20	9	16	> 20
21	12	20	> 20
22	15	> 20	> 20
23	19	> 20	> 20
> 24	> 20	> 20	> 20