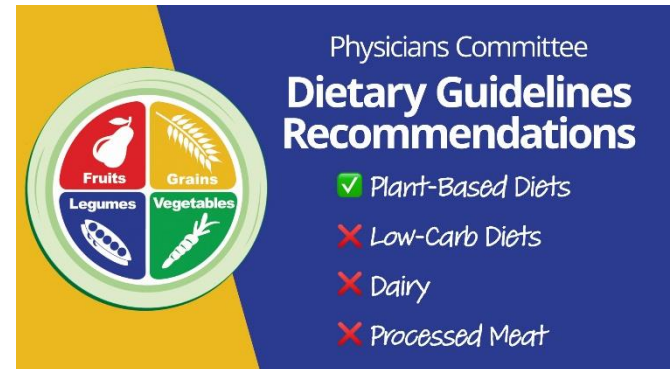
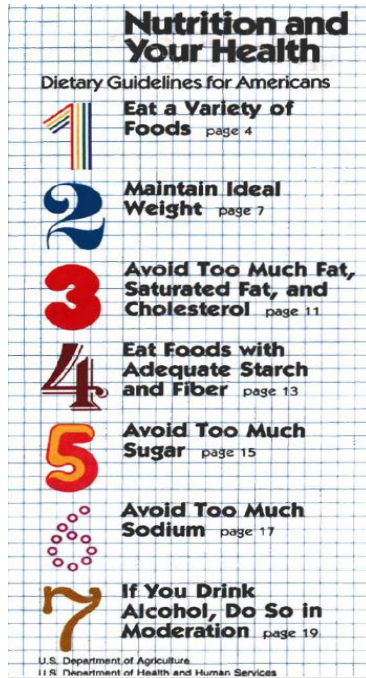


# Background

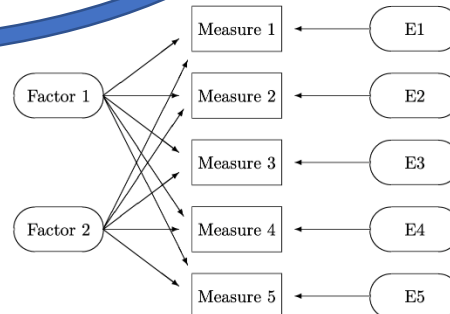
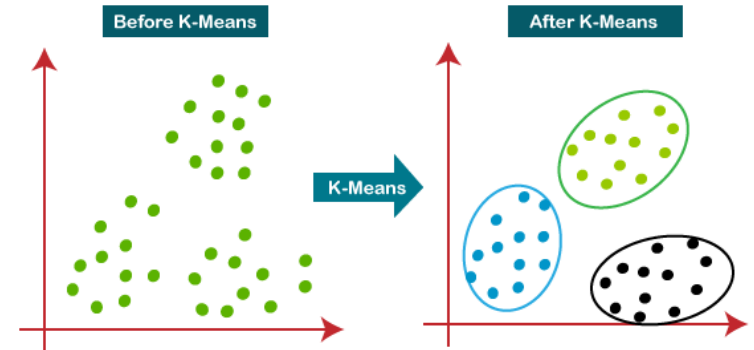
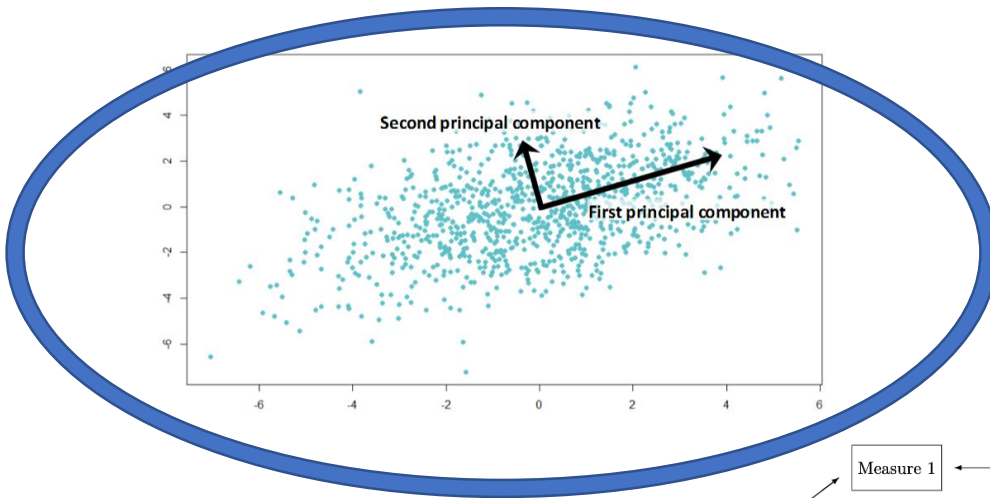
- 단일 식품/영양소 분석 → 식이패턴 분석으로의 전환
  - 개인의 식단은 보통 하나가 아닌 다수의 식품들로 구성되어 있음
  - 관심 식품과 다른 식품들과의 상호작용이 질병 위험에 영향을 미침
  - 한 개인의 식단 내에서 식품들이 서로를 대체함
  - 즉, 식품의 섭취량은 다른 식품의 섭취량에 의해서도 일부 결정됨 (다중공선성)



(Zhao et al, 2021)

# Background

- A-priori vs. A-posteriori 분석
  - A-priori: '양질의 식이패턴'에 대한 기존 이론을 기반으로 점수화
  - A-posteriori: 직접 식이 데이터에서 패턴 추출
    - 탐색적 요인분석, 주성분분석, k-means 군집분석 등 다양한 기법 적용 가능
    - 연구 목적에 맞게 사용해야 함



(Hu et al, 2002)

# Background

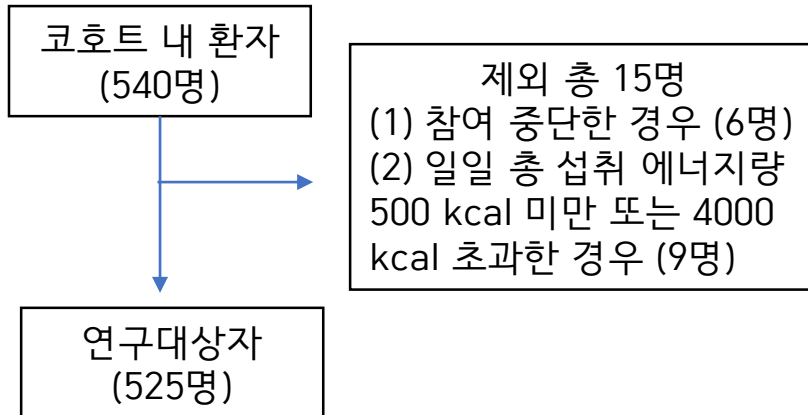
- 대장암 - 식이패턴 관련 선행 연구
  - Meyerhardt et al (2007)
    - 대장암 3기 환자 코호트 연구
    - 요인분석 → 'Healthy' & 'Western' 2가지 식이패턴 추출
    - 'Western' 패턴의 대장암 재발 및 사망 가능성이 유의미하게 높았음
  - 오세영, 이지현, 김효종 (2004)
    - 대장암 환자-대조군 연구
    - 주성분분석 → '웰빙식', '육류와 생선', '우유와 주스', '돼지고기와 술', '밥과 김치', '케이크와 커피' 총 6가지 식이패턴 추출
    - '웰빙식'과 '우유와 주스'는 낮은 발병 위험, '돼지고기와 술', '커피와 케이크'는 높은 발병 위험과 관련됨
  - Park et al (2016)
    - 대장암 환자-대조군 연구
    - 주성분분석 → 'Traditional', 'Prudent', 'Western' 의 3가지 식이패턴 추출
    - 다른 2개 패턴과 달리, Western 식이패턴은 대장암 발병 위험을 높이는 것으로 나타남

- 환자-대조군 연구 → 코호트 연구
- 발병 위험 → 생존/ 암 진행과의 관계

# Methods (1) – Data Used

## • 분석 자료

- 자료원: 서울대학교병원 대장암센터의 대장암 환자 코호트 데이터
- 조사 기간: 2014년 12월 1일~ 2021년 12월 31일
- 연구대상자 선정방법



- 환자 기본 특성 (Table1)

Table1. 환자 기본 특성

요인	전체
대상자 수, n	525
연령(년), mean±SD	61.16±11.17
성별, n(%)	
남자	313(56.62)
여자	212(40.38)
암 병기, n(%) <sup>a</sup>	
0	10(1.90)
I	107(20.38)
II	136(25.90)
III	207(39.43)
IV	65(12.38)
동반질환, <sup>b</sup> n(%)	
없음	192(36.57)
있음	333(63.43)
수술 후 합병증, n(%)	
없음	361(68.76)
있음	164(31.24)
체질량지수(kg/m <sup>2</sup> ), mean±SD	24.13±3.41
흡연력, n(%)	
비흡연 (0)	390(74.29)
과거흡연 (1)	52(9.90)
현재흡연 (2)	83(15.81)
음주력, n(%)	
비음주 (0)	334(63.52)
과거음주 (1)	0(0)
현재음주 (2)	191(36.38)

<sup>a</sup> AJCC(American Joint Committee on cancer)에 근거함

<sup>b</sup> 동반질환은 당뇨/고혈압/심장 질환/뇌 질환/간 질환/폐 질환/기타를 포함함

# Methods (1) – Data Used

- 진단 전 식이 정보
  - 총 106가지 식품으로 구성된 식품섭취빈도조사(Food Frequency Questionnaire)를 통해 수집
  - 환자들은 각 식품별로 진단 전 1년 간의 하루 평균 섭취빈도와 1회당 섭취량을 자율적으로 기입함 → CAN Pro로 식품별 일일 섭취량(g)과 에너지(kcal) 계산
  - 알코올은 이후 음주력이 보정변수 후보에 있으므로 분석에서 제외
  - 총 105가지의 식품들을 35가지 식품군으로 분류 (Table2)

Table2. 35가지 식품군별 포함된 식품 종류 수

Food groups	No. Subitems	Food groups	No. Subitems	Food groups	No. Subitems
Refined grains	19	Red meat	31	Fruits	26
Whole grains	26	Meat by-products	3	Fruit products	10
Tubers and roots	9	Processed meat	12	Green and yellow vegetables	88
Noodles	10	Poultry	13	Light-colored vegetables	71
Rice cakes	17	Fish	28	Pickled vegetables	7
Bread	7	Seafood products	4	Kimchi	11
Cereals and snacks	14	Other seafood	30	Mushrooms	12
Pizza and hamburger	2	Salted fermented seafood	17	Oil and fat	20
Cakes and sweets	27	Seaweeds	11	Condiments and seasonings	64
Legumes	10	Eggs	8	Carbonated beverages	7
Tofu and soy milk	9	Milk	4	Coffee and tea	6
Nuts and seeds	10	Dairy products	15		

(Hoang et al, 2021)

# Methods (2) – PCA

- Why PCA?

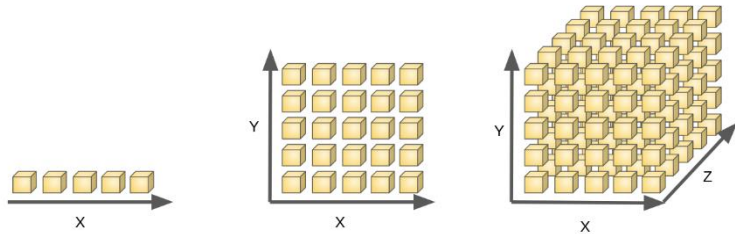
- 다중공선성

- : 설명 변수들 간 상호 관계가 복잡함

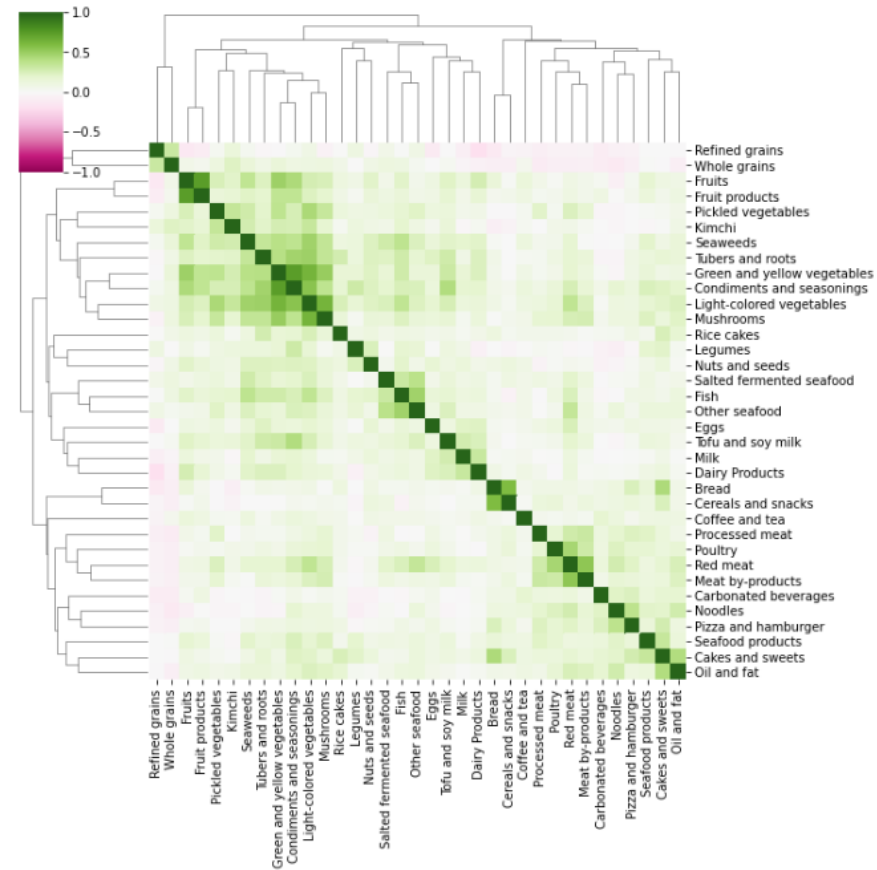
- ex) 초록채소와 해조류 / 붉은육류와 가공육류  
섭취량 간 양의 상관 有

- 차원의 저주

- : 데이터 차원이 높아질수록 공간 상에서 급격히 멀어짐

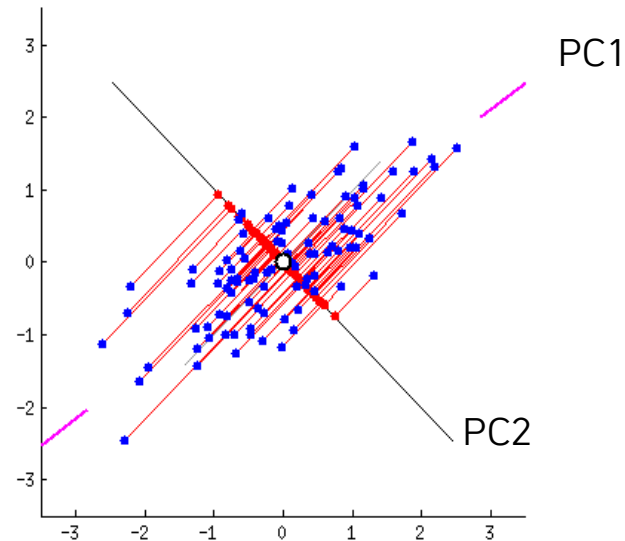
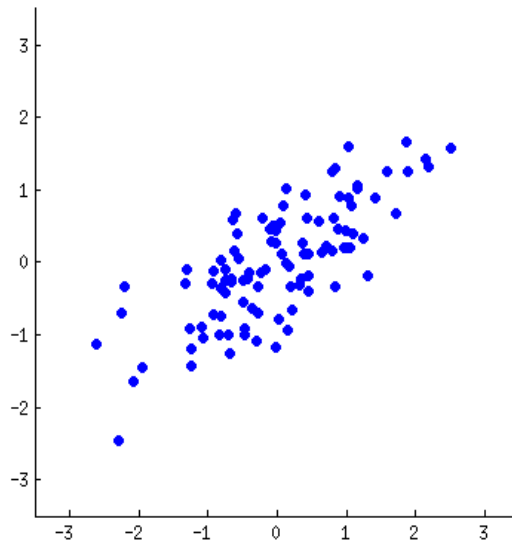


→ 차원 축소 필요



# Methods (2) – PCA

- PCA란?
  - Principal Component Analysis, 주성분분석
  - $p$ 개의 변수로 구성된 데이터를 서로 독립적인  $k(\leq p)$ 개의 변수(=주성분)로 요약하는 기법
  - 요약된  $k$ 개의 주성분은 기존 변수의 선형결합으로 표현됨
  - 원래 데이터의 분산을 최대한 보존하는 새로운 축을 찾고, 그 축에 데이터를 정사영(projection)하는 방법으로 진행

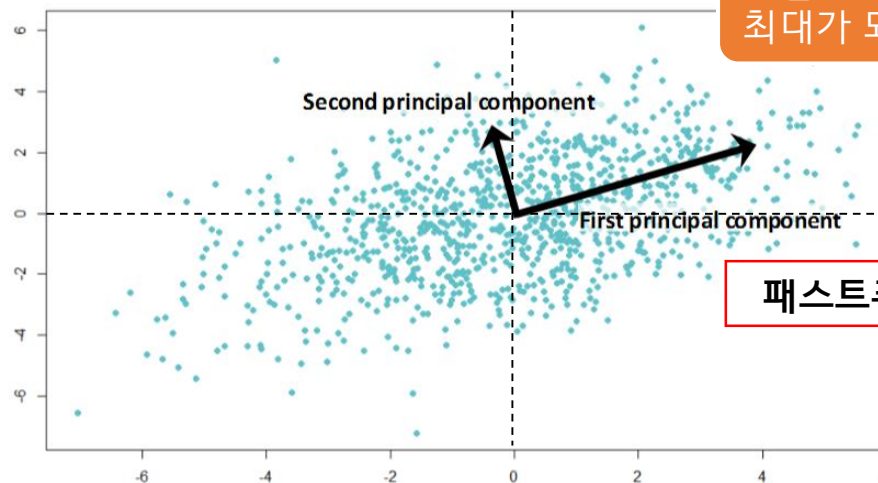


# Methods (2) – PCA

환자 ID	잡곡밥	채소	생선	햄버거	탄산음료	고기
A001	80	40	30	40	10	30
A002	5	15	0	200	40	150
A003	100	50	60	0	0	20



탄산음료 섭취량 (표준화)



원점으로부터 거리 제곱합(분산)이 최대가 되는 방향

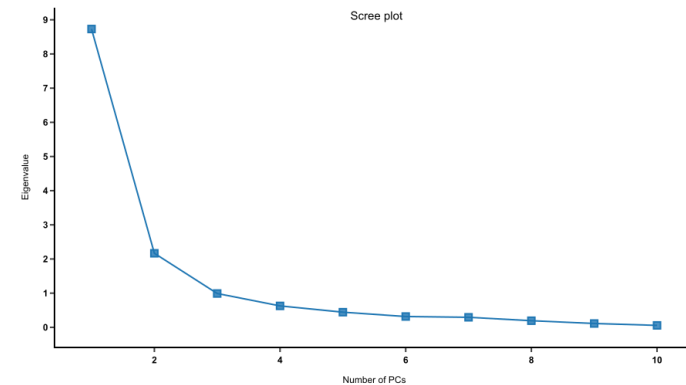
패스트푸드

햄버거 섭취량 (표준화)



# Methods (2) – PCA

- $X = \begin{bmatrix} 80 & 40 & 30 & 40 & 10 & 20 \\ 5 & 15 & 0 & 100 & 40 & 200 \\ 100 & 50 & 50 & 30 & 0 & 0 \end{bmatrix}$
- 공분산행렬  $\Sigma = (X - E(X))^T (X - E(X))$  를 계산
- $\Sigma \vec{e}_i = \lambda_i \vec{e}_i$ 인  $\lambda_i, \vec{e}_i$  들을 고유값 분해로 찾기
- 단위 고유벡터( $\vec{e}_i$ )가 i번째 주성분의 부하량 (=방향), 고유값 ( $\lambda_i$ )이 해당 주성분의 설명력을 나타냄
- $X_{scaled}[\vec{e}_1 \ \vec{e}_2 \ \vec{e}_3 \ .. ] \rightarrow$  개인별 주성분값 계산



환자 ID	PC1	PC2		
A001	5.0	0.1		
A002	-0.2	3.9		
A003	2.9	0.3		

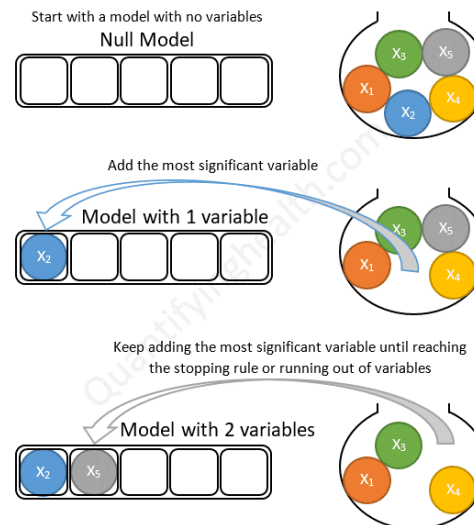
Healthy 패턴 점수

Unhealthy 패턴 점수

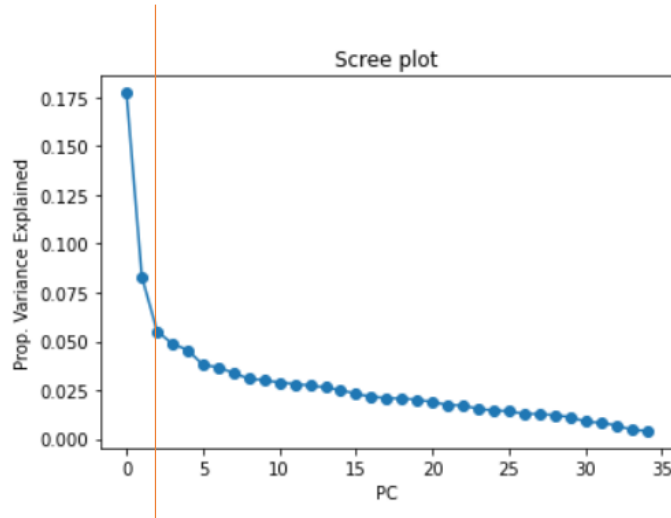
# Methods (3) – Survival Analysis

- 생존 기간(Overall Lifetime), 무진행 질병 보존 기간(Progression-free Lifetime) 각각에 Cox 비례위험모형 적용
- 검정력 높이기 위해 집단 나누지 않고, 패턴별 점수 (=주성분 값) 그대로 설명변수로 사용  
→ HR 1단위 증가에 따른 위험비로 해석
- 보정변수는 후보들 중 Forward Stepwise Selection 적용하여 AIC 가장 높은 모형에 들어간 변수 조합으로 선택

Forward stepwise selection example with 5 variables:



# Results



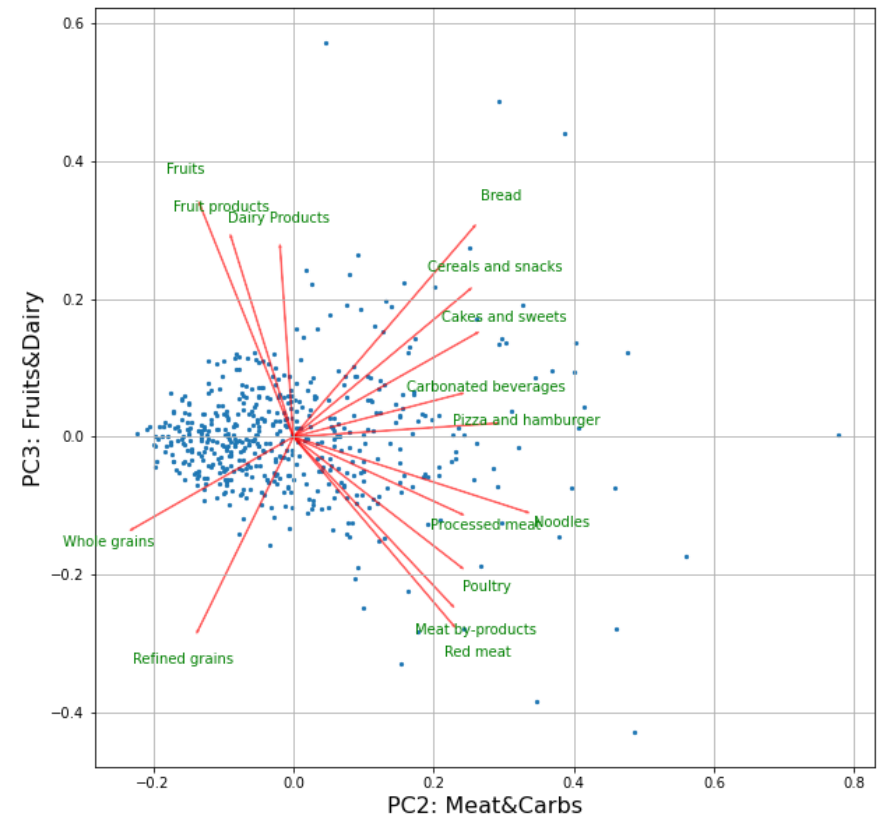
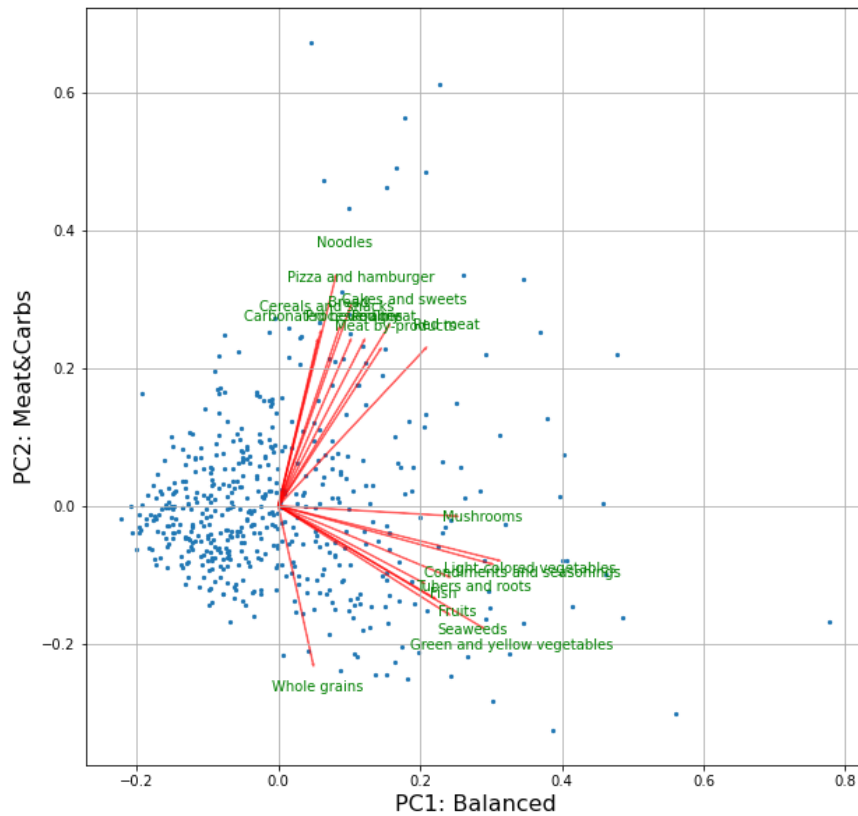
- Elbow method에 따라 주성분 개수 3개로 선정
- PCA를 통해 식품군별 섭취량 데이터를 총 3가지의 식이패턴 점수로 축소
- 부하량 높은 식품군에 따라 각 패턴 명명 (Table3)
  - Balanced: 각종 채소, 생선, 육류 ↑
  - Meat & Carbs: 육류, 패스트푸드 ↑ 통곡물 ↓
  - Fruits & Dairy: 과일, 유제품 ↑ 육류, 정제곡물 ↓

Table3. 식품군별 주성분 부하량

	food group	Balanced	Meat & Carbs	Fruits & Dairy
0	Refined grains	-	-	-0.28
1	Whole grains	-	-0.23	-
2	Tubers and roots	0.24	-	-
3	Noodles	-	0.33	-
4	Rice cakes	-	-	-
5	Bread	-	0.26	0.3
6	Cereals and snacks	-	0.25	0.21
7	Pizza and hamburger	-	0.29	-
8	Cakes and sweets	-	0.26	-
9	Legumes	-	-	-
10	Tofu and soy milk	-	-	-
11	Nuts and seeds	-	-	-
12	Red meat	0.21	0.23	-0.27
13	Meat by-products	-	0.23	-0.24
14	Processed meat	-	0.24	-
15	Poultry	-	0.24	-
16	Fish	0.2	-	-
17	Seafood products	-	-	-
18	Other seafood	-	-	-0.2
19	Salted fermented seafood	-	-	-
20	Seaweeds	0.24	-	-
21	Eggs	-	-	-
22	Milk	-	-	-
23	Dairy Products	-	-	0.28
24	Fruits	0.22	-	0.34
25	Fruit products	-	-	0.29
26	Green and yellow vegetables	0.29	-	-
27	Light-colored vegetables	0.31	-	-
28	Pickled vegetables	-	-	-
29	Kimchi	-	-	-
30	Mushrooms	0.25	-	-
31	Oil and fat	-	0.2	-
32	Condiments and seasonings	0.3	-	-
33	Carbonated beverages	-	0.24	-
34	Coffee and tea	-	-	-

# Results

- Biplot으로 시각화



# Results

- 생존과의 관계

- 전진 단계적 선택법 의해, 나이, BMI, 병기 보정변수로 선정
  - : 나이 많을수록, BMI 낮을수록, 병기 높을수록 사망 위험 유의미하게 높음
- Meat & Carbs 패턴점수의 사망 위험비가 1보다 크고 다른 두 패턴에 비해 높게 나타났지만, 결과는 유의미하지 않았음
  - : 대장암 환자의 생존과 유의미하게 관련된 식이패턴 찾지 못하였음

Table4. 환자 생존에 대한 Cox모형 적합 결과

	log(HR)	HR	SE(log(HR))	Z	p-value
나이	0.03	1.03	0.01	2.95	<.001
<u>BMI</u>	-0.07	0.92	0.04	-2.14	<.05
병기	1.21	3.36	0.16	7.77	<.05
Balanced 패턴점수	0.01	1.01	0.04	0.26	0.79
Meat & Carbs 패턴점수	0.10	1.11	0.08	1.34	0.18
Fruits & Dairy 패턴점수	0.07	1.07	0.08	0.88	0.38

# Results

- 암 진행과의 관계

- 전진 단계적 선택법 의해, 나이, BMI, 병기 보정변수로 선정
  - : 나이 많을수록, BMI 낮을수록, 병기 높을수록 암 진행 위험 유의미하게 높음
- 마찬가지로 Meat & Carbs 패턴점수의 암 진행 위험비가 1보다 크고 다른 두 패턴에 비해 높게 나타났지만, 결과는 유의미하지 않았음
  - : 대장암 환자의 암 진행과 유의미하게 관련된 식이패턴을 찾지 못하였음

Table5. 암 진행에 대한 Cox모형 적합 결과

	log(HR)	HR	SE(log(HR))	Z	p-value
나이	0.02	1.03	0.01	2.81	<.01
<u>BMI</u>	-0.07	0.93	0.03	-2.25	<.05
병기	0.92	2.52	0.12	7.57	<.001
Balanced 패턴점수	0.00	1.00	0.04	0.01	0.99
Meat & Carbs 패턴점수	0.06	1.06	0.07	0.84	0.40
Fruits & Dairy 패턴점수	0.01	1.01	0.07	0.08	0.93

# Discussion

- 수집된 식이 정보의 한계
  - 식품섭취빈도조사: 설문지를 통한 자기보고 형식
    - 참여자들의 회상 바이어스, 섭취량 계산 과정에서 오차 발생했을 가능성 有
  - 진단 전 식이 정보
    - 진단 후 식생활습관이 바뀐 경우, 진단 전 섭취보다 진단 후 섭취가 생존이나 진행에 더 유의미하게 작용했을 수 있음
- 주성분분석 적용의 한계
  - Scree plot을 활용한 주성분 수 채택에 있어서 연구자의 주관이 들어감
  - 독립 변수들의 차원축소가 목적이기 때문에 종속 변수인 질병 관련 위험에 대한 기존 정보는 활용할 수 없음
    - Reduced Rank 회귀모형
  - 결과가 유의해도 패턴점수가 모든 식품군 섭취량의 선형결합 형태이기 때문에 해석이 어려움
    - k-means 군집분석

# References

- 세계암연구재단. <https://www.wcrf.org/cancer-trends/worldwide-cancer-data/>.
- 국가암등록사업 연례 보고서(2019년 암등록통계). 국립암센터.
- 오세영, 이지현, 김효종.(2004).대장암 위험도와 관련된 식생활 행동 분석.한국영양학회지,37(3),202-209.
- Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol*. 2002;13(1):3-9. doi:10.1097/00041433-200202000-00002
- Hoang T, Kim H, Kim J. Dietary Intake in Association with All-Cause Mortality and Colorectal Cancer Mortality among Colorectal Cancer Survivors: A Systematic Review and Meta-Analysis of Prospective Studies. *Cancers (Basel)*. 2020;12(11):3391. Published 2020 Nov 16. doi:10.3390/cancers12113391
- Zhao, J., Li, Z., Gao, Q. *et al*. A review of statistical methods for dietary pattern analysis. *Nutr J* 20, 37 (2021). <https://doi.org/10.1186/s12937-021-00692-7>
- 미국인 식생활 가이드라인 홈페이지. <https://www.dietaryguidelines.gov/>.
- Meyerhardt JA, Niedzwiecki D, Hollis D, et al. Association of dietary patterns with cancer recurrence and survival in patients with stage III colon cancer. *JAMA*. 2007;298(7):754-764. doi:10.1001/jama.298.7.754
- Ginter MA, McCullough ML, Gapstur SM, Campbell PT. Associations of Pre- and Postdiagnosis Diet Quality With Risk of Mortality Among Men and Women With Colorectal Cancer [published online ahead of print, 2018 Oct 19]. *J Clin Oncol*. 2018;36(34):JCO1800714. doi:10.1200/JCO.18.00714
- Park Y, Lee J, Oh JH, Shin A, Kim J. Dietary patterns and colorectal cancer risk in a Korean population: A case-control study. *Medicine (Baltimore)*. 2016;95(25):e3759.
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning : with Applications in R. New York :Springer, 2013



# References

- Fox J & Weisberg S. Cox Proportional-Hazards Regression for Survival Data in R.
- Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: current state of the art and future prospects. *F1000Res*. 2017;6:926. Published 2017 Jun 16. doi:10.12688/f1000research.10703.1
- Hoffmann K, Schulze MB, Schienkiewitz A, Nöthlings U, Boeing H. Application of a New Statistical Method to Derive Dietary Patterns in Nutritional Epidemiology. *American Journal of Epidemiology*. 2004;159(10):935-944. doi:10.1093/aje/kwh134
- Batis C, Mendez MA, Gordon-Larsen P, Sotres-Alvarez D, Adair L, Popkin B. Using both principal component analysis and reduced rank regression to study dietary patterns and diabetes in Chinese adults. *Public Health Nutr*. 2016;19(2):195-203. doi:10.1017/S1368980014003103
- Hoang T, Lee J, Kim J. Differences in Dietary Patterns Identified by the Gaussian Graphical Model in Korean Adults With and Without a Self-Reported Cancer Diagnosis. *J Acad Nutr Diet*. 2021;121(8):1484-1496.e3. doi:10.1016/j.jand.2020.11.006
- Northstone K, Ness AR, Emmett PM, Rogers IS. Adjusting for energy intake in dietary pattern investigations using principal components analysis. *Eur J Clin Nutr*. 2008;62(7):931-938. doi:10.1038/sj.ejcn.1602789