**여러 변수들이 죽음에 영향을 끼치는 영향도**

**요약**

총 9가지의 요소(성별, 인종, 나이, 입원, 외래, 응급, 내원일수, 진단, 의약품)의 죽음에 관한 영향도의 우선순위를 매겨보았다. 결론은 높은 순으로 입원, 나이, 진단, 의약품, 내원 일수, 응급, 외래, 성별, 인종 순이었다. 이와 같은 결과를 내기 위해 psycopg2, numpy, matplotlib.pyplot, pandas, from sklearn.linear\_model import LinearRegression, from sklearn.preprocessing import OneHotEncoder 등의 모듈을 사용하였으며 각각 PostgreSQL과의 연동, 데이터의 시각화, 데이터의 연관도 분석, 데이터의 전처리 등을 위해 사용하였다.

**본론**

여러 변수들이 죽음에 얼마나 영향을 끼치는지 영향도를 확인하기로 하였다. 그 변수들은 총 9가지로 성별, 인종, 나이, 입원, 외래, 응급, 내원일수, 진단, 의약품이 있다. 우선 하나하나 분석해 보기로 했다.

1. 성별

성별은 두가지로 남성, 여성이 있다. 죽은 남성은 전체의 11%정도이였고 죽은 여성은 18%정도 였다. 남성이 여성보다 살짝 높아 남성이 여성보다 죽을 가능성이 조금 더 높다. 하지만 이는 유의미한 차이는 아니다.

1. 인종

인종은 4가지로 white, black or african American, Asian, no matching이 있다. 비율로 보았을 때, 순서대로, 15%, 16%, 15%정도로 black or african American이 1%높지만 의미가 없는 차이이다.

1. 나이

나이를 bar chart로 그려보았을 때 죽은 사람들은 약간 오른쪽으로 쏠린 정규분포를 그린다. 전체의 사람들을 보았을 때는 죽은 사람들의 그래프를 마치 좌우로 반전한 듯한 모습이 보인다. 이를 비율로 보았을 때, 예상외로 높은 유아기의 죽음과 높은 비율의 고령층의 죽음이 눈에 띈다. 20대의 경우 10%대로 매우 낮은 수치가 눈에 띈다.

1. 방문 유형

방문 유형은 총 3가지로 입원, 외래, 응급이 그에 해당한다. 외래와 응급은 24%, 28%로 비교적 비슷한 수치를 띄었지만 입원은 50%정도로 상당히 높은 수치를 보였다.

1. 내원 일수

내원 일수는 각 환자가 여태까지 병원을 방문한 날 수이다. 죽은 사람은 매일이 거의 비슷하지만 초반이 더 높은 것으로 보아 초반이 상당한 고비로 보이고 전체 인원은 초반으로 급격히 쏠린 그래프로 보인다. 하지만 이를 비율로 보면 달라진다. 초반에는 오히려 10%정도로 매우 낮고 100일이 넘어갈수록 살 확률은 희박하다고 보인다.

1. 진단과 의약품

이 둘은 그래프나 비율로 보았을 때 크게 눈에 띄는 점은 보이지 않았다. 하지만 이 둘은 상당한 연관관계가 있다고 생각했다. 예를 들어 감기에 걸리면 감기약을 진단 받는 것처럼 말이다. 그래서 가설을 세웠다.

1.의사가 병을 알아보고 제대로 약을 주었음.

2.의사가 병을 알지만 이상한 약을 주었음.

3.의사가 병을 몰랐지만 제대로 약을 주었음.

4.의사가 병도 모르고 약도 이상한 약을 주었음.

그래서 같은 병에 대해서 준 약들과 그 약들의 죽음에 주는 영향을 보았다.

1. 사망일

사망일의 경우 눈에 띄는 점은 2020에 상당한 사람이 죽었단 점이다. 이는 요즘 유행하는 코로나로 인한 영향으로 보인다.

**결론**

이와 같은 여러 분석을 통한 결과로 나는 입원, 나이, 진단, 의약품, 내원 일수, 응급, 외래, 성별, 인종 순으로 죽음에 영향을 준다고 생각한다.