Phongsathorn Röser
09.12.2023

Preparing for Influenza Season

Interim Report

Project Overview

Motivation:
The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff.
Objective:
Determine when to send staff, and how many, to each state.
Scope:
The agency covers all hospitals in each of the 50 states of the United States, and the project will plan for the upcoming influenza season.

Research Hypothesis:

1. Areas with big population size display higher numbers of death cases related by Influenza.
2. Older population group will likely die from death related by influenza.
3. Higher numbers of death by influenza are observed during winter months.

Data Overview:

Death cases related to influenza:

Data set contains numbers of death cases by influenza and the respective information such as name of the states, year, month and age groups of population until 85+ years from 2009 until 2017.

Consensus data:

Data sets includes demographic information about different counties of each states and the respective population size which is categorize in age groups until 85+ years from 2009 to 2017.

Limitation of the data sets:

Death cases related to influenza:

- Because no further information apart from the death status of the patients exist, we can not assume any pre-condition of the patients health.
- We also can´t say if the patient is already vaccinated before death.
- Apart from this, information about Puerto Rico is not present in the data set, therefore we can not make plans for Puerto Rico.

- The record of “Sum death cases” were either described as numbers or “suppressed”. Hence, these rows needs to be replaced by random numbers using randomization. By doing this, we won´t know how close we are to the reality.

Consensus data:

- The numbers of population are estimates, hence the sum of the numbers from the different age groups may not sum up to the total in the first, which gives us the wrong impression of the total population size.
- Using the data from “Death by Influenza”, we would not cover the real percentage of death cases for the entire population.
- No determination of high and low populated areas in USA. Because population density is not the same as the number of population which might impact the infection rate/ death case by influenza. A decision needs to be make to describe a high population area in order to compare influenza outbreaks between low population and high population size.

Variables	**Mean**	**Standard deviation**
Death of Population under 65 years	39.61	15.38
Death of Population over 65 years	74.70	86.74
Grand Total Death	119.26	97.49
Total Popultion	5973848.67	6799372.13
Population with <1 billion population	9067.40	3595.78
Population with > 1 billion population	28270.80	12877.93

Variables	***Correlation coeffecient***	***Interpretation***
Death of Population > 65 & Total Population size	0.88	strong correlation which means that, with the increasing population size, the sum of death among population over 65 will also increase
Grand Total Death & Total Population size	0.88	strong correlation which means, that the number of death cases grows with the increasing population size
Death of Population >65 & Death Total	0.99	strong correlation which means that, the death cases among population contributes to the increasing number of total death cases

Results and Insights:

<u>Hypothesis 1:</u>

Population over 65 years old are likely to die from death by influenza

H0: Death cases of Population older than 65 years are equal or less than population under 65 years

H1: Death cases of Population older 65 years old are higher than population under 65 years.

Reults:

P-value is <0.05. H0 rejected, H1 accepted. Population under 65 years have a chance of less than 5% to die from influenza in comparison to population over 65 years

<u>Hypothesis 2:</u>

Bigger Population size leads to higher number in death cases by influenza

H0: States with high numbers in Population show equal or less number of death cases in total

H1: States with high numbers in Population show higher numbers of death cases in total

Results:

P-value is significantly lower than 0.05, therefore H0 will be rejected and H1 accepted. The results of the statistical test suggest that the chance of people under 65 to die from influenza is less than 5% in comparison to population over 65 years.

<u>Hypothesis 3:</u>

Death by influenza are more observed during winter months.

H0: Death cases during December, January and February are equal or less than during other months.

H1: More death cases are observed during December, January and February than during other months

Results:

Death cases between winter months and the rest of the remaining months of the year show significant differences in terms of death cases. H0 needs to be rejected and H1 accepted. The results shows that, during winter months, there is a > 99% chance that death cases are higher than during the remaining months of the year.

Remaining Analysis and Next Steps:

Hypothesis 1:

Furhter analysis should investigate if vaccinations against flu has impact on the sum of death cases or not. As well, what kind of pre-health condition leads to the lethal outcome of influenza?

We can conduct such analysis for every age group who shows susceptibility to influenza symptoms. This will help us to relocate resources and also inform us if young children are under the same danger like elderly population or not. We should also keep track of areas with higher number of population over the age of 65 years, to decrease/ specify our area of interest.

Hypothesis 2:
Since birth rate and death rates are constantly changing, we should keep track on population size across USA and focus on areas that we consider "big population" areas. By doing this we can at least decrease the areas of interest which needs more resources in terms of influenza protection.

Hypothesis 3:
Knowing this, we are able to make assumption of the next influenza outbreak season. Further analysis could investigate the death cases of states that display high numbers of death cases during these winter months. This way, we can increase our resources during outbreak months and ease during other months such as June, July and August. The numbers of patient and nurses/ hospital staff should be investigated for every month to keep track at which ratio the efficiency of each hospital wont be negatively affected during flu season.

Appendix

Project overview:
https://images.careerfoundry.com/public/courses/data-immersion/A1-A2_Influenza_Project/A1-A2-data-immersion-project-brief.pdf

Hypothesis:

- Which location displays the most cases of death by influenza in the history of the USA? How many people were vaccinated during the events?
- Do adjacent state also display high cases of death by influenza?
- What time of the year displayed to highest number of death?
-
- 2. Where is the location with the most vulnerable population? Which population group are the most vulnerable one in this are?
- Are locations from colder region more affected than hotter region?
- What is the population size of the affected location?
-
- 3. What are the ratio between nurse/patient in every state? Why are the ratio the way they are?
- Could the ratio be changed without affecting the functionality of the hospitals?
-
- 4. How did the flu season develop in the last decades? What year showed the highest number of cases?
- How many of the cases end up in death?
- Does the development of flu season correlate to changes in the environment?
- Is the flu season always at the same time period?
-
- 5. How many people are vaccinated in each state during the flu season? From all the vaccinated people, how many needed to be hospitalized?
- From all the vaccinated people, how many died?
- Which population group was most vaccinated?

Data overview

Influenza deaths by geography: Collected by the CDC

Variable	*Mean*	*Median*	*SD*	*Type*
Death < 65	39.61765	37	15.37585	Numeric
Death > 65	74.70171	46	86.73812	Numeric
Death Total	119	87	97.49242	Numeric

Population data by geography, time, age, and gender: Collected by US Census Bureau.

Variables	Minimum	Maximum	Mean	Roundup
Total population	41.00	10105722.00	107854.38	107855
Male Total population	23.00	4979641.00	53027.66	53028
Female Total population	15.00	5126081.00	54826.73	54827
Under 5 years	0.00	733897125.00	4723638.27	4723639
5 to 9 years	0.00	644044896.00	4793742.94	4793743
10 to 14 years	0.00	702594432.00	4786003.75	4786004
15 to 19 years	0.00	753656519.00	5052080.15	5052081
20 to 24 years	0.00	777987834.00	5199112.95	5199113
25 to 29 years	0.00	814629555.00	5022169.43	5022170
30 to 34 years	0.00	754286625.00	4740059.25	4740060
35 to 39 years	0.00	753467715.00	4737274.92	4737275
40 to 44 years	4.00	733897125.00	4944940.15	4944941
45 to 49 years	0.00	704717784.00	5098427.52	5098428
50 to 54 years	0.00	682610384.00	5143017.03	5143018
55 to 59 years	17.00	612341668.00	4619334.51	4619335
60 to 64 years	0.00	488735947.00	3920712.71	3920713
65 to 69 years	0.00	381458744.00	3029334.36	3029335
70 to 74 years	0.00	271036476.00	2236269.07	2236270
75 to 79 years	0.00	211200255.00	1723856.10	1723857
80 to 84 years	0.00	160614208.00	1368557.12	1368558
85 years and over	0.00	170971635.00	1291784.17	1291785

Results and Insight

The sample size of the second hypothesis test was only 10 for each variable level, for more realistic representation, data from more record should be used for future analysis.