c

PREPARING FOR INFLUENZA SEASON

INTERIM REPORT

Graham Field

3/21/23

1.10: Consolidating Analytical Insights

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 the temporary staff.
- **Objectives –** 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.

Hypothesis

- People Older than 65 years old are more at risk of dying from influenza.

Data Overview

- **Influenza deaths data set –** This data represents the annual deaths from Influenza by location and age.
  Source: CDC Download Data Set

- **Census Data –** This data show people that are alive categorized by Gender, location, and age.
  US Census Bureau

Data Limitations

Influenza deaths data set

- The data is suppressed of there are less than 10 individuals, this is done to prevent individuals from being identified.
- There is also only one cause of death listed, this may hide contributing factors such as Immunodeficiency syndromes or other underlying conditions such as obesity which may be major contributing factors to people developing secondary illnesses which kill them.

Census Data

- Data Is Captured manually and entered manually, this can lead to inaccuracies twice in the process which can be very hard to detect and rectify.

Descriptive Analysis

Data Spread Combi standardised data			
Dataset Name	65>Years CDC	65<Years CDC	Grand Total cdc
Sample or Population ?	Sample	Sample	Population
Normal Distribution ?	Normal	Normal	Normal
Variance	14950,48953	952349,8858	0,00081719
Standard Deviation 1	122,2721944	975,8841559	0,028586538
Standard Deviation 2	244,5443889	1951,768312	0,057173077
Standard Deviation 3	366,8165833	2927,652468	0,085759615
Mean	0,001768866	0,00808447	0,002944668
Number of outliers STD1	58	43	57
Number of outliers STD2	18	29	19
Number of outliers STD3	0	0	0
Outlier Percentage STD1	13%	9%	12%
Outlier Percentage STD2	4%	6%	4%
Outlier Percentage STD3	0%	0%	0%
+1 Standard deviation upper	1865,632522	613,1153317	2522,249933
+2 Standard deviation upper	2841,516678	735,3875261	3609,908362
+3 Standard deviation upper	6745,053302	45342,31173	7960,542079
-1 Standard deviation lower	-86,13578991	368,5709428	346,933074
-2 Standard deviation lower	-1062,019946	246,2987484	-740,7253552
-3 Standard deviation lower	-4965,55657	-44360,62546	-5091,359072
Correlation Combi Data Deathy Rate			
Variables	65>Years Death Rate	65< Years Death Rate	Grand Total Death Rate
Proposed Relationship	People over retirement age are likely to die from influenza		
Correlation Coefficient	0,972	0,972	
Strength of Correlation	Strong	Strong	
Usefulness / Interpretation	From this data we can see there is a strong correlation between being old and dying from influenza		

Notes:
- The population in every case is over 30 and so the distribution will be considered normal.
- For the lower standard deviations the Numbers are negative, this is impossible for this type of data so we will only use the upper when counting the outliers.
- The Census and CDC data for over 65 and in total would be with in the 2nd standard deviation as the data does not exceed 95%.

Results and Insights

Statistical Hypothesis	
Research Hypothesis	The Older you get the more likely you are to die from Influenza
Independent Variable	Death Rate
Dependant Variable	Age Groupings
Null Hypothesis	the death rate remains the same or is greater the older you get (<65 years Death Rate : "=" "<" 0.05)
Alternative Hypothesis	the death rate decreases the older you get (<65 years Death rate : ">" 0.05)
One Vs Two Tailed Test	Given that people can only get older and die this would suggest that we are only interested in results in in one direction. A possible solution to bringing down the mortality isn't making people younger as its impossible. This test would be one tailed as we are only considering data on one side of the mean. given that our lower standard deviations are negative which in this case is also impossible. we are technically looking at right tail P-value.
Alpha	= 0.05
P- Value	0,042
Comparison of P-Value to Significance Level	Considering that the P- value is Below the alpha value we can eliminate the null hypothesis. With a significance value of less than 5% we can see that there is a significant difference in death rate between over 65 and under 65 population groups. This would suggest that the more vulnerable older group is at significantly greater risk of dying.

t-Test: Paired Two Sample for Means

	65<Death Rate	*65> Death Rate*
Mean	0,001768866	0,00808447
Variance	0,000330038	0,007072895
Observations	459	459
Pearson Correlation	0,971935732	
Hypothesized Mean Difference	0	
df	458	
	-	
t Stat	2,032229759	
P(T<=t) one-tail	0,02135346	
t Critical one-tail	1,648187415	
P(T<=t) two-tail	0,042706921	
t Critical two-tail	1,965157098	

Remaining Procedural steps

- Identify areas that are most at risk.
- Identify times where outbreaks are most likely to occur.
- Visualize the spread of Influenza over the country.
- Visualize the spread over the areas with the highest infection rates.
- Identify factors from areas where infection rates are low and differentiate between areas with high infection rates.
- Develop a strategy to help prevent or slow infection.
- Consolidate information into a concise and cohesive presentation to present to steak holders.

Appendix

Project Deliverables

- identify trends in staffing need and create a report and strategy to make sure that hospitals and clinics are not over or understaffed with given demand.
- identify vulnerable groups of people. Help high light trends and high-risk behaviors that may lead to higher rates of infection. Additionally identify trends and behaviors that lead to lower risk of infection. Develop guideline to help high risk individuals avoid becoming sick.
- High light the potential risks and rewards of being vaccinated.
- Develop a marketing campaign to elevate public awareness the current situation with some basic guidelines. Include access to refined data sets and more detailed strategies.
- Create reporting tools to develop larger and more relevant data sets.

Audience Definition

Stakeholders

- Medical agency frontline staff (nurses, physician assistants, and doctors)
- Hospitals and clinics using the staffing agency's services
- Influenza patients
- Staffing agency administrators

High Risk and Vulnerable Populations

- Adults over 65 years of age
- Children under 5 years
- Pregnant women
- Individuals with HIV/AIDs, Cancer, Heart Disease, Stroke, Diabetes, Asthma
- Children with Neurological Disorders
- Health care workers

Success Factors

The project's success will be based on:

- A staffing plan that utilizes all available agency staff per state requirements, without necessitating additional resources.
- Minimal instances of understaffing and overstaffing across states (a state can be considered understaffed if the staff-to-patient ratio is lower than 90% of the required ratio and overstaffed if greater than 110%.

Assumptions & Constraints

Assumptions:

- Vulnerable populations suffer the most-severe impacts from the flu and are the most likely to end up in the hospital.
- Flu shots decrease the chance of becoming infected with the flu.

Constraints:

- The staffing agency has a limited number of nurses, physician assistants, and doctors on staff.
- There's no money to hire additional medical personnel.

Data Profile Census			
Variable	Min	Max	Mean
Year	2009	2017	2012,999655
Total population	41	10105722	97842,18223
Male Total population	23	4979641	48121,57243
Female Total population	15	5126081	49720,6098
Under 5 years	0	733897,125	6309,255033
5 to 9 years	0	665400,06	6377,695594
10 to 14 years	0	724111,83	6377,695594
15 to 19 years	0	753656,519	6804,061966
20 to 24 years	0	777987,834	6908,612746
25 to 29 years	0	831276	6714,164599
30 to 34 years	0	762619	6436,455928
35 to 39 years	0	753467,715	6344,996549
40 to 44 years	1,001	733897,125	6613,688983
45 to 49 years	0	704717,784	6900,900678
50 to 54 years	0	683886,54	6921,971039
55 to 59 years	2,976	628513	6292,770023
60 to 64 years	0	535357	5416,440305
65 to 69 years	0	415243	4189,202945
70 to 74 years	0	295420	3128,959231
75 to 79 years	0	215181	2394,108835
80 to 84 years	0	161647	1821,318912
85 years and over	0	177493	1774,313884

Variable	Time Variant	Structure	Data Type
County	Invariant	Structured	Qualitative/ Nominal
State	Invariant	Structured	Qualitative/ Nominal
Year	Invariant	Structured	Qualitative/ Nominal
Total population	Variant	Structured	Quantitative/ Discrete
Male Total population	Variant	Structured	Quantitative/ Discrete
Female Total population	Variant	Structured	Quantitative/ Discrete
Under 5 years	Variant	Structured	Quantitative/ Discrete
5 to 9 years	Variant	Structured	Quantitative/ Discrete
10 to 14 years	Variant	Structured	Quantitative/ Discrete
15 to 19 years	Variant	Structured	Quantitative/ Discrete
20 to 24 years	Variant	Structured	Quantitative/ Discrete
25 to 29 years	Variant	Structured	Quantitative/ Discrete
30 to 34 years	Variant	Structured	Quantitative/ Discrete
35 to 39 years	Variant	Structured	Quantitative/ Discrete
40 to 44 years	Variant	Structured	Quantitative/ Discrete
45 to 49 years	Variant	Structured	Quantitative/ Discrete
50 to 54 years	Variant	Structured	Quantitative/ Discrete
55 to 59 years	Variant	Structured	Quantitative/ Discrete
60 to 64 years	Variant	Structured	Quantitative/ Discrete
65 to 69 years	Variant	Structured	Quantitative/ Discrete
70 to 74 years	Variant	Structured	Quantitative/ Discrete
75 to 79 years	Variant	Structured	Quantitative/ Discrete
80 to 84 years	Variant	Structured	Quantitative/ Discrete
85 years and over	Variant	Structured	Quantitative/ Discrete

Data Accuracy
The Data Is captured as Continuous which is impossible for the type of data captured this will be addressed in cleaning
when converting the CSV file to excel, I was challenged with the state and county separating and forming their own data sets so I added state to the csv file before translating it into excel

Data Consistency	
County	28985
State	28985
Year	28985
Male Total population	1394803777
Female Total population	1441151875
Total Male + Female	2835955652
Total population	2835955652

Data Profile CDC Deaths				
	Variable	Time variant	Structure	Data Type
1	State	Invariant	Structured	Qualitative / Nominal
2	State Code	Invariant	Structured	Qualitative / Ordinal
3	Year	Invariant	Structured	Qualitative / Ordinal
4	Month	Invariant	Structured	Qualitative / Ordinal
5	Month Code	Invariant	Structured	Qualitative / Ordinal
6	Ten-Year Age Groups	Invariant	Structured	Qualitative / Ordinal
7	Ten-Year Age Groups Code	Invariant	Structured	Qualitative / Ordinal
8	Deaths	Variant	Structured	Quantitative / Discrete

Variable	Column	Changes
State	A	A column was added with cleaned data (State Cleaned)
State Code	C	No Changes
Year	D	A column was added with clead data (Year Cleaned)
Month	F	No Changes
Month Code	G	No Changes
Ten-Year Age Groups	H	No Changes
Ten-Year Age Groups Code	I	No Changes
Deaths	J	A Column was added with cleaned data (Deaths Cleaned)
State Cleaned	B	Abreviations for states were changes to state name, N/A was chanegd to correct state name. please refer to the abrevation table for more details.
Year Cleaned	E	The incorrect year was changed D66081 to D66097 were changed from 20133 to 2013 a total of 17 changes were made.
Deaths Averaged	K	Deaths were average to 4.5 as the CDC supressd data that is between 0-9

Variable	Column	Count Per Variable	Consistency	Total Count
State	A	114/1152	Inconsistant	66096
State Code	C	1296	Consistent	66096
Year	D	7344 / 7327 / 17	Inconsistant	66096
Month	F	612	Consistent	66096
Month Code	G	612	Consistent	66096
Ten-Year Age Groups	H	5508	Consistent	66096
Ten-Year Age Groups Code	I	5508	Consistent	66096
Deaths	J	Total : 415419		66096
State Cleaned	B	1296	Consistent	66096
Year Cleaned	E	7344	Consistent	66096
Deaths Averaged	K	Total : 658477,5		66096