Project 2: Dataset Research Piece

The data set below gives data on the rates of COVID-19 cases and deaths over 4 weeks over the months of April 2021 to September 2022, by their vaccination status and age group.

Note: When viewing in Excel, age group labeled 12-17 appeared as 17-Dec, and age group 5-11 appears as 11-May.

Dataset: <https://data.cdc.gov/Public-Health-Surveillance/Rates-of-COVID-19-Cases-or-Deaths-by-Age-Group-and/3rge-nu2a/explore/query/SELECT%0A%20%20%60outcome%60%2C%0A%20%20%60month%60%2C%0A%20%20%60mmwr_week%60%2C%0A%20%20%60age_group%60%2C%0A%20%20%60vaccine_product%60%2C%0A%20%20%60vaccinated_with_outcome%60%2C%0A%20%20%60fully_vaccinated_population%60%2C%0A%20%20%60unvaccinated_with_outcome%60%2C%0A%20%20%60unvaccinated_population%60%2C%0A%20%20%60crude_vax_ir%60%2C%0A%20%20%60crude_unvax_ir%60%2C%0A%20%20%60crude_irr%60%2C%0A%20%20%60age_adj_vax_ir%60%2C%0A%20%20%60age_adj_unvax_ir%60%2C%0A%20%20%60age_adj_irr%60%2C%0A%20%20%60continuity_correction%60/page/filter>

Section 1.2: Exercise 1.5

Given is the relative frequency histogram of the number of cases one week in July 2022 across different age groups

1. Which age group on the horizontal axis are associated with the largest number of cases.

30-49 with 91,440 cases

1. How many cases occurred in each age group.

5-11: 5,954

12-17: 9,202

18-29: 45,267

30-49: 91,440

50-64: 74,852

65-79: 53,175

80+: 19,144

1. How many cases were there between ages 5 and 49.

5,149 + 9,202 + 45,267 + 91,440 = 151,058 cases

Section 1.3: Exercise 1.21

One week during April 2021, the CDC reported that the number of cases where a vaccinated person tested positive over the age groups of 12-17, 18-29, 30-49, 50-64, 65-79, and 80+ is approximately normally distributed with a mean of 1415.3 and a standard deviation of 953. What is the probability that there were more than 462 cases in one day in April 2021.

By Empirical Rule P(y > 462) = 0.84

µ + - σ = 68%

µ + - σ = 95%

Section 2.3: Exercise 2.8

Out of 88,310,076 people who were fully vaccinated in May 2021, it was found that 10,733,721 were in the age group of 18-29, and out of 88,804,567 unvaccinated people in May 2021, 23,355,004, were 18-29 and of those 17,423 tested positive. Find the number of people that …

1. vaccinated in May 2021, age 18-29, or both

(88,310,076 + 10,733,721) = 99,043,797

1. were unvaccinated and not 18-29 in May 2021

88,804,567 – 23,355,004 = 65,449,563

1. 18-29, unvaccinated, but did not test positive in May 2021,

23,355,004 – 17,423 = 23,337,581

Section 2.4: Exercise 2.14

A survey classified on week day in August 2022, that over a population of 41,328,354, the proportions of testing positive if you were vaccinated or not.

|  |  |  |
| --- | --- | --- |
|  | Tested Positive | |
| Vaccinated | Yes | No |
| Yes | 0.00143 | 0.99857 |
| No | 0.00397 | 0.99602 |
|  |  |  |

If a random case is selected from this survey, find the probability that the person was,

1. Testing Positive

.0054

1. Was vaccinated but did not test positive

.99857

1. Was not vaccinated whether they tested positive or not.

1

Section 2.5: Exercise 2.33

The CDC reports that over the month of January 2022, The median number of children age 12-17 over 10 days who were vaccinated and tested positive was 139,505. Suppose two weeks were randomly selected on this day.

1. List the points in the sample space.

E1 = {N,N,N,N} E4 = {P,P,P,N} E7 = {N,N,P,N} E10 = {N,P,N,P}

E2 = {P,N,N,N} E5 = {P,P,P,P} E8 = {N,N,N,P} E11 = {N,N,P,P}

E3 = {P,P,N,N} E6 = {N,P,N,N} E9 = {P,N,P,N} E12 = {N,P,P,P}

1. Identify the simple events for each of the following events.

A: At least two had greater than 139,595

E3, E4, E5, E9, E10, E11, E12

B: Exactly two had greater than 139,595

E3, E9, E10, E11

C: Exactly one had less than 139,595

E4, E12

1. Find the probabilities for each event in part b. P(A), P(B), and P(C)

P(Ei) = 1/10

P(A) = (1/10)\*7 = 7/10

P(B) = (1/10)\*4 = 4/10

P(C) = (1/10)\*2 = 2/10

Section 2.6: Exercise 2.54

One week in April 2021, 4 people ages 30-49 die from COVID who were vaccinated, and 16 people ages 50-64 died from COVID who were vaccinated. If 3 people are chosen, what is the probability that exactly 2 are ages 50-64.

C(16,2) / C(20,3) = 120/1140 = .1053

Section 2.7: Conditional Probability

If there were 42,248,013 people age 30-49 who were vaccinated, and 12,909,109 were unvaccinated, and 148,091 people tested positive, and of those 98,278 were vaccinated. Given that a person was unvaccinated, what is the probability that they tested positive.

P(A∩B) = 49,813/12,909,109

P(B) = 12,909,109/55,157,122

P(A|B) = P(A∩B)/P(B) = .01649

Section 2.8: Two Laws of Probability

Using the last problem, what is the probability that the person was did not test positive

P(B’) = 1 – P(B)

= 1 - .23404 = .76596

Section 2.9: Calculating the Probability of an Event: Event-Composition Method

It was determined that over the first week of February 2022, people age 80+ had a .00372 chance of testing positive. Over the course of the next three weeks, what is the probability that a person age 80+ will test positive all three weeks.

.00372^3 = 5.15 x 10^-8

Section 2.10: Exercise 2.124

A person is randomly selected from a population of people age 65-79 in December 2021 , 91% are vaccinated, and 8% are unvaccinated. 0.17% of vaccinated people and 1.15% of unvaccinated people tested positive. This person is found to have tested positive. Find the conditional probability that this person is unvaccinated.

P(A|B) = .0015

P(B) = .08

P(A) = .0172

P(B|A) = .0015 \* .08 / .0172 = .0069

Section 2.11: Exercise 2.142

A person can be any one of four, vaccinated and not positive, vaccinated and positive, unvaccinated and not positive or unvaccinated and positive. Two people are chosen, one after the other, and their type is noted. Let Y denote the types that the two people *are not*. Compute the probabilities of each value of Y.

P(Y=3) = ¼\*1/4\*1/4 = 1/16

P(Y=2) = ¼\*1/4 = 1/8

Section 3.2: Probability Distribution for a Discrete Random Variable

A hospital has three vaccinated and three unvaccinated patients. A doctor must select two to treat, but as to not be biased, he selects two at random and is not made aware of their vaccination status. Let Y denote the number of unvaccinated patients, find the probability distribution for Y.

p(0)=P(Y=0)= C(3,0)\*C(3,2)/C(6,2) = 1/5

p(1)=P(Y=1)=C(3,1)\*C(3,1)/C(6,2) = 3/5

p(2)=P(Y=2)=C(3,2)\*C(3,0)/C(6,2) = 1/5

Section 3.3 Expected Value or Function of a Random Variable

One week in May 2021 has positive cases amongst the different age groups respectively. 25, 719, 1925, 1711, 1479, 699. What is the expected value and variance over all the age groups over one week?

Mean = 1093 = E(Y)

Standard Deviation = 729.2, V(Y) = √729.2 = 27.004

Section 3.4 Binomial Probability Distribution

One week in July 2021, a team of doctors wanted to choose 10 people in the ages of 18-29 out of the people who were vaccinated and see if they were positive. It is known that 3,462 people tested positive out of 15,097,896 vaccinated people in that age group. What is the probability that 4 people came back positive?

P = 3,462 / 15,097,896 = .00023

Q = 1 – P = .99977

P(Y=4) = C(10,4) \* .00023^4 \* .99977^6 = 5.86 x 10^-13

Section 3.5 Geometric Distribution

From the last problem, What is the probability that the sixth person selected will be the first positive case? Probability that the first positive case is before the sixth person?

P(Y=6) = .99977^5 \* .00023 = 2.2 x 10^-4

P(Y < 6) = 1- (1 - .00023)^5 = .00115

Section 3.7 Hyper Geometric Distribution

In May 2021, There were 56 deaths due to COVID, 41 were vaccinated with the Moderna vaccine, and 15 were vaccinated with the Janssen vaccine. If a researcher was to pick 10 deaths at random, what is the probability that all of them would be vaccinated with the Janssen vaccine.

N = 56, n = 10, r = 15, y = 10

P(10) = C(15,10) \* C(41, 0) / C(56,10) = 3,003 / 3.56 x 10^10 = 8.4 x 10^-8

Section 3.8 Poisson Distribution

In May 2021, there are an average of 8 deaths in the age group 30-49 of people who were vaccinated. What is the probability that in the next three months, there will be 6 deaths in each month.

P(Y=6) = 8^6 \* e^-8 / 6! = .122

Section 3.11 Tchebysheff’s Theorem

The mean and standard deviation of cases of all the age groups of vaccinated people testing positive in one week in April 2021 is 1,283.8 and 850.9 respectively. Find the percentage of values that are between 318.8 and 2,248.8 cases.

2,248.8 – 1,283.8 = 965 1,283 – 318.8 = 965

1,283.8 + k(850.9) = 2,248.8

K = 1.13

1 – 1/1.12^2 = .305555 or 30.5%

Section 4.2 Probability Distribution for a Continuous Random Variable

The probability distribution function for the number of cases over all age groups in one week in June 2021 is f(y) = y/6 0≤y≤1. Find the distribution function F(y) over the interval (0,1/3)

F(y) = .00926

Section 4.3 Expected Values for Continuous Random Variables

Using the last problem, find E(Y).

E(Y) = = .00103

Section 4.4 Uniform Distribution

Using the same problem, what is P(0 ≤ y ≤ 1/6)

P(0 ≤ y ≤ 1/6) = (1/6 – 0)/(1 – 0) = 1/6

Story:

While studying over this data set. There are a couple of conclusions that become obvious. This dataset suggests that your probability of testing positive for COVID, under any circumstance, is extremely low. Using the geometric or binomial distributions, it would theoretically be extremely difficult to simply go out and find someone who has COVID. Additionally, the probability of contracting COVID was slightly higher (only by a couple of thousandths of a decimal) if you are unvaccinated, compared to if you were vaccinated. Both conclusions are the case within both deaths and cases. This data set could be used to argue the true effectiveness of the vaccines, or show how unlikely it is for someone to contract COVID in the first place.