Using R in Field Epidemiology

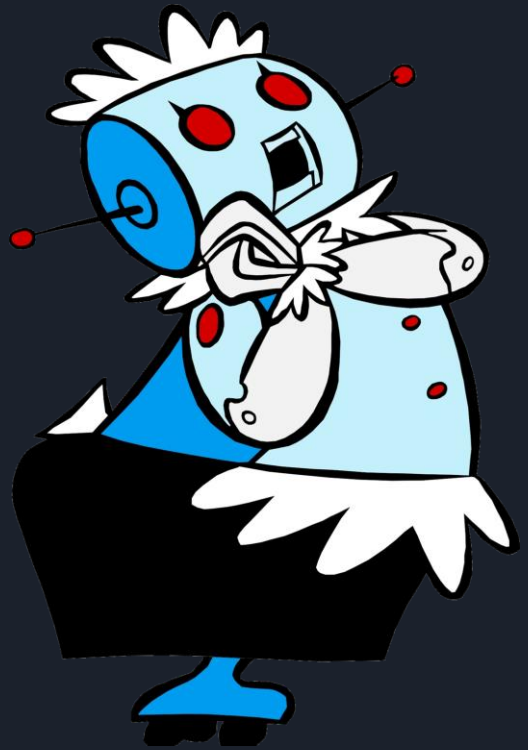
By Travis Oishi

About Me

- Central Shenandoah Health District Epidemiologist
- Jr Epi for Central Shenandoah from from Oct 2019 - March 2022
- Limited experience in R when I first started in 2019
- Started coding around Spring of 2020 to improve epi processes

Benefits of R

- Completing tasks that you need to do more than once
- Processing complex equations
- Navigating large datasets
- Large scale data manipulation
- Less prone to human error



https://hero.fandom.com/wiki/Rosie_The_Robot_Maid

2x2 Tables

- A compact summary of data for 2 variables from a study
- Used to look at the exposure and the health outcome
- Can be used in outbreak settings

Useful Calculations

- Attack Rate (Risk)
- Risk Difference
- Relative Risk (Risk Ratio)
- Odds Ratio

	DISEASE(+)	DISEASE(-)	**TOTALS**
EXPOSURE (+)	6	3	9
EXPOSURE (-)	3	4	7
TOTALS	9	7	16

https://open.oregonstate.education/epidemiology/chapter/introduction-to-2x2-tables-epidemiologic-study-design-and-measures-of-association/

Attack Rate (Risk)

- The risk of getting the disease during a specified period, such as the duration of an outbreak.
- In an outbreak setting, the term attack rate is often used as a synonym for risk
- **Overall attack rate**: Total number of new cases divided by the total population.
- **Food-specific attack rate:** Number of persons who ate a specified food and became ill divided by the total number of persons who ate that food

Food Specific Attack Rate

$$\frac{\text{\# Sick people who ate a specific food}}{\text{Total \# people who ate a specific food}} \times 100$$

https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section2.html#:~:text=In%20the%20outbreak%20setting%2C%20the,divided%20by%20the%20total%20population.

Risk Difference

- Difference in risk of between risk of exposed group and risk of unexposed group.

Risk of exposed group – Risk of not exposed group

Risk Ratio (Relative Risk)

- Compares the risk of a health event among one group with the risk among another group.
- RR = 1.0 indicates identical risk among the two groups.
- RR> 1.0 indicates an increased risk for the group in the numerator, usually the exposed group.
- RR< 1.0 indicates a decreased risk for the exposed group, indicating that perhaps exposure actually protects against disease occurrence.

$$\frac{\text{Risk of disease (incidence proportion, attack rate) in group of primary interest}}{\text{Risk of disease (incidence proportion, attack rate) in comparison group}}$$

https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section5.html

Odds Ratio

The odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure

The odds ratio can also be used to determine whether a particular exposure is a risk factor for a particular outcome, and to compare the magnitude of various risk factors for that outcome.

- OR=1 Exposure does not affect odds of outcome
- OR>1 Exposure associated with higher odds of outcome
- OR<1 Exposure associated with lower odds of outcome

a = Number of exposed cases
b = Number of exposed non-cases
c = Number of unexposed cases
d = Number of unexposed non-cases

$$\mathrm{OR} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

$$\mathrm{OR} = \frac{(\mathrm{n})\ \text{exposed cases}/(\mathrm{n})\ \text{unexposed cases}}{(\mathrm{n})\ \text{exposed non-cases}/(\mathrm{n})\ \text{unexposed non-cases}} = \frac{(\mathrm{n})\ \text{exposed cases} \times (\mathrm{n})\ \text{unexposed non-cases}}{(\mathrm{n})\ \text{exposed non-cases} \times (\mathrm{n})\ \text{unexposed cases}}$$

Summary Tables

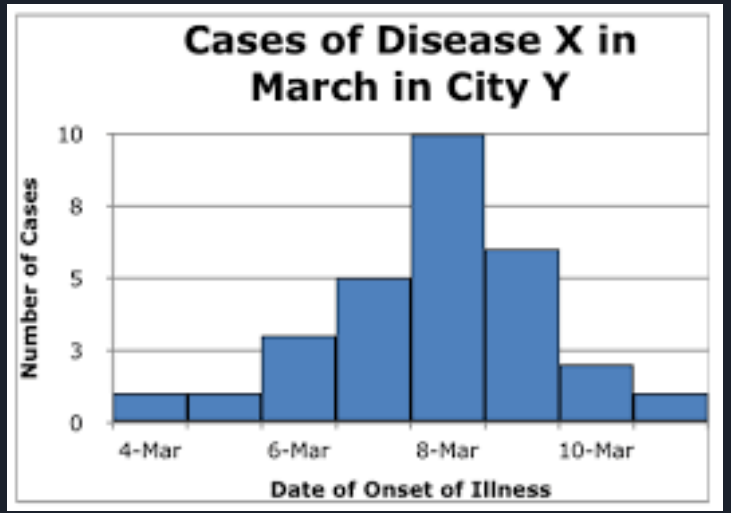
- Summarize the datasets into a digestible format
- Summarize data
- Can be used to create plots

Useful R commands

- mutate(): Create new variables / edit existing variables
- group_by(): Stratify the data by the selected columns
- summarize(): Create the calculations to summarize the grouped data

Epidemic (Epi) Curves

- Shows the progression of illness onset over time
- x axis: Date/time of onset
- y axis: Total count



Useful R commands

- ggplot(): create custom plots

http://www.pandemsim.com/beta/epicurve.html

Useful links

Stats Resources

https://www.openepi.com/TwobyTwo/TwobyTwo.htm

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938757/

https://open.oregonstate.education/epidemiology/chapter/introduction-to-2x2-tables-epidemiologic-study-design-and-measures-of-association/

R Resources

https://dplyr.tidyverse.org/

https://dplyr.tidyverse.org/reference/group_by.html

http://r-statistics.co/ggplot2-Tutorial-With-R.html