Exam I Solutions
Administered: Wednesday, September 14, 2020
20 points

For each problem part: 0 points if not attempted or no work shown,
1 point for partial credit, if work is shown,
2 points for correct numerical value of solution

All 95 counties of Tennessee have experienced some COVID-19 cases. In the Excel data file on the course website, you will find the following data for all 95 counties. A small sample of the data is shown below.

County	Cases	Deaths	Population	Death/Cases	Income	% white
Anderson	1237	10	76978	0.008084	52585	91.7
Bedford	2345	28	49713	0.01194	34946	87
Benton	1770	43	16160	0.024294	30193	94.7
...	...	...	...	...	...	...
White	2289	26	27345	0.011359	34863	95.5
Williamson	1918	13	238412	0.006778	51914	88.7
Wilson	2104	21	144657	0.009981	43314	88.6

The columns are defined as follows.

Cases	definition:	total confirmed positive cases of COVID-19 per 100,000 in county as of 09/08/20
	source:	https://www.nytimes.com/interactive/2020/us/tennessee-coronavirus-cases.html#county
Deaths	definition:	total confirmed deaths due to COVID-19 per 100,000 in county as of 09/08/20
	source:	https://www.nytimes.com/interactive/2020/us/tennessee-coronavirus-cases.html#county
Population	definition:	county population
	source:	https://www.tennessee-demographics.com/counties_by_population
Death/Case	definition:	calculated as the ratio of deaths to cases above
	source:	https://www.nytimes.com/interactive/2020/us/tennessee-coronavirus-cases.html#county
Income	definition:	median household income
	source:	https://en.wikipedia.org/wiki/List_of_Tennessee_locations_by_per_capita_income
% white	definition:	percentage of the population who report as white
	source:	https://www.indexmundi.com/facts/united-states/quick-facts/tennessee/white-population-percentage#table

Based on this data, answer the following questions.
(a) What is the mean value of the number of cases per 100,000 in a county?
(b) What is the mean value of the number of deaths per 100,000 in a county?
(c) What is the standard deviation of the number of cases per 100,000 in a county?
(d) What is the correlation coefficient between the number of cases in a county and its population? Based on this answer, do urban (high population) or rural (low population) counties experience more cases on a per capita basis?
(e) What is the correlation coefficient between the number of cases and deaths? Interpret this correlation.
(f) What is the correlation coefficient between county population and income? Interpret this correlation.
(g) What is the correlation coefficient between county population and the percent of the population that is white? Based on this correlation, are urban (high population) or rural (low population) counties more diverse?
(h) What is the correlation coefficient between deaths and the percent of the population that is white? Based on this correlation, does this support the reported observation that communities of color are more impacted by Covid-19?
(i) What is the correlation coefficient between income and the percent of the population that is white? Carefully interpret this correlation, keeping in mind the correlations calculated in parts (f) and (g).
(j) Do masks reduce the likelihood of transmitting Covid-19?

Solution:

(a) What is the mean value of the number of cases per 100,000 in a county?

$$E[c] = \frac{\sum_{i=1}^{n} c_i}{n} = 2389$$

(b) What is the mean value of the number of deaths per 100,000 in a county?

$$E[d] = \frac{\sum_{i=1}^{n} d_i}{n} = 25.0$$

(c) What is the standard deviation of the number of cases per 100,000 in a county?

$$\sigma_c^2 = E\left[c^2\right] - E[c]^2 = 9.500 \cdot 10^6 - (2389)^2 = 3.792 \cdot 10^6$$

$$\sigma_c = \sqrt{\sigma_c^2} = 1947.4$$

(d) What is the correlation coefficient between the number of cases in a county and its population? Based on this answer, do urban (high population) or rural (low population) counties experience more cases on a per capita basis?

$$\sigma_{c \cdot p} = E[c \cdot p] - E[c]E[p] = -1.261 \cdot 10^6$$

$$\sigma_p = \sqrt{\sigma_p^2} = 1.336 \cdot 10^5$$

$$\rho_{c \cdot p} = \frac{\sigma_{c \cdot p}}{\sigma_c \cdot \sigma_p} = 0.00$$

Based on this answer, there is no correlation between the cases per capita and population. So, urban (high population) or rural (low population) counties experience the same cases on a per capita basis.

(e) What is the correlation coefficient between the number of cases and deaths? Interpret this correlation.

$$\sigma_{c \cdot d} = E[c \cdot d] - E[c]E[d] = 1.462 \cdot 10^4$$

$$\sigma_d = \sqrt{\sigma_d^2} = 18.88$$

$$\rho_{c \cdot d} = \frac{\sigma_{c \cdot d}}{\sigma_c \cdot \sigma_d} = 0.40$$

Based on this answer there is a positive correlation between the number of cases and the number of deaths. This seems reasonable as the number of cases in a county increases, the number of deaths also increases.

(f) What is the correlation coefficient between county population and income? Interpret this correlation.

$$\sigma_{p \cdot i} = E[p \cdot i] - E[p]E[i] = 7.142 \cdot 10^8$$

$$\sigma_i = \sqrt{\sigma_i^2} = 8674.5$$

$$\rho_{p \cdot i} = \frac{\sigma_{p \cdot i}}{\sigma_p \cdot \sigma_i} = 0.62$$

Based on this answer there is a positive correlation between the county population and median income. This seems to accord with the intuitive understanding that cities are more expensive to live in and provide higher median income.

(g) What is the correlation coefficient between county population and the percent of the population that is white? Based on this correlation, are urban (high population) or rural (low population) counties more diverse?

$$\sigma_{p\cdot w} = E[p \cdot w] - E[p]E[w] = 5.647 \cdot 10^6$$

$$\sigma_w = \sqrt{\sigma_w^2} = 11.02$$

$$\rho_{p\cdot w} = \frac{\sigma_{p\cdot w}}{\sigma_p \cdot \sigma_w} = -0.53$$

Based on this answer there is a negative correlation between the county population and the percent of the population who identify as white. This seems to accord with the intuitive understanding that cities with high population are home to a more diverse group of people, as measure by a lower fraction of white residents, compared to rural counties with low population.

(h) What is the correlation coefficient between deaths and the percent of the population that is white? Based on this correlation, does this support the reported observation that communities of color are more impacted by Covid-19?

$$\sigma_{d\cdot w} = E[d \cdot w] - E[d]E[w] = 2.136 \cdot 10^3$$

$$\rho_{d\cdot w} = \frac{\sigma_{d\cdot w}}{\sigma_d \cdot \sigma_w} = -0.47$$

Based on this answer there is a negative correlation between Covid-19 and the percent of the population who identify as white. This data accords with the reports that Covid-19 disproportionately impacts communities of color.

(i) What is the correlation coefficient between income and the percent of the population that is white? Carefully interpret this correlation, keeping in mind the correlations calculated in parts (f) and (g).

$$\sigma_{i\cdot w} = E[i \cdot w] - E[i]E[w] = 3.458 \cdot 10^6$$

$$\rho_{i\cdot w} = \frac{\sigma_{i\cdot w}}{\sigma_i \cdot \sigma_w} = -0.19$$

Based on this answer there is a negative correlation between median income and the percent of the population who identify as white. It is important to remember the difference between correlation and causation. This does not indicate that because a greater portion of a population are white that the median income is lower. However, this correlation does exist. We know that there are many well-established studies that find a wage gap between white and black workers. Therefore, we might suppose that this correlation should be positive. However if we look at part (f), we found that, in Tennessee, income is positively correlated with population (cities are expensive). Also we saw in part (g) that the percent of people who identify as white is negatively correlated with population (cities are diverse). Therefore, it remains a possibility that the negative correlation between fraction of population that is white and income is a result of (f) and (g), cities are diverse and expensive, while rural counties are less diverse and have lower income.

(j) Do masks reduce the likelihood of transmitting Covid-19?

Yes! Always where your mask on campus!