

Statistical Thinking for Forensic Practitioners

Excel Lab on Part 2: The Language of Probability

We will practice calculating various types of probabilities discussed in lecture using Excel. You can perform the calculations in Excel and keep track of your answers on this sheet (or simply type in the spreadsheet - this is for your own practice). Note that your performance on this lab will not affect whether you pass the course, but is meant to provide you with practice. Solutions keys are provided on the course website as well.

Download the file **deathrow.xlsx** from the course website and open this file in Excel (your computer should automatically open it in Excel if you have Excel installed). The table shown in the top left is called a *2-way table*. Each cell represents the number of individuals on death row whose race corresponds to the column label (White, Black, or Latino) and whose alleged victim's race corresponds to the row label (also White, Black, or Latino). These data were collected in 1982/83.

1 Probability calculations in Excel

We will practice using Excel to calculate probabilities. A *formula* is used to perform a calculation in Excel. Refer to [this webpage](#) for an introduction to creating formulas in Excel. If you have never used formulas in Excel, it is recommend that you first complete this introduction.

Note that some cells contain a number followed by 3 question marks (e.g., 1. ???). These correspond to questions given below. You can replace the contents of these cell by clicking on the cell and typing. Also, take note of the formula in cell G8 that was used to calculate the probability that a victim was white. In particular, the Excel function 'SUM' was used to avoid individually add up each of these cells. We will be using such functions often in this course.

1. What is the probability calculated in cell G9? Use similar mathematical notation as shown in cell F8 (e.g., $\Pr(\text{Event})$).

$\Pr(\text{Black Victim})$

2. Under the constraint that the 3 probabilities in G8, G9, and G10 must sum to 1, what is the probability that needs to be calculated in cell G10? Use similar mathematical notation as you used above.

$\Pr(\text{Latino Victim})$

3. Calculate the missing probability in cell G10. Note that you can either use a similar formula as those used in cells G8 or G9 or use the property of complementary events. (Note: refer to slides 10 in the Part 2 lecture slides)

Approximately .064

4. All of these probabilities are associated with the probability of a victim being a particular race. Are these marginal, conditional, or joint probabilities? (Note: refer to slides 27 in the Part 2 lecture slides)

Marginal

5. Calculate the odds that a victim is Latino.

$$O_{\text{L. Victim}} = \frac{\Pr(\text{Latino Victim})}{1 - \Pr(\text{Latino Victim})} = .068 \text{ (rounding to 3 decimal places)}$$

6. Calculate the following conditional probabilities:

- (a) $\Pr(\text{White Defendant} \mid \text{Black Victim})$

Approximately .095

- (b) $\Pr(\text{Black Defendant} \mid \text{Black Victim})$

Approximately .886

- (c) $\Pr(\text{Latino Defendant} \mid \text{Black Victim})$

Approximately .019

7. Suppose a prosecutor considers the conditional probabilities calculated in question 6 and makes the claim: “Based on these probabilities it is clear that when a defendant on death row is black, it is extremely likely that their victim is also black.” Comment on the validity of this statement. You may do so by either explaining why the claim is true or by calculating another probability that contradicts the claim.

The claim made by the prosecutor is associated with $\Pr(\text{Black Victim} \mid \text{Black Defendant})$ which is not equal to any of the probabilities calculated above. Rather, $\Pr(\text{Black Victim} \mid \text{Black Defendant}) = .343$ (rounding to 3 decimal places). This is not nearly as large as $\Pr(\text{Black Defendant} \mid \text{Black Victim})$, making the prosecutor’s claim unfounded based on the data. (Note: this is an example of the Prosecutor’s Fallacy.)

8. Consider the formula in cell G19.

- (a) What is the probability that is represented by the calculation in cell G19? Use similar mathematical notation as in question 1 (e.g., $\Pr(\text{Event})$).

$\Pr(\text{Black Defendant and Black Victim})$

- (b) Excel did not calculate the formula shown in cell G19 because the cell was formatted to contain only “text”. Change the formatting of the cell by right clicking on the cell, selecting “Format Cells” from the menu, and changing the Category of the cell to “Number”. After changing the format of the cell, re-enter the formula ($=G9*G16$) to see that Excel now actually calculates the probability.

Approximately .119

9. Calculate $\Pr(\text{Black Defendant})$.

Approximately .348

10. Are the events “Black Defendant” and “Black Victim” independent? (Hint: your previous answers can be used to solve this question)

For events A and B to be independent, it must be true that $\Pr(A \text{ and } B) = P(A)P(B)$. From question 8(b), we know that $\Pr(\text{Black Defendant and Black Victim}) = .119$ approximately. From questions 1 and 8, we know that $\Pr(\text{Black Defendant})\Pr(\text{Black Victim}) = (.135)(.348) = .047$ approximately. Since these two quantities are not equal, the events “Black Defendant” and “Black Victim” are not independent.

11. Consider the probabilities calculated in cells G23, G24, and G25.

- (a) Calculate $\Pr(\text{Black Victim} \mid \text{White Defendant})$.

Approximately .022

- (b) By re-formatting cell G25 to a "Number" as you did in question 8(b), calculate the probability represented by the formula in cell G25. Based on the resulting value (and compared to the result in cell G24), what is the probability calculated in cell G25? Use similar mathematical notation as was used in problem 1 (e.g., $\Pr(\text{Event})$).

It is also $\Pr(\text{Black Victim} \mid \text{White Defendant})$.

- (c) Name the theorem that justifies your findings in part (b).

Bayes' Theorem states that

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}.$$