

Conditional Probability and Independence

Preparation Tasks

1. Before beginning Homework 3 you should have completed the following OLI modules:

Module 8: Introduction to Probability

Module 9: Finding the Probability of Events

Module 10: Conditional Probability

2. From Canvas, download the following datafile:

- auto_premiums.RDATA

RMarkdown Instructions

1. Create a setup code chunk just below the header that loads the auto_premiums.RDATA file.

```
```{r setup, message = FALSE}
load("auto_premiums.RDATA")
```
```

2. Show probability calculations for the questions below in R code chunks.

Homework Questions

1. In a sample of 50 people, 21 had type O blood, 22 had type A blood, 5 had type B blood and, 2 had type AB blood. Find the probability that,
 - a. A person has type O blood.
 - b. A person has type A or type B blood.
 - c. A person has neither type A nor type O blood.
 - d. A person does not have type AB blood.
2. A group of visitors to Ashesi University consisted of 15 students from Oxford University and 20 students from Arizona State University. Among the students from Oxford were 8 females and among the Arizona State students were 5 males. A student was selected at random to give a vote of thanks at the end of the visit.
 - a. What is the probability that the student is a female?
 - b. Suppose the student selected is a female, what is the probability the student comes from Arizona State University?
3. A box contains 24 surgical screws, 4 of which are defective. If 4 are sold at random, find the probability that,
 - a. Exactly 2 are defective.
 - b. None is defective.
 - c. All are defective.

d. At least 1 is defective.

4. Hospital records indicated that knee replacement patients stayed in the hospital for the number of days shown in the distribution.

| Number
of days
stayed | Frequency |
|-----------------------------|-----------|
| 3 | 15 |
| 4 | 32 |
| 5 | 56 |
| 6 | 19 |
| 7 | 5 |
| | <hr/> 127 |

Find the probability that,

- A patient stayed exactly 5 days.
 - A patient stayed less than 6 days.
 - A patient stayed at most 4 days.
 - A patient stayed at least 5 days.
5. Each year, ratings are compiled concerning the performance of new cars during the first 90 days of use. Suppose that the cars have been categorized according to whether the car needs warranty-related repair (yes or no) and the country in which the company manufacturing the car is based (United States or not United States). Based on the data collected, the probability that the new car needs a warranty repair is 0.04, the probability that the car is manufactured by a U.S.-based company is 0.60, and the probability that the new car needs warranty repair and was manufactured by a U.S.-based company is 0.025.
- Suppose you know that a company based in the United States manufactured a particular car. What is the probability that the car needs warranty repair?
 - Suppose you know that a company based in the United States did not manufacture a particular car. What is the probability that the car needs warranty repair?
 - Are need for warranty repair and location of the company manufacturing the car statistically independent?
- 6.
- In a club, there are 7 women and 5 men. A committee of 3 women and 2 men is to be chosen. How many different possibilities are there?
 - The manager of a department store chain wishes to make four-digit identification cards for her employees. How many different cards can be made if she uses the digits 1, 2, 3, 4, 5, 6 and repetitions are permitted?

7. Look at the auto premiums data frame in the Environment tab to identify the variable names and types. The data shows the number of years of driving experience, gender, and insurance premium (monthly cost of automobile insurance in dollars) for 50 randomly selected US adults.
- Create a boxplot of the data\$Premium.
 - Is the distribution of auto insurance premiums symmetric or skew? Do you expect the data to follow the empirical rule (standard deviation rule)? Explain.
 - Use values for the mean and standard deviation to describe the central tendency and variability of the sample distribution of auto premiums.
 - Determine the probability that an insurance premium chosen at random is within one standard deviation of the mean. Is this close to the value you would have expected based on the empirical rule?
 - Determine the probability that an insurance premium chosen at random is within two standard deviations of the mean. Is this close to the value you would have expected based on the empirical rule?