Conditional Probability

2024-09-24

Math 2265 Section 3.2

- Work as a group!
- You will need to replace <> in the source code or answer questions.
- Update your name in L5.

Load Packages

```
## Loading required package: airports
```

```
## Loading required package: cherryblossom
```

```
## Loading required package: usdata
```

We will use the `smallpox` data set. From text book p. 99,

> The `smallpox` data set provides a sample of 6,224 individuals from the year 1721 who were exposed to smallpox in Boston. Doctors at the time believed that inoculation, which involves exposing a person to the disease in a controlled form, could reduce the likelihood of death.   Each case represents one person with two variables: `inoculated` and `result`. The variable `inoculated` takes two levels: `yes` or `no`, indicating whether the person was inoculated or not. The variable `result` has outcomes `lived` or `died`.

First, use `str` to glance through the data set `smallpox`

```
str(<>)
```

```
## Error: <text>:1:5: unexpected '<'
## 1: str(<
##         ^
```

Question 1. What are the names of the variables?

Answer:

Question 2. Use `table` to tally each variable.

For the first variable:

```
table(smallpox$<>)
```

```
## Error: <text>:1:16: unexpected '<'
## 1: table(smallpox$<
##                    ^
```

For the second variable:

```
table(smallpox$<>)
```

```
## Error: <text>:1:16: unexpected '<'
## 1: table(smallpox$<
##                    ^
```

Question 3.

(a) Make a contingency table (two-way table) of the variables `result` and `inoculated` in the data set `smallpox` and save it to a variable `t`.

```
t <- table(smallpox$<>, smallpox$<>)
t
```

```
## Error: <text>:1:21: unexpected '<'
## 1: t <- table(smallpox$<
##                         ^
```

(b) Make a proportional table.

```
options(digits = 1)
prop.table(<>)
```

```
## Error: <text>:2:12: unexpected '<'
## 1: options(digits = 1)
## 2: prop.table(<
##               ^
```

Summary of code

Compared to the Tables 3.15 and 3.16 in the textbook,

- we miss the totals and
- the order of categories are reversed

We can add `addmargins` for the former, but the latter is merely technical (which we won't put much emphasis). The following code will display the table in the same format.

```
smallpox$result <- factor(smallpox$result, levels = c("lived", "died"))
smallpox$inoculated <- factor(smallpox$inoculated, levels = c("yes", "no"))

my_table <- table(smallpox$result, smallpox$inoculated)

print(addmargins(my_table))
```

```
##
##          yes   no  Sum
##   lived  238 5136 5374
##   died     6  844  850
##   Sum    244 5980 6224
```

```
options(digits = 2)
print(addmargins(my_table)/sum(my_table))
```

```
##
##             yes      no     Sum
##   lived 0.03824 0.82519 0.86343
##   died  0.00096 0.13560 0.13657
##   Sum   0.03920 0.96080 1.00000
```

Example/Suggested answer format:

What is the probability that a randomly selected person lived, given that they were inoculated?

$$P(\text{lived} \mid \text{yes}) = \frac{238}{244}$$

```
238/244
```

```
## [1] 0.98
```

Question 4

---

1. **What is the probability that a randomly selected person was inoculated?**

$$P(<>) = \frac{<>}{<>} =$$

---

2. **What is the probability that a person lived and was inoculated?**

$$P(<> \cap <>) = \frac{<>}{<>} =$$

---

3. **What is the probability that a person was inoculated, given that they lived?**

$$P(<> \mid <>) = \frac{<>}{<>} =$$

---

4. **If a randomly selected person died, what is the probability that they were inoculated?**

$$P(<> \mid <>) = \frac{<>}{<>} =$$

Upload your work to Canvas