Lab1: A Revied of Probability Theory

WCM Modern Methods for Causal Inference

Tianran Zhang

Simpson's paradox

1. Simpson's Paradox is a phenomenon in which a trend appears in several different groups of data but disappears or reverses when these groups are combined.
2. An example of Simpson's Paradox:

	Group A	**Group B**
V1	10/40 = 25%	2/10 = 20%
V2	20/40 = 50%	4/10 = 40%
Total	30/80 = 37.5%	6/20 = 60%

Probability

3. Estimate the following marginal and conditional probabilities from the data:

a. P(infection within 7 days) = (58 + 101 + 76 + 68)/(405 + 150 + 308 + 81 + 58 + 101 + 76 + 68) = 0.243

b. P(infection within 7 days OR Lung Transplant) = (58 + 101 + 76 + 68 + 150)/(405 + 150 + 308 + 81 + 58 + 101 + 76 + 68) = 0.363

c. P(Infection within 7 days | Lung Transplant) = 101/(150 + 101) = 0.42

d. P(Lung Transplant | Infection within 7 days) = 101/(58 + 101 + 76 + 68) = 0.33

Properties of Random Variables

4. An example of a random variable and a value.

- In words: the outcome of rolling a dice is a random variable
- In mathematical notation: X = 1, 2, ..., 6.

5. If a random variable A is independent of B, then $P(A)P(B) = P(A \wedge B)$. i.e., $A \perp\!\!\!\perp B$
6. Suppose random variable A is conditionally independent of B given C, we can write: $A \perp\!\!\!\perp B|C$

Bayes Theorem

7. Bayes Theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

Expectation

8. $E[X] = \sum_{i=1}^{n} X_i P(X = X_i)$

   In this problem, $E[x] = 0 * 0.9 + 1 * 0.07 + 2 * 0.02 + 3 * 0.01 = 0.14$

Variance and covariance

9. $Var(X) = E[(X - E[X])^2]$

   $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$

   $Corr(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

10. Compute the following quantities.

    a.

$$P(x) = \begin{cases} 3/4 & \text{if } x = 1 \\ 1/4 & \text{if } x = 0 \end{cases}$$

$$P(y) = \begin{cases} 1/2 & \text{if } y = 1 \\ 1/2 & \text{if } y = 0 \end{cases}$$

$$P(x, y) = \begin{cases} 1/4 & \text{if x = 1, y = 1} \\ 1/2 & \text{if x = 1, y = 0} \\ 1/4 & \text{if x = 0, y = 1} \\ 0 & \text{if x = 0, y = 0} \end{cases}$$

$$P(y|x) = \begin{cases} 1/3 & \text{if x = 1, y = 1} \\ 2/3 & \text{if x = 1, y = 0} \\ 1 & \text{if x = 0, y = 1} \\ 0 & \text{if x = 0, y = 0} \end{cases}$$

$$P(x|y) = \begin{cases} 1/2 & \text{if x = 1, y = 1} \\ 1/2 & \text{if x = 1, y = 0} \\ 1/2 & \text{if x = 0, y = 1} \\ 1/2 & \text{if x = 0, y = 0} \end{cases}$$

b.

$E[X] = 3/4$

$E[Y] = 1/2$

$$E(Y|X = x) = \begin{cases} 1/3 & \text{if x = 1} \\ 1 & \text{if x = 0} \end{cases}$$

$$E(X|Y = y) = \begin{cases} 1/2 & \text{if y = 1} \\ 1/2 & \text{if y = 0} \end{cases}$$

$Var(X) = E[X^2] - E[X]^2 = 3/16$

$Var(Y) = 1/4$

$Cov(X, Y) = E[XY] - E[X]E[Y] = 1/4 - 3/8 = -1/8$

$\rho_{XY} = -1/\sqrt{3}$

c.

$$P(x|y=1) = \begin{cases} 1/2 & \text{if } x=1 \\ 1/2 & \text{if } x=0 \end{cases}$$

So, the two gusses are equal.

d.

$$P(y|x=1) = \begin{cases} 1/2 & \text{if } y=1 \\ 1/2 & \text{if } y=0 \end{cases}$$

The two gusses are equal

e.

No. Set x = 1, y = 0, $P(X=1, Y=0) = 2/3$, while $P(X=1)P(Y=0) = 3/8$. So, X and Y are not mutually independent.

Regression

11. Estimate the following marginal and conditional expectations using the 32 cars in the `mtcars` data set.

a.

```
> mean(mtcars$hp)
[1] 146.6875
```

b.

```
> mtcars %>%
+ group_by(cyl) %>%
+ summarise(exp_hp = mean(hp))
# A tibble: 3 x 2
    cyl exp_hp
  <dbl>  <dbl>
1     4   82.64
2     6   122.29
3     8   209.21
```

c.

```
> lm(hp ~ 1, data = mtcars)

Call:
lm(formula = hp ~ 1, data = mtcars)
```

```
Coefficients:
(Intercept)
      146.7

> lm(hp ~ factor(cyl) - 1, data = mtcars)

Call:
lm(formula = hp ~ factor(cyl) - 1, data = mtcars)

Coefficients:
factor(cyl)4  factor(cyl)6  factor(cyl)8
       82.64        122.29        209.21
```

$E[H] = 146.7$

$E[H|C] = 82.64I(cyl = 4) + 122.29I(cyl = 6) + 209.21I(cyl = 8)$

These results are the same with mine.

d.

```
> lm(hp ~ factor(cyl) + mpg, data = mtcars)

Call:
lm(formula = hp ~ factor(cyl) + mpg, data = mtcars)

Coefficients:
 (Intercept)  factor(cyl)6  factor(cyl)8           mpg
     171.349        16.623        88.105        -3.327
```

So, $E[H|C, M] = 171.35 + 16.623I(cyl = 6) + 88.105I(cyl = 8) - 3.327 * mpg$