3.1 Determining association between Categorical variables

Monday, 10 October 2022 11:39

Summary

- Contingency table
- Joint, marginal and conditional probabilities

- When we infer, it's about the sample data or population?

Determining and inferring association

- Determining the association: Given a **sample data**, how do you determine that the variables are associated?
- Inferring the association: We infer about the **population.** Once you have determined the association, how can you extend that to the population?

- What do we calculate to look if there was a gender discrimination in the admission process?

Example

- Consider a B-school which shortlisted 1200 candidates (960 men and 240 women) for its post-graduate management program. Out of these, 324 candidates were given offer letters for admission. The data is included here:

	Male	Female	Total
Offers made	288	36	324
Not offered	672	204	876
Total	960	240	1200

Contingency table

- After reviewing the record, a women's forum raised the issue of gender discrimination on the basis that 288 male candidates were offered admission against only 36 female candidates.

- 1200 is not the population. It's a sample from the entire list of candidates. It's a sample data.

Example

- To this, the B-school management replied that it was not a case of discrimination, but was because of the fact that only 240 women candidates appeared for the examination.

Let us review the case using probability.

- Let *M* be the event that the candidate is a male.
- Let *F* be the event that the candidate is a female.
- Let *A* be the event that the candidate is offered admission.
- Let A^c be the event that the candidate is not offered admission.

(Event A^c is called compliment of event *A*.

One can see that $Pr(A) + Pr(A^c) = 1$.)

Example

- Probability that a randomly observed candidate is a male and is offered the admission.

$Pr(M \cap A) = 288/1200 = 0.24$

- Probability that randomly observed candidate is a male and is not offered the admission. $Pr(M \cap A^c) = 672/1200 = 0.56$
- Similarly,

$Pr(W \cap A) = 36/1200 = 0.03$

$Pr(W \cap A^c) = 204/1200 = 0.17$

- W(woman) = F (female) (It's a typo here)

Example

- In terms of probabilities, the previous table can now be rewritten as:

	Male	Female	Total
Offers made	0.24	0.03	0.27
Not offered	0.56	0.17	0.73
Total	0.8	0.2	1.0

- Joint probabilities (of what?) appear in the main body of the table (e.g. 0.24, 0.03).
- Marginal probabilities (of what?) appear in the margin of the table (e.g. 0.8, 0.2).

Example

- What will be $Pr(A|M)$?
- First of all, what does this mean?
- This conditional probability tells us that we are concerned with admission status of only males!
- We know that out of the 960 male candidates, 288 were offered admission. So probability that a male candidate is offered admission will be 288/960 = 0.3.
- Also observe that:

	$$\Pr(A \mid M) = \frac{288}{960} = \frac{288/1200}{960/1200} = \frac{0.24}{0.8} = 0.3.$$
• Relation between conditional, joint and marginal probabilities?	**Example** • Now, the numerator, *0.24* is the joint probability of events *A* and *M*. that is *Pr(A∩M) = 0.24*. And 0.8 is the marginal probability of the event *M*, i.e. *Pr(M) = 0.8*. $$\Pr(A \mid M) = \frac{\Pr(A \cap M)}{\Pr(M)}.$$ • This is, precisely, the definition of conditional probability. • Back to the problem at hand: $$\Pr(A \mid W) = \frac{\Pr(A \cap W)}{\Pr(W)} = \frac{0.03}{0.2} = 0.15.$$ • $\textbf{Conditional Probability} = \frac{\textbf{Joint Probability}}{\textbf{Marginal Probability}}$
$P(A \mid M) = 0.3$ $P(A \mid F) = 0.15$ • From these probabilities, what can you say about the discrimination?	**Example: Conclusion** • The probability of admission offer given the candidate is a male is 0.3, twice of 0.15 probability of admission offer given the candidate is a woman. • Although the use of conditional probability does not, in itself, prove discrimination, there is support for the argument!