Probability and Two(+1) Way Tables

Berkeley gender discrimination case

| | Admit | Deny | Total |
|---|---|---|---|
| Men | 3,738 | 4,704 | 8,442 |
| Women | 1,494 | 2,827 | 4,321 |
| Total | 5,232 | 7,531 | 12,763 |

**Joint Probability**

The prob of two specific outcomes of the two variables considered

P(Var 1 and Var 2)

$$\text{Joint Prob} = \frac{\text{Value from middle}}{\text{The grand total}}$$

$$P(\text{Man and Admitted}) = \frac{3{,}738}{12{,}763} = 29.3\%$$

Joint prob distribution

| | Admit | Den |
|---|---|---|
| Men | $\frac{3{,}738}{12{,}763} = 29.3\%$ | $\frac{4{,}704}{12{,}763} = 36.9\%$ |
| Women | $\frac{1{,}494}{12{,}763} = 11.7\%$ | $\frac{2{,}827}{12{,}763} = 22.1\%$ |

**Marginal Probability**

The prob of a single variable taking a specific value while ignoring the other variable

$$\text{Marginal Prob} = \frac{\text{Value from the margins}}{\text{The grand total}}$$

Marginal Prob distributions

Gender

$$P(\text{Men}) = \frac{8{,}442}{12{,}763} = 66.1\%$$

$$P(\text{Women}) = \frac{4{,}321}{12{,}763} = 33.9\%$$

Acceptance

$$P(\text{Accepted}) = \frac{5{,}232}{12{,}763} = 41.0\%$$

$$P(\text{Deny}) = 1 - P(\text{Accepted}) = 59.0\%$$
$$= \frac{7{,}531}{12{,}763} = 59.0\%$$

**Conditional Probability**

Conditioned on a specific value of one variable, what is the probability of the other variable

$$\text{Conditional Prob} = \frac{\text{Value from the center}}{\text{Value from the margins}}$$

Conditional Prob Distributions

$$P(\text{Accept} \mid \text{Man}) = \frac{3{,}738}{8{,}442} = 44.3\%$$

$$P(\text{Deny} \mid \text{Men}) = 1 - P(A \mid M) = 55.7\% = \frac{4{,}704}{8{,}442}$$

$$P(\text{Accept} \mid \text{Women}) = \frac{1{,}494}{4{,}321} = 34.6\%$$

$$P(\text{Deny} \mid \text{Women}) = 65.4\%$$

The trends that we observe in any data can change or at least become

The trends that we observe in any data can change or at least become more neuanced when we consider other variables.

3-way table: Gender, Admittance, Department

| | | Men | | Women | |
|---|---|---|---|---|---|
| | | Accept | Deny | Accept | Deny |
| Department | A | 511 | 314 | 86 | 22 |
| | B | 353 | 207 | 17 | 8 |
| | C | 120 | 205 | 202 | 391 |

Conditional Prob

Condition on department and Gender

$P(\text{Accept} \mid A, \text{Men}) = \frac{511}{825} = 62\%$ ☆

$P(\text{Accept} \mid A, \text{Women}) = \frac{86}{108} = 82\%$ ☆

$P(\text{Accept} \mid B, M) = \frac{353}{560} = 63\%$

$P(A \mid B, W) = \frac{17}{25} = 68\%$ ☆

$P(\text{Accept} \mid C, M) = \frac{120}{325} = 37\%$ ☆

$P(A \mid C, W) = \frac{202}{593} = 34\%$

Simpson's Paradox

The relationship between acceptance and gender reverses when we consider a third variable.

Example: Survival rates for ambulatory helicopters are lower than rates for ground based ambulances.

If you consider severity of injury helicopters are better.