

Date.....

12/05/2020

LESSON 3: Admissions case study

University of California Berkeley

→ wanted to know whether admission procedure is gender biased?

→ Male

	Applied	Admitted	Rate
Major A	900	480	53%
Major B	100	10	10%
Both	1000	460	46%

Female

	Applied	Admitted	Rate
Major A	100	80	80%
Major B	900	180	20%
Both	1000	260	26%

Is there a gender bias?

• Yes • No

who is being favoured?

Male Female

above

→ As we can see, the statistics for single major looks like females are more preferred but if we look at total statistics we found that males are more preferred.

↓

This is known as Simpson's paradox.

Case-Study

1) loaded admission- data.csv

2) Proportion & admission rate for each gender.

Proportion of students that are female

```
(len(df[df['gender'] == 'female']) / df.shape[0])
```

Proportion of students that are male

```
(len(df[df['gender'] == 'male']) / df.shape[0])
```

Admission rate for females

```
(len(df[(df['gender'] == 'female') & (df['admitted'] == 1)]) / len(df[df['gender'] == 'female']))
```

Admission rate for males

```
(len(df[(df['gender'] == 'male') & (df['admitted'] == 1)]) / len(df[df['gender'] == 'male']))
```

3) Proportion & admission rate for physics majors of each gender

what proportion of female students are majoring in physics?

```
fem_phy_rate = df.query("gender == 'female' & major == 'Physics').
```

```
count() / df.query("gender == 'female'").count()
```

```
print(fem_phy_rate)
```

what proportion of male students are majoring in physics?

```
(male_phy_rate = df.query("gender == 'male' & major == 'Physics').
```

```
count() / df.query("gender == 'male'").count())
```

```
print(male_phy_rate)
```


4) Proportion and admission

Admission rate for female physics major

$$\frac{\text{len(df[(df['gender'] == 'female') \& (df['major'] == 'Physics') \& df['admitted']])}}{\text{len(df[(df['gender'] == 'female') \& (df['major'] == 'Physics')])}}$$

Admission rate for male physics major

$$\frac{\text{len(df[(df['gender'] == 'male') \& (df['major'] == 'Physics') \& df['admitted']])}}{\text{len(df[(df['gender'] == 'male') \& (df['major'] == 'Physics')])}}$$

4) Proportion & admission rate for chemistry majors of each gender.

→ In this we have to perform all 4 methods which we have performed in previous step i.e., Step 3.

5) Admission rate for each major

Admission rate for physics majors

$$\frac{\text{len(df[(df['major'] == 'Physics') \& df['admitted']])}}{\text{len(df[(df['major'] == 'Physics')])}}$$

Admission rate for chemistry majors

$$\frac{\text{len(df[(df['major'] == 'Chemistry') \& df['admitted']])}}{\text{len(df[(df['major'] == 'Chemistry')])}}$$