Simpson's Paradox ¶

Use admission_data.csv for this exercise.

In [2]:
```
# Load and view first few lines of dataset
import pandas as pd
admits = pd.read_csv('admission_data.csv')
admits.head()
```

Out[2]:

	student_id	gender	major	admitted
0	35377	female	Chemistry	False
1	56105	male	Physics	True
2	31441	female	Chemistry	False
3	51765	male	Physics	True
4	53714	female	Physics	True

Proportion and admission rate for each gender

In [12]:
```
print (len(admits[admits['gender']=='female']))
print (admits.shape[0])
```

```
257
500
```

In [13]:
```
# Proportion of students that are female
(len(admits[admits['gender']=='female']))/admits.shape[0]
```

Out[13]: 0.514

In [14]:
```
# Proportion of students that are male
(len(admits[admits['gender']=='male']))/admits.shape[0]
```

Out[14]: 0.486

In [15]:
```
# Admission rate for females
len(admits[(admits['gender']=='female') & (admits['admitted'])])/(len(admits[admits['gender']=='female']))
```

Out[15]: 0.28793774319066145

In [16]:
```
# Admission rate for males
len(admits[(admits['gender']=='male') & (admits['admitted'])])/(len(admits[admits['gender']=='male']))
```

Out[16]: 0.48559670781893005

Proportion and admission rate for physics majors of each gender

In [17]:
```
# What proportion of female students are majoring in physics?
fem_phys_rate = admits.query("gender == 'female' & major == 'Physics'").count()/ \
    (admits.query("gender == 'female'").count())
print (fem_phys_rate)
```

```
student_id    0.120623
gender        0.120623
major         0.120623
admitted      0.120623
dtype: float64
```

In [18]:
```
# What proportion of male students are majoring in physics?
fem_phys_rate = admits.query("gender == 'male' & major == 'Physics'").count()/ \
    (admits.query("gender == 'male'").count())
print (fem_phys_rate)
```

```
student_id    0.925926
gender        0.925926
major         0.925926
admitted      0.925926
dtype: float64
```

In [26]:
```
# Admission rate for female physics majors
len(admits[(admits["gender"]=='female') & (admits["major"] == 'Physics') & admits["admitted"]]) / len(admits[(admit
s["gender"]=='female') & (admits["major"] == 'Physics')])
```

Out[26]: 0.7419354838709677

In [27]:
```
# Admission rate for male physics majors
len(admits[(admits["gender"]=='male') & (admits["major"] == 'Physics') & admits["admitted"]]) / len(admits[(admits
["gender"]=='male') & (admits["major"] == 'Physics')])
```

Out[27]: 0.5155555555555555

Proportion and admission rate for chemistry majors of each gender

In [28]:
```
# What proportion of female students are majoring in chemistry?
len(admits[(admits['gender']=='female') & (admits['major'] == 'Chemistry')]) / len(admits[admits['gender']=='femal
e'])
```

Out[28]: 0.8793774319066148

In [29]:
```
# What proportion of male students are majoring in chemistry?
len(admits[(admits['gender']=='male') & (admits['major'] == 'Chemistry')]) / len(admits[admits['gender']=='male'])
```

Out[29]: 0.07407407407407407

In [30]:
```
# Admission rate for female chemistry majors
len(admits[(admits['gender']=='female') & (admits['major'] == 'Chemistry') & admits['admitted']]) / len(admits[(adm
its['gender']=='female') & (admits['major'] == 'Chemistry')])
```

Out[30]: 0.22566371681415928

In [31]:
```
# Admission rate for male chemistry majors
len(admits[(admits['gender']=='male') & (admits['major'] == 'Chemistry') & admits['admitted']]) / len(admits[(admit
s['gender']=='male') & (admits['major'] == 'Chemistry')])
```

Out[31]: 0.1111111111111111

Admission rate for each major

In [32]:
```
# Admission rate for physics majors
len(admits[(admits['major'] == 'Physics') & admits['admitted']]) / len(admits[(admits['major'] == 'Physics')])
```

Out[32]: 0.54296875

In [33]:
```
# Admission rate for chemistry majors

len(admits[(admits['major'] == 'Chemistry') & admits['admitted']]) / len(admits[(admits['major'] == 'Chemistry')])
```

Out[33]: 0.21721311475409835