# Inferential Analysis

```
from google.colab import drive
drive.mount('/content/drive')
```

```
import pandas as pd
import seaborn as sns
```

## Simpson's Paradox

```
data = pd.read_csv("/content/drive/MyDrive/[Lecture]/빅데이터개론/BigData_Python/09_Inferentia
```

```
data.head()
```

| | Admit | Gender | Dept | n |
|---|---|---|---|---|
| **0** | Admitted | Male | A | 512 |
| **1** | Rejected | Male | A | 313 |
| **2** | Admitted | Female | A | 89 |
| **3** | Rejected | Female | A | 19 |
| **4** | Admitted | Male | B | 353 |

```
data_sum = data.groupby(['Gender','Admit']).sum().reset_index()
data_sum
```

| | Gender | Admit | n |
|---|---|---|---|
| **0** | Female | Admitted | 557 |
| **1** | Female | Rejected | 1278 |
| **2** | Male | Admitted | 1198 |
| **3** | Male | Rejected | 1493 |

```
#data_sum_wide = data_sum.pivot_table(index=['Gender'],columns='Admit',values='n').reset_index(
data_sum_wide = data_sum.pivot(index=['Gender'],columns='Admit',values='n').reset_index()
#data_sum_wide = data_sum_wide[['Gender','Admitted','Rejected']]
data_sum_wide
```

| Admit | Gender | Admitted | Rejected |
|---|---|---|---|
| **0** | Female | 557 | 1278 |
| **1** | Male | 1198 | 1493 |

```
data_sum_wide['total']=data_sum_wide['Admitted']+data_sum_wide['Rejected']
data_sum_wide
```

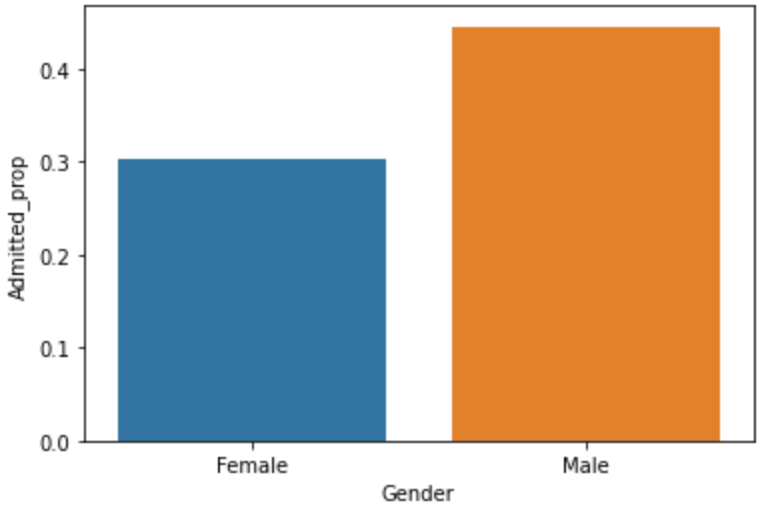
| Admit | Gender | Admitted | Rejected | total |
|---|---|---|---|---|
| **0** | Female | 557 | 1278 | 1835 |
| **1** | Male | 1198 | 1493 | 2691 |

```
data_sum_wide['Admitted_prop'] = data_sum_wide['Admitted'] / data_sum_wide['total']
data_sum_wide
```

| | Admit | Gender | Admitted | Rejected | total | Admitted_prop |
|---|---|---|---|---|---|---|
| **0** | Female | | 557 | 1278 | 1835 | 0.303542 |

```
sns.barplot(x='Gender',y='Admitted_prop', data=data_sum_wide)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f81f69f30d0>
```




```
data.head()
```

| | Admit | Gender | Dept | n |
|---|---|---|---|---|
| **0** | Admitted | Male | A | 512 |
| **1** | Rejected | Male | A | 313 |
| **2** | Admitted | Female | A | 89 |
| **3** | Rejected | Female | A | 19 |
| **4** | Admitted | Male | B | 353 |

```
#data_dept_wide = data.pivot(index=['Gender','Dept'],columns='Admit',values='n').reset_index()
data_dept_wide = data.pivot_table(index=['Gender','Dept'],
                                  columns='Admit',values='n').reset_index()
data_dept_wide
```

```
data_dept_wide['total']=data_dept_wide['Admitted']+data_dept_wide['Rejected']
data_dept_wide
```

| Admit | Gender | Dept | Admitted | Rejected | total |
|---|---|---|---|---|---|
| **0** | Female | A | 89 | 19 | 108 |
| **1** | Female | B | 17 | 8 | 25 |
| **2** | Female | C | 202 | 391 | 593 |
| **3** | Female | D | 131 | 244 | 375 |
| **4** | Female | E | 94 | 299 | 393 |
| **5** | Female | F | 24 | 317 | 341 |
| **6** | Male | A | 512 | 313 | 825 |
| **7** | Male | B | 353 | 207 | 560 |
| **8** | Male | C | 120 | 205 | 325 |
| **9** | Male | D | 138 | 279 | 417 |
| **10** | Male | E | 53 | 138 | 191 |
| **11** | Male | F | 22 | 351 | 373 |

```
data_dept_wide['Admitted_prop'] = data_dept_wide['Admitted'] / data_dept_wide['total']
data_dept_wide
```

| | Admit | Gender | Dept | Admitted | Rejected | total | Admitted_prop |
|---|---|---|---|---|---|---|---|
| **0** | Female | A | 89 | 19 | 108 | 0.824074 |
| **1** | Female | B | 17 | 8 | 25 | 0.680000 |
| **2** | Female | C | 202 | 391 | 593 | 0.340641 |
| **3** | Female | D | 131 | 244 | 375 | 0.349333 |
| **4** | Female | E | 94 | 299 | 393 | 0.239186 |
| **5** | Female | F | 24 | 317 | 341 | 0.070381 |
| **6** | Male | A | 512 | 313 | 825 | 0.620606 |
| **7** | Male | B | 353 | 207 | 560 | 0.630357 |
| **8** | Male | C | 120 | 205 | 325 | 0.369231 |
| **9** | Male | D | 138 | 279 | 417 | 0.330935 |
| **10** | Male | E | 53 | 138 | 191 | 0.277487 |
| **11** | Male | F | 22 | 351 | 373 | 0.058981 |

```
sns.barplot(x='Dept',y='Admitted_prop', hue='Gender', data=data_dept_wide)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f83835eea90>
```