

Gender Bias

June 4, 2022

```
[45]: import pandas as pd
```

```
[46]: data=pd.read_csv('student_records.csv')
```

```
[47]: data.shape
```

```
[47]: (30, 4)
```

```
[48]: data.head()
```

```
[48]:
```

	S.No	Gender	Score_Football_Question	Total_Examination_Score
0	1	M	1	80
1	2	M	0	90
2	3	M	0	60
3	4	M	1	65
4	5	F	0	80

```
[29]: data[(data['Score_Football_Question']==1) & (data['Gender']=='M')]
```

```
[29]:
```

	S.No	Gender	Score_Football_Question	Total_Examination_Score
0	1	M	1	80
3	4	M	1	65
15	16	M	1	95
16	17	M	1	40
17	18	M	1	50
21	22	M	1	90
29	30	M	1	90

```
[32]: data[(data['Score_Football_Question']==1) & (data['Gender']=='M')].shape
```

```
[32]: (7, 4)
```

```
[30]: data[(data['Score_Football_Question']==1) & (data['Gender']=='F')]
```

```
[30]:
```

	S.No	Gender	Score_Football_Question	Total_Examination_Score
7	8	F	1	90
8	9	F	1	92
9	10	F	1	65

10	11	F	1	70
13	14	F	1	85
14	15	F	1	90
22	23	F	1	55
23	24	F	1	60
24	25	F	1	65

```
[33]: data[(data['Score_Football_Question']==1) & (data['Gender']=='F')].shape
```

```
[33]: (9, 4)
```

```
[35]: data.groupby('Gender').agg({'Score_Football_Question':
    ↳ ['count', 'mean'], 'Total_Examination_Score': 'mean'})
```

```
[35]:
```

	Score_Football_Question		Total_Examination_Score
	count	mean	mean
Gender			
F	16	0.5625	72.625
M	14	0.5000	75.000

```
[ ]: #Ho: There is no gender bias
    #H1: There is a gender bias
```

```
[36]: import statsmodels.api as sm
```

```
[39]: data['Gender_F'] = (data['Gender']=='F').astype(float)
```

```
[41]: Xtrain = data[['Gender_F', 'Total_Examination_Score']]
    ytrain = data['Score_Football_Question']

    # building the model and fitting the data
    log_reg = sm.Logit(ytrain, Xtrain).fit()
```

```
Optimization terminated successfully.
    Current function value: 0.688961
    Iterations 4
```

```
[42]: print(log_reg.summary())
```

```

                                Logit Regression Results
=====
===
Dep. Variable:    Score_Football_Question    No. Observations:
30
Model:                                Logit    Df Residuals:
28
Method:                                MLE    Df Model:
1
```

Date: Wed, 01 Jun 2022 Pseudo R-squ.:
0.002841
Time: 13:13:09 Log-Likelihood:
-20.669
converged: True LL-Null:
-20.728
Covariance Type: nonrobust LLR p-value:
0.7315

```
=====
=====
              coef      std err          z      P>|z|      [0.025
0.975]
-----
-----
Gender_F      0.2397      0.710      0.338      0.736      -1.151
1.631
Total_Examination_Score 0.0002      0.007      0.023      0.981      -0.013
0.014
=====
=====
```

```
[ ]: # Here pvalue is higher that means we accept the null hypothesis. Therefore
      ↳ there is no gender bias
```