CSCI 535 Assignment 3

Mehak Piplani

Question 1:

To investigate bias on traits due to race factors, I extracted a subset from the data belonging to the three races: White (1), Black (2), East Asian (3) and calculated mean, median, and performed ANOVA tests (using f_oneway from scipy.stats).

.

Mean for all traits:

Race	Confident	Egotistic	Intelligent	Kind	Responsible	Trustworthy	Aggressive	Caring	Emotional	Friendly	Sociable
1	5.925109	4.057195	5.863676	5.715496	5.905479	5.645065	3.636961	5.580169	4.869951	5.867516	5.851906
2	5.967909	4.322545	5.528636	5.404500	5.514227	5.346682	4.128409	5.305455	4.732318	5.498636	5.647455
3	5.836190	3.604444	6.317937	5.857302	6.146032	5.866508	3.217619	5.751429	4.845238	6.043333	5.830635

Median for all traits:

Race	Confident	Egotistic	Intelligent	Kind	Responsible	Trustworthy	Aggressive	Caring	Emotional	Friendly	Sociable
1	6.03	4.00	6.00	5.930	6.07	5.80	3.40	5.80	4.87	6.19	6.07
2	6.13	4.20	5.70	5.835	5.80	5.67	3.73	5.73	4.73	5.93	6.13
3	5.87	3.33	6.47	6.190	6.27	6.00	2.93	6.00	4.93	6.33	6.07

One-way ANOVA Test:

TRAIT	P VALUE
Confident	0.5125652685487156
Egotistic	1.4412211501758323e-07
Intelligent	1.6711926732092035e-15
Kind	0.00012878104123732718
Responsible	8.005057161350754e-11
Trustworthy	3.672594877204227e-06
Aggressive	2.7078370672104353e-11
Caring	0.00041893692225684995
Emotional	0.02319487807768113
Friendly	4.9774464052824096e-05

Sociable	0.03449616164486741

The bias is observed in all the traits except Confident trait based on the p values observed from the one-way ANOVA test.

Analysis: The ANOVA test results reveal that in each of the above traits except Confident, at least two of the racial groups had a significant different in their mean values.

Question 2:

To investigate bias on traits due to genders: Female (0), Male (1), I calculated mean, median, and t-tests (using ttest_ind from scipy.stats assuming independent samples).

Mean for all traits:

Gender	Confident	Egotistic	Intelligent	Kind	Responsible	Trustworthy	Aggressive	Caring	Emotional	Friendly	Sociable
0	6.13	3.5	6.00	6.33	6.20	6.13	3.00	6.27	5.33	6.53	6.40
1	5.93	4.4	5.87	5.53	5.87	5.47	3.88	5.33	4.53	5.73	5.67

Median for all traits:

Gender	Confident	Egotistic	Intelligent	Kind	Responsible	Trustworthy	Aggressive	Caring	Emotional	Friendly	Sociable
0	5.994449	3.601511	5.947671	6.172770	6.118059	6.008562	3.127545	6.081123	5.289098	6.314092	6.257156
1	5.865760	4.432829	5.752246	5.303822	5.667762	5.309094	4.110788	5.139165	4.528345	5.450024	5.488266

T- Test:

TRAIT	P VALUE
Confident	0.000249067
Egotistic	5.724535735483598e-103
Intelligent	6.617541566954897e-10
Kind	3.6772754632964984e-84
Responsible	5.314276336798082e-34
Trustworthy	5.706506511591532e-73
Aggressive	6.29311355462135e-101
Caring	6.745136923994714e-105
Emotional	1.029377924430242e-164
Friendly	3.7108186188490396e-64
Sociable	3.1946748840509533e-62

The bias is observed in all the traits based on the p values observed from the t-test.

Analysis: : The t-test results reveal that in each of the above traits both the genders had a significant different in their mean values.

Question 3:

The 5 most significant biases among <u>gender and Overall (i.e., among gender and race):</u>

	Type, trait	p-value
1)	Gender, Emotional,	1.029377924430242e-166
2)	Gender, Caring,	6.745136923994714e-105
3)	Gender, Egotistic,	5.724535735483598e-103
4)	Gender, Aggressive,	6.29311355462135e-101
5)	Gender, Kind,	3.6772754632964984e-84

Analysis: Based on mean values, we can say Women are more Emotional, Caring, and Kind whereas Men are more Egoistic and Aggressive.

The 5 most significant biases among <u>race:</u>

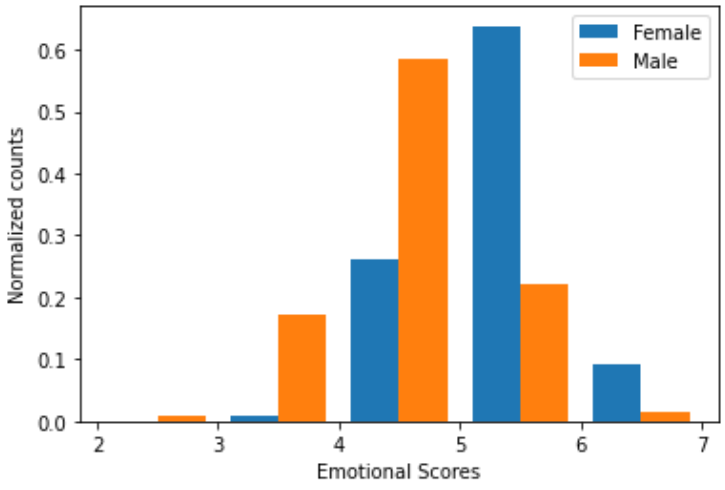
	Type, trait	p-value
1)	Race, Intelligent,	1.6711926732092035e-15
2)	Race, Aggressive,	2.7078370672104353e-11
3)	Race, Responsible,	8.005057161350754e-11
4)	Race, Egotistic,	1.4412211501758323e-07
5)	Race, Trustworthy,	3.672594877204227e-06

Analysis: Based on mean values, we can say East Asian people are more Intelligent, Responsible, and Trustworthy whereas Black people are more Egoistic and Aggressive.

Question 4:

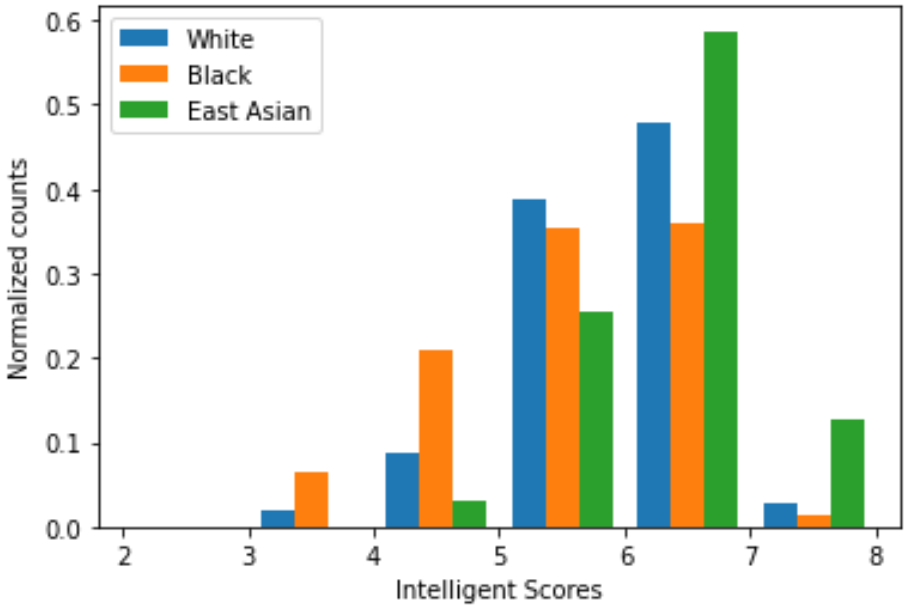
The histograms are plotted using hist function from matplotlib.pyplot and density parameter which returns a probability density: each bin will display the bin's raw count divided by the total number of counts and the bin width.

The histogram for the most significant biases among **<u>gender and Overall (i.e., among gender and race):</u>**



Analysis: Women are more emotional than men.

The histogram for the most significant biases among **race:**



Analysis: East Asian people are intelligent as compared to the other two races. White people are more intelligent compared to Black.

Question 5:

GENDER:

Gender	Selection Rate
0	0.118573
1	0.208826

Analysis: The value of selection rate of females is less than 4/5th of the value of selection rate of males, hence male group is an adversity towards female group.

RACE:

Race	Selection Rate
1	0.183007
2	0.122727
3	0.174603

Analysis: The value of selection rate of Black race group is less than 4/5th of the value of selection rate of White race group as well as East Asian group, hence both White and East Asian races are adversities towards Black race group.

Question 6:

The other contributing factors to bias in ML systems, aside from human labeler bias are:

1) **Sample bias:** This type of bias is observed when the collected dataset does not consider the facts of the environment in which the model would run. An example of this is certain facial recognition systems trained primarily on images of white men.

2) **Exclusion Bias:** This type of bias mostly occurs in the stage of data pre-processing when a valuable part of data/feature is removed thinking it is unimportant.

3) **Observer Bias:** This type of bias happens because of seeing what you expect to see or want to see in data in data. This means that the observer/experimenter is working with conscious/unconscious prejudices.

4) **Measurement Bias:** This type of bias occurs when data distortion is observed due to faulty measurements or when the data distribution of the collected data is different from real world samples.

5) **Racial Bias:** This type of bias can occur when data skews in favor of a particular demographics.

6) **Association Bias:** This type of bias occurs when the data for a machine learning model reinforces and/or multiplies a cultural bias. For example, if a dataset consists of a collection of jobs in which all men are doctors, and all women are nurses. In reality, this does not imply that women cannot be doctors, and men cannot be nurses. However, as far as your machine learning model is concerned female doctors and male nurses do not exist.

Question 7:

The ways to mitigate such biases in machine learning are:

1) Take in account all the cases you expect your model to be exposed to. This can be done by examining the domain of each feature and make sure we have balanced evenly distributed data covering all of it. Otherwise, you'll be faced by erroneous results and outputs the don't make sense will be produced.
2) Investigate before discarding features by doing sufficient analysis on them.
3) Ensure that observers (people conducting experiments) are well trained and screening them for potential biases.
4) Create a gold standard for your data labeling. A gold standard is a set of data that reflects the ideal labeled data for your task. It enables you to measure your team's annotations for accuracy.
5) Use multi-pass annotation for any project where data accuracy may be prone to bias. Examples of this include sentiment analysis, content moderation, and intent recognition.
6) Enlist the help of someone with domain expertise to review your collected and/or annotated data. Someone from outside of your team may see biases that your team has overlooked.090
7) Choose the right learning model for the problem. Each problem requires a different solution and provides varying data resources. There's no single model to follow that will avoid bias, but there are parameters that can inform your team as its building.

8) Choose a representative training data set.

9) Monitor performance using real data.