Corey Capooci
cvcapooc@ncsu.edu
HW4

1

There are two combinations of genotypes that produce the close to the dosage we are looking for.
CYPTC9: *2/*3, VKORC1: A/G, Dosage: 21.242
CYPTC9: *3/*3, VKORC1: unknown, Dosage: 20.982
The combination of CYPTC9 *3/*3 and VKORC1 unknown is closer and I would chose this one.

2

a

X	Accuracy (when N=3)
3	1173/10000 (12%)
5	1318/10000 (13%)
7	894/10000 (9%)
10	1224/10000 (12%)

b

N	Accuracy (when X=5)
3	831/10000 (8%)
5	1258/10000 (13%)
10	915/10000 (9%)

c

There is no clear definitive pattern according to the tables in **a** and **b**. Nevertheless, for both **a** and **b**, it appears the accuracy improves with larger X or larger N. Note for **a** that X=5 and X=10 performed better than X=3. It did not occur in a linear fashion but there a slight upward trend. For **b**, N=5 and N=10 performed better than N=3. The best performer is X=5 for **a** and N=5 for **b**.

3

No statistical parity for Female group. Difference is 0.1300
Statistical parity for Asian-Pacific-Islander group. Difference is 0.0300

P(Income>$50K | whole dataset) = 0.2393
P(Income>$50K | Female) = 0.1093
P(Income>$50K | Asian-Pacific-Islander) = 0.2693