Wasfi Momen

CSC 6223 Assignment 2

1. 1. k=2; records 1 and 4 are an anonymity set and records 2,3,5,6, and 7 are an anonymity set
   2. k=3; there are at least 3 male records 5,6,7 and three plus one female records 12,3,4.
   3. L=2 for Dataset 1 since each record has at least one other record that is the same. There is no l-Diversity for Dataset 2 since Fav. Show has unique variables that compromise the quasi-identifiers.
2. (P || Q) = .1\* ln(.1/.4) + .2 \* ln(.2/.3) + .3 \* ln(.3/.3) + .4 \* ln(.4/.1) = **0.33479528671**

Count query local sensitivity of D:

Global value is three since there are 3 eights.

There is a count of 2 for one and 2 for two. All other counts are 1.

So the max count value being 3, we do 3-1 and 3-2 and take the max. This means the count query local sensitivity is **2**.

Median query local sensitivity of D:

Med(D) = x­m, m = (n+1)/2

n is 11 so 12/2 is 6. 6th element in D is 6 =xm

LSf(D) = max(xm – xm-1 , xm+1 – xm)

= max(6 – 4, 7- 6)

= max(2,1)

= **2**



With differential privacy as

We can substitute the values as

And ­solve for ε. To be at least equal to .3, eε has to be at least episolon value **ln(.3)**. So our differential privacy protection level is about **-1.204**.



Since satisfies differential privacy.

The probability of yes answers is ¾ and the probability of no answers is ¼

.75/.25 is 3. So randomized response gives **ln(3)** differential privacy.