Assignment2

Q1:

All possible items are: {1,2,3,4,5}, so $C_4 = F_3 \times F_1 = \{\{1,2,3,4\},\{1,2,3,5\},\{1,2,4,5\},\{1,3,4,5\},\{2,3,4,5\}\}$.

Q2:

The F_3 is sorted, so we can do Priori by traversing the F_3 in $O(n^2)$.

For {1,2,3}, we can combine it with {1,2,4} and {1,2,5} to get {1,2,3,4} and {1,2,3,5};

For {1,2,4}, we can combine it with {1,2,5} to get {1,2,4,5}. {1,3,4} is the first element in F_3 that we can not combine, and because F_3 is sorted, the following elements can not combine neither.

For {1,3,4} , nothing to combine.

For{2,3,4}, it can be combined with {2,3,5} to get {2,3,4,5};

For {2,3,5} and {3,4,5}, nothing to combine.

So C_4={{1,2,3,4},{1,2,3,5},{1,2,4,5},{2,3,4,5}}.

Q3:

Survived: {1,2,3,4}

For {1,2,3,4} we have no {1,3,5}, for {1,2,4,5} we have no {1,4,5}, for(2,3,4,5} we have no {2,4,5}

Q4.a:

The cardinality of the data is 7, so the number of rules is $3^7 - 2^{7+1} + 1 = 1932$

Q4.b:

$$\frac{\sigma(\{Milk, Diapers, Butter\})}{\sigma(\{Milk, Diapers\})} = \frac{2}{4} = 0.5$$

Q4.c:

$$\frac{\sigma(\{Milk, Diapers, Butter\})}{|T|} = \frac{2}{10} = 0.2$$

Q5:

True. $|\{a,b\}| \geq |\{a,b,c,d\}|$, so if {a,b} is not a frequent list, which means $|\{a,b\}| < threshold$, $|\{a,b,c,d\}| < threshold$.

Q6:

False. There can be no {a,b,c} even if all {a,b},{a,c},{b,c} reached the threshold.

Q7:

False. May be {b} occurs along for 50 times.

Q8:

False. If the cardinality of a set is 5 and minsup = 1, then there can be 10 qualified frequent set at most.

Q9: