1.1 Tan, Ch. 5 (Association Analysis)

1-a:

Answer: Milk −→ Bread. Such obvious rule tends to be uninteresting.

1-b:

Answer: Milk −→ Tuna. While the sale of tuna and milk may be higher than the support threshold, not all transactions that contain milk also contain tuna. Such low-confidence rule tends to be uninteresting.

1-c:

Answer: Cooking oil −→ Laundry detergent. Such low confidence rule tends to be uninteresting.

1-d:

Answer: Vodka −→ Caviar. Such rule tends to be interesting.

2-a:

s({e}) =8/10=0.8
s({b,d}) =2/10=0.2
s({b,d,e}) =2/10=0.2

2-b:

c(bd −→ e) =0.2/0.2 = 100%
c(e−→bd) =0.2/0.8=25%

2-c:

s({e}) =4/5= 0.8
s({b,d}) =5/5=1
s({b,d,e}) =4/5=0.8

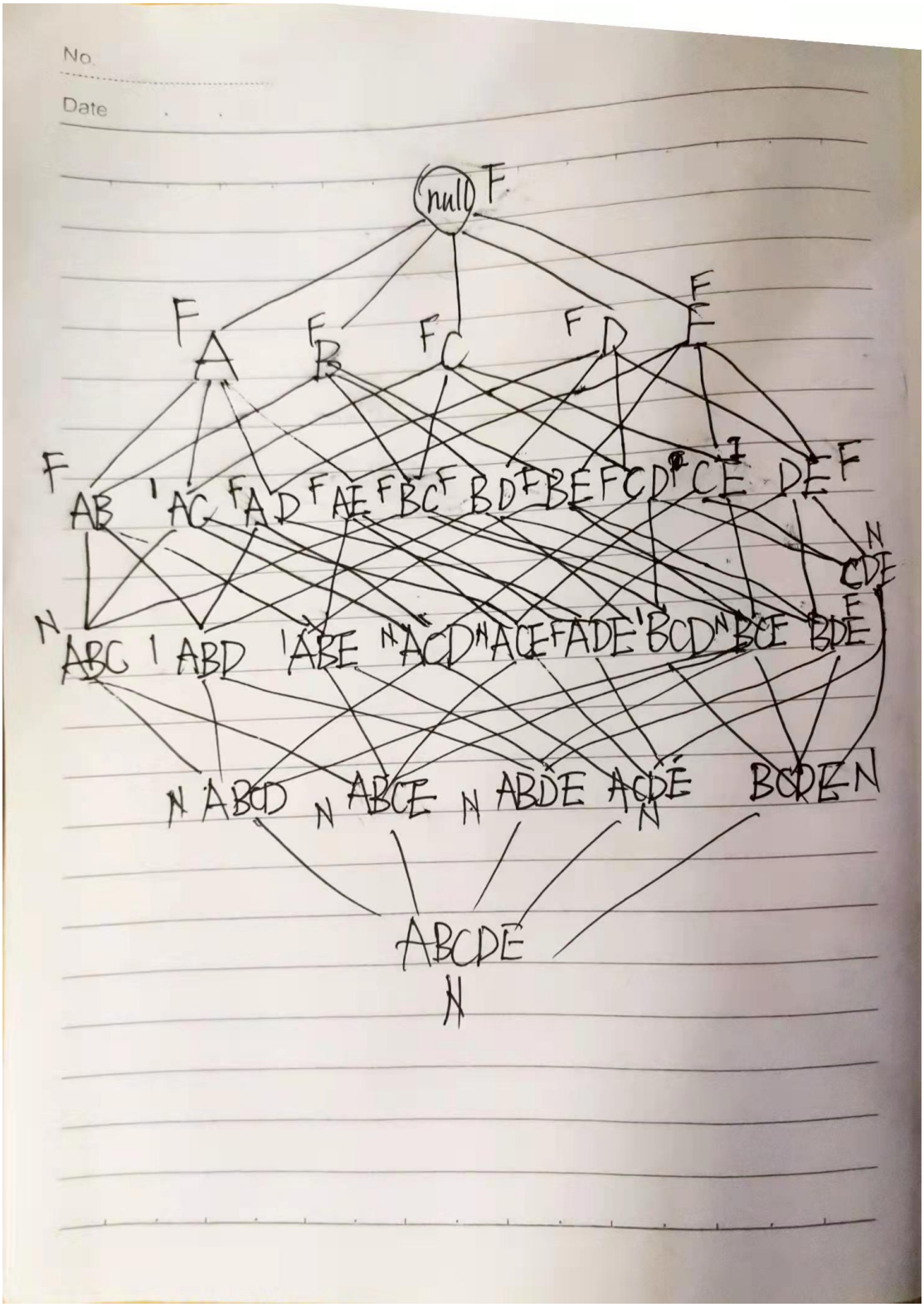
2-d:

c(bd−→e) =0.8/1=80%
c(e −→ bd) =0.8/0.8= 100%

2-e:

There are no apparent relationships between s1, s2, c1, and c2.

9-a:



9-b:

Percentage of frequent itemsets = 16/32 = 50.0% (including the null set).

15-a:

Answer: Data set (e) because it has to generate the longest frequent itemset along with its subsets.

15-b:

Answer: Data set (d) which does not produce any frequent itemsets at 10% support threshold.

15-c:

Answer: Data set (e).

15-d:

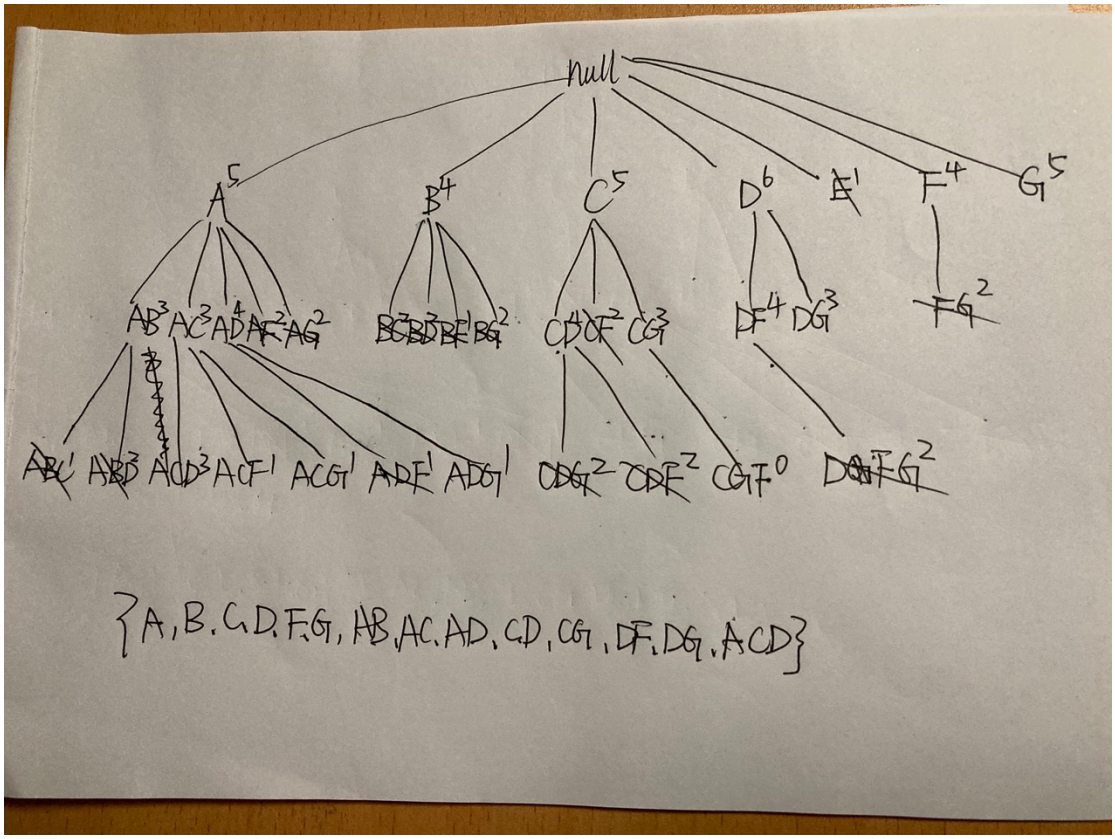
Answer: Data set (b).

15-e:

Answer: Data set (e).

1.2 1.2 Zaki, Chapter 8 (Frequent Pattern Mining)

Q1-a:



Q4:

{ABE}==>{C}
{ABE}==>{D}
{ABE}==>{CD}