

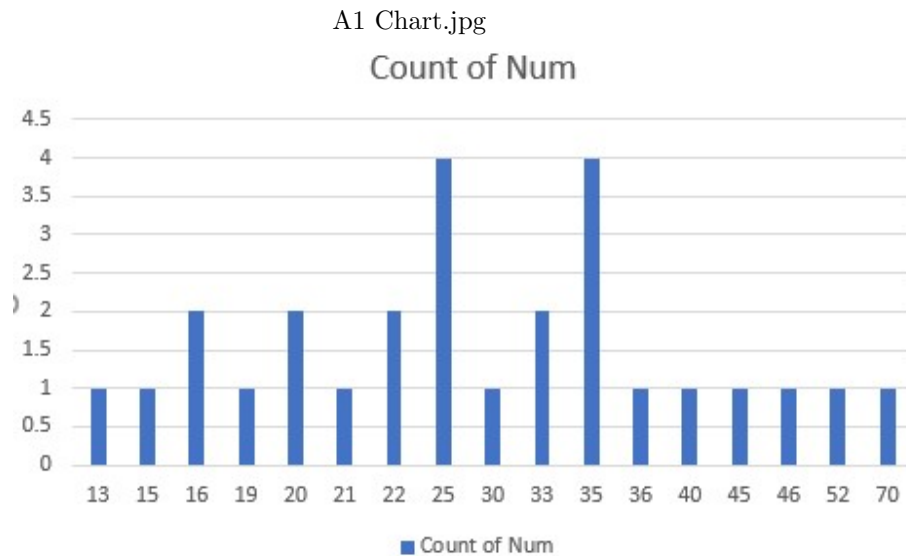
CP421 a01

Max Niebergall
nieb3100@mylaurier.ca
160623100

January 2019

1. (a) Mean ≈ 29.96
Median = 27.5

(b) Mode

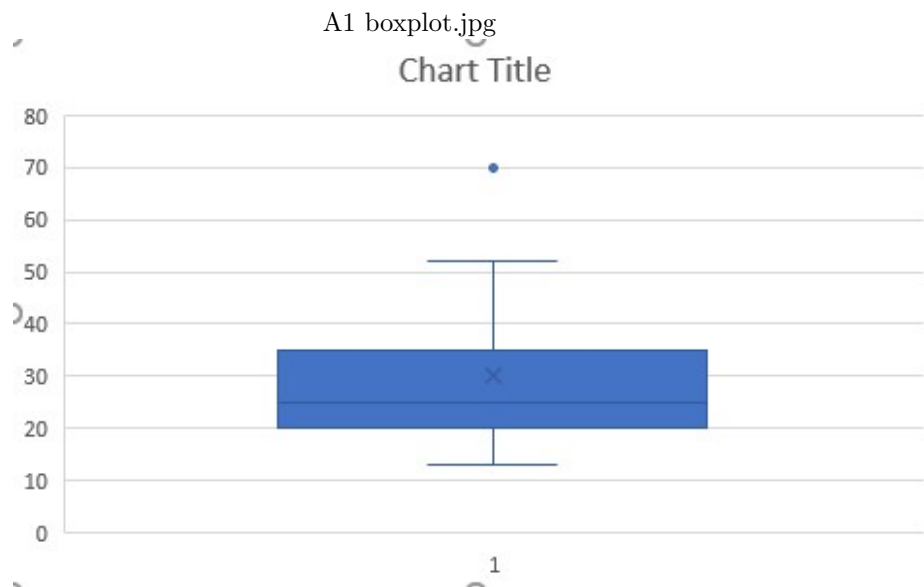


This data set is bimodal, with local maxima at 25, and 35

- (c) Midrange = 41.5
- (d) first quartile = 20
third quartile = 35
- (e) Minimum: 13
Quartile Q1: 20

Median: 25
 Quartile Q3: 35
 Maximum: 70

(f) Boxplot



2. (a) i. A1 euclidian distance.jpg

	x1	x2	x3	x4	x5	x
x1	0					
x2	0.538516	0				
x3	0.141421	0.412311	0			
x4	0.360555	0.894427	0.5	0		
x5	0.7	1.029563	0.806226	0.583095	0	
x	0.141421	0.67082	0.282843	0.223607	0.608276	0

ii. Cosine Similarity

x1	x2	x3	x4	x5
0.999991391	0.995752261	0.999969484	0.999028235	0.965363393

(b) A1 Norm euclidian distance.jpg

	x1	x2	x3	x4	x5	x
x1	0					
x2	0.088026	0				
x3	0.003663	0.084366	0			
x4	0.048234	0.136187	0.051896	0		
x5	0.259084	0.171549	0.255452	0.306836	0	
x	0.004149	0.092171	0.007812	0.044085	0.263198	0

3. (a) item: 1 count: 100
 item: 2 count: 50
 item: 3 count: 33
 item: 4 count: 25
 item: 5 count: 20
 item: 6 count: 16
 item: 7 count: 14
 item: 8 count: 12
 item: 9 count: 11
 item: 10 count: 10
 item: 11 count: 9
 item: 12 count: 8
 item: 13 count: 7
 item: 14 count: 7
 item: 15 count: 6
 item: 16 count: 6
 item: 17 count: 5
 item: 18 count: 5
 item: 19 count: 5
 item: 20 count: 5
- (b) i. $1/2$
 ii. $1/8$

4. This program operates by applying the apriori algorithm to prune infrequent elements and their subsets. This program first reads through the file of transactions and counts the occurrences of each unique item. It then prunes all items which were counted less than 100 times. Next, the program searches through the transactions again and counts only the pairs of items which were not pruned in the previous step. All pairs of items which occur ed less than 100 times are then pruned. Finally, the program generates all possible combinations of the unpruned set of pairs, which are stored in a dictionary. The program counts all occurances of each triple in a each transaction. All triples with less than 100 ocurances are pruned. The confidence of each triple and pair association are calcaulted and written to files.

- (a) A file "pairs.txt" has been included in this submission and contains a list of all pair associations of two items, and the confidence of the association.
- (b) A file "triples.txt" has been included in this submission and contains a list of all "triple" associations of two items, and the confidence of the association.

Top 5:

('DAI23334', 'ELE92920', 'DAI62779') 1.0
('DAI31081', 'GRO85051', 'FRO40251') 1.0
('DAI55911', 'GRO85051', 'FRO40251') 1.0
('DAI75645', 'GRO85051', 'FRO40251') 1.0
('ELE17451', 'GRO85051', 'FRO40251') 1.0