1) A local retailer has a database that stores 10,000 transactions of last summer. After analysing the data, a data science team has identified the following statistics:

- {battery} appears in 6000 transactions.
- {sunscreen} appears in 5000 transactions.
- {sandals} appears in 4000 transactions.
- {bowls} appears in 2000 transactions.
- {battery, sunscreens} appears in 1500 transactions.
- {battery, sandals} appears in 1000 transactions.
- {battery, bowls} appears in 250 transactions.
- {battery, sunscreen, sandals} appears in 600 transactions

Answer the following questions:

a) What are the support values of the preceding itemsets?

b.) Assuming the minimum support is 0.05, which itemsets are considered frequent?

c) What are the confidence values of {battery} → {sunscreen} and {battery, sunsceen} → {sandals}? Which of these two rules is more interesting (ie have higher value of confidence)?

d) List all the candidate rules that can be formed from the statistics. Which rules are considered interesting at the minimum confidence 0.25? Out of these interesting rules, which rule is considered the most useful (that is, least coincidental).

2) Suppose for three products A, B and C,

support (A) = .6, support (B) = .6,

Confidence $(B \rightarrow A) = .9$ and confidence $(C \rightarrow \{AB\}) = .5$.

Compute the following quantities.

a) Lift $(A \rightarrow B)$.

b) Leverage $(A \rightarrow B)$.

c) Confidence $(A \rightarrow B)$.

d) Lift $(\{AB\} \rightarrow C)$.

# Logistic regression

1. If the probability of an event occurring is 0.4, then
   (i) What is the odds ratio?
   (ii) What is the log odds ratio?
2. If $\beta_3 = -0.5$ is an estimated coefficient in a logistic regression model, what is the effect on the odds ratio for every one unit increase in the value of $x_3$ ?