

# Association Analysis (Market Basket Analysis)



Analyze transactional data such as product purchases and occurrence of events to identify those that are dependent upon each other or tend to occur together. Metrics such as the likelihood of items/events occurring based on the occurrence of other items/events, among others are produced.

Note that the data must be in list format, where each row identifies the customer or transaction ID (in one column) and the item or event (in another column).

## Association Analysis

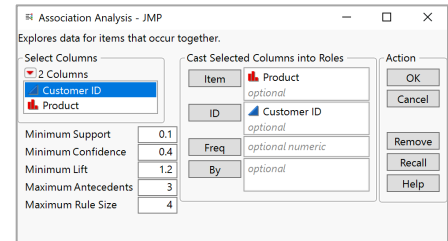
- From an open JMP® data table, select **Analyze > Screening > Association Analysis**.
- Select a categorical variable from **Select Columns** and click **Item** (categorical variables have red or green bars).
- Select the variable that represents the transaction ID and click **ID**.
- Accept the default settings (lower left), and click **OK**.

By default, JMP provides the following tables of results:

- Frequent Item Sets: the list of item sets whose support (proportion of occurrences) exceeds the minimum value specified in the launch window.
- Rules: Association rules that meet the minimum support, minimum confidence, minimum lift, maximum antecedents, and maximum rule size requirements specified in the launch dialog.

Here, we have sorted the tables in ascending order on Item Set and Condition (respectively). To sort a table, right-click on the table, select **Sort by Column**, select the column of interest, check **Ascending**, and click **OK**.

Example: Grocery Purchases.jmp (Help > Sample Data Folder)



| Frequent Item Sets                     |         |         |  |
|----------------------------------------|---------|---------|--|
| Item Set                               | Support | N Items |  |
| {apples, avocado, baguette}            | 11%     | 3       |  |
| {apples, avocado}                      | 14%     | 2       |  |
| {apples, baguette}                     | 15%     | 2       |  |
| {apples, corned beef, herring, olives} | 11%     | 4       |  |
| {apples, corned beef, herring}         | 11%     | 3       |  |
| {apples, corned beef, olives}          | 12%     | 3       |  |
| {apples, corned beef, steak}           | 10%     | 3       |  |
| {apples, corned beef}                  | 15%     | 2       |  |
| {apples, Heineken}                     | 10%     | 2       |  |
| {apples, herring, olives}              | 11%     | 3       |  |

| Rules                        |                 |            |       |
|------------------------------|-----------------|------------|-------|
| Condition                    | Consequent      | Confidence | Lift  |
| apples                       | avocado         | 44%        | 1.221 |
| apples                       | corned beef     | 48%        | 1.231 |
| apples                       | peppers         | 41%        | 1.4   |
| apples                       | sardines        | 40%        | 1.357 |
| apples, avocado              | baguette        | 81%        | 2.058 |
| apples, baguette             | avocado         | 76%        | 2.101 |
| apples, corned beef          | herring         | 75%        | 1.541 |
| apples, corned beef          | olives          | 77%        | 1.626 |
| apples, corned beef          | steak           | 68%        | 2.979 |
| apples, corned beef          | herring, olives | 72%        | 2.823 |
| apples, corned beef, herring | olives          | 96%        | 2.041 |

### Interpretation:

The minimum support is 0.1 and the minimum confidence is 0.4 (see the default settings above). Thus, an item set will not appear in either list unless that item set is purchased by at least 10% of the customers (this is support). Furthermore, it won't appear in the rules list unless a consequent item is purchased at least 40% of the time (this is confidence). For example, apples and avocado were purchased together 14% of the time. And, when apples and avocados were both purchased by a customer, that customer also purchased a baguette 81% of the time.

The maximum antecedents is 3 and the maximum rule size is 4. Therefore, the rules list only contains conditions of three or fewer items and no rule includes more than 4 condition and consequent items (combined).

The minimum lift is 1.2. Lift, which is the ratio of the confidence to the support, measures how much more (or less) likely a customer is to purchase the consequent item when they are already purchasing the condition items, compared to the probability of buying the consequent item in general. For example, the lift for apples, avocado and the consequent baguette is 2.058, indicating that customers are more than twice as likely to buy a baguette when they are also buying apples and avocados.

Visit **Predictive and Specialized Models > Association Analysis** in JMP Help to learn more.