JMP Instructional data set 030

# Child Resistant Packaging

jmp STATISTICAL DISCOVERY

## JMP platforms and statistical methods

**Graph Builder:** Mosaic plot; scatter plot; linear regression
**Distribution:** Summary statistics; confidence intervals for proportions
**Fit Y by X:** Linear regression
**Fit Model:** Logistic regression

## Problem statement

There are many products found in a household, such as prescription drugs, cleaning solutions, and pesticides, that pose a serious danger to children if ingested. The Food and Drug Administration (FDA), Consumer Product Safety Commission (CPSC), and other regulatory agencies establish standards that packaging must meet to be certified as child-resistant packaging (CRP). One requirement for a package to be considered CRP is that at least 85% of a panel of 200 children tested between the ages of 42-51 months of age would be unable to open it after five minutes. Another requirement is that at least 80% of these children would still be unable to open it, even after the process for doing so is demonstrated to the child and they are given another five minutes to try.

To test the safety level of these four different packaging designs, a test was conducted with a panel of 200 children between 42 to 51 months in age. For each age, 10 boys and 10 girls were selected for the test. Each child tested all four packages with two days in between testing of each package. The order of testing was chosen at random. The child was given the package and simply instructed to open it to access the contents inside. After five minutes, the administrator of the test demonstrates to the child how to open the package without any verbal description. The child is then given another five minutes to open the package, after which the testing stops.

The focus of the analysis will be to evaluate and compare how effective these four different packaging designs are at preventing a child from opening them.

## Data

**IDS_030.jmp**

| | |
|---|---|
| **Child ID** | Identification of test subject |
| **Age** | Age (months) |
| **Gender** | Gender of test subject (Boy, Girl) |
| **Design** | Package design (A, B, C, D) |
| **Testing Order** | Order subject tested the four designs |
| **Demonstration** | Trial before or after opening of the package was demonstrated to subject |
| **Open** | Binary indicator (Yes, No) if subject was able to open package |
| **Time to Open** | Time (mm:ss) for subject to open package.<br>*Note: Missing value if subject was unable to open within 5 minutes* |

## Exercises

1. Create a mosaic plot displaying the percentage of children who are able/unable to open up the 4 different package designs with and without demonstration. Provide a few short sentences summarizing the results. How do these results compare to the requirements for CRP?

   ***Instructions:***
   *Graph > Graph Builder. Place the variable 'Open' on the Y axis. Place 'Design' on X axis. Also place 'Demonstration' on the X axis above 'Design'.*
   *Select the Mosaic Plot from the Graph Palette.* 
   *Choose to 'Label by Percent'.*

2. Create a Bar Chart and Frequency Table of the percentage of the open rate for each design and before/after demonstration. Note: These display the same information as the Mosaic Plot. The Distribution Platform will allow other statistical analyses to be done.

   ***Instructions:***
   *Analyze > Distribution. Place 'Open' as the Y role. Place both 'Design' and 'Demonstration' as the By Variable. Click OK.*

3. The analyses done in Exercises 1 and 2 summarized the results of the test. They do not, however, provide estimates of a population of children who would be unable to open the packages incorporating statistical uncertainty. To do so, create 99% confidence interval estimates for the proportion of children able and unable to open the package. Provide an interpretation of these intervals. Based upon these results, which package designs achieve the results of >85% unable to open prior to demonstration and >80% after demonstration making inference to a population a children not just those tested?

   ***Instructions:***
   *Under the red triangle next to each bar chart in the Distribution report, select Confidence Interval > 0.99.*

4. Create a Heat Map displaying the results of each child tested. Provide a few short sentences summarizing what the visualization shows.

***Instructions:***
*Graph > Graph Builder. Place the variable 'Child ID' on the X axis. Place 'Design' and 'Demonstration' on the Y axis. Place the variable 'Open' in the Color Role. Select the Heat Map from the Graph Palette.* 

5. Fit the following set of logistic regression models:
   a) 'Open' vs 'Age' for those able to open package A before demonstration
   b) 'Open' vs 'Age' for those able to open package A after demonstration
   c) 'Open' vs 'Age' for those able to open package B before demonstration
   d) 'Open' vs 'Age' for those able to open package B after demonstration
   e) 'Open' vs 'Age' for those able to open package C before demonstration
   f) 'Open' vs 'Age' for those able to open package C after demonstration
   g) 'Open' vs 'Age' for those able to open package D before demonstration
   h) 'Open' vs 'Age' for those able to open package D after demonstration

   Provide an interpretation of these logistic regression equations. Explain why it's not possible to fit a logistic regression model for all conditions? For those conditions where a model could be created, which of the models do you feel most and least confident in using the results to generalize beyond the children used in the study? Why?

***Instructions:***
*Analyze > Fit Model. Place 'Open' in the Y role and 'Age' in the X. Click OK.*
*Note: The initial model is for all the data. To examine subsets of the data, choose 'Local Data Filter' under the red triangle in the report. Choose to use the variables 'Design' and 'Demonstration' in the data filter. Now a select design and demonstration stage can be chosen, and the model and report will update using only those data.*

6. Use the logistic regression models in Exercise 5 to estimate the probability that a child 4 ½ years of age would be able to open package B and package C with and without a demonstration.

***Instructions:***
*Add the Profiler to the report by selecting 'Profiler' under the top red triangle. Enter the value 54 for the Age. Note: Be sure to choose the specific Design and Demonstration stage in the Data Filter before interpreting the results.*

7. Build multiple variable logistic regression models for each condition of ‘Design’ and ‘Demonstration’ to determine if there is any statistical evidence indicating that there is a difference between the genders in terms of the likelihood of being able to open the packages.

***Instructions:***
*Analyze > Fit Model. Place ‘Open’ in the Y role. Select ‘Age’ and ‘Gender’ from the column list, and choose Macros > Degree 2 from Construct Model Effects section.*
*Note: This will fit a model that includes the main effects of ‘Age’ and ‘Gender’ as well as the ‘Age*Gener’ interaction. This equation if for all the package designs and demonstration stages. To examine subsets of the data, choose ‘Local Data Filter’ under the red triangle and choose to use the variables ‘Design’ and ‘Demonstration’. Now a select design and demonstration stage can be chosen, and the report will update using only those data. For each ‘Design’ and ‘Demonstration’ stage, examine the p-values testing the significant of the ‘Age*Gender’ interaction term as well as the ‘Gender’ main effect.*

8. Create a scatterplot along with linear regression equations showing the relationship between ‘Time to Open’ and ‘Age’ for those children able to open the packages. Provide a few sentences summarizing the results revealed in the graph.

***Instructions:***
*Graph > Graph Builder. Place ‘Time to Open’ on the Y axis and ‘Age’ on the X axis. Place ‘Demonstration’ in the Overlay role and ‘Design’ in the Color role.*

9. Fit the following set of linear regression models:
   a) ‘Time to Open’ vs ‘Age’ for those able to open package A before demonstration
   b) ‘Time to Open’ vs ‘Age’ for those able to open package A after demonstration
   c) ‘Time to Open’ vs ‘Age’ for those able to open package B before demonstration
   d) ‘Time to Open’ vs ‘Age’ for those able to open package B after demonstration
   e) ‘Time to Open’ vs ‘Age’ for those able to open package C before demonstration
   f) ‘Time to Open’ vs ‘Age’ for those able to open package C after demonstration
   g) ‘Time to Open’ vs ‘Age’ for those able to open package D before demonstration
   h) ‘Time to Open’ vs ‘Age’ for those able to open package D after demonstration

   Provide an interpretation of these linear regression equations. Explain why it’s not possible to fit a regression model for all conditions? For those conditions where a model could be created, which of the models do you feel most and least confident in using the results to generalize beyond the children used in the study? Why?

***Instructions:***
*Analyze > Fit Y by X platform. Place ‘Time to Open’ in the Y role and ‘Age’ in the X. From the red triangle at the top of the report, choose ‘Fit Line’.*
*Note: The initial model is for all the packages that have been opened. To examine subsets of the data, choose ‘Local Data Filter’ under the red triangle and choose to use the variables ‘Design’ and ‘Demonstration’. Now a select design and demonstration stage can be chosen and the model and report will update using only those data.*

10. For both package B and C, estimate the 'Time to Open' for children 4 ½ years of age who are able to open the package within 5 minutes without a demonstration. Why should you not use this equation to make predictions of 'Time to Open' for 4 ½ year old children in general?

***Instructions:***
*Add the Profiler to the report by selecting 'Profiler' under the top red triangle. Enter the value 54 for the Age. Note: Be sure to choose the specific Design and Demonstration stage in the Data Filter before interpreting the results.*

11. Create a mosaic plot displaying the percentage of children who are able/unable to open up the packages by 'Testing Order'. Is there any evidence to indicate that the likelihood of a child being able to open a package increases if it occurs later in the testing?

***Instructions:***
*You'll first need to create a new variable for 'Testing Order' that is Ordinal. Do so by choosing Cols > New Cols. Copy and Paste the 'Testing Order' column into the new column created. Right-click on the column and choose Column Info. Change the Modeling Type to Ordinal. Name this new column 'Testing Order_Ordinal'.*
*Graph > Graph Builder. Place the variable 'Open' on the Y axis. Place 'Testing Order_Ordinal' on the X axis. Select the Mosaic Plot from the Graph Palette.* 

*Note: The initial graph is for all the packages and stages of demonstration. To examine subsets of the data, choose 'Local Data Filter' under the red triangle and choose to use the variables 'Design' and 'Demonstration'. Select all the different design and demonstration stages and examine the proportions of open/unopen.*

12. Create a scatterplot displaying the 'Time to Open' vs. 'Testing Order'. Is there any evidence to indicate that the 'Time to Open' decreases if it occurs later in the testing?

***Instructions:***
*Graph > Graph Builder. Place the 'Time to Open' on the Y axis and 'Testing Order' on the X axis. Place 'Demonstration' in the Overlay role and 'Design' in the Color role. Note: This initial graph is for all the packages and stages of demonstration. To examine subsets of the data, choose 'Local Data Filter' under the red triangle and choose to use the variables 'Design' and 'Demonstration'. Select all the different design and demonstration stages and examine the 'Time to Open' across the 'Testing Order'.*

13. What are some ideas for further testing that would help better understand the safety of these different designs?