JMP Academic Case Study 033

Durability of Mobile Phone Screen – Part 2

Inferential Statistics: Confidence Intervals, Contingency Analysis, Comparing Proportions via Difference, Relative Risk and Odds Ratio

Produced by

Kevin Potcner, JMP Global Academic Team

kevin.potcner@jmp.com

jmp STATISTICAL DISCOVERY

Durability of Mobile Phone Screen – Part 2

Inferential Statistics: Confidence Intervals, Contingency Analysis, Comparing Proportions via Difference, Relative Risk and Odds Ratio

Key Ideas

This case study involves performing a contingency table analysis to test if there are significant differences between more than two proportions. To describe those differences, pairwise comparisons using difference in proportions, relative risk and odds ratios are conducted.

This case study is an extension to the case study, JMP032 Durability of Mobile Phone Screen - Part 1. Though it isn't necessary to have completed that case study first, it is recommended since they present a more comprehensive description of each stage of the study and subsequent analyses when viewed together.

Background

The durability of a product is clearly an important quality characteristic for both the end user and the manufacturer. For end users, durability is especially important for mobile phones. Dropping a phone on a hard surface, for example, can result in the screen cracking or even breaking, rendering the phone unusable. To evaluate the durability of these screens, manufacturers subject a sample of screens to a variety of tests to simulate typical wear and tear by a user, such as dropping the phone onto a concrete surface.

In JMP032 Durability of Mobile Phone Screen - Part 1, material scientists for a screen manufacturer experimented with two new formulations of an aluminosilicate glass (A and B). These two formulations were produced by making a change to a final processing step that uses a specific level of potassium nitrate to strengthen the glass.

A sample of 10 screens of each type was developed for testing. Each screen was installed into the same style of phone. The phones were then dropped in a controlled identical manner from a height of 1 meter onto a concrete surface. A binary variable "Success" (no damage) and "Fail" (screen damage) was recorded.

One of the company's goals is that 97% of the screens manufactured would be able to experience a drop of 1 meter without becoming damaged (i.e., the Population Success Rate).

The analyses illustrated in JMP032 Durability of Mobile Phone Screen - Part 1, which were based on only 10 phones tested of each Screen Type, failed to generate the statistical evidence needed to demonstrate the desired Success Rate. The analyses also failed to find any statistically significant difference between the two Screen Types.

The engineers decided to test an additional 40 phones for each of the two Screen Types. Analyzing the results from this expanded test was the objective of the exercises for that case study.

To continue the work in creating a highly durable material to use for mobile phone screens, the engineers developed a third formulation (Type C), which requires a more expensive process. In the test, 50 specimens of Type C were subjected to a 1.0m drop. In an effort to gain a more comprehensive understanding of the durability of these screens, the engineers also conducted the drop test at two additional heights (0.5 and 1.5 meters). At 1.5m, 50 phones of each Screen Type were tested; 40 were used for the 0.5m test. These new tests, along with the previous tests, provide us with the following data for analysis:

	0.5m	1.0m	1.5m
Type A	n=40 S=36 F=4	n=50 S=41 F=9	n=50 S=28 F=22
Type B	n=40 S=40 F=0	n=50 S=48 F=2	n=50 S=40 F=10
Type C	n=40 S=39 F=1	n=50 S=47 F=3	n=50 S=39 F=11

The Task

The primary objectives of this analysis are:

1. Describe the durability of each of the three Screen Types at the three different Drop Heights.
2. Describe any statistical differences across the Screen Types and Drop Heights.
3. Determine if the durability of Screen Type C warrants the more expensive process (i.e., is the durability of Screen Type C better than that of A and B?).

The Data **drop-test-2.jmp**

The data is stored in what's referred to as Outcome/Frequency Table format.

Screen	Three Screen Types (A, B, C)
Drop Height	Height from which screens were dropped (0.5m, 1.0m, 1.5m)
Tested	Number of screens tested
Outcome	Two outcomes (Success, Fail)
Count	Number of screens tested that resulted in either Success or Fail
Rate	Proportion of specimens that resulted in either Success or Fail

JMP Tips

To perform an analysis on a subset of the data in the data table, which will be used throughout these analyses, you'll need to filter the data. It can be done either through the Global Data Filter at the data table level before running an analysis or with the Local Data Filter within the output window from an analysis.

For Global Data Filter, select Rows>Data Filter. To choose a variable to filter, select the variable and click +. Choose Add to filter by additional variables. Select the values of the variables you wish to include in your analysis and then select the Show and Include boxes. Perform the analysis.

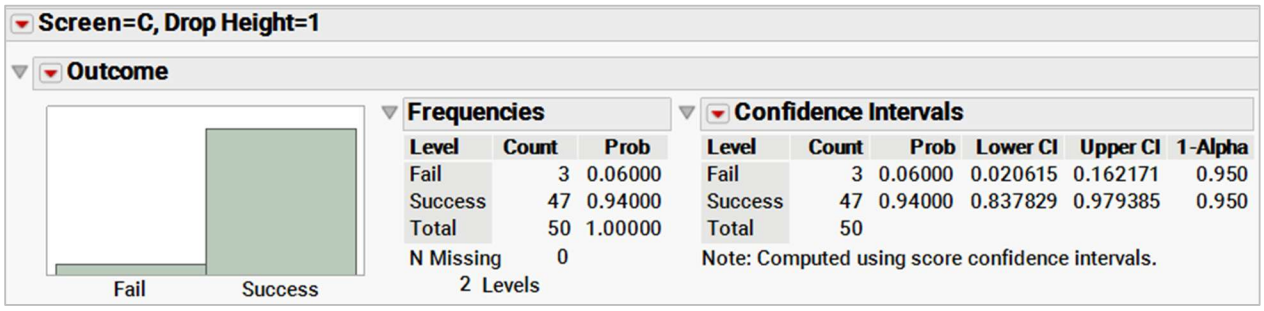
For Local Data Filter, first run the analysis. Then select Local Data Filter under the red triangle at the top of the output. To choose a variable to filter, select the variable and click +. Choose Add to filter by additional variables. Select the values of the variables you wish to subset the data by. The numerical and graphical results will change to correspond to the data being included in the analysis. The tools and analysis options available may change based on attributes of the data being included (e.g., analysis based on two groups versus three or more).

Analysis

Confidence interval for population proportion

We begin by calculating a 95% confidence interval for $p_{C,1.0}$ (i.e., the Population Success Rate for Screen Type C at the 1.0 meter Drop Height). Exhibit 1 shows a 95% confidence interval for the Population Success Rate as well as the Population Failure Rate.

Exhibit 1 Confidence for Success Rate



(To create, Analyze > Distribution. Select Outcome as the Y Variable, Count as the Freq Variable, and Screen and Drop Height as the By Variables. Choose Confidence Interval >0.95 from the red triangle next to the bar chart for each Screen type.)

Note: Holding down the Command key for Mac or Ctrl key for Windows while selecting an operation under the red triangle will perform that operation for all the groups specified in the By Variable.

The confidence interval [0.838, 0.979] translates to estimating with 95% confidence that the Population Success Rate for Screen Type C is between 88.3% and 98.0%.

Technical Note: The confidence intervals shown in Exhibit 1 are based upon a statistical procedure that uses the binomial distribution. Another common method used to estimate a population proportion is based upon the normal distribution. Results using that technique will be not identical to those shown here.

The exercises for this case study will ask you to create 95% confidence intervals of the Population Success Rate for all the remaining combinations of Screen Type and Height resulting in 3 x 3 = 9 confidence intervals.

Contingency table analyses for differences in population proportions

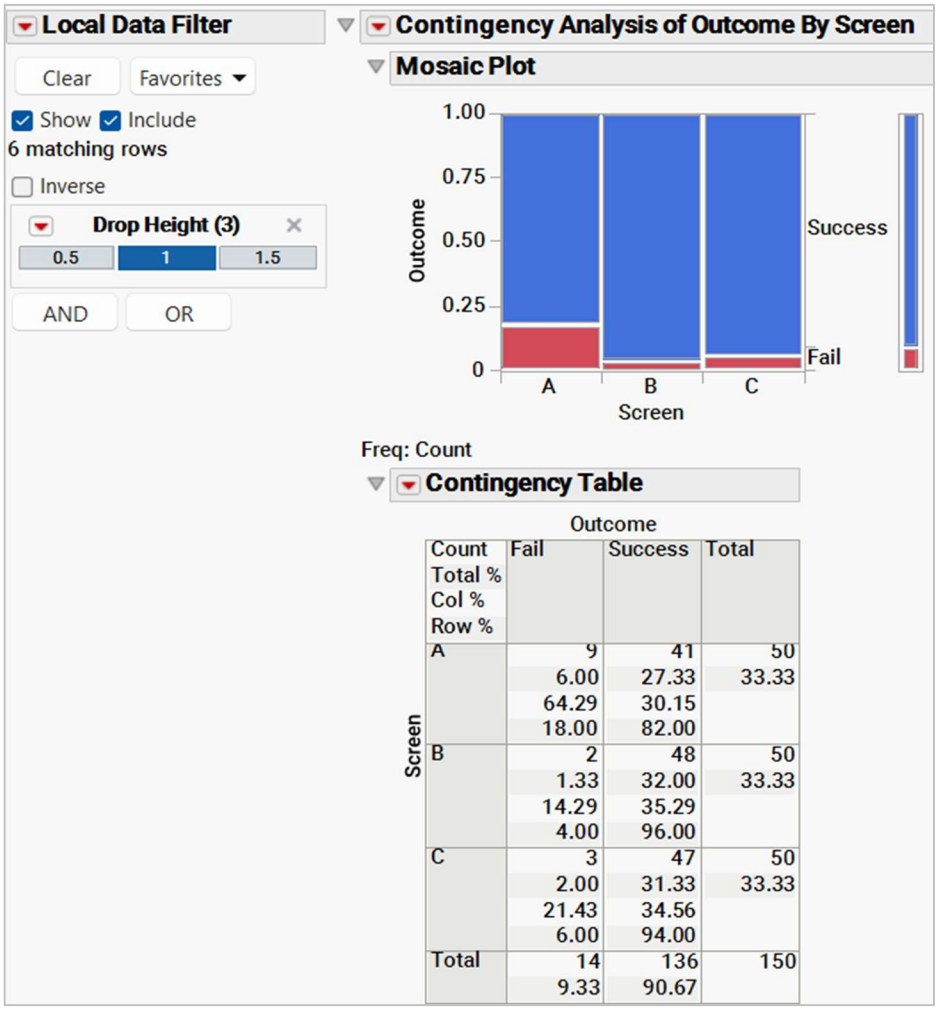
Comparison across the three Screen Types at a given Drop Height is another important analysis suitable for these data. The hypothesis test for that inquiry can be stated as:

$$H_O : p_{A,1.0} = p_{B,1.0} = p_{C,1.0}$$
$$H_A : p_{A,1.0}, p_{B,1.0}, \text{ and } p_{C,1.0} \text{ are not all equal}$$

where $p_{A,1.0}$, $p_{B,1.0}$ and $p_{C,1.0}$ are the Population Success Rates for Screens Types A, B, and C respectively at the 1.0m Drop Height.

Note that the alternative hypothesis H_A does not state that "*all three Success Rates are not equal,*" rather that *"the three Success Rates are not all equal."* Exhibit 2 displays a mosaic plot, contingency table analysis, and hypothesis tests for equality of the Success Rates across the three Screen Types.

Exhibit 2 Mosaic Plot and Contingency Table Analysis



Tests

N	DF	-LogLike	RSquare (U)
150	2	3.2121692	0.0690

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	6.424	0.0403*
Pearson	6.775	0.0338*

Warning: 20% of cells have expected count less than 5, ChiSquare suspect.

Fisher's Exact Test	Table Probability (P)	Two-Sided Prob ≤ P
	0.0034*	0.0630

(Subset the data either through the Global Data Filter at the data table level before running the analysis or Local Data Filter within the output window from the analysis. The subset of the data for this analysis should include all Screen Types and only the 1.0m Drop Height. See JMP Tips on page 3 for instructions.)

(To create, choose Analyze>Fit Y by X. Place Outcome in the Y Response Category, Screen in the X Factor Category, and Count in the Freq field. The red triangle in the contingency table output provides choices for the information to be displayed in the table cells.)

The p-values corresponding to a Chi-square test for independence of the Success Rate across the three Screen Types (0.0403 for the likelihood ratio method; 0.0338 for the Pearson method) are smaller than 0.05, a common significant level used in practice. Note, however, that JMP will provide a warning: *20% of cells have expected count less than 5. ChiSquare suspect*. This warning indicates that the test statistics and corresponding p-values may not be accurate for these data.

JMP Pro provides an alternative method known as the Fisher's Exact Test that can be more accurate in these scenarios. That test is also shown in Exhibit 2. The p-value for the Fisher's Exact Test is 0.0630. Though not at the 0.05 level of significance, it is quite close and would be considered significant at a more liberal cutoff level (e.g., 0.07). This is a common dilemma in statistical analyses – a significant result is not obtained at one level of significance commonly used (e.g., 0.05) but is significant at another (e.g., 0.10). It is best not to think of results of statistical analyses as a binary "yes" or "no" conclusion without any further inquiry, but instead as a continuum of an amount of evidence to support or refute a hypothesis. Many practicing statistical scientists would consider this as enough evidence to conclude a difference, just perhaps not at a level of confidence they would prefer. It is certainly enough evidence for more data collection and further study.

An examination of the mosaic plot provides an indication of where that difference might be, specifically Screen Type A performs the worst in the study with a Success Rate of 82%, while Screen Types B and C perform at 96% and 94% respectively.

In the exercises for this case study, you will perform a similar analysis for the 0.5m and 1.5m Drop Heights.

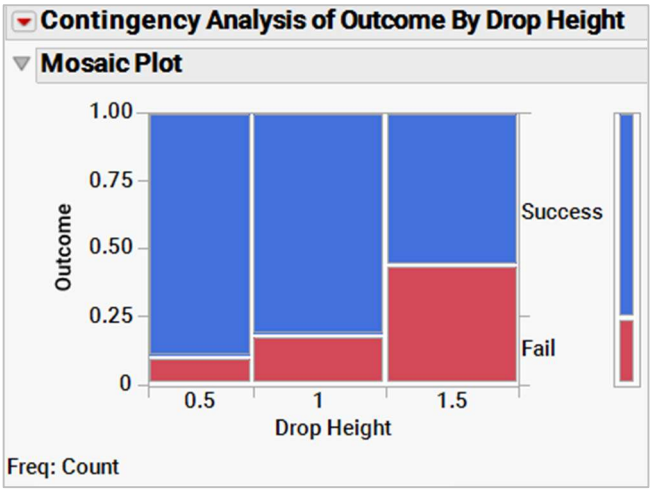
Another analysis that could be useful is to compare the Success Rates across the three Drop Heights for each of the Screen Types. The hypothesis for such a comparison for Screen Type A could be written as:

$H_O : p_{A,0.5} = p_{A,1.0} = p_{A,1.5}$
$H_A : p_{A,0.5}$, $p_{A,1.0}$, and $p_{A,1.5}$ are not all equal

where $p_{A,0.5}$, $p_{A,1.0}$ and $p_{A,1.5}$ are the Population Success Rates for Screen Type A at the 0.5m, 1.0m, and 1.5m Drop Heights.

Exhibit 3 shows the results of a Chi-square test for independence of the Success Rate across the three Drop Heights for Screen Type A.

Exhibit 3 Mosaic Plot and Contingency Table Analysis



Contingency Table

Count / Total % / Col % / Row %	Fail	Success	Total
Drop Height 0.5	4	36	40
	2.86	25.71	28.57
	11.43	34.29	
	10.00	90.00	
1	9	41	50
	6.43	29.29	35.71
	25.71	39.05	
	18.00	82.00	
1.5	22	28	50
	15.71	20.00	35.71
	62.86	26.67	
	44.00	56.00	
Total	35	105	140
	25.00	75.00	

Tests

N	DF	-LogLike	RSquare (U)
140	2	7.8574370	0.0998

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	15.715	0.0004*
Pearson	15.733	0.0004*

(Subset the data either through the Global Data Filter at the data table level before running the analysis or Local Data Filter within the output window from the analysis. The subset of the data for this analysis should include all Heights but only Screen Type A. See JMP Tips on page 3 for instructions.)

(To create, choose Analyze>Fit Y by X. Place Outcome in the Y Response Category, Height in the X Grouping Category, and Frequency in the Freq field. The red triangle in the contingency table output provides choices for the information to be displayed in the table cells. Fisher's Exact Test is available under the red triangle in JMP Pro (not shown here).)

The p-values corresponding to a Chi-square test of equality of the Success Rates across the three Drop Heights for Screen Type A is 0.0004 for the likelihood ratio method and 0.0004 for the Pearson method, which are both much smaller than the 0.05 significant level commonly used in practice. Examining the mosaic plot provides some details in that difference. Notice that the Success Rates in these data for Screen Type A is 90% at the 0.5m Drop Height, 82% at 1.0m, but only 56% at the 1.5m Drop Height.

In the exercises for this case study, you will perform a similar analysis for Screen Types B and C.

When a Chi-square test for independence produces the statistical evidence to indicate a difference across three or more categories, it can be useful to make comparisons for each of the possible pairs. To illustrate a pairwise comparison, we'll perform a hypothesis test and create a confidence interval for the difference in Success Rates between Screen Types A and C at the 1.0m Drop Height. The hypothesis of interest can be stated one of three ways, with each requiring a different statistical analysis technique and each providing a particular way to interpret the comparison.

Confidence intervals and hypothesis tests for differences in two proportions

One way to state the hypothesis of interest is:

$H_O : p_{A,1.0} = p_{C,1.0}$ or equivalently $H_O : p_{A,1.0} - p_{C,1.0} = 0$

$H_A : p_{A,1.0} \neq p_{C,1.0}$ $H_A : p_{A,1.0} - p_{C,1.0} \neq 0$

where $p_{A,1.0}$ and $p_{C,1.0}$ are the Population Success Rates for Screen Types A and C at the 1.0 meter Drop Height.

Exhibit 4 shows the results of a two proportions test.

Exhibit 4 Two Proportions Test

Two Sample Test for Proportions

Description	Proportion Difference	Lower 95%	Upper 95%
P(Success|A)-P(Success|C)	-0.12	-0.24469	0.013921

Adjusted Wald Test (Null Hypothesis)	Prob
P(Success|A)-P(Success|C) ≤ 0	0.9599
P(Success|A)-P(Success|C) ≥ 0	0.0401*
P(Success|A)-P(Success|C) = 0	0.0803

Response Outcome category of interest

○ Fail
● Success

(Subset the data either through the Global Data Filter at the data table level before running the analysis or Local Data Filter within the output window from the analysis. The subset of the data for this analysis should include Screen Types A and C, and only the 1.0m Drop Height. See JMP Tips on page 3 for instructions.)

(To create, choose Analyze>Fit Y by X. Place Outcome in the Y Response Category, Screen in the X Factor Category, and Frequency in the Freq field. Choose Two Sample Test for Proportions from the top red triangle.)

The p-value for the two-sided test is 0.0803, which is not significant at the 0.05 level but significant at the 0.10. This result suggests not a complete rejection of no difference, rather one that perhaps is but not at a desired level of confidence.

The confidence interval will lead us to the same conclusion. The confidence interval for the parameter $p_{A,1.0} - p_{C,1.0}$ is [-0.245, 0.014], which translates to estimating with 95% confidence that the Population Success Rate for Screen A is 0.245 units smaller and up to 0.014 units larger than the Success Rate of Screen C. Since 0 is a plausible value for $p_{A,1.0} - p_{C,1.0}$, these data do not demonstrate the statistical evidence at the 95% confidence level that there is a difference in the Success Rates. Notice, however, that the upper bound of the confidence interval is quite close to 0. A 90% confidence interval (output now shown) is [-0.224, -0.007], an interval not containing 0.

This is another example of a when a significant result is not obtained at one level of confidence commonly used (95%) but is significant at another confidence level (90%).

The exercises will ask you to use this technique to perform pairwise comparisons for the potential difference in the Success Rates between the three Screen Types for each of the three Drop Heights, as

well as for all pairwise comparisons between the three Drop Heights for each Screen Type. It will result in a total of (3 x 3) + (3 x 3) = 18 hypothesis tests and confidence intervals.

Confidence intervals for relative risk

Another approach for conducting a two-sample comparative analysis is to examine the ratio of the Success Rates, which is known as the relative risk. A hypothesis test framework for this ratio can be written as:

$$H_0: \frac{p_{A,1.0}}{p_{C,1.0}} = 1$$

$$H_A: \frac{p_{A,1.0}}{p_{C,1.0}} \neq 1$$

where $p_{A,1.0}$ and $p_{C,1.0}$ are the Population Success Rates for Screen Types A and C at the 1.0m Drop Height.

This approach phrases the parameter of interest as a percent increase or decrease between the two Success Rates versus the difference between them. It can be a useful way to describe the difference between two proportions. Exhibit 5 shows the results for that analysis.

Exhibit 5 Relative Risk

Relative Risk

Description	Relative Risk	Lower 95%	Upper 95%
P(Success|A)/P(Success|C)	0.87234	0.752677	1.011028

(To create, choose Relative Risk from the top red triangle. Choose desired response outcome and category in the numerator.)

The confidence interval [0.75, 1.01], which translates to estimating with 95% confidence that the Population Success Rate for Screen Type A is between 75.3% and 1.1% the size of the Population Success Rate for Screen Type C. Though this interval contains 1, the upper bound is above 1.0 by only 0.1. A 90% confidence interval for relative risk (not shown here) is [0.77, 0.99] indicating significance at the 90% confidence level, which is a similar result from our analysis of the absolute difference between the two Success Rates.

In the exercises, you will use this technique to perform pairwise comparisons for the potential difference in the Success Rates between the three Screen Types for each of the three Drop Heights, as well as all pairwise comparisons between the three Drop Heights for each Screen Type. It will result in a total of 18 confidence intervals.

Confidence intervals for the odds ratio

The last approach we'll illustrate is to evaluate the odds ratio. The odds for an outcome is the probability of the outcome occurring divided by the probability of it not occurring; it is written mathematically as:

$$\frac{p_A}{(1 - p_A)}$$

In this case, p_A would be the probability that Screen Type A would not be damaged in the drop test and $(1 - p_A)$ is the probability that it would be. For example, if the probability of a screen not being damaged is 0.90, then the probability of the screen being damaged is 0.10 and the odds would be 0.90/0.10 = 9. In

other words, the probability of a screen not being damaged is 9 times more likely than the probability of it being damaged.

The odds ratio is the ratio of two different odds. If, for example, the probability of Screen Type B not being damaged is 0.80, the odds of the screen not being damaged is 0.80/0.20 = 4. The odds ratio between the two Screen Types would be 9/4 = 2.25, meaning that the odds of Screen Type A not being damaged is 2.25 the size of the odds of screen Type B not being damaged.

An odds ratio equal to 1 occurs when the odds of success for the two Screen Types are the same, in other words, when the Success Rates are the same.

A hypothesis evaluating the odds ratio between the Population Success Rates of Screen Type A and C at the 1.0m Drop Height can be written as:

$$H_0: \frac{p_{A,1.0}/(1-p_{A,1.0})}{p_{C,1.0}/(1-p_{C,1.0})} = 1$$

$$H_A: \frac{p_{A,1.0}/(1-p_{A,1.0})}{p_{C,1.0}/(1-p_{C,1.0})} \neq 1$$

where $p_{A,1.0}$ and $p_{C,1.0}$ are the Population Success Rates for Screen Types A and C at the 1.0m Drop Height.

Exhibit 6 shows a 95% confidence interval for this odds ratio.

Exhibit 6 Odds Ratio

Odds Ratio

Odds Ratio	Lower 95%	Upper 95%
3.439024	0.87202	13.56263

(To create, choose Odds Ratio from the top red triangle. Choose desired response outcome and category in the numerator.)

The confidence interval for the odds ratio is [0.87, 13.56], thus we estimate with 95% confidence that the odds of Screen A succeeding at the 1.0m Drop Height is between 0.87 and 13.56 than the odds of Screen C succeeding. Note that 1 is a plausible value for the odds ratio, suggesting that there is not enough statistical evidence at the 95% confidence level indicating a difference between the two odds and ,consequently, the two Success Rates. A 90% confidence interval for the odds ratio (not shown here) is [1.09, 10.88] – an interval not containing 1.0, indicating significance at the 90% confidence level. This is another example of how an analysis may not generate enough statistical evidence at one level of confidence level but enough evidence at another level.

In the exercises for this case study, you will calculate confidence intervals for the odds ratios for all pairwise comparisons of the three Screen Types for each of the three Drop Heights, as well as all pairwise comparisons between the three Drop Heights for each Screen Type. This will result in a total of 18 confidence intervals.

Summary

Statistical insights

An analysis of the results for Screen Type A generated the statistical evidence to conclude that the Success Rates are not all equal across the three Drop Heights.

Three types of approaches were used to compare the Success Rates between Screen Types A and C at the 1.0m Drop Height:

- Difference in two proportions
- Relative risk
- Odds ratio

Our analyses showed that our test results did not produce the necessary statistical evidence at the 95% confidence level to conclude a difference in the Success Rates but did at the 90% confidence level. Many practicing statistical scientists would consider these results as sufficient evidence to conclude a difference, just perhaps not at a level of confidence they would prefer. This example illustrates that it is not uncommon in an analysis to lack the evidence of a statistical difference at one level of confidence but to have that evidence at a lower confidence level. It is certainly enough evidence to warrant additional data collection and study.

Implications

Recall that the tests generated the following data:

	0.5m	1.0m	1.5m
Type A	n=40 S=36 F=4	n=50 S=41 F=9	n=50 S=28 F=22
Type B	n=40 S=40 F=0	n=50 S=48 F=2	n=50 S=40 F=10
Type C	n=40 S=39 F=1	n=50 S=47 F=3	n=50 S=39 F=11

There are other additional comparative analyses that can be done with these data beyond the ones illustrated.

1. All possible comparisons between the three Screen Types at the 0.5m and 1.5m Drop Heights.
2. All possible comparisons between the three Drop Heights for each of the three Screen Types.

For each, we can perform these analyses based upon difference in two proportions, relative risk and odds ratios. In the following exercises, you will perform all of these analyses and summarize the findings.

Exercises

Use the **drop-test-2.jmp** data set to answer the following questions:

1. Calculate 95% confidence intervals for the Population Success Rates for the nine different combinations of Screen Type and Drop Height. Provide the statistical interpretation of the confidence intervals. Describe how the Success Rates compare across the three Screen Types and three Drop Heights. Create a graph that displays the point estimate of the Success Rates, as well as the lower and upper bounds of the confidence intervals.

   Hint: This table will need to be made manually. A good way to present the table is to have an individual row for each of the nine combinations of Screen Type and Drop Height and columns for each of the three estimates (lower bound, point estimate, and upper bound). To create the graph, however, the data needs to be in a stacked format, meaning all the values for the estimates need to be in one column, Drop Height in another column, and Screen Type in a third column. There are estimates for nine Success Rates, so this data table should have 27 rows. This table can be created from the first via the Stack Command.

2. Perform a contingency table analysis for each Drop Height to evaluate the hypothesis that the Success Rates are the same across the three Screen Types. Which Drop Heights show a statistically significant difference between Screen Types at the 0.05 significance level? Are there any results that are borderline significant? If so, what might be a recommendation?

3. Perform a contingency table analysis for each Screen Type to evaluate the hypothesis that the Success Rates are the same across the three Drop Heights. Which Screen Types show a statistically significant difference in Success Rates between Drop Heights at the 0.05 significance level? Are there any results that are borderline significant? If so, what might be a recommendation?

4. Calculate 95% confidence intervals and conduct a hypothesis tests for the difference between the Population Success Rates for each possible pairwise comparison between the three Screen Types for each of the three Drop Heights (e.g., A vs. B, A vs. C, and B vs. C for each 0.5 m, 1.0m, and 1.5m Drop Heights). Create a table and graphical display that summarize the results (i.e., table and graph for the p-values and a table and graph for the confidence intervals). Interpret the results. *Hint: Create data tables similar to those created in Exercise 1.*

5. Calculate 95% confidence intervals and conduct hypothesis tests for the difference between the Population Success Rates for each possible pairwise comparison between the three Drop Heights for each of the three Screen Types (e.g., 0.5m vs. 1.0m , 0.5m vs. 1.5m, and 1.0m vs. 1.5m for Screen Types A, B, and C). Create a table and graphical display that summarize the results (i.e., table and graph for the p-values and a table and graph for the confidence intervals). Interpret the results. *Hint: Create data tables similar to those created in Exercise 4.*

6. Calculate 95% confidence intervals for the population relative risk for each possible pairwise comparison between the three Screen Types for each of the three Drop Heights (e.g., A vs. B, A vs. C, and B vs. C for each 0.5 m, 1.0m, and 1.5m Drop Heights). Interpret the results for the comparisons that are considered statistically significant.

7. Calculate 95% confidence intervals for the population odds ratios for each possible pairwise comparison between the three Screen Types for each of the three Drop Heights (e.g., A vs. B, A vs. C, and B vs. C for each 0.5 m, 1.0m, and 1.5m Drop Heights). Can the odds ratios be calculated for every comparison? If not, why? Interpret the results for the comparisons that are considered statistically significant.

8. Provide a one-page executive summary of your conclusions by choosing only one visualization and writing no more than five brief bullet points. Do not use any statistical terminology.

9. What ideas and recommendations do you have for further study to improve evaluating the durability of the screens?