

JMP INTRO® Lab Activities

Lab Activity – Describing Categorical Data Graphically

Data Set: Denim

Denim is a data set for a fictitious company that manufactures jeans. The denim fabric they use is treated to make it appear “worn”. Although appearance of the fabric is important, the company is also concerned about the strength of the treated fabric.

Thread Wear Measured rates the wear of the fabric on a numeric scale from 1 to 10. The variable **Thread Wear** converts the numerical values into categories, with 1 to 4 being Low, 5 to 9 as Moderate and 10 as Severe. The variable **Thread Wear** is an Ordinal variable - Low, Moderate and Severe are categories with a natural order. If you display the distribution of a categorical variable (either ordinal or nominal), the displays offered will be different from the displays for a continuous variable.

Create a display of the distribution of **Thread Wear** using the **Analyze® Distribution** command. Categorical variables show a multicolored mosaic plot (which may be unfamiliar to you) under the bar graph.

Create a new report from your JMP INTRO display including:

- the bar graph (histogram) for **Thread Wear** with a count axis;
- the table of counts and probabilities (proportions in each category).

Describe the distribution of the fabric thread wear trial results, referring to specific aspects of the graph or numerical displays to support your answer. Remember to write your description in the context of the situation. Also include a description of what the mosaic plot shows.

Create a new display to analyze the distribution of **Thread Wear Measured** in this file using the **Analyze® Distribution** command. **Thread Wear Measured** contains the original results from the fabric thread wear tests that were measured on a scale from 1 to 10. Copy these displays into your report, and describe the shape of the distribution of the fabric thread wear as originally measured. Use appropriate graphical displays and numerical summaries to support your answer.

Briefly discuss the differences between the ordinal version of the results (**Thread Wear**) and the numerical version of the results (**Thread Wear Measured**). Might it have been better to use different categorical divisions to create Thread Wear, such as 1 to 3, 4 to 7, and 8 to 10?

Print a copy of your lab report. Proofread your copy, make corrections as needed, close JMP INTRO and the word processing program, and log off.

JMP INTRO® Lab Activities

Teacher Notes

Lab Activity – Describing Categorical Data Graphically

Objectives:

- Gain familiarity with some of the basic operational tools of JMP INTRO.
- Use JMP INTRO to display categorical data with a histogram, mosaic plot, table of frequencies and probabilities.
- Use JMP INTRO to review how to display a histogram, boxplot, stem and leaf plot, and summary statistics.

Time Required: Approximately 20 to 30 minutes (may be done as a homework assignment following Describing Numerical Data Graphically)

Materials:

- Describing Categorical Data Graphically student activity directions
- Denim data set

Prerequisites:

- Students should have experience producing numerical summaries and graphical displays by hand and/or with graphing calculators.
- Students should have basic knowledge of how to use a word processing program.
- Students should have completed Describing Categorical Data Graphically.

JMPINTRO Notes:

- Categorical variables (either ordinal or nominal) show a mosaic plot rather than a boxplot with the histogram (bar graph). A mosaic plot is essentially a stacked bar chart where each bar is proportional to the number of cases at each level of the variable. The probabilities in the bar have a sum of one (1).
- The default setting in JMP INTRO is for the table of frequencies and probabilities to be automatically displayed with the histogram.
- **Thread Wear Measured** is a numeric variable. By converting **Thread Wear Measured** to an ordinal variable, **Thread Wear**, the graphical and numerical output we get from JMP INTRO are different.
- The histogram for **Thread Wear Measured** has gaps between the bars like a bar chart. This is because JMP INTRO has created two bins for each integer increment of 1 on the axis. For instance, the two bins that represent the values between 1 and 2 are 1 to <1.5 and 1.5 to < 2. With only integer data values, the second bin (here 1.5 to < 2) is empty, creating the gap in the histogram.

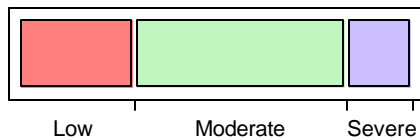
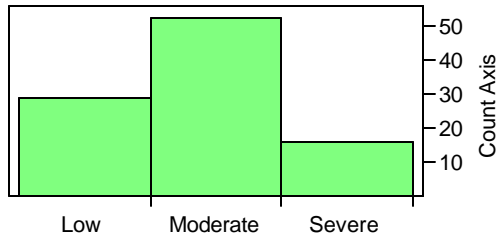
Discussion Questions:

- What are the differences between numerical and categorical data descriptions?
- What graphs and numerical descriptors are appropriate to use with each?

JMP INTRO® Lab Activities Answer Key

The following are suggested answers to Describing Categorical Data Graphically:

Thread Wear:



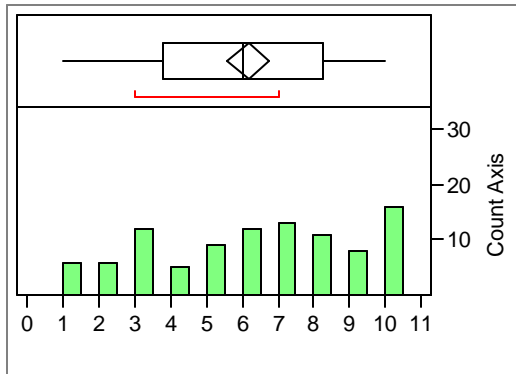
Frequencies

Level	Count	Prob
Low	29	0.29592
Moderate	53	0.54082
Severe	16	0.16327
Total	98	1.00000

3 Levels

The most common thread wear is Moderate, comprising about 54% of the results from the fabric swatches. Almost 30% of the treated fabric had low thread wear while over 16% had severe thread wear. It is not appropriate to describe the shape and numerical properties of ordinal data. You would be more interested in investigating the proportions and frequencies of each particular result as we have just done. One way to do this graphically is with a mosaic plot, which is essentially a stacked bar chart where each bar is proportional to the number of cases at each level of the variable.

Thread Wear Measured:



Quantiles

100.0%	maximum	10.000
75.0%	quartile	8.250
50.0%	median	6.000
25.0%	quartile	3.750
0.0%	minimum	1.000

Moments

Mean	6.1428571
Std Dev	2.8101436
N	98

Stem and Leaf

Stem	Leaf	Count
11		
10	0000000000000000	16
9	00000000	8
8	00000000000	11
7	0000000000000	13
6	000000000000	12
5	000000000	9
4	00000	5
3	000000000000	12
2	000000	6
1	000000	6

There appear to be three groups of responses to the fabric treatment: one centered around 3, one around 7 and a third at 10. The mean and median thread wear are approximately equal.