***Analyzing Data – Descriptive Statistics***

**Review**

What is a variable?

Levels of measurement

Types of research questions

**Reporting Data for Descriptive Research Questions – Means**

A variable must be interval or ratio level of measurement to calculate a mean.

**Example:**

A chain store that specializes in outdoor clothing is examining the young professionals' target market. Specifically, it wants to know how often young professionals purchase clothing designed for outdoor activities such as running, hiking, and winter sports.

**Research question:** How often do young professionals purchase clothing for outdoor activities?

The store surveys 338 randomly chosen young professionals and asks them to indicate how many outdoor clothing purchases they’ve made in the past six months.

**Analysis:** You would calculate the mean, minimum value, and maximum value for this variable in Excel.

**Excel results:**

Your calculated mean is 3.44, with a minimum of 0 and a maximum of 12

**Answer to the research question:** On average, young professions purchase 3.44 outdoor clothing purchases in a six-month period. In the sample of 338 young professionals, the minimum number of purchases in a six-month period was 0 and the maximum was 12.

**Example continued:**

The store also wants to adjust its merchandise to make it more appealing to young professionals, so it needs to know which of several brands are most appealing to this target market.

**Research question**: How appealing are the following brands to young professionals—Patagonia, Columbia, and North Face?

The survey asked young professionals to rate how appealing each of the brands is on a 7-point scale (1=not very appealing to 7=very appealing).

**Analysis:** You would calculate the mean, minimum value, and maximum value for the appealingness rating of each of the tree brands.

**Excel results:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Minimum | Maximum | Mean |
| Patagonia | 2.0 | 7.0 | 5.85 |
| Columbia | 2.0 | 7.0 | 4.81 |
| North Face | 3.0 | 7.0 | 5.43 |

**Answer to the research question:** Among the 338 young professionals surveyed, Patagonia scored highest in appeal of the three brands assessed (average = 5.85 on a 7-point scale) and Columbia faced the lowest (average = 4.81).

**Graphing a series of means**

**Things to note:**

**Reporting Percentages**

Percentages are calculated for variables that are nominal or ordinal level of measurement.

**Example:**

A local restaurant with two locations believes it should add another salad dressing to its salad bar and has narrowed it to three options: honey mustard, balsamic vinaigrette, and poppy seed.

**Research question**: Restaurant management wants to know which flavor is most preferred by customers.

The restaurant asks 275 randomly chosen customers to state which one of the three dressings they most prefer. Five people didn’t answer the question.

**Analysis:** You would obtain a frequency distribution of customer choices and calculate the percentage of people who chose each option.

**Excel results:**

|  |  |
| --- | --- |
| **Row Labels** | **Sum of Dressing Preference** |
| Honey mustard | 34.07% |
| Balsamic vinaigrette | 37.78% |
| Poppy seed | 28.15% |
| **Grand Total** | **100.00%** |

**Answer to the research question:** Among the 270 customers surveyed, balsamic vinaigrette was most frequently chosen as the dressing that should be added (by 37.78% of those surveyed) and poppy seed was chosen least frequently (by 28.15%).

**Graphing percentages** (next page)

**Things to note:**

***Crosstabulation for Comparing Subgroups (Differences RQ)***

Crosstabulation is useful for comparing subgroups, for example:

* different market segments
* different demographic groups
* different geographic areas

Crosstabulations are calculated for variables that are nominal or ordinal level of measurement.

**Salad dressing example**

We saw, above, that balsamic vinaigrette was the most preferred salad dressing overall.

However, restaurant management would also like to know if the customers at the two restaurant locations differ in which salad dressing they prefer.

**Research question**: Are dressing preferences the same in the two restaurant locations?

Crosstabulation allows us to find out.

***Honey Balsamic***

***Mustard Vinaigrette Poppy Seed Total***

South side location 37 60 43 140

West side location 55 42 33 130

Total 92 102 76 270

**Converting to percentages**

Percentages are usually easier to interpret than raw numbers

***Honey Mustard B. Vinaigrette Poppy Seed Total***

South side 37 60 43 140 (n)

**26.4% 42.9% 30.7% 100% (row %)**

40.2% 58.8% 56.6% (column %)

West side 55 42 33 130 (n)

**42.3% 32.3%** **25.4% 100% (row %)**

59.8% 41.2% 43.4% (column %)

Total 92 102 76 270

**How to choose whether to calculate row or column percentages**

* If the groups you’re comparing are represented in rows (as in our example), calculate row percentages.
* If the groups are represented in columns (as below), calculate column percentages.

***South side West side Total***

Honey mustard 37 55 92

Balsamic vinaigrette 60 42 102

Poppy seed 43 33 33

Total 140 130 270

**Excel results:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of Location** | **Column Labels** |  |  |  |
| **Row Labels** | **Honey mustard** | **Balsamic vinaigrette** | **Poppy seed** | **Grand Total** |
| South side | 26.42% | 42.86% | 30.71% | 100.00% |
| West side | 42.31% | 32.31% | 25.38% | 100.00% |
| **Grand Total** | **34.07%** | **37.78%** | **28.15%** | **100.00%** |

**Answer to the research question:** Among the 270 customers surveyed at the two restaurant locations, balsamic vinaigrette was most frequently chosen as the dressing that should be added by customers at the south side location, and honey mustard was most frequently chosen by customers at the west side location.

**Graphing a crosstabulation**

**Things to note:**