***Exploring Associations between Categorical Variables***

**Consider the statistical question: “Is there an association between texting while driving and gender?”**

**Data has already been collected from the teachers in this group through the CAMP Survey on the two variables of interest, *texting while driving* and *gender*. We will use our data to address this statistical question.**

**When analyzing possible relationships between two variables, it is helpful to identify one variable as the *independent* (or *explanatory*) *variable* and the other as the *dependent* (or *response*) *variable* if possible.**

* **Which of our two variables would you consider as the independent variable and which would you consider as the dependent variable? Why?**

**Tables and graphs can help us organize the data in a way that may reveal a possible association between the variables.**

* **Construct a two-way frequency table for the data.**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Gender** |  |
|  |  | **Female** | **Male** | **Total** |
| **Text while driving?** | **Yes** |  |  |  |
| **No** |  |  |  |
|  | **Total** |  |  |  |

* **What does the information in this table tell us about the two variables, *texting* *while driving* and *gender*, and their relationship?**
* **Construct *parallel bar graphs* and a *stacked* (or *segmented*) *bar graph* using appropriate conditional relative frequencies from the two-way table.**

**100%**

 **90%**

 **80%**

 **70%**

 **60%**

 **50%**

 **40%**

 **30%**

 **20%**

 **10%**

 **0%**

**100%**

 **90%**

 **80%**

 **70%**

 **60%**

 **50%**

 **40%**

 **30%**

 **20%**

 **10%**

 **0%**

* **What does the information in the bar graphs tell us about the two variables, *texting* *while driving* and *gender*, and their relationship?**

**What do the data suggest about our original statistical question: “Is there an association between texting while driving and gender?”**

***Which Hospital Would You Choose?***

(*adapted from Rossman and Chance, 2012*)

**The following two-way table classifies hypothetical hospital patients according to the hospital that treated them and whether they survived or died.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Survived** | **Died** | **Total** |
| **Hospital A** | **800** | **200** | **1000** |
| **Hospital B** | **900** | **100** | **1000** |

* **Based on this information, which hospital would you rather go to if you were ill? Why?**

**Suppose that when you further categorized each patient according to whether they were in fair condition or poor condition prior to treatment you obtain the following two-way tables.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Fair condition** | **Survived** | **Died** | **Total** |
| **Hospital A** | **590** | **10** | **600** |
| **Hospital B** | **870** | **30** | **900** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Poor condition** | **Survived** | **Died** | **Total** |
| **Hospital A** | **210** | **190** | **400** |
| **Hospital B** | **30** | **70** | **100** |

* **Based on this information, which hospital would you rather go to if you were ill? Why?**

**This phenomenon is called Simpson’s Paradox, which refers to the fact that aggregate proportions can reverse the direction of the relationship seen in the individual pieces.**