Last time we discussed:

* models and variability
* types of plots
* five-number summary

# Turning a Statistical Double Play

In the previous class we focused on the use of plots to summarize data, especially plots that have numerical data for both axes. Another tool of descriptive statistics is the *frequency table*, which is especially useful for categorical data. Frequency tables are more than just lists of numbers—they provide information that allows the data to be generalized and easily compared to other data or to models. There are two types: one-way and two-way frequency tables. Although both are useful, the two-way frequency table, once mastered, becomes a powerful method for testing models.

**One-Way Frequency Table**

A one-way frequency table lists data that can be **summed in one direction**.

It can also be used to list the *relative frequency*, *cumulative frequency*, and the *percentage* for each category.

**Activity 1:** Favorite colors of 1st graders (redux)

Based on previous data, how many students in a class of 30 would have each color as their favorite?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Color** | **Frequency** | **Relative Frequency** | **Cumulative Frequency** | **Percentage** (%) | **Prediction** (class of 30) |
| Blue | 6 | 0.30 | 0.30 | 30 |  |
| Red | 4 | 0.20 | 0.50 | 20 |  |
| Green | 3 |  |  |  |  |
| Pink | 2 | 0.10 |  | 10 |  |
| Purple | 2 |  |  |  |  |
| Yellow | 2 |  |  |  |  |
| Orange | 1 |  |  |  |  |
| **Total** | 20 | 1.0 |  | 100 | 30 |

**Frequency**—number of occurrences of each category. Column sums to the total number of occurrences.

**Relative Frequency**—frequency divided by total. Column sums to 1.0.

**Cumulative Frequency**—sum of all previous and current relative frequencies. Column is not summed.

**Percentage**—relative frequency multiplied by 100%.

**Question:** What model is used to make the prediction for the class of 30 students?

**Two-Way Frequency Table**

A two-way frequency table lists data that can be **summed in two directions**. A two-way table is used to study the relationship between two categorical variables. As with a one-way frequency table, entries can be given as frequencies, relative frequencies, or percentages.

**Example:** Preferred color of 20 1st graders by gender (redux-redux)

Which color do you like better, red or blue?

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Red** | **Blue** | **Gender Total** |
| **Girls** | 5 | 7 | 12 |
| **Boys** | 2 | 6 | 8 |
| **Color Total** | 7 | 13 | 20 |

**Marginal**

**Frequencies**

**Joint**

**Frequencies**

What is the relative frequency in the class of girls who prefer red?

What percentage of the class are boys who prefer blue?

**Two-way frequency tables are useful for determining conditional probabilities.**

Probability is the *expected percentage* for a given outcome.

What is the probability that a boy will like red?

What is the probability that a 1st grader who likes blue is a girl?