****

**Was “Stop and Frisk” Biased?**

Today, we’ll analyze 2011 data from New York’s “Stop and Frisk” program. The program allowed police officers to stop people on the street and search them for weapons or contraband. The program was controversial. Critics alleged that it led to heightened police discrimination of people of color. We will explore that claim using a **chi-square test for independence**.

Open the following links:

* The program data (may take some time to load): [tinyurl.com/stop-and-frisk-data](http://www.tinyurl.com/stop-and-frisk-data)
* Random number generator: [random.org/integers](http://www.random.org/integers)
* NYPD precinct map: [tinyurl.com/nypd-precincts](http://www.tinyurl.com/nypd-precincts)

*The data we’re using contains every stop made by NYPD police officers in 2011 (more than half a million stops in total). Each row represents a single stop. The second tab in the spreadsheet contains a data key and further information. Note: These data were reported by the police officers who made the stops.*

1. Use the random number generator to obtain 10 random integers between the values 2 and 632722 (the number of dataset rows). Find the dataset row numbers that correspond with your 10 random integers and record the following information for each selected stop:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Race |  |  |  |  |  |  |  |  |  |  |
| Force Level |  |  |  |  |  |  |  |  |  |  |

2. For one of your selected stops, look at all the variables listed: race, gender, and age of suspect along with the suspected crime and whether an arrest was actually made. If you’d like, use the precinct number and precinct map (linked above) to see the general area in which this stop took place. Write a two-sentence description of this stop, as if you’re writing a news brief for an article:

3. Often, police interactions are portrayed at an individual level in the news media, with vivid details given about the people involved and the interactions themselves. We are about to conduct a statistical analysis of many interactions, focusing solely on the relationship between race and force level used. What are the strengths and weaknesses of each type of analysis (individual & statistical)?

4. As a class, combine your samples to fill in the **observed** table. Then, calculate the **expected** counts. Note: all force levels other than “hands” (including push to wall/ground, handcuffs, draw/point weapon, pepper spray, baton) are all considered “higher level” force. The NYPD did not include police shootings in this dataset.

**Observed Expected**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Force** | Black | Hispanic | White | **Total** |
| None |  |  |  |  |
| Hands Only |  |  |  |  |
| Higher Level |  |  |  |  |
| **Total** |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Force** | Black | Hispanic | White | **Total** |
| None |  |  |  |  |
| Hands Only |  |  |  |  |
| Higher Level |  |  |  |  |
| **Total** |  |  |  |  |

5. Do the data provide convincing evidence of an association between race of suspects and the levels of force used by police officers? Use α = 0.05.

**STATE:**  Hypotheses: Significance level:

**PLAN:**  Name of procedure: **chi-square test for independence**

Check conditions:

**DO:**  Specific Formula:

Work: Picture:

Test statistic:

P-value:

**CONCLUDE**:

The **population** data (all 2011 stops):

**Raw Counts Conditional Distribution (By Race)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Force level** | **Black** | **Hispanic** | **White** |
| None |  | 76.7% |  |
| Hands Only |  | 16.9% |  |
| Higher Level |  | 6.4% |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Force Level** | **Black** | **Hispanic** | **White** | **Total** |
| None | 274,260 | 170,633 | 52,040 | 496,933 |
| Hands Only | 52,676 | 37,620 | 5,904 | 96,200 |
| Higher Level | 21,739 | 14,321 | 3,528 | 39,588 |
| **Total** | 348,675 | 222,574 | 61,472 | 632,721 |

*There is a substantial association between race of suspects and the force levels used in the population. Specifically, White suspects receive “no force” at a higher rate, and they receive higher level force at a lower rate.*

**Discussion Questions**

1. Were the results of your chi-square test consistent with the above conclusion about the population data? Explain.

2. Do these data provide enough evidence to prove that New York police are racially biased (in terms of use of force)? Why or why not?

3. (**Optional Extension Question**) Information about the residential population of NYC in 2011 is displayed below. Compare these counts to the data above. If police stops were done completely at random, which groups would be stopped proportionally more often? Less often? What conclusions can you draw from this?

|  |  |  |  |
| --- | --- | --- | --- |
| **Race** | **Black** | **Hispanic** | **White** |
| Population  Data source: 2011 American Community Survey (data.census.gov) | 2,054,101 | 2,373,304 | 2,731,173 |

Chi-Square Test for Independence

Important ideas:

Check Your Understanding

*For each of the following situations decide what type of chi square test is appropriate. Explain.*

1. A random sample of 200 students was asked to sample a new type of pizza that the school was considering using as a replacement for the current pizza. Each student stated if they were a freshman, sophomore, junior, or senior and also if they liked the new pizza more than the current pizza (or not). The school would like to know if there is a relationship between grade level and pizza opinion.
2. Another school is also considering changing their pizza vendor. This school selects separate random samples of 50 freshmen, 50 sophomores, 50 juniors, and 50 seniors. Each student tries the new pizza and states whether they like it more than the current pizza (or not). The school would like to know if the distribution of opinion differs across the grade levels.
3. A pizza shop claims that 30% of orders are placed on Fridays, 20% are placed on Saturdays, and 10% of orders are placed on the other days of the week. A global pandemic may have changed this distribution. The manager investigates so he knows how to staff the pizza shop appropriately. He selects a random sample of 300 orders and classifies each one according to the day of the week the order was placed. He wants to know if the distribution of orders is the same as it was before the global pandemic.