# Lesson 2

Tuesday 1/30/25

# Comment on Crime-Related Statistics

- We've spent some time discussing the limitations of important crime-related statistics.

- This is to underscore that all statistics have limitations and all interpretations of statistics require that we make some assumptions.

- An important scientific principle: any statistic or scientifically interesting quantity must be accompanied by a measure or description of the uncertainty about that statistic or quantity.

# Three Ways of Bringing Uncertainty Into Our Analysis

- (1) Considering how our answer might change depending on different kinds of data or collecting new data (triangulation & replication)

- (2) Bounds on our estimates due to missing information or measurement error.

- (3) Most frequently used: accounting for sampling error.

- Sampling error is what we usually focus on introductory statistics classes and is the focus of your textbook.

**U.S. Department of Justice**
Office of Justice Programs
*Bureau of Justice Statistics*

Program Report

July 2014 | NCJ 247060

# The Nation's Two Measures of Homicide

The United States uses two national data collection systems to track detailed information on homicides: the Federal Bureau of Investigation's Supplementary Homicide Reports and the Centers for Disease Control and Prevention's Fatal Injury Reports. Both measures were developed as part of a federal effort to improve national statistical systems in the early twentieth century and have gone through a number of changes since then to improve their consistency and coverage. Each program provides valuable information on the nature, trends, and patterns of homicides in the United States. Although the two measures generally capture information on the same types of events, they are designed for distinct purposes and collect different types of information. In combination, however, they produce a fairly comprehensive understanding of homicide, the most serious form of violence.

## Supplementary Homicide Reports

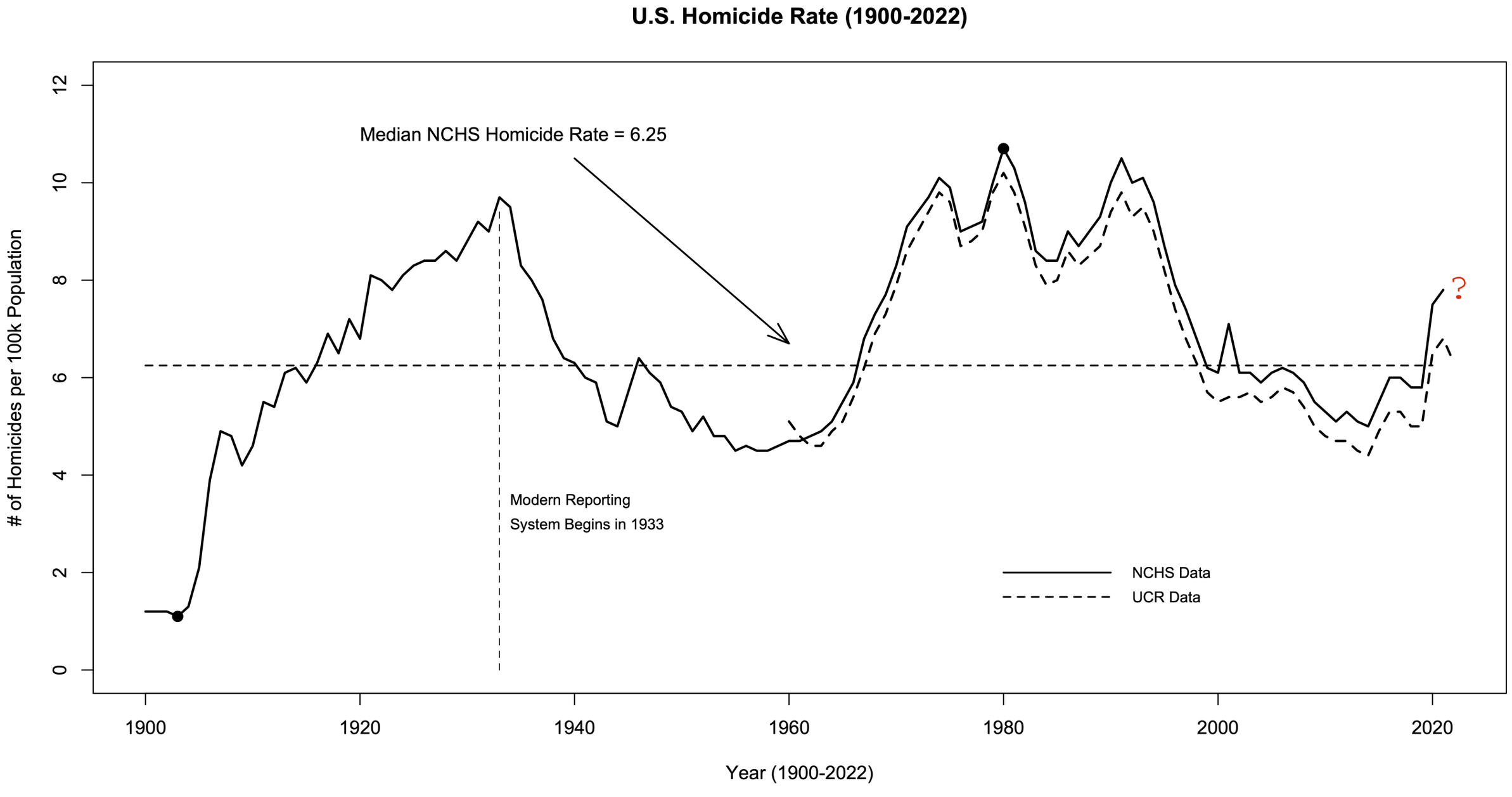
### Federal Bureau of Investigation

Supplementary Homicide Reports (SHR) are part of the

## National Vital Statistics System, Fatal Injury Reports

### Centers for Disease Control and Prevention

# Murder/Homicide Rates



# Missing Data/Nonresponse

- Brame et al. (2012) measured the fraction of people who self-reported that they had ever been arrested by the police for something besides a minor traffic offense -- by age 23.
- This is the data from the National Longitudinal Survey of Youth (1997) collected by the Bureau of Labor Statistics and the National Opinion Research Center (NORC).

| | Not Arrested | Arrested | Missing |
|---|---|---|---|
| Number of People | 4,299 | 1,858 | 1,178 |

- Analysis 1: assume the arrest rate for the observed and missing cases is the same: 1858/(4299+1858) = 0.302 or 30.2%.
- Analysis 2: calculate maximum and minimum possible arrest rate. Minimum = 1858/(4299+1858+1178) = 0.253 (25.3%); maximum = (1858+1178)/(4299+1858+1178) = 0.414 (41.4%).

# Random Sampling & Sampling Error

- The U.S. Bureau of Justice Statistics has, over the past 40 years, conducted several recidivism studies of people who have been released from U.S. state prisons.

- Recidivism refers to new offending among people who have offended in the past. One commonly used definition of recidivism in these studies is a new arrest that occurs within 3 years of prison release.

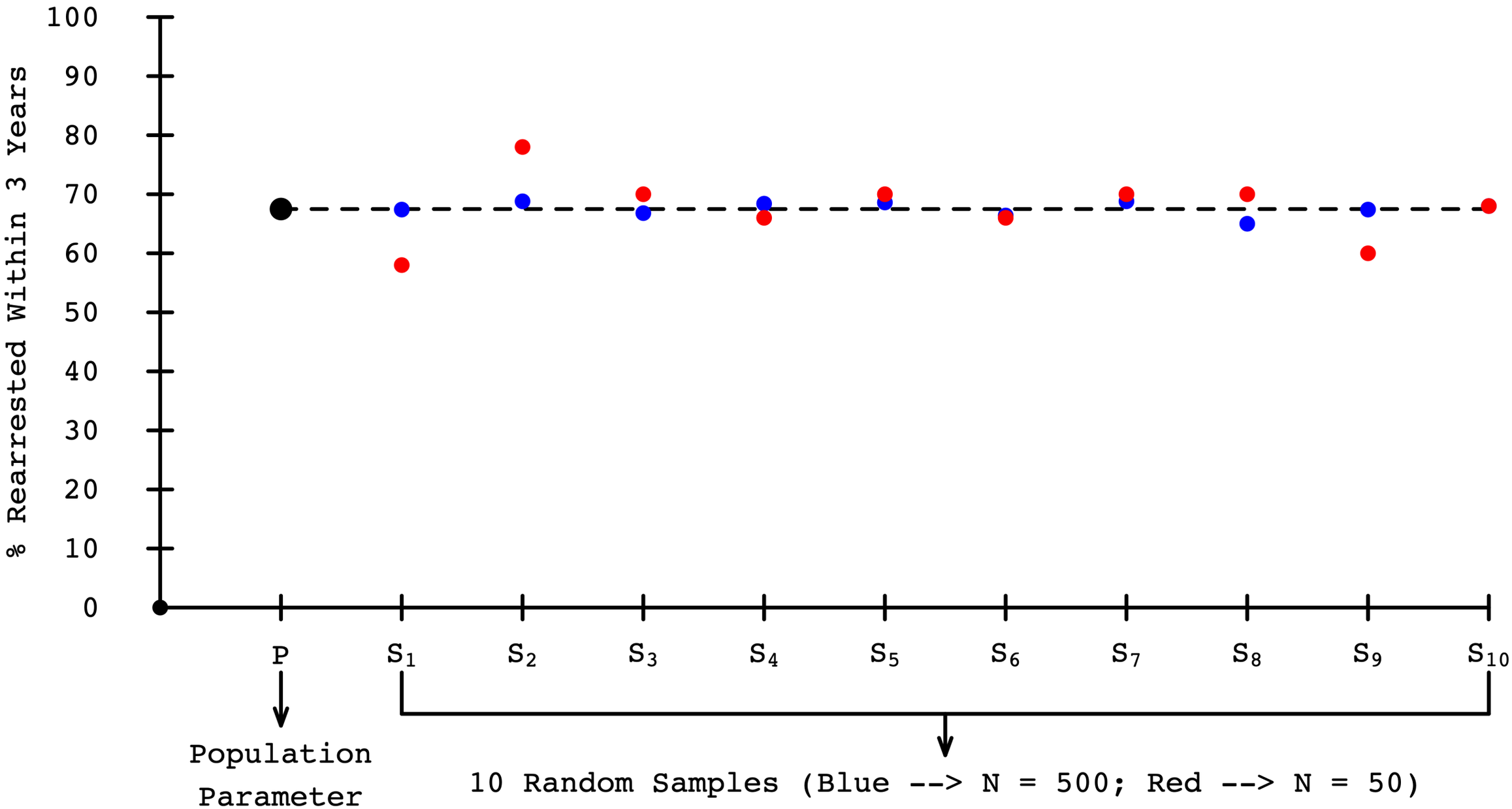
- A recidivism rate in these studies is based on the number of people arrested divided by the number of people released from prison; the fraction is then multiplied by 100 to place it on a percentage scale.

- The raw data for these studies is not publicly available but we can approximate it by thinking of a population of 1M prison releasees where 675,000 of these people got rearrested for a new offense within 3 years of release (a 67.5% recidivism rate; this is the "population parameter"). So, this is made-up data but it is based on realistic conditions.

- Now, let's suppose it is not realistic to study all 1M releasees because it would be too expensive and time consuming. Instead, we select a random sample of prison releasees and study the recidivism rate in the sample. What might we expect to see?

- The difference between the recidivism rate in the sample and the recidivism rate in the population is what we call "sampling error."

# Recidivism Rate Estimates From 20 Simple Random Samples




Note: a simple random sample means that every case in the population has the same chance of being selected to be in the sample.