Welcome to

The Session Will Begin Shortly

START

Statistics in R with Tidyverse

Session 7: Sampling

instats

Statistics vs Parameters

- Definitions
    - Statistics: estimates from a sample
    - Parameters: true values in the population
- Importance of distinction
    - Accuracy and reliability of inferences
    - Application to real-world data analysis

instats

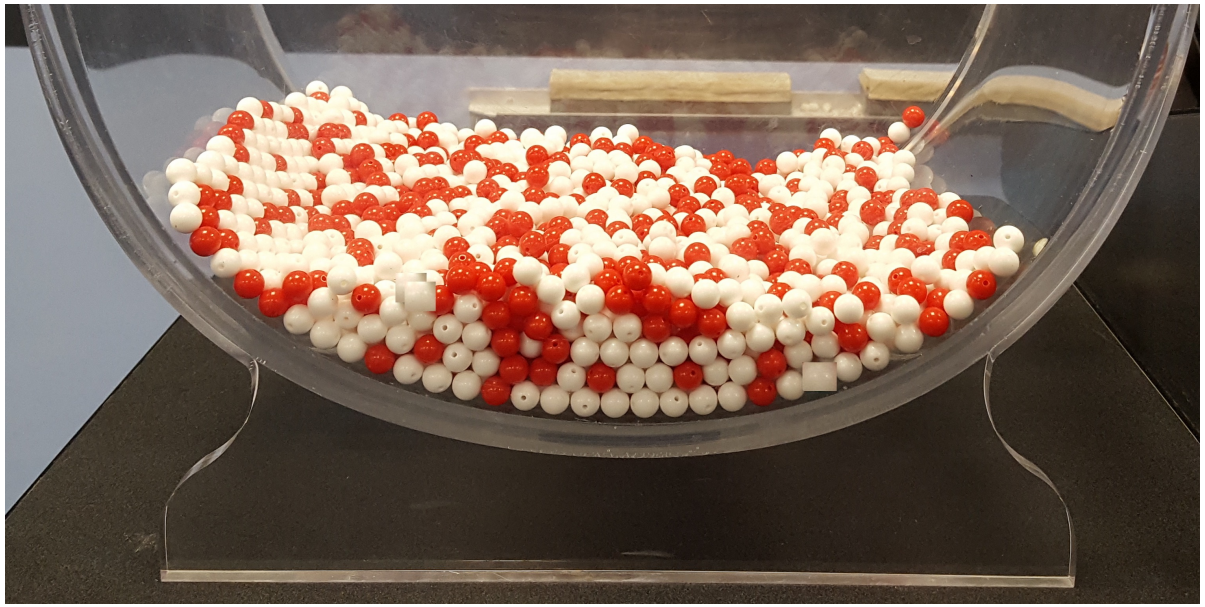
Sampling

- Randomly selecting samples from a population
- Focuses on drawing multiple samples to study variability
- Key concept: population vs. sample

instats

Population Proportion

- The ratio of subjects with a characteristic to the entire population
- Compute population proportion using `dplyr`
- **Tip**: Use `mean()` to find proportions of logical values



instats

Sampling Variation

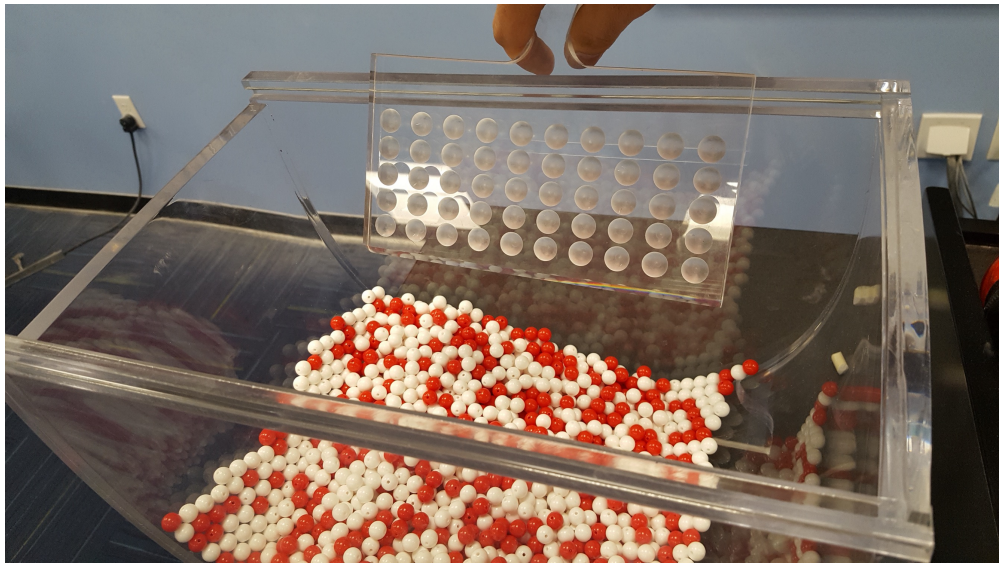
- Different samples produce different statistics due to random selection
- Key concept: Sampling with replacement and random mixing
- Explore sampling distribution with multiple replicates
- **Tip**: Histograms visualize the variation across samples

Central Limit Theorem (CLT)

- States that sample means will be approximately normally distributed
- Applies even if the population distribution isn’t normal
- Key concept: The larger the sample, the better the normal approximation
- **Tip**: Look for bell-shaped histograms to confirm CLT

Sample Proportion

- The proportion of a particular outcome in a sample
- Important to understand sampling distribution of proportions
- **Tip**: Use `rep_slice_sample()` for repeated sampling



instats

Standard Error

- A measure of how much sample statistics (e.g., sample means or proportions) vary from the population parameter
- Decreases with increasing sample size
- **Tip**: The smaller the SE, the more precise the sample statistic

Virtual Sampling with Different Sample Sizes

- Study how the sampling distribution changes with varying sample sizes
- Smaller samples lead to greater variability
- Larger samples reduce variability and give a better estimate of the population parameter
- **Tip**: Visualize how the width of the distribution narrows with larger sample sizes

instats

Demo & Exercises

Q & A

STOP