

---

**MATH1401**

**Fall 2021**

# Lecture 15

---

Sampling

---

# Discussion Question

---

A population has 100 people, including Rick and Morty. We sample two people at random without replacement.

- (a)  $P(\text{both Rick and Morty are in the sample})$
  - (b)  $P(\text{neither Rick nor Morty is in the sample})$
-

# Discussion Question

---

A population has 100 people, including Rick and Morty. We sample two people at random without replacement.

(a)  $P(\text{both Rick and Morty are in the sample})$

$= P(\text{first Rick, then Morty}) + P(\text{first Morty, then Rick})$

$= (1/100) * (1/99) + (1/100) * (1/99) = 0.0002$

(b)  $P(\text{neither Rick nor Morty is in the sample})$

$= (98/100) * (97/99) = 0.9602$

---

# Sampling

# Random Samples

---

- Deterministic sample:
  - Sampling scheme doesn't involve chance
- Random sample:
  - Before the sample is drawn, you have to know the selection probability of every group of people in the population
  - NOTE: Not all individuals / groups have to have equal chance of being selected

(Demo)

---

# Sample of Convenience

---

- Example: sample consists of whoever walks by
  - Just because you think you're sampling "randomly", doesn't mean you have a random sample.
  - If you can't figure out ahead of time
    - what's the population
    - what's the chance of selection, for each group in the population then you don't have a random sample
-

# Distributions

# Probability Distribution

---

- Random quantity with various possible values
  - “Probability distribution”:
    - All the possible values of the quantity
    - The probability of each of those values
  - If you can do the math, you can work out the probability distribution without ever simulating it
  - But... simulation is often easier!
-



# Empirical Distribution

---

- “Empirical”: based on observations
- Observations can be from repetitions of an experiment
- “Empirical Distribution”
  - All observed values
  - The proportion of times each value appears

(Demo)

---

# Large Random Samples

# Law of Averages / Law of Large Numbers

---

If a chance experiment is repeated many times, independently and under the same conditions, then the proportion of times that an event occurs gets closer to the theoretical probability of the event

As you increase the number of rolls of a die, the proportion of times you see the face with five spots gets closer to  $1/6$

---

# Empirical Distribution of a Sample

---

If the sample size is large,  
then the empirical distribution of a uniform random sample  
resembles the distribution of the population,  
with high probability

(Demo)

---

# A Statistic

# Inference

---

- **Statistical Inference:**

Making conclusions based on data in random samples

- **Example:**

fixed

Use the data to guess the value of an unknown number

depends on the random sample

Create an **estimate** of the unknown quantity

---

# Terminology

---

- **Parameter**
  - A number associated with the population
- **Statistic**
  - A number calculated from the sample

A statistic can be used as an **estimate** of a parameter

(Demo)

---

# Probability Distribution of a Statistic

---

- Values of a statistic vary because random samples vary
  - “Sampling distribution” or “probability distribution” of the statistic:
    - All possible values of the statistic,
    - and all the corresponding probabilities
  - Can be hard to calculate
    - Either have to do the math
    - Or have to generate all possible samples and calculate the statistic based on each sample
-



# Empirical Distribution of a Statistic

---

- Empirical distribution of the statistic:
    - Based on simulated values of the statistic
    - Consists of all the observed values of the statistic,
    - and the proportion of times each value appeared
  - Good approximation to the probability distribution of the statistic
    - if the number of repetitions in the simulation is large
-