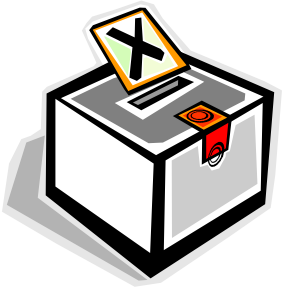
Sampling Designs

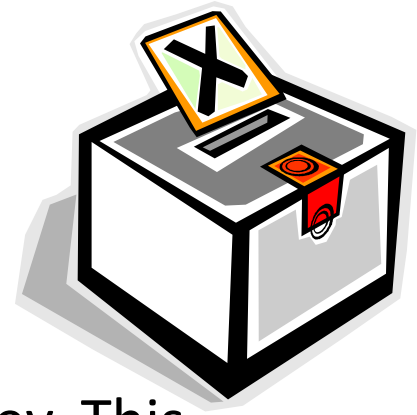
Poll on Course Scheduling System



- The Student Association at a given university wishes to poll the students on how they feel about a new course scheduling system proposed by the university administration.
- With this system most courses would be offered in 2-3 hour blocks only 2 days per week rather than in 50 minute blocks 4-5 days per week.
- With the assistance of the Registrar's office, the Student Association is able to receive anonymous responses from 500 students randomly selected among students registering for classes.

Sampling Designs

Poll on Course Scheduling System (cont)



- At another university they are doing a similar survey. This university has 1000 seniors, 1200 juniors, 1300 sophomores and 1500 freshmen.
- With the assistance of the Registrar's office, the Student Association is able to receive anonymous responses from 100 seniors, 120 juniors, 130 sophomores and 150 freshmen, each sample randomly selected within the corresponding subpopulation.

Sampling Designs

Poll on Course Scheduling System (cont)

- At a third university they are doing a similar survey, but are unable to get assistance from the Registrar's office.
- They randomly select 5 buildings in which classes are offered. In each building they visit every 4th classroom. In each classroom they randomly select 10 male and 10 female students.

Ethics

BASIC DATA ETHICS

The organization that carries out the study must have an **institutional review board** that reviews all planned studies in advance in order to protect the subjects from possible harm.

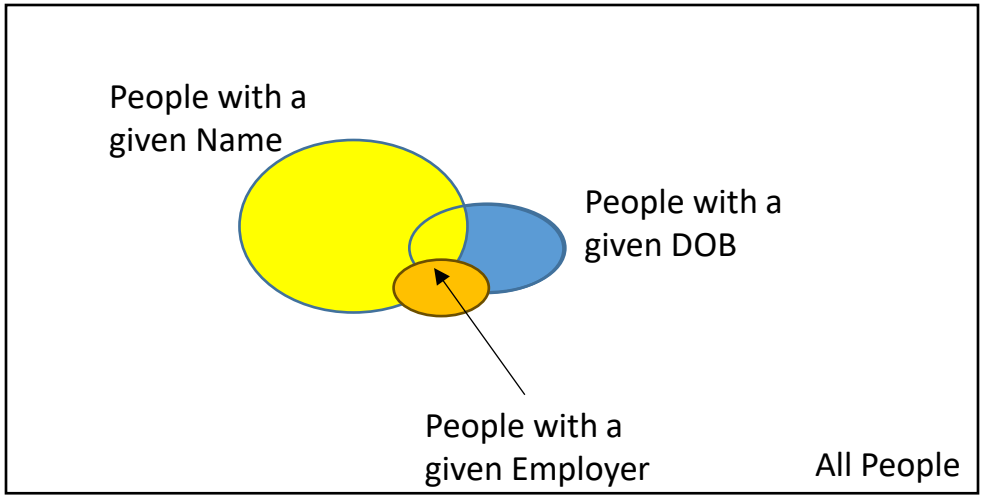
All individuals who are subjects in a study must give their **informed consent** before data are collected.

All individual data must be kept **confidential.** Only statistical summaries for groups of subjects may be made public.

Personally Identifiable Information (PII)

The PII in the Sky

- Personally identifiable information (PII) is any data that could potentially be used to identify a particular person.
- Not all PII uniquely identify a person, but combined with other PII may offer “clues” to possibly make an identification.
- PII is often stored in the Cloud, thus the “PII in the Sky”
- Examples:
  - Personal ID Number
  - DNA
  - Fingerprint
  - Name
  - Date of Birth
  - Nationality
  - Address
  - Age
  - Employer



PII Regulation

- Each individual country has its own set of regulations when it comes to protecting PII
- In the United States PII is protected through the following:
  - Gramm-Leach-Bliley Act – for financial information;
  - Fair Credit Reporting Act (FCRA) – regulating how consumer reporting agencies use credit information;
  - Health Insurance Portability and Accountability Act (HIPAA) and Health Information Technology for Economic and Clinical Health Act (HITECH) – for healthcare related information;
  - The Family Educational Rights and Privacy Act (FERPA) – relating to PII protections for student educational records;
  - The Children's Online Privacy Protection Act (COPPA) – relating to the privacy of children under 13;
  - The Privacy Act of 1974 – requiring fair information practices regarding PII held by federal agencies;
- The European General Data Protection Regulation (GDPR) is undoubtedly the most restrictive set of regulations among all countries in the world.

SOURCE: Regerlaw.com

https://inside.ewu.edu/records-and-registration/ferpa/

EWU FERPA policy WAC 172-191

Family Educational Rights and Privacy Act of 1974 (FERPA)

In 1974, Senator James Buckley of New York was concerned with the misuse and abuse of student records, particularly at elementary and secondary schools. In several instances, some schools revealed confidential information about their students to people who did not have a legitimate interest in the material. Some of these student files contained comments that were irrelevant or unsupported, such as statements about a student's psychological makeup by a person not qualified to make such an assessment. Sen. Buckley introduced legislation to protect the confidentiality and accuracy of student records.

FERPA for Post-Secondary Students

The Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99), also known as the Buckley Amendment, is a Federal law that protects the privacy of student education records. The law applies to all schools that receive funds under an applicable program of the U.S. Department of Education.

FERPA rights transfer from the parent/guardian to the student when he or she reaches the age of 18 or attends a school beyond the high school level.

- Directory Information +
- Adding a Directory Restriction +
- Release of Information (ROI) +
- Removing a Directory Restriction +
- Student Rights Under FERPA +
- FERPA Applies Only To Educational Records +
- Additional FERPA Information +

Clinical Trials

- Clinical trials study the effectiveness of medical treatments on actual patients – these treatments can harm as well as heal.
- Points for a discussion:
  - Randomized comparative experiments are the only way to see the true effects of new treatments.
  - Most benefits of clinical trials go to future patients. We must balance future benefits against present risks.
  - The interests of the subject must always prevail over the interests of science and society.
- In the 1930s, the Public Health Service Tuskegee study recruited 399 poor blacks with syphilis and 201 without the disease in order to observe how syphilis progressed without treatment. The Public Health Service prevented any treatment until word leaked out and forced an end to the study in the 1970s.

Behavioral and Social Science Experiments

- Many behavioral experiments rely on hiding the true purpose of the study.
- Subjects would change their behavior if told in advance what investigators were looking for.
- The "Ethical Principles" of the American Psychological Association require consent unless a study merely observes behavior in a public space.

The Three-Step Process in Statistical Analysis.

Step 1: Collecting the Data

- *Sampling design*
- *Experimental design*
- *Observational study*

Step 2: Summarizing/Organizing the Data

- *Descriptive statistics*
- *Graphical displays*
- *Numerical measures*

Step 3: Drawing Conclusions from the Data

- *Inferential Statistics*
- *Estimation of unknown parameters*
- *Hypothesis testing*

Some Basic Probability Terminology

Random Phenomenon: A phenomenon is random if the individual outcomes are uncertain, but nonetherless there is a regular distribution of outcomes in a large number of repetitions.

Sample Space: The collection of all possible outcomes.

Symbol: **S = {.,.,.,.,.}**

Event: A collection of outcomes (a subset of the sample space).

Symbols: **A, B, C** …

Probability: P(A) = A measure of the likelihood that A will occur
(a number between 0 and 1)

Probability Model: A sample space and the probabilities of all associated outcomes

A Random Phenomenon

Roll a balanced Die

Sample Space

Events:
A = An even number is rolled
B = A number greater than 4 is rolled

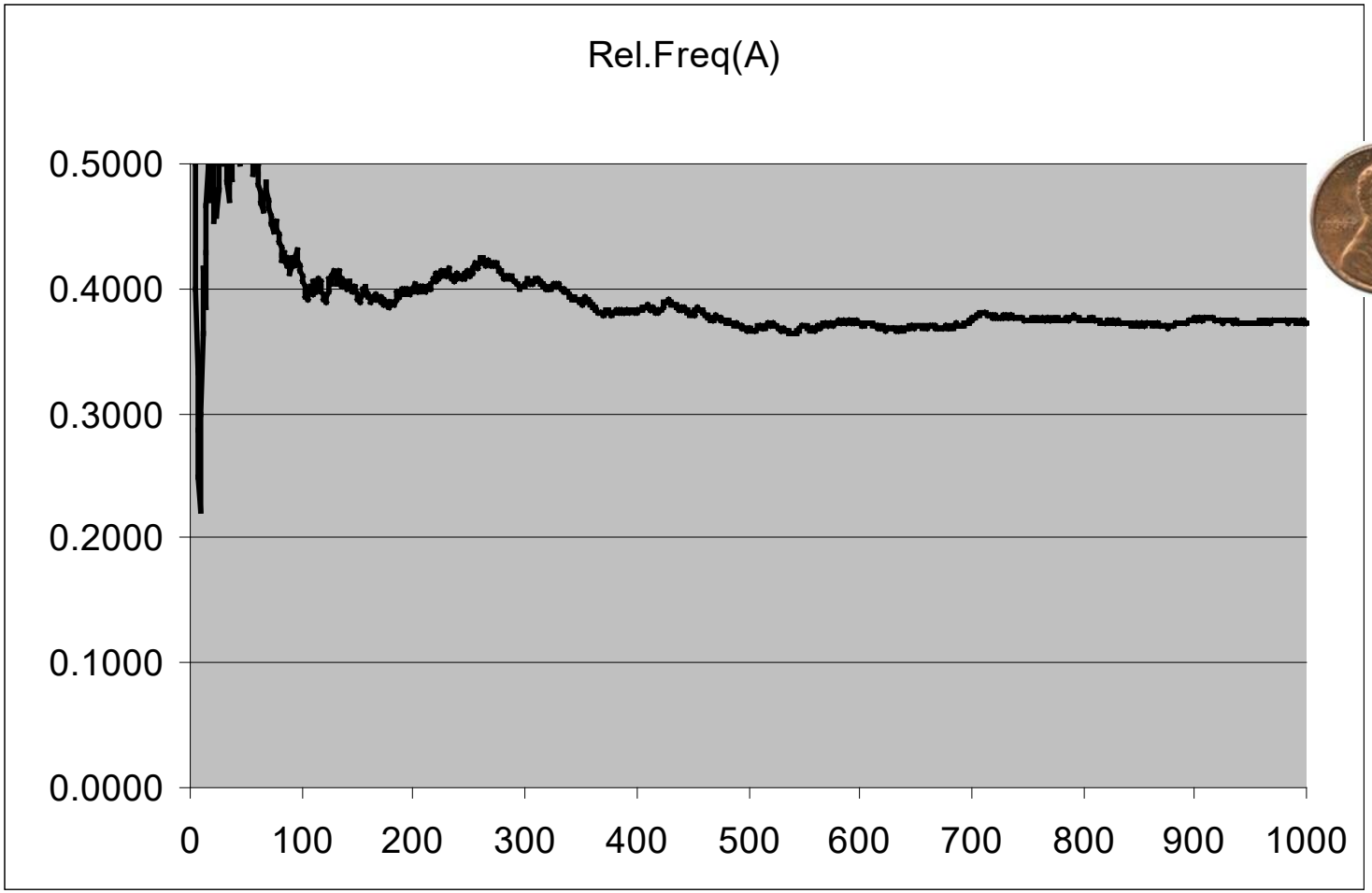
Probability:
P(A) =
P(B) =

A Random Phenomenon

- Toss a coin four times
- Define the event:
  A = Two heads are observed

$P(A)$ =

Long-term Relative Frequency



Experiment: Toss a Coin 4 times A = Two heads

Two Definitions of Probability

Classical Definition (Blaise Pascal): The probability of an event A represents the long-term relative frequency of the event when the experiment is repeated many times.

Bayesian Definition : The probability of an event A represents the degree of belief that the event A will occur.