STAT 3743: Probability and Statistics

G. Jay Kerns, Youngstown State University

Fall 2010

Notes

What are statistics?

- L. *status* $\longrightarrow$ a “standing”, or “condition”
- 1700’s Germans: “Statistik” $\rightsquigarrow$ Political Science
- each datum $\longrightarrow$ *statistic*
- all data $\longrightarrow$ statistics

Notes

What are statistics?

Statistics (loosely): *decision making under uncertainty*

Definition.
Statistics is that branch of knowledge which deals with the *multiplicity of data*, its

1. *collection*,
2. *analysis*, and
3. *interpretation*[a]

[a] *Information-Statistical Data Mining: Warehouse Integration with Examples of Oracle Basics (The Springer International Series in Engineering and Computer Science) by Bon K. Sy and Arjun K. Gupta (Nov 30, 2003)*

Notes

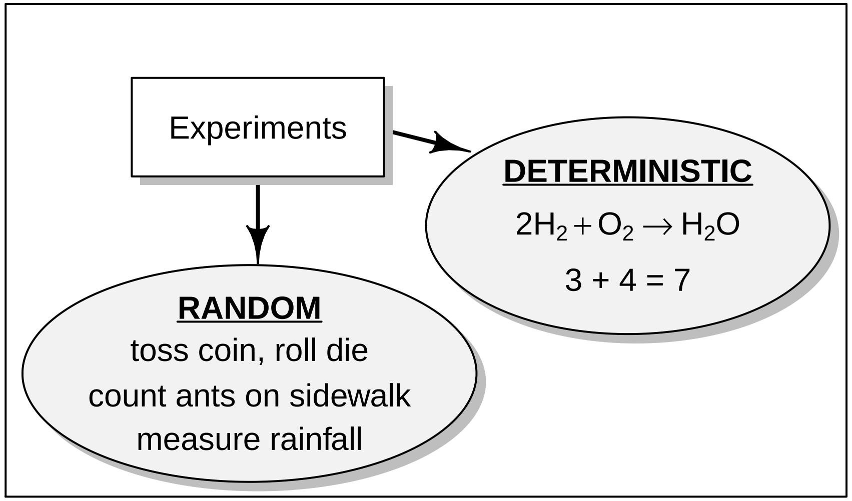

Figure: Two types of experiments

Notes

Random experiments

Definition.
The **sample space** is the set of *all possible* outcomes. It is denoted by S.

Examples

- Toss a coin
  $S =$
- number of pets owned by students in class
  $S =$
- color of a student's eyes
  $S =$

Notes

Random experiments

- outcomes associated w/ random experiments called **random variables**: X, Y, Z, *etc.*
- observed values: x, y, z

Do a Random Experiment:

Notes

Random experiments

Make a Frequency Table:

Notes

Random experiments

Make a Frequency Histogram:

Notes

Random experiments

- In general, do random experiment n times
- For outcome x, get frequency f_x
- Turns out, f_x can be crazy for small values of n
- However,
$$\lim_{n \to \infty} \frac{f_x}{n} = p(x),$$
where $p(x)$ is the "probability of outcome x"
- p is the **probability mass function** (PMF) of X,
$$p_X(x) = \mathbb{P}(X = x), \quad \text{for } x \in S$$

Notes

Random variable characteristics

Let X be a r.v. taking values in the sample space

$$S = \{x_1, x_2, \ldots, x_k\}$$

Then
$$p(x_i) = \mathbb{P}(X = x_i), \quad i = 1, 2, \ldots, k$$

And
$$\begin{aligned} \sum_{i=1}^{k} p(x_i) &= p(x_1) + p(x_2) + \cdots + p(x_k), \\ &= 1 \end{aligned}$$

Notes

Random variable characteristics

The **mean** of the r.v. X is

$$\mu = \sum_{i=1}^{k} x_i p(x_i), = x_1 p(x_1) + x_2 p(x_2) + \cdots + x_k p(x_k)$$

Remarks:

- the mean is a center or average value of X
- μ can be any number or decimal
- μ is the "first moment about the origin"

Notes

Random variable characteristics

Example

Pick a chip out of an urn. The urn has ____ chips labeled "1", ____ labeled "2" and ____ labeled "3".

$$X = \text{number listed on chip}$$

Notes

Random variable characteristics

The **variance** of the r.v. X is

$$\sigma^2 = \sum_{i=1}^{k} (x_i - \mu)^2 p(x_i),$$

Compute it for above example:

Notes

Random variable characteristics

- mean measures CENTER, variance measures SPREAD
- $\sigma^2 \geq 0$
- $\sigma = \sqrt{\sigma^2}$ is the **standard deviation**

Shortcut:

Notes

Other characteristics

Do random experiment n times, observe $x_1, x_2, \ldots, x_n$.

Definition.
The **empirical distribution** puts mass $1/n$ on each of the values $x_1, x_2, \ldots, x_n$.

The mean of the EDstn:

Notes

Other characteristics

Variance of the Empirical Distribution:

$$v = \sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot \frac{1}{n}$$

Sample variance:

$$v = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Sample standard deviation:

$$s = \sqrt{s^2}$$

Notes

Examples of random variables

Write down a bunch of random variables:

Notes

Notes