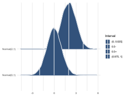
Chapter 9 (Part 2) Notes

Random Variables and Density Curves
STP-231

Arizona State University

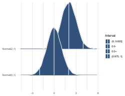


Relative Frequency Hist./Density Curves

- Consider a relative frequency histogram as an approximation of the underlying true

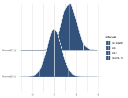
- It is often desirable to describe a population frequency distribution with a smooth curve (especially for continuous variables)

- We can idealize a density curve as a relative frequency histogram with very narrow classes



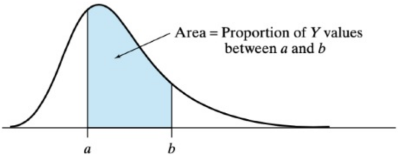
Density Curves

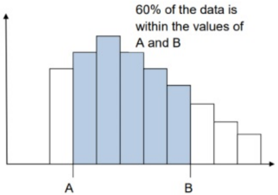
- **Density Curve**: a smooth curve representing a frequency distribution
- **Density Scale**:
- The plot of the vertical coordinates on the density curve are plotted on this scale
- Relative frequencies are represented as areas under the curve

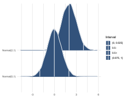


Relative Frequency Expansion

- Relative frequencies are probabilities
- Relative frequencies add up to one
- The area under a density curve or a relative frequency histogram adds up to one
- We can use this to find proportions between particular values

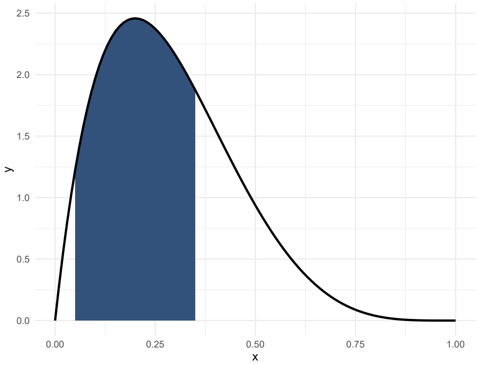






Distribution

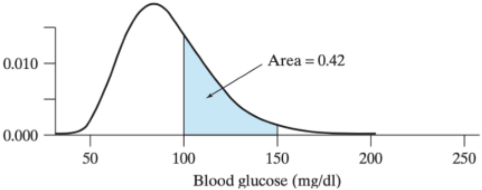
- For any two numbers A and B, the area under the density curve between A and B is the same as the proportion of $y-$values between A and B
- The total area under the curve is 1



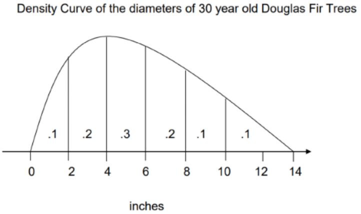
Probabilities and Density Curves

- The relative frequency of a specific Y value is zero, i.e. $\Pr(Y = \#) = 0$. Therefore, we do not discuss the relative frequency of a single Y value
- However, we can assign probabilities to intervals, defined as the area under the curve

$$\Pr(A \le Y \le B) = \Pr(A < Y < B)$$

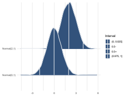


Example



- Find the probabilities $\Pr(d = 4)$, $\Pr(0 < d < 4)$, $\Pr(d < 6)$, $\Pr(d > 8)$, and $\Pr(d < 10)$
- Now assume we take a sample of two trees, which we consider as independent events. Find the probability:
- both trees have diameter less than 4".
- Diameter of first tree less than 8", 2nd tree greater than 8".
- Exactly one tree has diameter less than 4" and exactly one tree has diameter greater than 8".

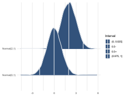
Random Variables



Random Variables

A **random variable** is a variable whose value is a numerical outcome of a random phenomenon

- Can either represent discrete or continuous values
- Example: In a population of flies, the random variable X represents the amount of flies that are still alive after 24 hours. X takes on values $\{0, 1, 2, \ldots\}$
- Random number generatror generates numbers in the interval [0,1]. The random variable Y represents a number chosen and takes on values in [0,1].



Probability Distribution

The **probability distribution** of a random variable X tells us what values X can take and how to assign probabilities to those values

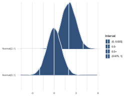
- Example: Roll a die. Y represents the number of spots on the side facing you. Therefore, Y=1,2,3,4,5,6. We do not know what Y will be till we toss the die. For all we know it could be a weighted dice.
- Note, this is a **discrete** probability distribution, so the random variable Y can take specific values with a probability attached.

Fair Dice

y_i	1	2	3	4	5	6
$\Pr(Y = y_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Weighted Die

y_i	1	2	3	4	5	6
$\Pr(Y = y_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$



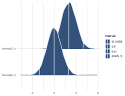
Probability Distribution (Continued)

- 3.5.2 from Statistics for the Life sciences: Y denotes nymber of children in a family chosen at random. $Y = 0, 1, 2, 3, \ldots$. The probability Y has a particular value is equal to the %-age of families with that many children. For example, if 23% of families have 2 kids, then

$$\Pr(Y = 2) = 0.23$$

- Example 3.5.3 from text. Y is random variable denoting number of medications a patient is given following cardiac surgery. If 52% of all patients are given 2,3,4, or 5 medications, then

$$\Pr(2 \leq Y \leq 5) = 0.52$$

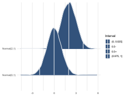


Mean of a Discrete Random Variable

- Let X be a discrete random variable with k elements $x_1, x_2, \ldots, x_k$ in its sample space $\mathcal{S}$. The mean of X is

$$\sum_{i=1}^{k} x_i \times \Pr(X = x_i)$$

- This is also called the expected value: $E(X) = \mu_X$
- Named expected value because over several repetitions of the random event we expect the value of X to be μ_X



Variance & std. deviation (Discrete R.V.)

- Let X be a discrete random variable with k elements $x_1, x_2, \ldots, x_k$ in its sample space $\mathcal{S}$. The variance of X is:

$$\sigma^2 = \sum_{i=1}^{k} (x_i - \mu)^2 \times \Pr(X = x_i)$$

- The standard deviation is the square root of the variance

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

- The standard deviation is a measure of dispersion away from the mean that takes into account both how far each value is from the mean and how likely each value is