Percentiles

are measures of location. There are 99 **percentiles** denoted P_1, P_2, . . . P_{99}, which divide a set of data into 100 groups with about 1% of the values in each group.

Finding the Percentile of a Data Value

$$\text{Percentile of value } x = \frac{\text{number of values less than } x}{\text{total number of values}} \bullet 100$$

Round it off to the nearest whole number

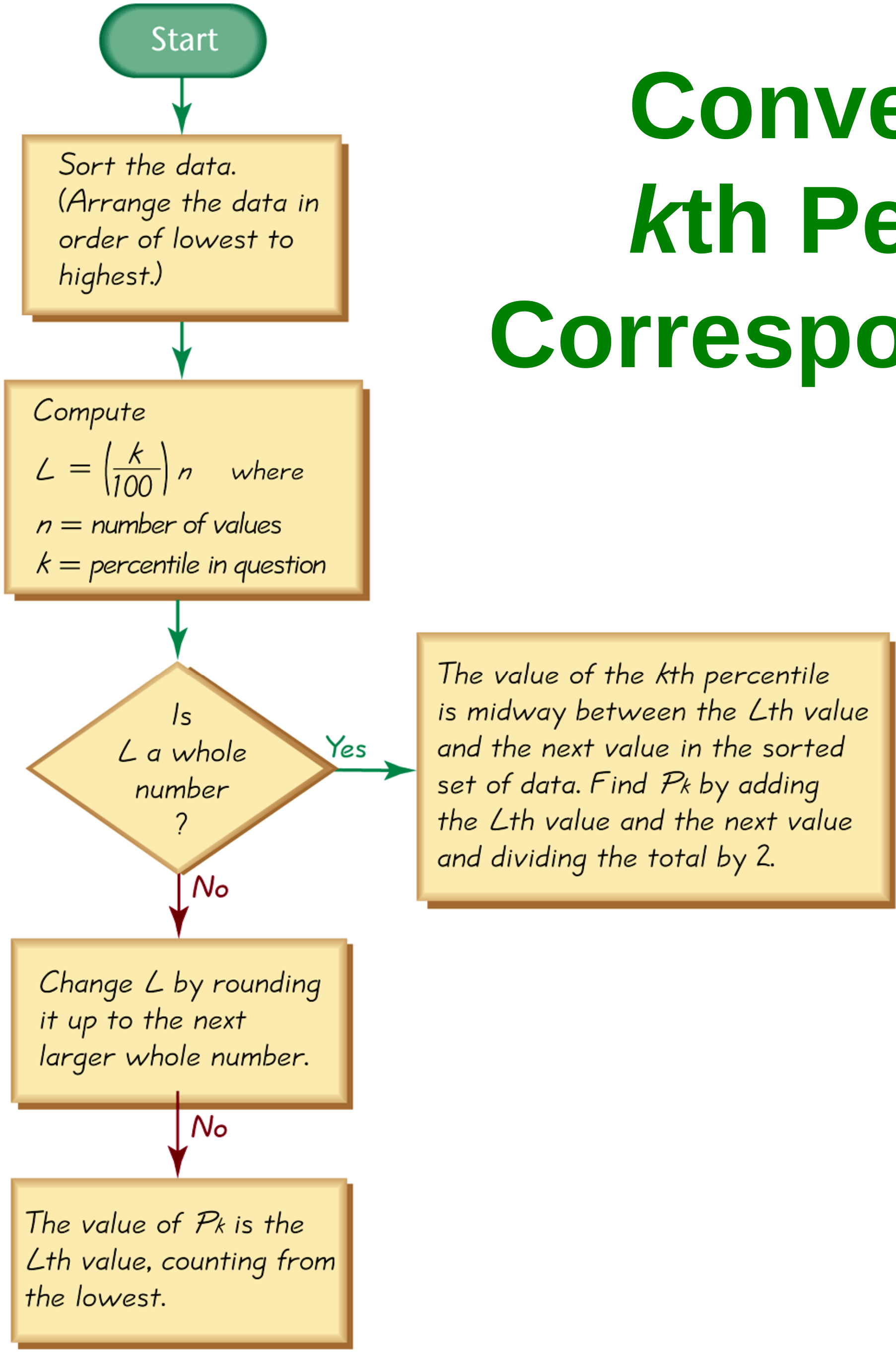
Converting from the *k*th Percentile to the Corresponding Data Value

Notation

$$L = \frac{k}{100} \cdot n$$

n	total number of values in the data set
k	percentile being used
L	locator that gives the position of a value
P_k	*k*th percentile

Converting from the *k*th Percentile to the Corresponding Data Value



Quartiles

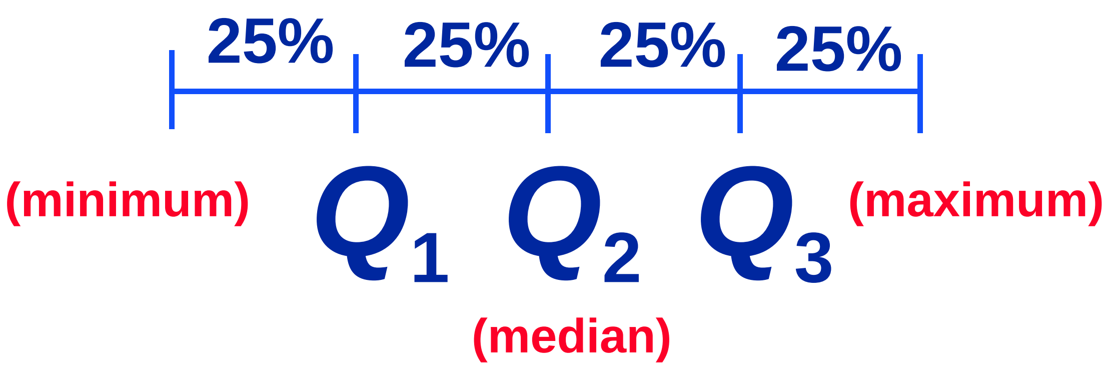
Are measures of location, denoted Q_1, Q_2, and Q_3, which divide a set of data into four groups with about 25% of the values in each group.

- **Q_1 (First Quartile) separates the bottom 25% of sorted values from the top 75%.**
- **Q_2 (Second Quartile) same as the median; separates the bottom 50% of sorted values from the top 50%.**
- **Q_3 (Third Quartile) separates the bottom 75% of sorted values from the top 25%.**

Quartiles

Q_1, Q_2, Q_3

divide ranked scores into four equal parts



Some Other Statistics

- **Interquartile Range (or IQR):** $Q_3 - Q_1$
- **Semi-interquartile Range:** $\dfrac{Q_3 - Q_1}{2}$
- **Midquartile:** $\dfrac{Q_3 + Q_1}{2}$
- **10 - 90 Percentile Range:** $P_{90} - P_{10}$

5-Number Summary

- For a set of data, the 5-number summary consists of the minimum value; the first quartile Q_1; the median (or second quartile Q_2); the third quartile, Q_3; and the maximum value.

Basic Principle of Statistics: Rare Event Rule

If, under a given assumption, the probability of a particular observed event is exceptionally small, we conclude that the assumption is probably not correct.

Example: if you flip a coin 10 times and observe 10 Heads and 0 Tails, would you believe that it is a normal (balanced) coin? Or would you rather have a doubt and have the coin checked out?

The need to compute probabilities

To use the rare even rule, we need to be able to compute probabilities under given assumptions.

Basic concepts of Probabilities

- **Event**

  any *collection of results or outcomes* of a procedure, or an experiment, or a game

- **Simple Event**

  an outcome or an event that cannot be further broken down into simpler components

- **Sample Space**

  (for a procedure) the sample space consists of <u>all</u> possible simple events; that is, the sample space consists of all outcomes that cannot be broken down any further

Notation for Probabilities

P - **denotes a probability.**

A*, *B*, and *C - **denote specific events.**

***P*(*A*)** - **denotes the probability of event *A* occurring.**

Classical Approach to Probability
(Requires Equally Likely Outcomes)

Assume that a given procedure has n different simple events and that each of those simple events has an equal chance of occurring. If event A can occur in s of these n ways, then

$$P(A) = \frac{s}{n} = \frac{\text{number of ways } A \text{ can occur}}{\text{total number of different simple events}}$$

EXAMPLE 7 **Gender of Children** Find the probability that when a couple has 3 children, they will have exactly 2 boys. Assume that boys and girls are equally likely and that the gender of any child is not influenced by the gender of any other child.

How big is this sample space?

8 possible events.

P(2 boys in 3 births) =3 /8

Law of Large Numbers

As a procedure is repeated again and again, the relative frequency of an event tends to approach its actual probability.

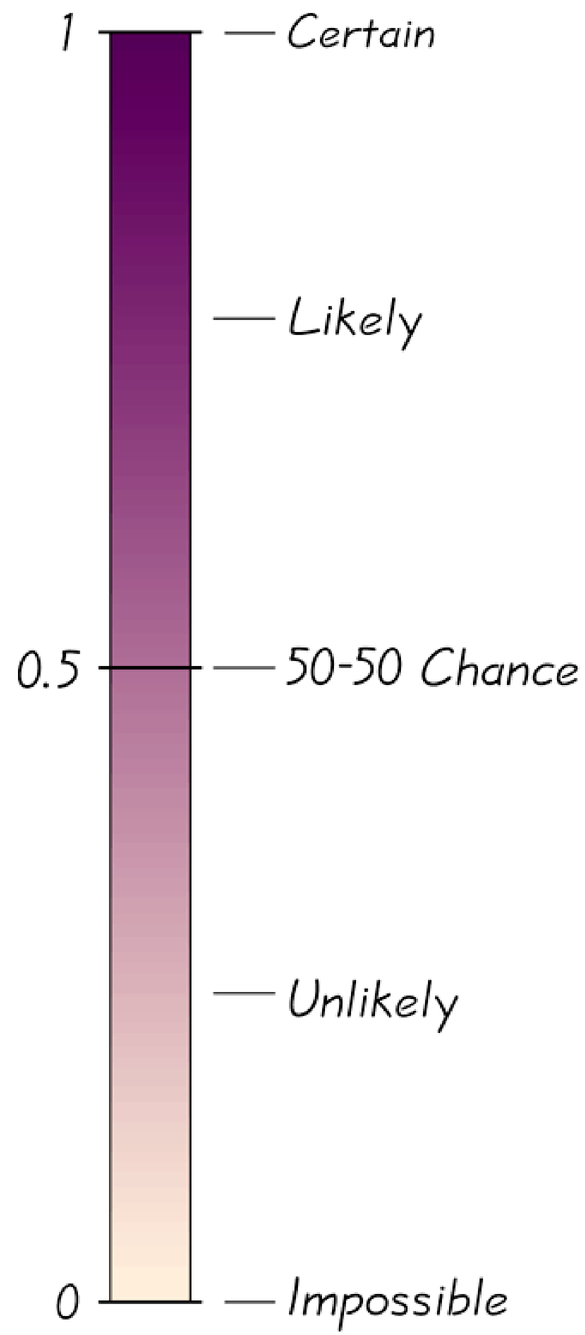
$$\textit{Relative frequency} = \frac{\text{\# of times } A \text{ occurred}}{\text{\# of times procedure was repeated}}$$

Probability Limits

Always express a probability as a fraction or decimal number between 0 and 1.

- **The probability of an impossible event is 0.**
- **The probability of an event that is certain to occur is 1.**
- **For any event *A*, the probability of *A* is between 0 and 1 inclusive.**
  That is, $0 \leq P(A) \leq 1$.

Possible Values for Probabilities



Rounding Off Probabilities

When expressing the value of a probability, either give the exact fraction or decimal or round off final decimal results to three significant digits.

(*Suggestion*: When a probability is not a simple fraction such as 1/3 or 5/9, express it as a decimal so that the number can be better understood.)

Aprox. : 1/3 to .333

5/9 to .556

Compound Event

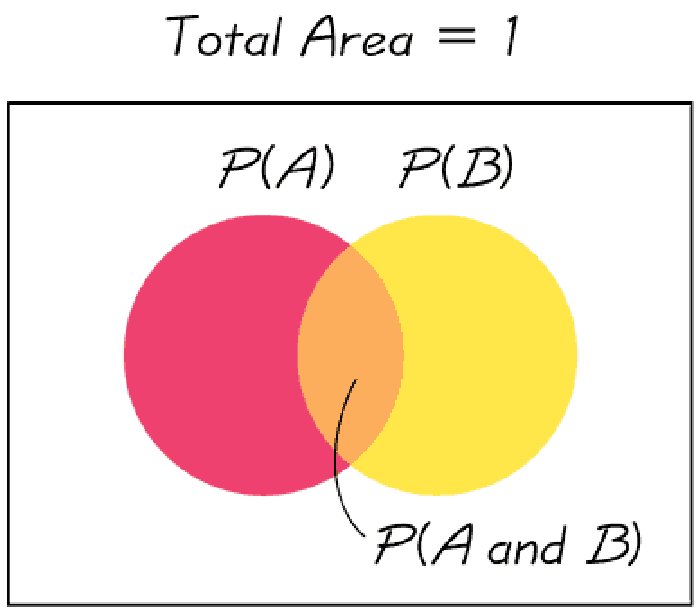
$P(A \text{ or } B) = P$ (event *A* occurs or event *B* occurs or they both occur)

Formal Addition Rule

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

where $P(A \text{ and } B)$ denotes the probability that A and B both occur at the same time

Illustration (Venn Diagram)

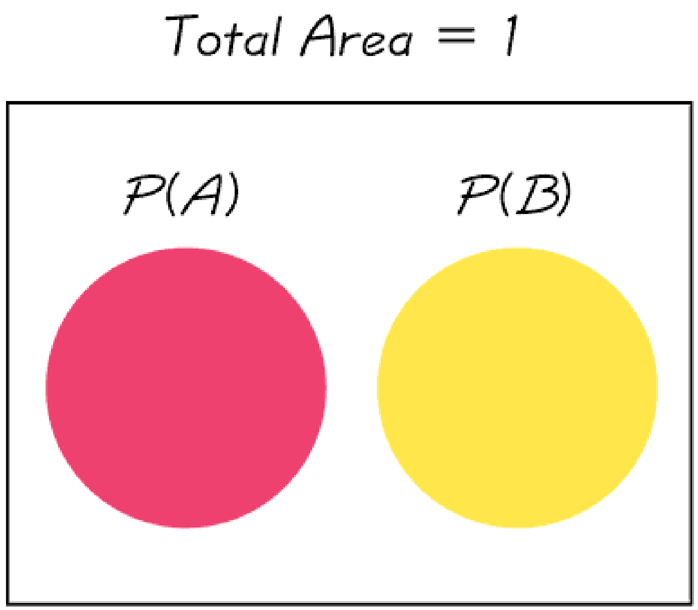


A **is the red disk,** ***B*** **is the yellow disk**

A **or** ***B*** **is the total area covered by both disks.**

Disjoint Events

Events *A* and *B* are disjoint (or mutually exclusive) if they cannot occur at the same time. (That is, disjoint events do not overlap.)



Venn Diagram for Disjoint Events

Formal Addition Rule for disjoint events

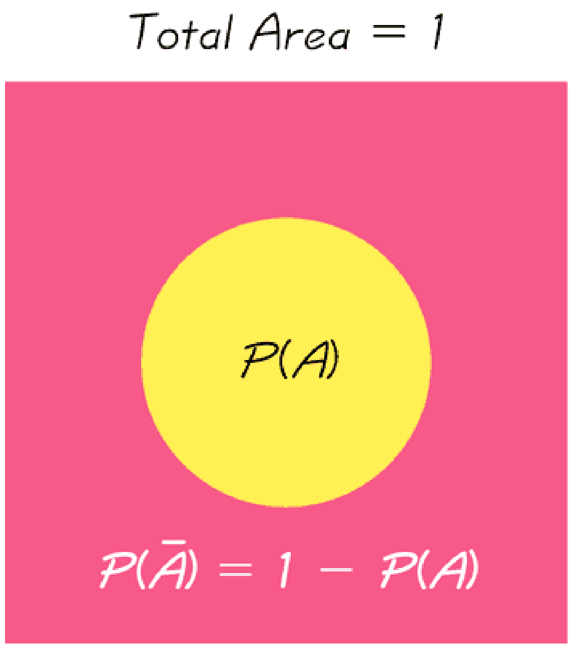
$$P(A \text{ or } B) = P(A) + P(B)$$

(only if *A* and *B* are disjoint)

Complementary Events

The complement of event *A*, denoted by $\overline{A}$, consists of all outcomes in which the event *A* does not occur.

Venn Diagram for the Complement of Event *A*



A is yellow, $\bar{A}$ is pink

Rules for Complementary Events

$$P(A) + P(\bar{A}) = 1$$

$$P(\bar{A}) = 1 - P(A)$$

$$P(A) = 1 - P(\bar{A})$$

Fundamental Counting Rule (multiplication rule)

For a sequence of two events in which the first event can occur *m* ways and the second event can occur *n* ways, the events together can occur a total of $m \cdot n$ ways.

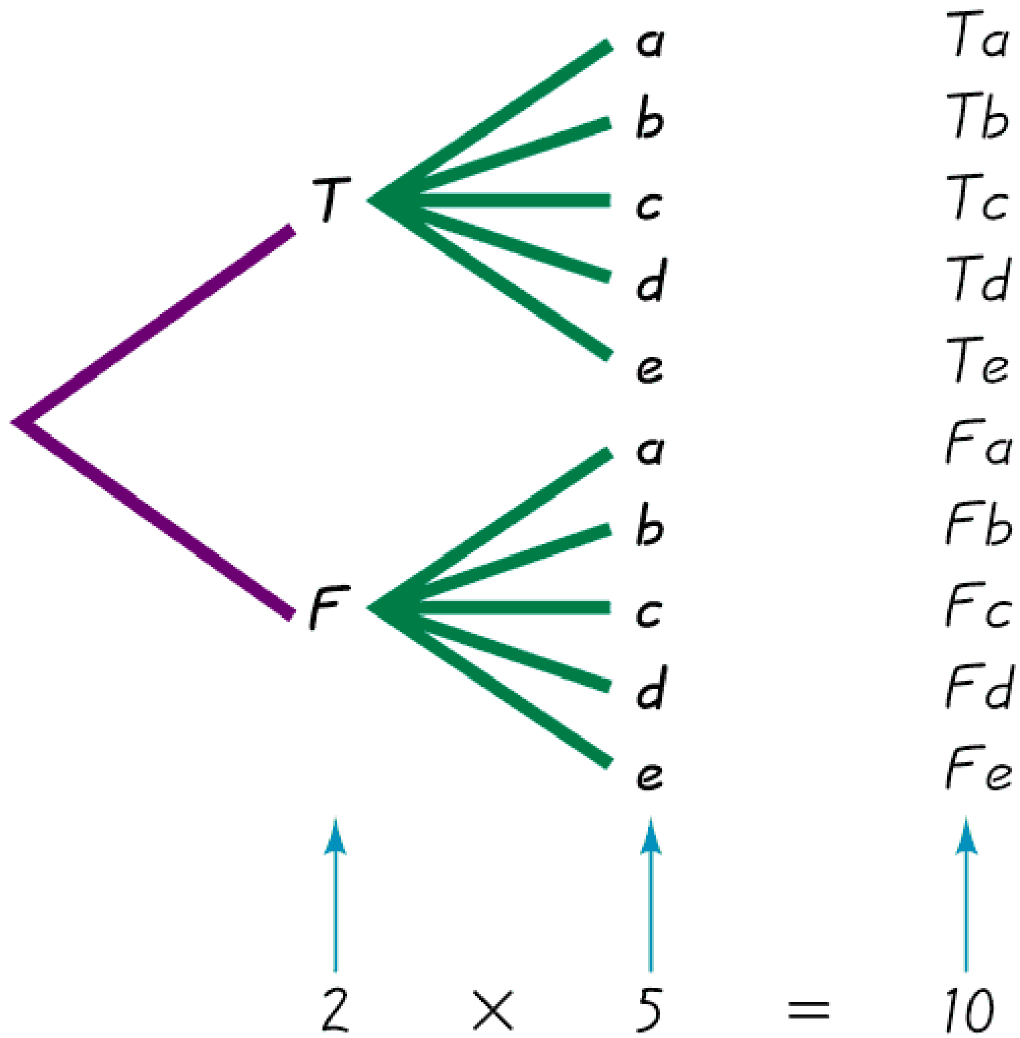
Example: a test consisting of a true/false question followed by a multiple choice question, where the choices are a,b,c,d,e.

Then there are 2 x 5 = 10 possible combinations.

Tree Diagrams

This figure summarizes the possible outcomes for a true/false question followed by a multiple choice question.

Total: 10 possible combinations.



Key Concept

The basic multiplication rule is used for finding $P(A \text{ and } B)$, the probability that event A occurs in a first trial and event B occurs in a second trial.

Notation

***P*(*A* and *B*) =**

***P*(event *A* occurs in a first trial and event *B* occurs in a second trial)**

Dependent and Independent

Two events *A* and *B* are independent if the occurrence of one does not affect the *probability* of the occurrence of the other.

If *A* and *B* are not independent, they are said to be dependent.

Formal Multiplication Rule for Independent Events

$$P(A \text{ and } B) = P(A) \bullet P(B)$$

Only if *A* and *B* are independent

- **Example:** A deck of 52 cards has 13 spades. If two cards are drawn from the deck at random, what is the chance that both are spades?

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B/A) = \frac{1}{4} \cdot \frac{12}{51} = \frac{12}{204}.$$

Dependent Events

Two events are dependent if the occurrence of one of them affects the *probability* of the occurrence of the other.

(But this does not necessarily mean that one of the events is a *cause* of the other.)

Key Concept

If the outcome of the first event *A* somehow affects the probability of the second event *B*, its is important to adjust the probability of *B* to reflect the occurrence of event *A*.

Conditional Probability Important Principle

The probability for the second event *B* should take into account the fact that the first event *A* has already occurred.

Notation for Conditional Probability

$P(B|A)$ represents the probability of event B occurring after it is assumed that event A <u>has already occurred</u>

(read $B|A$ as "B given A")

Formal Multiplication Rule for Dependent Events

- $P(A \text{ and } B) = P(A) \cdot P(B|A)$
- **Note that if A and B are independent events, P(B|A) is really the same as P(B).**

- **Example.** Roll a die twice. If the first roll is a six, what is the chance the second roll will be a six?

$$\mathbb{P}(A/B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/36}{1/6} = \frac{1}{6}.$$