Important Concepts and Formulas For

DATA AND DECISIONS

Part I
Voting

February 9, 2017

1. Social Choice Method

A **social choice method** (also known as **social choice function**) is a procedure that takes as input the preferences among the candidates expressed by the voters, and gives as output either a single winning candidate, or tied winning candidates, or a statement that there is no winner.

2. Plurality Voting

Suppose an election has more than two candidates. Every voter votes for one candidate. The method of **Plurality Voting** is that the candidate with the most votes wins the election.

3. Single Runoff Voting with Two Ballots

Suppose an election has more than two candidates. Every voter votes for one candidate. The method of **Single Runoff Voting** is that if a candidate has more than 50% of the votes, that candidate wins; if no candidate has more than 50% of the votes, then there is a runoff vote between the two candidates with the highest numbers of votes in the first round.

4. Single Runoff Voting with One Ballot

Suppose an election has more than two candidates. Every voter ranks the candidates without ties. The method of **Single Runoff Voting** is that if a candidate has more than 50% of the first place rankings, that candidate wins; if no candidate has more than 50% of the first place rankings, then all the candidates except the two with the largest number of the first place rankings are dropped from the rankings, and the candidate who has more than 50% of the first place rankings among the remaining two candidates is the winner.

5. Instant Runoff Voting

Suppose an election has more than two candidates. Every voter ranks the candidates without ties. The method of **Instant Runoff Voting** (also called the **Hare System**, among other names) is that if a candidate has more than 50% of the first place rankings, that candidate wins; if no candidate has more than 50% of the first place rankings, then the candidate with the least number of the first place rankings is dropped from the rankings, and the process is repeated, as many times as needed, until a candidate has more than 50% of the first place rankings among the remaining candidates, and that candidate is the winner.

6. Borda Count Voting

Suppose an election has more than two candidates. Suppose that there are n candidates. Every voter ranks the candidates without ties. The method of **Borda Count Voting** is that each candidate receives n point for each first place ranking, and $n-1$ points for each second place ranking, and so forth, concluding with 1 point for each last place ranking, and the candidate with the highest total number of points wins the election.

7. Sequential Pairwise Voting with Fixed Agenda

Suppose an election has more than two candidates. Every voter ranks the candidates without ties. The method of **Sequential Pairwise Voting with Fixed Agenda** is as follows. First, the candidates are listed in some order (called an "agenda"). Next, the first two candidates in the given order are compared, where the winner is the candidate who is ranked higher than the other candidate on a majority of the ballots. Next, the winner among the first two candidates is the compared with the third candidate in the given order, and the process is repeated until the last candidate is compared with the winner of the previous comparisons, and the final winner wins the election.

8. Condorcet Voting

Suppose an election has more than two candidates. Every voter ranks the candidates without ties. The method of **Condorcet Voting** is that every pair of candidates are compared, where the winner is the candidate who is ranked higher than the other candidate on a majority of the ballots, and if a candidate defeats everyone else, that candidate wins the election; if no candidate defeats everyone else, then there is no winner.

9. Copeland Voting

Suppose an election has more than two candidates. Every voter ranks the candidates without ties. The method of **Copeland Voting** is that every pair of candidates are compared, where the winner is the candidate who is ranked higher than the other candidate on a majority of the ballots, and in every pairwise comparison, the candidate who defeats the other receives 1 point, and tied candidates receive $\frac{1}{2}$ point each, and the candidate with the highest total number of points wins the election.

10. Dictatorship

Suppose an election has more than two candidates, and that one voter is chosen as the "dictator." Every voter ranks the candidates without ties. The method of **Dictatorship** is that the candidate with the highest ranking on the dictator's ballot wins the election.

11. Majority Criterion

A social choice method in which the voters rank the candidates without ties satisfies the **Majority Criterion**, abbreviated **MAJ**, if the following holds: if candidate A receives a majority of first place votes, then A is the winner.

12. Always a Winner Criterion

A social choice method in which the voters rank the candidates without ties satisfies the **Always a Winner Criterion**, abbreviated **AAW**, if the following holds: the method always produces a winner or a group of candidates tied for winner.

13. Condorcet Winner

Suppose an election has more than two candidates. Every voter ranks the candidates without ties. A **Condorcet winner** is a candidate who, when compared with every other candidates, is ranked higher than the other candidate on a majority of the ballots.

14. Condorcet Winner Criterion

A social choice method in which the voters rank the candidates without ties satisfies the **Condorcet Winner Criterion**, abbreviated **CWC**, if the following holds: if candidate A is a Condorcet winner, then A is the winner by the social choice method.

15. Monotonicity Criterion

A social choice method in which the voters rank the candidates without ties satisfies the **Monotonicity Criterion**, abbreviated **MON**, if the following holds: if candidate A is the winner (or tied for winner), and if one or more voters changes her ballot by exchanging A with the candidate she previously ranked just above A, then A would still be the winner (or tied for winner).

16. Pareto Criterion

A social choice method in which the voters rank the candidates without ties satisfies the **Pareto Criterion**, abbreviated **PAR**, if the following holds: if all voters rank candidate A higher than candidate B, then candidate B is not the winner (or tied for winner).

17. Independence of Irrelevant Alternatives Criterion

A social choice method in which the voters rank the candidates without ties satisfies the **Independence of Irrelevant Alternatives Criterion**, abbreviated **IIA**, if the following holds: if candidate A is the winner (or tied for winner) and candidate B is not the winner (or tied for winner), and if one or more voters changes her ballot but does not change which of A or B is ranked higher than the other, then B would still not be the winner (or tied for winner).

18. Simple Impossibility Theorem

There is no social welfare method for three or more candidates in which the voters rank the candidates without ties, and that satisfies Always a Winner Criterion, Independence of Irrelevant Alternatives Criterion and Condorcet Winner Criterion.

19. Social Welfare Method

A **social welfare method** (also known as **social welfare function**) is a procedure that takes as input the preferences among the candidates expressed by the voters, and gives as output a ranking of the candidates, perhaps with ties.

20. Pareto Criterion for Social Welfare Methods

A social welfare method in which the voters rank the candidates without ties satisfies the **Pareto Criterion for Social Welfare Methods** (also called **Unanimity Criterion**), abbreviated **PARSWM**, if the following holds: if all voters rank candidate A higher than candidate B, then candidate A is ranked higher than candidate B in the social welfare ranking.

21. Arrow's Impossibility Theorem

There is no social choice method for three or more candidates in which the voters rank the candidates without ties, and that satisfies Pareto Criterion for Social Welfare Methods, Independence of Irrelevant Alternatives Criterion and Monotonicity Criterion, other than dictatorship.

22. Approval Voting

Suppose an election has two or more candidates. Every voter votes for as many candidates as she wants. The method of **Approval Voting** is that each candidate receives 1 point for each vote, and the candidate with the highest total number of points wins the election.

23. Range Voting

Suppose an election has two or more candidates. Every voter gives each candidate some number of points within a preset range of possible scores. The method of **Range Voting** is that the candidate with the highest total number of points wins the election.

24. Intensity of Independence of Irrelevant Alternatives Criterion

A social choice method in which the voters give each candidate some number of points within a preset range of possible scores satisfies the **Intensity of Independence of Irrelevant Alternatives Criterion**, abbreviated **IIIA**, if the following holds: if candidate A is the winner (or tied for winner) and candidate B is not the winner (or tied for winner), and if one or more voters changes her ballot but does not change her intensity of preference for A over B, then B would still not be the winner (or tied for winner).

25. Non-Manipulable Criterion

A social choice method in which the voters rank the candidates without ties satisfies the **Non Manipulable Criterion**, abbreviated **NM**, if the following holds: there is no voter who can change her ranking of the candidates from her sincere ranking to an insincere ranking and by doing so cause there to be a winner who is higher ranked on her sincere ranking than occurred when she voted sincerely.

26. Gibbard-Satterthwaite Theorem

There is no social choice method for three or more candidates in which the voters rank the candidates without ties, and that satisfies Pareto Criterion and Non-Manipulable Criterion, other than dictatorship.

27. Standard Quota

Suppose an apportionment has n states, denoted $S_1, \ldots, S_n$, and H representatives. Suppose that the states have populations $P_1, \ldots, P_n$, respectively. The total population is $P = P_1 + \cdots + P_n$.

1. The **standard quota** for state S_k, denoted Q_k, is $\frac{P_k}{P} H$.
2. The **rounded-down-standard-quota** for state S_k, denoted D_k, is the result of rounding Q_k down to the nearest whole number.
3. The **standard quota remainder** for state S_k, denoted R_k, is $Q_k - D_k$.

28. Hamilton's Method

Suppose an apportionment has n states, denoted $S_1, \ldots, S_n$, and H representatives. Suppose that the states have populations $P_1, \ldots, P_n$, respectively.

Hamilton's Method allocates the representatives in the following steps.

1. For each state, find its standard quota, rounded-down-standard-quota and standard quota remainder.
2. Allocate representatives according to the rounded-down-standard-quotas.
3. Add up the representatives allocated in the previous step, and find the remaining number of representatives.
4. Allocate the remaining representatives by giving one representative at a time, starting with the state with the largest standard quota remainder, then the state with the second largest standard quota remainder, etc., until all the remaining representatives are allocated.

29. Standard Divisor

Suppose an apportionment has n states, denoted $S_1, \ldots, S_n$, and H representatives. Suppose that the states have populations $P_1, \ldots, P_n$, respectively. The total population is $P = P_1 + \cdots + P_n$.

1. The **standard divisor** for these states, denoted T, is $\frac{P}{H}$.
2. The standard quota for state S_k equals $\frac{P_k}{T}$.

30. Rounding

Let x be a number. Let n be a whole number such that $n \leq x \leq n + 1$.

1. The number x **rounded down** is n.
2. The number x **rounded up** is $n + 1$.
3. The number x **rounded** (also called **standardly rounded**) is n if $x \leq n + \frac{1}{2}$, and is $n + 1$ if $n + \frac{1}{2} < x$.

31. Jefferson's Method

Suppose an apportionment has n states, denoted $S_1, \ldots, S_n$, and H representatives. Suppose that the states have populations $P_1, \ldots, P_n$, respectively.

Jefferson's Method allocates the representatives in the following steps.

1. Find the standard divisor, denoted T.

2. For each state S_k, find its standard quota, which equals $\frac{P_k}{T}$, and its rounded-down-standard-quota, which is the whole number that is the result of doing rounding down to the standard quota.

3. Allocate representatives according to the rounded-down-standard-quotas.

4. Add up the representatives allocated using the rounded-down-standard-quotas. If the number of representatives allocated using the rounded-down-standard-quotas equals H, the allocation is complete.

5. If the number of representatives allocated using the rounded-down-standard-quotas does not equal H, choose a modified divisor, denoted $\hat{T}$, which is different from the standard divisor. For each state S_k, find its modified quota, which equals $\frac{P_k}{\hat{T}}$, and its rounded-down-modified-quota, which is the whole number that is the result of doing rounding down to the modified quota.

6. Allocate representatives according to the rounded-down-modified-quotas.

7. Add up the representatives allocated using the rounded-down-modified-quotas. If the number of representatives allocated using the rounded-down-modified-quotas equals H, the allocation is complete.

8. If the number of representatives allocated using the rounded-down-modified-quotas does not equal H, choose a new modified divisor, and try again, and keep trying until a modified divisor is found so that the number of representatives allocated using the rounded-down-modified-quotas equals H

32. Adams' Method

Suppose an apportionment has n states, denoted $S_1, \ldots, S_n$, and H representatives. Suppose that the states have populations $P_1, \ldots, P_n$, respectively.

Adams' Method allocates the representatives in the following steps.

1. Find the standard divisor, denoted T.

2. For each state S_k, find its standard quota, which equals $\frac{P_k}{T}$, and its rounded-up-standard-quota, which is the whole number that is the result of doing rounding up to the standard quota.

3. Allocate representatives according to the rounded-up-standard-quotas.

4. Add up the representatives allocated using the rounded-up-standard-quotas. If the number of representatives allocated using the rounded-up-standard-quotas equals H, the allocation is complete.

5. If the number of representatives allocated using the rounded-up-standard-quotas does not equal H, choose a modified divisor, denoted $\hat{T}$, which is different from the standard divisor. For each state S_k, find its modified quota, which equals $\frac{P_k}{\hat{T}}$, and its rounded-up-modified-quota, which is the whole number that is the result of doing rounding up to the modified quota.

6. Allocate representatives according to the rounded-up-modified-quotas.

7. Add up the representatives allocated using the rounded-up-modified-quotas. If the number of representatives allocated using the rounded-up-modified-quotas equals H, the allocation is complete.

8. If the number of representatives allocated using the rounded-up-modified-quotas does not equal H, choose a new modified divisor, and try again, and keep trying until a modified divisor is found so that the number of representatives allocated using the rounded-up-modified-quotas equals H

33. Webster's Method

Suppose an apportionment has n states, denoted $S_1, \ldots, S_n$, and H representatives. Suppose that the states have populations $P_1, \ldots, P_n$, respectively.

Webster's Method allocates the representatives in the following steps.

1. Find the standard divisor, denoted T.
2. For each state S_k, find its standard quota, which equals $\frac{P_k}{T}$, and its standardly-rounded-standard-quota, which is the whole number that is the result of doing standard rounding to the standard quota.
3. Allocate representatives according to the standardly-rounded-standard-quotas.
4. Add up the representatives allocated using the standardly-rounded-standard-quotas. If the number of representatives allocated using the standardly-rounded-standard-quotas equals H, the allocation is complete.
5. If the number of representatives allocated using the standardly-rounded-standard-quotas does not equal H, choose a modified divisor, denoted $\hat{T}$, which is different from the standard divisor. For each state S_k, find its modified quota, which equals $\frac{P_k}{\hat{T}}$, and its standardly-rounded-modified-quota, which is the whole number that is the result of doing standard rounding to the modified quota.
6. Allocate representatives according to the standardly-rounded-modified-quotas.
7. Add up the representatives allocated using the standardly-rounded-modified-quotas. If the number of representatives allocated using the standardly-rounded-modified-quotas equals H, the allocation is complete.
8. If the number of representatives allocated using the standardly-rounded-modified-quotas does not equal H, choose a new modified divisor, and try again, and keep trying until a modified divisor is found so that the number of representatives allocated using the standardly-rounded-modified-quotas equals H

34. Geometric Mean Rounding

1. Let a and b be two non-negative whole numbers. The **geometric mean** of a and b is $\sqrt{ab}$.
2. Let x be a number. Let n be a whole number such that $n \leq x \leq n+1$. The number x **geometric mean rounded** is n if x is less than or equal to the geometric mean of n and $n+1$, and is $n+1$ if x is greater than the geometric mean of n and $n+1$.

35. Huntington-Hill Method

Suppose an apportionment has n states, denoted $S_1, \ldots, S_n$, and H representatives. Suppose that the states have populations $P_1, \ldots, P_n$, respectively.

The **Huntington-Hill Method** allocates the representatives in the following steps.

1. Find the standard divisor, denoted T.

2. For each state S_k, find its standard quota, which equals $\frac{P_k}{T}$, and its geometric-mean-rounded-standard-quota, which is the whole number that is the result of doing geometric mean rounding to the standard quota.

3. Allocate representatives according to the geometric-mean-rounded-standard-quotas.

4. Add up the representatives allocated using the geometric-mean-rounded-standard-quotas. If the number of representatives allocated using the geometric-mean-rounded-standard-quotas equals H, the allocation is complete.

5. If the number of representatives allocated using the geometric-mean-rounded-standard-quotas does not equal H, choose a modified divisor, denoted $\hat{T}$, which is different from the standard divisor. For each state S_k, find its modified quota, which equals $\frac{P_k}{\hat{T}}$, and its geometric-mean-rounded-modified-quota, which is the whole number that is the result of doing geometric mean rounding to the modified quota.

6. Allocate representatives according to the geometric-mean-rounded-modified-quotas.

7. Add up the representatives allocated using the geometric-mean-rounded-modified-quotas. If the number of representatives allocated using the geometric-mean-rounded-modified-quotas equals H, the allocation is complete.

8. If the number of representatives allocated using the geometric-mean-rounded-modified-quotas does not equal H, choose a new modified divisor, and try again, and keep trying until a modified divisor is found so that the number of representatives allocated using the geometric-mean-rounded-modified-quotas equals H

36. Quota Criterion

An apportionment method satisfies the **Quota Criterion**, abbreviated **QUO**, if the following holds: the number of representatives allocated to each state is either its rounded-down-standard-quota or its rounded-up-standard-quota.

37. House Monotonicity Criterion

An apportionment method satisfies the **House Monotone Criterion**, abbreviated **HMON**, if the following holds: if the number of representatives is increased, no state loses representatives.

38. Population Criterion

An apportionment method satisfies the **Population Criterion**, abbreviated **POP**, if the following holds: if the population of state A increases and the population of state B decreases, then it cannot happen that A loses representatives and B gains representatives or stays the same.

39. One Version of Balinski-Young Theorem

An apportionment method that satisfies House Monotone Criterion and Population Criterion will not satisfy Quota Criterion, and an apportionment method that satisfies Quota Criterion will not satisfy House Monotone Criterion and Population Criterion.

40. Shortest Splitline Method

Suppose a state is to be divided into n districts.

The **Shortest Splitline Method** divides the state in the following steps.

1. Suppose n is even. For the first step, divide the state into two parts by finding the shortest straight line that separates the state into two parts with equal population. Each of these two parts will subsequently be divided up into $\frac{n}{2}$ districts.

2. Suppose n is odd. Let d be the result of rounding down $\frac{n}{2}$, and let u be the result of rounding up $\frac{n}{2}$. For the first step, divide the state into two parts by finding the shortest straight line that separates the state into two parts with populations in proportion $d : u$. Each of these two parts will subsequently be divided up into d and u districts, respectively.

3. For the next step, take each of the two parts created above, and divide each of them into two districts, using the same methods as above depending upon whether each part has to be divided into an even or odd number of districts.

4. Continue until each part represents one district.

Important Concepts and Formulas For

DATA AND DECISIONS

Part II
Networks

February 9, 2017

1. Graphs

1. A **graph** G is a collection of objects, called **vertices** (singular: vertex), together with another collection of objects, called **edges**, such that each edge connects two vertices (which are not necessarily different).
2. A **loop** in a graph is an edge of the graph that has both endpoint the same.
3. The collection vertices of the graph G is denoted $V(G)$.
4. The collection edges of the graph G is denoted $E(G)$.

2. Graph Definitions

Let G be a graph.

1. Let v and w be vertices of G. The vertices v and w are **adjacent** is there is an edge containing them.
2. Let v be a vertex of G. The **degree** of the vertex v, denoted $\deg(v)$, is the number of edges that contain v. If a loop contains v, it is counted twice in the degree.
3. A graph H is a **subgraph** of G if all the vertices of H are vertices of G and all the edges of H are edges of G.

3. Other Types of Graphs

1. A **directed graph** G is a collection of objects, called **vertices** (singular: vertex), together with another collection of objects, called **directed edges**, such that each directed edge starts at a vertex and ends at a vertex (which is not necessarily different from the starting vertex).
2. A **weighted graph** G is a graph (undirected) such that every edge is assigned a number, called the **weight** of the edge.

4. Directed Graph Definitions

Let G be a directed graph.

1. Let v be a vertex of G. The **in-degree** of the vertex v, denoted $\deg^{in}(v)$, is the number of directed edges that contain v as the end point.
2. Let v be a vertex of G. The **out-degree** of the vertex v, denoted $\deg^{out}(v)$, is the number of directed edges that contain v as the starting point.

5. Total Weight

Let G be a weighted graph. Let H be a subgraph of G. The **total weight** of H is the sum of the weights of the edges of H.

6. Paths and Circuits

Let G be a graph.

1. A **path** in the graph G is an alternating sequence

$$v_0, e_0, v_1, e_1, v_2, \ldots, v_{k-1}, e_{k-1}, v_k$$

of vertices and edges of G, which begins and ends with vertices, such that the two vertices of each edge of the sequence are the vertices in the sequence before and after the edge.

2. A **circuit** in the graph G is a path that starts and ends at the same vertex.

7. Components

Let G be a graph.

1. The graph G is **connected** if for every two vertices x and y, there is a path in G that starts at x and ends at y.

2. A **component** of the graph G is a subgraph of G that is connected, and is that is not contained in a larger subgraph of G that is connected.

8. Euler Paths and Circuits

Let G be a graph.

1. An **Euler path** in the graph G is a path in the graph that includes each edge of the graph exactly once.

2. An **Euler circuit** in the graph G is a circuit in the graph that includes each edge of the graph exactly once.

9. Euler Paths Theorem

Let G be a graph.

1. The graph G has an Euler circuit if and only if every vertex of G has even degree; an Euler circuit can start and end at any vertex of G.
2. The graph G has an Euler path that is not a circuit if and only if two vertices of G have odd degree and all other vertices have even degree; an Euler past must start at one of the vertices of odd degree and end at the other vertex of odd degree.

10. Vertex Coloring

Let G be a graph. A **vertex coloring** (also called a **coloring**) of the graph G is an assignment of colors (often written as numbers) to the vertices of G such that adjacent vertices are assigned different colors.

11. Chromatic Number

Let G be a graph.

1. Let k be a positive integer. A **k-coloring** of the graph G is a coloring of G with exactly k colors.
2. The **chromatic number** of the graph G, denoted $\chi(G)$, is the smallest whole number k such that G has a k-coloring.

12. Welsh and Powell Algorithm

Let G be a graph. This algorithm, called the **Welsh and Powell Algorithm**, finds a vertex coloring of the graph; this vertex coloring might not have the smallest possible number of colors.

1. Make a list of all the vertices of the graph G such that the degrees of the vertices are in decreasing or equal order.
2. Color the first vertex on the list with color number 1.
3. Color the second vertex on the list with color number 1 if possible, and with color number 2 otherwise.
4. Continue coloring one vertex at a time in the list, where each vertex is colored with the same color as the previous vertex if possible, and with a new color otherwise.
5. Repeat the above procedure until all the vertices are colored.

13. Cycle and Tree

Let G be a graph.

1. A **cycle** in the graph G is a circuit with no repeated vertices other than the starting and ending vertex.
2. The graph G is a **tree** if it is connected and it does not have any cycles.

14. Spanning Trees

Let G be a graph. A **spanning tree** of the graph G is subgraph that is a tree and that contains all the vertices of G.

15. Spanning Tree Theorem

Let G be a graph. If G is connected, then it has a spanning tree.

16. Minimum Spanning Trees

Let G be a weighted graph. Suppose that G is connected. A **minimum spanning tree** of the graph G is a spanning tree of G that has the smallest total weight of any spanning tree of G.

17. Prim's Algorithm

Let G be a weighted graph. Suppose that G is connected. This algorithm, called **Prim's Algorithm**, finds a minimum spanning tree of the graph.

1. Choose an edge of G that has the smallest weight among all the edges of G; if there is more than one such edge, choose one. Mark the chosen edge and its vertices.
2. Choose an edge of G that has the smallest weight among all the unmarked edges of G that have one vertex marked and one vertex unmarked; if there is more than one such edge, choose one. Mark the chosen edge and its vertices.
3. Repeat the above procedure until all the vertices are marked.

18. Domination

Let G be a graph.

1. A **dominating set** of the graph G is a collection W of vertices of G such that every vertex in G is either in W or is adjacent to a vertex in W.

2. The **domination number** of the graph G, denoted $\gamma(G)$, is the smallest whole number k such that G has a dominating set with k vertices.

19. Summation Notation

1. Let $x_1, x_2, x_3, \ldots, x_n$ be a collection of n numbers. Then the summation

$$x_1 + x_2 + x_3 + \cdots + x_n$$

is represented in **summation notation** by

$$\sum_{i=1}^{n} x_i.$$

The number i in the above summation notation is the **index** of the summation.

2. The index in a summation notation does not need to start at the value 1; it can start and end at any numbers. The step size in summation notation is always 1.

3. If it is clear from the context what the possible values of the index are, the notation

$$\sum_{i=1}^{n} x_i$$

can be abbreviated as

$$\sum x_i.$$

If it is clear from the context what the possible values $x_1, x_2, x_3, \ldots, x_n$ are, the notation

$$\sum_{i=1}^{n} x_i$$

can be abbreviated even further as

$$\sum x.$$

20. Average Value

Let $x_1, x_2, x_3, \ldots, x_n$ be a collection of n numbers. The **average value** of $x_1, x_2, x_3, \ldots, x_n$ is

$$\frac{x_1 + x_2 + x_3 + \cdots + x_n}{n},$$

which can also be written

$$\frac{1}{n}(x_1 + x_2 + x_3 + \cdots + x_n).$$

Using summation notation, the average value can be written as any of

$$\frac{\sum_{i=1}^{n} x_i}{n} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \frac{\sum x_i}{n} \quad \text{and} \quad \frac{1}{n}\sum x_i \text{and} \quad \frac{\sum x}{n} \quad \text{and} \quad \frac{1}{n}\sum x.$$

21. Handshake Theorem

Let G be a graph. Suppose that G has N vertices and L edges. Let $v_1, \ldots, v_N$ be the vertices of G, and let $k_1, \ldots, k_N$ denote the degrees of the vertices, respectively. Then

$$\sum_{i=1}^{N} k_i = 2L.$$

22. Vertices with Odd Degree

Let G be a graph. Then G has an even number of vertices with odd degree.

23. Average Degree

Let G be a graph. Suppose that G has N vertices. Let $v_1, \ldots, v_N$ be the vertices of G, and let $k_1, \ldots, k_N$ denote the degrees of the vertices, respectively. The **average degree** of G, denoted $\bar{k}$ (and also $\langle k \rangle$), is defined by

$$\bar{k} = \frac{1}{N}\sum_{i=1}^{N} k_i.$$

24. Shortest Path

Let G be a graph, and let v and w be vertices of G. A **shortest path** from v to w is a path from v to w that has the fewest edges among all paths from v to w.

25. Edge Distance

Let G be a graph, and let v and w be vertices of G. The **edge distance** (also called **distance**) from v to w is the number of edges of a shortest path from v to w.

26. Breadth-First Search Algorithm

Let G be a graph. Suppose that G is connected. Let A be a vertex of G. This algorithm, called **Breadth-First Search Algorithm** (abbreviated **BFS**), finds the edge distance from A to every other vertex of G.

The algorithm starts with the vertices of the graph unlabeled. The label on each vertex at the end of the algorithm is the edge distance from A to that vertex.

1. Label A with value 0.
2. For each vertex that is adjacent to A, label it with value 1.
3. For each vertex that is unlabeled and is adjacent to a vertex labeled 1, label it with value 2.
4. For each vertex that is unlabeled and is adjacent to a vertex labeled 2, label it with value 3.
5. Continue the above process until all vertices are labeled.
6. The numerical label of each vertex is the edge distance from A to that vertex.

27. Breadth-First Search Algorithm for Shortest Paths

Let G be a graph. Suppose that G is connected. Let A be a vertex of G. The Breadth-First Search Algorithm can be used to find a shortest path from A to every other vertex of G.

First, do Breadth-First Search Algorithm starting with the vertex A. The algorithm will produce a numerical label of each vertex of G, which is the edge distance from A to that vertex.

Let B be a vertex of G other than A.

1. Suppose B has value r.
2. The vertex B must be adjacent to at least one vertex with label $r-1$; choose one such vertex, and call it v_{r-1}.
3. The vertex v_{r-1} must be adjacent to at least one vertex with label $r-2$; choose one such vertex, and call it v_{r-2}.
4. The vertex v_{r-2} must be adjacent to at least one vertex with label $r-3$; choose one such vertex, and call it v_{r-3}.
5. Continue the above process until a vertex with value 0 is reached, and that vertex must be A.
6. The path with vertices $A, v_1, v_2, \ldots, v_{r-1}, B$ is a shortest path from A to B.

28. Average Distance and Diameter

Let G be a graph.

1. The **average distance** of G is the average of the edge distances between all pairs of distinct vertices of G.
2. The **diameter** of G is the longest edge distance between any pair of distinct vertices of G.

29. Breadth-First Search Algorithm for Spanning Trees

Let G be a graph. Suppose that G is connected. This algorithm, called **Breadth-First Search Algorithm for Spanning Trees**, finds a spanning tree of G.

The algorithm starts with both the vertices and the edges of the graph unmarked. The highlighted edges at the end of the algorithm is a spanning tree.

1. Choose an arbitrary vertex of G, say A.

2. Label A with value 0.

3. For each vertex that is adjacent to A, label it with value 1.

4. For each vertex with label 1, choose a single edge that from it to the vertex labeled 0, and highlight that edge.

5. For each vertex that is unlabeled and is adjacent to a vertex with label 1, label it with value 2.

6. For each vertex with label 2, choose a single edge that from it to a vertex with label 1, and highlight that edge.

7. For each vertex that is unlabeled and is adjacent to a vertex labeled 2, label it with value 3.

8. For each vertex with label 3, choose a single edge that from it to a vertex with label 2, and highlight that edge.

9. Continue the above process until all vertices are labeled, and every vertex has a highlighted edge from it to a vertex with a label with value one less.

10. The highlighted edges are a spanning tree of G.

30. Shortest Weighted Path

Let G be a weighted graph, and let v and w be vertices of G. A **shortest weighted path** from v to w is a path from v to w that has the smallest total weight among all paths from v to w.

31. Weighted Distance

Let G be a graph, and let v and w be vertices of G. The **weighted distance** (also called **distance**) between v and w is the total weight of edges of a shortest weighted path from v to w.

32. Dijkstra's Algorithm

Let G be a weighted graph. Suppose that all the edge weights are positive or zero. Suppose that G is connected. Let A be a vertex of G. This algorithm, called **Dijkstra's Algorithm**, finds the weighted distance to every other vertex of G, and finds a shortest path from A to every other vertex.

At every step of the algorithm, each vertex will be labeled with a numerical value from 0 to ∞, including possibly either of those. After each step of the algorithm, the number labeling some of the vertices might be updated to a smaller number than the previous label. The label on each vertex at the end of the algorithm is the weighted distance from A to that vertex.

1. Label A with value 0, and label every other vertex with value ∞.

2. Circle vertex A. This vertex is the currently visited vertex.

3. For each vertex that is adjacent to A, update its numerical label to the distance to A, and draw a small arrow at the vertex pointing to A.

4. Among all the vertices on G other than A, choose the one with the smallest current numerical label. Circle that vertex, and designate it as the currently visited vertex.

5. Main Step: Suppose that the currently visited vertex is labeled X. Look at all the un-circled vertices of G that are adjacent to X, one at a time (in any order). Let U be such a vertex. Add the distance from X to U to the numerical label of X. If this sum is less than the numerical label of U, then update the numerical label of U to the sum, draw a small arrow at U pointing to X, and erase the previous small arrow at U pointing to another vertex if there is such an arrow; if the sum is not less than the numerical label of U, then do not change anything for U. Do that for all un-circled vertices that are adjacent to X.

6. Among all the un-circled vertices of G, choose the one with the smallest current numerical label. Circle that vertex, and designate it as the currently visited vertex.

7. Repeat the Main Step until all the vertices are circled.

8. When every vertex is circled, each vertex will have a numerical label that is less than ∞, and each vertex other than A will have a one small arrow pointing to another vertex.

9. The final numerical label of each vertex is the weighted distance from A to that vertex.

10. To find a shortest path from A to another vertex, start at that other vertex, and follow the arrows back to A.

33. PageRank Algorithm

Let G be a directed graph. This algorithm, called the **PageRank Algorithm**, ranks the vertices of G by assigning a number to each vertex, where the vertices are then ranked from highest assigned number to lowest assigned number.

The algorithm works by making an initial assignment of numbers to the vertices, and then revising the numbers repeatedly until the desired numbers are found.

Let p be a number between 0 and 1. The number p is typically 0.85.

1. Suppose that G has n vertices. Assign each vertex an initial value of $\frac{1}{n}$.

2. If a vertex has m edges going out of it (that is, the out-degree of the vertex is m), then give every edge going out of that vertex weight $\frac{1}{m}$.

3. Redistribute the number assigned to each vertex as follows. Suppose a vertex is assigned the number x, and it has m edges going out of it. Then transfer $x \cdot \frac{1}{m}$ to each of the vertices that are reached by the m edges going out of that vertex. Do that simultaneously to all the vertices.

4. Modify the number assigned to each vertex as follows. Suppose that after the redistribution a vertex is assigned the number y. Recall that there are n vertices. Then modify the number assigned to this vertex to be $p \cdot y + (1 - p) \cdot \frac{1}{n}$. Do that simultaneously to all the vertices.

5. Repeat this two-step process (redistribution and modification, redistribution and modification, and so on) until it appears that the numbers assigned to the vertices do not change with each new repetition of the two-step process.

6. The numbers assigned to each vertex in the previous step are the final numbers assigned to each vertex.

7. A way to check for errors in calculation at each stage of the process is to use the fact that at each stage of the process, the sum of the numbers assigned to the vertices is always 1. If the sum is ever not 1, that indicates an error.

34. Matrices Definitions

1. A **matrix** (plural: **matrices**) is a rectangular array of numbers enclosed in square brackets.

2. The size of the matrix is determined by the number of rows and the number of columns.

3. If a matrix has m rows and n columns, it is called an $m \times n$ **matrix**.

4. Let A be an $m \times n$ matrix. Then A is written in general as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

   where the entry a_{ij} is the entry in the i^{th} row and j^{th} column.

5. The entry in the i^{th} row and j^{th} column of a matrix is referred to as the (i, j) entry in the matrix.

6. A **vector** (also called a **column vector**) is an $m \times 1$ matrix.

35. Adjacency Matrix of a Graph

Let G be a graph. Suppose that G has N vertices. Let $v_1, \ldots, v_N$ be the vertices of G. The **adjacency matrix** of G is the $N \times N$ matrix that has the (i, j) entry in the matrix equal to 1 if the vertices v_i and v_j are adjacent and equal to 0 if the vertices v_i and v_j are not adjacent.

36. Adjacency Matrix of a Directed Graph

Let G be a directed graph. Suppose that G has N vertices. Let $v_1, \ldots, v_N$ be the vertices of G. The **adjacency matrix** of G is the $N \times N$ matrix that has the (i, j) entry in the matrix equal to the number of edges from v_i to v_j (which is 0 if there is no edge from v_i to v_j).

37. Matrices: Addition and Scalar Multiplication

Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$, and let c be a number.

1. $A + B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}$.

2. $A - B = \begin{bmatrix} a_{11}-b_{11} & a_{12}-b_{12} & \cdots & a_{1n}-b_{1n} \\ a_{21}-b_{21} & a_{22}-b_{22} & \cdots & a_{2n}-b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}-b_{m1} & a_{m2}-b_{m2} & \cdots & a_{mn}-b_{mn} \end{bmatrix}$.

3. $-A = \begin{bmatrix} -a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & -a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ -a_{m1} & -a_{m2} & \cdots & -a_{mn} \end{bmatrix}$.

4. $cA = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{bmatrix}$.

38. Matrices: Multiplication

Row times Column

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1b_1 + a_2b_2 + \cdots a_nb_n.$$

General If A is an $m \times p$ matrix and B is a $p \times n$ matrix, then AB is an $m \times n$ matrix obtained by multiplying each row in A by each column in B.

39. Transition Matrix of a Directed Graph

Let G be a directed graph. Suppose that G has N vertices. Let $v_1, \ldots, v_N$ be the vertices of G, and let $k_1, \ldots, k_N$ denote the degrees of the vertices, respectively. The **transition matrix** of G is the $N \times N$ matrix that has the (j, i) entry in the matrix equal to $\frac{1}{k_i}$ if there is an edge from v_i to v_j and equal to 0 if there is no edge from v_i to v_j.

40. Probability Vector and Stochastic Matrix

1. A vector is a **probability vector** (also called a **stochastic vector**) if none of the values in the vector is negative, and the sum of the values in the vector is 1.

2. A matrix is a **stochastic matrix** if none of the values in the matrix is negative, and the sum of the values of every column in the matrix is 1.

41. PageRank Algorithm Using Matrices

Let G be a directed graph. This algorithm, called the **PageRank Algorithm**, ranks the vertices of G by assigning a number to each vertex, where the vertices are then ranked from highest assigned number to lowest assigned number.

In the matrix version of the PageRank algorithm, rather than directly assigning a number to each vertex, a probability vector is found that has as many entries as there are vertices in G; the first entry of the probability vector is then assigned to the first vertex, and so on.

The matrix version of the PageRank algorithm works by defining an initial probability vector, and then repeatedly multiplying it by a certain stochastic matrix, until the desired probability vector is found.

Let p be a number between 0 and 1. The number p is typically 0.85.

1. Suppose that G has n vertices. The initial probability vector v is the vector with n entries, and with each entry equalling $\frac{1}{n}$.

2. Let A be the transition matrix of the graph G. The matrix A is an $n \times n$ stochastic matrix.

3. Let B be the $n \times n$ matrix that has all its entries equal to $\frac{1}{n}$. The matrix B is an $n \times n$ stochastic matrix.

4. Let $M = pA + (1 - p)B$. The matrix M is an $n \times n$ stochastic matrix.

5. Compute Mv, then compute M^2v, then compute M^3v, and so on. Keep going, until it appears that the resulting vector does not change with each new multiplication by M. All these vectors are probability vectors.

6. The probability vector that does not change when multiplied by M, found in the previous step, is used to assign a number to each vertex, where the first entry in the probability vector is assigned to the first vertex, and so on.

7. A quicker way to find the desired probability vector is to find a probability vector $\bar{v}$ such that $M\bar{v} = \bar{v}$, though doing so requires a bit more algebra than the above method.

42. The Square of the Adjacency Matrix

Let G be a directed graph. Suppose that G has N vertices. Let $v_1, \ldots, v_N$ be the vertices of G. Let A be the adjacency matrix of G.

1. The (i, i) element of A^2, which is on the diagonal, is the degree of vertex v_i.
2. If i and j are two different numbers, the (i, j) element of A^2 is the number of paths of length 2 from v_i to v_j.

43. Powers of the Adjacency Matrix

Let G be a directed graph. Suppose that G has N vertices. Let $v_1, \ldots, v_N$ be the vertices of G. Let A be the adjacency matrix of G. Let r be a positive whole number. The (i, j) element of A^r is the number of paths of length r from v_i to v_j.

44. Clique

Let G be a graph. A **clique** in G is a collection Q of vertices in G that satisfies the following three conditions.

1. The collection Q has at least three vertices.
2. Every vertex in Q is adjacent to every other vertex in Q.
3. The collection Q is not contained in a larger collection of vertices such that very vertex in the larger collection is adjacent to every other vertex in the larger collection.

45. Cliques and Matrices

Let G be a directed graph. Suppose that G has N vertices. Let $v_1, \ldots, v_N$ be the vertices of G. Let A be the adjacency matrix of G. The vertex v_i is in a clique if and only if the (i, i) element of A^3 is non-zero.

Important Concepts and Formulas For

DATA AND DECISIONS

Part III
Statistics

February 9, 2017

1. Statistics Terminology

1. A **population** is the complete collection of objects of interest.
2. A **parameter** of the population is a number that summarizes some aspect of the population.
3. A **sample** of the population is a subcollection of the population.
4. A **statistic** of the sample is a number that summarizes some aspect of the sample.

2. Frequency Table

Suppose that in a given set of numerical data, the values that appear (possibly more than once each) are $x_1, x_2, \ldots, x_n$. A **frequency table** for the data is a chart of the form

$$\begin{array}{c|cccc} x & x_1 & x_2 & \cdots & x_n \\ \hline f & f_1 & f_2 & \cdots & f_n \end{array},$$

where $f_1, f_2, \ldots, f_n$ are the number of times each of $x_1, x_2, \ldots, x_n$ occurs, in the data, respectively.

3. Mean

Let $x_1, x_2, x_3, \ldots, x_n$ be a collection of n numbers. The **mean** of these numbers, denoted $\bar{x}$, is defined by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x}{n} = \frac{1}{n}\sum x.$$

4. Mean from Frequency Table

Suppose that numerical data is given in a frequency table of the form

$$\begin{array}{c|cccc} x & x_1 & x_2 & \cdots & x_n \\ \hline f & f_1 & f_2 & \cdots & f_n \end{array}.$$

The mean of this data is computed by

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \cdots + f_n x_n}{f_1 + f_2 + f_3 + \cdots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum fx}{\sum f}.$$

5. Sample Mean vs. Population Mean

Suppose that a sample is taken from a population.

1. The **sample mean**, which is the mean of the sample, is denoted $\bar{x}$.
2. The **population mean**, which is the mean of the whole population, is denoted μ.

6. Median

Let $x_1, x_2, x_3, \ldots, x_n$ be a collection of n numbers.

1. The **median** of these numbers, denoted $\tilde{x}$, is the number such that half of $x_1, x_2, x_3, \ldots, x_n$ are above it and half of $x_1, x_2, x_3, \ldots, x_n$ are below it.
2. To find the median, first list the numbers $x_1, x_2, x_3, \ldots, x_n$ in increasing order.
    1. Suppose n is odd. Then the median is middle number, which is the entry x_i, where i is obtained by rounding the fraction $\frac{n}{2}$ up to the nearest whole number.
    2. Suppose n is even. Then the median is the average of the two middle numbers, which is $\frac{x_i + x_{i+1}}{2}$, where i is the whole number $\frac{n}{2}$.

7. Mode

Let $x_1, x_2, x_3, \ldots, x_n$ be a collection of n numbers. The **mode** of these numbers is the value that occurs most frequently among the numbers.

8. Range

Let $x_1, x_2, x_3, \ldots, x_n$ be a collection of n numbers. The **range** of these numbers is the difference between the largest value among these numbers and the smallest value among these numbers.

9. Sample Variance and Standard Deviation

Let $x_1, x_2, x_3, \ldots, x_n$ be a sample.

1. The **sample variance** of these numbers, denoted s^2, is defined by

$$\begin{aligned} s^2 &= \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1} \\ &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{\sum(x-\bar{x})^2}{n-1}. \end{aligned}$$

2. The **sample standard deviation** of these numbers, denoted s, is defined by

$$\begin{aligned} s &= \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1}} \\ &= \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}. \end{aligned}$$

3. Note that the denominator of the above fractions is $n-1$.

10. Sample Variance and Standard Deviation from Frequency Table

Suppose that numerical data is given in a frequency table of the form

x	x_1	x_2	$\cdots$	x_n
f	f_1	f_2	$\cdots$	f_n

1. The sample variance of this data is computed by

$$\begin{aligned} s^2 &= \frac{f_1(x_1-\bar{x})^2+f_2(x_2-\bar{x})^2+f_3(x_3-\bar{x})^2+\cdots+f_n(x_n-\bar{x})^2}{f_1+f_2+f_3+\cdots+f_n-1} \\ &= \frac{\sum_{i=1}^{n} f_i(x_i-\bar{x})^2}{\left(\sum_{i=1}^{n} f_i\right)-1} = \frac{\sum f(x-\bar{x})^2}{\left(\sum f\right)-1}. \end{aligned}$$

2. The sample standard deviation of this data is computed by

$$\begin{aligned} s &= \sqrt{\frac{f_1(x_1-\bar{x})^2+f_2(x_2-\bar{x})^2+f_3(x_3-\bar{x})^2+\cdots+f_n(x_n-\bar{x})^2}{f_1+f_2+f_3+\cdots+f_n-1}} \\ &= \sqrt{\frac{\sum_{i=1}^{n} f_i(x_i-\bar{x})^2}{\left(\sum_{i=1}^{n} f_i\right)-1}} = \sqrt{\frac{\sum f(x-\bar{x})^2}{\left(\sum f\right)-1}}. \end{aligned}$$

11. Population Variance and Standard Deviation

Let $x_1, x_2, x_3, \ldots, x_n$ be a population.

1. The **population variance** of these numbers, denoted σ^2, is defined by

$$\begin{aligned}\sigma^2 &= \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}\\ &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} = \frac{\sum(x-\bar{x})^2}{n}.\end{aligned}$$

2. The **population standard deviation** of these numbers, denoted σ, is defined by

$$\begin{aligned}\sigma &= \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}}\\ &= \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}} = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}.\end{aligned}$$

3. Note that the denominator of the above fractions is n.

12. Probability Terminology

1. A **random variable** is a numerical variable the value of which depends upon a random phenomenon.
2. A **sample space** for a random variable is the set of all possible distinct outcomes for the random variable.
3. An **event** for a random variable is a subcollection of the sample space.
4. The **probability** of an event is a number between 0 and 1 (including those two numbers) that measures the likelihood of the event occurring.
5. If E is an event, the probability of E is denoted $P(E)$.

13. Discrete Probability

Suppose that a random variable has a finite sample space, in which all elements of the sample space are equally likely. Let E be an event for this random variable. Then

$$\begin{aligned}P(E) &= \frac{\text{number of elements in } E}{\text{number of elements in the sample space}}\\ &= \frac{\text{number of successful outcomes}}{\text{total number of possible outcomes}}.\end{aligned}$$

14. Set

1. A **set** is any collection of objects.
2. An **element** of a set is any object contained in the set.
3. A finite set S with elements $a_1, a_2, \ldots, a_n$ is written

$$S = \{a_1, a_2, \ldots, a_n\}.$$

15. Subset

Let A and S be sets. The set A is a **subset** of S if every element of A is also an element of S.

16. Complement

Let A and S be sets. Suppose that A is a subset of S. The **complement** of A in S, denoted A^c, is the set consisting of all elements of S that are not in A.

17. Probability of a Complement

Let S be a sample space and let E be an event. Then

$$P(\text{not } E) = 1 - P(E).$$

18. Intersection

Let A, B and S be sets. Suppose that A and B are subsets of S. The **intersection** of A and B in S, denoted $A \cap B$, is the set consisting of all elements of S that are in both A and B.

19. Probability of an Intersection

Let S be a sample space and let A and B be events. Suppose that A and B are independent. Then

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

This formula is also written

$$P(A \cap B) = P(A) \cdot P(B).$$

20. Union

Let A, B and S be sets. Suppose that A and B are subsets of S. The **union** of A and B in S, denoted $A \cup B$, is the set consisting of all elements of S that are in either A or B or both.

21. Disjoint

Let A, B and S be sets. Suppose that A and B are subsets of S. The sets A and B are **disjoint** (also called **mutually exclusive**) if there are no elements of S that are in both A and B.

22. Probability of a Disjoint Union

Let S be a sample space and let A and B be events. Suppose that A and B are disjoint. Then

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

This formula is also written

$$P(A \cup B) = P(A) + P(B).$$

23. Probability of a Union

Let S be a sample space and let A and B be events. Then

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

This formula is also written

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

24. Probability Density Function

A **probability density function** (abbreviated **PDF**) is a function that allows us to compute the probability of a continuous random variable X.

A probability density function satisfies the following conditions.

1. The probability density function is continuous.
2. The probability density function is never negative.
3. The area under the whole probability density function and above the x-axis is 1.
4. If a and b are numbers such that $a < b$, then $P(a < X < b)$ is the area under the probability density function and above the x-axis between the values $x = a$ and $x = b$.

25. Normal Distribution

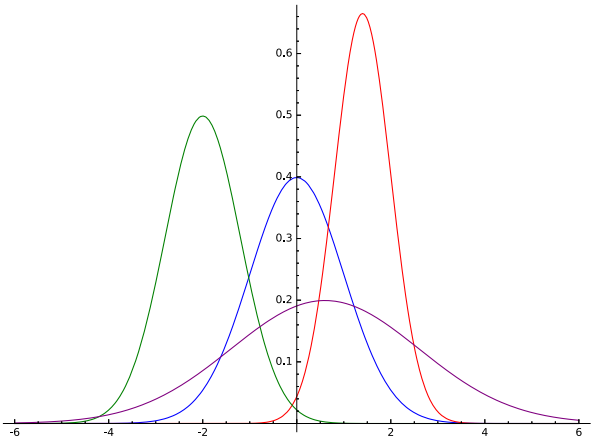
The **normal distribution** (also knows as the **Gaussian distribution**, or **bell-shaped curve**), is a type of probability density function that is very widely used in probability and statistics.

1. Each normal distribution is determined by two parameters, which are the mean, denoted μ, and the standard deviation, denoted σ. The value of μ can be any number, and the value of σ must be positive.

2. The normal distribution with mean μ and standard deviation σ is denoted $N(\mu, \sigma)$.

3. All normal distributions have similar shapes. The choice of μ moves the normal distribution to the right or left. The choice of σ makes the normal distribution either taller and thinner, or shorter and wider.

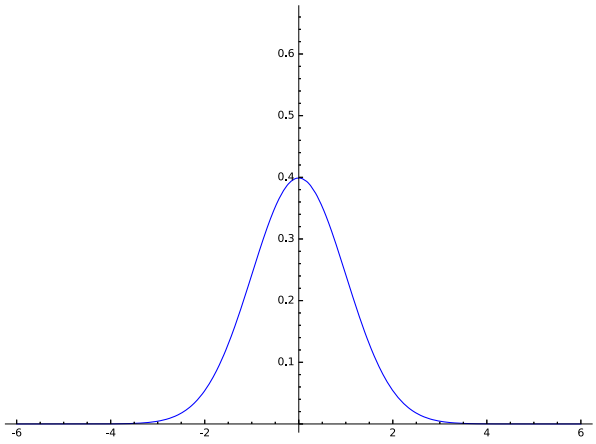
4. Some normal distributions are shown in below.



26. Standard Normal Distribution

1. The **standard normal distribution** is the normal distribution with $\mu = 0$ and $\sigma = 1$. The standard normal distribution is denoted $N(0, 1)$, also written $N(\mu = 0, \sigma = 1)$.

2. The graph of the standard normal distribution is shown below.



27. Z-Score Cutoff Points

Let c be a positive number less than 0.5.

1. The number $\mathbf{z_c}$ is defined to be the number such that $P\left(Z > z_c\right) = c$.

2. By symmetry, it follows that $P\left(Z < -z_c\right) = c$.

3. To find z_c, it is easier to find $-z_c$ first, and then obtain z_c by negating $-z_c$.

28. Standard Z-Scores

Here are the Z-scores for some standard intervals.

Size of Interval: 90%

Name	Top	Middle	Bottom
Interval	$[-1.282, \infty)$	$[-1.645, 1.645]$	$(-\infty, 1.282]$

Size of Interval: 95%

Name	Top	Middle	Bottom
Interval	$[-1.645, \infty)$	$[-1.960, 1.960]$	$(-\infty, 1.645]$

Size of Interval: 99%

Name	Top	Middle	Bottom
Interval	$[-2.326, \infty)$	$[-2.576, 2.576]$	$(-\infty, 2.326]$

29. X-Value to Z-Score: Generic

Let X be a random variable. Suppose X follows a normal distribution with mean μ and standard deviation σ. This normal distribution is transformed to the standard normal distribution via the transformation

$$Z = \frac{X - \mu}{\sigma}.$$

The value of Z obtained in this way is called the Z-score for X.

30. Probability for X-Values

Let X be a random variable. Suppose X follows a normal distribution with mean μ and standard deviation σ. Let a and b be numbers. Suppose $a < b$.

1.

$$P(X < b) = P\left(Z < \frac{b - \mu}{\sigma}\right).$$

2.

$$P(a < X) = P\left(\frac{a - \mu}{\sigma} < Z\right) = 1 - P\left(Z < \frac{a - \mu}{\sigma}\right).$$

3.

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right) \\ &= P\left(Z < \frac{b - \mu}{\sigma}\right) - P\left(Z < \frac{a - \mu}{\sigma}\right). \end{aligned}$$

31. 68-95-99.7 Rule

Let X be a random variable. Suppose X follows a normal distribution with mean μ and standard deviation σ. The following probabilities are called the **68-95-99.7 Rule**.

1. $$P(\mu - \sigma < X < \mu + \sigma) \approx 0.68 = 68\%.$$

2. $$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95 = 95\%.$$

3. $$P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997 = 99.7\%.$$

32. Z-Score to X-Value: Generic

Let X be a random variable. Suppose X follows a normal distribution with mean μ and standard deviation σ. Let Z be a Z-score for the standard normal distribution. Then Z is transformed to the corresponding X-value via the transformation

$$X = \mu + Z\sigma.$$

33. Distribution of Averages

Let X be a random variable (with any type of distribution). Suppose X has mean μ and standard deviation σ.

Suppose that X is sampled n times, and the mean $\bar{X}$ is computed. If sampling and computing the mean is done repeatedly, it leads to a new distribution, called the **sampling distribution**, which is the distribution of the mean $\bar{X}$.

1. The mean of the sampling distribution is called the **sample mean** and is denoted $\mu_{\bar{X}}$.

2. The standard deviation of the sampling distribution is called the **sample standard deviation** (also known as the **standard error**) and is denoted $\sigma_{\bar{X}}$.

34. Central Limit Theorem

Let X be a random variable (with any type of distribution). Suppose X has mean μ and standard deviation σ.

The **Central Limit Theorem** states that for large sample sizes n (generally 30 or more), the sampling distribution is approximately a normal distribution, and that

$$\mu_{\bar{X}} = \mu \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

35. Point Estimate

A **point estimate** is an estimate of a single parameter regarding the population.

36. Confidence Interval: Basic Idea

Let X be a random variable. Suppose a sample is taken, with sample mean $\bar{x}$.

1. The **level of confidence** (aka **confidence level**) for confidence intervals is a percentage strictly between 0% and 100%. The larger the percentage, the more confident we are in the result, though the harder it is to achieve that level of confidence. The level of confidence is specified by a positive number α that is less than 1, where the level of confidence equals $100(1-\alpha)\%$.

2. A **confidence interval** for the population mean μ at the $100(1-\alpha)\%$ level of confidence is an interval of the form $[\bar{x}-E, \bar{x}+E]$, for some number E, such that the probability that the interval $[\bar{x}-E, \bar{x}+E]$ actually contains the true value of μ is $100(1-\alpha)\%$.

3. The **margin of error** of a confidence interval of the form $[\bar{x}-E, \bar{x}+E]$ is the number E.

37. Confidence Intervals: Standard Z-Scores

Here are Z-scores for some standard confidence levels for confidence intervals.

Confidence level: 90%, $\alpha = 0.10$

Name	Top	Middle	Bottom
Interval	$[-1.282, \infty)$	$[-1.645, 1.645]$	$(-\infty, 1.282]$

Confidence level: 95%, $\alpha = 0.05$

Name	Top	Middle	Bottom
Interval	$[-1.645, \infty)$	$[-1.960, 1.960]$	$(-\infty, 1.645]$

Confidence level: 99%, $\alpha = 0.01$

Name	Top	Middle	Bottom
Interval	$[-2.326, \infty)$	$[-2.576, 2.576]$	$(-\infty, 2.326]$

38. Confidence Intervals: Z-Scores

The Z-score confidence interval at the $100(1-\alpha)\%$ level of confidence is

$$\left[-z_{\alpha/2}, z_{\alpha/2}\right].$$

39. Confidence Intervals: Z-Score to X-Value

Let X be a random variable. Suppose X follows a normal distribution with mean μ and standard deviation σ. Suppose a sample of size n is taken. Let Z be a Z-score for the standard normal distribution. Then for computing confidence intervals, the score Z is transformed to the corresponding X-value via the transformation

$$X = \bar{x} + Z \cdot \frac{\sigma}{\sqrt{n}}.$$

If the population standard deviation σ is not know, it is replaced by the sample standard deviation s, yielding the transformation

$$X = \bar{x} + Z \cdot \frac{s}{\sqrt{n}}.$$

40. Confidence Interval: Known Population Standard Deviation

Let X be a random variable. Suppose X follows a normal distribution with standard deviation σ. Suppose a sample of size n is taken, with sample mean $\bar{x}$. The **confidence interval** for the population mean μ at the $100(1-\alpha)\%$ level of confidence is

$$\left[\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right].$$

41. Confidence Interval: Unknown Population Standard Deviation

Let X be a random variable. Suppose X follows a normal distribution. Suppose a sample of size n is taken, with sample mean $\bar{x}$ and sample standard deviation s. The **confidence interval** for the population mean μ at the $100(1-\alpha)\%$ level of confidence is

$$\left[\bar{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right].$$

42. Margin of Error: Known Population Standard Deviation

Let X be a random variable. Suppose X follows a normal distribution with standard deviation σ. Suppose a sample of size n is taken, with sample mean $\bar{x}$. The **margin of error** for the population mean μ at the $100(1-\alpha)\%$ level of confidence is

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

The confidence interval can be written

$$[\bar{x} - E, \bar{x} + E].$$

43. Margin of Error: Unknown Population Standard Deviation

Let X be a random variable. Suppose X follows a normal distribution. Suppose a sample of size n is taken, with sample mean $\bar{x}$ and sample standard deviation s. The **margin of error** for the population mean μ at the $100(1-\alpha)\%$ level of confidence is

$$E = z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}.$$

The confidence interval can be written

$$[\bar{x} - E, \bar{x} + E].$$

44. Confidence Interval: Sample Size

Let X be a random variable. Suppose X follows a normal distribution with standard deviation σ. Let E be a positive number.

1. The minimum sample size n needed to obtain a confidence interval with margin of error E and $100(1-\alpha)\%$ level of confidence is

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}.$$

2. The sample size obtained using the above formula must be *rounded up* to be a whole number

45. Hypothesis Testing: Basic Idea

Let X be a random variable. Suppose X follows a normal distribution. Suppose a sample of size n is taken, with sample mean $\bar{x}$.

1. The **level of significance** (aka **significance level**) for hypothesis testing is a percentage strictly between 0% and 100%. The smaller the percentage, the more confident we are in the result, though the harder it is to achieve that level of significance. The level of significance is specified by a positive number α that is less than 1, where the level of significance equals $100\alpha\%$.

2. The **null hypothesis**, denoted H_0, is the hypothesis we assume to be true unless there is sufficient evidence to reject it.

3. The **alternative hypothesis**, denoted H_a, is the hypothesis we would support if there is sufficient evidence, and only if there is sufficient evidence, to reject the null hypothesis.

4. A **hypothesis testing** for the sample mean $\bar{x}$ at the $100\alpha\%$ level of significance is a procedure to determine whether there is sufficient evidence to deduce that the sample mean $\bar{x}$ is sufficiently different from what the null hypothesis states that the null hypothesis should be rejected.

46. Three Cases of Hypothesis Testing

There are three cases of hypothesis testing for the mean, depending upon the form that the null hypothesis takes.

Let μ_0 be a number.

	Left tail	Two-sided	Right tail
Actual Null Hypothesis	$H_0 : \mu \geq \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu \leq \mu_0$
Used Null Hypothesis	$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$
Alternative Hypothesis	$H_a : \mu < \mu_0$	$H_a : \mu \neq \mu_0$	$H_a : \mu > \mu_0$

The name of each of the three cases corresponds to the nature of where the region to reject the null hypothesis is located.

47. Hypothesis Testing: Z-Scores

Let μ_0 be a number. Let α be the level of significance for hypothesis testing.

The Z-scores for rejecting and not rejecting the null hypothesis are as follows.

	Left tail	Two-sided	Right tail
Actual Null Hypothesis	$H_0 : \mu \geq \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu \leq \mu_0$
Used Null Hypothesis	$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$
Alternative Hypothesis	$H_a : \mu < \mu_0$	$H_a : \mu \neq \mu_0$	$H_a : \mu > \mu_0$
Non-Rejection Region	$\left[-z_\alpha, \infty\right)$	$\left[-z_{\alpha/2}, z_{\alpha/2}\right]$	$\left(-\infty, z_\alpha\right]$
Rejection Region	$\left(-\infty, -z_\alpha\right]$	$\left(-\infty, -z_{\alpha/2}\right]$ & $\left[z_{\alpha/2}, \infty\right)$	$\left[z_\alpha, \infty\right)$

48. Hypothesis Testing: Standard Z-Scores

Here are Z-scores for some standard levels of significance for hypothesis testing.

Significance level: 10%

	Left tail	Two-sided	Right tail
Null Hypothesis	$H_0 : \mu \geq \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu \leq \mu_0$
Alternative Hypothesis	$H_a : \mu < \mu_0$	$H_a : \mu \neq \mu_0$	$H_a : \mu > \mu_0$
Non-Rejection Region	$[-1.282, \infty)$	$[-1.645, 1.645]$	$(-\infty, 1.282]$
Rejection Region	$(-\infty, -1.282]$	$(-\infty, -1.645]$ & $[1.645, \infty)$	$[1.282, \infty)$

Significance level: 5%

	Left tail	Two-sided	Right tail
Null Hypothesis	$H_0 : \mu \geq \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu \leq \mu_0$
Alternative Hypothesis	$H_a : \mu < \mu_0$	$H_a : \mu \neq \mu_0$	$H_a : \mu > \mu_0$
Non-Rejection Region	$[-1.645, \infty)$	$[-1.960, 1.960]$	$(-\infty, 1.645]$
Rejection Region	$(-\infty, -1.645]$	$(-\infty, -1.960]$ & $[1.960, \infty)$	$[1.645, \infty)$

Significance level: 1%

	Left tail	Two-sided	Right tail
Null Hypothesis	$H_0 : \mu \geq \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu \leq \mu_0$
Alternative Hypothesis	$H_a : \mu < \mu_0$	$H_a : \mu \neq \mu_0$	$H_a : \mu > \mu_0$
Non-Rejection Region	$[-2.326, \infty)$	$[-2.576, 2.576]$	$(-\infty, 2.326]$
Rejection Region	$(-\infty, -2.326]$	$(-\infty, -2.576]$ & $[2.576, \infty)$	$[2.326, \infty)$

49. Hypothesis Testing: Z-Score to X-Value

Let X be a random variable. Suppose X follows a normal distribution with mean μ and standard deviation σ. Suppose a sample of size n is taken. Let Z be a Z-score in the standard normal distribution. Then for hypothesis testing with Null Hypothesis $\mu = \mu_0$ (or $\mu \leq \mu_0$ or $\mu \geq \mu_0$), the score Z is transformed to the corresponding X-value via the transformation

$$X = \mu_0 + Z \cdot \frac{\sigma}{\sqrt{n}}.$$

If the population standard deviation σ is not know, it is replaced by the sample standard deviation s, yielding the transformation

$$X = \mu_0 + Z \cdot \frac{s}{\sqrt{n}}.$$

50. Hypothesis Testing: X-Values

Let μ_0 be a number. Let α be the level of significance for hypothesis testing.

The X-values for rejecting and not rejecting the null hypothesis are as follows.

If $\bar{x}$ is in the non-rejection region, the null hypothesis is not rejected, and if $\bar{x}$ is in the rejection region, the null hypothesis is rejected.

Left tail

Null Hypothesis: $\mu \geq \mu_0$

Alternative Hypothesis: $\mu < \mu_0$

Non-Rejection Region: $\left[\mu_0 - z_\alpha \cdot \frac{s}{\sqrt{n}}, \infty\right)$

Rejection Region: $\left(-\infty, \mu_0 - z_\alpha \cdot \frac{s}{\sqrt{n}}\right]$.

Two-sided

Null Hypothesis: $\mu \geq \mu_0$

Alternative Hypothesis: $\mu < \mu_0$

Non-Rejection Region: $\left[\mu_0 - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \mu_0 + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right]$

Rejection Region: $\left(-\infty, \mu_0 - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right]$ & $\left[\mu_0 + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \infty\right)$.

Right tail

Null Hypothesis: $\mu \geq \mu_0$

Alternative Hypothesis: $\mu < \mu_0$

Non-Rejection Region: $\left(-\infty, \mu_0 + z_\alpha \cdot \frac{s}{\sqrt{n}}\right]$

Rejection Region: $\left[\mu_0 + z_\alpha \cdot \frac{s}{\sqrt{n}}, \infty\right)$.

51. Hypothesis Testing: Errors

The possible errors in hypothesis testing are summarized in the following chart.

		Actually True or False	
		H_0 is true	H_0 is false
Our Decision	H_0 is not rejected	Good	Type 2 Error
	H_0 is rejected	Type 1 Error	Good

1. A **type 1 error** for hypothesis testing is when we reject the null hypothesis, even though it is actually true.

2. A **type 2 error** for hypothesis testing is when we do not reject the null hypothesis even though it is false.