The test will cover Chapters 7, 13, and 12.

CHAPTER 7: THE MATHEMATICS OF NETWORKS

Sample Exercises: 1, 2, 3, 5, 19, 20, 21, 22, 23, 24, 27, 29, 30, 35, 39, 41, 42, 54, 57, 58, 59

Concepts: Understand the following.

- **Trees.** A tree is a special kind of graph. Sometimes it sort of looks like a real world tree, but it might not. Know the four properties of a tree. Be able to tell me if a graph is a tree or not. (See page 273–275, and Exercises 1, 2, 3, 5 on pages 293–294.)

- **Minimum Spanning Tree.** The trees that we care most about are the spanning trees (that connect to all the vertices). In networking, we usually want a spanning tree with the smallest possible total weight, called a minimum spanning tree (MST). We use Kruskal’s Algorithm to find the MST. (See pages 276–278, and Exercises 19-24 on pages 297–298.)

- **Steiner Points.** The only way to do better than the minimum spanning tree (MST) is to add new vertices and use them to form new networks. Sometimes you cannot improve on the MST, but sometimes you can. If you can improve on the MST the best type of points to add are Steiner points. These have three edges connected to them and the angles are all 120 degrees. (See pages 279–282.)

- **Non-Steiner Points** You can try adding non-Steiner points, but the results will not be as good. (But they still can be better than the MST). (See page 280, Figure 7-15, page 285 Figure 7-22 (c), and page 286 figure 7-24. See also Exercise 41 on page 302, or Exercise 57 on page 305.)

- **The 120 degree rule.** If the largest angle in a triangle is 120 or more, then you do not want to use a Steiner point. In this case, just choose the two shortest sides of the triangle (which is the MST). The moral: sometimes the MST cannot be improved upon. (See pages 283–284, and Exercises 27, 29, 30 on page 299.)

- **Shortest Network.** Suppose you want the shortest network connecting N points. Then sometimes you need to add Steiner points. If N is large, then finding the best Steiner points and the best connections between these points is a difficult problem. You just need to know the concept, you do not need to actually find the shortest network except for triangles and rectangles (including squares). You should know that sometimes you want no Steiner points, but usually you want several. (See page 287-288 and the picture on page 291.)

Methods: Be able to use the following.

- **Kruskal’s Algorithm.** An easy way to find the minimum spanning tree (MST) is Kruskal’s Algorithm. This is easy to learn, and you should know it well. You
should know that this algorithm is both efficient and optimal. (See pages 276–278, and Exercises 19-24 on pages 297–298.)

- **Torricelli’s Construction.** You can use a freehand sketch, or a ruler and compass. I might give you a triangle, and ask you to draw the approximate location of the Steiner point, and illustrate the shortest network. (See page 282. See Exercise 35 on page 300.)

- **30-60-90 Triangles.** If I give you one side of such a triangle, be able to find the other two sides. (See class notes, and Exercise 39 on Page 301. See also Exercise 41 and 42 on page 302, and Exercise 58 and 59 on page 305.)

- **45-45-90 Triangles.** If I give you one side of such a triangle, be able to find the other two. (See class notes, Exercise 57 on page 305, and page 285 Figure 7-22 (c.).)

- **Pythagorean Theorem.** If I give you two sides of a right triangle, be able to find the other side using the Pythagorean Theorem. (See Exercise 54 on page 304. This exercise also requires Kruskal’s algorithm.)

- **Steiner Points in Equilateral Triangles.** Be able to find the length of the shortest network of an equilateral triangle by working out all the distances to the Steiner point. (See Exercise 42 on Page 302, or Figure 7-15 (c) on page 280.)

- **Steiner Points in Squares and other Rectangles.** Here there are two Steiner points you need to use. Be able to find the length of the shortest network of a square or rectangle by working out all the distances to the Steiner points. (See Exercises 58 and 59 on page 305.)

- **Non-Steiner Points.** Be able to find the length of networks even if they involve non-Steiner points. I will stick to easy situations where you can use 30-60-90 triangles, 45-45-90 triangles, or the Pythagorean Theorem. (See page 280 Figure 7-15 and page 285 Figure 7-22 (c). See also Exercise 41 on page 302, or Exercise 57 on page 305.)

**Results:** Know the following.

- **Approximation Theorem.** Know that the MST (which uses no Steiner points) is guaranteed to be within 13.4% of the shortest network. So if that percent increase is acceptable, you do not need to bother with Steiner points. Know that this was proved very recently (in 1990). (See page 289.)

- **Shortest Network Rule.** (See page 288.)

---

**Chapter 13: Sampling, Census, and Clinical Studies**

**Sample Exercises:** 2, 15–16, 17–20, 23, 25, 26, 27, 29, 30, 39

**Concepts:** Understand the following.

- **Population.** This is the individuals or objects that we are interested in. Keep in mind it can be animate or inanimate: for example, the coins in a jar can be the population. We usually use $N$ to denote the size of the population. (See pages 518–519.)

- **Census.** This is where you try to get information from every member of the population. In practice, there might be some members that are not counted. There are statistical methods for estimating the number that are not counted, but the Supreme Court ruled in 1999 that they cannot be used for the U.S. Census for the purposes of apportionment.
• **Differential Undercount.** This is where certain groups are undercounted in a Census more than others. This was a big concern in the 1990 census, leading to a Supreme court case. (See page 520.)

• **Surveys and Sampling.** This is where you try to get information about the population by only sampling a fraction of the population. If this is done correctly, you usually get better results with a relatively small number of samples than using a biased large sample or a flawed census. We use the letter \( n \) to denote the size of the sample. The fraction \( n/N \) is called the *sampling rate.* A larger sampling rate does not necessarily result in a more accurate survey.

• **Chance Error.** When you sample a fraction of the population, there is a chance that you are not getting a representative sample. Mathematicians and statisticians can tell you what your chance error is likely to be. For example, if your sample size is \( n = 1500 \) the chance error is usually small, even if the population is a large city or the whole country. But if you used \( n = 15 \), you would expect a much larger error. (See page 529.)

• **Selection Bias.** There is another way in which your sample might not be a representative sample. That is if you employ a sampling method that tends to select some subsets of the population at the expense of others. (See page 523, page 529, Exercise 15 on Page 539, and Exercise 29d on page 541.)

• **Nonresponse Bias.** This is a form of selection bias where some subsets of the population choose not to respond to the survey at a higher rate that other subsets. (See page 523.)

• **Quota Sampling.** This is a method for trying to avoid selection bias. In this method the collector of data is assigned certain numbers for various groups called the *quotas,* and must sample the assigned quota from each group. History has shown that this method does not get rid of all the selection bias since it still gives the data collector too much freedom in choosing the sample that is convenient for him or her. (See page 524–525. What about Exercise 15 on Page 539?)

• **Random Sampling.** This is a method for avoiding selection bias. In this method you employ a random method to decide the sample. You do not give the data collectors any choice on deciding the sample. (See pages 526–527.)

• **Simple Random Sampling.** This is a form of random sampling where you make full list of the population, and randomly choose a subset for sampling. This is the best way to get rid of selection bias (especially if you can avoid non-response bias), but it can be expensive if done on a national level. (See pages 526–527, and Exercise 19 on page 540.)

• **Statified Random Sampling.** This is the main method used today in public opinion polling. It is cheaper than simple random sampling. In this method the population is divided into subsets, call them *communities.* Then a certain number of communities are chosen at random. Then only the chosen communities are subdivided into zones called *wards* say. From each chosen community, we choose at random a certain number of wards. Only chosen wards are subdivided. This subdivision process continues a few more times (perhaps the wards are divided into *precinct,* and the randomly selected precincts are divided into households). The final sample is
still randomly determined, but not in a simple one-step process. (There are many variants of this method, and some are less random than others). (See page 527–528).

- **Clinical study, Clinical trial and Confounding Variables.** One tries to determine if a factor (or variable) increases or decreases the chance or risk of a certain outcome. For example, does a certain drug increase the chances of recovering from a certain illness? Does a certain food additive increase the risk of cancer? What makes answering these questions difficult is the presence of other factors, called *confounding variables* that can also potentially cause or prevent the outcome. In Clinical studies, one tries to eliminate confounding variables. (See page 532 and Exercise 39 on Page 542).

- **Controlled Study** (or Controlled Experiment). One way to eliminate the confusion caused by confounding variables is to divide the subjects into two groups: the *control group* and the *treatment group*. The control group does not get the treatment. This method does not try to eliminate confounding variables, but the confounding variables should be present to the same degree in both groups (hopefully). So the differences between the groups can be attributed to the treatment. Such studies are called *controlled studies*. (See page 532–533 and Exercise 39 on Page 542. What about Exercise 29d on page 541?)

- **Control Group.** In a controlled study, this is the group that does not receive treatment. (See page 532–533.)

- **Treatment Group.** In a controlled study, this is the group that does receive treatment. (See page 532–533 and Exercise 39 on Page 542.)

- **Placebo and Placebo Effect and Controlled Placebo Study.** In many situations, you do not want the subject to know if he or she is in the control group or the treatment group. A common way to do this is to give the subjects in the control group a fake treatment called a *placebo*. Such studies are called *controlled placebo studies*. One reason this is necessary is due to the *placebo effect* where a patient feels better when he or she thinks that he or she is receiving treatment. (See pages 532–533 and Exercise 39 on Page 542.)

- **Blind Study.** This is a clinical study where subjects do not know what group they are in: they do not know if they are in the control group or the treatment group. A study that uses placebos is an example of a blind study when subjects do not know if they are getting the placebo or the real treatment. (See pages 532–533 and Exercise 39 on Page 542.)

- **Double-Blind Study.** Even the person treating the patients can introduce confounding variables by treating the control group slightly differently than the treatment group. The way to control for this is to conduct a double-blind study, where the person treating the subjects doesn’t even know which group each subject is in. (See page 333. What about Exercise 39 on Page 542?)

- **Randomized Controlled Study.** In order to cause the confounding variables to be equally present in both the control and treatment groups, one should randomly distribute the subjects into each group. (See page 532)
Methods: Be able to use the following.

- **Sample Rate.** The sample rate is \( n/N \) where \( n \) is the sample size (that is, the number of responses) and \( N \) is the size of the population. (See Page 529, Exercise 2b on page 537, and Exercise 18 on page 539)

- **Capture-Recapture Method.** First you capture a certain number of objects: \( n_1 \). Next you mark or tag these \( n_1 \) objects and put them back. After the objects are mixed into the population, you recapture a certain number \( n_2 \). Count the number \( k \) of marked objects in this sample. Then you estimate that the fraction marked in the recapture is approximately equal to the fraction marked in the whole population. In symbols

\[
\frac{k}{n_2} \approx \frac{n_1}{N}.
\]

Now solve for \( N \) to get an estimate for the population.

For example, if you have a jar of pennies, and you grab a handful of \( n_1 = 30 \) pennies. Then you would mark them, and mix them back in the jar. Then grab a new handful of \( n_2 = 25 \), say. Suppose \( k = 5 \) of these are marked. Then \( 5/25 = k/n_2 \) is the proportion marked in the sample. You know that \( 30/N = n_1/N \) is the fraction marked in the whole population, but you haven’t figured out the population \( N \) yet. You estimate that

\[
\frac{5}{25} \approx \frac{30}{N}.
\]

Solving for \( N \) gives you

\[
5N \approx 30 \times 25 \quad \text{so} \quad N \approx 150.
\]

Thus you estimate that there are 150 pennies in the jar. (See pages 529–530 and Exercises 23, 25, 26, 27 on Pages 540–541.)

History: Know the following.

- **The 1936 Election.** (See pages 522-524.) Before the 1936 election, the Literary Digest had a perfect record for predicting the winner of presidential races (since 1916). The Literary Digest was one of the most respected magazines of the time. It mailed mock ballots to all its subscribers, as well as to all people it could find from phone books, club membership lists, professional organization lists etc. This was a massive, expensive undertaking. About 24% of these people returned their sample ballots: around 2.4 million people in all. It predicted that the Republican Landon would defeat Roosevelt (FDR) in a landslide. When the election day came, FDR won in a landslide. This massive mistake cost the magazine its credibility, and it soon went out of business. Why do you suppose its prediction was so far off? What sort of selection bias was involved?

  During this election, George Gallup used quota sampling to give an accurate prediction. He only sampled 50,000 people, a small fraction of the size of the Literary Digest’s sample. The moral is that a larger sample is not always a better sample.

- **The 1948 Election.** (See pages 524-527.) In this election George Gallup and others used Quota Sampling to try to eliminate selection bias. Dewey was predicted to beat Truman. In the actual election Truman won. After this quota sampling fell out of favor, and was replaced by random sampling. (See pages 524–526.)
• **The 1990 Census.** The 1990 census undercounted so many people (about 1.6% or 4 million people) that it led to a Supreme court case. One side advocated that statistical methods be applied to the raw numbers to estimate the number of people who were not counted. The Supreme court decided that, for the purposes of apportionment at least, only raw numbers could be used. The last census in 2000 undercounted fewer people than the 1990 census, but still a significant number. (See pages 519–520.)

**Chapter 12: Fractals**

*Sample Exercises:* 1, 3, 9, 10, 11, 12, 17, 33, 35, 41, 53, 58, 60

**Concepts:** Understand the following.

- **Self Similarity** (or Similarity of Scale). Where small pieces of an object or shape look like bigger pieces. In other words, the small scale structure of the object or shape is similar to the small-scale structure. (See page 483.)
- **Fractals.** The three main features of a fractal are (i) they are generated by repeating simple rules, for example recursive replacement rules, (ii) they exhibit self-similarity, (iii) they often have an infinite perimeter or area in a finite space.
- **Koch snowflake.** Has infinite perimeter and finite positive area. Know the recursive replacement rule and how to draw it. (See pages 478–480.)
- **Serpenski Gasket.** Has infinite perimeter and zero area. Know the recursive replacement rule and how to draw it. (See page 483–485.)
- **Other Fractals.** These use other recursive replacement rules or other types of rules. These can be used to make all kinds of interesting shapes, including some that look like mountains, rivers, coastlines, clouds, snowflakes, and trees. For example, a fractal cloud might have infinite surface area and finite volume. (See Page 485-489 for other types of rules. See Pages 486–489 for a natural looking fractals.)
- **Mandelbrot Sequences.** If \( s \) is a number, then the Mandelbrot sequence is defined by the rule \( s_{n+1} = s + s_n^2 \). If the \( s_n \) keep repeating the same numbers, the seed \( s \) is *periodic*. If \( s_n \) gets closer to a certain number, the seed \( s \) is *attracted*. If \( s_n \) gets larger and larger in size, the seed \( s \) is *escaping*. The number \( s \) can be a real number or a complex number. (See pages 494–496.)
- **The Mandelbrot Set.** This is made up of complex numbers \( a + bi \). If \( s = a + bi \) is periodic or attracted, then \( s \) is inside the Mandelbrot set. If \( a + bi \) is escaping, then it is outside the Mandelbrot set. (See pages 491–497.)
- **Naturalistic Fractals.** By adding some randomness to the fractal rule, fractals can be made to look like natural objects such as mountains, rivers, coastlines, clouds, and trees. (See Pages 486–489 for a natural looking fractals.)
- **Complex Numbers.** Traditional real numbers are only one-dimensional numbers: they fit on a line. Complex numbers are two-dimensional numbers. They look like \( a + bi \) where \( a \) is the real or horizontal part, and \( b \) is the vertical part. The number \( i \) has the special property that \( i^2 = -1 \). (See page 494.)

**Methods:** Be able to do the following:

- **Sketch Fractals.** Be able to follow a recursive replacement rule to sketch the first few stages of a fractal. (See Exercises 1, 9, 17 on pages 500 – 503.)
• **Find Patterns and Formulas for Perimeter.** From these patterns or formulas, be able use the concept of exponential growth to show perimeter is infinite. (See Page 480, and Exercises 3 and 11 on pages 500–501.)

• **Find Patterns and Formulas for Area.** In certain cases, like the Sierpinski Gasket, you should be able to take the pattern or formula you found, and use the concept of exponential decay to show area is zero. (See Exercises 12 on page 501 and Exercise 41 on page 508. Also, for extra credit: Exercise 58 on page 510.)

• **Find Other Patterns and Formulas.** (See Exercise 10 on page 501.)

• **Mandelbrot Sequence.** Compute the Mandelbrot sequence $s_0, s_1, s_2, \ldots$ from a certain seed $s$. From this, be able to tell if a seed is escaping, periodic, or attracting. Be able to tell if a seed is in or out of the Mandelbrot set. (See page 494–497, and Exercises 33, 35, 53, 60 on pages 507–510.)

• **Complex Numbers.** Be able to work with complex numbers. (See page 494, and Exercise 60 on page 510.)

• **Infinite Series** (Extra Credit). Be able to add up an infinite number of areas in order to find total area. (See Exercise 58 on page 501. See also pages 480–482.)