GEMS 100 – Section Summaries

8: Regression and Correlation

A **least squares regression line** is a line that best fits two-variable data by minimizing the sum of the squares of the vertical distances between the points in the scatterplot and the line. Although the equation for this line can be found by using a complicated formula, it is more easily found with a graphing calculator.

Correlation measures both the strength and the direction of a linear relationship between two numerical variables. Correlation is usually denoted by the letter r, and values for correlations are between -1 and 1 inclusive. A correlation close to 1 means that there is a strong positive relationship between the two variables. In other words, as one variable gets larger, so does the other. A correlation close to -1 means that there is a strong negative relationship. In other words, as one variable gets larger, the other gets smaller. A correlation close to zero indicates that there is no relationship between the variables.

9: Exponential Functions

Functions that increase (or decrease) by a fixed percentage are exponential functions. An **exponential function** is a function where the growth factor is constant.

Symbolically, an exponential function is

$$f(x) = ba^x.$$

where $a > 0$ and $a \neq 1$. The y-intercept (or starting value) is b, and the **growth factor** is a. So, $b = f(0)$ and $a = f(x+1)/f(x)$. The growth factor, a, is also called the **base.**

Exponential functions of the form $f(x) = ba^x$ contain the points $(0, b)$ and $(1, ba)$. If b is positive, then the domain is all the real numbers and the range is $f(x) > 0$. Again assuming b is positive, an exponential function will be increasing if the growth factor is greater than 1 and decreasing if the growth factor is between 0 and 1. In either case, the function is always concave up with the x-axis as a **horizontal asymptote.**

When exponential functions are given verbally, they often include a growth rate. A **growth rate** is always 1 less than the corresponding growth factor.

10: Logarithmic Functions

The **magnitude** of a number a is the exponent b such that $10^b = a$. Magnitudes can be defined using other numbers besides powers of 10, but for this section, we chose 10 because our number system is in powers of 10. A magnitude is just another word asking for an exponent with a given base.

The function whose input is a number and whose output is its magnitude is called the **logarithm** function or, simply, the **log** function. In this section, we only looked at logarithm functions with a base of 10. Therefore,

$$\log x = y \qquad \text{is equivalent to} \qquad 10^y = x.$$

Some basic properties of $f(x) = \log x$ are the following.

- $\log(1) = 0$.
- $\log x$ is negative for $0 < x < 1$ and positive for $x > 1$.

- $f(x) = \log x$ has a domain of $x > 0$ and a range of all real numbers.
- The graph of $f(x) = \log x$ is increasing and concave down.

Two other important properties of logarithms are the following.

- $\log(ab) = \log a + \log b$ for any positive numbers a and b.
- $\log(a^r) = r \log a$ for any real number r and positive number a.

11: Periodic Functions

Repetitious behavior can be described mathematically as being periodic. To be periodic, something must not only repeat itself, but repeat itself at regular intervals. A **periodic function** is a function that gives the same output for inputs a fixed distance apart. Symbolically, we say that f is periodic if for some real number p, $f(x) = f(x + p)$ for all x. The **period** of this function is the smallest value of p for which this relationship is always true; that is, the period is the *minimum* fixed distance where the outputs are always the same.

The period is associated with the input of a function, whereas the amplitude is associated with the output. Suppose a horizontal line is drawn through a graph of a periodic function halfway between the function's maximum and minimum values; the **amplitude** is the distance from this horizontal line to the maximum (or minimum) value. The amplitude can also be defined as $\frac{1}{2}(M - m)$, where M is the function's maximum value and m is the function's minimum value.

We often use the sine and cosine functions to describe periodic behavior symbolically. These functions can be defined using a unit circle (a circle with radius 1). Consider an arbitrary point P on a unit circle, centered at the origin. The sine and cosine functions are defined in terms of the position of point P. The input is the distance on the circle from $(1, 0)$ *counterclockwise* to P. In other words, x is the arc length from (1, 0) to the point P. **Sine** of x (abbreviated sin x) is the vertical position of P, and **cosine** of x (cos x) is the horizontal position of P. The coordinates of P are therefore (cos x, sin x).

12: Power Functions

A **power function** is a function of the form $f(x) = kx^p$.

For a power function of the form $f(x) = x^p$, where p is a positive integer, $f(0) = 0$ and $f(1) = 1$. If p is even, then the function has a domain of all real numbers and a range of $f(x) \geq 0$. If p is odd, then the function has a domain of all real numbers and a range of all real numbers.

Suppose f is a power function of the form $f(x) = kx^p$, where $k > 0$. If $p > 1$ then the function is increasing and concave up in the first quadrant. If $0 < p < 1$ then the function is increasing and concave down in the first quadrant. If $p < 0$ then the function is decreasing and concave up in the first quadrant. Also, instead of $f(1) = 1$ we now have $f(1) = k$.

The exponent of a power function may not be an integer. It may be a rational number instead. A **rational** number is one that can be written as the quotient of two integers. These types of exponents can also be written as radicals. We define $x^{1/n} = \sqrt[n]{x}$ and $x^{m/n} = \sqrt[n]{x^m}$.

13: Probability

For two events, the **general counting method** states that if event A can occur a different ways and after it has occurred event B can occur b different ways, then the joint event A and B

(or event A followed by B) can occur $a \cdot b$ different ways. The general counting method is used when items are being selected from more than one group (like sandwiches, side orders, and drinks) or items are being selected from one group and the items can be repeated (like numbers on two dice).

Finding probabilities of equally likely events is connected with finding the sample space. The **sample space** is the set of all equally possible outcomes. An **event** is a subset of outcomes of the sample space. The **probability of an event** (where each item in the sample space has an equally likely chance of occurring) is defined as:

$$\frac{\text{the number of outcomes of an event}}{\text{the total number of outcomes in the sample space.}}$$

Whatever the event is, the probability of that event, $P(E)$, is a number between 0 and 1 (i.e. $0 \leq P(E) \leq 1$). The **complement** of an event is the set of all outcomes in the sample space that are not in the event. For any event, E, and its complement, E', we have $P(E) + P(E') = 1$. This means that $P(E') = 1 - P(E)$.

Two events are **independent** if the occurrence of one does not change the probability of the other occurring. If two events, A and B, are independent, the probability of both A and B occurring is $P(A \text{ and } B) = P(A) \cdot P(B)$.

A **probability distribution** is the collection of all outcomes of a random phenomenon together with their associated probabilities. To find the **mean of a probability distribution**, you multiply the outcomes by the respective probability and add the products. The mean of a probability distribution is also called the **expected value**.

14: Random Samples

To obtain information about a population, it is often difficult to actually examine the entire population. One therefore must look at a sample of the population and from that sample make inferences about the population. A **population** consists of the entire group of people or objects from which you would like some information. A **sample** is the part of the population that you actually examine. A **parameter** is a number that describes a population. A **statistic** is a number computed from a sample of the population.

A **confidence interval** is an interval estimate computed from sample data such that the interval has a given probability of containing the population parameter. The formula used to determine a confidence interval for a population proportion is:

$$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

In this formula, $\hat{p}$ is the sample proportion and n is the sample size. The **confidence level** is the probability that the interval contains the population parameter. The most commonly used confidence level is 95%. In that case, $z = 1.96$.

Two frequently used types of samples are a simple random sample and a stratified random sample. A **simple random sample** of size n consists of n units taken from a population in such a way that every set of n units has an equal probability of being in that sample. In a **stratified random sample**, the population is first divided into different groups and then simple random samples are taken from each group.

When selecting a sample to determine information about the entire population, there are two things that should be minimized - bias and variability. **Bias** exists when the results obtained from an experiment are in some way prejudiced. **Variability** exists when the results you obtain vary significantly when you use a different sample.