Chapter 5 Notes:

- Random Variable
  - A **random variable** is a variable (typically represented by *x*) that has a single numerical value, determined by chance, for each outcome of a procedure.
  - A **probability distribution** is a description that gives the probability for each value of the random variable. It is often expressed in the format of a table, formula, or graph.

A **discrete random variable** has a collection of values that is finite or countable.

Example: computer ownership in a random classroom with 20 students:

X (num de computers)	Count	P(x)
0	1	0.05
1	8	0.40
2	7	0.35
3	4	0.20

Deduce the formula for the mean and the standard deviation for a probability distribution:

$$\mu = \sum [x \cdot P(x)]$$

$$\sigma^2 = \sum \left[(x-\mu)^2 \cdot P(x) \right]$$ (This format is easier to understand.)

$$\sigma = \sqrt{\sum \left[x^2 \cdot P(x) \right] - \mu^2}$$

Another example:

The experiment, described by Spiegel, 2011, consists of five pennies that were tossed 1000 times. At each toss the number of heads was observed. The results were summarized in the following table.

Number of heads	Count	P(x)
0	38	0.038
1	144	0.144
2	342	0.342
3	287	0.287
4	164	0.164
5	25	0.025

- Deducing that the sum of p(x) = 1.
- Rational of the formula for the mean and standard deviation of the freq. distribution.
- Relate the empirical mean of the experiment and the law of large numbers.

Range Rule of Thumb for Identifying Significant Values

- **Significantly low** values are ($\mu - 2\sigma$) or lower.
- **Significantly high** values are ($\mu + 2\sigma$) or higher.
- **Values not significant**: Between ($\mu - 2\sigma$) and ($\mu + 2\sigma$).
- **Identifying Significant Results with Probabilities:**

  Significantly high number of successes:

  - *x* successes among *n* trials is a **significantly high** number of successes if the probability of *x* or more successes is 0.05 or less. That is, *x* is a significantly high number of successes if *P*(*x* **or more**) ≤ 0.05.

The value 0.05 is not absolutely rigid. Other values, such as 0.01, could be used to distinguish between results that are significant and those that are not significant.

- Significantly low number of successes:

  - *x* successes among *n* trials is a **significantly low** number of successes if the probability of *x* or fewer successes is 0.05 or less. That is, *x* is a significantly low number of successes if *P(x* **or fewer***)* ≤ 0.05.

The value 0.05 is not absolutely rigid. Other values, such as 0.01, could be used to distinguish between results that are significant and those that are not significant.

The Rare Event Rule for Inferential Statistics:

If, under a given assumption, the probability of a particular outcome is very small and the outcome occurs **significantly less than or significantly greater than** what we expect with that assumption, we conclude that the assumption is probably not correct.

Binomial Probability Distribution:

A **binomial probability distribution** results from a procedure that meets these four requirements:

- The procedure has a **fixed number of trials.** (A trial is a single observation.)
- The trials must be **independent,** meaning that the outcome of any individual trial doesn't affect the probabilities in the other trials.
- Each trial must have all outcomes classified into exactly **two categories,** commonly referred to as **success** and **failure.**
- The probability of a success remains the same in all trials.

Binomial Probability Formula:

$$P(x) = \frac{n!}{(x-n)!\,x!} \cdot p^x \cdot q^{n-x} \quad \text{for } x = 0, 1, 2, 3, \ldots, n$$

where

n = number of trials

x = number of successes among n trials

p = probability of success in any one trial

q = probability of failure in any one trial ($q = 1 - p$)

Explain the rationale of the formula stated above.

For Binomial Distributions:

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

Standard Deviation: $\sigma = \sqrt{npq}$