I. Inferential statistics: make inferences (statements) about populations based on the samples.

To make inferences we need to know how to find the probability of obtaining a particular sample outcome.

II. What is probability?
Classical interpretation (from games of chance) - numerical values of probabilities arise from the nature of the game.

- Outcome - each possible distinct result
- Event - collection of outcomes

\[ P(E) = \frac{N_e}{N} \]

Relative frequency interpretation (empirical approach) - if an experiment is conducted \( n \) times and an event \( E \) occurred \( n_e \) times then probability of the even \( E \) is approximately

\[ P(E) \approx \frac{n_e}{n} \]
III. After a large number of tries the relative frequency will be equal to the true probability.

We can check it using a large number of tries in a situation when the true probability is known:

Example 4.2 (pp.125-128) - tossing two coins: a penny and a dime.

There are four possible outcomes:
TT - tails for both coins;
TH - a tail for the penny, a head for the dime;
HT - a head for the penny, a tail for the dime;
HH - heads for both coins
What is the probability of having exactly one head?

Classical interpretation:
There are four possible outcomes \((N=4)\) there are two outcomes when we will observe one head \((N_e=2)\), hence

\[
P(\text{exactly one head}) = \frac{N_e}{N}=\frac{2}{4}=0.5
\]

Relative frequency interpretation:
Out of 500 computer simulated coin tosses:

<table>
<thead>
<tr>
<th>Event</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT</td>
<td>129</td>
<td>129/500=0.258</td>
</tr>
<tr>
<td>TH</td>
<td>117</td>
<td>117/500=0.234</td>
</tr>
<tr>
<td>HT</td>
<td>125</td>
<td>125/500=0.250</td>
</tr>
<tr>
<td>HH</td>
<td>129</td>
<td>129/500=0.258</td>
</tr>
</tbody>
</table>

\[
P(\text{exactly one head}) = \frac{n_e}{n}=(117+125)/500=0.484
\]

Very close to the theoretical probability of 0.5
Relative frequency interpretation provides a practical interpretation of the probability for most events of interest in biology.

IV. Some probability rules

1. $0 < P(E) < 1$

2. If events $A$ and $B$ are mutually exclusive (if $A$ occurs $B$ can not occur at the same time)

$$P(\text{either } A \text{ or } B \text{ occur}) = P(A) + P(B)$$
- (addition rule)

Example with flipping two coins - mutually exclusive events TT and TH

$P(\text{TT}) = 0.25$  $P(\text{TH}) = 0.25$

What is the probability of getting either two tails or a tail on the penny and a head on the dime?

$P(\text{TT} + \text{TH}) = P(\text{TT}) + P(\text{TH}) = 0.25 + 0.25 = 0.5$
3. If events A and B are independent, then the probability that in a series of independent trials both A and B will occur

\[ P(A \text{ and } B) = P(A) \times P(B) \]
- (multiplication rule)

Example with two coins - the coins are flipped several times (each flip is independent from another).
What is the probability that you get a TT and after that a TH?
\[ P(TT \text{ and } TH) = P(TT) \times P(TH) = 0.25 \times 0.25 = 0.0625 \]

Example with one coin - what is the probability that when I flip a coin 5 time, I’ll get 5 heads in a row?
Discrete variable - observations on a quantitative random variable can assume only a countable number of values. We will compute probability of specific individual values.

Continuous variable - observations on a quantitative random variable can assume any of the uncountable number of values on a line interval. We will compute probability of an interval of values.
V. Probability distributions for discrete random variables - display probability associated with each value of $y$.

Example with two coins: $y$ - is the number of heads, $y$ can assume three values: 0 (both tails), 1 (one head in any coin), 2 (both heads). Tossed them 500 times.

<table>
<thead>
<tr>
<th>Value of $y$</th>
<th>Frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>129</td>
<td>0.258</td>
</tr>
<tr>
<td>1</td>
<td>242</td>
<td>0.484</td>
</tr>
<tr>
<td>2</td>
<td>129</td>
<td>0.258</td>
</tr>
</tbody>
</table>

Probability histogram
Properties of discrete random variables:

1. Probability associated with every value of $y$ lies between 0 and 1.

2. The sum of the probabilities of all values of $y$ is equal to 1.

3. The probabilities for a discrete random variable are additive. Probability the $y = 1$ or $2$ equals to $P(1) + P(2)$.

We generate probability distribution using relative frequencies
VI. Binomial distribution

Data that can be viewed as sets of 0s and 1s:
  yes/no responses
  male/female
  infected/not infected
  success/failure
  head/tail in coin tossing

Properties of a binomial experiment:
1. The experiment consists of n identical trials.
2. Each trial results in one of the two possible outcomes.
3. The probability of success $\pi$ remains the same from trial to trial.
4. The trials are independent.
5. The random variable is the number of successes observed in n trials.
We will use coin tossing as an example:

Flip a coin $n$ times
[we get $n$ values of 1s (if heads) and 0s (if tails)]

The random variable of interest $y$ - number of times we got heads

Probability of success $\pi$ ($\pi$ of getting heads = 0.5)

Example 4.5 p.145 - check if the study satisfies the properties of a binomial experiment.
If the study is a binomial experiment, the probability distribution of the binomial variable \( y \) can be calculated by using formula for binomial probabilities.

\[
P(y) = \frac{n!}{y!(n-y)!} \pi^y (1-\pi)^{n-y}
\]

Where
\( n \) number of trials
\( \pi \) probability of success in a single trial
\( 1-\pi \) probability of failure in a single trial
\( y \) number of successes in \( n \) trials
\( n! = (n)(n-1)(n-2)...(3)(2)(1) \)

e.g.
3! = 3*2*1 = 6
4! = 4*3*2*1 = 24
0! = 1
Example with a coin - flip a coin 5 times \((n=5)\) and find what is the probability of getting heads 3 times \((v=3)\). For the coin \(\pi\) of success (getting heads) = \(\pi\) of failure (getting tails) = 1-\(\pi\) = 0.5

\[
P(3) = \frac{5!}{3!(5 - 3)!} (0.5)^3 (1 - 0.5)^{5-3} = \frac{120}{6*2} \times 0.125 \times 0.25 = 0.3125
\]

Example 4.7 p.147 - germination rate of a new turf grass variety.
Mean and standard deviation of the binomial probability distribution:

$$\mu = n \pi$$

$$\sigma = \sqrt{n \pi (1 - \pi)}$$

If we know \( \pi \) and the sample size, \( n \), we can calculate \( \mu \) and \( \sigma \) to locate the center and describe the variability for a particular binomial probability distribution.

Determine those values of \( y \) that are probable and those that are improbable.

Example - the turf grass seed germination:

$$\mu = n \pi = 20 \times (0.85) = 17$$

$$\sigma = \sqrt{n \pi (1 - \pi)} = \sqrt{20 \times (0.85) \times (1 - 0.85)} = 1.60$$
Probability distribution for the number of seeds that germinate in the sample of 20 seeds.

The binomial distribution for $n=20$ and $p=0.85$
VII. Probability distributions for continuous random variables

Continuous variable - observations on a quantitative random variable can assume any of the uncountable number of values on a line interval. We will compute probability of an interval of values.
\[ P(y=5) = 0.13 \]

\[ P(4.5 < y < 5.5) = \text{some number} \]
The total area under the probability curve is equal to one.

Example Fig. 4.6 p.155
The general shape of the probability distribution is important since it will affect the inferences about the population parameters.

Example Fig.4.8 p.156

Many of the variables in nature are normally distributed (bell-shaped, mound-shaped).
VIII. Normal (Gaussian) distribution. To make inference about the population based on a sample we need to know the probability of observing a particular outcome.

Frequency distributions for many variables in nature as well as several statistics can be approximated by a normal probability distribution.
Equation for $f(y)$ of the normal distribution:

$$f(y) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

$\mu$ – population mean

$\sigma^2$ – population variance
For a continuous random variable, the probability that a particular outcome falls in an interval from $a$ to $b$ is equal to the area under the probability distribution curve for that interval.
Area under the $f(y)$ curve can be found by integration:

$$\int_{a}^{b} f(y) \, dy = \int_{a}^{a} \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \, dy$$

Fortunately, we can use the tabulated values for the areas under a normal curve (Table 1 in Ott&Longnecker).

$z$-score associated with the value of $y$ is calculated as

$$z = \frac{y - \mu}{\sigma}$$

$$y = \mu + z\sigma$$
Table 1 (pp. 1091-1092)
Example: \( z = -2.25 \)

\[ Pr (Z \leq -2.25) = 0.0122 \]
Example: \( z = 1.66 \)

\[ Pr (Z \leq 1.66) = 0.9515 \]
Example 4.12

Normal distribution with $\mu = 20$ and $\sigma = 2$. What is the probability that the measurement will be less than 23?

\[ z = \frac{y - \mu}{\sigma} = \frac{23 - 20}{2} = 1.5 \]

\[ Pr (Z \leq 1.5) = 0.9332 \]
Example 4.13

What is the probability that $y$ will be less than 16 ($\mu=20$ and $\sigma=2$)?

$$z = \frac{y - \mu}{\sigma} = \frac{16 - 20}{2} = -2$$

$$Pr \ (Z \leq -2.0) = 0.0228$$
Probability that the measurement will be greater than ..?
Probability that the measurement will be between \( a \) and \( b \)?

**Example 4.14**
The mean daily milk production = 70 lb, with standard deviation = 13 lb (\( \mu=70 \) and \( \sigma=13 \)).

b. What is the probability that the milk production for a cow chosen at random will be greater than 90 lb?

\[
z = \frac{y - \mu}{\sigma} = \frac{90 - 70}{13} = 1.54
\]

\( Pr (Z \leq 1.54) = 0.9382 \) then

\( Pr (Z \geq 1.54) = 1 - Pr(Z \leq 1.54) = 1 - 0.9382 = 0.0618 \)
c. What is the probability that the milk production for a cow chosen at random will be between 60 and 90 lb?

\[ z_{60} = \frac{y - \mu}{\sigma} = \frac{60 - 70}{13} = -0.77 \]

\[ z_{90} = \frac{y - \mu}{\sigma} = \frac{60 - 70}{13} = 1.54 \]

\[ Pr (-0.77 \leq Z \leq 1.54) = Pr (Z \leq 1.54) - Pr (Z \leq -0.77) = 0.9382 - 0.2206 = 0.7176 \]
How to find percentiles of the normal distribution?

Percentile – the $p^{th}$ percentile is the value that has $p\%$ of measurements below it and $(100-p)\%$ measurements above it.

Use Table 1 in the reverse order:

First, we find the probability $p$ in Table 1.
Second, we find the $z_p$ value corresponding to the probability.
Third, we calculate the $p^{th}$ percentile value $y_p$ as $y_p = \mu + z_p \sigma$
Example: find $80^{th}$ percentile.

From Table 1: $z_{0.8} = 0.84$

$y_{0.80} = \mu + 0.84 \sigma$

Example 4.15

Based on the results of the Scholastic Assessment Test determine the lower 10% of all scores (that is - 10$^{th}$ percentile) ($\mu=500$ and $\sigma=100$).

From Table 1: $z_{0.1} = -1.28$

$y_{0.10} = \mu - 1.28 \sigma = \mu - 1.28 \times 100 = 500 - 1.28 \times 100 = 372$

10% of the SAT scores are less than 372.