Introduction to Statistics for the Life Sciences

Fall 2014
Chapter 5:
Example

Magnesium

- Mg - chemical symbol
- 12 protons
- 9th most abundant element in the known universe
- 11th most abundant element in the human body (manipulating ATP, DNA, RNA)
- Used in laxatives, antacids, for calming abnormal nerve excitation
- At the center of chlorophyll (metallic ion)
Isotopes are variations of an element with differing numbers of neutrons

- Elements are named by how many protons they have (12 for Mg)
- Mg usually has 12 neutrons, but can have 13 or 14 and maintain stability
- The weight of an element is the sum of the number of protons and neutrons
- Mg’s weight is usually 24, but sometimes 25 or 26.
Mg Atomic Weight Population Probabilities

Probability vs. Atomic Weight

- Atomic Weight: 24, Probability: 0.8
- Atomic Weight: 25, Probability: 0.2
- Atomic Weight: 26, Probability: 0.0
Example

- Periodic table gives 24.32 as weight of Mg
- On Earth, 78.99% of Mg weighs 24
- On Earth, 10.00% of Mg weighs 25
- On Earth, 11.01% of Mg weighs 26

\[(24)(.7899) + (25)(.10) + (26)(.1101) = 24.32\]

Thinking in terms of this class, what does this calculation look like?
Example

- Periodic table gives 24.32 as weight of Mg
- On Earth, 78.99% of Mg weighs 24
- On Earth, 10.00% of Mg weighs 25
- On Earth, 11.01% of Mg weighs 26

\[(24)(.7899) + (25)(.10) + (26)(.1101) = 24.32\]

Thinking in terms of this class, what does this calculation look like?
Example

Let $Y$ be a random variable characterized by the following table:

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$P(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>.7899</td>
</tr>
<tr>
<td>25</td>
<td>.10</td>
</tr>
<tr>
<td>26</td>
<td>.1101</td>
</tr>
</tbody>
</table>

Calculate the mean, $\mu$, of $Y$:

$$
\mu = \sum_y yP(Y = y) \\
= (24)(.7899) + (25)(.10) + (26)(.1101) \\
= 24.32
$$
Example

- Let $Y$ be the weight of a given Mg element drawn at random from the Mg currently on Earth.
- $\mu = 24.32$ is the element weight of Mg given on the periodic table.
- The atomic weight given on the periodic table is an average, in particular, the mean $\mu$.

Aside:
- Can you think of a better “measure of center” to put on the periodic table of elements?
Example

- Let $Y$ be the weight of a given Mg element drawn at random from the Mg currently on Earth.
- $\mu = 24.32$ is the element weight of Mg given on the periodic table.
- The atomic weight given on the periodic table is an average, in particular, the mean $\mu$.

Aside:
- Can you think of a better “measure of center” to put on the periodic table of elements?
Could we calculate the std. dev. $\sigma$ of $Y$, the atomic weight of Mg? Solution: Use

\[
\sigma^2 = \left[ \sum_y y^2 P(Y = y) \right] - \mu^2.
\]

Then

\[
\sum_y y^2 P(Y = y) = (24)^2(.7899) + (25)^2(.10) + (26)^2(.1101) = 591.91
\]

So that $\sigma^2 = 591.91 - 24.32^2 = 0.4476$, giving

\[
\sigma = \sqrt{0.4476} = .67
\]

Is this reasonable?
Could we calculate the std. dev. $\sigma$ of $Y$, the atomic weight of Mg? Solution: Use

$$\sigma^2 = \left[ \sum_y y^2 P(Y = y) \right] - \mu^2.$$ 

Then

$$\sum_y y^2 P(Y = y) = (24)^2(0.7899) + (25)^2(0.10) + (26)^2(0.1101) = 591.91$$

So that $\sigma^2 = 591.91 - 24.32^2 = 0.4476$, giving

$$\sigma = \sqrt{0.4476} = .67$$

Is this reasonable?
Example

Could we calculate the std. dev. $\sigma$ of $Y$, the atomic weight of Mg? Solution: Use

$$\sigma^2 = \left[ \sum_y y^2 P(Y = y) \right] - \mu^2.$$

Then

$$\sum_y y^2 P(Y = y) = (24)^2(.7899)+(25)^2(.10)+(26)^2(.1101) = 591.91$$

So that $\sigma^2 = 591.91 - 24.32^2 = 0.4476$, giving

$$\sigma = \sqrt{0.4476} = .67$$

Is this reasonable?
Example

Suppose the scientist wishes to randomly sample 2 Mg molecules. Enumerate the outcomes and their probabilities:

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24, 24</td>
<td>(.7899)(.7899) = .62</td>
</tr>
<tr>
<td>24, 25</td>
<td>(.7899)(.10) = .078</td>
</tr>
<tr>
<td>24, 26</td>
<td>(.7899)(.1101) = .09</td>
</tr>
<tr>
<td>25, 24</td>
<td>(.10)(.7899) = .078</td>
</tr>
<tr>
<td>25, 25</td>
<td>(.10)(.10) = .01</td>
</tr>
<tr>
<td>25, 26</td>
<td>(.10)(.1101) = .01</td>
</tr>
<tr>
<td>26, 24</td>
<td>(.1101)(.7899) = .09</td>
</tr>
<tr>
<td>26, 25</td>
<td>(.1101)(.10) = .01</td>
</tr>
<tr>
<td>26, 26</td>
<td>(.1101)(.1101) = .01</td>
</tr>
</tbody>
</table>
Consider the following alternative table:

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Mean Outcome</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24, 24</td>
<td>24</td>
<td>.62</td>
</tr>
<tr>
<td>24, 25</td>
<td>24.5</td>
<td>.078</td>
</tr>
<tr>
<td>24, 26</td>
<td>25</td>
<td>.09</td>
</tr>
<tr>
<td>25, 24</td>
<td>24.5</td>
<td>.078</td>
</tr>
<tr>
<td>25, 25</td>
<td>25</td>
<td>.01</td>
</tr>
<tr>
<td>25, 26</td>
<td>25.5</td>
<td>.01</td>
</tr>
<tr>
<td>26, 24</td>
<td>25</td>
<td>.09</td>
</tr>
<tr>
<td>26, 25</td>
<td>25.5</td>
<td>.01</td>
</tr>
<tr>
<td>26, 26</td>
<td>26</td>
<td>.01</td>
</tr>
</tbody>
</table>

Can we simplify this table? Why?
Consider the following alternative table:

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Mean Outcome</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24, 24</td>
<td>24</td>
<td>.62</td>
</tr>
<tr>
<td>24, 25</td>
<td>24.5</td>
<td>.078</td>
</tr>
<tr>
<td>24, 26</td>
<td>25</td>
<td>.09</td>
</tr>
<tr>
<td>25, 24</td>
<td>24.5</td>
<td>.078</td>
</tr>
<tr>
<td>25, 25</td>
<td>25</td>
<td>.01</td>
</tr>
<tr>
<td>25, 26</td>
<td>25.5</td>
<td>.01</td>
</tr>
<tr>
<td>26, 24</td>
<td>25</td>
<td>.09</td>
</tr>
<tr>
<td>26, 25</td>
<td>25.5</td>
<td>.01</td>
</tr>
<tr>
<td>26, 26</td>
<td>26</td>
<td>.01</td>
</tr>
</tbody>
</table>

Can we simplify this table? Why?
Let $Y_1$ and $Y_2$ be two random variables with the same distribution as $Y$. Consider

$$\bar{Y} = \frac{1}{n}(Y_1 + Y_2).$$

The table above is the distribution of $\bar{Y}$, called the **Sampling Distribution**.
Example

Simplified Table

<table>
<thead>
<tr>
<th>Mean Outcome</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>.62</td>
</tr>
<tr>
<td>24.5</td>
<td>.156</td>
</tr>
<tr>
<td>25</td>
<td>.19</td>
</tr>
<tr>
<td>25.5</td>
<td>.02</td>
</tr>
<tr>
<td>26</td>
<td>.01</td>
</tr>
</tbody>
</table>

Let $Y_1$ and $Y_2$ be two random variables with the same distribution as $Y$. Consider

$$\bar{Y} = \frac{1}{n}(Y_1 + Y_2).$$

The table above is the distribution of $\bar{Y}$, called the **Sampling Distribution**.
Think of \( Y_1, \ldots, Y_n \) as \textbf{independent copies} of some random variable \( Y \).

**Sampling Distribution**

Let \( Y_1, \ldots, Y_n \) be the random variables that correspond to the values of a sample of size \( n \). A table that gives the possible values of

\[
\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i
\]

is called the \textbf{sampling distribution} of \( \bar{Y} \).
\( \bar{y} \) versus \( \bar{Y} \)

- \( \bar{y} \) is the sample mean after the sample has been observed.
- \( \bar{Y} \) is the sample mean before the sample has been observed.
- When an experiment is conducted, the random variable \( \bar{Y} \) takes a value \( \bar{y} \) (out of several possible values for \( \bar{y} \)).
Expected value of \( \bar{Y} \)

Consider the \( n \) random variables \( Y_1, \ldots, Y_n \), which are independent copies of \( Y \). Denote the mean of \( Y \) by \( \mu_Y \). Then

\[
\mu_{\bar{Y}} = \mu_Y.
\]

In other words, the mean of \( \bar{Y} \) is the same as the mean of \( Y \).
We have \( n = 2 \) and the sampling distribution for \( \bar{Y} \):

<table>
<thead>
<tr>
<th>( Y )</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>.62</td>
</tr>
<tr>
<td>24.5</td>
<td>.156</td>
</tr>
<tr>
<td>25</td>
<td>.19</td>
</tr>
<tr>
<td>25.5</td>
<td>.02</td>
</tr>
<tr>
<td>26</td>
<td>.01</td>
</tr>
</tbody>
</table>

The mean \( \mu_{\bar{Y}} \) can be calculated by multiplying across the rows and then adding. Or

\[
\mu_{\bar{Y}} = (24)(.62) + (24.5)(.156) + (25)(.19) + (25.5)(.02) + (26)(.01)
\]

\[= 24.222\]

WAIT! This is not equal to \( \mu_Y = 24.32 \). What happened?
We have \( n = 2 \) and the sampling distribution for \( \bar{Y} \):

<table>
<thead>
<tr>
<th>( Y )</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>.62</td>
</tr>
<tr>
<td>24.5</td>
<td>.156</td>
</tr>
<tr>
<td>25</td>
<td>.19</td>
</tr>
<tr>
<td>25.5</td>
<td>.02</td>
</tr>
<tr>
<td>26</td>
<td>.01</td>
</tr>
</tbody>
</table>

The mean \( \mu_{\bar{Y}} \) can be calculated by multiplying across the rows and then adding. Or

\[
\mu_{\bar{Y}} = (24)(.62) + (24.5)(.156) + (25)(.19) + (25.5)(.02) + (26)(.01)
\]

\[= 24.222\]

WAIT! This is not equal to \( \mu_{\bar{Y}} = 24.32 \). What happened?
Calculate $\sigma_{\bar{Y}}$ in the Mg Example

Calculate $\sigma_{\bar{Y}}$ in the Mg example:

\[
\sum_y y^2 P(\bar{Y} = y) = (24)^2(.62) + (24.5)^2(.156) + (25)^2(.19) + (25.5)^2(.02) + (26)^2(.01)
\]

\[= 591.6911 \]

\[\sigma_{\bar{Y}}^2 = 591.6911 - (24.3202)^2 = .2189.\]

\[\sigma_{\bar{Y}} = \sqrt{.2189} = 0.4679.\]

Recall $\sigma_Y$ from the beginning of lecture! $\sigma_Y = .67.$
Calculate $\sigma_{\bar{Y}}$ in the Mg Example

Calculate $\sigma_{\bar{Y}}$ in the Mg example:

\[
\sum_y y^2 P(\bar{Y} = y) = (24)^2(.62) + (24.5)^2(.156) \\
+ (25)^2(.19) + (25.5)^2(.02) \\
+ (26)^2(.01) \\
= 591.6911
\]

\[\sigma_{\bar{Y}}^2 = 591.6911 - (24.3202)^2 = .2189.\]

\[\sigma_{\bar{Y}} = \sqrt{.2189} = 0.4679.\]

Recall $\sigma_Y$ from the beginning of lecture! $\sigma_Y = .67$. 
Calculate $\sigma_{\bar{Y}}$ in the Mg Example:

\[
\sum_y y^2 P(\bar{Y} = y) = (24)^2(.62) + (24.5)^2(.156) + (25)^2(.19) + (25.5)^2(.02) + (26)^2(.01)
\]
\[
= 591.6911
\]

\[
\sigma_{\bar{Y}}^2 = 591.6911 - (24.3202)^2 = .2189.
\]

\[
\sigma_{\bar{Y}} = \sqrt{.2189} = 0.4679.
\]

Recall $\sigma_Y$ from the beginning of lecture! $\sigma_Y = .67.$
Calculate $\sigma_Y$ in the Mg example:

\[
\sum_{y} y^2 P(\bar{Y} = y) = (24)^2(0.62) + (24.5)^2(0.156) + (25)^2(0.19) + (25.5)^2(0.02) + (26)^2(0.01)
\]

\[
= 591.6911
\]

\[
\sigma_Y^2 = 591.6911 - (24.3202)^2 = 0.2189.
\]

\[
\sigma_Y = \sqrt{0.2189} = 0.4679.
\]

Recall $\sigma_Y$ from the beginning of lecture! $\sigma_Y = 0.67$. 
Consider the $n$ random variables $Y_1, \ldots, Y_n$, which are independent copies of $Y$. Denote the std. dev. of $Y$ by $\sigma_Y$. Then

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}.$$ 

Also

$$\sigma^2_{\bar{Y}} = \frac{\sigma^2_Y}{n}.$$
Check $\sigma_{\bar{Y}} = \sigma_Y / \sqrt{n}$ in Mg Example

From two slides ago: $\sigma_{\bar{Y}} = 0.4679$ and $\sigma_Y = .67$. Compute:

$$\frac{\sigma_Y}{\sqrt{n}} = \frac{.67}{\sqrt{2}} = \frac{.67}{1.414214} = 0.4737615 \approx \sigma_{\bar{Y}}$$

Close enough.
Questions about $n$

- What if we repeated everything for $n = 3$ Mg molecules?
- What if we repeated everything for $n = 10$ Mg molecules? Wouldn’t the sampling distribution get too complicated to calculate?
- How would the sampling distribution of $\bar{Y}$ change as $n$ gets large?
- What happens to $\sigma_{\bar{Y}}$ when $n$, the sample size, gets large?
What happens to $\sigma \bar{Y}$ when $n$, the sample size, gets large?

$$\sigma \bar{Y} = \frac{\sigma_Y}{\sqrt{n}}$$

If $n$ gets extremely big, what happens to $\sigma \bar{Y}$?

Answer: It gets extremely small. Therefore, the values of $\bar{Y}$ will...?

- Recall that $\sigma$ is the “true average distance of observations from $\mu$”
- Values of $\bar{Y}$ will start to collect tighter around $\mu$. 
What happens to $\sigma \bar{Y}$ when $n$, the sample size, gets large?

$$\sigma \bar{Y} = \frac{\sigma Y}{\sqrt{n}}$$

If $n$ gets extremely big, what happens to $\sigma \bar{Y}$?

Answer: It gets extremely small. Therefore, the values of $\bar{Y}$ will...

- Recall that $\sigma$ is the “true average distance of observations from $\mu$”
- Values of $\bar{Y}$ will start to collect tighter around $\mu$. 
What happens to $\sigma \bar{Y}$ when $n$, the sample size, gets large?

$$\sigma \bar{Y} = \frac{\sigma Y}{\sqrt{n}}$$

If $n$ gets extremely big, what happens to $\sigma \bar{Y}$?

Answer: It gets extremely small. Therefore, the values of $\bar{Y}$ will...?

- Recall that $\sigma$ is the “true average distance of observations from $\mu$”
- Values of $\bar{Y}$ will start to collect tighter around $\mu$. 
Let’s view the sampling distribution of $\bar{Y}$ in the Mg example as $n$ gets larger.

- We can’t do this by hand – but a computer can do it.
Mg Atomic Weight Population Probabilities

Probability

Atomic Weight

0.0 0.2 0.4 0.6 0.8

24 25 26
Mg Atomic Weight Population Probabilities

<table>
<thead>
<tr>
<th>Atomic Weight</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>0.6</td>
</tr>
<tr>
<td>25</td>
<td>0.1</td>
</tr>
<tr>
<td>24.5</td>
<td>0.2</td>
</tr>
<tr>
<td>25.5</td>
<td>0.1</td>
</tr>
<tr>
<td>26</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Sampling Distribution of Y-bar with n = 10

Atomic Weight

<table>
<thead>
<tr>
<th>Probability</th>
<th>Atomic Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>24.0</td>
</tr>
<tr>
<td>0.5</td>
<td>24.2</td>
</tr>
<tr>
<td>1.0</td>
<td>24.4</td>
</tr>
<tr>
<td>1.5</td>
<td>24.6</td>
</tr>
<tr>
<td>2.0</td>
<td>24.8</td>
</tr>
<tr>
<td></td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td>25.2</td>
</tr>
</tbody>
</table>
Sampling Distribution of Y-bar with n = 15

Atomic Weight

Probability

24.0 24.2 24.4 24.6 24.8 25.0 25.2

0.0 0.5 1.0 1.5 2.0 2.5

Atomic Weight
Sampling Distribution of Y-bar with n = 20

<table>
<thead>
<tr>
<th>Atomic Weight</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.0</td>
<td>0.0</td>
</tr>
<tr>
<td>24.2</td>
<td>0.0</td>
</tr>
<tr>
<td>24.4</td>
<td>0.5</td>
</tr>
<tr>
<td>24.6</td>
<td>1.0</td>
</tr>
<tr>
<td>24.8</td>
<td>1.5</td>
</tr>
<tr>
<td>25.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

![Graph showing Sampling Distribution of Y-bar with n = 20](image-url)
Sampling Distribution of $Y\bar{}$ with $n = 100$

<table>
<thead>
<tr>
<th>Atomic Weight</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.1</td>
<td>0</td>
</tr>
<tr>
<td>24.2</td>
<td>1</td>
</tr>
<tr>
<td>24.3</td>
<td>3</td>
</tr>
<tr>
<td>24.4</td>
<td>4</td>
</tr>
<tr>
<td>24.5</td>
<td>5</td>
</tr>
<tr>
<td>24.6</td>
<td>0</td>
</tr>
</tbody>
</table>
Probability plot with n = 5
Probability plot with \( n = 10 \)
Probability plot with n = 15

Value of Y-bar

Z quantile
Probability plot with n = 100

Value of Y-bar

Z quantile
Reflection

How does the sampling distribution of $\bar{Y}$ change as $n$ grows large?

It becomes more normal. This is called the Central Limit Theorem.
How does the sampling distribution of $\bar{Y}$ change as $n$ grows large?

It becomes more normal. This is called the **Central Limit Theorem**.
Central Limit Theorem

Suppose $Y_1, \ldots, Y_n$ are independent copies of a random variable $Y$ with mean $\mu_Y$ and standard deviation $\sigma_Y$.

- If $Y$ is already $N(\mu_Y, \sigma_Y)$,
  \[ \bar{Y} \text{ follows } N(\mu_Y, \sigma_Y / \sqrt{n}) \]
  for any value of $n$

- If $Y$ is any* other population, then as $n$ gets large
  \[ \bar{Y} \text{ approximately follows } N(\mu_Y, \sigma_Y / \sqrt{n}) \]
Is the distribution of $Y$ in the Mg example normal?
Can we apply the CLT?
How large does $n$ have to be in order for the CLT to kick in?