5.5.3 Means and Variances for Linear Combinations

Let $U = a_0 + a_1X + a_2Y + a_3Z + \cdots$

Means of Linear Combinations

$E(U) = a_0 + a_1E(X) + a_2E(Y) + a_3E(Z) + \cdots$

Example

$X, Y, Z =$ values of three dice

You win $U = 5 + 2X + 3Y + 4Z$ dollars

$E(X) = E(Y) = E(Z) = 3.5$

Your total expected winnings are

$E(U) = 5 + 2(3.5) + 3(3.5) + 4(3.5) = 36.50$

Expected value(constant) = constant

$E(5) = 5$

$E(\Sigma) = \Sigma$ (expected values)


$E(constant \cdot X) = constant \cdot EX$

$E(3X) = 3E(X) \quad E(4Y) = 4E(Y) \quad E(5Z) = 5E(Z)$

$E(a_0 + a_1X + a_2Y + a_3Z) = E(a_0) + E(a_1X) + a_2E(Y) + E(a_3Z)$

$= a_0 + a_1E(X) + a_2E(Y) + a_3E(Z)$

Variances of Linear Combinations

$Var(a_0 + a_1X + a_2Y + a_3Z) = a_1^2 Var(X) + a_2^2 Var(Y) + a_3^2 Var(Z)$

$Var(a_0) = 0$

$Var(constant) = 0$

$Var(a_0 + X) = Var(X)$

Adding a constant doesn’t change variance
\[ \text{Var}(a_1X) = a_1^2 \text{Var}(X) \]

Variances are in \textit{units}^2.

\[ \text{St.Dev}(a_1X) = |a_1| \text{St.Dev}(X) \]

If \(X, Y, Z\) are independent or uncorrelated,
\[ \text{Var}(X + Y + Z) = \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z) \]

Putting these facts together

\[ \begin{align*}
\text{Var}(a_0 + a_1X + a_2Y + a_3Z) \\
= (\text{if independent}) \text{Var}(a_0) + \text{Var}(a_1X) + \text{Var}(a_2Y) + \text{Var}(a_3Z) \\
= a_1^2 \text{Var}(X) + a_2^2 \text{Var}(Y) + a_3^2 \text{Var}(Z)
\end{align*} \]

For independent \(X, Y\)

\[ \sigma_{X+Y} = \sqrt{\text{Var}(X) + \text{Var}(Y)} \]

We can think of independent random variables as orthogonal vectors with the squared lengths being the variances.

\[ \begin{align*}
\sigma_{X+Y}^2 &= \sigma_X^2 + \sigma_Y^2 \\
\sigma_{X+Y} &= \sqrt{\sigma_X^2 + \sigma_Y^2}
\end{align*} \]

Like Pythagorean Theorem
\[ c^2 = a^2 + b^2, \quad c = \sqrt{a^2 + b^2} \]
Most useful facts:

If $X_1, X_2, X_3, \ldots, X_n$ are independent measurement each with 
$E(X_i) = \mu$ and $Var(x_i) = \sigma^2$ then

$$E(X) = \mu \quad Var(X) = \frac{\sigma^2}{n} \quad St.Dev(X) = \frac{\sigma}{\sqrt{n}}$$

To see this:

$$E\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n} E(\sum X_i) = \frac{1}{n} \sum E(x_i) = \frac{1}{n} \sum \mu = \frac{1}{n} (n\mu) = \mu$$

The sample mean $\bar{X}$ is an unbiased estimator of the population mean $\mu$.

$$Var\left(\frac{1}{n} \sum X_i\right) = (\frac{1}{n})^2 Var(\sum X_i) = \frac{1}{n^2} \sum Var(x_i) = \frac{1}{n^2} \sum i \sigma^2 = \frac{1}{n^2} (n\sigma^2) = \frac{\sigma^2}{n}$$
5.5.4 Propagation of Error

Often done in chemistry or physics labs.

\[ E(a + bX) = a + bE(X) \]

For \( g(X) = a + bX \), \( E(g(X)) = g(E(X)) \)

But \( E(\log X) \neq \log(EX) \)

For \( g(X) = \log(X) \), \( E(g(X)) \neq g(E(X)) \)

<table>
<thead>
<tr>
<th>( X )</th>
<th>( P(X) )</th>
<th>( \log_{10} X )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \frac{1}{3} )</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>( \frac{1}{3} )</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>( \frac{1}{3} )</td>
<td>2</td>
</tr>
</tbody>
</table>

\[ E(X) = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 10 + \frac{1}{3} \cdot 100 = 37 \]

\[ \log(EX) = 1.58 \]

\[ E(\log X) = \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 2 = 1 \neq \log(E(X)) \]

For a nonlinear function except for very special cases \( E(g(X)) \neq g(E(x)) \)

\[ E(\log X) \neq \log(EX) \quad E(\frac{1}{EX}) \neq \frac{1}{E(X)} \quad E(X^2) \neq (E(X))^2 \]

If \( X \) has small variance, so that we are in a very local part of \( g(X) \), then approximately \( E(g(X)) \approx g(E(X)) \).

From a Taylor series approximation, \( g(X) \approx g(\mu) + (X - \mu)g'(\mu) \)

\[
E(g(x)) \approx E(g(\mu)) + E(g'(\mu)(x - \mu)) \\
= g(\mu) + g'(\mu)E(X - \mu) \\
= g(\mu) + g'(\mu)(EX - E\mu) \\
= g(\mu) + g'(\mu)(\mu - \mu) \\
= g(\mu)
\]

\[
Var(g(X)) \approx Var(g(\mu) + g'(\mu)(X - \mu)) \\
= 0 + (g'(\mu))^2 Var(X - \mu) \\
= (g'(\mu))^2 Var(X)
\]
Example:

\[
\text{Var}(\ln X) \approx \left( \frac{1}{\mu} \right)^2 \text{Var}(X) \\
\approx \frac{\sigma^2}{\mu^2} \\
= (\text{CV})^2
\]

\(\text{CV} = \text{Coefficient of variation.}\)

If \(X, Y, Z\) are independent,
\[
\text{Var}(g(X,Y,Z)) \approx \left( \frac{\partial g}{\partial X} \right)^2 \text{Var}(X) + \left( \frac{\partial g}{\partial Y} \right)^2 \text{var}(Y) + \left( \frac{\partial g}{\partial Z} \right)^2 \text{Var}(Z)
\]
with partial derivatives equal at \((X = \mu_X, Y = \mu_Y, Z = \mu_Z)\).

Example 24 (p. 311) Independent resistors

\[
R = R_1 + \frac{R_2 R_3}{R_2 + R_3} \\
= R_1 + \left( \frac{1}{R_2} + \frac{1}{R_3} \right)^{-1}
\]

This is a harmonic mean back-transformed reciprocals.

\[
E(R_i) = 100\Omega \\
\text{St.Dev}(R_i) = 2\Omega \\
\text{Var}(R_i) = 4\Omega^2 \\
E(R) \approx 100 + \left( \frac{1}{100} + \frac{1}{100} \right)^{-1} \\
= 100 + \left( \frac{2}{100} \right)^{-1} \\
= 100 + 50 \\
= 150
\]
\[ \text{Var}(R_1 + \left(\frac{1}{R_2} + \frac{1}{R_3}\right)^{-1}) = (\text{since ind't}) \text{Var}(R_2) + \text{Var}(\left(\frac{1}{R_2} + \frac{1}{R_3}\right)^{-1}) \]

\[
\frac{\partial}{\partial R_2} \left( \frac{R_2 R_3}{R_2 + R_3} \right) = \frac{R_3^2}{(R_2 + R_3)^2}
\]

\[
\frac{\partial}{\partial R_2} \left( \frac{R_2 r_3}{R_2 + R_3} \right) = \frac{r_3^2}{(R_2 + R_3)^2}
\]

\[
\text{Var} \left( \frac{R_2 R_3}{R_2 + R_3} \right) \approx \left( \frac{100^2}{(100 + 100)^2} \right)^2 \text{Var}(R_2) + \left( \frac{100^2}{(100 + 100)^2} \right)^2 \text{Var}(R_3)
\]

\[
= \left( \frac{100^4}{16 \times 100^4} \right) \cdot 4 \times 2
\]

\[
= \frac{1}{2}
\]

\[
\text{Var}(R) = \text{Var}(R_1) + \text{Var} \left( \frac{R_2 r_3}{R_2 + R_3} \right)
\]

\[
\approx 4 + 0.5
\]

\[
= 4.5
\]

\[
\text{St.Dev}(R) \approx \sqrt{4.5}
\]

\[
= 2.12 \Omega
\]
5.5.5 The Central Limit Effect

If $X_1, X_2, \ldots, X_n$ are independent random variables with mean $\mu$ and variance $\sigma^2$, then for large $n$, the variable $\bar{X}$ is approximately normally distributed.

![Figure 4.30](image)

(a) Histogram of ages for 500 pennies

(b) Sampling distribution of $\bar{y}$ for $n = 5$

(c) Sampling distribution of $\bar{y}$ for $n = 10$

(d) Sampling distribution of $\bar{y}$ for $n = 25$
Even if the underlying distribution is not normal, probabilities involving means can be approximated with the normal table.

** HOWEVER, ** if transformation, e.g. \( \ln(X) \), makes the distribution more normal or if another distribution, e.g. Weibull, fits better, then relying on the Central Limit Theorem a normal approximations is less precise. We will do a better job if we use the more precise method.

If \( X_1, X_2, \ldots, X_n \) are normal, then \( \bar{X} \) is exactly normal.

Example: Bags of potatoes weigh

\( \mu = 5.0 \text{ pounds} \)
\( \sigma = 0.1 \text{ pounds} \)

If we buy 4 bags, what is \( P(\bar{X} < 4.9) \)?

\[
E(\bar{X}) = \mu = 5 \quad \sqrt{\text{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}} = \frac{0.1}{\sqrt{4}} = 0.05
\]

\[
P(\bar{X} < 4.9) = P\left(\frac{\bar{X} - \mu}{\sqrt{\text{Var}(\bar{X})}} < \frac{4.9 - 5.0}{0.05}\right) = P(z < 2) = 0.0228
\]