STAT 234 Lecture 11
Covariance and Correlation
Section 5.2

Yibi Huang
Department of Statistics
University of Chicago

Covariance

Covariance

The **covariance** of X and Y, denoted as $\mathrm{Cov}(X, Y)$ or σ_{XY}, is defined as

$$\mathrm{Cov}(X, Y) = \sigma_{XY} = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)],$$

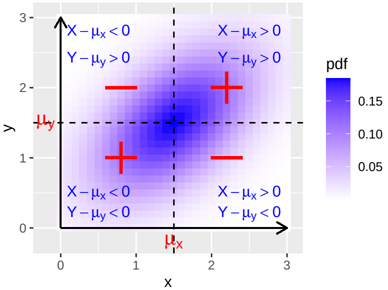
in which $\mu_X = \mathrm{E}(X)$, $\mu_Y = \mathrm{E}(Y)$

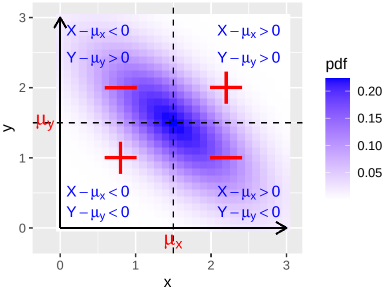
- Covariance is a generalization of variance as the variance of a random variable X is just the covariance of X with itself.

$$\mathrm{Var}(X) = \mathrm{Cov}(X, X) = \mathrm{E}[(X - \mu_X)^2]$$

Sign of Covariance Reflects the Direction of (X, Y) Relation

- $\mathrm{Cov}(X, Y) > 0$ means a *positive* relation between X, Y
  - When X increases, Y tends to increase
- $\mathrm{Cov}(X, Y) < 0$ means a *negative* relation between X, Y
  - When X increases, Y tends to decrease





Shortcut Formula for Covariance

$$\boxed{\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)}$$

- Like the Shortcut Formula for Variance $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$.
- If X & Y are indep., then $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$, which implies $\mathrm{Cov}(X, Y) = 0$.
- However $\mathrm{Cov}(X, Y) = 0$ does **not** imply the independence of X and Y. In this case, we say X and Y are **uncorrelated.**

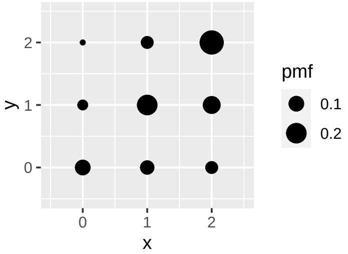
Proof of the Shortcut Formula for Covariance

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] \\
&= \mathrm{E}(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\
&= \mathrm{E}(XY) - \mu_X \underbrace{\mathrm{E}(Y)}_{=\mu_Y} - \mu_Y \underbrace{\mathrm{E}(X)}_{=\mu_X} + \mu_X \mu_Y \\
&= \mathrm{E}(XY) - \mu_X \mu_Y
\end{aligned}$$

Example (Gas Station) — $\mathrm{E}(XY)$

For the Gas Station Example in L09, recall the joint pmf is

	$p(x,y)$	Y (full-service) 0	1	2
X	0	0.10	0.04	0.02
self-	1	0.08	0.20	0.06
service	2	0.06	0.14	0.30

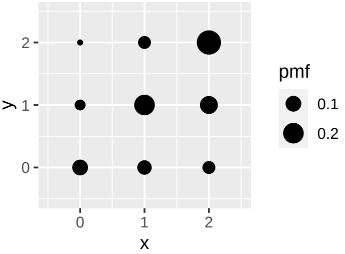


Guess $\mathrm{Cov}(X,Y) > 0$ or < 0?

Example (Gas Station) — $\mathrm{E}(XY)$

For the Gas Station Example in L09, recall the joint pmf is

	$p(x,y)$	Y (full-service) 0	1	2
X	0	0.10	0.04	0.02
self-	1	0.08	0.20	0.06
service	2	0.06	0.14	0.30



Guess $\mathrm{Cov}(X,Y) > 0$ or < 0?

$$\begin{aligned}
\mathrm{E}(XY) &= \sum_{xy} xyp(x,y) \\
&= 0\cdot 0\cdot 0.10 + 0\cdot 1\cdot 0.04 + 0\cdot 2\cdot 0.02 \\
&\quad + 1\cdot 0\cdot 0.08 + 1\cdot 1\cdot 0.20 + 1\cdot 2\cdot 0.06 \\
&\quad + 2\cdot 0\cdot 0.06 + 2\cdot 1\cdot 0.14 + 2\cdot 2\cdot 0.30 \\
&= 1.8
\end{aligned}$$

Example (Gas Station) — Covariance

Recall in L09, we obtained the marginal pmfs for X and Y:

x	0	1	2
$p_X(x)$	0.16	0.34	0.50

, $\mathrm{E}(X) = 0 \cdot 0.16 + 1 \cdot 0.34 + 2 \cdot 0.5 = 1.34$

y	0	1	2
$p_Y(y)$	0.24	0.38	0.38

, $\mathrm{E}(Y) = 0 \cdot 0.24 + 1 \cdot 0.38 + 2 \cdot 0.38 = 1.14$

By the shortcut formula, the covariance is

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) \\ &= 1.8 - 1.34 \times 1.14 = 0.2724 > 0. \end{aligned}$$

When one service island has more hoses in-use, the other also tends to have more in-use.

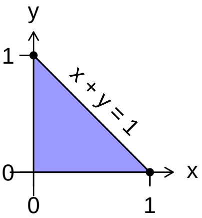
Example 5.5 (Mixed Nuts) — Covariance

Recall in Lecture 9, the joint pdf for

$X =$ the weight of almonds, and $Y =$ the weight of cashews

in a can of mixed nuts is

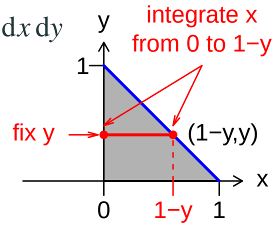
$$f(x,y) = \begin{cases} 24xy & \text{if } 0 \le x, y \le 1, x + y < 1 \\ 0 & \text{otherwise} \end{cases}$$



Before we calculate it, guess $\mathrm{Cov}(X,Y) > 0$ or < 0?

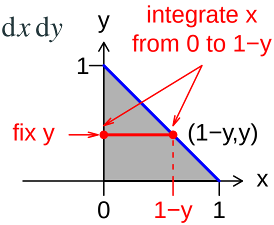
Example 5.5 (Mixed Nuts) — $\mathrm{E}(XY)$

$$\mathrm{E}(XY) = \iint xyf(x,y)\mathrm{d}x\mathrm{d}y = \int_0^1\int_0^{1-y} 24x^2y^2 \,\mathrm{d}x\,\mathrm{d}y$$



Example 5.5 (Mixed Nuts) — $\mathrm{E}(XY)$

$$\mathrm{E}(XY) = \iint xyf(x,y)\mathrm{d}x\mathrm{d}y = \int_0^1 \int_0^{1-y} 24x^2y^2 \,\mathrm{d}x\,\mathrm{d}y$$



where

$$\int_0^{1-y} 24x^2y^2 \,\mathrm{d}x = 8x^3y\Big|_{x=0}^{x=1-y} = 8(1-y)^3y^2$$

Putting it back to the double integral,

$$\mathrm{E}(XY) = \int_0^1 \int_0^{1-y} 24x^2y^2 \,\mathrm{d}x\,\mathrm{d}y = \int_0^1 8(1-y)^3y^2\mathrm{d}y = \frac{2}{15}.$$

Example 5.5 (Mixed Nuts) — Covariance

Recall in L09, we calculated the marginal pdf's for X and for Y:

$$f_X(x) = 12x(1-x)^2, \quad f_Y(y) = 12y(1-y)^2, \text{ for } 0 \le x, y \le 1.$$

using which we can calculate

$$\begin{aligned} \mathrm{E}(X) &= \int_0^1 x f_X(x) \mathrm{d}x = \int_0^1 12x^2(1-x)^2 \mathrm{d}x \\ &= \int_0^1 12x^2 - 24x^3 + 12x^4 \mathrm{d}x = 4x^3 - 6x^4 + \frac{12}{5}x^5 \Big|_0^1 = \frac{2}{5} \end{aligned}$$

Likewise, $\mathrm{E}(Y) = 2/5$. By the shortcut formula, the covariance is

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = \frac{2}{15} - \frac{2}{5} \times \frac{2}{5} = -\frac{2}{75}$$

When the amount of almond is increased, the amount of cashew is likely reduced.

$\mathrm{Cov}(aX + c, bY + d) = ab\ \mathrm{Cov}(X, Y)$

Recall variance has the scaling property

$$\mathrm{Var}(aX + b) = a^2\ \mathrm{Var}(X).$$

The scaling property for Covariance is

$$\begin{aligned}
&\mathrm{Cov}(aX + c, bY + d) \\
&= \mathrm{E}\{[aX + c - \mathrm{E}(aX + c)][bY + d - \mathrm{E}(bY + d)]\} \\
&= \mathrm{E}[(aX + c - a\,\mathrm{E}(X) - c)(bY + d - b\,\mathrm{E}(Y) - d)] \\
&= \mathrm{E}[(aX - a\,\mathrm{E}(X))(bY - b\,\mathrm{E}(Y))] \\
&= ab\ \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] \\
&= ab\ \mathrm{Cov}(X, Y)
\end{aligned}$$

Correlation

How Large the Covariance Indicates a Strong Relation?

One can prove the Cauchy Inequality for covariance

$$[\mathrm{Cov}(X, Y)]^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)$$

How Large the Covariance Indicates a Strong Relation?

One can prove the Cauchy Inequality for covariance

$$[\mathrm{Cov}(X, Y)]^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)$$

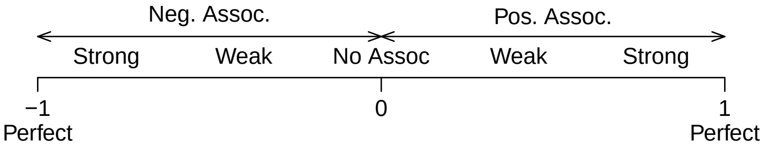
Moreover, the *covariance reaches its maximum possible magnitude if and only if X and Y has a perfect linear relation $Y = aX + b$, $a \neq 0$*.

Thus, one can assess the strength of linear relation between X, Y by comparing $\mathrm{Cov}(X, Y)$ with $\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$.

Correlation

$$\text{Correlation} = \rho_{XY} = \text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

- $-1 \le \rho_{XY} \le 1$ since $\text{Cov}(X,Y) \le \sqrt{\text{Var}(X)\,\text{Var}(Y)}$
- The closer ρ_{XY} is to 1 or to -1, the stronger the linear relation between X and Y



- $\rho_{XY} = 1$ or -1 if and only if $Y = aX + b$ and $a \neq 0$, i.e., X and Y has an perfect linear relation

Covariance Is NOT Scale Invariant but Correlation Is!

Example. Let

- X = amount of time studying STAT 234 per week, and
- Y = grade in STAT 234

If X is measured in minutes rather than in hours, $\mathrm{Cov}(X, Y)$ would be 60 times as large.

The *strength* of XY relation should be the same no matter X is measured in minutes or in hours.

Correlation ρ_{XY} is *scale invariant* and has no unit.

$$\begin{aligned}
\mathrm{Corr}(aX + c, bY + d) &= \frac{\mathrm{Cov}(aX + c, bY + d)}{\sqrt{\mathrm{Var}(aX + c)\,\mathrm{Var}(bY + d)}} \\
&= \frac{ab\,\mathrm{Cov}(X, Y)}{\sqrt{a^2\,\mathrm{Var}(X) b^2\,\mathrm{Var}(Y)}} = (\text{sign of } ab)\,\mathrm{Corr}(X, Y)
\end{aligned}$$

Example (Gas Station) — Correlation

Recall in L09, we obtained the marginal pmfs for X and Y:

x	0	1	2
$p_X(x)$	0.16	0.34	0.50

,

y	0	1	2
$p_Y(y)$	0.24	0.38	0.38

$$\mathrm{E}(X^2) = 0^2 \cdot 0.16 + 1^2 \cdot 0.34 + 2^2 \cdot 0.5 = 2.34$$

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 2.34 - 1.34^2 = 0.5444$$

$$\mathrm{E}(Y^2) = 0^2 \cdot 0.24 + 1^2 \cdot 0.38 + 2^2 \cdot 0.38 = 1.9$$

$$\mathrm{Var}(Y) = \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 = 1.9 - 1.14^2 = 0.6004$$

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{0.2724}{\sqrt{0.5444 \times 0.6004}} \approx 0.476.$$

Example 5.5 (Mixed Nuts) — Correlation

Recall in L09, we calculated the marginal pdf's for X and for Y:

$$f_X(x) = 12x(1-x)^2, \quad f_Y(y) = 12y(1-y)^2, \text{ for } 0 \le x, y \le 1.$$

using which we can calculate

$$\begin{aligned} E(X^2) &= \int_0^1 x^2 f_X(x)\mathrm{d}x = \int_0^1 12x^3(1-x)^2\mathrm{d}x \\ &= \int_0^1 12x^3 - 24x^4 + 12x^5\mathrm{d}x = 3x^4 - \frac{24x^5}{5} + 2x^6\Big|_0^1 = \frac{1}{5} \\ \mathrm{Var}(X) &= E(X^2) - (E(X))^2 = \frac{1}{5} - (\frac{2}{5})^2 = \frac{1}{25} \end{aligned}$$

Similar, one can calculate $\mathrm{Var}(Y) = 1/25$

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{-2/75}{\sqrt{(1/25)(1/25)}} = -\frac{2}{3} \approx -0.667.$$

More Properties of Covariance

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))] = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$$

In the following, a, b are constants. X, Y, Z are random variables

- Symmetry: $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
- Homogeneity: $\mathrm{Cov}(aX, bY) = ab\,\mathrm{Cov}(X,Y)$
- Right-linearity: $\mathrm{Cov}(X+Y, Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$
- Left-linearity: $\mathrm{Cov}(X, Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$
- $\mathrm{Cov}(a,X) = 0$.

Proofs for Properties of Covariance

The proofs for these propertie are all straightforward from definition. We just prove the Right-linearity as an example.

$$\begin{aligned}
\mathrm{Cov}(X+Y,Z) &= \mathrm{E}((X+Y)Z) - \mathrm{E}(X+Y)\,\mathrm{E}(Z) \\
&= \mathrm{E}(XZ) + \mathrm{E}(YZ) - [\mathrm{E}(X) + \mathrm{E}(Y)]\,\mathrm{E}(Z) \\
&= \underbrace{\mathrm{E}(XZ) - \mathrm{E}(X)\,\mathrm{E}(Z)}_{\mathrm{Cov}(X,Z)} + \underbrace{\mathrm{E}(YZ) - \mathrm{E}(Y)\,\mathrm{E}(Z)}_{\mathrm{Cov}(Y,Z)} \\
&= \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)
\end{aligned}$$

Note in the proof above, we used the property of expected value that

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$$

$$\mathrm{E}(XZ+YZ) = \mathrm{E}(XZ) + \mathrm{E}(YZ)$$

Variance of Linear Combinations of Two Random Variables

Recall that expectation has the following linear property:

$$\mathrm{E}(aX + bY) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y).$$

We also have shown that $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$.

How about $\mathrm{Var}(aX + bY)$?

$$\boxed{\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + 2ab\,\mathrm{Cov}(X, Y) + b^2\,\mathrm{Var}(Y)}$$

- If X is independent of Y, $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Proof of $\operatorname{Var}(aX + bY)$

$$\operatorname{Var}(aX + bY) = \operatorname{Cov}(aX + bY, aX + bY)$$

$$= \underbrace{\operatorname{Cov}(aX, aX + bY)}_{\downarrow} + \underbrace{\operatorname{Cov}(bY, aX + bY)}_{\downarrow} \qquad \text{(right-linearity)}$$

$$= \overbrace{\operatorname{Cov}(aX, aX) + \operatorname{Cov}(aX, bY)} + \overbrace{\operatorname{Cov}(bY, aX) + \operatorname{Cov}(bY, bY)} \qquad \text{(left-linearity)}$$

$$= \operatorname{Var}(aX) + 2\operatorname{Cov}(aX, bY) + \operatorname{Var}(bY) \qquad \text{(symmetry)}$$

$$= a^2 \operatorname{Var}(X) + 2ab \operatorname{Cov}(X, Y) + b^2 \operatorname{Var}(Y) \qquad \text{(homogeneity)}$$

Linear Combinations of Random Variables

For any random variables $X_1, X_2, \ldots, X_n$., a *linear combination* of $X_1, X_2, \ldots, X_n$ is

$$a_1 X_1 + a_2 X_2 + \cdots + a_n X_n,$$

where $a_1, a_2, \ldots, a_n$ are constant numbers. For example,

- The sum $X_1 + X_2 + \cdots + X_n$ is a linear combination of $X_1, \ldots, X_n$ with all a_i's $= 1$.
- The average
$$\frac{X_1 + X_2 + \cdots + X_n}{n}$$
is a linear combination of $X_1, X_2, \ldots, X_n$ with all a_i's $= 1/n$.
- The difference $X - Y$ is a linear combination of X and Y with $a_1 = 1$, $a_2 = -1$

Example (Total Bus Fare)

Suppose the bus fare is

\$2 for senior citizens, \$1 for children, and \$3 for all other people

Let

$$X = \text{the number of senior citizens on the bus,}$$
$$Y = \text{the number of children on the bus,}$$
$$Z = \text{the number of all other passengers on the bus}$$

The total amount of bus fares collected is then

$$2X + Y + 3Z$$

which is a linear combination of X, Y, Z.

Expected Values for Linear Combinations of RV's

For the linear combination

$$a_1X_1 + a_2X_2 + \cdots + a_nX_n,$$

the expected value is

$$\begin{aligned}
& \mathrm{E}(a_1X_1 + a_2X_2 + \cdots + a_nX_n) \\
= {} & \mathrm{E}(a_1X_1) + \mathrm{E}(a_2X_2) + \cdots + \mathrm{E}(a_nX_n) && \text{by linearity of expected value} \\
= {} & a_1\,\mathrm{E}(X_1) + a_2\,\mathrm{E}(X_2) + \cdots + a_n\,\mathrm{E}(X_n) && \text{since } \mathrm{E}(aX) = a\,\mathrm{E}(X)
\end{aligned}$$

Variance of a Linear Combination of RV's

$$\operatorname{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{Var}(X_i) + 2 \sum_{i<j} a_i a_j \operatorname{Cov}(X_i, X_j)$$

- There is a covariance term for every pair of X_i and X_j
- When $X_1, \ldots, X_n$ are independent, then

$$\operatorname{Var}(X_1 + X_2 + \ldots + X_n) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + \ldots + \operatorname{Var}(X_n).$$

- When $\operatorname{Var}(X_i) = \sigma^2$ for $i = 1, \ldots, n$,
  and $\operatorname{Cov}(X_i, X_j) = \rho$ for $1 \le i \neq j \le n$, then

$$\operatorname{Var}(X_1 + \ldots + X_n) = n\sigma^2 + n(n-1)\rho.$$

Example: Variance of the Binomial Distribution

In Lecture 6A, we computed the expected value for the Binomial distribution Bin(n, p), but the variance is given without proof.

$$\mu = E(X) = np, \quad \sigma^2 = \text{Var}(X) = np(1-p).$$

Here we will to prove the formulas using linear combinations.

First for the special case $n = 1$, $X \sim$ Bin$(n = 1, p)$, X only takes value 0 and 1 with the pmf below

x	0	1
$p(x)$	$1-p$	p

Hence

$$\text{E}(X) = \sum\nolimits_{x=0,1} xp(x) = 0 \cdot (1-p) + 1 \cdot p = \boxed{p},$$

$$\text{E}(X^2) = \sum\nolimits_{x=0,1} x^2 p(x) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

$$\text{Var}(X) = \text{E}(X^2) - (\text{E}(X))^2 = p - p^2 = \boxed{p(1-p)}$$

For general n, recall a Binomial random variable $X \sim \text{Bin}(n, p)$ is the total number of successes obtained in n independent Bernoulli trials. For each of the n trials, define

$$X_i = \begin{cases} 1 & \text{if success in the } i\text{th trial} \\ 0 & \text{if failure in the } i\text{th trial} \end{cases} \qquad \Rightarrow X_i \sim \text{Bin}(n = 1, p).$$

$$\begin{aligned} \text{Then } X &= \text{the number of successes obtained in the } n \text{ trials} \\ &= X_1 + X_2 + \ldots + X_n, \end{aligned}$$

The expected value and variance of X are thus

$$\text{E}(X) = \underbrace{\text{E}(X_1)}_{=p} + \cdots + \underbrace{\text{E}(X_n)}_{=p} = np$$

$$\text{Var}(X) = \underbrace{\text{Var}(X_1)}_{=p(1-p)} + \cdots + \underbrace{\text{Var}(X_n)}_{=p(1-p)} = np(1-p)$$

since X_i's are indep. and each with mean p and variance $p(1-p)$ as $X_i \sim \text{Bin}(n = 1, p)$.

Example (Sample Mean)

Suppose $X_1, \dots, X_n$ are *i.i.d.* rv's with mean μ and variance σ^2.

- *i.i.d.* = "independent and have an identical distribution"

Consider the *sample mean*

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

Then

$$\mathrm{E}(\overline{X}) = \frac{1}{n}[\mathrm{E}(X_1) + \ldots + \mathrm{E}(X_n)] = \frac{1}{n}(\underbrace{\mu + \ldots + \mu}_{n \text{ copies}}) = \mu.$$

$$\begin{aligned}
\mathrm{Var}(\overline{X}) &= \frac{1}{n^2}\,\mathrm{Var}(X_1 + X_2 + \ldots + X_n) \quad \text{since } \mathrm{Var}(aX) = a^2 V(X) \\
&= \frac{1}{n^2}[\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)] \quad \text{as all } X_i\text{'s are indep.} \\
&= \frac{1}{n^2}(\underbrace{\sigma^2 + \ldots + \sigma^2}_{n \text{ copies}}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned}$$