Review of Linear Models Class 1-3

Robert Gulotty

University of Chicago

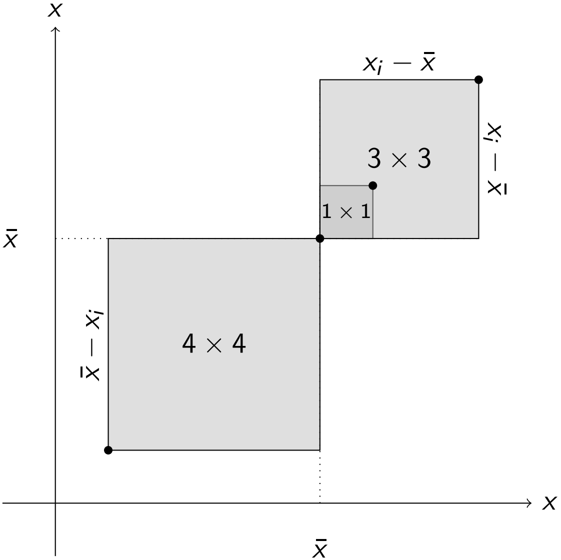
April 6, 2023

Often the sample (co)variance appears in an alternative form:

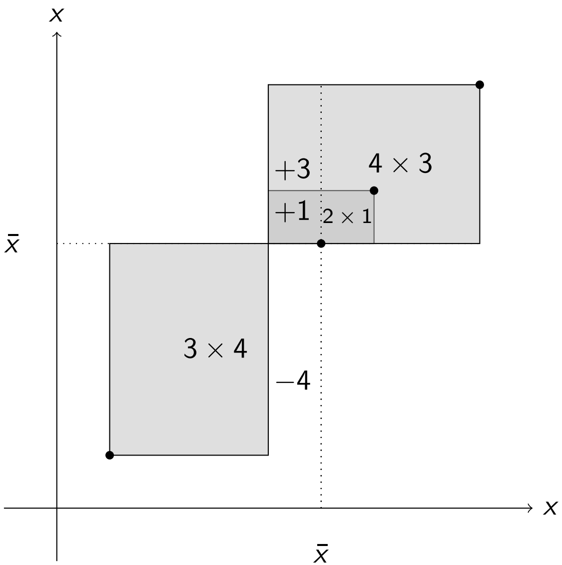
$$
\begin{aligned}
\widehat{Var}(X) &= \frac{1}{N}\sum_i (x_i - \bar{x})(x_i - \bar{x}) \\
&= \frac{1}{N}\sum_i (x_i - \bar{x})x_i \\
\widehat{Cov}(X, Y) &= \frac{1}{N}\sum_i (x_i - \bar{x})(y_i - \bar{y}) \\
&= \frac{1}{N}\sum_i (x_i - \bar{x})y_i
\end{aligned}
$$

Algebra Demonstration

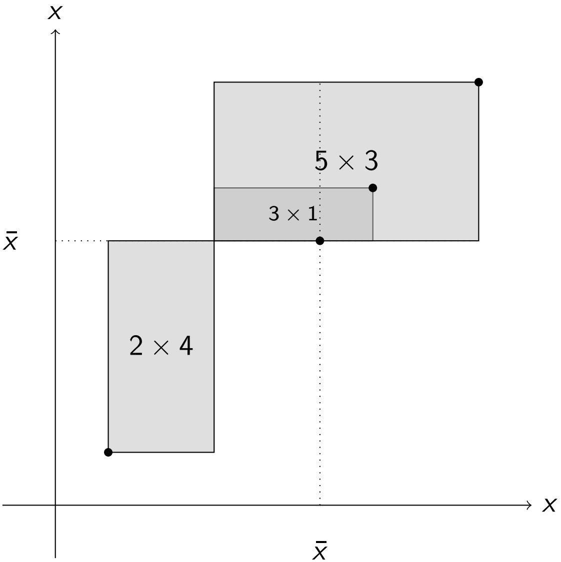
$$\begin{aligned}
\widehat{Var}(X) &= \frac{1}{N}\sum_i (x_i - \bar{x})(x_i - \bar{x}) \\
&= \frac{1}{N}\sum_i x_i^2 - \bar{x}x_i - x_i\bar{x} + \bar{x}\bar{x} && \text{(Expand the square)} \\
&= \frac{1}{N}\sum_i x_i^2 - \bar{x}\frac{1}{N}\sum_i x_i - \bar{x}\frac{1}{N}\sum_i x_i + \frac{1}{N}\sum_i \bar{x}\bar{x} && \text{(Distribute the } \textstyle\sum\text{)} \\
&= \frac{1}{N}\sum_i x_i^2 - \bar{x}\frac{1}{N}N\bar{x} - \bar{x}\frac{1}{N}\sum_i x_i + \bar{x}\bar{x} && (\textstyle\sum_i x_i = N\bar{x}) \\
&= \frac{1}{N}\sum_i x_i^2 - \bar{x}\frac{1}{N}\sum_i x_i && (\bar{x}\bar{x} \text{ cancel}) \\
&= \frac{1}{N}\sum_i (x_i - \bar{x})x_i
\end{aligned}$$



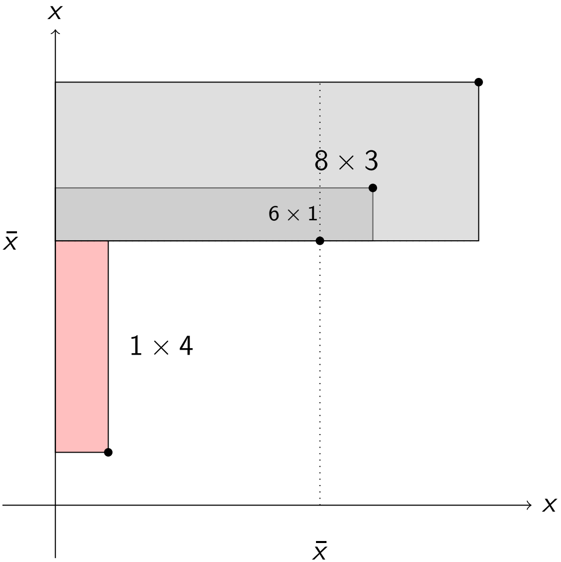
$$
\begin{aligned}
\widehat{Var}(X) &= \frac{1}{N}\sum_i (x_i - \bar{x})(x_i - \bar{x}) \\
&= \frac{(-4 * -4) + (1 * 1) + (3 * 3)}{4} \\
&= \frac{26}{4}
\end{aligned}
$$



$$\widehat{Var}(X) = \frac{1}{N}\sum_i (x_i - \bar{x})(x_i - (\bar{x} - 1))$$
$$= \frac{(-4 * -3) + (1 * 2) + (3 * 4)}{4}$$
$$= \frac{26}{4}$$



$$\widehat{Var}(X) = \frac{1}{N}\sum_i (x_i - \bar{x})(x_i - (\bar{x} - 2))$$
$$= \frac{(-4 * -2) + (1 * 3) + (3 * 5)}{4}$$
$$= \frac{26}{4}$$



$$\widehat{Var}(X) = \frac{1}{N}\sum_i (x_i - \bar{x})(x_i - (\bar{x} - \bar{x}))$$

$$\widehat{Var}(X) = \frac{1}{N}\sum_i (x_i - \bar{x})(x_i)$$

$$= \frac{(-4 * 1) + (1 * 6) + (3 * 8)}{4}$$

$$= \frac{26}{4}$$

How does y_k affect $Cov(X, Y)$?

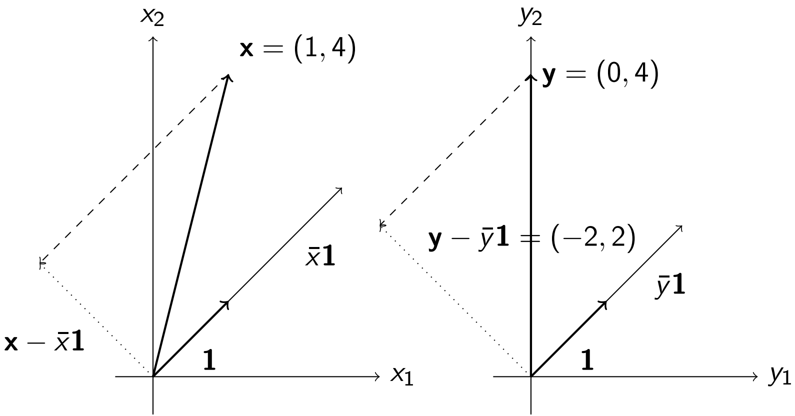
$$
\begin{aligned}
\widehat{Cov}(X,Y) &= \frac{1}{N}\sum_i (x_i - \bar{x})(y_i - \bar{y}) \\
&= \frac{1}{N}\sum_i (x_i - \bar{x})(y_i - \frac{1}{N}\sum_j y_j) \\
&= \frac{1}{N}\sum_i (x_i - \bar{x})y_i
\end{aligned}
$$

We can take the derivative with respect to y_k in only one term of the summation.

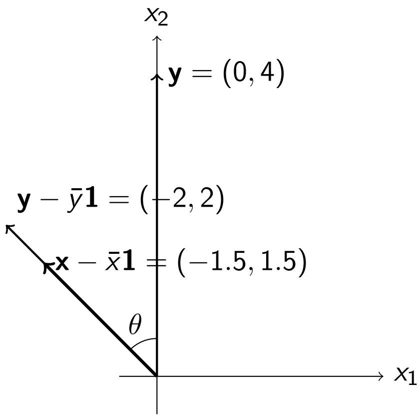
Geometry and the Dot product

Suppose
$\boldsymbol{x} = (1,4)$, $\bar{x} = 2.5$,
$\boldsymbol{y} = (0,4)$, $\bar{y} = 2$

$$(\boldsymbol{x} - \bar{x}\mathbf{1}) \cdot (\boldsymbol{y}) =$$
$$-1.5 * 0 + 1.5 * 4 = 6$$
$$(\boldsymbol{x} - \bar{x}\mathbf{1}) \cdot (\boldsymbol{y} - \bar{y}\mathbf{1}) =$$
$$-1.5 * (-2) + 1.5 * 2 = 6$$



$\boldsymbol{a} \cdot \boldsymbol{b} = ||\boldsymbol{a}||||\boldsymbol{b}|| \cos\theta$



Note, $\boldsymbol{y} - \bar{y}\mathbf{1} \perp \bar{y}\mathbf{1}$, so $[(0,4)-(0,0)-(-2,2)]$ is a right triangle and $\cos\theta = \frac{\text{Adjacent}}{\text{Hypotenuse}}$

$$||\boldsymbol{y} - \bar{y}\mathbf{1}|| = ||\boldsymbol{y}|| \cos\theta$$

Any vector minus its means lies on a hyperplane $\perp \mathbf{1}$ and so are parallel to one another ($\theta = 0$ or π)

$$\begin{aligned} \boldsymbol{y} - \bar{y}\mathbf{1} \parallel \boldsymbol{x} - \bar{x}\mathbf{1} \rightarrow \cos(0) = 1 \rightarrow \\ ||\boldsymbol{x} - \bar{x}\mathbf{1}||||\boldsymbol{y}|| \cos(\theta) = \\ \sqrt{4.5}\sqrt{16} * cos(\pi/4) = 6 \\ ||\boldsymbol{x} - \bar{x}\mathbf{1}||||\boldsymbol{y} - \bar{y}\mathbf{1}|| \cos(\theta) = \\ \sqrt{4.5} * \sqrt{8} \cos(0) = 6 \end{aligned}$$