An R^2 Lesson from a walk through Manhattan

Ryan Brill*

August 2025

Setup. Think of yourself as a pedestrian in Manhattan. Think of Manhattan as a grid of equally spaced city blocks in the four cardinal directions (yes, this is a simplification). You begin at $(0,0)$. Horizontally you walk some fraction of blocks $X \sim \text{Unif}[-1,1]$ and vertically you walk $Y \sim \text{Unif}[-10,10]$ blocks, where X and Y are independent. $X > 0$ is East/right, $X < 0$ is West/left, $Y > 0$ is North/up, and $Y < 0$ is South/down. $U := |X| \sim \text{Unif}[0,1]$ is the horizontal distance traveled and $V := |Y| \sim \text{Unif}[0,10]$ is the vertical distance traveled. $D = U + V$ is the total city-blocks distance traveled.

Regression. Consider predicting the total distance traveled from just the horizontal distance traveled. This amounts to regressing D on just U, the simple linear regression $D = \beta_0 + \beta_1 U + \varepsilon$. Since $D = U + V$ with $U \perp V$,

$$\hat{\beta}_1 = \frac{\text{Cov}(D,U)}{\text{Var}(U)} = \frac{\text{Var}(U)}{\text{Var}(U)} = 1 \quad \text{and} \quad \hat{\beta}_0 = \mathbb{E}[D] - \hat{\beta}_1 \mathbb{E}[U] = \mathbb{E}[V] = 5.$$

We also have

$$\text{Var}(U) = \tfrac{1}{12}, \qquad \text{Var}(V) = \tfrac{100}{12}, \qquad \text{Var}(D) = \text{Var}(U) + \text{Var}(V) = \tfrac{101}{12}.$$

Therefore,

$$R^2 = \frac{\text{Var}(\hat{D})}{\text{Var}(D)} = \frac{\hat{\beta}_1^2 \text{Var}(U)}{\text{Var}(D)} = \frac{\frac{1}{12}}{\frac{101}{12}} = \frac{1}{101} \approx 0.0099 \approx 0.010 \quad \text{and} \quad R = \frac{1}{\sqrt{101}} \approx 0.0995 \approx 0.10.$$

Implication. Despite the *exact* and interpretable slope $\hat{\beta}_1 = 1$ (each extra horizontal block adds one block to D), both R^2 and R are tiny because variability in V dominates. Low R^2 (or low R) reflects a low fraction of explained variance, *not* an unimportant or biased effect.

Better things to think about than R^2. Think about the effect size $\hat{\beta}_1 = 1$ and its uncertainty. Think about out-of-sample error metrics in the original units (e.g., RMSE or MAE). I like these more than R^2.

*Adi Wyner mentioned this idea at lunch.