JSC270 Bonus Question

Let X, Y be random variables. Suppose that we have n observations $\{(x_i, y_i)\}_{i=1}^{n}$ for $i \in \{1, 2, \ldots, n\}$, where each observation is sampled from these variables.

Furthermore, suppose that we performed a linear regression to determine the relationship of these variables. Thus, we can write $Y = \hat{\beta}_0 + \hat{\beta}_1 X + \epsilon$, for constants $\hat{\beta}_0, \hat{\beta}_1 \in \mathbb{R}$, and a variable error term ϵ.

Per the formula from the slides, we know that $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{Cov(X,Y)}{Var(X)}$, by the statistical definition of the covariance and variance for these variables.

Furthermore, we know that $Corr(X, Y) = \frac{Cov(X,Y)}{SD(X) \cdot SD(Y)} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$, again, by the definition of this function on the random variables, and the definition of standard deviation as the square root of the variance of a variable.

Suppose that $Cov(X, Y) \neq 0$, and that X and Y are not constants (and hence $Var(X) > 0, Var(Y) > 0$).

$$
\begin{aligned}
\frac{\hat{\beta}_1}{Corr(X,Y)} &= \frac{\frac{Cov(X,Y)}{Var(X)}}{\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}} && \text{(As shown above)} \\
&= \Big(\frac{Cov(X,Y)}{Var(X)}\Big)\Big(\frac{\sqrt{Var(X)}\sqrt{Var(Y)}}{Cov(X,Y)}\Big) && \text{(Cross-multiplying)} \\
&= \frac{\sqrt{Var(X)}\sqrt{Var(Y)}}{Var(X)} && \text{(By assumption that } Cov(X,Y) \neq 0\text{)} \\
&= \frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}} && \text{(Simplifying)} \\
&= \frac{SD(Y)}{SD(X)} && \text{(Definition of standard deviation)} \\
\hat{\beta}_1 &= \frac{Corr(X,Y)SD(Y)}{SD(X)} && \text{(Multiplying through by } Corr(X,Y)\text{)}
\end{aligned}
$$

Thus, we see that the estimated slope for a univariate linear regression is equal to the correlation of the two variables multiplied by the ratio of the standard deviations of the ‘independent’ and ‘dependent’ variables.