Exercise B

In the linear model $\mathbb{E}[y_i] = \beta_1 + \beta_2 x_i$, suppose that instead of observing x_i we observe $x_i^* = x_i + u_i$, where u_i is independent of x_i and $\mathrm{var}(u_i) = \sigma_u^2$, with $\mathbb{E}[u_i] = \mu_u$. Recall that

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \qquad \text{(By assumption)}$$

Add and subtract $\beta_2 u_i$ →

$$= \beta_1 + \beta_2 u_i - \beta_2 u_i + \beta_2 x_i + \varepsilon_i$$

$$= \beta_1 + \beta_2 (x_i - u_i) + \beta_2 u_i + \varepsilon_i$$

Add and subtract $\beta_2 \mu_u$ →

$$= \beta_1 + \beta_2 \mu_u + \beta_2 x_i^* + \overbrace{\beta_2 (u_i - \mu_u) + \varepsilon_i}^{\varepsilon_i^*}$$

$$\boxed{y_i = \beta_1^* + \beta_2^* x_i^* + \varepsilon_i^*.}$$

$\varepsilon_i^* \overset{iid}{\sim} \mathcal{F}_*$, with

$$\mathbb{E}[\varepsilon_i^*] = \beta_2 \mathbb{E}[u_i - \mu_u] + \mathbb{E}[\varepsilon_i] = 0$$

$$\mathrm{var}(\varepsilon_i^*) = \beta_2^2 \sigma_u^2 + \sigma^2 = \sigma_*^2$$

$$\beta_1^* = \beta_1 + \beta_2 \mu_u; \qquad \beta_2^* = \beta_2.$$

This is a linear model with covariate x_i^* and errors ε_i^*. The estimator

$$\hat{\beta}_2^* = \frac{\mathrm{cov}(x^*, y)}{\mathrm{var}(x^*)} = \frac{\frac{1}{m}\sum_{i=1}^{m} (x_i^* - \bar{x}^*)(y_i - \bar{y})}{\frac{1}{m}\sum_{i=1}^{m} (x_i^* - \bar{x}^*)^2}$$

is unbiased and consistent for $\beta_2^* = \beta_2$, even under a "perturbed" covariate.

$$\mathbb{E}[\hat{\beta}_2^*] = \beta_2^* = \beta_2.$$

(This follows from the equation above, but you can prove it directly).

<u>However</u>, the estimator for β_1 is <u>problematic</u>

$$\mathbb{E}[\hat{\beta}_1^*] = \mathbb{E}[\bar{y} - \bar{x}^* \hat{\beta}_2^*] = \boxed{\beta_1^* \neq \beta_1}$$

because $\beta_1^* = \beta_1 + \beta_2 \mu_u$.

Therefore the estimator is <u>inconsistent</u> for β_1 unless

1. $\beta_2 = 0$ $\left(\beta^*_1 = \beta_1 + 0 = \beta_1 \right)$. The covariate is irrelevant a non-interesting situation.
2. $\mu_v = 0$ $\left(\beta^*_1 = \beta_1 + 0 = \beta_1 \right)$. The measurement error is unbiased.