Regression

8/22

$$\underbrace{y_i}_{\text{random}} = \underbrace{\beta_0 + \beta_1 x_i}_{\text{fixed}} + \underbrace{\varepsilon_i}_{\text{random}}$$

How to estimate the values of β_0 & β_1.

↳ Use LS Principle

↳ minimize $Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$

Call

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} Q(\beta_0, \beta_1)$$

↳ Closed form:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

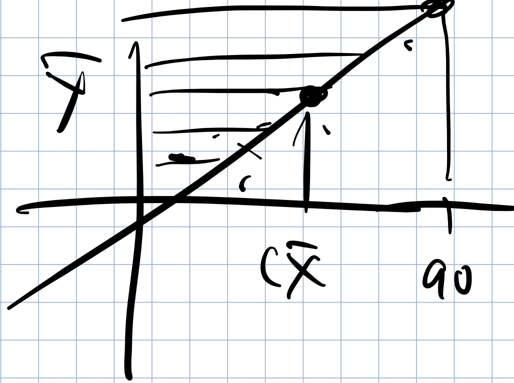
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

LS Estimates

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

$=$

$(\bar{x}, \bar{y})$



Ex: If I want to predict the y-value for a given x-value (e.g. $x = 90$)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = \hat{\beta}_0 + \hat{\beta}_1 (90)$$

$$35.5 + 0.44(90)$$

Are our LS estimators any good?

Gauss-Markov Theorem:

The LS estimators $\hat{\beta}_0$ & $\hat{\beta}_1$ are unbiased,

i.e. $E(\hat{\beta}_0) = \beta_0$

$E(\hat{\beta}_1) = \beta_1$.

And they are the

best linear unbiased estimators (BLUE)

for the true parameters β_0 & β_1.

BLUE: best linear unbiased est.
① ② ③

① best: smallest possible variance out of all linear & unbiased est.

② linear: the estimator is a linear combination of the y_i's

③ unbiased: $\text{bias}(\hat{\theta}) = 0$

$E(\hat{\theta}) = \theta$

Let's focus on the slope est. $\hat{\beta}_1$.

I claimed that

$\hat{\beta}_1$ was linear in y_i

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^n (x_i-\bar{x})^2}$$

why?

$\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})$

$\sum_{i=1}^n (x_i-\bar{x})y_i - \sum_i (x_i-\bar{x})\bar{y}$

$'' - \bar{y}\sum_i (x_i-\bar{x})$

$'' - \bar{y}(\sum_i x_i - n\bar{x})$

$'' - \bar{y}(0)$

$n\bar{x} = \cancel{n}\frac{1}{\cancel{n}}\sum_i x_i$

$$= \frac{\sum_{i=1}^n (x_i-\bar{x})y_i}{\sum_{i=1}^n (x_i-\bar{x})^2}$$

call this SSX

$$= \sum_{i=1}^n \left[\frac{(x_i-\bar{x})}{SSX}\right] y_i$$

$$= \underbrace{\sum_{i=1}^n k_i y_i}$$

$\Rightarrow \hat{\beta}_1$ is linear in y_i's.

$\sum_i (x_i-\bar{x})y_i$

$\sum_i x_i y_i - \sum_i \bar{x} y_i$

$\sum_i x_i y_i - \bar{x}\sum_i y_i$

$$SSX = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$k_i = \frac{x_i - \bar{x}}{SSX}$$

Next let's show that $\hat{\beta}_1$ is unbiased

$$E(\hat{\beta}_1) = E\left(\sum_{i=1}^{n} k_i y_i\right)$$

$$= \sum_{i=1}^{n} k_i \, E(y_i)$$

$$= \sum_{i=1}^{n} k_i \, (\beta_0 + \beta_1 x_i)$$

$$= \sum_{i=1}^{n} k_i \beta_0 + \sum_{i=1}^{n} k_i \beta_1 x_i$$

$$= \beta_0 \underbrace{\sum_{i=1}^{n} k_i} + \beta_1 \underbrace{\sum_{i=1}^{n} k_i x_i}$$

(want this to be β_1)

$$= \beta_0 (0) + \beta_1 (1) = \beta_1$$

$$\sum_{i=1}^{n} k_i = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{SSX}$$

$$= \frac{1}{SSX} \underbrace{\sum_i (x_i - \bar{x})}_{0} = 0$$

$$\sum_{i=1}^{n} k_i x_i = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{SSX} x_i$$

$$= \frac{1}{SSX} \sum_{i=1}^{n} (x_i - \bar{x}) x_i = \frac{SSX}{SSX} = 1$$

$$\sum_{i=1}^{n} (x_i-\bar{x})\, x_i \overset{?}{=} \sum_{i=1}^{n} (x_i-\bar{x})^2 \leftarrow SSX$$

$$= \sum_{i=1}^{n} (x_i^2 - \bar{x}\, x_i)$$

$$= \sum_i x_i^2 - \bar{x} \sum_i x_i$$

$$= \underline{\sum_i x_i^2 - n\bar{x}^2}$$

$$= \sum_{i=1}^{n} (x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= \sum_i x_i^2 - 2\bar{x}\sum_i x_i + n\bar{x}^2$$

$$= \sum_i x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$= \underline{\sum_i x_i^2 - n\bar{x}^2}$$

Next we want to show that $\hat{\beta}_1$ has the smallest variance of all linear est.

$$Var(\hat{\beta}_1) = Var\left(\sum_{i=1}^{n} k_i y_i\right)$$

$$= \sum_{i=1}^{n} Var(k_i y_i)$$

by $Cov(y_i, y_j) = 0$ for $i \neq j$

$$= \sum_{i=1}^{n} k_i^2 \, Var(y_i)$$

$$= \sum_{i=1}^{n} k_i^2 \, Var(\underbrace{\beta_0 + \beta_1 X_i}_{\text{fixed}} + \varepsilon_i)$$

$$= \sum_{i=1}^{n} k_i^2 \, Var(\varepsilon_i)$$

$$= \sum_{i=1}^{n} k_i^2 \sigma^2$$

$$= \sigma^2 \sum_{i=1}^{n} k_i^2$$

$$= \sigma^2 \sum_{i=1}^{n} \left[\frac{(x_i - \bar{x})}{SSX} \right]^2$$

$$= \sigma^2 \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{(SSX)^2} = \frac{\sigma^2}{(SSX)^2} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{\sigma^2 \, SSX}{(SSX)^2}$$

$$= \frac{\sigma^2}{SSX} = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}$$

Why is this the smallest possible variance (among linear estimators)?

Sketch:

Consider an alternative linear estimator $\tilde{\beta}_1$.

Want to show that

$$\mathrm{Var}(\hat{\beta}_1) \le \mathrm{Var}(\tilde{\beta}_1) \text{ always.}$$

Idea:

Define $\tilde{\beta}_1 = \sum_{i=1}^{n} \tilde{k}_i y_i$

where $\tilde{k}_i = \hat{k}_i + d_i$ where

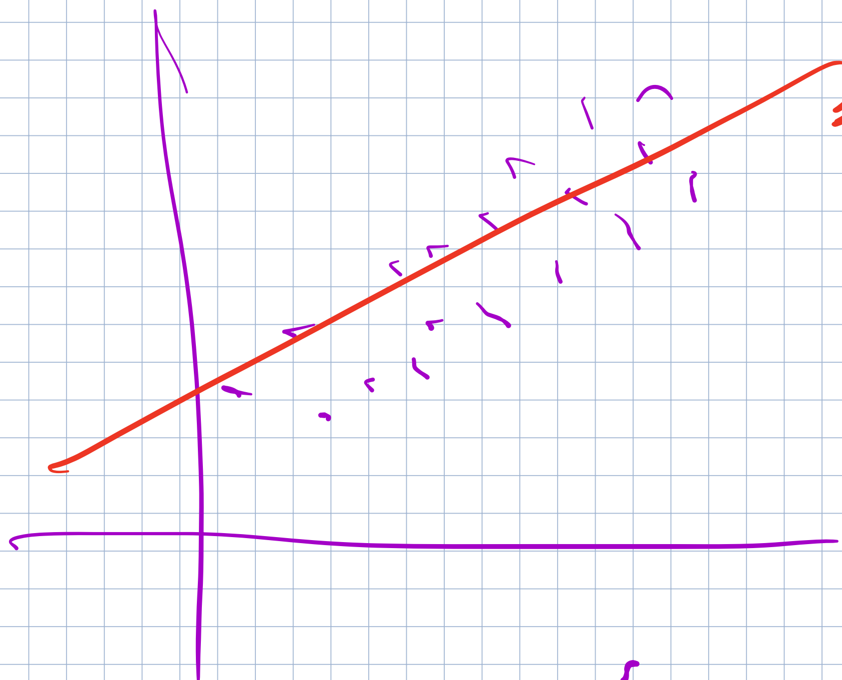
$d_i \neq 0$
$\forall i$.

$$\mathrm{Var}(\tilde{\beta}_1) = \mathrm{Var}\left(\sum_i (k_i + d_i)\, y_i\right)$$

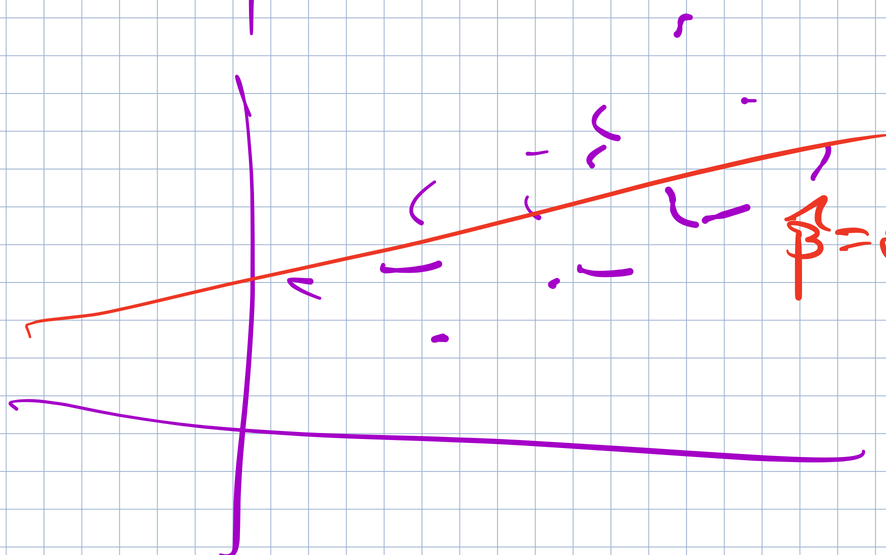
$$= \vdots$$

$$= \mathrm{Var}(\hat{\beta}_1) + \underbrace{\qquad}_{\geq 0}$$

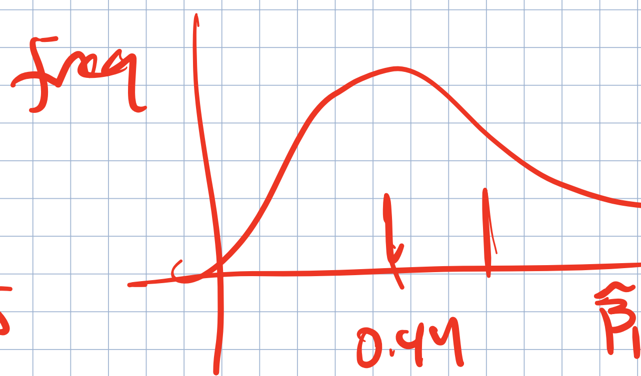
What does var($\hat{\beta}_1$) mean intuitively?



$\hat{\beta}_1 = 0.44$ sample of 300 movies



$\hat{\beta} = 0.5$

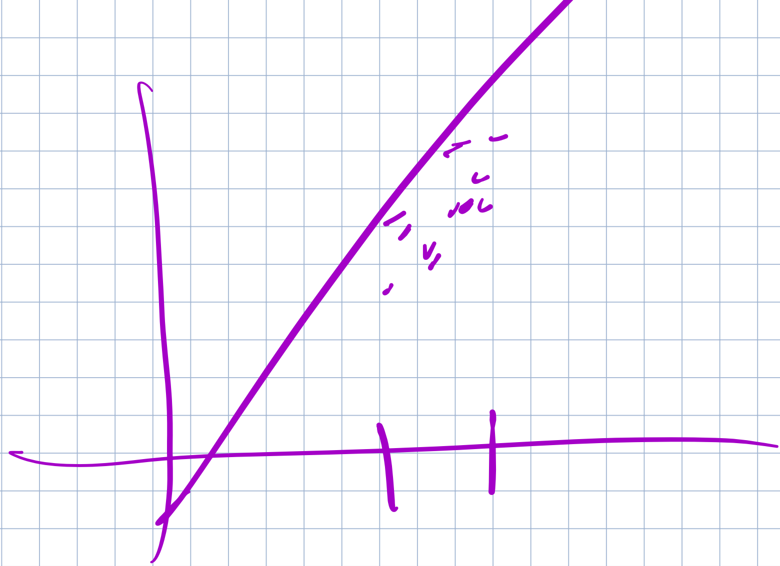


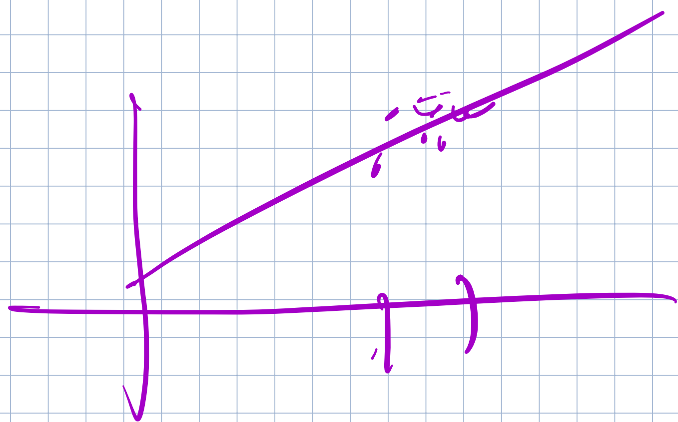
If we assume (iv) then

$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{SSx})$

Reduce the spread of x's:

$\Rightarrow$ higher variance in $\hat{\beta}_1$





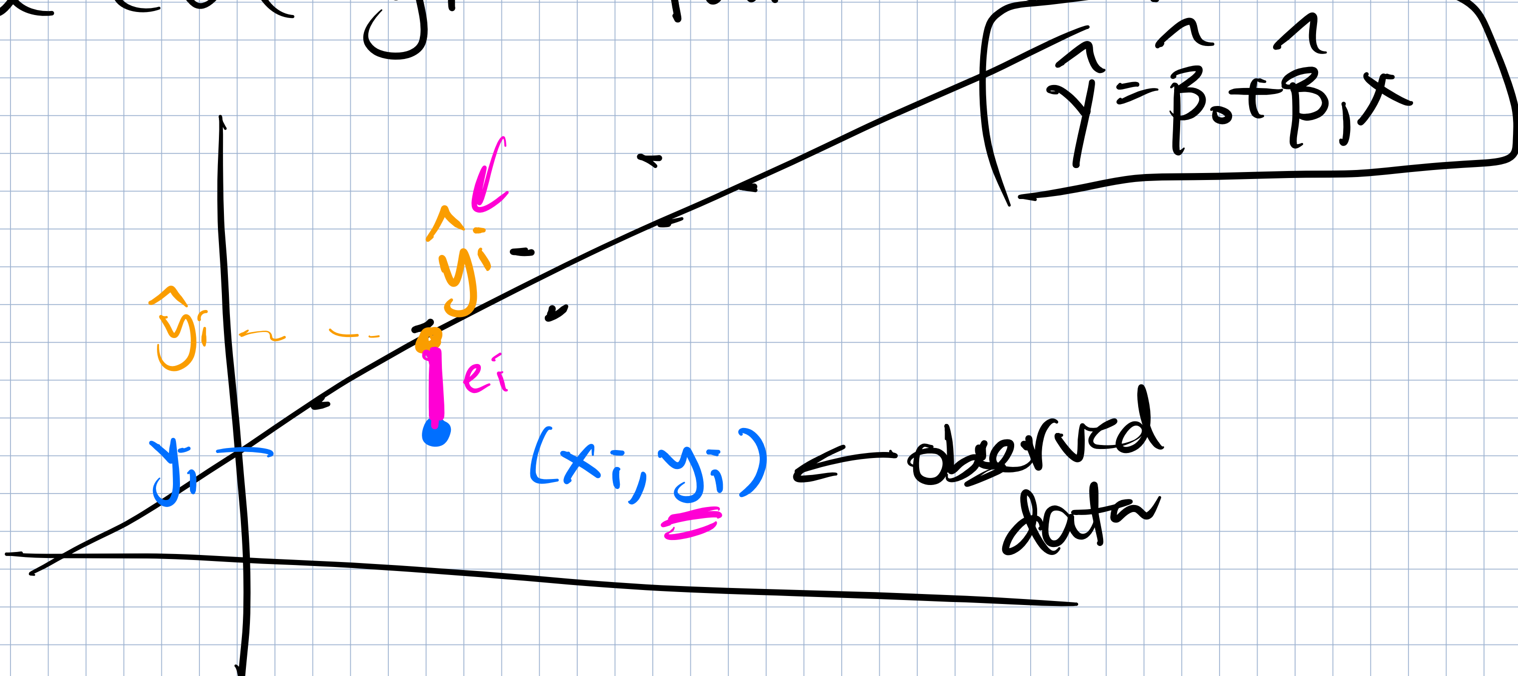
If I fit my LS Estimators & get values for $(\hat{\beta}_0, \hat{\beta}_1)$ I can use them to predict the response @ any given predictor level:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$ ← Fitted LSRL

If I'm specifically talking about the predictions for my sample:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

I call $\hat{y}_i$ a "fitted value"



- $\hat{y}_i$ predicts the value of Y @ $X = x_i$

single observation

- $\hat{y}_i$ estimates $E(Y | X = x_i) = g(x_i)$

$$\underbrace{\beta_0 + \beta_1 x_i}_{\text{population qty}}$$

Def

The residual for the i^{th} observation is the diff. b/w the observed response & the fitted value:

$$e_i = y_i - \hat{y}_i$$

The residuals try to quantify

our prediction error.

Note: there is a difference b/w the residuals e_i & the model errors ε_i

Notice the difference:

$$\varepsilon_i = Y_i - \beta_0 - \beta_1 X_i$$

vs.

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

e_i predicts ε_i

Recall:

$$\varepsilon_i \overset{iid}{\sim} N(0, \underline{\sigma^2})$$

It can be shown

$$e_i \sim N\left(0, \sigma^2\left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{SSx}\right]\right)$$

Notice:

$Var(e_i) \leq Var(\varepsilon_i)$

Properties of e_i:

- e_i is a RV: $e_i = y_i - \hat{y}_i$
- e_i is Normally distributed [under (iv)]
- e_i is estimated error of the $\boxed{\text{in sample}}^{*}$ y_i's

Let's characterize the mean & variance.

(want to show)
WTS:

$E(e_i) = 0$ (up to you)

$Var(e_i) = \sigma^2 \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{SSX} \right]$

start:

$$\mathrm{Var}(e_i) = \mathrm{Var}(y_i - \hat{y}_i)$$

$$= \mathrm{Var}\Big(\underbrace{\beta_0 + \beta_1 x_i}_{\text{fixed}} + \overset{\text{random}}{\varepsilon_i} - \overbrace{(\hat{\beta}_0 + \hat{\beta}_1 x_i)}^{\text{random}}\Big)$$

$$= \mathrm{Var}(\varepsilon_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

$$= \mathrm{Var}(\varepsilon_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i)$$

$$= \mathrm{Var}\big(\varepsilon_i - \bar{y} + \hat{\beta}_1(\bar{x} - x_i)\big) \quad ★$$

$$= \mathrm{Var}(\varepsilon_i) + \mathrm{Var}(\bar{y}) + (x_i - \bar{x})^2 \mathrm{Var}(\hat{\beta}_1)$$

$$- 2\mathrm{Cov}(\varepsilon_i, \bar{y}) \quad ①$$

$$- 2\mathrm{Cov}(\varepsilon_i, \hat{\beta}_1 (x_i - \bar{x})) \quad ②$$

$$+ 2\mathrm{Cov}(\bar{y}, \hat{\beta}_1 (x_i - \bar{x})) \quad ③$$

$$= \sigma^2 + \frac{\sigma^2}{n} + (x_i - \bar{x})^2 \frac{\sigma^2}{SSX}$$

$$- ① - ② + ③$$ (worked out below!)

$$= \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SSX} \right)$$

$$- 2 \left(\frac{\sigma^2}{n} + \frac{\sigma^2 (x_i - \bar{x})^2}{SSX} + 0 \right)$$

$$= \sigma^2 \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{SSX} \right)$$!!

① $\operatorname{Cov}(\varepsilon_i, \bar{Y}) =$

$$\operatorname{Cov}\left(\varepsilon_i, \frac{1}{n}\sum_j y_j\right) =$$

$$\frac{1}{n}\sum_j \operatorname{Cov}(\varepsilon_i, y_j) =$$

$$\frac{1}{n}\sum_j \operatorname{Cov}(\varepsilon_i, \underbrace{\beta_0 + \beta_1 x_j}_{} + \varepsilon_j) =$$

$$\frac{1}{n}\sum_i \operatorname{Cov}(\varepsilon_i, \varepsilon_j) =$$

$$\frac{1}{n}\Big(\cancel{\operatorname{Cov}(\varepsilon_i, \varepsilon_1)}^{0} + \cancel{\operatorname{Cov}(\varepsilon_i, \varepsilon_2)}^{0} +$$

$$\dots + \operatorname{Cov}(\varepsilon_i, \varepsilon_i) +$$

$$\dots + \cancel{\operatorname{Cov}(\varepsilon_i, \varepsilon_n)}^{0} \Big)$$

$$= \sigma^2 / n$$

② $\mathrm{Cov}(\varepsilon_i, (x_i-\bar{x})\hat{\beta}_1)$

$$= \mathrm{Cov}(\varepsilon_i, (x_i-\bar{x})\sum_{j=1}^{n} k_j y_j)$$

$$= \mathrm{Cov}(\varepsilon_i, \underline{(x_i-\bar{x})k_i}\, y_i)$$

$$= (x_i-\bar{x})k_i \mathrm{Cov}(\varepsilon_i, \varepsilon_i) = (x_i-\bar{x})k_i\sigma^2$$

$$= \frac{(x_i-\bar{x})^2}{SSX}\sigma^2$$

③ $\operatorname{Cov}\left(\bar{Y}, (x_i - \bar{x})\hat{\beta}_1\right)$

$$= \operatorname{Cov}\left(\frac{1}{n}\sum_{j=1}^{n} y_j, (x_i-\bar{x})\sum_{l=1}^{n} k_l y_l\right)$$

$$= \left(\frac{x_i-\bar{x}}{n}\right)\sum_{j=1}^{n}\sum_{l=1}^{n} \underbrace{\operatorname{Cov}(y_j, k_l y_l)}_{0 \text{ for all } j \neq l}$$

$$= \frac{(x_i-\bar{x})}{n}\sum_{j=1}^{n} k_j \sigma^2$$

$$= \frac{(x_i-\bar{x})}{n}\sigma^2 \cancelto{0}{\sum_{j=1}^{n} k_j} = 0$$