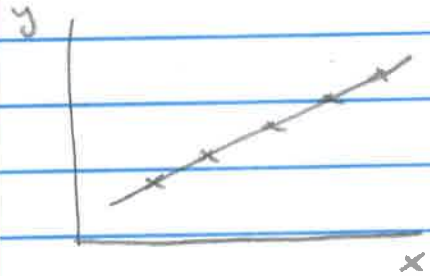
Ch. 11 Linear Models and Estimation by Least Squares

Introduction

Deterministic model : $y = \beta_0 + \beta_1 x$

e.g. x = speed of car

y = distance traveled



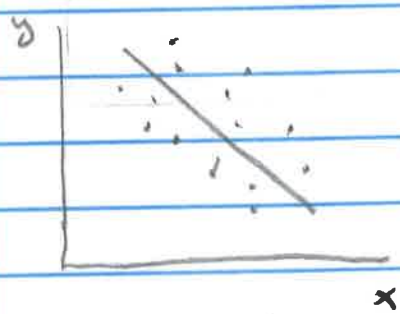
In this case, x (speed of car) determines y (distance traveled) i.e. there is no error or variability about the line.

Probabilistic model: $y = \beta_0 + \beta_1 x + \varepsilon$

↳ RV that is usually assumed to $\varepsilon \sim N(0, \sigma^2)$

e.g. x = wt. of car

y = gas mileage



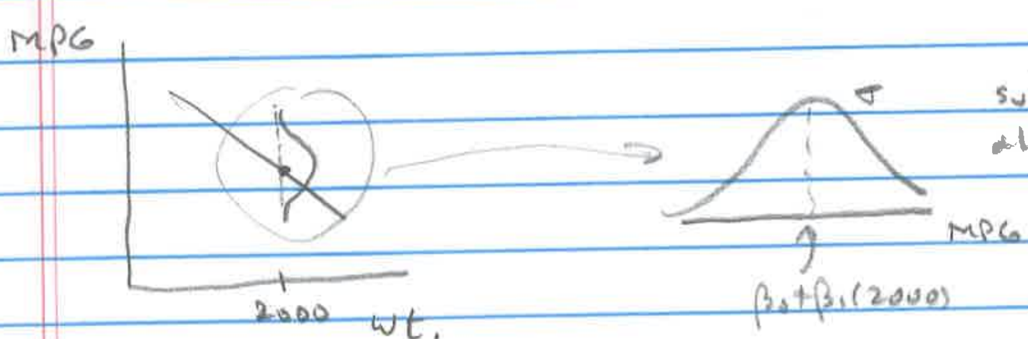
note: $E(y) = E(\beta_0 + \beta_1 x + \varepsilon)$

$= \beta_0 + \beta_1 x + E(\varepsilon) = \boxed{\beta_0 + \beta_1 x}$

mean of Y depends on X

$V(y) = V(\beta_0 + \beta_1 x + \varepsilon) = V(\varepsilon) = \sigma^2$

does not depend on X.



subpop. of MPG for all cars that weigh 2000 lbs

11.2 Linear Statistical Models

Def a linear statistical model relating a random response Y to a set of independent variables $x_1 \cdots x_k$

$$\boxed{Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon}$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are unknown parameters, ε is a RV, and $x_1 \cdots x_k$ assume known values. Assume $E(\varepsilon) = 0$ (usually $\varepsilon \sim N(0, \sigma^2)$) and hence

$$\boxed{E(Y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}$$

Simple Linear Regression

$$Y = \beta_0 + \beta_1 x + \varepsilon \quad , \quad \varepsilon \sim N(0, \sigma^2)$$

(one covariate or one predictor, x)

Multiple Linear Regression

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon \quad , \quad \varepsilon \sim N(0, \sigma^2)$$

(k covariates)

note:

① By "linear", we mean linear in terms of the parameters

$Y = \beta_0 + \beta_1 x^2 + \beta_2 \sqrt{x} + \varepsilon$ is still a linear model.

② Linear models can still fit curved surfaces

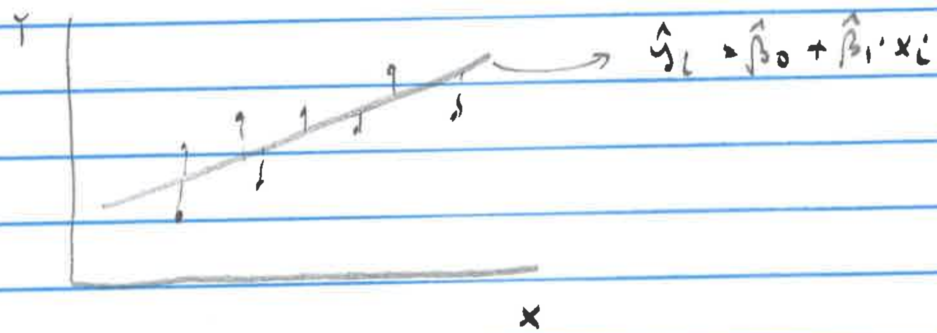
(Fig. 11.4 p.568)

11.3 Method of Least Squares

Spse we are fitting the simple linear regression model:

$Y = \beta_0 + \beta_1 x + \varepsilon$, $E(\varepsilon) = 0$

What are the "best" estimators for β_0 and β_1?



The method of least squares, minimizes the sum of the squared errors (residuals or vertical distance from the pt. to the line)

minimize $Q = SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$

$$\frac{\partial SSE}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) \overset{SET}{=} 0$$

$$\frac{\partial SSE}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) \overset{SET}{=} 0$$

① $\sum y_i - n \cdot \hat{\beta}_0 - \hat{\beta}_1 \cdot \sum x_i = 0$

② $\sum x_i y_i - \hat{\beta}_0 \cdot \sum x_i - \hat{\beta}_1 \sum x_i^2 = 0$

$$n\hat{\beta}_0 = \Sigma y_i - \hat{\beta}_1 \Sigma x_i \Rightarrow \boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}}$$

$$\Sigma x_i y_i - [\bar{y} - \hat{\beta}_1 \cdot \bar{x}] \cdot \Sigma x_i - \hat{\beta}_1 \cdot \Sigma x_i^2 = 0$$

$$\Sigma xy - \frac{(\Sigma y)(\Sigma x)}{n} + \hat{\beta}_1 \frac{(\Sigma x_i)^2}{n} - \hat{\beta}_1 (\Sigma x_i^2) = 0$$

$$\hat{\beta}_1 \left(\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n} \right) = \Sigma xy - \frac{(\Sigma y)(\Sigma x)}{n}$$

$$\boxed{\Rightarrow \hat{\beta}_1 = \frac{\Sigma xy - \frac{(\Sigma y)(\Sigma x)}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \frac{\Sigma (y-\bar{y})(x-\bar{x})}{\Sigma (x-\bar{x})^2} = \frac{S_{xy}}{S_{xx}}}$$

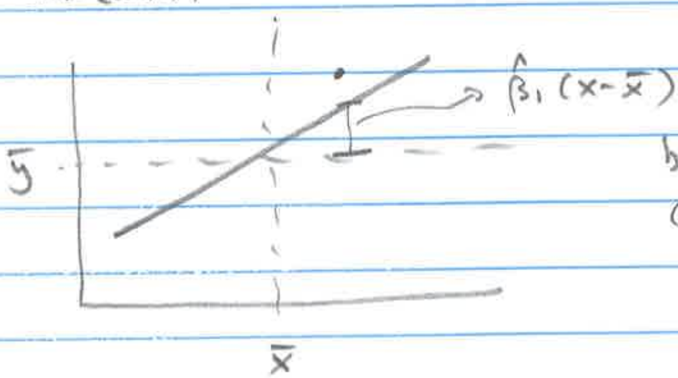
note:

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

$$\hat{y} = (\bar{y} - \hat{\beta}_1 \cdot \bar{x}) + \hat{\beta}_1 \cdot x$$

$$\boxed{\hat{y} = \bar{y} + \hat{\beta}_1 (x - \bar{x})}$$

i.e. The LS line always goes through the center of the data.



$\Rightarrow$ adjusted predicted Y up, because x above avg. (positive association) e.g. Computer Repair Svc.

x = no. of components to repair

y = length of svc (minutes)

$\boxed{\text{see excel}}$

$n = 14 \quad \Sigma X = 84 \quad \Sigma Y = 1361 \quad \Sigma X^2 = 618 \quad \Sigma Y^2 = 160{,}077 \quad \Sigma XY = 9{,}434$

$\bar{x} = 6 \quad \bar{y} = 97.2$

$$S_{xy} = 9{,}434 - \frac{(84)(1361)}{14} = 1{,}768$$

$$\Rightarrow \hat{\beta}_1 = 15.5$$

$$S_{xx} = 618 - \frac{(84)^2}{14} = 114$$

$$\Rightarrow \hat{\beta}_0 = 97.2 - 15.5(6) = 4.16$$

$$\boxed{\hat{y} = 4.16 + 15.5x}$$

$\boxed{\text{interpret}}$

or

$$\boxed{\hat{y} = 97.2 + 15.5(x - 6)}$$

11.4 Properties of LS Estimators: (SLR)

Least squares estimation gives us point estimates for Y given X. To perform statistical inference (CIs, hypothesis tests) we need to find the distribution of $\hat{\beta}_0, \hat{\beta}_1$.

In regression, we assume X is a known constant, i.e. it is not a RV.

$$\hat{\beta}_1 = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{\sum(x-\bar{x})\cdot y - \bar{y}\cdot \overbrace{\sum(x-\bar{x})}^{0}}{\sum(x-\bar{x})^2}$$

$$= \frac{\sum(x-\bar{x})\cdot y}{\sum(x-\bar{x})^2}$$

$$\Rightarrow E(\hat{\beta}_1) = E\left[\frac{\sum(x_i-\bar{x})\cdot y_i}{\sum(x_i-\bar{x})^2}\right] = \frac{1}{\sum(x_i-\bar{x})^2}\cdot\sum(x_i-\bar{x})\cdot E(Y_i)$$

assume $\varepsilon \sim N(0,\sigma^2)$ $\Rightarrow E(Y_i) = E[\beta_0+\beta_1 x_i+\varepsilon] = \beta_0+\beta_1 x_i$

$$\Rightarrow E(\hat{\beta}_1) = \frac{\sum(x_i-\bar{x})(\beta_0+\beta_1 x_i)}{S_{xx}}$$

$$= \frac{\beta_0\cdot \overbrace{\sum(x_i-\bar{x})}^{0} + \beta_1\cdot\sum(x_i-\bar{x})\cdot x_i}{S_{xx}}$$

$\sum x_i^2 - \bar{x}\sum x_i$

$= \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

$= S_{xx}$

$$= \beta_1\cdot\frac{S_{xx}}{S_{xx}} = \boxed{\beta_1} \Rightarrow \hat{\beta}_1 \text{ is unbiased}$$

②

$$V(\hat{\beta}_1) = \frac{1}{\left[\Sigma(x_i-\bar{x})^2\right]^2} \overbrace{\Sigma(x_i-\bar{x})^2 \cdot V(Y_i)}^{\text{ind.}}$$

$V(Y_i) = V(\beta_0+\beta_1 X_i+\varepsilon)$
$= V(\varepsilon) = \sigma^2$

$$= \frac{\sigma^2 \cdot \Sigma(x_i-\bar{x})^2}{\left[\Sigma(x_i-\bar{x})^2\right]^2} = \frac{\sigma^2}{\Sigma(x_i-\bar{x})^2} = \boxed{\frac{\sigma^2}{S_{xx}}}$$ (see 3710 notes!)

How about $\hat{\beta}_0$?

$V(\hat{\beta}_0) = V(\bar{Y} - \hat{\beta}_1 \cdot \bar{X})$ **note:** $\bar{X}$ is a constant in linear regression

$$= V(\bar{Y}) - 2\bar{X}\,\mathrm{cov}(\bar{Y},\hat{\beta}_1) + \bar{X}^2 \cdot \underbrace{V(\hat{\beta}_1)}_{\sigma^2/S_{xx}}$$

Find $V(\bar{Y})$ and $\mathrm{cov}(\bar{Y},\hat{\beta}_1)$

$V(\bar{Y})$:

$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$

$$\Rightarrow \bar{Y} = \frac{\Sigma Y_i}{n} = \frac{n\beta_0 + \beta_1 \Sigma x_i + \Sigma \varepsilon_i}{n} = \beta_0 + \beta_1 \cdot \bar{X} + \bar{\varepsilon}$$

$$\Rightarrow E(\bar{Y}) = \beta_0 + \beta_1 \bar{X} + E(\bar{\varepsilon}) = \boxed{\beta_0 + \beta_1 \bar{X}}$$

$$\Rightarrow V(\bar{Y}) = V(\beta_0 + \beta_1 \bar{X} + \bar{\varepsilon}) = V(\bar{\varepsilon}) = \frac{\sigma^2}{n}$$

$\mathrm{cov}(\bar{Y},\hat{\beta}_1)$:

From before ... $\hat{\beta}_1 = \dfrac{\Sigma(x_i-\bar{x}) \cdot y_i}{\Sigma(x_i-\bar{x})^2} = \Sigma c_i \cdot y_i$

$c_i = \frac{x_i - \bar{x}}{S_{xx}}$ <u>note:</u> $\sum c_i = \frac{\sum (x_i - \bar{x})}{S_{xx}} = 0$

$\Rightarrow cov(\bar{Y}, \hat{\beta}_1) = cov\left(\sum (\tfrac{1}{n}) \cdot Y_i , \sum c_i Y_i \right)$

$\frac{1}{n} \sum Cov(Y_i, c_i Y_i) + 2 \cdot \frac{1}{n} \sum\sum_{i \neq j} cov(Y_i, c_j Y_j)$

$\frac{1}{n} \sum c_i \cdot V(Y_i) + 2 \cdot \frac{1}{n} \sum\sum_{i \neq j} c_i \cdot \underbrace{cov(Y_i, Y_j)}_{0}$ (Y_i, Y_j are ind.)

$= \frac{\sigma^2}{n} \cdot \sum c_i = 0$

$\Rightarrow cov(\bar{Y}, \hat{\beta}_1) = 0$

$\Rightarrow E(\hat{\beta}_0) = E(\bar{Y} - \hat{\beta}_1 \cdot \bar{x}) = E(\bar{Y}) - \bar{x} \cdot E(\hat{\beta}_1)$

$= \beta_0 + \beta_1 \bar{x} - \bar{x} \cdot \beta_1 = \boxed{\beta_0}$

$\Rightarrow V(\hat{\beta}_0) = V(\bar{Y}) - 2\bar{x}\, cov(\bar{Y}, \hat{\beta}_1) + \bar{x}^2 \cdot V(\hat{\beta}_1)$

$= \frac{\sigma^2}{n} + \bar{x}^2 \cdot \frac{\sigma^2}{S_{xx}} = \boxed{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}$ (see 3710 notes!!)

$cov(\hat{\beta}_0, \hat{\beta}_1) = cov(\bar{Y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1)$

$= \underbrace{cov(\bar{Y}, \hat{\beta}_1)}_{0} - \bar{x}\, V(\hat{\beta}_1)$

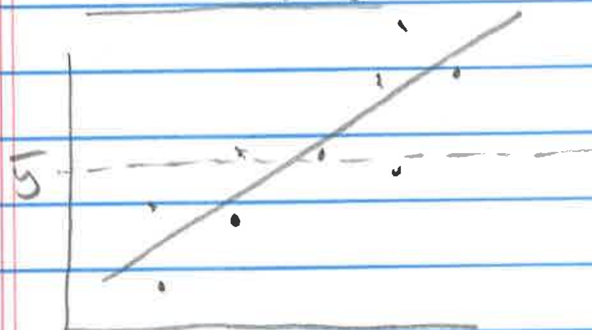
$= \boxed{\frac{-\bar{x} \cdot \sigma^2}{S_{xx}}}$

L.S. regression model: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim iid\ N(0, \sigma^2)$

So far we have: $\hat{\beta}_0$, $E(\hat{\beta}_0)$, $V(\hat{\beta}_0)$

$\hat{\beta}_1$, $E(\hat{\beta}_1)$, $V(\hat{\beta}_1)$

Now we need $\hat{\sigma}^2$



If we have no info. on X, we estimate the $E(Y) = \mu$ with $\bar{Y}$. Since $\sigma^2 = E[(Y-\mu)^2]$ it seems natural to estimate σ^2 with

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n}$$

Of course, we have shown $E(\hat{\sigma}^2) = \frac{(n-1)}{n} \cdot \sigma^2$ ($\hat{\sigma}^2$ biased)

Therefore, we use $s^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1}$

In regression, we have info regarding X as well. Our model states that the mean of Y is conditional on X.

$$E(Y|x) = \beta_0 + \beta_1 x$$

σ^2 measures the variability about the conditional mean of Y.

⑤

Therefore, we replace $\bar{Y}$ with $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$

$\hat{\sigma}^2 \overset{?}{=} \frac{\sum(Y_i - \hat{Y}_i)^2}{n}$. Let's find the $E(\hat{\sigma}^2)$

$$\sum(Y_i - \hat{Y}_i)^2 = \sum(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$= \sum(Y_i - (\bar{Y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i)^2$$

$$= \sum([Y_i - \bar{Y}] - \hat{\beta}_1 (x_i - \bar{x}))^2$$

$$= \sum(Y_i - \bar{Y})^2 - 2\hat{\beta}_1 \sum(Y_i - \bar{Y})(x_i - \bar{x}) + \hat{\beta}_1^2 \sum(x_i - \bar{x})^2$$

$$= \sum(Y_i - \bar{Y})^2 - 2 \cdot \hat{\beta}_1^2 \cdot S_{xx} + \hat{\beta}_1^2 \cdot S_{xx}$$

$$= \sum Y_i^2 - n\bar{Y}^2 - \hat{\beta}_1^2 \cdot S_{xx}$$

$$\Rightarrow E\left[\sum(Y_i - \hat{Y}_i)^2\right] = E\left[\sum Y_i^2 - n\bar{Y}^2 - \hat{\beta}_1^2 \cdot S_{xx}\right]$$

$$= \sum E(Y_i^2) - nE(\bar{Y}^2) - S_{xx} \cdot E(\hat{\beta}_1^2)$$

$$= \sum\left[\sigma^2 + (\beta_0 + \beta_1 x_i)^2\right] - n \cdot \left[\frac{\sigma^2}{n} + (\beta_0 + \beta_1 \bar{x})^2\right] - S_{xx}\left(\frac{\sigma^2}{S_{xx}} + \beta_1^2\right)$$

$$= n\sigma^2 + \sum(\beta_0 + \beta_1 x_i)^2 - \sigma^2 - n(\beta_0 + \beta_1 \bar{x})^2 - \sigma^2 - S_{xx} \cdot \beta_1^2$$

$$= (n-2)\sigma^2 + \sum\left(\beta_0^2 + 2\beta_0\beta_1 x_i + \beta_1^2 x_i^2\right) - n\left(\beta_0^2 + 2\beta_0\beta_1\bar{x} + \beta_1^2\bar{x}^2\right)$$

$$- \sum(x_i - \bar{x})^2 \cdot \beta_1^2$$

6

$$(n-2)\sigma^2 + \cancel{n\beta_0^2} + \cancel{2\beta_0\beta_1 n\bar{x}} + \beta_1^2 \sum x_i^2$$
$$- \cancel{n\beta_0^2} - \cancel{2\beta_0\beta_1 n\bar{x}} - n\beta_1^2\bar{x}^2$$
$$- \left[\sum x_i^2 - n\bar{x}^2\right]\cdot\beta_1^2$$

$$= (n-2)\sigma^2 + \cancel{\beta_1^2 \sum x_i^2} - \cancel{n\beta_1^2\bar{x}^2}$$
$$- \cancel{\beta_1^2 \sum x_i^2} + \cancel{n\beta_1^2\bar{x}^2}$$

$$= (n-2)\sigma^2$$

$$\Rightarrow \text{Let } \boxed{\hat{\sigma}^2 = \frac{\sum(y_i - \hat{y})^2}{n-2}} = \frac{SSE}{n-2} \quad \Rightarrow E(\hat{\sigma}^2) = \sigma^2$$

note: $n-2$ = (sample size) − (no. of estimated parameters)

Back to Computer Repair Data:

$$SSE = S_{yy} - \hat{\beta}_1 \cdot S_{xy} \quad (\text{Ex. } 11.15)$$

$$= \underbrace{\left(160{,}077 - \frac{1361^2}{14}\right)}_{27{,}768.36} - 15.5(1{,}768) = 364.36$$

$$\Rightarrow \hat{\sigma}^2 = \frac{364.36}{14-2} = 30.363$$

$$\hat{\beta}_1 = 15.5 \quad \hat{V}(\hat{\beta}_1) = \frac{30.363}{114} = .2663 \quad \hat{SD}(\hat{\beta}_1) = .516$$

$$\hat{\beta}_0 = 4.16 \quad \hat{V}(\hat{\beta}_0) = 30.363\left(\frac{1}{14} + \frac{6^2}{114}\right) = 11.757 \quad \hat{SD}(\hat{\beta}_0) = 3.43$$

We now have unbiased estimators of β_0 and β_1, and we have variances for these estimators as well. In order to perform inference using these estimators, we need to know the dist. of $\hat{\beta}_0$ and $\hat{\beta}_1$ too.

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0, \sigma^2)$$

<u>note</u>: Generalized Linear Models allow other distributions for errors (residuals)

If $\varepsilon_i \sim N(0,\sigma^2)$ then $Y_i \sim N(\mu = \beta_0 + \beta_1 x_i, \sigma^2)$

From before,

$$\hat{\beta}_1 = \frac{\Sigma(x_i - \bar{x}) \cdot Y_i}{\Sigma(x_i - \bar{x})^2} = \sum_{i=1}^{n} a_i \cdot Y_i \quad , \quad a_i = \frac{x_i - \bar{x}}{S_{xx}}$$

Since $\hat{\beta}_1$ is a <u>linear combination</u> of <u>normal RVs</u>, it also has a normal dist.

$$\Rightarrow \boxed{\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \cdot \bar{x} = \frac{\sum_{i=1}^{n} Y_i}{n} - \left(\sum_{i=1}^{n} a_i \cdot Y_i\right) \cdot \bar{x} \quad , \quad a_i = \frac{x_i - \bar{x}}{S_{xx}}$$

$$= \sum_{i=1}^{n} c_i \cdot Y_i \quad , \quad c_i = \frac{1}{n} + \frac{(x_i - \bar{x})}{S_{xx}} \cdot \bar{x}$$

i.e. $\hat{\beta}_0$ is also a linear combination of normal RVs

$$\Rightarrow \boxed{\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)}$$

11.21 Find $cov(\hat{\beta}_0, \hat{\beta}_1)$

$$= cov(\bar{Y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1)$$

$$= cov(\bar{Y}, \hat{\beta}_1) - \bar{x}\, var(\hat{\beta}_1)$$

from before

$$= 0 - \frac{\bar{x} \cdot \sigma^2}{S_{xx}} = \frac{-\bar{x} \cdot \sigma^2}{S_{xx}}$$

note: If $\Sigma x_i = 0$ then $cov(\hat{\beta}_0, \hat{\beta}_1) = 0$

Find the MLE estimators for $\beta_0, \beta_1, \sigma^2$

Let $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim iid\ N(0, \sigma^2)$

$\Rightarrow Y_i \sim N(\mu = \beta_0 + \beta_1 x_i, \sigma^2)$

$$\Rightarrow f(y_i) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2}$$

$$\Rightarrow L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2}$$

$$= \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\Sigma(y_i - (\beta_0 + \beta_1 x_i))^2}$$

$$\ln L = -n/2 \ln(2\pi) - n \ln\sigma - \frac{1}{2\sigma^2}\Sigma(y_i - (\beta_0 + \beta_1 x_i))^2$$

note: for σ^2, find the MLE for σ ($\hat{\sigma}$) and then use the invariance property of MLEs to find $\hat{\sigma}^2 = (\hat{\sigma})^2$

① $\frac{\partial \ln L}{\partial \beta_0} = +\frac{1}{\sigma^2}\Sigma(y_i - \beta_0 - \beta_1 x_i) \overset{SET}{=} 0$

② $\frac{\partial \ln L}{\partial \beta_1} = \frac{1}{\sigma^2}\Sigma(y_i - \beta_0 - \beta_1 x_i)\cdot x_i \overset{SET}{=} 0$

③ $\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3}\Sigma(y_i - \beta_0 - \beta_1 x_i)^2 \overset{SET}{=} 0$

① and ② can be rewritten as

$$\left.\begin{array}{l} \Sigma y_i - n\beta_0 - \beta_1 \Sigma x_i = 0 \\ \Sigma x_i y_i - \beta_0 \Sigma x_i - \beta_1 \Sigma x_i^2 = 0 \end{array}\right\}$$ normal eqns we developed finding the least squares estimators for β_0, β_1

$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

i.e. the MLEs for β_0, β_1 are the same as the LS estimators

$\Rightarrow$ MLEs for β_0, β_1 are unbiased

What about σ^2? substitute $\hat{\beta}_0$ and $\hat{\beta}_1$ into Eqn ③ for β_0 and β_1 ...

$$-\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3}\Sigma(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 = 0$$

L.S. estimate too

$$-\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3} \sum (Y_i - \hat{Y}_i)^2 = 0$$

$$\Rightarrow \frac{1}{\hat{\sigma}^3} \sum (Y_i - \hat{Y})^2 = \frac{n}{\hat{\sigma}} \Rightarrow \hat{\sigma}^2 = \frac{\sum (Y_i - \hat{Y})^2}{n}$$

$$\Rightarrow \boxed{\hat{\sigma}^2 = \frac{SSE}{n}}$$

note: the MLE for σ^2 is biased $\left(E(\hat{\sigma}^2) = \frac{n \cdot \sigma^2}{n-2} \right)$

This often happens, but the bias is negligible for large n (asymptotically unbiased)

11.5 Inferences Concerning β_i

Result:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim iid\ N(0, \sigma^2)$$

It can be shown...

1. $$\frac{(n-2)\cdot s^2}{\sigma^2} \sim \chi^2_{(n-2)}, \quad s^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}$$

2. s^2 is independent of $\hat{\beta}_0, \hat{\beta}_1$

Combining this result w/ results from the last section:

- $\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$
- $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$

We can form T-distributions to perform inference on β_0, β_1.

$\hat{\beta}_0$:

$$z = \frac{\hat{\beta}_0 - \beta_0}{\sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim N(0,1) \qquad \frac{(n-2)s^2}{\sigma^2} \sim \chi^2_{(n-2)}$$

$$\Rightarrow T = \frac{\dfrac{\hat{\beta}_0 - \beta_0}{\sigma\cdot\sqrt{\frac{1}{n} + \bar{x}^2/S_{xx}}}}{\sqrt{\dfrac{(n-2)s^2}{\sigma^2}\Big/(n-2)}} = \frac{\dfrac{\hat{\beta}_0 - \beta_0}{\sigma\cdot\sqrt{\frac{1}{n} + \bar{x}^2/S_{xx}}}}{s/\sigma} =$$

$$T = \frac{\hat{\beta}_0 - \beta_0}{\sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \cdot \frac{\sigma}{s} = \boxed{\frac{\hat{\beta}_0 - \beta_0}{s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim t_{(n-2)}}$$

__note:__ Just as w/ univariate stats, we replace σ w/ s and use a T-dist.

__note:__ If d.f > 30 then we can basically use a z

__$\hat{\beta}_1$:__

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{S_{xx}}} \sim N(0,1)$$

$$\Rightarrow T = \frac{\frac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{S_{xx}}}}{\sqrt{\frac{(n-2) \cdot s^2}{\sigma^2} / (n-2)}} = \frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{S_{xx}}} \cdot \sigma/s = \boxed{\frac{\hat{\beta}_1 - \beta_1}{s/\sqrt{S_{xx}}} \sim t_{(n-2)}}$$

In general, the primary interest in regression is β_1. The intercept, β_0, often does not have a valid interpretation.

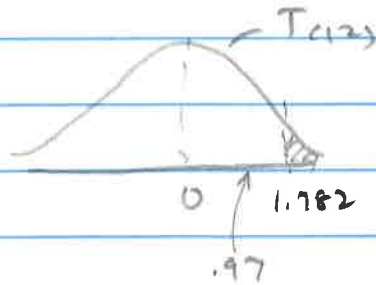
For Computer Repair Data . . .

$$\hat{y} = 4.16 + 15.5x$$

Test $H_0: \beta_1 = 15$
$H_a: \beta_1 > 15$

$$T = \frac{15.5 - 15}{\sqrt{30.363}/\sqrt{114}} = .97$$

$t_{.05} = 1.782$ (d.f. = 12)



don't reject Ho

Not enough evidence (at $\alpha = .05$) to conclude that avg. repair time goes up by at least 15 mins, for every additional part to be repaired.

Find a 95% lower CI for β_1:

$$\hat{\beta}_1 - t_{.05} \cdot \hat{\sigma}/S_{xx}$$

$$15.5 - 1.782\sqrt{\frac{30.363}{114}} = 14.58 \quad \text{i.e.} \quad \boxed{C(\beta_1 > 14.58) = .95}$$

(11.28) Show under $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim iid\ N(0, \sigma^2)$ that the likelihood ratio test for $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$ is equivalent to the T-test.

$$\Omega_0 = \{\beta_0, \sigma^2, \beta_1 = 0\}$$

$$L(\Omega_0) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(y_i - \beta_0)^2}$$

$$= \frac{1}{(2\pi)^{n/2} \cdot \sigma^n} e^{-\frac{1}{2\sigma^2}\Sigma(y_i - \beta_0)^2}$$

Now find MLEs under Ω_0:

$$\ln L(\Omega_0) = -n/_2 \ln(2\pi) - n \ln\sigma - \frac{1}{2\sigma^2}\Sigma(y_i - \beta_0)^2$$

(1) $$\frac{\partial \ln L}{\partial \beta_0} = +\frac{1}{\sigma^2}\Sigma(y_i - \beta_0) \overset{SET}{=} 0$$

(2) $\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{2}{2\sigma^3} \Sigma (y_i - \beta_0)^2 \overset{SET}{=} 0$

Eqn (1) yields $\Sigma(y_i - \hat{\beta}_0) = 0 \Rightarrow \boxed{\hat{\beta}_0 = \bar{y}}$

substituting $\hat{\beta}_0 = \bar{y}$ into (2) ...

$$\frac{-n}{\sigma} + \frac{1}{\sigma^3} \Sigma (y_i - \bar{y})^2 = 0$$

$$\boxed{\frac{\Sigma (y_i - \bar{y})^2}{n} = \hat{\sigma}_0^2}$$

from before, the MLEs from the unrestricted space are:

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

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

$\hat{\sigma}^2 = \frac{\Sigma (y_i - \hat{y}_i)^2}{n}$, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

$$\lambda = \frac{L(\Omega_0)}{L(\Omega)} = \frac{\frac{1}{\cancel{(2\pi)^{n/2}} \cdot \hat{\sigma}_0^n} e^{-\frac{1}{2\hat{\sigma}_0^2} \Sigma (y_i - \bar{y})^2}}{\frac{1}{\cancel{(2\pi)^{n/2}} \cdot \hat{\sigma}^n} e^{-\frac{1}{2\hat{\sigma}^2} \Sigma (y_i - \hat{y}_i)^2}}$$

$$\lambda = \left(\frac{\hat{\sigma}}{\hat{\sigma}_0}\right)^n \cdot \frac{e^{-\frac{n}{2} \cdot \frac{\Sigma (y_i - \bar{y})^2}{\Sigma (y_i - \bar{y})^2}}}{e^{-\frac{n}{2} \cdot \frac{\Sigma (y_i - \hat{y}_i)^2}{\Sigma (y_i - \hat{y}_i)^2}}} = \left(\frac{\hat{\sigma}}{\hat{\sigma}_0}\right)^n$$

$$\lambda = \left(\frac{\sqrt{\frac{\sum (y_i - \hat{y})^2}{n}}}{\sqrt{\frac{\sum (y_i - \bar{y})^2}{n}}} \right)^n \Rightarrow \lambda = \left(\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \right)^{n/2}$$

$$\Rightarrow \lambda = \left(\frac{SSE}{S_{yy}} \right)^{n/2} \quad \text{reject } H_0 \text{ if } \left(\frac{SSE}{S_{yy}} \right)^{n/2} \leq k$$

How is this equivalent to T-test?

$$\left(\frac{S_{yy}}{SSE} \right)^{n/2} \geq \frac{1}{k} \Rightarrow \frac{S_{yy}}{SSE} \geq \left(\frac{1}{k} \right)^{n/2} = k^*$$

$$\Rightarrow \frac{SSE + \hat{\beta}_1 \cdot S_{xy}}{SSE} \geq k^* \Rightarrow 1 + \frac{\hat{\beta}_1 \cdot S_{xy}}{SSE} \geq k^*$$

under $H_0: \beta_1 = 0$

$$T = \frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{S_{xx}}} = \frac{\hat{\beta}_1 \cdot \sqrt{S_{xx}}}{\sqrt{\frac{SSE}{n-2}}} \Rightarrow T^2 = \frac{(n-2) \cdot \hat{\beta}_1^2 \cdot S_{xx}}{SSE}$$

$$\lambda = 1 + \frac{\hat{\beta}_1 \cdot S_{xy}}{SSE} = 1 + \frac{\hat{\beta}_1 (\hat{\beta}_1 \cdot S_{xx})}{SSE} = 1 + \frac{\hat{\beta}_1^2 \cdot S_{xx}}{SSE} = 1 + \frac{T^2}{n-2} \geq k^*$$

i.e. Reject H_0 if T^2 get large, which is equivalent to reject H_0 if $|T|$ gets large.

11.6 Inferences on fcts of β_0 and β_1

Spse we would like to estimate $E(Y) = \beta_0 + \beta_1 x$. The LS regression line will estimate $E(Y)$, but if we want CIs or tests concerning $E(Y)$, we need to look at the properties of our estimator.

For the general case, let $\Theta = a_0 \cdot \beta_0 + a_1 \cdot \beta_1$ (linear comb. of β_0, β_1) note: usually $a_0 = 1$, $a_1 = x^*$

$$\hat{\Theta} = a_0 \cdot \hat{\beta}_0 + a_1 \hat{\beta}_1$$

$$E(\hat{\Theta}) = E[a_0 \cdot \hat{\beta}_0 + a_1 \hat{\beta}_1] = a_0 E(\hat{\beta}_0) + a_1 E(\hat{\beta}_1) = a_0 \beta_0 + a_1 \beta_1 = \Theta$$

(unbiased)

$$V(\hat{\Theta}) = V[a_0 \cdot \hat{\beta}_0 + a_1 \hat{\beta}_1] = a_0^2 V(\hat{\beta}_0) + a_1^2 V(\hat{\beta}_1) + 2a_0 a_1 Cov(\hat{\beta}_0, \hat{\beta}_1)$$

$$= a_0^2 \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) + a_1^2 \cdot \frac{\sigma^2}{S_{xx}} + 2a_0 a_1 \sigma^2 \left(\frac{-\bar{x}}{S_{xx}}\right)$$

$$= \sigma^2 \left(\frac{a_0^2 \cdot S_{xx} + n\bar{x}^2 a_0^2 + n a_1^2 + 2a_0 a_1 \cdot n\bar{x}}{n \cdot S_{xx}}\right)$$

$$= \sigma^2 \left(\frac{a_0^2 \left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right) + \frac{(\Sigma x)^2}{n} \cdot a_0^2 + n a_1^2 - 2n a_0 a_1 \bar{x}}{n \cdot S_{xx}}\right)$$

$$= \sigma^2 \left(\frac{a_0^2 \cdot \Sigma x^2 + n a_1^2 - 2n a_0 a_1 \bar{x}}{n \cdot S_{xx}}\right)$$

$$= \boxed{\sigma^2 \left(\frac{a_0^2 \frac{\Sigma x^2}{n} + a_1^2 - 2a_0 a_1 \bar{x}}{S_{xx}}\right)}$$

Recall: $Y = \beta_0 + \beta_1 x + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$

What is the dist of $\hat{\theta}$? $\hat{\theta}$ is a linear combination of normal RVs so it is also normally dist.

$$\hat{\theta} \sim N\left(\mu = a_0\beta_0 + a_1\beta_1 \,,\ var = \sigma^2 \left(\frac{a_0^2 \frac{\Sigma x^2}{n} + a_1^2 - 2a_0 a_1 \bar{x}}{S_{xx}} \right)\right)$$

$$\Rightarrow z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \sim N(0,1)$$

In practice, we won't know $\sigma_{\hat{\theta}}$ because we don't know σ^2

From before,

$$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2_{(n-2)}$$

$$\Rightarrow T = \frac{\frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}}{\sqrt{\frac{(n-2)s^2}{\sigma^2} \Big/ (n-2)}} \sim t_{(n-2)}$$

$$T = \frac{\hat{\theta} - \theta}{\sigma \sqrt{\frac{a_0^2 \frac{\Sigma x^2}{n} + a_1^2 - 2a_0 a_1 \bar{x}}{S_{xx}}}} \cdot \frac{\sigma}{s}$$

$$\boxed{T = \frac{\hat{\theta} - \theta}{s \sqrt{\frac{a_0^2 \frac{\Sigma x^2}{n} + a_1^2 - 2a_0 a_1 \bar{x}}{S_{xx}}}}}$$

③

This result can be used to test:

$H_0: \Theta = \Theta_0$ vs. $H_1: \Theta \underset{>}{\overset{\neq}{<}} \Theta_0$

And develop $100(1-\alpha)\%$ CI for $\Theta = a_0\beta_0 + a_1\beta_1$:

$$(a_0 \cdot \hat{\beta}_0 + a_1\hat{\beta}_1) \pm \underset{(d.f.=n-2)}{t_{\alpha/2}} \cdot s \cdot \sqrt{\frac{a_0^2 \frac{\Sigma x^2}{n} + a_1^2 - 2a_0a_1\bar{x}}{S_{xx}}}$$

Back to $E(Y) = \beta_0 + \beta_1 X^* = \Theta$

$\Rightarrow a_0 = 1 \quad a_1 = x$

$$\frac{a_0^2 \cdot \frac{\Sigma x^2}{n} + a_1^2 - 2a_0a_1\bar{x}}{S_{xx}} = \frac{\frac{\Sigma x^2}{n} + x^2 - 2x\cdot\bar{x}}{S_{xx}}$$

$$= \frac{\frac{\Sigma x^2}{n} + (x-\bar{x})^2 - \bar{x}^2}{S_{xx}} = \frac{\frac{1}{n}\overbrace{(\Sigma x^2 - n\bar{x}^2)}^{S_{xx}} + (x-\bar{x})^2}{S_{xx}}$$

$$= \boxed{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$$

$\Rightarrow 100(1-\alpha)\%$ CI for $E(Y) = \beta_0 + \beta_1 x$ is:

$$\boxed{\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{\alpha/2} \cdot s\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}}$$ → Check Graybill book

note: $\frac{(x-\bar{x})^2}{S_{xx}}$ term gives us narrower CIs near the center of the data.

Back to Computer Repair Data:

$n = 14$ $S_{xx} = 114$ $S_{xy} = 1{,}768$ $\bar{x} = 6$ $\bar{y} = 97.2$

$\hat{\beta}_0 = 4.16$ $\hat{\beta}_1 = 15.5$ $s^2 = 30.363$

Estimate the avg. repair time w/ 95% confidence if 9 parts need to be repaired

$\hat{y} = 4.16 + 9(15.5) = 143.66$ $t_{12} = 2.179$

$$143.66 \pm 2.179\sqrt{30.363}\sqrt{\frac{1}{14} + \frac{(9-6)^2}{114}}$$

$143.66 \pm \underbrace{4.66}_{E} \Rightarrow \boxed{(139,\ 148.32)}$ on avg

Estimate the avg. repair time w/ 95% confidence if 6 parts need to be repaired.

$\hat{y} = 4.16 + 6(15.5) = 97.16$ (note: $\bar{y}$ is the prediction at $x = \bar{x}$)

$$97.2 \pm 2.179\sqrt{30.363}\sqrt{\frac{1}{14} + \cancel{\frac{(6-6)^2}{114}} \to 0}$$

$97.2 \pm \underbrace{3.21}_{E} \Rightarrow \boxed{(93.99,\ 100.41)}$

note: narrower CI for $E(y)$ at $x = \bar{x}$

note: let $a_0 = 0 \quad a_1 = 1 \quad \Rightarrow \theta = \beta_1$

$$V(\hat{\theta}) = V(\hat{\beta}_1) = \sigma^2 \left(\frac{0 \cdot \frac{\Sigma x^2}{n} + 1 - 2(0)(1)\bar{x}}{S_{xx}} \right)$$

$$= \boxed{\frac{\sigma^2}{S_{xx}}}$$ (as before)

For β_0, let $a_0 = 1 \quad a_1 = 0$

$$\Rightarrow V(\theta) = V(\hat{\beta}_0) = \sigma^2 \left(\frac{\frac{\Sigma x^2}{n} + 0^2 - 2(1)(0)\bar{x}}{S_{xx}} \right)$$

$$= \sigma^2 \left(\frac{\frac{\Sigma x^2}{n}}{S_{xx}} \right) = \sigma^2 \left(\frac{\frac{1}{n} \overbrace{(\Sigma x^2 - n\bar{x}^2)}^{S_{xx}} + \bar{x}^2}{S_{xx}} \right)$$

$$= \boxed{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}$$

11.7 Predicting a future Observation

Spse a repairmen has just shown up for a job and finds out there are 9 parts to be repaired. How long can we expect it to take?

Note that in this case we are not interested in the avg. time of repair for all service calls with 9 parts to be repaired. We have one call to predict and it's a RV. not a parameter

For our model, $Y = \beta_0 + \beta_1 x + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$

$\Rightarrow Y \sim N(\mu = \beta_0 + \beta_1 x, \sigma^2)$

Spse we want to predict Y^* when $x = x^*$

$\Rightarrow Y^* \sim N(\mu = \beta_0 + \beta_1 x^*, \sigma^2)$

a good estimator $\boxed{\hat{Y}^* = \hat{\beta}_0 + \hat{\beta}_1 x^*}$ which is the estimate for the center of the dist. of Y^*

note: This is also the same estimator for $E(Y)$ when $x = x^*$

What kind of error do we have in our estimate?

$\boxed{\text{Error} = Y^* - \hat{Y}^*}$ what are the properties of this error?

$$E(\text{Error}) = E[Y^* - \hat{Y}^*] = E(Y^*) - E(\hat{Y}^*)$$

$$= (\beta_0 + \beta_1 x^*) - E(\hat{\beta}_0) - x^* \cdot E(\hat{\beta}_1)$$

$$= \beta_0 + \beta_1 x^* - (\beta_0 + \beta_1 x^*) = 0 \quad \underline{\text{unbiased}}$$

$$V(\text{Error}) = V(Y^* - \hat{Y}^*) = V(Y^*) + V(\hat{Y}^*) - 2\,\mathrm{cov}(Y^*, \hat{Y}^*)$$

Since Y^* is a future obs. in this context and $\hat{Y}^*$ is an estimator developed from prior data, we assume Y^* and $\hat{Y}^*$ are independent

$\Rightarrow cov(Y^*, \hat{Y}^*) = 0$

$$\Rightarrow V(\text{error}) = V(Y^*) + V(\hat{Y}^*) = \sigma^2 + V(\hat{\beta}_0 + \hat{\beta}_1 x^*)$$

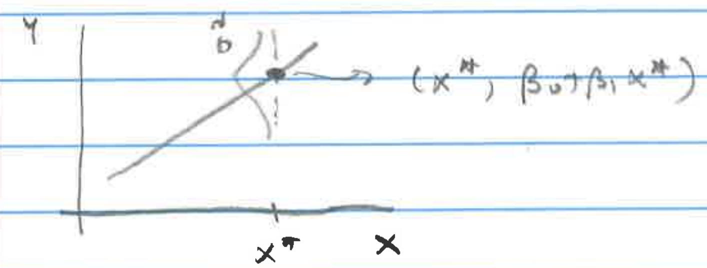
$$= \sigma^2 + \sigma^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right)$$

$$= \boxed{\sigma^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right)}$$ (see Graybill book)

Look at what we have:

V(error) = V(estimating the center)

+ V(observations about the center)



Since Y^* and $\hat{Y}^*$ are both normally dist., it means that $E = Y^* - \hat{Y}^*$ is also normally dist.

$$\Rightarrow E \sim N\left(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right)\right)$$

$$\Rightarrow z = \frac{Y^* - \hat{Y}^*}{\sigma\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}} \sim N(0,1)$$