* Maximum Likelihood Estimators of normal Error regression model

Recall : Likelihood function:

Likelihood function is the joint pdf, when we consider it as a function of parameters.

Eg:- Let $X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then the likelihood function is

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{\sum (x_i-\mu)^2}{2\sigma^2}}.$$

Defn MLE

Values of the parameters which maximize the likelihood function (ie $L(\mu,\sigma^2)$) are called maximum likelihood estimators (ie MLEs).

$$\text{ie } (\hat{\mu}_{MLE}, \hat{\sigma}^2_{MLE}) = \underline{\arg}\max \left(L(\mu,\sigma^2) \right)$$

Note:
The likelihood function and it's log function have their maximums at same value of the parameters.

Since $Y_i \overset{\text{inde}}{\sim} N(\beta_0+\beta_1 X_i, \sigma^2)$, $i=1,2,\dots n$.

the likelihood function:

$$L(\beta_0,\beta_1,\sigma^2) = \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}} e^{-\frac{1}{2\sigma^2}(Y_i-(\beta_0+\beta_1X_i))^2} \quad (\because Y_i\text{s are independent}).$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \operatorname{Exp}\left[-\frac{1}{2\sigma^2}\Sigma(Y_i-\beta_0-\beta_1X_i)^2\right]$$

$$\Rightarrow \ln[L(\beta_0,\beta_1,\sigma^2)] = -\tfrac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\Sigma(Y_i-\beta_0-\beta_1X_i)^2$$

$$\frac{\partial}{\partial\beta_0}[\ln[L(\beta_0,\beta_1,\sigma^2)]] = -\frac{1}{2\sigma^2}2\Sigma(Y_i-\beta_0-\beta_1X_i)(-1) \overset{\text{set}}{=} 0 \longrightarrow 1$$

$$\frac{\partial}{\partial\beta_1}[\ln[L(\beta_0,\beta_1,\sigma^2)]] = -\frac{1}{2\sigma^2}2\Sigma(Y_i-\beta_0-\beta_1X_i)(-X_i) \overset{\text{set}}{=} 0 \longrightarrow 2$$

$$\frac{\partial}{\partial\sigma^2}[\ln[L(\beta_0,\beta_1,\sigma^2)]] = -\frac{n}{2}\cdot\frac{1\cdot(2\pi)}{2\pi\sigma^2} + \frac{\Sigma(Y_i-\beta_0-\beta_1X_i)^2}{2(\sigma^2)^2} \overset{\text{set}}{=} 0 \longrightarrow 3$$

Here note that 1 and 2 are same as the normal equations of the least square method. So maximum likelihood estimators for β_0 and β_1 are same as the least square estimators.

i.e $\hat{\beta}_{1_{MLE}} = b_1 = \dfrac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2}$ and,

$$\hat{\beta}_{0_{MLE}} = b_0 = \bar{Y} - b_1\bar{X}.$$

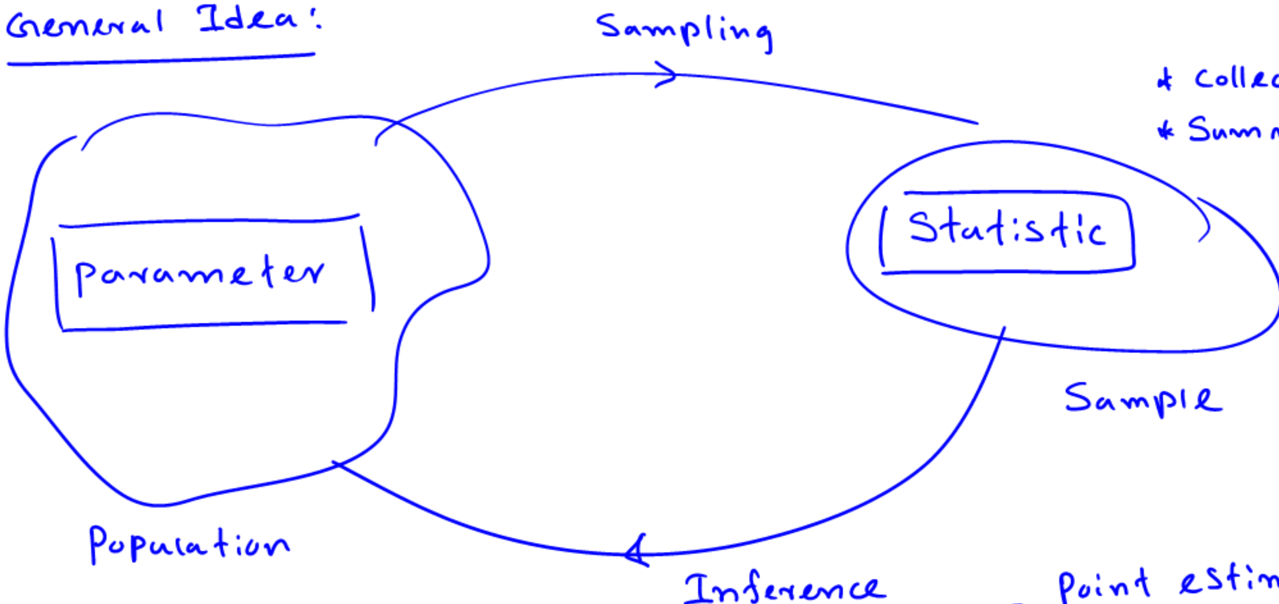
By 3,

$$\text{MLE of } \sigma^2 = \hat{\sigma}^2_{MLE} = \frac{\Sigma(Y_i - \overbrace{b_0-b_1X_i}^{\hat{Y}_i})^2}{n} = \frac{\Sigma(Y_i-\hat{Y}_i)^2}{n} = \frac{\Sigma e_i^2}{n}$$

So MLE of σ^2 is different than the least squre estimator of σ^2 (i.e $\frac{\Sigma e_i^2}{n-2}$).

Chapter - 2 : Inferences in Regression Analysis

General Idea:



* Collecting data
* Summarizing data.

1. Estimation → Point estimation; Interval estimation
2. Testing
3. Predicting.

Inference concerning β_1

Point estimator of $\beta_1 = b_1 = \dfrac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum (x_i-\bar{x})^2}$

Note:

$$b_1 = \sum k_i y_i \,, \quad \text{where} \quad k_i = \frac{x_i-\bar{x}}{\sum (x_i-\bar{x})^2}$$

$\bar{x} = \frac{\sum x_i}{n}$

$\sum x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0$

Proof:

$$b_1 = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum (x_i-\bar{x})^2} = \sum \underbrace{\frac{(x_i-\bar{x})}{\sum (x_i-\bar{x})^2}}_{k_i} y_i - \bar{y}\,\underbrace{\frac{\sum (x_i-\bar{x})}{\sum (x_i-\bar{x})^2}}_{=0}$$

$$= \sum k_i y_i \,, \quad \text{where} \quad k_i = \frac{(x_i-\bar{x})}{\sum (x_i-\bar{x})^2}.$$

* **Properties of k_i**

1) $\sum k_i = 0$

Proof:

$$\sum k_i = \sum\left[\frac{(X_i-\bar{X})}{\sum(X_i-\bar{X})^2}\right] = \frac{n\bar{X}-n\bar{X}}{\sum(X_i-\bar{X})^2} = 0.$$

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

Proof - HW.

3) $\sum k_i^2 = \dfrac{1}{\sum(X_i-\bar{X})^2}$

Proof - HW

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum a_i E(X_i)$$

$$E(aX+b) = aE(X)+b$$

* $E(b_1) = E\left(\sum k_i Y_i\right)$

$$= \sum k_i E(Y_i)$$

$$= \sum k_i(\beta_0+\beta_1 X_i)$$

$$= \beta_0 \underbrace{\sum k_i}_{=0} + \beta_1 \underbrace{\sum k_i X_i}_{=1}$$

$$= \beta_1.$$

ie b_1 is unbiased estimator for β_1

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \, var(X_i)$$

if $X_1, X_2, \ldots X_n$ are indepen...

$$Var(aX+b) = a^2 var(X)$$

$$Var(b_1) = \sigma^2\{b_1\} = var\left(\sum k_i Y_i\right)$$

$$= \sum k_i^2 \underbrace{var(Y_i)} \quad (\because Y_i\text{s are independent})$$

$$= \sum k_i^2 (\sigma^2)$$

$$= \sigma^2 \sum k_i^2$$

$$= \frac{\sigma^2}{\sum(X_i-\bar{X})^2}.$$

unknown.

* when σ^2 is unknown, the estimated variance of b_1 is

$$\widehat{\mathrm{Var}}(b_1) = \hat{\sigma}^2\{b_1\} = S^2\{b_1\} = \frac{MSE}{\sum (X_i-\bar{X})^2}$$

<u>Sampling distribution</u>

Since b_1 is a linear combination of independent Y_i and $Y_i \overset{ind}{\sim} N(\beta_0+\beta_1 X_i, \sigma^2)$,

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum (X_i-\bar{X})^2}\right).$$

*When σ^2 is known.

pivot

$$\underbrace{\frac{b_1-\beta_1}{\sigma\{b_1\}}}_{\text{Standardize Statistic}} \sim N(0,1) \leftarrow \text{Standard normal distribution.}$$

* when σ^2 is unknown

$$\underbrace{\frac{b_1-\beta_1}{S\{b_1\}}}_{\text{Studentized Statistic}} \sim t_{n-2} \quad (t \text{ distribution with degrees of freedom } n-2),$$

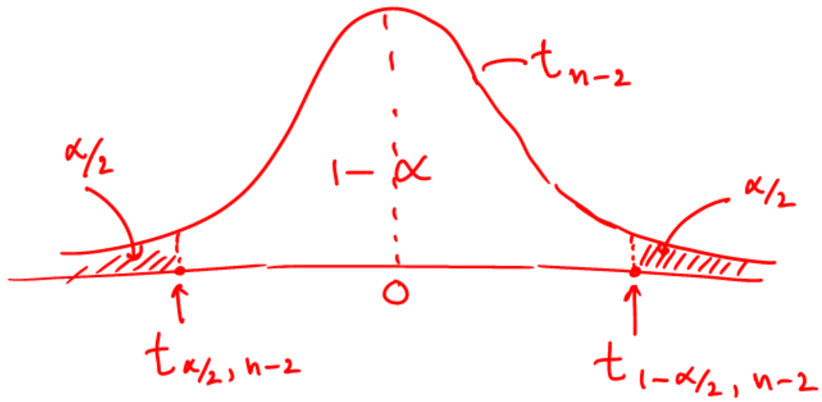
where $S\{b\} = \sqrt{\dfrac{MSE}{\sum (X_i-\bar{X})^2}}$.

<u>Recall:</u>

If $P(X_1 \le \mu \le X_2) = 95\%$, then $[X_1, X_2]$ is called 95% confidence interval.

<u>idea:</u>

In a long run, the interval contains the parameter of interest at least 95% of the times.



* $t_{1-\alpha/2, n-2} = -t_{\alpha/2, n-2}$ ($\because$ Symmetric)

So $P\left(t_{\alpha/2, n-2} \le \dfrac{b_1-\beta_1}{S\{b_1\}} \le t_{1-\alpha/2, n-2}\right) = 1-\alpha$

$\Rightarrow P\left(t_{\alpha/2, n-2}\, S\{b_1\} - b_1 \le -\beta_1 \le -b_1 + t_{1-\alpha/2, n-2}\, S\{b_1\}\right) = 1-\alpha$

$\Rightarrow P\left(b_1 - t_{1-\alpha/2 : n-2}\, S\{b_1\} \le \beta_1 \le b_1 - t_{\alpha/2 : n-2}\, S\{b_1\}\right) = 1-\alpha$

$$P\Big(\underbrace{b_1 - t_{1-\alpha/2 : n-2}\, S\{b_1\}}_{X_1} \le \beta_1 \le \underbrace{b_1 + t_{1-\alpha/2 : n-2}\, S\{b_1\}}_{X_2}\Big) = 1-\alpha$$

$\therefore$ $100(1-\alpha)\%$ confidence interval for β_1:

$b_1 \pm t_{1-\alpha/2}\, S\{b_1\}$, where $S\{b_1\} = \sqrt{\dfrac{MSE}{\Sigma(x_i-\bar{x})^2}}$ and

$$MSE = \frac{\Sigma(y_i - \hat{y}_i)^2}{n-2}.$$

Test concerning β_1

$H_0: \mu < 5.5$ $\quad$ $H_1: \mu > 5.5$

Steps:

1) Statement of hypotheses:

Null hypothesis	Alternative hypothesis
$H_0 : \beta_1 = \beta_{10}$ — constant	$H_1 : \beta_1 \neq \beta_{10}$ (two sided)
$H_0 : \beta_1 \le \beta_{10}$	$H_1 : \beta_1 > \beta_{10}$ (one sided)
$H_0 : \beta_1 \ge \beta_{10}$	$H_1 : \beta_1 < \beta_{10}$ (one sided)

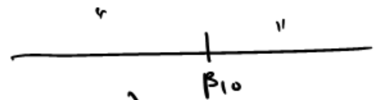
2) Test Statistic

under $H_0: \beta_1 = \beta_{10}$

$$T = \frac{b_1 - \beta_{10}}{s\{b_1\}} \sim t_{n-2}$$

3) calculating the critical value or calculating the p-value.

Suppose the observed value of T is t^*.



critical value:

$t(1-\alpha/2 : n-2)$ if $H_1: \beta_1 \neq \beta_{10}$ (two Sided)

$t(1-\alpha : n-2)$ if $H_1: \beta_1 > \beta_{10}$ or $H_1: \beta_1 < \beta_{10}$ (one Sided)

OR

$$\begin{aligned} \text{p-value} &= P(T > t^*) \quad \text{if } H_1: \beta_1 > \beta_{10}, \\ &= P(T < t^*) \quad \text{if } H_1: \beta_1 < \beta_{10}, \\ &= 2P(T > |t^*|) \quad \text{if } H_1: \beta_1 \neq \beta_{10}. \end{aligned}$$

4) conclusion:

* using the critical value:

$t^* > t(1-\alpha, n-2) \Rightarrow$ reject H_0 ; $H_1: \beta_1 > \beta_{10}$

$t^* < t(1-\alpha, n-2) \Rightarrow$ reject H_0 : $H_1: \beta_1 < \beta_{10}$

$|t^*| > t(1-\alpha/2, n-2) \Rightarrow$ reject H_0 : $H_1: \beta_1 \neq \beta_{10}$

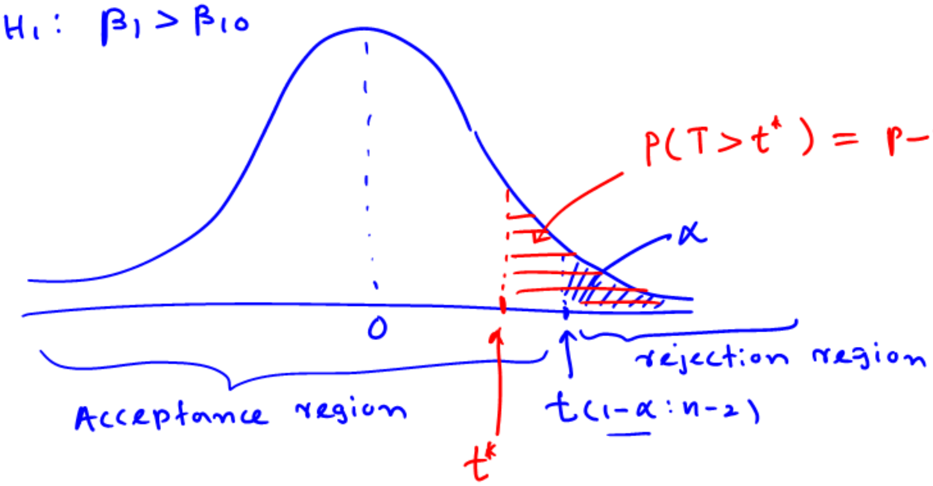
* using the p-value

Reject H_0 if p-value $< \alpha$.

* Make a non-technical Statement about your conclusion.

Idea about the conclusion:

Consider $H_1: \beta_1 > \beta_{10}$



Note:

The hypothesis

$H_0: \beta_1 = 0$ (ie there is no relationship between X and Y)

vs

$H_1: \beta_1 \neq 0$ (ie there is a relationship)

is tested frequently.

Inference on β_0

Point estimator: $b_0 = \bar{Y} - b_1 \bar{X}$, where $\bar{Y} = \underbrace{\frac{\sum Y_i}{n} - \sum k_i Y_i \bar{X}}$

$E(b_0) = \beta_0$ (proof HW).

$Var(b_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right]$ (proof HW)

$\therefore$ Estimated variance $= \hat{Var}(b_0) = S^2\{b_0\} = MSE \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right]$

Further since b_0 is a linear combination of independent Y_is,

$$b_0 \sim N\left(\beta_0 , \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right] \right).$$

$\therefore$ When σ^2 is known,

$$\underbrace{\frac{b_0-\beta_0}{\sigma\{b_0\}}}_{\text{Standardize Statistic}} \sim N(0,1)$$

when σ^2 is unknown,

$$\underbrace{\frac{b_0-\beta_0}{S\{b_0\}}}_{\text{Studentize Statistic}} \sim t_{n-2}$$

* Confidence interval for β_0 :

By following a Similar argument as for β_1, $(1-\alpha)100\%$ confidence interval for β_0 :

$$\left(b_0 - t_{1-\alpha/2:n-2}\, S\{b_0\}\ ,\ b_0 + t_{1-\alpha/2:n-2}\, S\{b_0\}\right),$$

where $S\{b_0\} = \sqrt{MSE\left[\frac{1}{n} + \frac{\bar{X}^2}{\Sigma(X_i-\bar{X})^2}\right]}$.

Note:

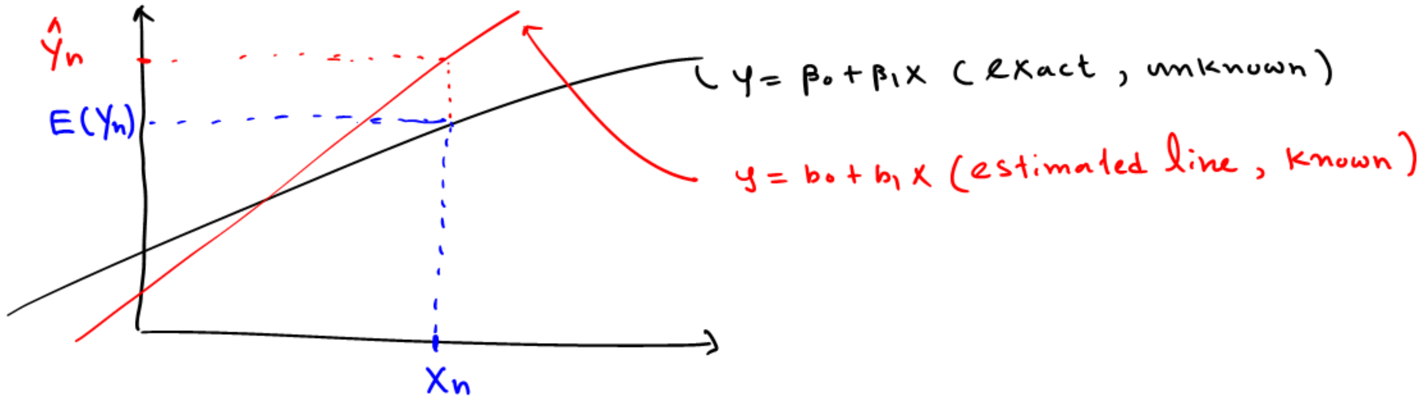
When the normality assumption is vialated, distribution of b_0 and b_1 are not normal. But b_0 and b_1 are approximately normal for large samples (central limit theorem).

Inference on the mean response

Consider the mean response at $X=X_h$,

$$E(Y_h) = \beta_0 + \beta_1 X_h \text{ — parameter}$$

$$\hat{Y}_h = b_0 + b_1 X_h \text{ — point estimator.}$$



Interval estimator of $E(Y_n)$

$$E(\hat{Y}_n) = E(b_0 + b_1 X_n) = E(b_0) + X_n E(b_1) = \beta_0 + \beta_1 X_n = E(Y_n)$$

$$\sigma^2\{\hat{Y}_n\} = \mathrm{var}(\hat{Y}_n) = \sigma^2\left[\frac{1}{n} + \frac{(X_n - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right] \quad (\text{proof - HW})$$

Further since $\hat{Y}_n$ is linear combination of Y_is

$$\hat{Y}_n \sim N\left(E(Y_n),\ \sigma^2\left[\frac{1}{n} + \frac{(X_n - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]\right).$$

* when σ^2 is known,

$$\frac{\hat{Y}_n - E(Y_n)}{\sigma\{\hat{Y}_n\}} \sim N(0,1)$$

* when σ^2 is unknown

$$\frac{\hat{Y}_n - E(Y_n)}{s'\{\hat{Y}_n\}} \sim t_{n-2}.$$

* $(1-\alpha)100\%$ confidence interval for the mean response $E(Y_n)$ at $X = X_n$ is

$$\left(\hat{Y}_n - t_{1-\alpha/2;n-2}\, s'\{\hat{Y}_n\},\ \hat{Y}_n + t_{1-\alpha/2;n-2}\, s'\{\hat{Y}_n\}\right),$$

where $S\{\hat{Y}_h\} = \sqrt{MSE\left[\frac{1}{n} + \frac{(X_h-\bar{X})^2}{\Sigma(X_i-\bar{X})^2}\right]}$.