Confidence Intervals :-

(σ - known)

$X_1, \dots, X_n$ (i.i.d.) Sample from Normal$(\mu, 3^2)$

$$\mathbb{P}\left(\frac{\sqrt{n}(\bar{x}-\mu)}{3} \le 1.96 \right) = 0.95$$

$$\mathbb{P}\left(\mu \in \left(\frac{-(1.96)3}{\sqrt{n}} + \bar{X}, \ \bar{X} + \frac{(1.96)3}{\sqrt{n}} \right) \right) = 0.95$$

We said

$$\left(\frac{-(1.96)3}{\sqrt{n}} + \bar{X}, \ \bar{X} + \frac{(1.96)3}{\sqrt{n}} \right)$$

is a 95% C.I for μ.

Consider the case when σ - unknown

let $X_1, \dots, X_n$ i.i.d Normal distribution

Fact

- $\frac{n-1}{\sigma^2} S^2$ has χ^2_{n-1} distribution
- $\bar{X}$ and S^2 are independent

- $\dfrac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t_{n-1}$ distribution.

where t_{n-1} – student's t distribution.

P.d.f of student's t- distribution t_k (for $k \geq 1$)

$$f(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \left(1+\frac{t^2}{k}\right)^{-\frac{(k+1)}{2}}$$

$t \in \mathbb{R}$

Take $t_{n-1}^{0.95}$ such that

$$P(|t_{n-1}| \leq t_{n-1}^{0.95}) \cong 0.95$$

$$\Rightarrow P\left(\mu \in \left(-\frac{t_{n-1}^{0.95})\,S}{\sqrt{n}} + \overline{X},\ \frac{t_{n-1}^{0.95}}{\sqrt{n}} S + \overline{X}\right)\right)$$

$$\cong 0.95$$

Confidence interval for μ (σ - unknown)

we say that

$$\left(-\left(t_{n-1}^{0.95}\right)\frac{s}{\sqrt{n}} + \bar{X} \,,\ \frac{t_{n-1}^{0.95}}{\sqrt{n}}\, s + \bar{X} \right)$$ is a

95% confidence interval for μ.

Example: • $\bar{X} = 10.2$, $\mu = ?$, $\sigma = 3$, $n = 16$

$$\left(-(1.96)\frac{3}{4} + 10.2 \,,\ +(1.96)\frac{3}{4} + 10.2 \right)$$

$$\cong (8.73,\ 11.67)$$

• $\bar{X} = 10.2$, $S = 3$ $\mu = ?$, $\sigma = ?$, $n = 16$

$$\left(-(2.13)\frac{3}{4} + 10.2 \,,\ (2.13)\frac{3}{4} + 10.2 \right)$$

$$\cong (8.6,\ 11.80)$$ [large interval]

Revisit Simple linear Regression

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

Immediate

$$E(\hat{\beta}_1) = \beta_1 \quad , \quad E(\hat{\beta}_0) = \beta_0$$

from $E[y_i] = \beta_0 + \beta_1 x_i$

Observe that

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

$$\Rightarrow \quad Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Thus :- $\hat{\beta}_1 \sim$ Normal distributed

[$\because$ y_i are independent Normal random variables]

$$\Rightarrow \quad \frac{\hat{\beta}_1 - E(\hat{\beta}_1)}{\sqrt{Var(\hat{\beta}_1)}} \overset{d}{=} N(0,1)$$

$\therefore$ If σ^2 is known then we can construct confidence intervals for β_1.

Question:- • How to estimate σ if it is not known?

• Can we construct confidence intervals if σ is estimated from the data?

$$RSS = \sum_{i=1}^{n} \underbrace{(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)}_{\text{residual}}^2$$

(residual sum of squares)

let $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

(See H.w.)

Ex :- $\hat{\varepsilon}$, $\hat{\beta}$ are independent

$\Rightarrow$ $\hat{\beta}_1$ and RSS are independent.

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/S_{xx}^2}} \stackrel{d}{=} N(0,1)$$

$\boxed{\text{\& independent.}}$

Ex: $$(n-2)\frac{RSS}{\sigma^2} \stackrel{d}{=} \chi^2_{n-2}$$

$$\Rightarrow \quad \frac{\hat{\beta} - \beta_1}{\sqrt{RSS/S_{xx}^2}} \stackrel{d}{=} t_{n-2}$$

[Facts from Probability] See H.w.

• 95% confidence interval for β_1

$$\left(\hat{\beta}_1 - \underbrace{t_{n-2,\, 0.025}}_{\text{appropriate quantile}} \sqrt{\frac{RSS}{S_{xx}}},\right.$$

$$\left.\hat{\beta}_1 + t_{n-2,\, 0.025} \sqrt{\frac{RSS}{S_{xx}}}\right)$$