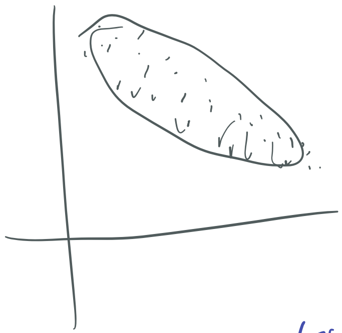
☆ Lecture #3 2023年9月6日(三).

Prediction

Recap.

SLR: $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$



$\mathbb{E}\varepsilon_i = 0$

$Var\,\varepsilon_i = \sigma^2$ (homoscedastic)

$Cov(\varepsilon_i, \varepsilon_j) = 0.$

$\Rightarrow Cov(y_i, y_j) = 0$

Given obs's $(X_i, y_i)_{i=1}$

$$\hat{\beta}^{LS} = b_1 = \frac{\Sigma(X_i - \bar{X})(y_i - \bar{y})}{\Sigma(X_j - \bar{X})^2} \leftarrow S_{xx}$$

$$\hat{\beta}_0^{LS} = b_0 = \bar{y} - b_1\bar{X}$$

Gauss Markov Thm → For the SLR model, b_0, b_1 were MVUE.

BLUE.

☆ Proof of GMT:

$\mathbb{E}(b_1) = \beta_1$

$\mathbb{E}(b_0) = \beta_0$

$b_1 = \Sigma k_i y_i$

$$b_1 = \Big[\sum_{i=1}^n (X_i - \bar{X})y_i - \underbrace{\sum_{i=1}^n (X_i - \bar{X})\bar{y}}_{=0}\Big]\frac{1}{S_{xx}}$$

$$\therefore b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})y_i}{S_{xx}}$$

$\bar{y}\sum_{i=1}^n (X_i - \bar{X})$

$= \bar{y}(\Sigma X_i - n\bar{X})$

$= \bar{y}(\Sigma X_i - n\frac{\Sigma X_i}{n}) = 0.$

$$\mathbb{E}(b_1) = \mathbb{E}(\Sigma_i k_i y_i) = \Sigma k_i\,\mathbb{E}(y_i)$$

linearity of $\mathbb{E}$.

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

from SLR model

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

$$= \beta_1$$

$$\sum_i k_i = \sum_{i=1}^n \frac{(x_i-\bar{x})}{S_{xx}} = 0.$$

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

$$= \frac{1}{S_{xx}} \sum (x_i-\bar{x})(x_i-\bar{x})$$

$$= \frac{S_{xx}}{S_{xx}}$$

$$= 1$$

☆ Similarly, $\mathbb{E}(b_0) = \beta_0$.

☆ OLS model gives the lowest variance.

proof: Consider any other $\hat{\beta}_1 = \sum_{i=1}^n \underbrace{c_i}_{\text{Constant}} y_i$

Need to show that $var(\hat{\beta}_1) \geq var(\underbrace{b_1}_{\text{OLS estimator}})$

where $b_1 = \sum k_i y_i$

Let us write: $\hat{\beta}_1 = \sum_{i=1}^n \overbrace{(k_i+d_i)}^{c_i} y_i$

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

$$= \sum (k_i+d_i)^2 var(y_i) + \underbrace{cov(y_i, y_j)}_{=0.}$$

$$= \sum \left(k_i^2 + 2k_i d_i + d_i^2\right) \delta^2 \quad (*)$$

$$\sum_i k_i d_i = \sum_i k_i (c_i - k_i) = \frac{1}{S_{xx}} \left[\underbrace{\sum (x_i-\bar{x}) c_i}_{\|} - \underbrace{\sum (x_i-\bar{x}) k_i}_{\|} \right] = 0$$

$$= \underbrace{\sum c_i x_i}_{1} - \bar{x} \underbrace{\sum c_i}_{=0} \qquad = \sum k_i x_i - \bar{x} \sum k_i$$

$\mathbb{E}(\hat{\beta}_1) = \beta_1 \Leftrightarrow \sum_i c_i = 0$ $\qquad = 1 - 0$

$\sum_i c_i x_i = 1$ $\qquad = 1.$

$$(*) = \sum (k_i^2 + d_i^2)\,\sigma^2 = \underbrace{\sigma^2 \sum_i k_i^2}_{Var(b_1)} + \sigma^2 \sum_i d_i^2 \geq 0.$$

$\therefore \; Var(\hat{\beta}_1) \geq Var(b_1)$

☆ Inference in SLR:

Hypothesis test, β_0, β_1

CI of β_0, β_1

CI of mean of resp. var.

$\varepsilon_i \sim$ i.i.d. $N(0, \sigma^2)$

1) Hypothesis test on β_1

$H_0: \beta_1 = 0, \quad H_a: \beta_1 \neq 0$

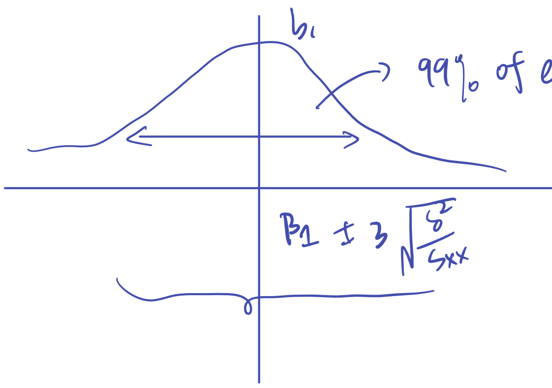
under H_0: $y_i = \beta_0 + \varepsilon_i \sim N(\beta_0, \sigma^2), \forall i$

under H_a: $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$

$\Rightarrow$ Need to know distribution of b_1

$$b_1 = \sum_{i=1}^{n} k_i y_i \sim N\left(\beta_1, \frac{1}{S_{XX}}\sigma^2\right) \text{ OR } N\left(\beta_1, \frac{1}{S_{XX}} \sum_j (X_j - \bar{X})^2\right).$$

Fact: $Var(b_1) = \sigma^2 \sum k_i^2 = \sigma^2 \frac{1}{S_{XX}}$



Test statistics:

t-stats.

$t^* = b_1 / s(b_1)$ → Sample standard deviation

where $s^2(b_1)$ – Sample variance

$var(\beta_1) = \sigma^2 / S_{xx}$.

How to estimate σ^2? Using MSE.

$\hat{\sigma}^2 = MSE$.

$s^2(b_1) = \hat{\sigma}^2 / S_{xx} = MSE / \Sigma_j (x_j - \bar{x})^2$

$MSE = \dfrac{\text{Sum of square error}}{n-2}$

$= \dfrac{1}{n-2} \left(\Sigma_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 \right)$

deg. of freedom

So $t^* = \dfrac{1}{s(b_1)} b_1 \sim t_{n-2}$ → t-distribution.

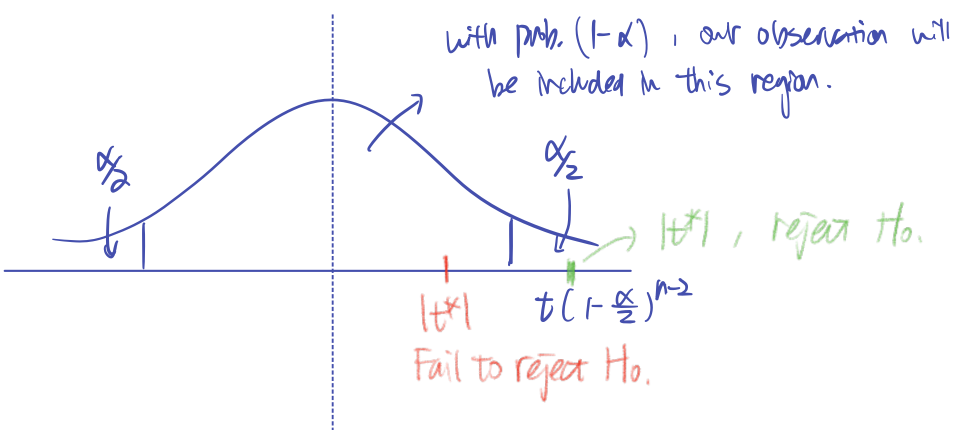
→ w/ specified significance level α

a) if $|t^*| < t(1-\alpha/2, n-2)$

↳ accept H_0

b) if $|t^*| > t(1-\alpha/2, n-2)$

↳ reject H_0.

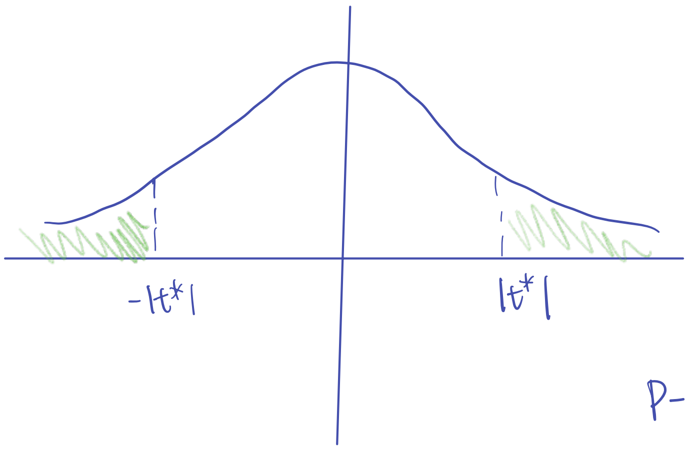


α : Prob. (Type I error).

reject H_0 when H_0 is TRUE.

P-value : "strength of your observation"

$\equiv$ prob. of seeing observation as extreme as t^*, or more extreme, in favor of H_a.



P-value : $2P(t > |t^*|)$

If the p-value is too low, Null hyp. (H_0) must go.

★ Confidence Interval

CI for β_1 : $b_1 \pm t(1-\frac{\alpha}{2}, n-2) \cdot s(b_1)$

duality : When we fail to reject H_0 w/ sig. level α $\Leftrightarrow$ $0 \in CI_{1-\alpha}$.

reject H_0 w/ sig. level α $\Leftrightarrow$ $0 \notin CI_{1-\alpha}$.

(!!) Interpretation of CI :

If we take samples $(X_i, Y_i)_{i=1}^{n}$ many times with fixed X_i, then $(1-\alpha) \times 100\%$ of the times, the CI will contain the true β_1.

☆ Section #2 2023年9月8日(五).

t-distribution

$Z \sim N(0,1)$, χ^2_v , $Z \perp\!\!\!\perp \chi^2_v$

$$t_v = \frac{Z}{\sqrt{\frac{\chi^2_v}{v}}}$$

$\downarrow$
deg. of freedom

random sample of n observation, $Y_1, \dots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$

① sample standard deviation

$$S = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{n-1}}$$

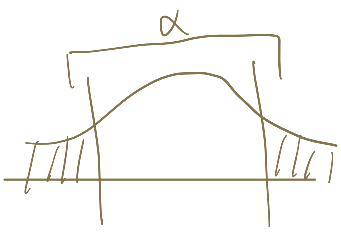
make sure this is the unbias estimator for θ

② $S\{\bar{Y}\} = \frac{S}{\sqrt{n}}$ estimator s.d. of $\bar{Y}$.

③ sample mean $\bar{Y} = \frac{1}{n} \sum Y_i$

$$\boxed{\frac{\bar{Y} - \mu}{S\{\bar{Y}\}} \sim t_{n-1}}$$

$$t^* = \frac{\bar{Y} - \mu_0}{S\{\bar{Y}\}}.$$



2-sided

① $H_0: \mu = \mu_0$, $|t^*| \leq t(1-\frac{\alpha}{2}; n-1)$ fail to reject H_0.

$H_a: \mu \neq \mu_0$ $|t^*| \geq t(1-\frac{\alpha}{2}; n-1) \Rightarrow H_a$

1-sided.

② $H_0: \mu \geq \mu_0$, $t^* \geq t(\alpha; n-1)$ fail to reject H_0.

$H_a: \mu < \mu_0$, otherwise $\Rightarrow H_a$.

③ $H_0: \mu \leq \mu_0$, $t^* \leq (1-\alpha; n-1) \Rightarrow H_0$.

$H_a: \mu > \mu_0$,

☆ Example:

$n=25 \quad \bar{Y}=5.7 \quad s=8 \quad \alpha=0.02$

2-sided:

$H_0: \mu=10$

$H_a: \mu \neq 10.$

$t(0.99; 24) = 2.492.$

$t^* = \dfrac{\bar{Y}-\mu_0}{s\{\bar{Y}\}}$

$s\{\bar{Y}\} = \dfrac{8}{\sqrt{25}} = \dfrac{8}{5} = 1.6$

$\bar{Y}-\mu_0 = -4.3$

$t^* = -\dfrac{4.3}{1.6} = -\dfrac{43}{16} = -2.6875$

$|t^*| > 2.492.$ reject H_0.

$CI = \bar{Y} \pm t(1-\frac{\alpha}{2}, n-1)\, s\{\bar{Y}\}.$

$= 5.7 \pm 2.492 \cdot 1.6.$

* If $\mu_0 = 10 \in CI \Leftrightarrow$ Not reject H_0.

P value = P(t_{24} is more extreme than t^*)

$= P(|t_{24}| > 2.69)$

If P value $< \alpha \Rightarrow$ reject H_0.

→ 1-side:

$H_0: \mu \leq 10$

$H_a: \mu > 10$

$t^* = -2.69$

$t(0.98, 24) = 2.1715$

$t(0.02, 24) = -2.1715$

$t^* \leq t(0.02, 24) \Rightarrow$ fail to reject H_0.

☆ Lecture #4. 2023年9月11日(一)。

- 2-sided hypothesis testing β_1.] duality.
- CI]

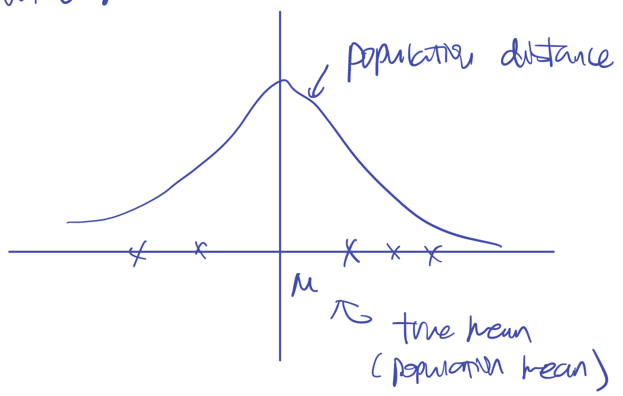
today: departure from normality
estimation of mean response $\widehat{E(Y_h)}$

☆ Recap of t-student.

sample: $Y_1, \ldots, Y_n \sim$ i.i.d. $N(\mu, \sigma^2)$

$$\hat{\mu} = \bar{Y} = \frac{\sum Y_i}{n}$$

We use random sample to estimate mean.

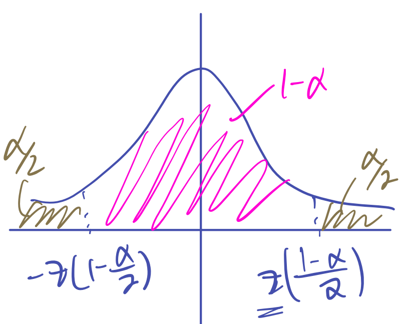


$\hat{\mu}$ is estimator of μ, μ is a fixed value, $\hat{\mu}$ is a r.v.

$$\hat{\mu} = \frac{\sum Y_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\Rightarrow \frac{\hat{\mu} - \mu}{\sqrt{Var(\hat{\mu})}} = \frac{\hat{\mu} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

↳ standardized r.v.



↳ notation of percentile of $N(0,1)$.

$$P\left(a \le \frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}} \le b\right) = 1-\alpha.$$

$$\mu \in \left(\hat{\mu} - \frac{\sigma}{\sqrt{n}} z\left(1-\frac{\alpha}{2}\right),\ \hat{\mu} + \frac{\sigma}{\sqrt{n}} z\left(1-\frac{\alpha}{2}\right)\right)$$ with sig. level α.

→ we need to estimate σ from the sample. (in case we don't know sigma)

$$S^2 = \frac{\sum_i (y_i - \bar{y})^2}{(n-1)}$$ (sample variance). y_i, $i=1,\dots,n$.

$$\bar{y} = \frac{1}{n}\sum_i y_i$$

$$\mathbb{E}(S) = \sigma.$$

$$\underbrace{\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}}}_{\text{standardized}} \xrightarrow{\sigma \to S} \underbrace{\frac{\hat{\mu}-\mu}{\frac{S}{\sqrt{n}}}}_{\text{studentized}} \sim t(n-1)$$

$n-1$ ↳ deg of freedom.

Fact: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ (chi-square).

Let:

sig. level $= \alpha$. CI $= ?$

$$\mu \in \left(\hat{\mu} - \frac{S}{\sqrt{n}} t\left(n-1, 1-\frac{\alpha}{2}\right),\ \hat{\mu} + \frac{S}{\sqrt{n}} t\left(n-1, 1-\frac{\alpha}{2}\right)\right)$$ w/ sig level.

☆ Back to linear Reg.

Two sided Hypothesis Testing: $\beta_1 = 0$ (H_0)

$\beta_1 \neq 0.$ (H_a)

One sided H.T. : $\left.\begin{array}{l}\beta_1 \le 0 \\ \beta_1 > 0\end{array}\right\}$ OR $\left\{\begin{array}{ll}\beta_1 \ge 0 & (H_0) \\ \beta_1 < 0. & (H_a)\end{array}\right.$

if $H_0: \beta_1 = \beta_0$ T.S. $\dfrac{b_1 - \beta_0}{s(b_1)}$ $s(b_1) = \sqrt{\dfrac{MSE}{S_{xx}}}$

$H_a: \beta_1 \neq \beta_0$

$MSE = \dfrac{(y_i - \bar{y})^2}{n-2}$

Note:

$\varepsilon_i \sim N(\mu, \sigma^2)$, i.i.d.

departure from normality is not a problem if

n is large, n=100, n=50.

$b_1 = \sum k_i y_i$ → still close to normal

even if ε_i not normal b/c

of CLT (central limit thm) as $n \to \infty$.

$b_1 = \sum k_i y_i$

$S_{xx} = \sum (X_i - \bar{X})^2$

$s(b_1) = \sqrt{\dfrac{(y_i - \bar{y})^2}{(n-2)\sum_j (X_j - \bar{X})^2}}$

⊗ Point estimation of $\varepsilon\, Y_h$ ← mean of the response var. at $X = X_h$.

$\mathbb{E}(Y_h) = \beta_0 + \beta_1 X_h$

$\mathbb{E}(\hat{Y}_h) = b_0 + b_1 X_h = \hat{y}_h$

$\hat{y}_h \sim N\left(\mathbb{E}(Y_h), \sigma^2\left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2}\right)\right)$

$\mathbb{E}(\hat{y}_h) = \mathbb{E}(b_0 + b_1 X_h) = \mathbb{E}(b_0) + \mathbb{E}(b_1) X_h$

$= \beta_0 + \beta_1 X_h = \mathbb{E}(Y_h)$ ✓

$Var(\hat{y}_h) = Var(b_0 + b_1 X_h) = Var(\underbrace{\bar{y} - b_1 \bar{X}}_{b_0} + b_1 X_h)$

$= Var(\bar{y}) + Var(b_1)(X_h - \bar{X})^2 + 2\,cov\underbrace{(\bar{y}, b_1(X_h - \bar{X}))}_{=0.}$

$= \frac{1}{n}\sigma^2 + \frac{1}{\sum (X_i - \bar{X})^2}(X_h - \bar{X})\sigma^2$

CI for the mean response at X_h.

$$\hat{y}_h \pm \sqrt{MSE\left(\frac{1}{n} + \frac{(X_h-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right)}\, t(n-2, 1-\tfrac{\alpha}{2})$$

OR $\hat{y}_h \pm s(\hat{y}_h) \cdot t(n-2, 1-\frac{\alpha}{2})$.

※ Prediction of a new observation y_h.

$\mathbb{E}(Y_h) = \beta_0 + \beta_1 X_h \quad \neq \quad y_h = \beta_0 + \beta_1 X_h + \varepsilon_h$.

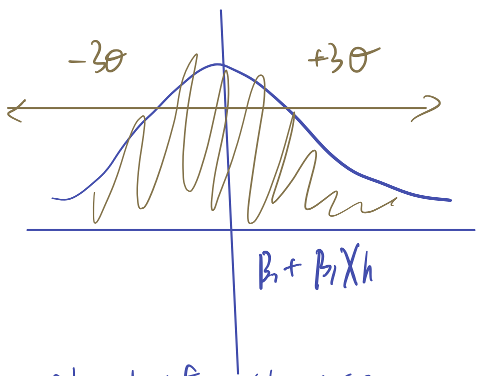
mean of response var. at X_h.

If β_0, β_1, σ are given:

$y_{h(new)} \sim N(\beta_0 + \beta_1 X_h, \sigma^2)$

$y_{h(new)}$ is inside $\beta_0 + \beta_1 X_h \pm 3\sigma$ with prob. 99.7%.

What if we don't know β_0, β_1, σ?



The spread of error term will be much larger than the chart show in left.

Chart for the case when β_0, β_1, σ is known.

Define new R.V. : $Y_{h(new)} - \hat{Y}_h \sim N(0, \sigma^2(1+\frac{1}{n}\frac{(X_h-\bar{X})^2}{\sum_i (X_i-\bar{X})^2}))$

☆ Lecture #5. 2023年 9月13日(三) .

- Finish CI for prediction of $Y_{h(new)}$.
- Confidence band for reg. line.
- ANOVA] Chap. 2.7.
- F-test
- General linear test chap. 2.8.

Prediction of new observation $Y_{h(new)}$.

Recall $\mathbb{E}(\hat{Y}_h) \neq Y_{h(new)}$

mean of the response var. at X_h | value of the new response var. at h.

$\hat{Y}_h = b_0 + b_1 X_h$

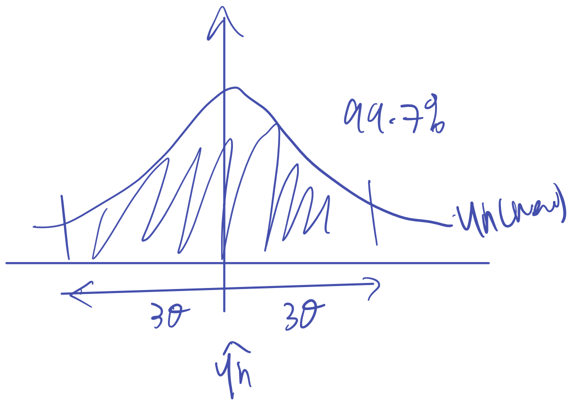
$Y_{h(new)} = \beta_0 + \beta_1 X_h + \varepsilon_h$.

$\mathbb{E}(\hat{Y}_h) = \mathbb{E}(Y_h)$

$\because \mathbb{E}(b_0) = \beta_0$

$\mathbb{E}(b_1) = \beta_1$.

☆ When $\beta_0, \beta_1, \sigma^2$ Known.



※ When $\beta_0, \beta_1, \sigma^2$ are unknown.

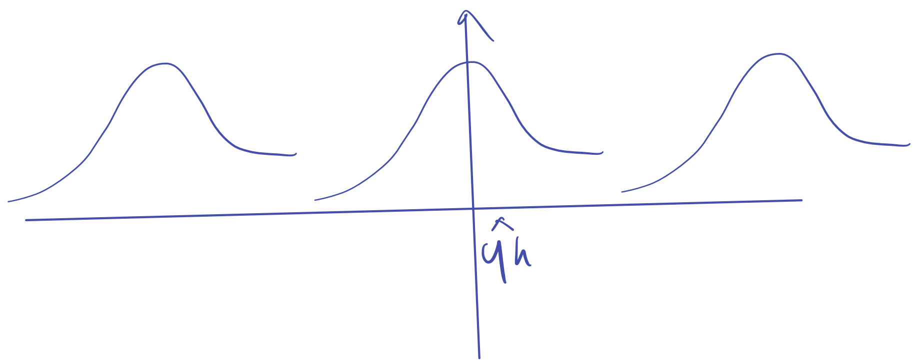
Recall: $\hat{Y}_h \sim N\left(E[Y_h], \sigma^2\left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right)\right)$ See lecture 4.

$\uparrow$

point estimator of the mean at X_h.

Back to $Y_{h(new)}$:

$$\left(Y_{h(new)} - \hat{Y}_h\right) \sim N\left(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right)^2\right)$$



※ $E(Y_{h(new)} - \hat{Y}_h) = 0$.

$$\begin{aligned} E(Y_{h(new)} - \hat{Y}_h) &= E(\beta_0 + \beta_1 X_h + \varepsilon_h - b_0 - b_1 X_h) \\ &= \beta_0 + \beta_1 X_h + 0 - E(b_0) - E(b_1 X_h) \\ &= \beta_0 + \beta_1 X_h - \beta_0 - \beta_1 X_h \\ &= 0 \end{aligned}$$

※ $var(Y_{h(new)} - \hat{Y}_h) : var(\underbrace{\beta_0+\beta_1 X_h}_{=0} + \varepsilon_h - b_0 - b_1 X_h)$

$= var(\varepsilon_h) + var(b_0+b_1X_h) - \underbrace{2cov(\varepsilon_h, b_0+b_1X_h)}_{=0}$

$= \sigma^2 + var(\hat{Y}_h)$

$cov(Y_i, Y_j) = 0, \forall i \neq j$

$= \sigma^2 + \sigma^2\left(\frac{1}{n} + \frac{(X_h-\bar{X})^2}{\sum_i (X_i-\bar{X})^2}\right)$

R.

↓

estimate by MSE.

$\hat{Y}_h \pm s(prediction)\, t(n-2, 1-\frac{\alpha}{2})$, α - sig. level.

↳ CI for $Y_{h(new)}$

$s^2(pred) = MSE\left(1 + \frac{1}{n} + \frac{(X_h-\bar{X})^2}{\sum_i (X_i-\bar{X})^2}\right)$

Notes: 1) Prediction interval ≠ confidence interval.

↓ for r.v.s e.g. $Y_{h(new)}$

↓ for parameters of distributions.

2) $\underbrace{\text{CI for } \mathbb{E}(Y_h)}_{\text{variability of the mean}}$ is always narrower than $\underbrace{\text{PI. for } Y_{h(new)}}_{\text{individual variability}}$

3) departure from normality.

$Y_{h(new)} = \beta_0 + \beta_1 X_h + \varepsilon_h$

↙ problem.

↳ 1 value.

but not for $\mathbb{E}(Y_h)$

Confidence Band for regression line (Working-Hotelling conf. band).

Region R in $\mathbb{R}^2$ s.t. $\mathbb{P}(\beta_0+\beta_1 X \subseteq R) = 1-\alpha$. $\forall X$.

W-H conf. Band : $\underline{\hat{Y}_h} \pm W \cdot s(\underline{\hat{Y}_h})$

$W = \sqrt{2F(1-\alpha, 2, n-2)}$ → constant.

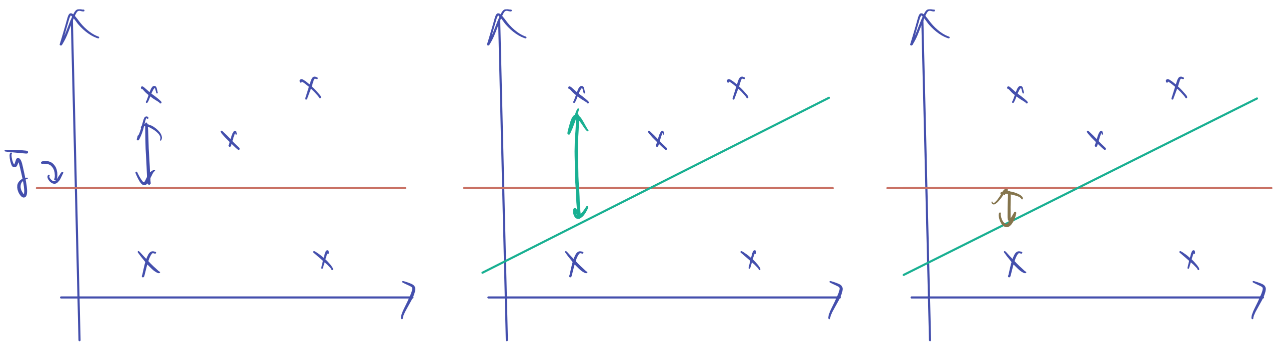
2 cols, df.

$$s(\hat{Y}_h) = \sqrt{MSE\left(\frac{1}{n} + \frac{(X_h-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right)}$$

F-distribution : If $U \sim \chi^2_m$, $V \sim \chi^2_l$, U, V indep. ($U \perp\!\!\!\perp V$)

$$\frac{U/m}{V/l} \sim F_{m,l}.$$

※ Chap. 2.7. KNNL – Analysis of variance approach to regression.



$$y_i - \bar{y} = \underbrace{y_i - \hat{y}_i}_{e_i - \text{residual}} + \hat{y}_i - \bar{y}$$

deviation of response var. from mean

e_i - residual.

deviation from fitted value

deviation of fitted value from its mean.

(Total sum of squares). $\sum (y_i - \bar{y})^2 = SSTO$.

(Error Sum of Squares) $\sum (y_i - \hat{y}_i)^2 = SSE$

(Regression sum of squares) $\sum (\hat{y}_i - \bar{y})^2 = SSR$.

Fundamental ANOVA identity:

$$SSTO = SSE + SSR.$$

Interpretation

SSTO: total variation in y_i

unexplained SSE: variation in y_i after accounting for X_i

explained SSR: variation in y_i due to X.

$$SSR = \sum (\hat{y}_i - \bar{y})^2$$

$$\bar{y} = \frac{1}{n} \sum_i \hat{y}_i$$

Source of Variation	SS	df	MS	E(MS)
Regression	SSR	1	$MSR = \frac{SSR}{1}$	$\sigma^2 + \beta_1^2 \sum (X_i - \bar{X})^2$
Error	SSE	$n-2$	$MSE = \frac{SSE}{(n-2)}$	σ^2
Total	SSTO	$n-1$		

H_0: If $\beta_1 = 0$, $\mathbb{E}(MSE) = \mathbb{E}(MSR) = \sigma^2$

$$MSE \approx MSR.$$

H_a: $\beta_1 \neq 0$. $\mathbb{E}(MSE) \geq \mathbb{E}(MSR)$

Model utility test (F-test)

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-2} \qquad \frac{SSR}{\sigma^2} \sim \chi^2_1 \qquad \text{under } H_0.$$

$SSE \perp\!\!\!\perp SSR$.

$$\frac{\frac{SSR}{\sigma^2}/1}{\frac{SSE}{\sigma^2}/n-2} = \frac{SSR}{SSE/n-2} = \frac{MSR}{MSE}. \sim F_{1,n-2} \text{ dfs}$$

When the model is useful?

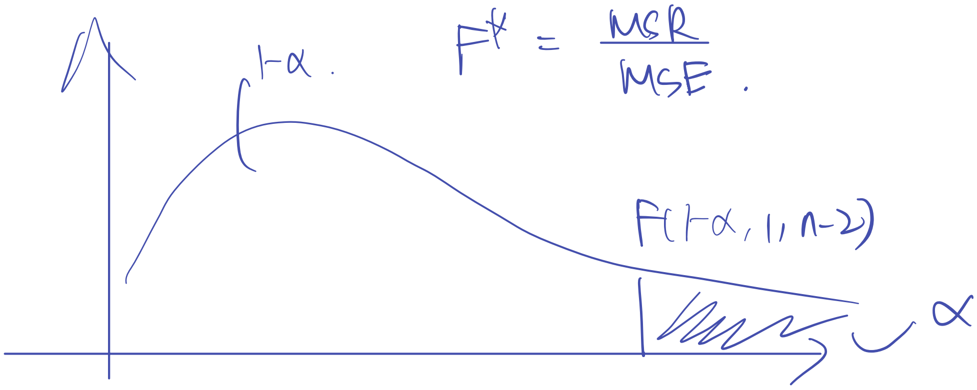
$\underset{\text{explained}}{MSR} \gg \underset{\text{unexplained}}{MSE}$. $\Rightarrow$ accept H_a $\beta_1 \neq 0$.

$MSR \approx MSE$ $\Rightarrow$ accept H_0 $\beta_1 = 0$.

rejection region:

$$F^* > F(1-\alpha, 1, n-2)$$

$$F^* = \frac{MSR}{MSE}.$$



☆ Section 2023年9月15日(五) Week #3.

Residual standard error = $\sqrt{MSE}$

$MSE = \frac{SSE}{n-p}$. p stands for # of variables + intercept Term. (usually 1).

$F\text{-stats} = \frac{SSR}{MSE}$. n = df + (# lines in the table)

$(t\text{-stats})^2 = F\text{-stats}$.

observed value

t-test P-value: $P(t\text{-stats} > t(x))$

F-test F-stats: $P(F\text{-stats} > F(x))$.

☆ Lecture 6 2023年9月18日(一).

Recall $SSTO = \sum (y_i - \bar{y})^2$

$SSE = \sum (y_i - \hat{y}_i)^2$

$SSR = \sum (\hat{y}_i - \bar{y})^2$ $\sum y_i = \sum \hat{y}_i$

ANOVA identity: $SSTO = SSE + SSR$

ANOVA Table:

Source of variation	SS	df	MS	$\mathbb{E}(MS)$
Regression	SSR	1	$MSR = \frac{SSR}{1}$	$\sigma^2 + \beta_1^2 \sum (X_i - \bar{X})$
Error	SSE	n-2	$MSE = \frac{SSE}{n-2}$	σ^2
Total	SSTO	n-1		

Model utility Test.

H_0: If $\beta_1 = 0$, $\mathbb{E}(MSR) = \mathbb{E}(MSE) = \delta^2$

H_a: If $\beta_1 \neq 0$, $\mathbb{E}(MSR) > \mathbb{E}(MSE)$.

$F^* = \frac{MSR}{MSE} \sim F_{1,n-2}$

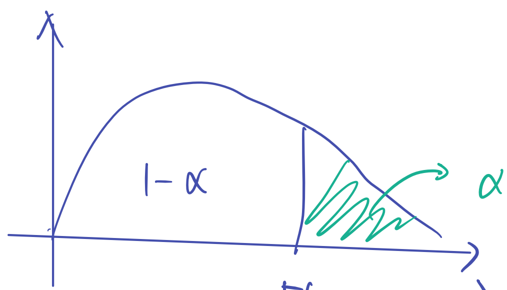
☆ Facts: $\frac{SSE}{\delta^2} \sim \chi^2_{n-2}$ → df for chi-square.

$\frac{SSR}{\delta^2} \sim \chi^2_1$ → df for chi-square

$SSE \perp\!\!\!\perp SSR$
↓
independent.

Rejection region. w/ sig level α

$F^* > F(1-\alpha, 1, n-2)$



What abt. t-test?

$t^* = \frac{b_1}{s(b_1)}$ $\quad F^* = \frac{MSR}{MSE} = \left(\frac{b_1}{s(b_1)}\right)^2$

Proof: Compute formula for SSTO, SSE, SSR

$SSTO = \sum y_i^2 - n\bar{y}^2$

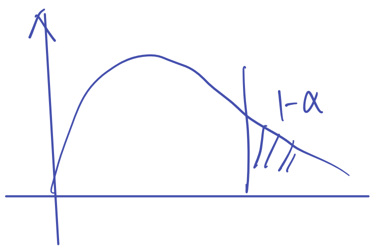
$SSR = b_1^2 \sum (X_i - \bar{X})^2 = b_1^2 S_{xx}$

$SSE = \sum y_i^2 - b_0(\sum y_i) - b_1 \sum y_i X_i$

(!!) T-test is more flexible than F-test.
(can be one-sided)

t-test

F-test



F-test is more generalizable.

$H_0: \beta_1 = \cdots = \beta_m = 0.$

★ General Linear Tests.

(F) Full model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

(R) Reduced model: $Y_i = \beta_0 + \varepsilon_i$

$SSE(F) = \Sigma (y_i - \hat{y}(F))^2$

$SSE(R) = \Sigma (y_i - b_0)^2$ in reduce model $\hat{y}_i = b_0$.

$= \Sigma (y_i - \bar{y})^2$ $\bar{y} = b_0$.

$= SSTO$

We know $SSE(F) \le \underset{\overset{\shortparallel}{SSTO.}}{SSE(R)}$.

★ General Linear Test:

H_a: If $SSE(F) \ll SSE(R)$

$\Rightarrow$ full model is better. accept H_a.

H_0: If $SSE(F) \approx SSE(R)$.

$\Rightarrow$ Full model is no better than the reduced model.

accept H_0.

✿ test statistics :

$$F^* = \frac{\dfrac{SSE(R) - SSE(F)}{df_R - df_F}}{\dfrac{SSE(F)}{df_F}}$$

reduction in the unexplained variance.

✿ Rej. Region.

$$F^* > F(1-\alpha,\ \underbrace{df_R - df_F}_{1},\ \underbrace{df_F}_{n-2})$$

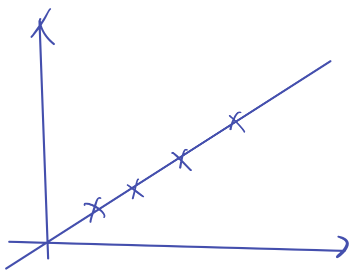
$df_R = n-1$

$df_F = n-2$

✿ R^2 = measure of linear association b/w X & Y.

$$R^2 = \frac{SSTO - SSE}{SSTO} = 1 - \frac{SSE}{SSTO}.$$

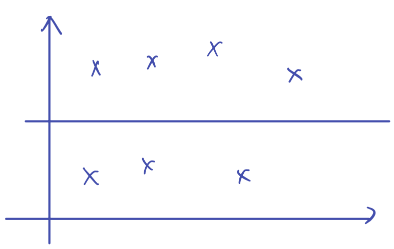
↳ " relative reduction in variation of y by accounting for x ".



$SSE = 0. = \Sigma(y_i - \hat{y}_i)^2$

$\Rightarrow R^2 = 1$

$\beta_1 \neq 0.$

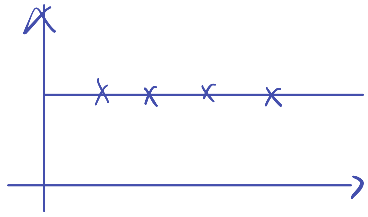


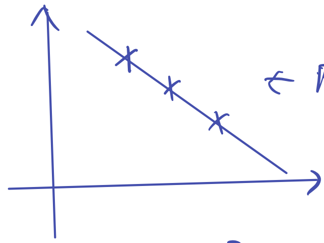
$SSTO = SSE$

$\Sigma(y_i - \bar{y})^2 = \Sigma(y_i - \hat{y}_i)^2$

$R^2 = 0.$

a)





$\leftarrow R^2 = 1$

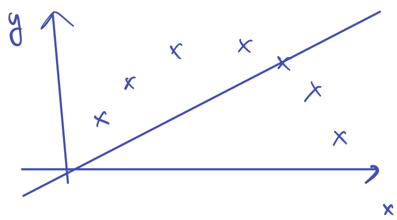
$\beta_1 \neq 0$.

SSTO = 0

$0 \le R^2 \le 1$.

Common misconceptions

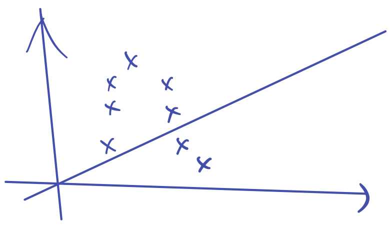
1) High R^2 is a good fit. False!



$R^2 = 0.7$

2) High R^2 = useful predictions.

3) low R^2 = X & y not related.

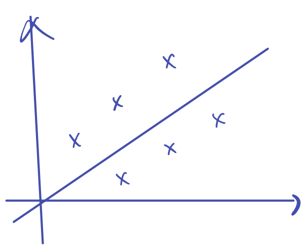


$R^2 = 0.02$

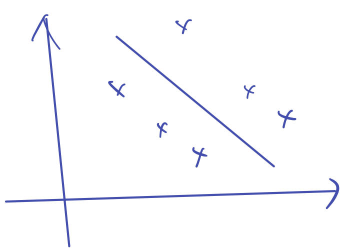
What we can say?

When R^2 is low, X & y not linearly related.

$r = \pm\sqrt{R^2}$ coefficient of correlation



$r = +\sqrt{R^2}$ $-1 \le r \le 1$.



$r = -\sqrt{R^2}$

$r \neq$ slope

has to do with fit.

※ Correlation model of regression

Previously, $Y_i = \beta_0 + \beta_1 \underbrace{X_i} + \varepsilon_i$

$\hookrightarrow$ fixed, known #s.

$Y_1 = X_i$ = temperature outside. Can't control it

$Y_2 = Y_i$ = ice cream sales.

Can we still do regression?

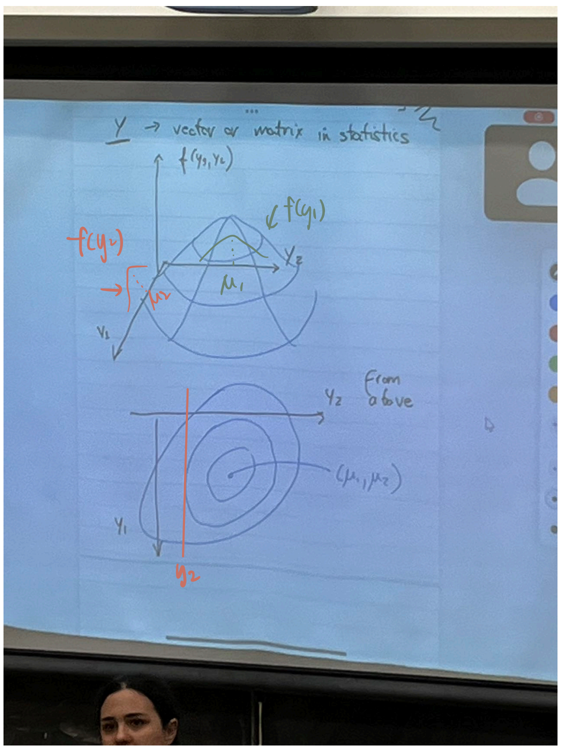
(Y_1, Y_2) are jointly distributed.

$$f(y_1, y_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho_{12}^2}} \exp\left\{ \frac{-1}{2(1-\rho_{12}^2)} \left[\frac{(y_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho_{12}(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2} \right] \right\}$$

$\uparrow$ bivariate normal

$$f(y_1, y_2) = \frac{1}{2\pi (\det \Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\underline{y} - \underline{\mu})^T \Sigma^{-1} \times (\underline{y} - \underline{\mu}) \right)$$

$\underline{y} \rightarrow$ vector or matrix in stats.



Fact: marginal dist of y_1, y_2:

$$f(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2$$

$$f(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1$$

$$\rho_{12} = corr(y_1, y_2) = \frac{\sigma_{12}}{\sigma_1 \cdot \sigma_2}$$

$$\sigma_{12} = cov(y_1, y_2) = \mathbb{E}((y_1 - \mathbb{E}[y_1])(y_2 - \mathbb{E}[y_2]))$$

Properties: $\rho_{12} = \rho_{21}$, $\sigma_{21} = \sigma_{12}$ (symmetric)

$y_1 \perp\!\!\!\perp y_2 \Rightarrow \sigma_{12} = \rho_{12} = 0$,

$\sigma_{12} = \rho_{12} = 0$ & y_1, y_2 jointly normal $\Rightarrow y_1 \perp\!\!\!\perp y_2$

☆ Conditional PDF. $f(y_1 \mid y_2)$.

$$\underbrace{f(y_1 \mid y_2)}_{\text{distribution of } y_1 \text{ given } y_2} = \frac{f(y_1, y_2)}{f(y_2)} \rightarrow \boxed{P(A \mid B) = \frac{P(A \cap B)}{P(B)}.}$$

It could be shown: $y_1 \mid y_2 = y_2 \sim N(\alpha_{1|2} + \beta_{1|2} y_2, \delta_{1|2}^2)$.

$$\alpha_{1|2} = \mu_1 - \mu_2 \rho_{12} \frac{\delta_1}{\delta_2}$$

$$\beta_{1|2} = \rho_{12} \frac{\delta_1}{\delta_2}$$

$$\delta_{1|2} = \delta_1^2 (1 - \rho_{12}^2)$$

Conditional mean of y_1 is linear function of y_2.
Var(y_1) when y_2 is known is constant. } Similar to the simple LR model.

Compare this to $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, ✓ ε_i independent $\Rightarrow$ y_i indep.

✓ $\mathbb{E}(\varepsilon_i) = 0 \Rightarrow \mathbb{E}[y_i] = \beta_0 + \beta_1 X_i$

✓ $Var(\varepsilon_i) = \delta^2 \Rightarrow Var(y_i) = \delta^2$.

① linearity
② normality
③ independence
④ non multicollinearity
⑤ homoscedasticity

☆ Point estimator of ρ_{12}:

ρ_{12} = population correlation coefficient

$$\hat{\rho}_{12}^{MLE} = r_{12} = \frac{\sum_i (y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)}{\sqrt{\left(\sum_i (y_{i1} - \bar{y}_1)^2\right)\left(\sum_i (y_{i2} - \bar{y}_2)^2\right)}}$$

$$(y_{i1}, y_{i2})_{i=1}^{n}$$

Sample Correlation Coeff. or Pearson product moment correlation.

r_{12} biased (unless $\rho_{12} = 0$, or $|\rho_{12}| = 1$)

$-1 \le r_{12} \le 1$.

☆ lecture #7 2023年9月20日(三).

$(y_1, y_2) \sim BN(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho_{12})$

$f(Y_1 | Y_2) = N(\alpha_{1|2} + \beta_{1|2} Y_2, \sigma^2_{1|2})$

$\alpha_{1|2} = \mu_1 - \mu_2 \rho_{12} \frac{\sigma_1}{\sigma_2}$ $\quad\quad$ $\sigma_{1|2} = \sigma_1^2(1-\rho_{12}^2)$.

$\beta_{1|2} = \rho_{12} \frac{\sigma_1}{\sigma_2}$

Compare this to: $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ $\quad$ ε_i independent $\Rightarrow$ y_i indep. ✓

$\mathbb{E}(\varepsilon_i) = 0 \Rightarrow \mathbb{E} y_i = \beta_0 + \beta_1 X_i$ ✓

$Var(\varepsilon_i) = \sigma^2 \Rightarrow Var(y_i) = \sigma^2$ ✓

point estimators of ρ_{12}: $\rightarrow$ population correlation coefficient.

$(y_{1i}, y_{2i})_{i=1}^n$

$$\hat{\rho}_{12}^{MLE} = r_{12} = \frac{\sum_i (y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)}{\sqrt{[\sum_i (y_{i1} - \bar{y}_1)^2][\sum_i (y_{i2} - \bar{y}_2)^2]}}$$

$\downarrow$

Sample correlation coefficient or Pearson product moment correlation.

→ r_{12} is biased (unless $\rho_{12} = 0$, or $|\rho_{12}| = 1$)

→ $-1 \le r_{12} \le 1$.

Hypothesis Testing on ρ_{12}.

H_0: $\rho_{12} = 0$

H_a: $\rho_{12} \neq 0$. (r_{12} is unbiased). $\Rightarrow$ is equivalent to

$\beta_{12} = \rho_{12} \cdot \frac{\sigma_1}{\sigma_2}$

$H_0: \beta_{12} = 0$

$H_a: \beta_{12} \neq 0.$

test stats:

$$t^* = \frac{r_{12}\sqrt{n-2}}{\sqrt{1-r_{12}^2}}$$

under H_0, $t^* \sim t(n-2)$ $\Rightarrow$ rejection region $|t^*| > t(1-\frac{\alpha}{2}, n-2)$

Interval Estimation of ρ_{12}

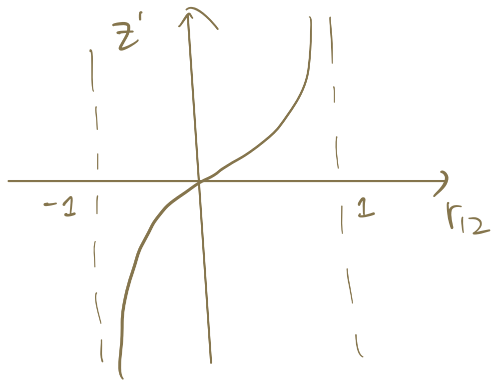
Support of the distribution of r_{12} is $[-1, 1]$.

transform $[-1, 1] \longrightarrow [-\infty, \infty] \longrightarrow$ use some asymptotics. (CLT)

transform back to $(-1, 1] \longleftarrow$ get confidence Interval

Fisher z-transform

$$z' = \frac{1}{2}\log\left(\frac{1+r_{12}}{1-r_{12}}\right)$$



For large n

$\hookrightarrow (n \geq 25)$

$$z' \approx N(\zeta, \frac{1}{n-3})$$

zeta

$$\zeta = \frac{1}{2}\log\left(\frac{1+\rho_{12}}{1-\rho_{12}}\right)$$

Standardize z': $\frac{z'-\zeta}{SD(z')}$ where $SD(z') = \sqrt{\frac{1}{n-3}}$

$\approx N(0,1)$

$CI_{1-\alpha}$ for ζ: $z' \pm z(1-\frac{\alpha}{2})SD(z')$ $(*)$

inv $\rho_{12} = \frac{e^{2\zeta}-1}{e^{2\zeta}+1} = \frac{\sinh \cdot \zeta}{\cosh \cdot \zeta} = \tanh(\zeta)$

$(*) \Rightarrow CI_{1-\alpha}$ for ρ_{12}: $[\tanh(z' - z(1-\frac{\alpha}{2})(\frac{1}{\sqrt{n-3}})), \tanh(z' + z(1-\frac{\alpha}{2})(\frac{1}{\sqrt{n-3}}))]$

Analog of CI for β_1.

Simple LR model

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$\mathbb{E} y_i = \beta_0 + \beta_1 X_i$$

Assumptions:

$$Var(y_i) = Var(\varepsilon_i) = \sigma^2$$

Hypothesis Testing:

$H_0: \beta_1 = 0$

$\beta_1 \neq 0.$

Is Y linearly related with X. $\Rightarrow$

CI for β_1:

$$t^* = \frac{b_1}{s(b_1)} \sim t(n-2)$$

$$\beta_1 = b_1 \pm t(1-\alpha, n-2) * s(b_1)$$

Corr. Model.

$(X_i, y_i) \sim BN(\underline{\mu}, \underline{\Sigma})$ $\rightarrow$ vector sign.

$\hookrightarrow$ bivariate normal

$$\mathbb{E}(y_i \mid X_i) = \alpha_{1|2} + \beta_{1|2} X_i$$

$$Var(y_i \mid X_i) = \sigma_{1|2}^2$$

$$\beta_{1|2} = \rho_{1|2} \frac{\sigma_1}{\sigma_2}$$

Hypothesis testing:

$H_0: \beta_{1|2} = \rho_{1|2} = 0$

$H_a: \beta_{1|2}, \rho_{1|2} \neq 0$

CI for $\rho_{1|2}$:

$$t^* = \frac{r_{12}\sqrt{n-2}}{\sqrt{1-r_{12}^2}},$$

$$r_{12} = \frac{\Sigma(X_i - \bar{X})(y_i - \bar{y})}{\sqrt{S_{xx} \cdot S_{yy}}}$$

$\hookrightarrow$ Fisher z-Transform.

※ All "ρ_{12}" are "$\rho_{1|2}$".

☆ section #4 2023年9月22日 (2).

☆☆ Traditional ANOVA ⇒ analysis of variance.

$H_0: \mu_1 = \mu_2 = \cdots = \mu_K$

☆☆ $SST_0 = SSR + SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$

Euclidean norm for $v \in \mathbb{R}^d$:

$\|v\| = \sqrt{\sum_{j=1}^d v_j^2} = \sqrt{v^T v}$

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

$= (\underline{Y} - \underline{\bar{Y}})^T (\underline{Y} - \underline{\bar{Y}})$

$= \|\underline{Y} - \underline{\bar{Y}}\|^2$

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

$= \|\underline{\hat{Y}} - \underline{\bar{Y}}\|^2$

$= \|\underline{Y} - \underline{\hat{Y}}\|^2$

will prove $\bar{\hat{Y}} = \bar{Y}$ in HW.

$\underline{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ $\quad \underline{\bar{Y}} = \begin{pmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix}$

(nx1) vector

$\underset{n\times 1}{Y} = \underset{n\times p}{\underline{X}} \underset{p\times 1}{\underline{\beta}} + \varepsilon \quad \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \quad \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$.

☆☆ want to show $\sum_{i=1}^n e_i (\hat{Y}_i - \bar{Y}) = 0$

$\sum_{i=1}^n e_i \hat{Y}_i - \bar{Y} \underbrace{\sum_{i=1}^n e_i}_{= 0 \text{ (showed in class)}}$

$= \sum_{i=1}^n e_i (b_0 + b_1 x_i)$

$= \underbrace{b_0 \sum_{i=1}^n e_i}_{=0} + b_1 \underbrace{\sum_{i=1}^n e_i x_i}_{=0}$

(!) β_0 and β_1 are unknown.

$\hat{Y}_i$ is known, thus use b_0, b_1.

b_0 and b_1 are random variable.

That's why we find Expectation of them.

(!) $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

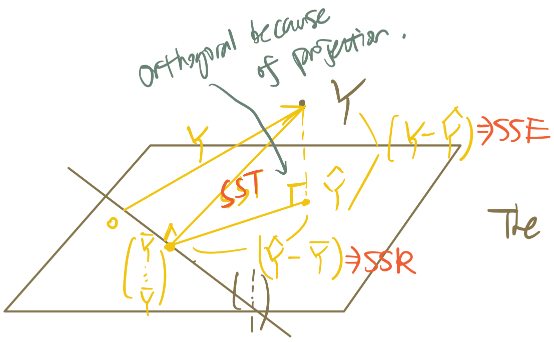
$Y_i = \hat{Y}_i + e_i$

$= b_0 + b_1 X_i + e_i$

☆☆ $\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2$

$= \sum_{i=1}^n [(Y_i - \hat{Y}_i)^2 + 2(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) + (\hat{Y}_i - \bar{Y})^2]$

$= \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{SSE} + \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{SSR} + 2\sum_{i=1}^n \underbrace{(Y_i - \hat{Y}_i)}_{= e_i} \underbrace{(\hat{Y}_i - \bar{Y})}_{= 0}$



The plane is Range $(X) = \{Y : AX = Y \text{ for all } X \in \mathbb{R}^n\}$.

Line is in the plane

Point Y is not in the plane. we want to ~~project it~~.

Recall proj $\vec{a}$ to $\vec{b}$ is $\frac{a^T b}{\|b\|^2} b$.

$\therefore \underbrace{\frac{Y^T \mathbf{1}}{n}}_{= \bar{Y}} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

$\min_{\hat{\beta}} \|\hat{Y} - Y\|^2$

$\hat{Y}$ is in the plane range(X).

$\therefore$ By the graph we can understand

$SST = SSR + SSE$

$(Y - \bar{Y})^2 = (\hat{Y} - \bar{Y})^2 + (Y - \hat{Y})^2$

☆☆ Review.

General liner test : used to compare two models, which one is true.

Compare SSE, choose the one with minimum SSE.

Set up : Simple Linear

Full model : $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

$\cup$

Reduced model : $Y_i = \beta_0 + \varepsilon_i$

$\uparrow$

$\min_{\beta_0} SSE.$

$= \arg\min_{\beta_0} \sum_{i=1}^{n} (y_i - \beta_0)^2$

$= \bar{Y}.$

$\begin{cases} SSE(\text{Full}) & \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\ SSE(\text{Reduced}) & \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \end{cases}$

H_1 : Reduced model is True. (i.e. $\beta_1 = 0$)

$SSE(\text{full}) \le SSE(\text{reduced})$

Test stats : $F^* = \dfrac{SSE(R) - SSE(F)}{df_R - df_F} \Big/ \dfrac{SSE(F)}{df_F}$

reject for large F^*

413/613 ASDA: Diagnostics and Remedial Measures

Fall 2023

Chapter 3. Diagnostics and Remedial Measures.

```
## Loading required package: leaps
```

```
## Loading required package: SuppDists
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

Diagnostics for Predictors → X_i

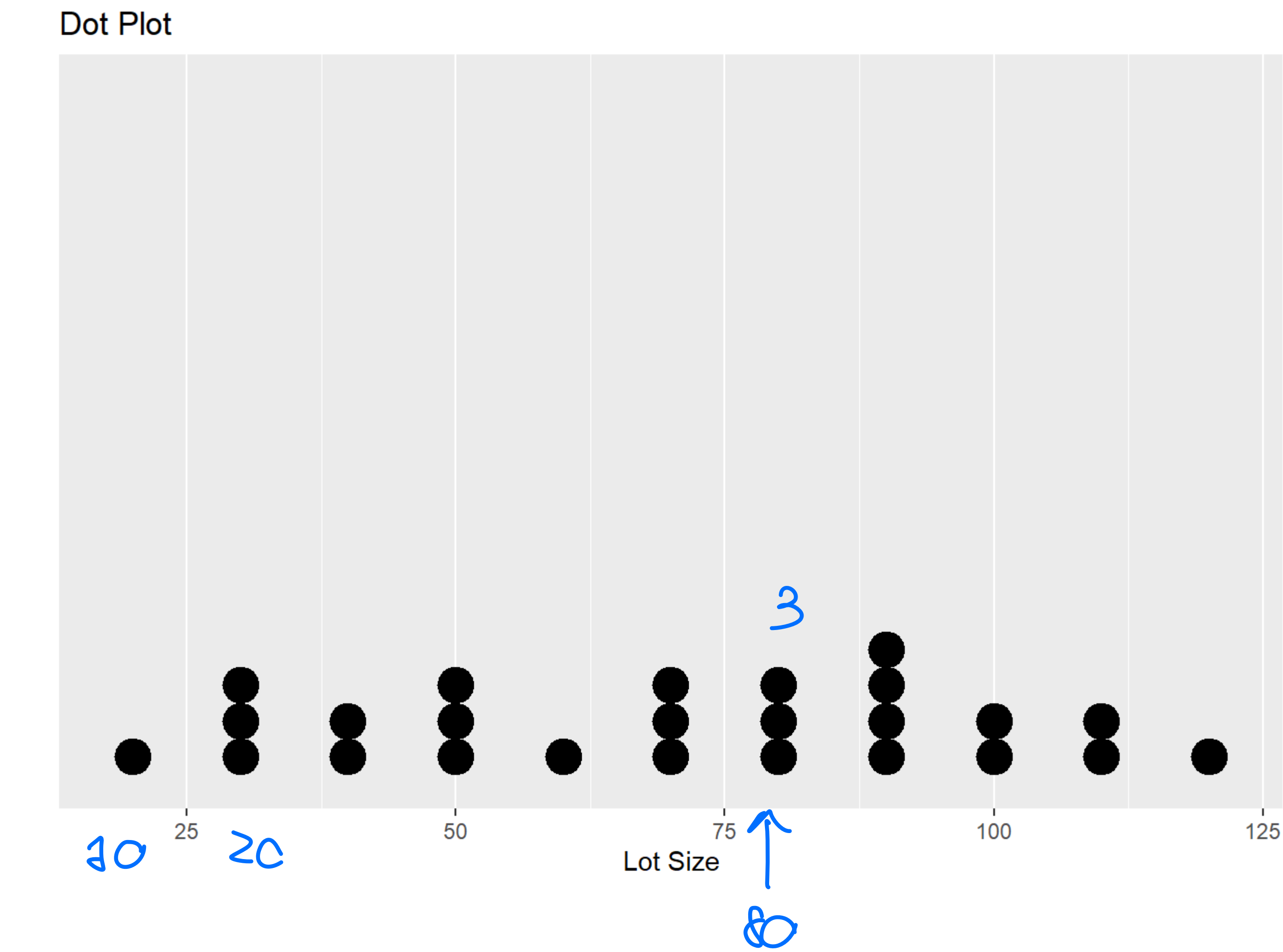
Dot plot, sequence plot, stem-and-leaf plot, box plot). This way one can check if there is anything wrong with the predictor variable

~~**Stem-and-leaf plot**~~ Toluca Company, KNNL

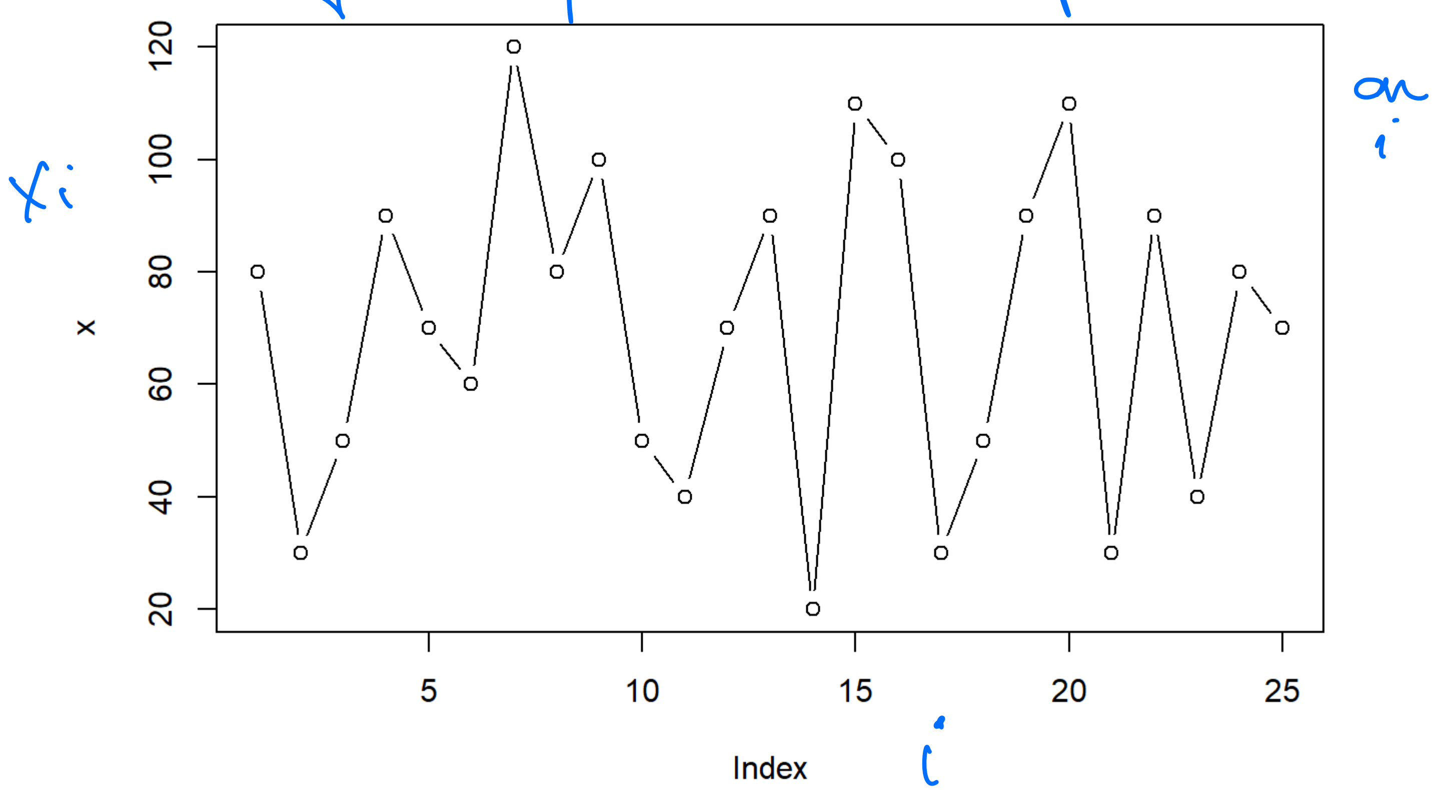
X_i =

```
##  [1]  80  30  50  90  70  60 120  80 100  50  40  70  90  20 110 100  30  50  90
## [20] 110  30  90  40  80  70
```

```
## Bin width defaults to 1/30 of the range of the data. Pick better value with
## `binwidth`.
```



Sequence plot = to see dependence



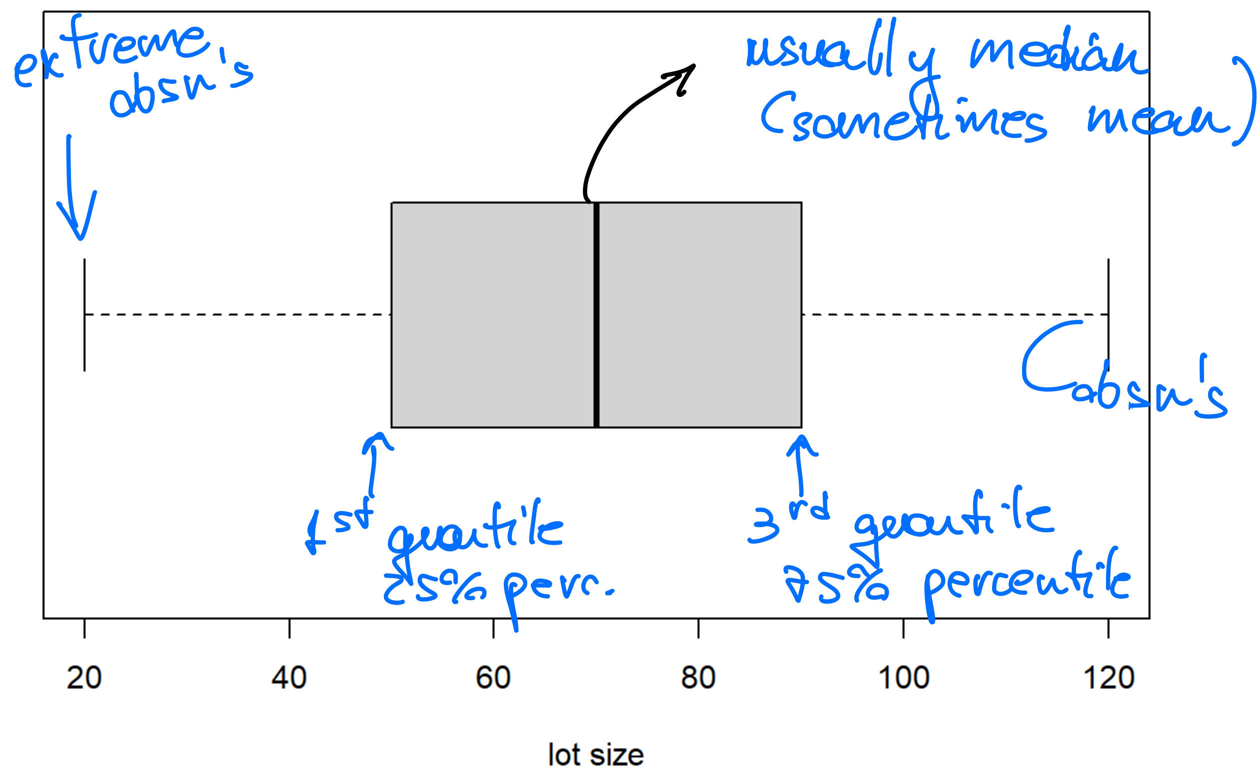
```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##    2 | 0
##    3 | 000
##    4 | 00
##    5 | 000
##    6 | 0
##    7 | 000
##    8 | 000
##    9 | 0000
##   10 | 00
##   11 | 00
##   12 | 0
```

$x_i = 20 \rightsquigarrow 1$ obs. [20]

[30, 30, 30, 30]

3|0001 [30, 30, 30, 30, 31]

Box plot



Section 3.2. Residuals.

Residuals and semi-studentized residuals will be used.

$$e_i = Y_i - \hat{Y}_i$$

$\hat{Y}_i = b_0 + b_1 X_i$

$$e_i^* = \frac{Y_i - \hat{Y}_i}{\sqrt{MSE}}$$

1) $E\, Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
2) $\operatorname{var} \varepsilon_i = \sigma^2$
3) ε_i independent
4) $\varepsilon_i \sim N(0, \sigma^2)$

Issues with the Model.

1. Non-linear function.
2. Non-constant variance.
3. Not independent error terms.
4. Outliers.
5. Error terms are not normally distributed.
6. Missing predictors.

1) $\hat{Y}_i = b_0 + b_1 X_i$
* 2) $\operatorname{var} e_i \approx \sigma^2$
** 3) e_i independent
4) $e_i \sim N(0, \sigma^2)$

Section 3.3 Diagnostics for Residuals

Possible plots one can make

a. e_i vs X_i
b. $|e_i|$ or e_i^2 vs X_i
c. e_i vs $\hat{Y}_i$
d. e_i vs time
e. e_i vs other predictors
f. Normal plot of e_i's (QQ-plot)

$e_i = \hat{\varepsilon}_i$, $\varepsilon_i = Y_i - E(Y_i)$

$E(\varepsilon_i) = 0$, $\operatorname{var} \varepsilon_i = \sigma^2$

*2) $\operatorname{var} e_i = \sigma^2 (1 - h_{ii})$

→ quantify from future

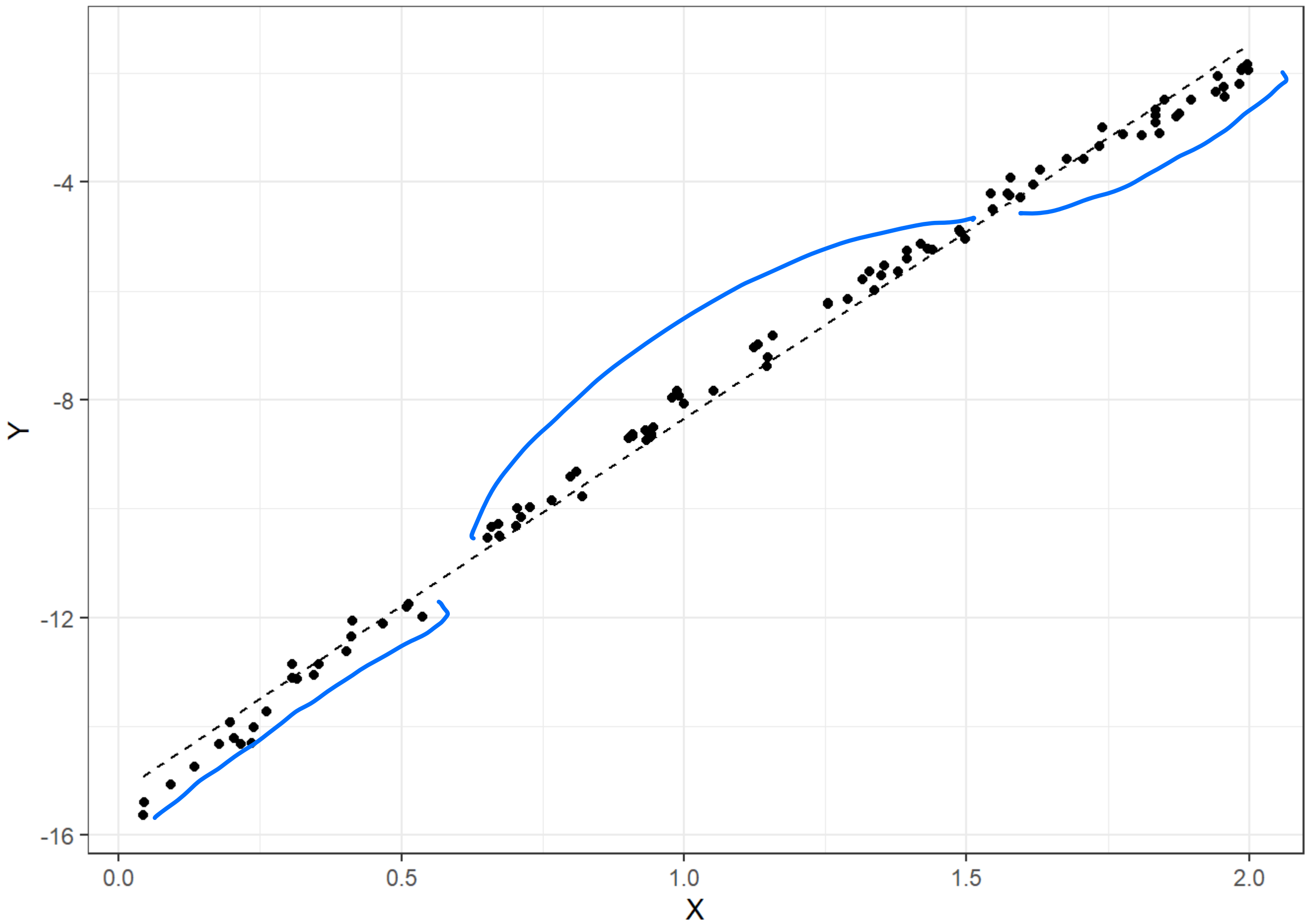
*3) e_i are not independent

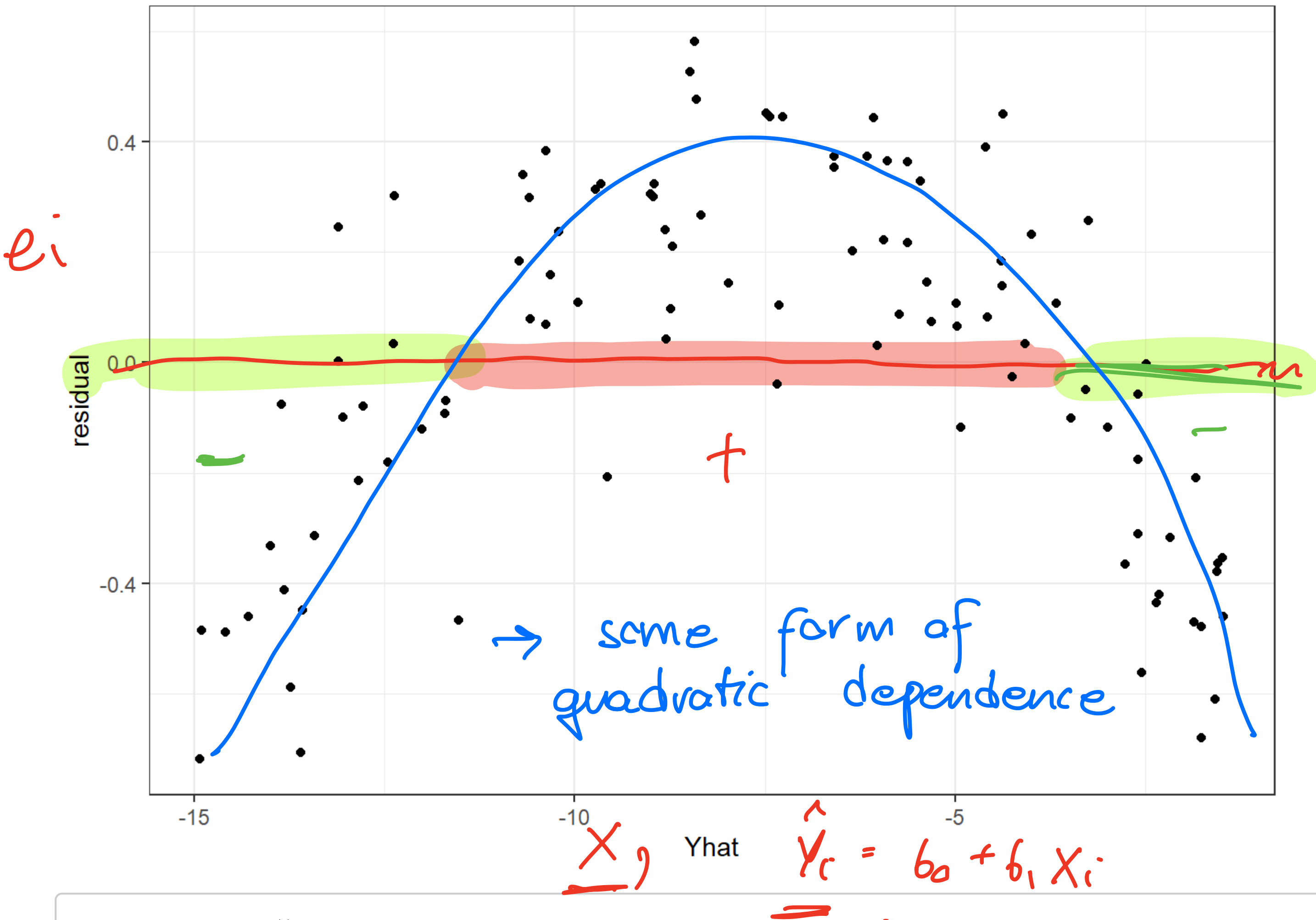
$e_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X_i$

$b_0 = \sum \eta_i Y_i \quad b_1 = \sum k_i Y_i$

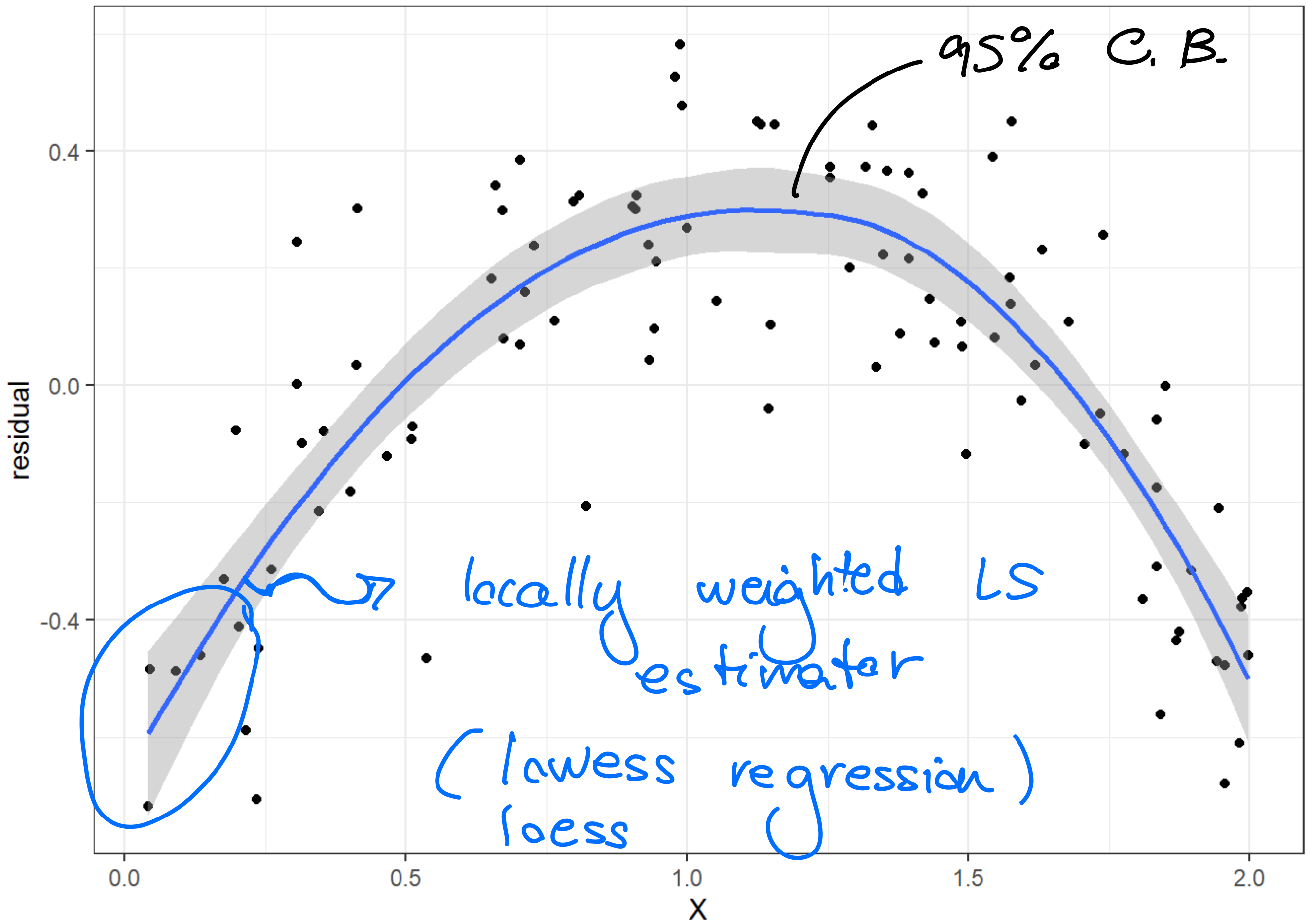
as $n \to \infty \Rightarrow \operatorname{cov}(e_i, e_j) \to 0$

Issue 1: nonlinear regression function






```
## `geom_smooth()` using formula = 'y ~ x'
```

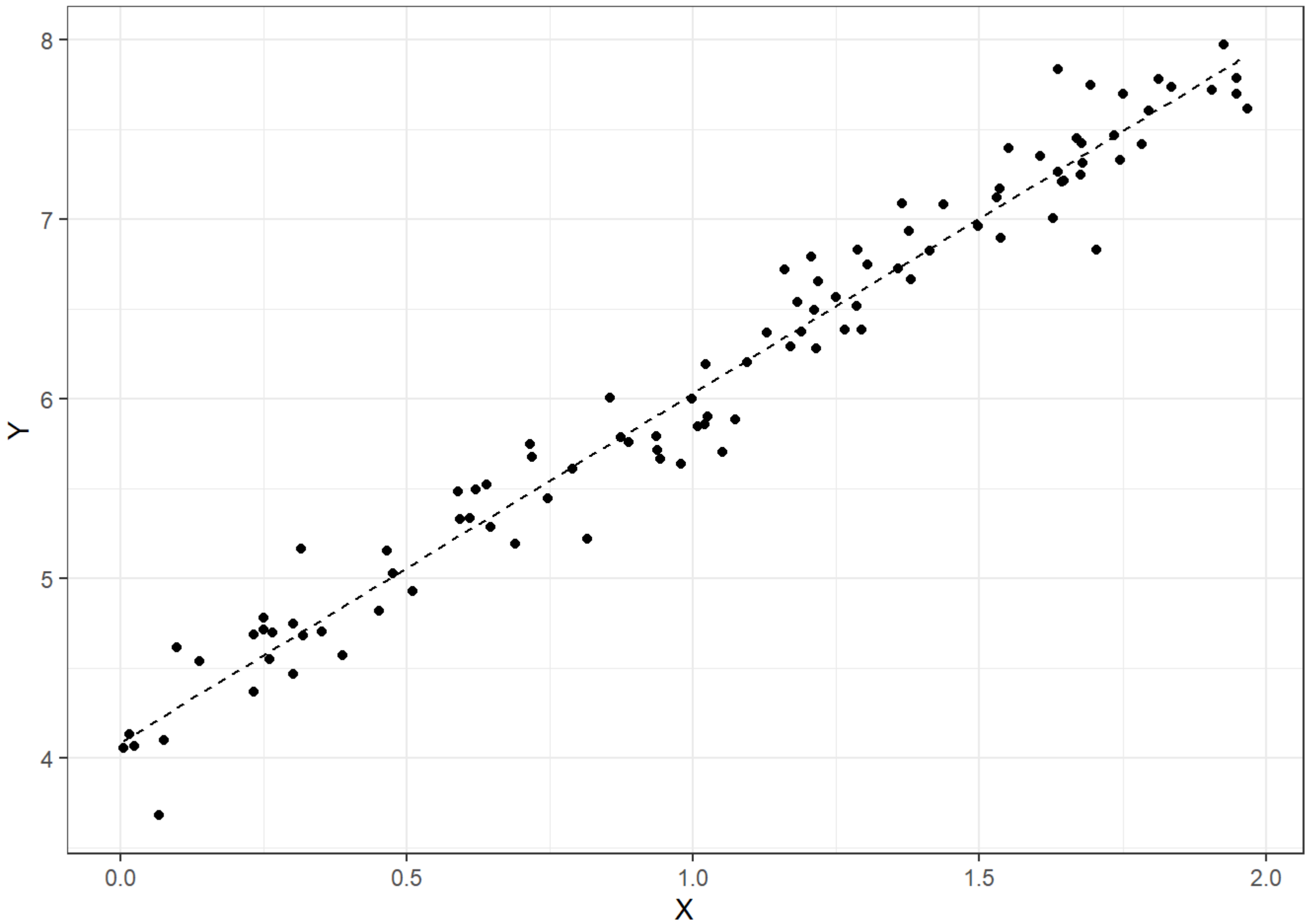


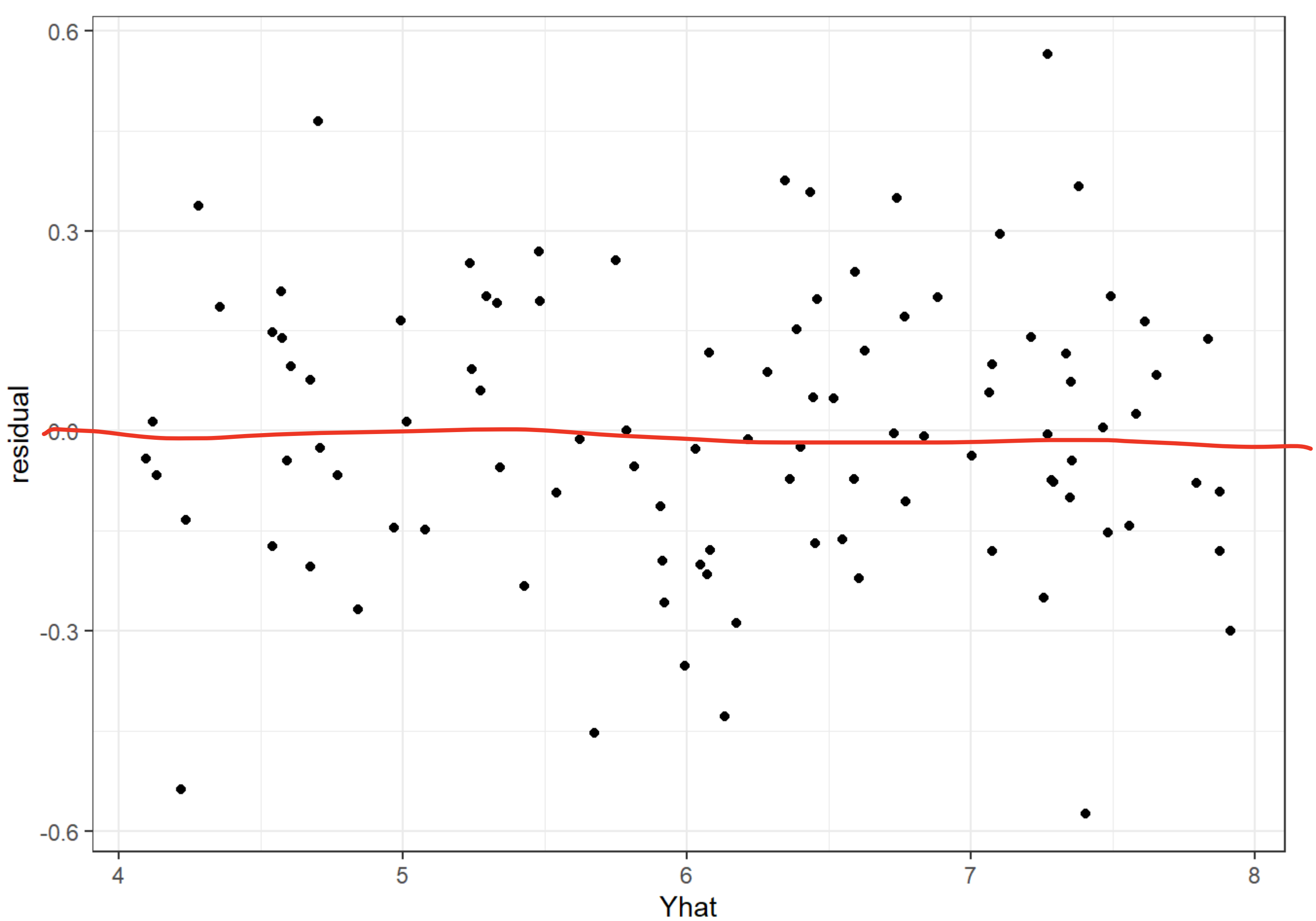
$e_i \propto a + bx + cx^2$

$Y_i = \beta_0 + \beta_1 X + \varepsilon_i$ quadratic

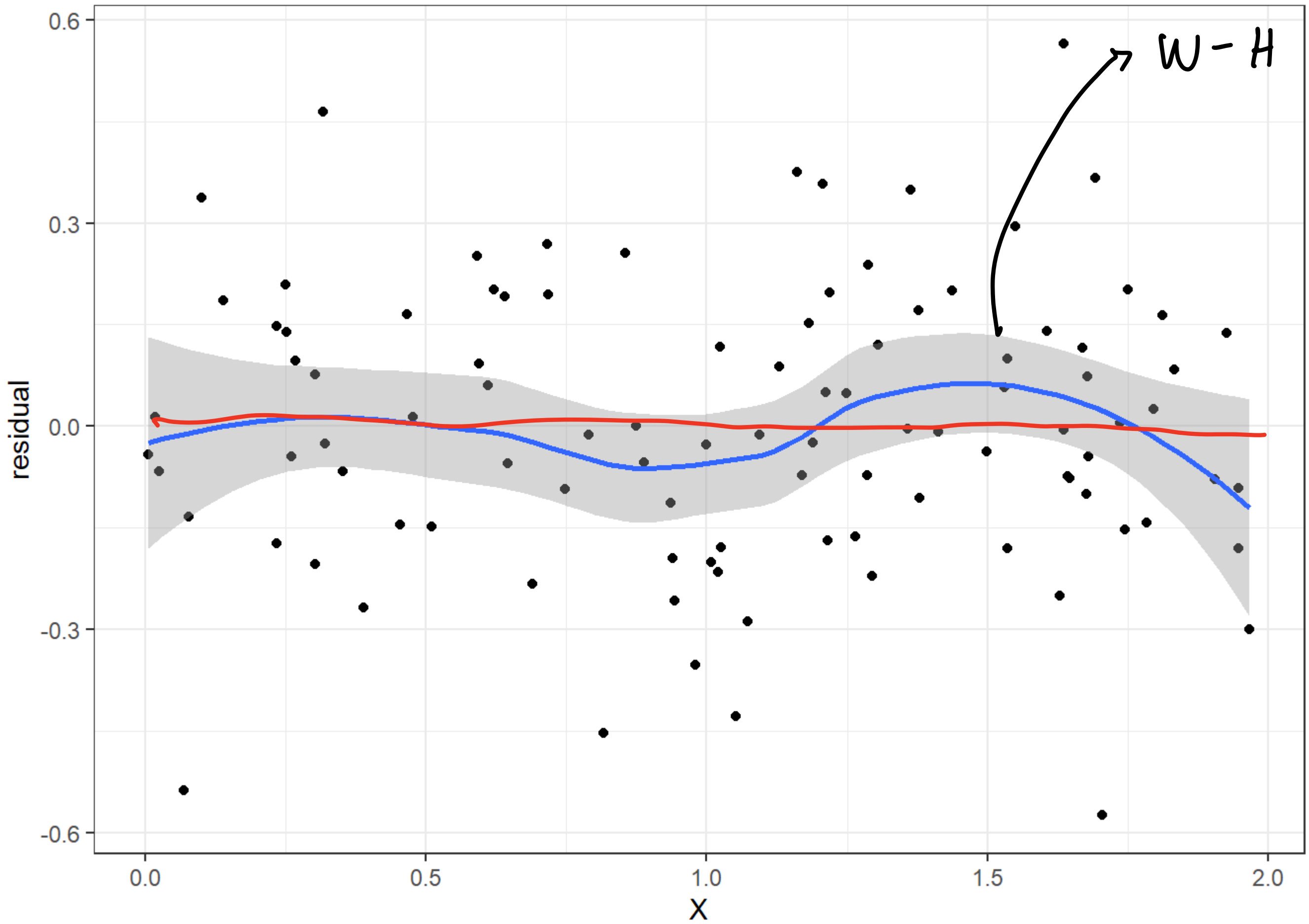
need to change the model!

$Y_i = \beta_0 + \beta_1 X + \beta_{11} X^2 + \varepsilon_i$


8
7
6
5
4
Y
0.0
0.5
1.0
1.5
2.0
X


0.6
0.3
0.0
-0.3
-0.6
residual
4
5
6
7
8
Yhat

```
## `geom_smooth()` using formula = 'y ~ x'
```

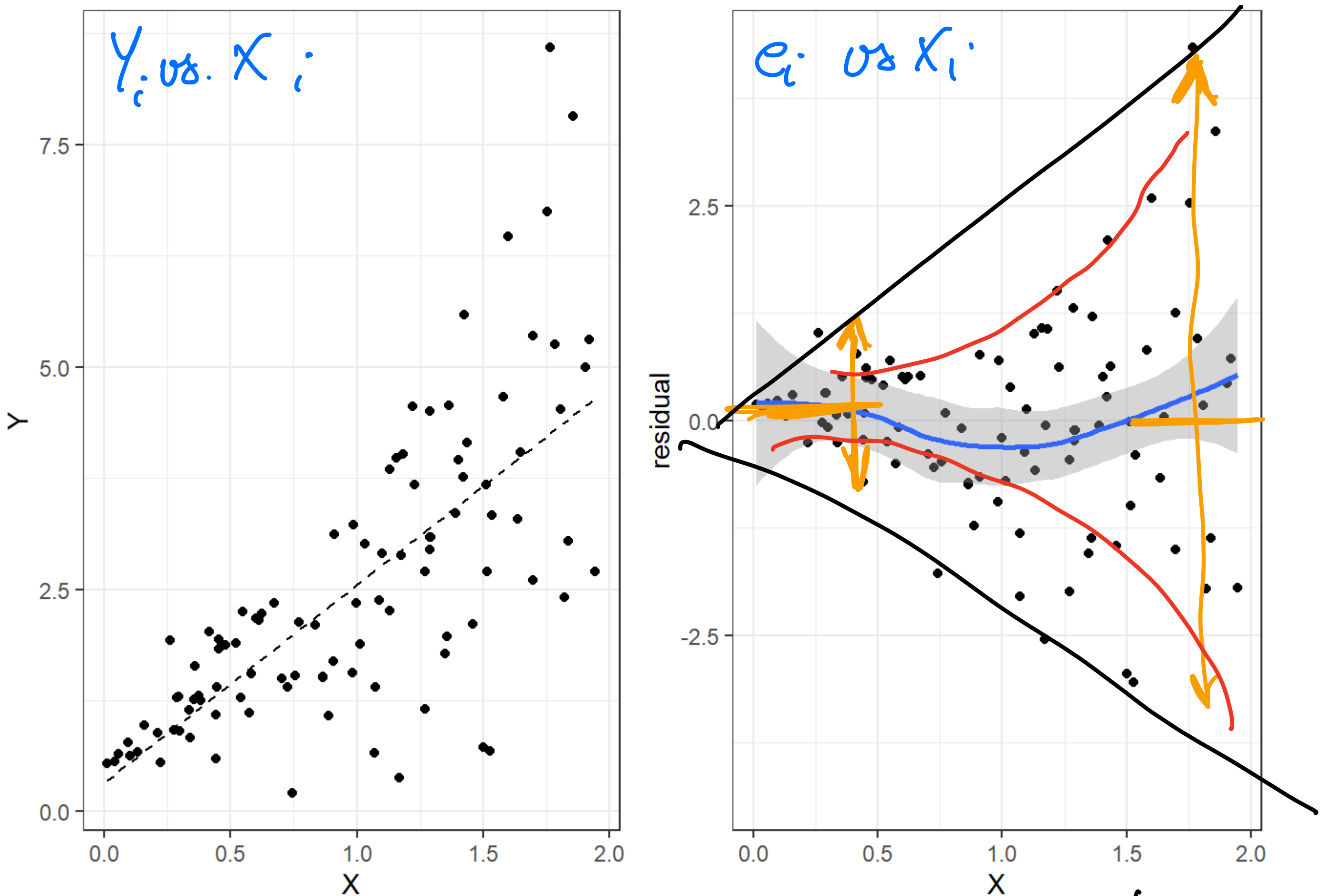


Issue 2: Heteroscedasticity

$Var(e_i) = \sigma^2$) → homoscedastic

```
## `geom_smooth()` using formula = 'y ~ x'
```

"variance is unequal"



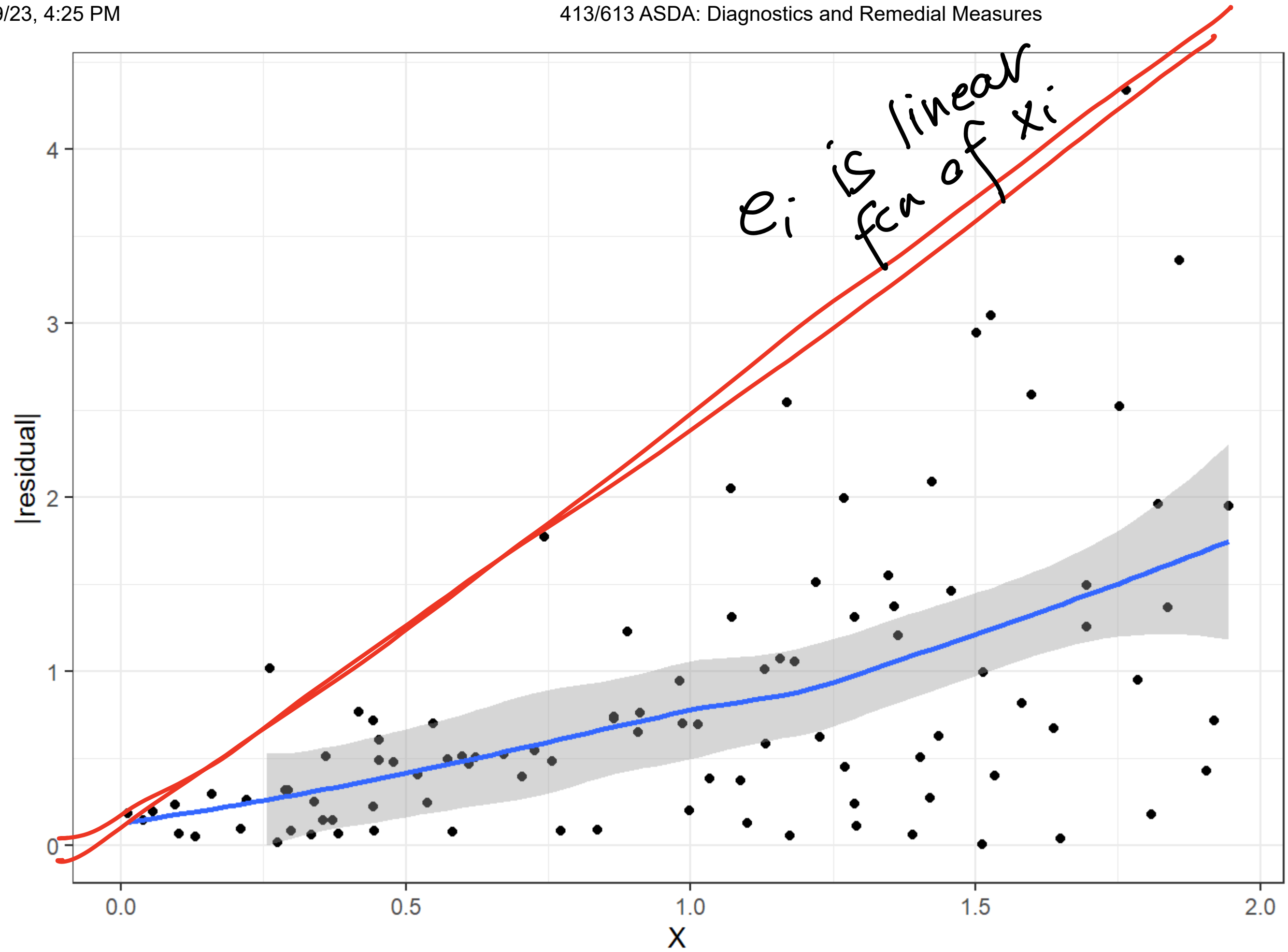
a=1

This is an example where the error does not have a constant variance. The true model is $Y = 0.5 + 2 * X + \varepsilon$, with $\varepsilon \sim N(0, X^2)$. sd = X

The least square regression line is the dashed line in the left plot. The plot of the residual showed that the variance of the ε_i might not be constant.

It might be clearer if we plot $|e_i|$ versus X_i

```
## `geom_smooth()` using formula = 'y ~ x'
```



Issue 3: Dependent Error Terms

```
##
## Attaching package: 'lawstat'
```

```
## The following object is masked from 'package:car':
##
##     levene.test
```

$e_i \sim N(0, \sigma(x_i))$

☆ Lecture #8 2023年9月25日 (一).

☆☆ Variance Stabilizing Transform.

→ to fix heteroscedasticity.

$sd(e_i) \approx sd(\varepsilon_i) = sd(Y_i)$ → want this.

→ $sd(y_i)$ changed linearly with X_i (or $\hat{Y}_i$).

Idea: find $h(Y)$ s.t. $Y' = h(Y)$ is a new responsive variable s.t.

$$Y_i' = \beta_0' + \beta_1' X_i + \underbrace{\varepsilon_i'}_{\sim N(0,\sigma^2)}$$

with more stable variance $\approx Var(e_i) = \sigma^2, \forall i$

→ $Var(h(Y)) \approx \underline{Var(h(EY)) + h'(EY)(Y - EY)}_{(*)}$. Taylor Expansion.

$h(Y) = h(Y_0) + h'(Y_0)(Y - Y_0) + \cdots$

$$(*) = Var(h'(EY)(Y-EY)) = (h'(EY))^2 Var(Y - EY)$$
$$= (h'(EY))^2 Var(Y).$$

$\therefore Var(h(Y)) \approx (h'(EY))^2 Var(Y)$

→ How to find $(h'(EY))^2$?

Find $(h'(EY))^2 \propto \frac{1}{Var(Y)}$.

We want $Var(h(Y)) = k$

↳ $Var(h(Y)) \approx (h'(EY))^2 Var(Y)$

Need $(h'(EY))^2 = k \cdot \frac{1}{Var(Y)}$

Then $Var(h(Y)) \approx k \cdot \frac{1}{\cancel{Var(Y)}} \cancel{Var(Y)} = k.$

$$(h'(E(y))) \propto \sqrt{\frac{1}{Var(y)}} = \frac{1}{sd(y)}$$

$$\Rightarrow h(y) = \int \frac{dy}{sd(y)} = \int \frac{1}{\sqrt{Var(y)}}\, dy$$

In our example, $sd(y) \approx y$

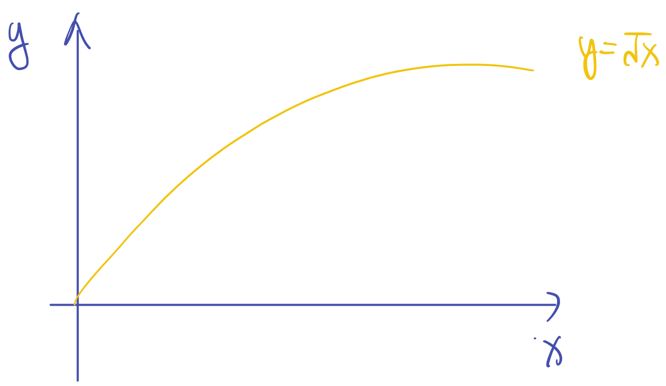
$$\therefore h(y) = \int \frac{1}{y}\, dy = \log(y)$$

$\therefore$ $\underline{\log(y_i) = y_i' = \beta_0' + \beta_1' X_i + \epsilon_i}$

we do this to make the variance constant.

After transforming, re-run equality of variance test (Brown-Forsythe).

☆☆ Box-Cox transform



option 1) transform y

$$y' = y^2 = (\sqrt{x})^2 \cong x.$$

option 2) transform x

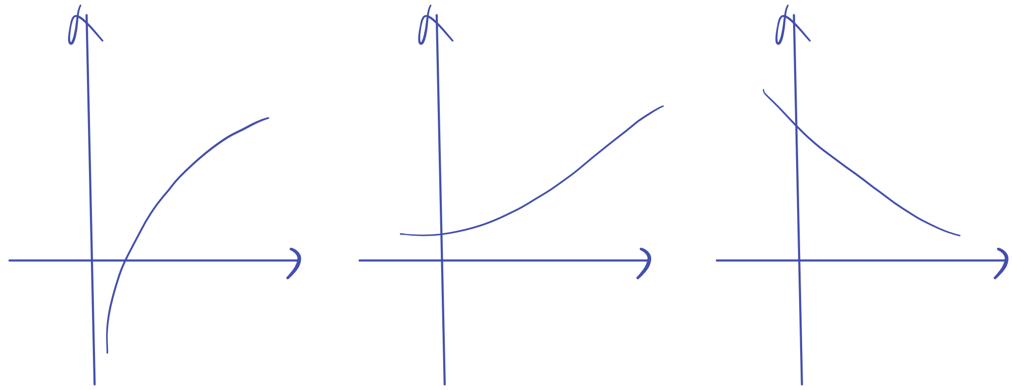
$$x' = \sqrt{x} \;;\; y \approx x'.$$

① $\underline{\text{use option 2}}$ because we already have constant variance for y.

If the variance of y is not constant, then we prefer to transform y.

option 3) transform both y and x.

Case: Already have linear fit of X & Y, but need to change Y to fix the Var(Y), however, it will change the linear fit, thus chang both X & Y to make sure Var(Y) is constant and X & Y stays linear fit.

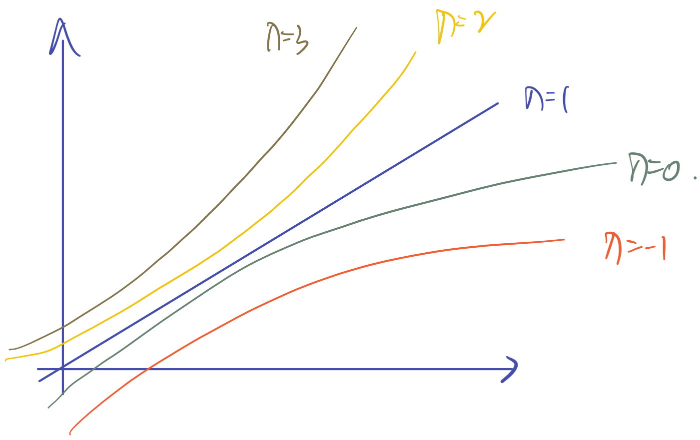


X transform	$\log x$, $\sqrt{x}$, $\log_{10} x$	x^2, $\exp(x)$	$\frac{1}{x}$, $\exp(-x)$
y transform	$\exp(y)$, y^2, 10^y	$\sqrt{y}$, $\log y$	$\frac{1}{y}$, $\log(y)$.

☆☆ Box-cox transform

☆☆☆ For y:

$$y' = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, & \lambda \neq 0. \\ \log y, & \lambda = 0. \end{cases}$$



→ λ is chosen from $\{\pm \frac{n}{2}\}$, $n \in \mathbb{N}$.

λ is chosen via Maximum Likelihood Estimation.

$$y_i^\lambda = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$y_i^\lambda - \beta_0 - \beta_1 X_i \sim N(0, \sigma^2)$$

Midterm
Lecture 1–8.5. HW#1,2.
see 1-4.

Lecture #9 2023年9月27日(三).

Joint estimation of β_0, β_1.

Idea: find a regression $R \subseteq \mathbb{R}^2$ s.t. $P((\beta_0, \beta_1) \in R) = 1-\alpha$.

∘) $CI^0_{1-\alpha}$: $b_0 \pm t(1-\frac{\alpha}{2}; n-2)\, sd(b_0)$.

∘) $CI^1_{1-\alpha}$: $b_1 \pm t(1-\frac{\alpha}{2}; n-2)\, sd(b_1)$.

Define events: event #1 : $A_1 = \{\beta_0 \notin CI^0_{1-\alpha}\}$

event #2 : $A_2 = \{\beta_1 \notin CI^1_{1-\alpha}\}$.

$P(A_1) = \alpha$

$P(A_2) = \alpha$.

We're interested in $A_1^c \cap A_2^c$ (both $\beta_0 \in CI^0$, $\beta_1 \in CI^1$).

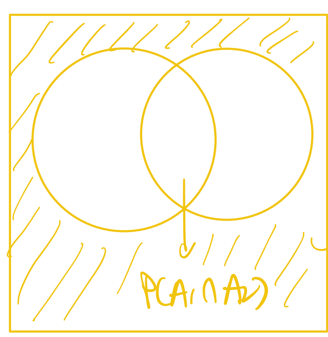
Wrong Derivation Example:

$$P(A_1^c \cap A_2^c) = P(A_1^c)\, P(A_2^c).$$
$$= (1-\alpha)^2$$

if $\alpha = 0.05$, $1-\alpha = 0.95$, $P(A_1^c \cap A_2^c) = 0.9025 < 0.95$

Fixed:

$$P(A_1^c \cap A_2^c) = 1 - P(A_1 \cup A_2).$$
$$\geq 1 - (P(A_1) + P(A_2))$$
$$= 1 - 2\alpha$$
$$= 0.9.$$



$P(A_1^c \cap A_2^c)$.

$P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$.

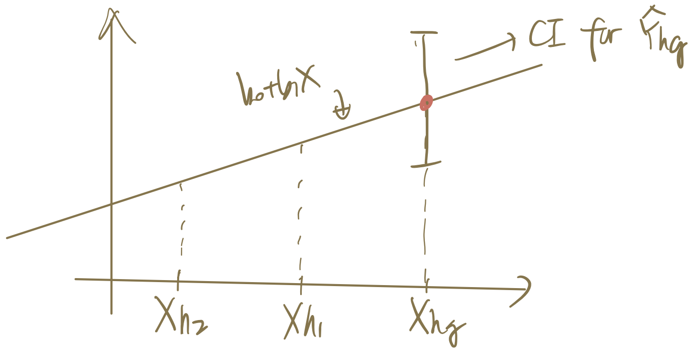
The joint CI has at least 90% significance.

$1 - 2\alpha = 0.95 \Rightarrow \alpha = 0.025$

General Rule : $\begin{cases} b_0 \pm t(1-\frac{\alpha}{4}; n-2)\, sd(b_0) \\ b_1 \pm t(1-\frac{\alpha}{4}; n-2)\, sd(b_1) \end{cases}$

joint CI for β_0 & β_1 with at least $(1-\alpha)$

☆☆ Simultaneous estimation of g mean responses.



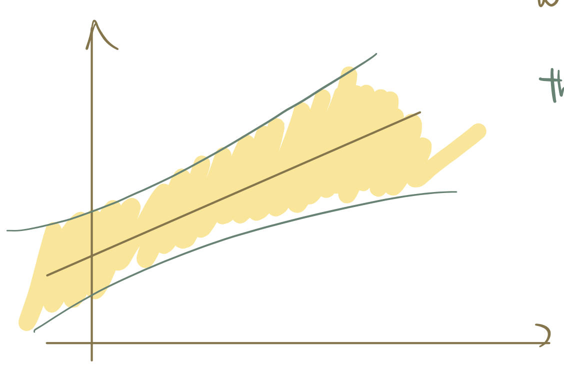
Bonferroni procedure: $\hat{Y}_h \pm B \cdot SD(\hat{Y}_h)$.

$B = t(1-\frac{\alpha}{2g}; n-2)$. $\quad 1-\frac{\alpha}{2g} = 0.95$

Working-Hoteling confidence band: $\hat{Y}_h \pm W \cdot sd(\hat{Y}_h)$

$W = \sqrt{2F(1-\alpha; 2, n-2)}$

the # of observations doesn't matter.



☆☆ Predict g responses. (predictions).

Bonferroni procedures.

CI for Y_h : $\hat{Y}_h \pm B \cdot sd(pred)$

$B = t(1-\frac{\alpha}{2g}; n-2)$.

g: # of observations.

Scheffe procedure ; $\hat{Y}_h \pm S \cdot sd(pred)$

$S = g \cdot F(1-\alpha; g; n-2)$. without derivation.

similar to W-H.

End for Midterm #1.

☆ (Chap 5) Matrix approach to LR.

$$\underline{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \quad \underset{\text{row}}{n} \times \underset{\text{col.}}{n}$$

Vectors : $(n \times 1)$ matrices.

$$\underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$A^T \rightarrow$ transpose : swap rows & cols.

Matrix $\underline{C}$, $\mathbb{E}(\underline{C}) = [\mathbb{E}(c_{ij})]$, $\underline{y}$, $\mathbb{E}\underline{y} = \begin{bmatrix} Ey_1 \\ \vdots \\ Ey_n \end{bmatrix}$

·) dot product $\underline{y} \cdot \underline{y}$ $(\langle y, y \rangle)$
(scalar)

$$\underline{y} \cdot \underline{y} = y_1^2 + y_2^2 + \cdots + y_n^2 = \|\underline{y}\|^2$$

$\|\underline{y}\|$: norm / length / magnitude of $\underline{y}$.

$\underline{u}, \underline{v} \in \mathbb{R}^n$, $\underline{u} \cdot \underline{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$.