L3. Simple Linear Regression

- Prediction Intervals and Confidence Intervals

- Check Model Assumptions

1. Prediction Intervals and Confidence Intervals

a) Point Prediction

we have mentioned that the fitted value given $X = x_0$ (new value of x)

$$\hat{y}|_{x=x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

gives the point estimate/prediction of

(i) $E(y|_{x=x_0}) = \beta_0 + \beta_1 x_0$ - Average of y for the whole population with $X = x_0$

(ii) $y|_{x=x_0} = \beta_0 + \beta_1 x_0 + \varepsilon_0$ - Actual value of y given $X = x_0$

The point estimates are the same, but the predictive varianes are different.

b) Confidence interval for estimating $E(y|x_0)$

To make a difference in notation, we denote the estimate as

$$\hat{\mu}_{y|x_0} = \widehat{E(y|x_0)} = \hat{\beta}_0 + \hat{\beta}_1 X_0$$

Then

(i) $E(\hat{\mu}_{y|x_0}) = \beta_0 + \beta_1 X_0 = E(y|x_0)$ unbiased.

(ii) look closer at the bias:

$$\hat{\mu}_{y|x_0} - E(y|x_0) = (\hat{\beta}_0 + \hat{\beta}_1 X_0) - (\beta_0 + \beta_1 X_0)$$

$$= (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) X_0$$

(iii) so the variance of the prediction bias is

$$\mathrm{Var}(\text{bias}) = \mathrm{Var}[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) X_0]$$

$$= \mathrm{Var}[(\hat{\beta}_0 + \hat{\beta}_1 X_0) - (\beta_0 + \beta_1) X_0]$$

$$= \mathrm{Var}(\hat{\beta}_0 + \hat{\beta}_1 X_0)$$

$$= Var(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_0)$$

$$= Var(\bar{y} + \hat{\beta}_1 (x_0 - \bar{x}))$$

$$= Var(\bar{y}) + (x_0 - \bar{x})^2 Var(\hat{\beta}_1) + cov(\bar{y}, \hat{\beta}_1) \cdot 2(x_0 - \bar{x})$$

$$= \frac{\sigma^2}{n} + \frac{(x_0 - \bar{x})^2 \sigma^2}{S_{xx}} + 0$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

$cov(\bar{y}, \hat{\beta}_1)$

$= cov(\Sigma \frac{y_i}{n}, \Sigma k_i y_i)$

$= \Sigma \, cov(\frac{1}{n} y_i, k_j y_j)$

$= \Sigma \frac{1}{n} k_j \, cov(y_i, y_j)$

$= \Sigma \frac{1}{n} k_j \sigma^2 = \frac{1}{n} \sigma^2 \Sigma k_j$

$= 0$

(iv) when σ^2 is unknown, estimate it by MSE.

$$\widehat{Var}_{prediction}(\hat{u}) = MSE \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

(v) Confidence interval / Prediction interval for $\hat{u}_{y|x_0} = E(Y|X=x_0)$

$$\frac{\hat{u}_{y|x_0} - E(Y|X=x_0)}{MSE \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]} \sim t(n-2)$$

so the $100(1-\alpha)\%$ C.I. for this case is:

$$\hat{\mu}_{Y|X_0} \pm t_{\alpha/2,\, n-2} \sqrt{MSE\left(\frac{1}{n} + \frac{(X_0-\bar{X})^2}{S_{XX}}\right)}$$

where $\hat{\mu}_{Y|X_0} = \hat{\beta}_0 + \hat{\beta}_1 X_0 = \hat{y}|_{X_0}$

$\boxed{\text{c) Prediction Interval for estimating } y|_{X=X_0}}$

(i) we denote the estimate by

$\hat{y}|_{X_0} = \hat{\beta}_0 + \hat{\beta}_1 X_0$ to estimate the actual

value of y given $X = X_0$. $(y|_{X_0})$

then $E(\hat{y}|_{X_0}) = \beta_0 + \beta_1 X_0 = E(y)|_{X_0}$ unbiased.

(ii) But when we look at the actual bias.

$$\hat{y}|_{X_0} - y|_{X_0} = (\hat{\beta}_0 + \hat{\beta}_1 X_0) - (\beta_0 + \beta_1 X_0 + \varepsilon_0)$$

Thus the variance of bias has to count the

extra variation from ε_0.

(iii) Variance of prediction bias is

$$Var(bias) = Var[(\hat{\beta}_0 + \hat{\beta}_1 X_0) - (\beta_0 + \beta_1 X_0 + \varepsilon_0)]$$

$$= Var[(\bar{y} - \hat{\beta}_1 \bar{X}) + \hat{\beta}_1 X_0 - \varepsilon_0 - (\beta_0 + \beta_1 X_0)]$$

$$= Var(\bar{y}) + (X_0 - \bar{X})^2 Var(\hat{\beta}_1) + Var(\varepsilon_0)$$

Hint: all the covariates are 0.

$$= \frac{\sigma^2}{n} + (X_0 - \bar{X})^2 \frac{\sigma^2}{S_{XX}} + \sigma^2$$

$$= \sigma^2 [1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{XX}}]$$

The extra part is exactly the $Var(\varepsilon_0)$. in the variance of prediction bias.

(iv) when σ^2 is unknown, estimate σ^2 by MSE

(v). prediction interval for $y|x_0$ is

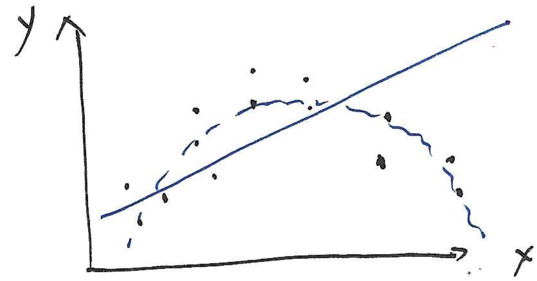
$$\frac{\hat{y}|x_0 - y|x_0}{MSE(1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{XX}})} \sim t(n-2)$$

so the $100(1-\alpha)\%$ C.I. for $y|x_0$ is: $\hat{y}|x_0 \pm t_{\alpha/2, n-2} \sqrt{MSE(1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{XX}})}$

2. Diagonosis of the SLR Model

When we apply a SLR model to a data, we usually are not certain in advance that a SLR is an "appropriate" approach.

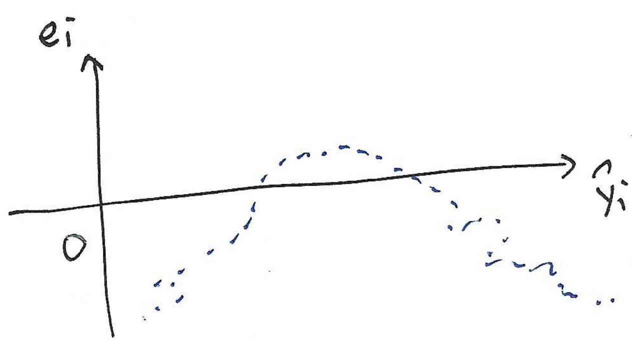
- Is a straight line the right choice?

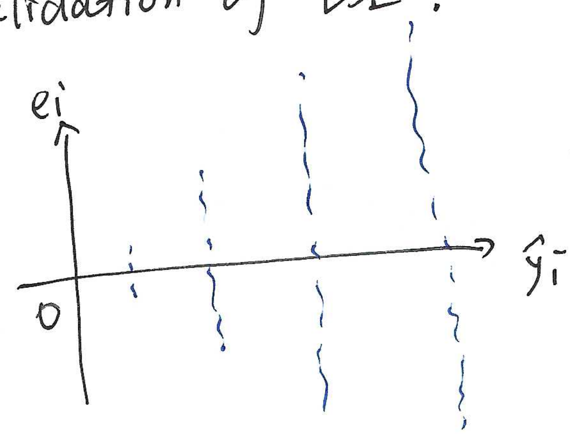


or some transformation might be needed?

- Are the assumptions $E(\varepsilon_i)=0$ or $Var(\varepsilon_i)=\sigma^2$ met?

If not, it will change the validation of LSE!





- Is the assumption of Normal distribution met?

If not, t-test and anova don't apply anymore!

In SLR case, we will simply introduce "Residual Plots" for the model diagonosis, later in MLR we will discuss more numeric measurements and tests.

$\boxed{\text{Model SLR Diagonosis}} \longleftrightarrow \boxed{\text{Examination of the residuals.}}$

Problems

(i) The regression function is not linear in X

(ii) The error terms do not have constant variance

(iii) The error terms are not independent

(iv) The error terms are not normally distributed

Plots for examination

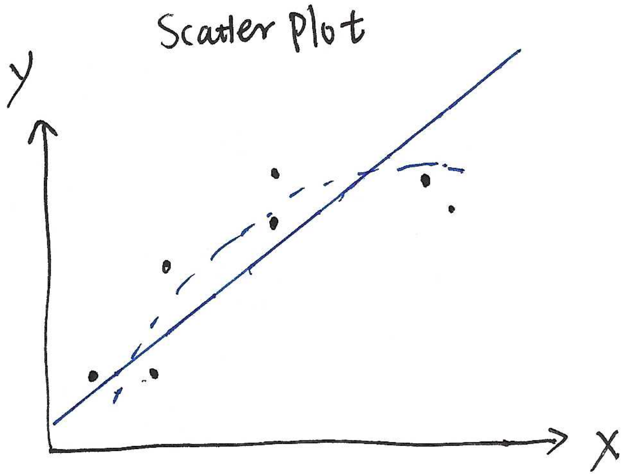
(a) Scatter plot of X v.s. Y.

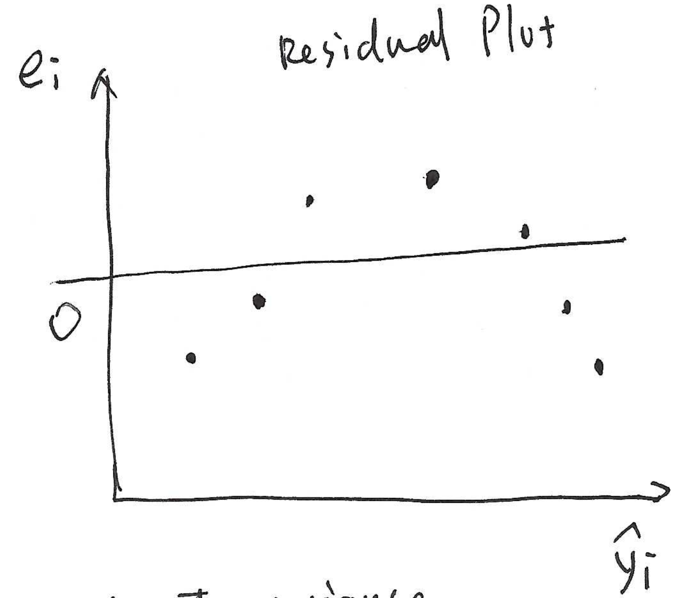
(b) Residual plot e_i against fitted values $\hat{Y}_i$

(c) Autocorrelation plot e_i v.s. time order of obs.

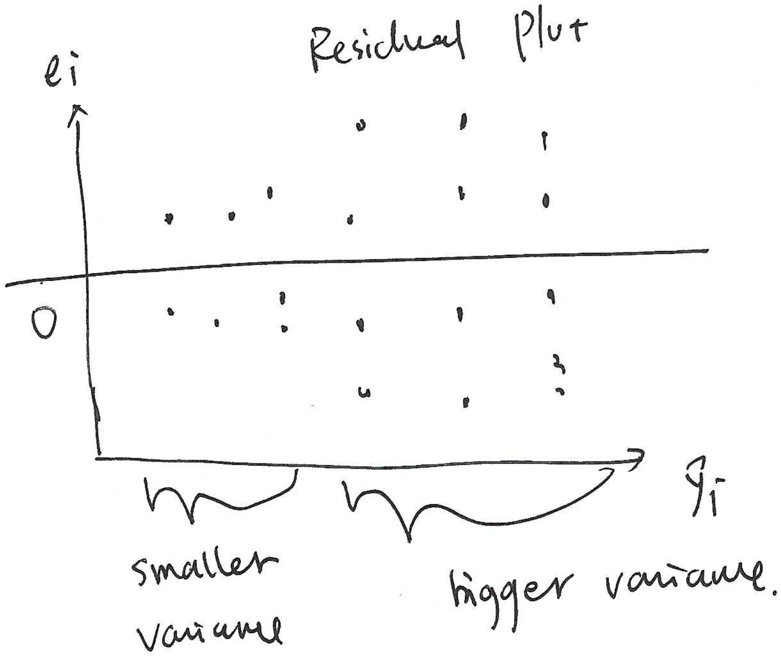
(d) Normality Probability Plot - Q-Q plot of e_i

(i). The regression function is not linear in X.

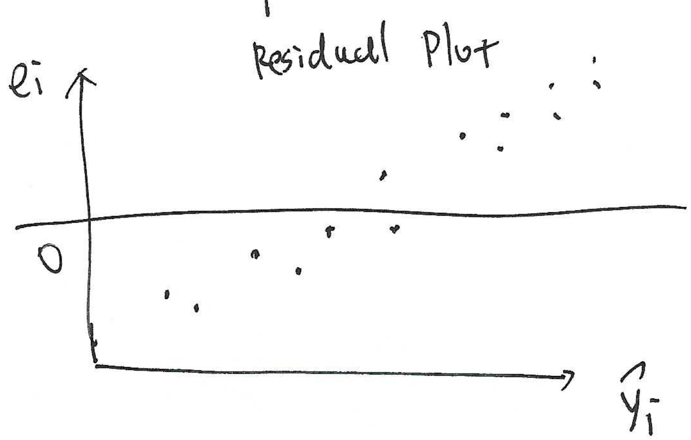


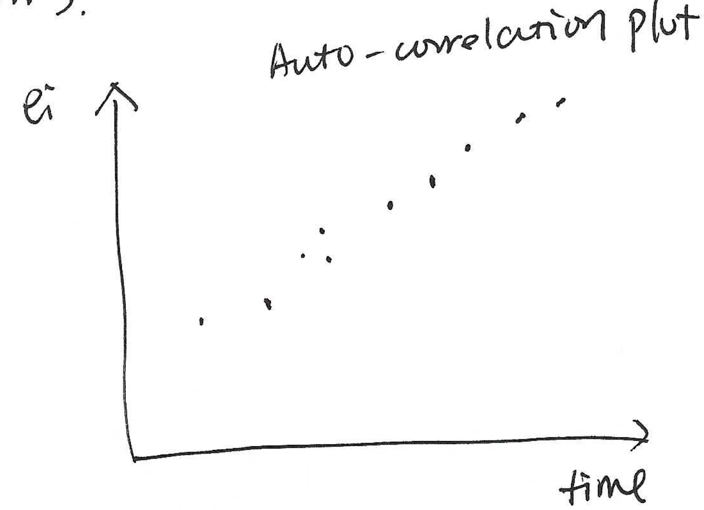


(ii) The error terms do not have constant variance.

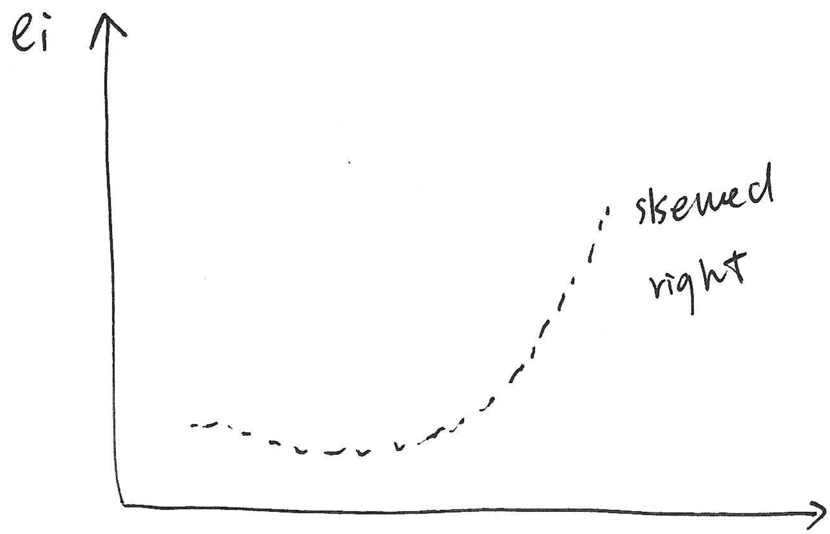


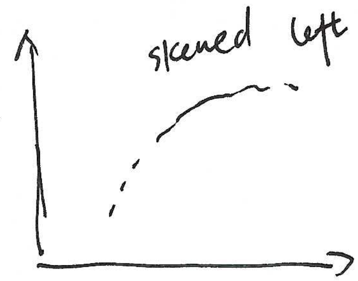
(iii) Non-independence of error terms.

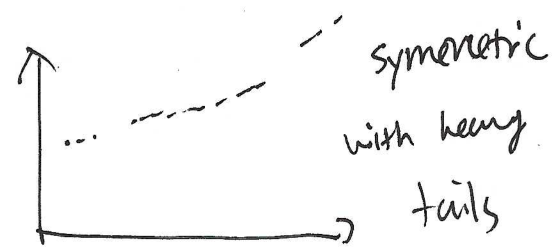




(iv) Non-normality of Error Terms.



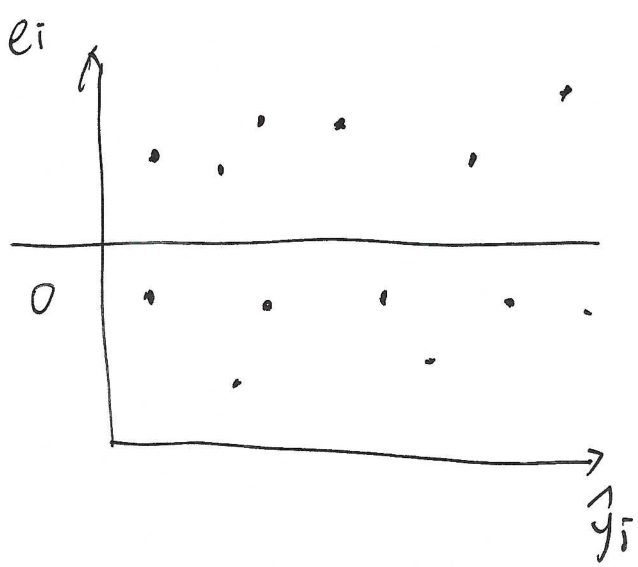




Expected value under Normality. (p110 ~ 111)

Note: Residual plot and Q-Q plot are the most important plots to look at when examinating assumptions.

"Good" Plots.



Distributed evenly, without obvious patterns around 0.

More or less diagonal.

Summary and Take-aways from this lecture.

1. Given $X = x_0$, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_0$ is a point estimate for both

$$E(y|x_0) = \beta_0 + \beta_1 x_0$$

$$y|x_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0$$

But the bias is different, so the variance of prediction is difference, which results in different prediction intervals:

For mean value: $\hat{y} \pm t_{\alpha/2,\, df=n-2} \sqrt{MSE\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$

For actual value: $\hat{y} \pm t_{\alpha/2,\, df=n-2} \sqrt{MSE\left(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$

↓

- The difference in prediction variance is σ^2, that's estimated by MSE. It's the variation from ε_0.

- The prediction variance is bigger for the actual value y, so the interval is wider.

2. Residual Plots: e_i v.s. $\hat{y}_i$ can check

- The regression function is linear in X or not
- The error terms have constant variance or not
- The error terms are independent or not.

QQ-plot: check the Normality assumption.

↓

Note:
When constant variance or independence are violated, the LSE result

$\hat{\beta}_1 = \sigma^2 [\frac{1}{n} + \frac{(X_0-\bar{X})^2}{S_{xx}}]$ are not "BLUE" anymore.

use the same variance,
but in reality, the data has different σ_i^2

This and violation of normal distribution also impact t-test and anova,

- we will discuss details in MLR.