Residuals and Fitted Values

Fitted Values: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad i=1,\cdots,n$.

where $\hat{\beta}_1 = S_{xy}/S_{xx}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}$

Residuals: $e_i = y_i - \hat{y}_i, \quad i=1,\cdots,n$

Two restrictions: (1) $\sum_i e_i = 0$, (2) $\sum_i e_i x_i = 0$

(1) $\sum_i e_i = \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$$= \sum_i \left[(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})\right]$$

$$= \sum_i (y_i - \bar{y}) - \hat{\beta}_1 \cdot \sum_i (x_i - \bar{x}) = 0 - \hat{\beta}_1 \cdot 0 = 0$$

(2) $\sum_i e_i x_i = \sum_i \left[(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})\right] \cdot x_i$

$$= \sum_i (y_i - \bar{y}) \cdot x_i - \hat{\beta}_1 \sum_i (x_i - \bar{x}) \cdot x_i$$

$$= \sum_i (y_i - \bar{y})(x_i - \bar{x}) - \hat{\beta}_1 \cdot \sum_i (x_i - \bar{x})(x_i - \bar{x})$$

$$= S_{xy} - \left(\frac{S_{xy}}{S_{xx}}\right) \cdot S_{xx} = 0$$

$e = (e_1, \cdots, e_n)'$, $x = (x_1, \cdots x_n)'$. $\quad e'x = 0 \Leftrightarrow e \perp x$, orthogonal

It is easy to show that

$$\sum_i e_i (\hat{y}_i - \bar{y}) = \sum_i e_i (\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})$$

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

$$= \hat{\beta}_1 \cdot \sum_i e_i (x_i - \bar{x}) = \hat{\beta}_1 \cdot \sum_i e_i x_i = 0$$

Cochran's Theorem – Decomposition of χ^2-distribution

(1) If $X_1 \sim \chi^2(r_1)$, $X_1+X_2 \sim \chi^2(r)$, X_1 is independent of X_2

then $X_2 \sim \chi^2(r-r_1)$.

(2) If $X_1 \sim \chi^2(r_1)$, $X_2 \sim \chi^2(r_2)$, and $X_1+X_2 \sim \chi^2(r_1+r_2)$

then X_1 and X_2 are independent.

Decomposition of Sum Square: $SST = SSR + SSE$

$$SSR = \sum_i (x_i-\bar{x})^2 \cdot \hat{\beta}_1^2 = S_{xx} \cdot \hat{\beta}_1^2$$

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

where $c_i = \dfrac{x_i-\bar{x}}{S_{xx}} = \dfrac{x_i-\bar{x}}{\sum_{i=1}^n (x_i-\bar{x})^2}$ with $\sum_{i=1}^n c_i = 0$, $\sum_{i=1}^n c_i x_i = \sum_{i=1}^n c_i(x_i-\bar{x}) = 1$

$$\sum_{i=1}^n c_i^2 = S_{xx}^{-1} = \left[\sum_i (x_i-\bar{x})^2\right]^{-1}.$$

$$\left.\begin{array}{l} \therefore\ \hat{\beta}_1 = \sum_{i=1}^n c_i \cdot y_i, \\ y_i \overset{ind.}{\sim} N(\beta_0+\beta_1 x_i, \sigma^2) \end{array}\right\} \Rightarrow \hat{\beta}_1 \sim N\left(\sum_{i=1}^n c_i\beta_0 + \sum_{i=1}^n c_i x_i \beta_1,\ \sum_{i=1}^n c_i^2\sigma^2\right)$$

$$\sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

(2) $$\therefore \frac{SSR}{\sigma^2} = \frac{\hat{\beta}_1^2 \cdot S_{xx}}{\sigma^2} = \frac{(\hat{\beta}_1-0)^2}{\sigma^2/S_{xx}} \overset{H_0:\beta_1=0}{\sim} \chi^2(1)$$

$$\frac{SST}{\sigma^2} = \frac{\sum_{i=1}^n (y_i-\bar{y})^2}{\sigma^2} = \frac{(n-1)\cdot S_y^2}{\sigma^2} \sim \chi^2(n-1)$$

In addition, $SSE \perp SSR$ as $Cov(e, \beta) = 0$

$$SSE = \sum_{i=1}^n (y_i-\hat{y}_i)^2 = \sum_{i=1}^n e_i^2 \underset{\text{Cochran's Theorem}}{\Rightarrow} \frac{SSE}{\sigma^2} \sim \chi^2(n-2)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \sim N(\mu_{\beta_0}, \ \sigma^2_{\beta_0})$$

$$\mu_{\beta_0} = E(\hat{\beta}_0) = E(\bar{y} - \hat{\beta}_1 \bar{x}) \qquad y_i \overset{ind.}{\sim} N(\beta_0 + \beta_1 x_i, \ \sigma^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i) - \hat{\beta}_1 \cdot \bar{x} = \beta_0 .$$

$$\sigma^2_{\beta_0} = Var(\hat{\beta}_0) = Var(\bar{y} - \hat{\beta}_1 \bar{x})$$

$$\left[\begin{aligned} \bar{y} - \hat{\beta}_1 \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} y_i - \Big(\sum_{i=1}^{n} c_i y_i\Big)\bar{x} \\ &= \sum_{i=1}^{n}\Big(\frac{1}{n} - c_i \bar{x}\Big) y_i \end{aligned} \right]$$

$$\hat{\beta}_1 = \sum_i c_i y_i, \qquad c_i = \frac{x_i - \bar{x}}{S_{xx}}$$

$$\therefore \ \sigma^2_{\beta_0} = Var\Big(\sum_{i=1}^{n}\Big(\frac{1}{n} - c_i \bar{x}\Big) y_i \Big)$$

$$= \sigma^2 \sum_{i=1}^{n}\Big(\frac{1}{n} - c_i \bar{x}\Big)^2 = \sigma^2\Big(\frac{1}{n} - \frac{2}{n}\bar{x}\sum_i c_i + \bar{x}^2 \sum_i c_i^2\Big)$$

$$= \sigma^2\Big(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\Big) = \frac{\sigma^2 \sum_i x_i^2}{S_{xx}} \qquad \Big[\sum_i (x_i - \bar{x})^2 + n\bar{x}^2 = \sum_i x_i^2 \Big]$$

$$\therefore \ \hat{\beta}_0 \sim N\Big(\beta_0, \ \frac{\sigma^2 \cdot \sum_i x_i^2}{S_{xx}}\Big)$$

Note that: $\hat{\beta}_1 \sim N\Big(\beta_1, \ \frac{\sigma^2}{S_{xx}}\Big)$

$$E(MSE) = E\Big[\frac{SSE}{n-2}\Big] = \sigma^2 \qquad \hat{\sigma}^2 = MSE$$

$$\therefore \ se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{MSE}{S_{xx}}} \qquad se(\hat{\beta}_0) = \sqrt{\frac{MSE \cdot \sum_i x_i^2}{S_{xx}}}$$

Sampling Distribution of $\hat{\beta}_1$

$$\frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\dfrac{\hat{\beta}_1 - \beta_1}{\sqrt{Var(\hat{\beta}_1)}}}{\sqrt{\dfrac{se^2(\hat{\beta}_1)}{Var(\hat{\beta}_1)}}} = \frac{\overset{\sim N(0,1)}{Z_{\beta_1}}}{\sqrt{\dfrac{MSE/S_{XX}}{\sigma^2/S_{XX}}}} = \frac{Z_{\beta_1}}{\sqrt{\dfrac{MSE}{\sigma^2}}}$$

$$\frac{MSE}{\sigma^2} = \frac{SSE}{\sigma^2} \Big/ (n-2) \sim \chi^2(n-2) \Big/ (n-2)$$

$$\therefore \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{Z_{\beta_1}}{\sqrt{\chi^2_{(n-2)} \big/ (n-2)}}, \qquad \hat{\beta}_1 \perp\!\!\!\perp SSE$$

$$\sim t(n-2)$$

$100(1-\alpha)\%$ C.I. for β_1:

$$\hat{\beta}_1 \pm t_{\frac{\alpha}{2}}(n-2) \cdot se(\hat{\beta}_1)$$

Confidence interval for β_0: $\quad \hat{\beta}_0 \pm t_{\frac{\alpha}{2}}(n-2) \cdot se(\hat{\beta}_0)$, $\quad se(\hat{\beta}_0) = \sqrt{\frac{MSE \cdot \sum_i x_i^2}{n S_{xx}}}$

C.I. for β_1: $\quad \hat{\beta}_1 \pm t_{\frac{\alpha}{2}}(n-2) \cdot se(\hat{\beta}_1)$, $\quad se(\hat{\beta}_1) = \sqrt{\frac{MSE}{S_{xx}}}$

* Confidence interval for $\beta_0 + \beta_1 x_i$ (Mean Response).

(1) LS estimators: $\quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

Fitted values: $\quad \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad i = 1, \cdots, n$

Mean Response $\quad E Y_i = \beta_0 + \beta_1 x_i, \quad i = 1, \cdots, n, \quad Y_i \overset{ind.}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$

Note that $\quad E(\hat{Y}_i) = \beta_0 + \beta_1 x_i = E(Y_i)$,

(2) $Var(\hat{Y}_i) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_i)$

$$= Var(\hat{\beta}_0) + x_i^2 \cdot Var(\hat{\beta}_1) + 2x_i \cdot Cov(\hat{\beta}_0, \hat{\beta}_1)$$

where $Cov(\hat{\beta}_0, \hat{\beta}_1) = Cov(\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1)$

$$= Cov(\bar{y}, \hat{\beta}_1) - \bar{x} \cdot Var(\hat{\beta}_1) = -\bar{x}\, Var(\hat{\beta}_1)$$

$$\hat{\beta}_1 = \sum_i c_i y_i = \sum_i \frac{x_i - \bar{x}}{S_{xx}} y_i, \quad \bar{y} = \sum_i \left(\frac{1}{n}\right) y_i$$

$$Cov(\bar{y}, \hat{\beta}_1) = Cov\left(\sum_{i=1}^{n} \left(\frac{1}{n}\right) y_i, \sum_{i=1}^{n} c_i y_i\right)$$

$y_1, \cdots, y_n$ are independent

$$Cov(\bar{y}, \hat{\beta}_1) = \sum_{i=1}^{n} \frac{c_i}{n} Var(y_i) = \frac{1}{n}\left(\sum_i c_i\right)\sigma^2 = 0$$

$$\Rightarrow Var(\hat{Y}_i) = Var(\hat{\beta}_0) + x_i^2 \cdot Var(\hat{\beta}_1) - 2x_i \cdot \bar{x} \cdot Var(\hat{\beta}_1)$$

$$= \frac{\sigma^2}{n S_{xx}} \left(\sum_i x_i^2 + n x_i^2 - 2x_i \cdot \sum_i x_i\right)$$

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

(3) Standard Error of $\hat{Y}_i$: $se(\hat{Y}_i)$

$$se^2(\hat{Y}_i) = \hat{\sigma}^2 \cdot \left(\frac{1}{n} + \frac{(x_i-\bar{x})^2}{S_{xx}}\right) = MSE \cdot \left[\frac{1}{n} + \frac{(x_i-\bar{x})^2}{S_{xx}}\right]$$

where $MSE = \frac{SSE}{n-2}$ is unbiased for σ^2.

(4) Sampling Distribution of $\hat{Y}_i$:

$$\frac{\hat{Y}_i - E(\hat{Y}_i)}{se(\hat{Y}_i)} = \frac{\hat{Y}_i - (\beta_0 + \beta_1 x_i)}{\sqrt{MSE \cdot \left[\frac{1}{n} + \frac{(x_i-\bar{x})^2}{S_{xx}}\right]}}$$

$$= \frac{\left[\hat{Y}_i - (\beta_0 + \beta_1 x_i)\right] \Big/ \sqrt{\sigma^2\left(\frac{1}{n} + \frac{(x_i-\bar{x})^2}{S_{xx}}\right)}}{\sqrt{\frac{SSE}{\sigma^2} \Big/ (n-2)}}$$

$$\overset{D}{=} \frac{Z \sim N(0,1)}{\sqrt{\chi^2(n-2)/(n-2)}} \qquad \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \perp\!\!\!\perp SSE$$

$$\sim t(n-2)$$

$100(1-\alpha)\%$ Confidence interval for $(\beta_0 + \beta_1 x_i)$ is

$$\hat{Y}_i \pm t_{\frac{\alpha}{2}}(n-2) \cdot se(\hat{Y}_i)$$

where $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $se(\hat{Y}_i) = \sqrt{MSE\left[\frac{1}{n} + \frac{(x_i-\bar{x})^2}{S_{xx}}\right]}$.

Sampling Distribution of Autocorrelation

Residuals of simple linear regression: $e_{1,...,}e_n$. Suppose data is collected over time $t = 1, ...n$ and normally distributed with constant variance σ^2, i.e. $e_{1,...,}e_n \sim i.i.d.N(0,1)$

Lag 1 autocorrelation

$$r_1 = \frac{\sum_{t=2}^{n} e_{t-1}e_t}{\sum_{t=1}^{n} e_t^2}$$

Under the hypothesed independence among the residuals, the numerator has mean $E\left[\sum_{t=2}^{n} e_{t-1}e_t\right] = 0$ and variance

$$\begin{aligned} var\left[\sum_{t=2}^{n} e_{t-1}e_t\right] &= E\left[\sum_{t=2}^{n} e_{t-1}e_t\right]^2 \\ = \quad E\left[\sum_{s=2}^{n}\sum_{t=2}^{n} e_{s-1}e_s e_{t-1}e_t\right] &= E\sum_{t=2}^{n} e_{t-1}^2 e_t^2 = (n-1)\sigma^4. \end{aligned}$$

For large n, we can show that $\sum_{t=2}^{n} e_{t-1}e_t$ is asymptotic normal $N(0,(n-1)\sigma^4)$, or $\sigma^{-2}(n-1)^{-1/2}\sum_{t=2}^{n} e_{t-1}e_t \sim AN(0,1)$.

Denominator $SSE = \sum_{t=1}^{n} e_t^2$. And $MSE = SSE/(n-2)$ is an estimate for σ^2. Hence

$$r_1 = \frac{\sum_{t=2}^{n} e_{t-1}e_t}{\sum_{t=1}^{n} e_t^2} \sim AN(0, \frac{1}{n}).$$

The 95% confidence interval of autocorrelation coefficient could be approximated by

$$r_1 \pm 2\sqrt{\frac{1}{n}}.$$