2. Regression Analysis

Data set : (y_i, x_i) $i = 1, \dots, 12$

1. In an experiment to investigate the variation of the specific heat (Y) of a certain chemical with the temperature (x), two specific heat measurements were made for each of the 6 temperatures under study.

Temperature (^oC)	50	60	70	80	90	100
Specific heat ($^oC/g$)	1.60	1.63	1.67	1.70	1.71	1.71
	1.64	1.65	1.67	1.72	1.72	1.74

$n = 12$

(a) Estimate the regression equation using the least squares method.

(b) Obtain a 95% confidence interval on the intercept β_0.

(c) Test the hypothesis $H_0 : \beta_1 = 0$ against the hypothesis $H_1 : \beta_1 \neq 0$.

(d) Get an estimate of the specific heat corresponding to temperature 75^oC. Determine the 95% confidence interval and prediction interval for this temperature value.

a) Model: $\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ $\quad (1 \;\; x)$

$$\hat{E}[Y|x] = \hat{\mu}_{Y|x} = \hat{\beta}_0 + \hat{\beta}_1 x = \underset{\sim}{x}^T \underset{\sim}{\hat{\beta}}$$

$$\underset{12\times 1}{\underset{\sim}{y}} = \begin{bmatrix} 1.60 \\ 1.64 \\ 1.63 \\ 1.65 \\ 1.67 \\ 1.67 \\ \vdots \\ 1.71 \\ 1.74 \end{bmatrix} \qquad \underset{12\times 2}{\underset{\sim}{X}} = \begin{bmatrix} 1 & 50 \\ 1 & 50 \\ 1 & 60 \\ 1 & 60 \\ \vdots & \vdots \\ 1 & 100 \\ 1 & 100 \end{bmatrix}$$

$$\underset{\sim}{X}^T \underset{\sim}{y} = \begin{bmatrix} \sum_{i=1}^{12} y_i \\ \sum_{i=1}^{12} x_i y_i \end{bmatrix} = \begin{bmatrix} 20.16 \\ 1519.9 \end{bmatrix}$$

$$\underset{\sim}{\hat{\beta}} = (\underset{\sim}{X}^T \underset{\sim}{X})^{-1} \underset{\sim}{X}^T \underset{\sim}{y}$$

$$\underset{\sim}{C}^{-1} = \underset{\sim}{X}^T \underset{\sim}{X} = \begin{bmatrix} n & \sum_{i=1}^{12} x_i \\ \sum_{i=1}^{12} x_i & \sum_{i=1}^{12} x_i^2 \end{bmatrix} = \begin{bmatrix} 12 & 900 \\ 900 & 71000 \end{bmatrix}$$

$$\underset{\sim}{C} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1} = \frac{1}{12 \times 71000 - 900^2}\begin{bmatrix} 71000 & -900 \\ -900 & 12 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{71}{42} & \frac{-3}{140} \\ \frac{-3}{140} & \frac{1}{3500} \end{bmatrix}$$

$$\underset{2\times 1}{\hat{\underset{\sim}{\beta}}} = \underbrace{\begin{bmatrix} \frac{71}{42} & -\frac{3}{140} \\ -\frac{3}{140} & \frac{1}{3500} \end{bmatrix}}_{\underset{\sim}{C}} \underbrace{\begin{bmatrix} 20.16 \\ 1519.9 \end{bmatrix}}_{\underset{\sim}{X}'\underset{\sim}{Y}} = \begin{bmatrix} 1.5107 \\ 0.002257 \end{bmatrix}$$

$$\hat{Y} = \underset{\hat{\beta}_0}{1.5107} + \underset{\hat{\beta}_1}{0.002257}\, x \quad , \forall x \in (50; 100)$$

(b) Obtain a 95% confidence interval on the intercept β_0.

b) $CI_{95\%}(\beta_0) = ?$

Pivotal quantity : $T = \dfrac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 c_{1,1}}} \overset{k=0}{\sim} t_{(12-2)} = t_{(10)}$ $\quad n=12,\ p=2$

where, $\hat{\sigma}^2 = \dfrac{SSE}{n-2} = \dfrac{SST - SSR}{n-2}$

and $\underset{\sim}{C} = \begin{bmatrix} \frac{71}{42} & \frac{-3}{140} \\ \frac{-3}{140} & \frac{1}{3500} \end{bmatrix} \rightarrow c_{1,1} = \frac{71}{42}$

Derivation of the C.I. (Random)

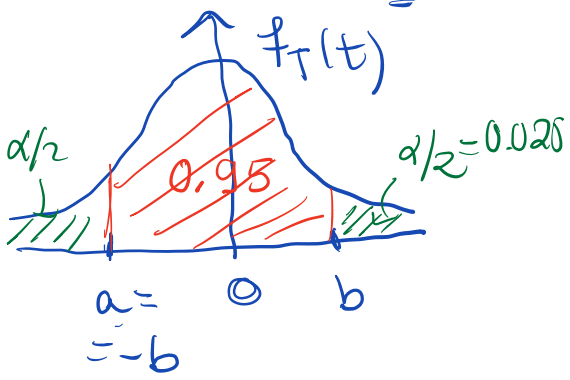
$$P(A \leq \beta_0 \leq B) = 0.95$$

Random variables

to get this we take the Pivotal Quantity and we find the constants a and b:

$$P(a \leq T \leq b) = 0.95$$

($0.95 = 1-\alpha$)



and taking the symmetrical Interval around 0 $\Rightarrow a = -b$

So, $b = F^{-1}_{t(10)}(0.975) = t_{0.975(10)}$ (97,5% quantile of the $t(10)$)

$$P\left(-t_{0.975(10)} < \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 c_{1,1}}} < t_{0.975(10)}\right) = 0.95 \Leftrightarrow$$

$$P\left(-t_{0.975(10)} \times \sqrt{\hat{\sigma}^2 c_{1,1}} \leq \hat{\beta}_0 - \beta_0 \leq t_{0.975(10)} \times \sqrt{\hat{\sigma}^2 c_{1,1}}\right) = 0.95$$

$\Leftrightarrow$

$\Leftrightarrow$

$$P\left(-\hat{\beta}_0 - t_{0.975(10)}\sqrt{\hat{\sigma}^2 c_{1,1}} \leq -\beta_0 \leq -\hat{\beta}_0 + t_{0.975(10)}\sqrt{\hat{\sigma}^2 c_{1,1}}\right) = 0.95$$

multiplying by (-1), we obtain:

$$P\left(\hat{\beta}_0 - t_{0.975(10)}\sqrt{\hat{\sigma}^2 c_{1,1}} \leq \beta_0 \leq \hat{\beta}_0 + t_{0.975(10)}\sqrt{\hat{\sigma}^2 c_{1,1}}\right) = 0.95$$

Random variable — A — random variable — Random variable — B — random variable

Concretization of the Random C.I:

$$\hat{\beta}_0 = 1.5107 \; ; \; t_{0.975(10)} = 2.2281 \; ; \; c_{1,1} = \frac{71}{42}$$

$$\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{SST - SSR}{n-2} \qquad SST = \underset{\sim}{y}^T \underset{\sim}{y} - n\bar{y}^2$$

$$SST - SSR = \underset{\sim}{y}^T \underset{\sim}{y} - \underset{\sim}{\beta}^T \underset{\sim}{X}^T \underset{\sim}{y} = 33.8894 - \begin{bmatrix} 1.5107 & 0.002257 \end{bmatrix} \begin{bmatrix} 20.16 \\ 1519.9 \end{bmatrix}$$

$$= 0.002769$$

$$\underset{\sim}{y}^T \underset{\sim}{y} = \sum_{i=1}^{12} y_i^2 = 33.8894$$

$$SSE = \underset{\sim}{\beta}^T \underset{\sim}{X}^T \underset{\sim}{y} - n\bar{y}^2$$

$$\hat{\sigma}^2 = MSE = \frac{0.002769}{10} = 0.0002769$$

$$CI_{95\%}(\beta_0) = \left[\hat{\beta}_0 \pm t_{0.975(10)}\sqrt{\hat{\sigma}^2 c_{1,1}}\right]$$

$\therefore \quad CI_{95\%}(\beta_0) = \left[1.5107 \pm 2.2281 \sqrt{\frac{0.0002769 \times 71}{42}}\right]$

$\hat{\beta}_0$

$= [1.4625 \; ; \; 1.5589]$

mean of C.I $= \hat{\beta}_0 = 1.5107$

(c) Test the hypothesis $H_0 : \beta_1 = 0$ against the hypothesis $H_1 : \beta_1 \neq 0$.

very Important test (see if the regression is significant)

x is Important to explain E[Y] ?

If $\beta_1 = 0$ then $E[Y|x] = \beta_0 + \beta_1 x$, $\beta_1 = 0$

$\beta_1 < 0 \vee \beta_1 > 0$

$H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$ (two sided test)

Pivotal Quantity: $T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 c_{2,2}}} \sim t(10)$ $\quad K = 1$

Under H_0 ($\beta_1 = 0$), we have the test statistic:

(H_0 true)

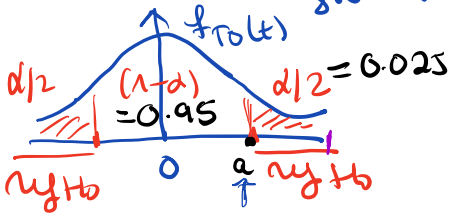
$T_{|H_0} = T_0 = \frac{\hat{\beta}_1 - 0}{\sqrt{\hat{\sigma}^2 c_{2,2}}} \sim t(10)$

Assuming that $\alpha = 0.05$

Note that $\alpha = P(\text{'type I error'}) = P(\text{'reject } H_0 | H_0 \text{ is true})$

$= 0.05$, we can obtain the critical region (where we reject H_0)

Critical region (C.R.)



$|T_0| > a$ and $P(|T_0| > a) = 0.05$

$\Leftrightarrow a = F^{-1}_{t(10)}(0.975) = t_{0.975(10)} = 2.2281$

$CR = \underset{0.05}{]-\infty, -2.2281[\cup]2.2281; +\infty[}$

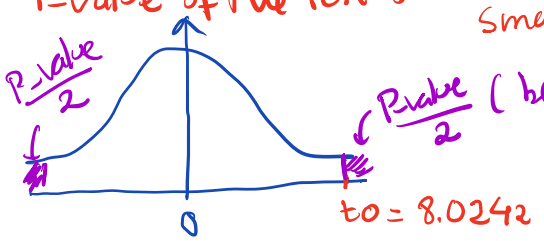
(critical region for $\alpha = 0.05$)

Observed value of the test statistic:

$$t_0 = \frac{0.002257}{\sqrt{\frac{0.0002769}{3500}}} \simeq 8.0242 \in C.R. \Rightarrow \text{reject } H_0 \text{ for } \alpha = 0.05,$$

so, for this dataset it seems that the temperature (x) it is important to explain the Expected value of heat (Y)

P-value of the test:



smallest α for reject H_0 $\Rightarrow$ $t_0 \in$ critical region

(because is a two sided test)

$$\text{P-value} = P(|T_0| \geq t_0) = 2P(T_0 \geq 8.0242) = 2(1 - t_{t(10)}(8.0242))$$

$$= 2(1 - 0.9999943) = 1.1467 \times 10^{-5} \Rightarrow \text{reject } H_0 \ \forall \alpha \text{ usual } \begin{cases} 1\% \\ 5\% \\ 10\% \end{cases}$$

(d) Get an estimate of the specific heat corresponding to temperature 75^oC. Determine the 95% confidence interval and prediction interval for this temperature value.

$x = 75°C \qquad x \in [50; 100]$

$\underset{\sim}{x}_0 = \begin{bmatrix} 1 \\ 75 \end{bmatrix}$

$\hat{E}[Y|\underset{\sim}{x}_0] = \hat{\mu}_{Y|\underset{\sim}{x}_0} = 1.5107 + 0.002257 \times 75 = 1.68$

$CI_{95\%}(\mu|\underset{\sim}{x}_0) = ?$

Pivotal Quantity: $T = \dfrac{\hat{\mu}_{Y|\underset{\sim}{x}_0} - \mu_{Y|\underset{\sim}{x}_0}}{\sqrt{\hat{\sigma}^2 \underset{\sim}{x}_0^T (\underset{\sim}{X}^T\underset{\sim}{X})^{-1} \underset{\sim}{x}_0}} \sim t(10)$

$CI_{95\%}(\mu_{Y|\underset{\sim}{x}_0}) = \left[\hat{\mu}_{Y|\underset{\sim}{x}_0} \pm t_{0.975(10)} \sqrt{\hat{\sigma}^2 \underset{\sim}{x}_0^T (\underset{\sim}{X}^T\underset{\sim}{X})^{-1} \underset{\sim}{x}_0}\right]$

$t_{0.975(10)} = 2.2281$

$\underset{\sim}{x}_0^T (\underset{\sim}{X}^T\underset{\sim}{X})^{-1} \underset{\sim}{x}_0 = (1\ 75) \begin{bmatrix} \frac{71}{42} & -\frac{3}{140} \\ -\frac{3}{140} & \frac{1}{3500} \end{bmatrix} \begin{bmatrix} 1 \\ 75 \end{bmatrix} =$

$= \left[\frac{71}{42} - \frac{3 \times 75}{140} \;;\; \frac{-3}{140} + \frac{75}{3500}\right] \begin{bmatrix} 1 \\ 75 \end{bmatrix} \simeq 0.083333$

$\hat{\sigma}^2 \underset{\sim}{x}_0^T (\underset{\sim}{X}^T\underset{\sim}{X})^{-1} \underset{\sim}{x}_0 = 0.000276 9 \times 0.083333 \simeq 2.3075 \times 10^{-5}$

$\sqrt{\hat{\sigma}^2 \underset{\sim}{x}_0^T (\underset{\sim}{X}^T\underset{\sim}{X})^{-1} \underset{\sim}{x}_0} \simeq 0.0048$

$CI_{95\%}(\mu_{Y|\underset{\sim}{x}_0}) = [1.68 \pm 2.2281 \times 0.0048]$

$= [1.669 \;;\; 1.691]$

$PI_{95\%}(Y_0) = ?$

Pivotal Quantity: $T = \dfrac{\hat{Y}_0 - Y_0}{\sqrt{\hat{\sigma}^2(1 + \underset{\sim}{x}_0^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{x}_0)}} \sim t(10)$

$PI_{95\%}(Y_0) = \left[\hat{Y}_0 \pm t_{0.975(10)} \sqrt{\hat{\sigma}^2(1 + \underset{\sim}{x}_0^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{x}_0}\right]$

$\hat{Y}_0 = \hat{E}[Y|\underset{\sim}{x}_0] = 1.68$

$t_{0.975(10)} = 2.2281$

$\sqrt{\hat{\sigma}^2(1 + \underset{\sim}{x}_0^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{x}_0} = \sqrt{0.0002769 + 2.3075 \times 10^{-5}}$

$\simeq 0.01732$

$PI_{95\%}(Y_0) = [1.68 \pm 2.2281 \times 0.01732]$

$= [1.641 ; 1.719]$ more wide then the C.I