17. $y_i=\beta_0+\beta_1 x_i+e_i \quad i=1,\cdots n$

$e_i \overset{iid}{\sim} N(0,\sigma^2)$

$$Var(\hat{\beta}_1)=\frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}, \quad Var(\hat{\beta}_0)=\frac{\sigma^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}$$

并求 $H_0:\beta_0=0$ 统计量及分布

由LSE，

$$\hat{\beta}_1=\frac{\sum_{i=1}^n (x_i-\bar{x})y_i}{\sum_{i=1}^n (x_i-\bar{x})^2} \qquad Var(\hat{\beta}_1)=\frac{\sum_{i=1}^n (x_i-\bar{x})^2 \, Var\, y_i}{\left(\sum_{i=1}^n (x_i-\bar{x})^2\right)^2}=\frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}$$

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=\frac{\sum_{i=1}^n y_i}{n}-\frac{\sum_{i=1}^n (x_i-\bar{x})\bar{x}y_i}{\sum_{i=1}^n (x_i-\bar{x})^2}$$

$$=\sum_{i=1}^n\left(\frac{1}{n}-\frac{\bar{x}(x_i-\bar{x})}{\sum_{j=1}^n (x_j-\bar{x})^2}\right)y_i$$

$$Var(\hat{\beta}_0)=\sum_{i=1}^n\left(\frac{1}{n}-\frac{\bar{x}(x_i-\bar{x})}{\sum_{j=1}^n (x_j-\bar{x})^2}\right)^2\sigma^2$$

$$=\frac{\sigma^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}$$

由此 $\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}\right)$

$$\hat{\sigma}^2=\frac{1}{n-2}SSE=\frac{1}{n-2}\sum_{i=1}^n (y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)^2$$

$H_0:\beta_0=0$ 下有

$$\frac{\hat{\beta}_0}{\hat{\sigma}\sqrt{\frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim t_{n-2}$$

设 $t\sim t_{n-2}$

$P(|t|>\lambda)=\alpha$

当 $|\hat{\beta}_0|>\lambda\hat{\sigma}\sqrt{\frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}}$ 拒绝假设

18. $\hat{\beta} = (X^TX)^{-1}X^Ty = \begin{pmatrix}\hat{\beta}_0 \\ \hat{\beta}_1\end{pmatrix}$

$X = [x_1 \cdots x_5]^T$

$y = [y_1 \cdots y_5]^T$

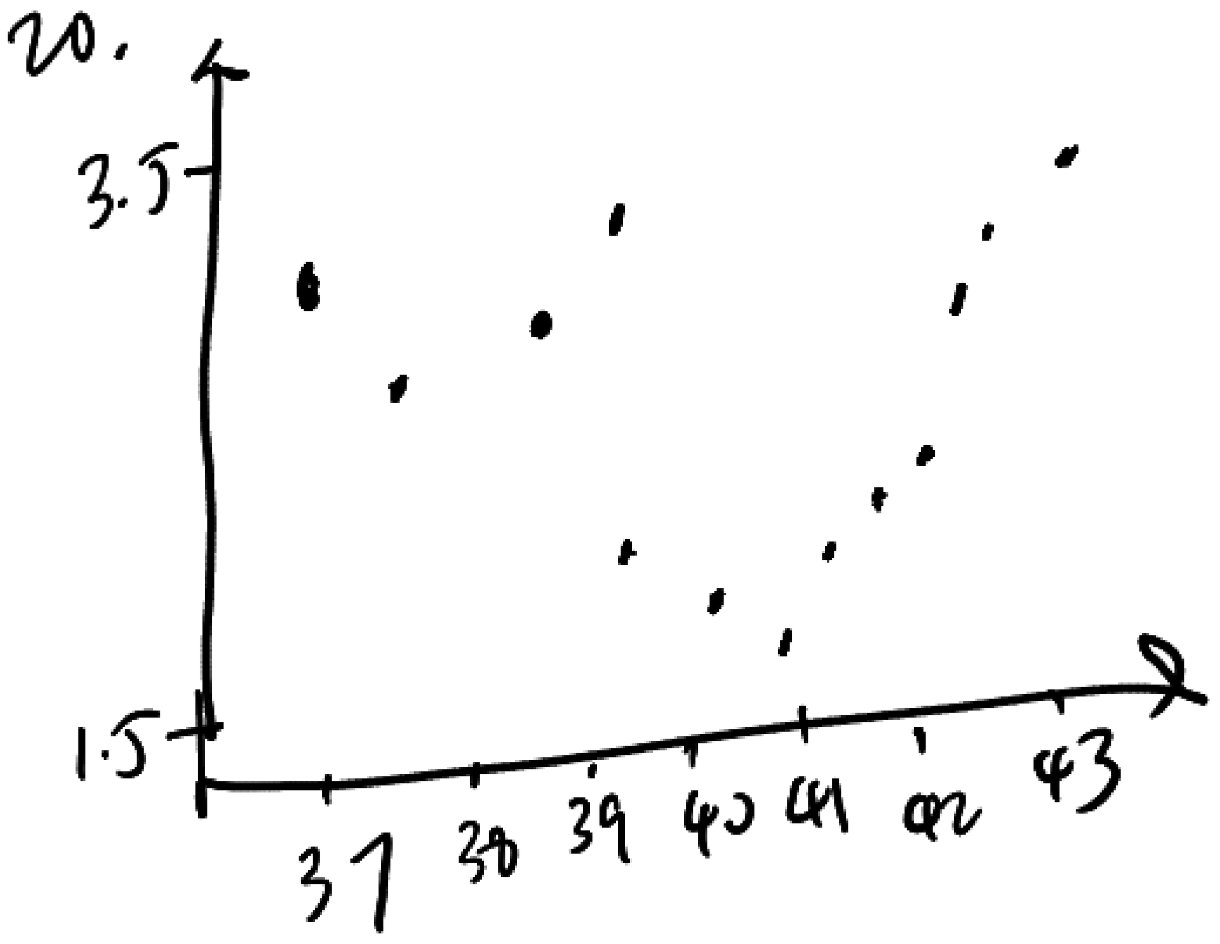
$\hat{\beta}_0 = -11.3 \quad \hat{\beta}_1 = 36.95$

$$T = \lambda\hat{\sigma}\sqrt{\frac{\sum_{i=1}^{n}(x_i)^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}} = 41.277$$

$|\hat{\beta}_0| < T$ 不拒绝 H_0

即认为 $y = \beta_1 x$，$\beta_0 = 0$

20.



从图上看出有相关关系

可以用二次函数 or 分段线性函数

2次函数拟合结果:

$$\hat{y} = 0.156x^2 - 12.62x + 257.07$$

对假设 $\beta_0 = \beta_1 = \beta_2 = 0$

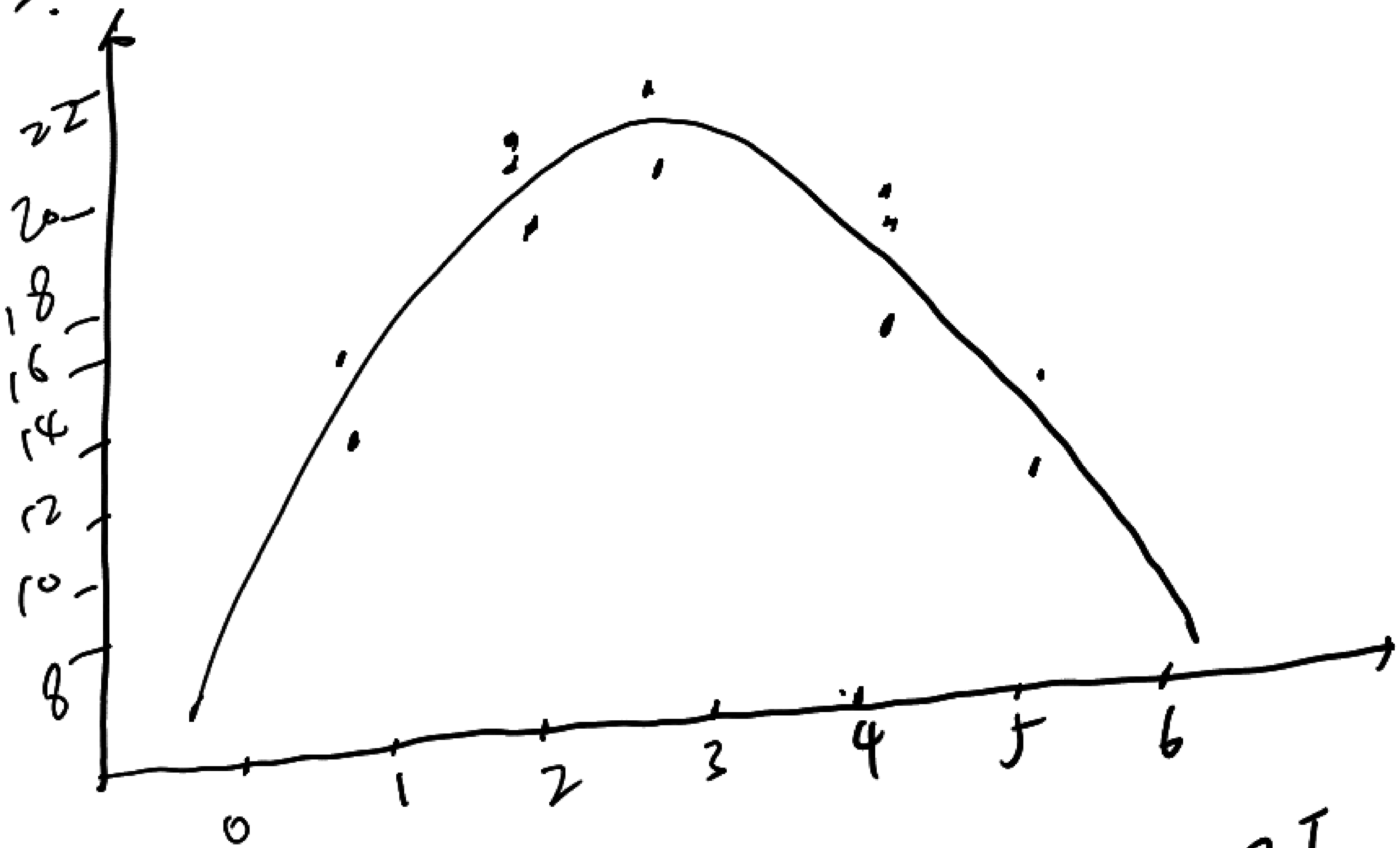
有 F statistic $= 0.13$

用分段线性拟合

$$\hat{y} = \begin{cases} -26.92 - 0.13x & x \le 40 \\ -18.485 + 0.495x & x > 40 \end{cases}$$

效果上这种拟合更好

23.



设 $X = [x_1 \cdots x_{12}]^T$ $y = [y_1 \cdots y_{12}]^T$ 拟合结果 $\hat{y} = 0.07x + 18.4$

一次拟合 $X^{(1)} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{12} \end{pmatrix}$ $r(X^{(1)}) = 2$

二次拟合 $X^{(2)} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{12} & x_{12}^2 \end{pmatrix}$ $r(X^{(2)}) = 3$, $\hat{y} = -1.5x^2 + 9.09x + 7.627$

三次拟合 $X^{(3)} = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ \vdots & \vdots & & \vdots \\ 1 & x_{12} & x_{12}^2 & x_{12}^3 \end{pmatrix}$ $r(X^{(3)}) = 4$, $\hat{y} = -0.063x^3 - 0.935x^2 + 7.589x + 8.7301$

分别计算 F statistic ($H_0: \beta_0 = \cdots \beta_{p-1} = 0$) $\alpha = 0.1$

1次: $F = 0.092$ 临界值 3.285

2次: $F = 0.324$ 临界值 2.923

3次: $F = 0.284$ 临界值 2.9605

故选2次拟合最合适