① a. Around 60 b. Right-skewed c. $200 \le T \le 400$ d. Expected mean > median

② Minimize $\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$ Objective is to minimize the sum of the squared residuals;

③

i	Q(p)	p
1	12	.05
2	15	.15
3	16	.25
4	16	.35
5	16	.45
6	17	.55
7	17	.65
8	17	.75
9	18	.85
10	25	.95

a. $Q(.5) = \frac{16+17}{2} = 16.5$

b. $Q(.88) = 18 + \frac{3}{10}(25-18) = 20.1$

c. $IQR = Q(.75) - Q(.25) = 17 - 16 = 1$

d. Rule 1: x is an outlier if outside $[Q_L, Q_U]$

$Q_L = Q(.25) - 1.5(IQR) = 16 - 1.5 = 14.5$

$Q_U = Q(.75) + 1.5(IQR) = 17 + 1.5 = 18.5$

* 12 & 25 are outliers, since outside $[14.5, 18.5]$

Rule 2: x is an outlier if outside $[\bar{x} - 3s, \bar{x} + 3s]$

$\bar{x} = 16.9 \quad s = 3.28 \rightarrow [7.06, 26.74]$

* no outliers using this rule

- Rules provide different results because extreme values affect rule 2, but not rule 1.

④ a. Response Y = Molecular weight (mg)
Explanatory x = Temp (°C)

b. Ave weight $\Delta = 22.86$ mg (slope of eqn)

c. $r = \sqrt{R^2}$ since linear model; $r = \sqrt{0.941} = 0.970$

d. Definition of $R^2 = 0.941 = 94.1\%$

e. $s = \sqrt{\frac{SSE}{n-k}} = 215.035 \quad (215.035)^2(8-2) = SSE = 277,440$

⑤ a.
- The points are not on the line
- Indicates that the residuals are <u>not</u> normally distributed, suggesting a problem with the linear model

b.
- The points make a pattern, i.e. parabola
- $\hat{y}$ (and therefore x) has inherent curvature (maybe need X^2 term)

⑥ a. $a_3 = \bar{y}_{3\cdot} - \bar{y}_{\cdot\cdot} = 43.7 - 34.4 = 9.3$ on ave, use of temp 3 causes an increase of 9.3 items in yield;

b. $b_1 = \bar{y}_{\cdot 1} - \bar{y}_{\cdot\cdot} = 28.7 - 34.4 = -5.7$

c. $\hat{y}_{31} = \bar{y}_{\cdot\cdot} + a_3 + b_1 = 34.4 + 9.3 - 5.7 = 38.0$ Fitted response

d. Residual $e_{31} = \bar{y}_{31} - \hat{y}_{31} = 44 - 38 = 6$

↑ This is the difference between the mean data response & the fitted response

Which also happens to be the interaction effect, ab_{31}