Lack-Of-Fit Practice problem: # Solution:

# of machines serviced (x_i)	Total # of minutes spent (y_{ij})	n_i	$\bar{y}_i$	$\sum_j (y_{ij} - \bar{y}_i)^2$
1	18, 17	2	17.5	0.5
2	33, 25	2	29	32
4	62, 39, 53, 55	4	52.25	278.75
5	95, 69, 65, 68	4	74.25	582.75
6	86	1	86	0
7	101, 105	2	103	8
8	102	1	102	0

(a) Decompose SSE into the sum of squares due to the pure error, SSPE, and sum of squares due to the lack of fit, SSLF.

- $$SSE = S_{yy} - \frac{S_{xy}^2}{S_{xx}} = \left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] - \frac{\left[\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}\right]^2}{\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]} =$$

$$\left[75187 - \frac{(993)^2}{16}\right] - \frac{(901.625)^2}{65.75} = 13558.94 - 12363.92 = \underline{\underline{1195.019}}$$

- $$SSPE = \sum_i \sum_j \left(y_{ij} - \bar{y}_i\right)^2 = \sum_i (18-17.5)^2 + (17-17.5)^2 + \ldots\ldots + (102-102)^2 =$$

  $$= 0.5 + 32 + 278.75 + 582.75 + 8 = \underline{\mathbf{902}}$$

$\therefore SSLF = SSE - SSPE = \underline{\mathbf{293.019}}$

(b) Test whether the linear model $y = \beta_0 + \beta_1 x + \varepsilon$ is appropriate. Use $\alpha = 0.05$.

H_0 : *model is appropriate* $\qquad \alpha = 0.05$

H_a : *model is not appropriate*

test-statistics: $F_{\text{LF}} = \dfrac{MSLF}{MSPE} = \dfrac{SSLF \Big/ \left[(n-2) - \sum_i (n_i - 1)\right]}{SSPE \Big/ \sum_i (n_i - 1)} = \dfrac{293.019/(14-9)}{902/9} =$

$$= \frac{58.6038}{100.2222} = \underline{\mathbf{0.584739}}$$

R.R: we reject H_0 **if** $F_{\text{LF}} > F_{\alpha(n-2-\sum_i(n_i-1),\sum_i(n_i-1))} = F_{0.05(5,9)} =$ **3.48**

Since F_{LF} **= 0.584739 ≯ 3.48, we do not reject** H_0 **and conclude that at 5% level of significance there is not enough evidence to say that a linear model is not appropriate.**