11.1.4 We can write table as:

Source	Df	Sum of squares	Mean squares	F-statistic	p-value
Treatments	**6**	**7.61**	**1.26**	**0.77**	**0.6**
Error	**77**	**125.57**	**1.63**		
Total	**83**	**133.18**			

You get **(−1 point)** for each empty or wrong value in cell.

11.1.8 (a) We have $s = \sqrt{MSE} = 2.22$ and $q_{0.05,3,30} = 3.49$. Thus, we have

$$\mu_1 - \mu_2 \in (48.05 - 44.74 - \frac{2.22 \times 3.49}{\sqrt{11}}, 48.05 - 44.74 + \frac{2.22 \times 3.49}{\sqrt{11}}) = (0.96, 5.63),$$

$$\mu_1 - \mu_3 \in (48.05 - 49.11 - \frac{2.22 \times 3.49}{\sqrt{11}}, 48.05 - 49.11 + \frac{2.22 \times 3.49}{\sqrt{11}}) = (-3.39, 1.27),$$

$$\mu_2 - \mu_3 \in (44.74 - 49.11 - \frac{2.22 \times 3.49}{\sqrt{11}}, 44.74 - 49.11 + \frac{2.22 \times 3.49}{\sqrt{11}}) = (-6.7, -2.04).$$

(b) Since $0 \in (-3.39, 1.27)$, μ_1 is indistinguishable from μ_3. The other cases are different.

(c) If the number of smaples is n, the length of the confidence intervals are given by

$$2 \times \frac{2.22 \times q_{0.05,3,3n-3}}{\sqrt{n}} < 2.$$

So, we have $4.96 \times q^2_{0.05,3,3n-3} < n$. If $n > 55$, we have the above inequality. Thus, we need additional $55 - 33 = 22$ smaples.

11.1.16 From DS 11.1.4,

$$n_1 = 12,\ \bar{x}_{1.} = 25.13,\ n_2 = 10,\ \bar{x}_{2.} = 29.11,\ n_3 = 11,\ \bar{x}_{3.} = 24.76$$

$$k = 3,\ n_T = 33,\ SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2 = 34.42$$

$$MSE = \frac{SSE}{n_T - k} = 1.147,\ \hat{\sigma} = \sqrt{MSE} = 1.071 \quad \textbf{(+3 points)}$$

Since $q_{0.05,3,30} = 3.486$ the 95% SCI for $\mu_1 - \mu_2$:

$$\left(25.13 - 29.11 - 1.071\frac{3.486}{\sqrt{2}}\sqrt{\frac{1}{12} + \frac{1}{10}}, 25.13 - 29.11 + 1.071\frac{3.486}{\sqrt{2}}\sqrt{\frac{1}{12} + \frac{1}{10}}\right)$$
$$= (-5.11, -2.85)$$

for $\mu_2 - \mu_3$: $(3.20, 5.50)$, and for $\mu_1 - \mu_3$: $(-0.73, 1.47)$. **(+3 points)**
Therefore, layout 2 takes more time to perform the task. **(+4 points)**

11.1.20 We have $n_1 = n_2 = n_3 = 5$ and $n_4 = 4$ and $n_5 = 6$. And we also have

$$\overline{x}_{1.} = 33.6,\ \overline{x}_{2.} = 40,\ \overline{x}_{3.} = 20.4,\ \overline{x}_{4.} = 26.5,\ \overline{x}_{5.} = 31,\ \overline{x}_{..} = 30.12$$

And it follows that

$$\begin{aligned}
\text{SST} &= \sum_{i=1}^{k}\sum_{j=1}^{n_j} x_{ij}^2 - n_T\overline{x}_{..}^2 = 1400.64 \\
\text{SST}_r &= \sum_{i=1}^{k} n_i(\overline{x}_{i.} - \overline{x}_{..})^2 = 1102.74 \\
\text{MST}_r &= \frac{\text{SST}_r}{k-1} = 275.69 \\
\text{SSE} &= \text{SST} - \text{SST}_r = 297.9 \\
\text{MSE} &= \frac{\text{SSE}}{n_T - k} = 14.90, \quad \text{where } k = 5 \quad \text{and} \quad n_T = \sum_{i=1}^{k} n_i = 25.
\end{aligned}$$

Under the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$, (assuming $x_{ij} \overset{\text{i.i.d.}}{\sim} N(\mu_i, \sigma^2)$) we have

$$F = \frac{\text{MST}_r}{\text{MSE}} \sim F_{k-1, n_T-k} \quad \text{and} \quad \text{P-value} = P\left(F_{k-1,n_T-k} \geq \frac{\text{MST}_r}{\text{MSE}}\right).$$

The F-statistics is 18.51 and the p-value is 0.000.**(+4 points)**
Set $\alpha = 0.05$. Let

$$(T_i, T_j) := \text{The Tukey intervals for } \mu_i - \mu_j.$$

Then

$$\begin{aligned}
(T_1, T_2) = (-13.70, 0.90),\ (T_1, T_4) = (-5.14, 10.34),\\
(T_3, T_5) = (-13.09, 0.89),\ (T_4, T_5) = (-2.95, 11.95).
\end{aligned}$$

And they (intervals written above) contain 0. But the following intervals do not contain 0:

$$\begin{aligned}
(T_1, T_3) = (5.89, 20.50),\ (T_1, T_5) = (0.11, 14.09),\\
(T_2, T_3) = (12.30, 26.90),\ (T_2, T_4) = (1.26, 16.47)\\
(T_2, T_5) = (6.51, 20.49),\ (T_3, T_4) = (-18.34, -2.86).
\end{aligned}$$

The treatment 2 and 1 are indistinguishable, as are treatment 1 and 4, but treatment 2 is known to have an greater mean level than that of treatment 4. And the treatment 3 and 5 are indistinguishable, as are treatment 5 and 4, but treatment 4 is known to have an greater mean level than that of treatment 5. **(+4 points)**

- The treatment with the largest mean is treatment 1 or treatment 2. **(+1 point)**
- The treatment with the smallest mean is treatment 3 or treatment 5. **(+1 point)**

11.1.24 We have $n_1 = n_2 = n_3 = n_4 = n_5 = 4$. And we also have

$$\overline{x}_{1.} = 29,\ \overline{x}_{2.} = \overline{x}_{3.} = 28.75,\ \overline{x}_{4.} = 37,\ \overline{x}_{5.} = 42, \overline{x}_{..} = 33.1$$

And it follows that

$$\begin{aligned}
\text{SST} &= \sum_{i=1}^{k}\sum_{j=1}^{n_j} x_{ij}^2 - n_T\overline{x}_{..}^2 = 687.8 \\
\text{SST}_r &= \sum_{i=1}^{k} n_i(\overline{x}_{i.} - \overline{x}_{..})^2 = 596.3 \\
\text{MST}_r &= \frac{\text{SST}_r}{k-1} = 149.1 \\
\text{SSE} &= \text{SST} - \text{SST}_r = 91.5 \\
\text{MSE} &= \frac{\text{SSE}}{n_T - k} = 6.1, \quad \text{where } k = 5 \quad \text{and} \quad n_T = \sum_{i=1}^{k} n_i = 20.
\end{aligned}$$

Under the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$, (assuming $x_{ij} \overset{\text{i.i.d.}}{\sim} N(\mu_i, \sigma^2)$) we have

$$F = \frac{\text{MST}_r}{\text{MSE}} \sim F_{k-1,n_T-k} \quad \text{and} \quad \text{P-value} = P(F_{k-1,n_T-k} \geq \frac{\text{MST}_r}{\text{MSE}}) = 0.000$$

The F-statistics is 24.44 and the p-value is 0.000. **(+4 points)**
Set $\alpha = 0.05$. Let

$$(T_i, T_j) := \text{The Tukey intervals for } \mu_i - \mu_j.$$

Then

$$\begin{aligned}
&(T_1, T_2) = (-4.97, 5.47),\ (T_1, T_3) = (-4.97, 5.47), \\
&(T_2, T_3) = (-5.22, 5.22),\ (T_4, T_5) = (-10.22, 0.22).
\end{aligned}$$

And they (intervals written above) contain 0. But the following intervals do not contain 0:

$$\begin{aligned}
&(T_1, T_4) = (-13.22, -2.75),\ (T_1, T_5) = (-18.22, -7.78), \\
&(T_2, T_4) = (-13.47, -3.03),\ (T_2, T_5) = (-18.47, -8.03) \\
&(T_3, T_4) = (-13.47, -3.03),\ (T_3, T_5) = (-18.47, -8.03).
\end{aligned}$$

Location 1,2,3 are one group and location 4,5 are the other. **(+4 points)**

- Location 4 or 5 has the highest E. Coli level **(+1 point)**
- Location 1, 2, or 3 has the smallest E. Coli level. **(+1 point)**