Jared Yu
Module 6 Discussion

1. Two different treatments are of interest. The first treatment has two levels ($A$ versus $B$). The second treatment has two levels ($a$ versus $b$). Denote by $y$ the targeted response variable.
    a. Construct a multiple linear regression model to estimate the difference in the expected value of $y$ between $A$ and $B$, the difference in the expected value of $y$ between $a$ and $b$, and the difference in the expected value of $y$ among the four treatment combinations $Aa$, $Ab$, $Ba$, $Bb$.

Ans:

Here, we have a regression model of the form,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon, \qquad (1)$$

where

$$x_1 = \begin{cases} 1 & \text{if the observation is from treatment } A \\ 0 & \text{if the observation is from treatment } B, \end{cases}$$
$$x_2 = \begin{cases} 1 & \text{if the observation is from treatment } a \\ 0 & \text{if the observation is from treatment } b. \end{cases}$$

This form of the model is going beyond the first order terms by including the interaction between $x_1$ and $x_2$ with $\beta_3$.

The first part asks us to estimate the difference in the expected value of $y$ between $A$ and $B$. This can be expressed as follows:

$$E(y|x_1 = 1, x_2) - E(y|x_1 = 0, x_2)$$
$$\rightarrow (\beta_0 + \beta_1 + \beta_2 x_2 + \beta_3 x_2) - (\beta_0 + \beta_2 x_2) = \boxed{\beta_1 + \beta_3 x_2}$$

Since in equation (1), we are including an interaction term, then it will include the impact of $x_2$. Therefore, if for example the observation belongs to treatment $a$, then the resulting expected value would be $\beta_1 + \beta_3$ and simply $\beta_1$ if the observation instead belonged to treatment $b$.

The second part asks us to estimate the difference in the expected value of $y$ between $a$ and $b$. This can be expressed as follows:

$$E(y|x_1, x_2 = 1) - E(y|x_1, x_2 = 0)$$
$$\rightarrow (\beta_0 + \beta_1 x_1 + \beta_2 + \beta_3 x_1) - (\beta_0 + \beta_1 x_1) = \boxed{\beta_2 + \beta_3 x_1}$$

Similarly, like before, since in equation (1), we are including an interaction term, then it will include the impact of $x_1$. Therefore, if for example the observation belongs to treatment $A$, then the resulting expected value would be $\beta_2 + \beta_3$ and simply $\beta_2$ if the observation instead belonged to treatment $B$.

The third part asks us to estimate the difference in the expected value of $y$ among the four treatment combinations $Aa$, $Ab$, $Ba$, $Bb$. These different models can be seen below in Table 1.

| First Treatment | Second Treatment | Expected Value |
|---|---|---|
| $A$ | $a$ | $E(y\|x_1 = 1, x_2 = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3$ |
| $A$ | $b$ | $E(y\|x_1 = 1, x_2 = 0) = \beta_0 + \beta_1$ |
| $B$ | $a$ | $E(y\|x_1 = 0, x_2 = 1) = \beta_0 + \beta_2$ |
| $B$ | $b$ | $E(y\|x_1 = 0, x_2 = 0) = \beta_0$ |

*Table 1 The above table shows the different resulting multiple linear regression models based on the combinations of treatment levels from equation (1).*

In such a scenario, there would be a total of $\binom{4}{2} = 6$ different combinations. These different scenarios can be seen below in Table 2. It is possible to add more to for example show $Ab - Aa$, but that will be ignored for simplicity since the results are roughly the same.

| Compared differences | Difference in expected value |
|---|---|
| $Aa - Ab$ | $\beta_0 + \beta_1 + \beta_2 + \beta_3 - (\beta_0 + \beta_1) = \beta_2 + \beta_3$ |
| $Aa - Ba$ | $\beta_0 + \beta_1 + \beta_2 + \beta_3 - (\beta_0 + \beta_2) = \beta_1 + \beta_3$ |
| $Aa - Bb$ | $\beta_0 + \beta_1 + \beta_2 + \beta_3 - (\beta_0) = \beta_1 + \beta_2 + \beta_3$ |
| $Ab - Ba$ | $\beta_0 + \beta_1 - (\beta_0 + \beta_2) = \beta_1 - \beta_2$ |
| $Ab - Bb$ | $\beta_0 + \beta_1 - (\beta_0) = \beta_1$ |
| $Ba - Bb$ | $\beta_0 + \beta_2 - (\beta_0) = \beta_2$ |

*Table 2 The above table shows the difference in expected value for the various combinations of treatment levels.*

b. Construct an analysis of variance model to estimate the difference in the expected value of $y$ between $A$ and $B$, the difference in the expected value of $y$ between $a$ and $b$, and the difference in the expected value of $y$ among the four treatment combinations $Aa$, $Ab$, $Ba$, $Bb$.

Ans:

Let the two-way analysis-of-variance model be represented as

$$y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk} \quad i = 1,2, \qquad j = 1,2, \qquad k = 1, \cdots, n, \tag{2}$$

where $\tau_i$ is the effect of level $i$ of the first treatment, $\gamma_j$ is the effect of level $j$ of the second treatment type, $(\tau\gamma)_{ij}$ is an interaction effect between the two treatment types, and $\varepsilon_{ijk}$ is an $NID(0, \sigma^2)$ random-error component. In this problem $i = 1,2$ corresponds with levels $A$ and $B$, while $j = 1,2$ corresponds with levels $a$ and $b$ respectively.

The first part asks us to estimate the difference in the expected value of $y$ between $A$ and $B$. This can be expressed as follows:

$$E(Y_{1jk}) - E(Y_{2jk}) = \mu + \tau_1 + \gamma_j + (\tau\gamma)_{1j} - [\mu + \tau_2 + \gamma_j + (\tau\gamma)_{2j}]$$
$$\boxed{= \tau_1 - \tau_2 + (\tau\gamma)_{1j} - (\tau\gamma)_{2j}}$$

Like in part a), this result is dependent on the second treatment due to the interaction term. Therefore, depending on if the observation is given treatment $a$ or $b$ the result will vary by the $j$ index where $j = 1$ or 2 respectively.

The second part asks us to estimate the difference in the expected value of $y$ between $a$ and $b$. This can be expressed as follows:

$$E(Y_{i1k}) - E(Y_{i2k}) = \mu + \tau_i + \gamma_1 + (\tau\gamma)_{i1} - [\mu + \tau_i + \gamma_2 + (\tau\gamma)_{i2}]$$
$$\boxed{= \gamma_1 - \gamma_2 + (\tau\gamma)_{i1} - (\tau\gamma)_{i2}}$$

Like before, this result is dependent on the second treatment due to the interaction term. Therefore, depending on if the observation is given treatment $A$ or $B$ the result will vary by the $i$ index where $i = 1$ or 2 respectively.

The third part asks us to estimate the difference in the expected value of $y$ among the four treatment combinations $Aa$, $Ab$, $Ba$, $Bb$. Similar to part a), there will be a total of $\binom{4}{2} = 6$ different combinations. These different scenarios can be seen below in Table 3. It is possible to add more to for example show $Ab - Aa$, but again that will be ignored for simplicity since the results are roughly the same.

| Compared differences | Difference in expected value |
|---|---|
| $Aa - Ab$ | $E(Y_{11k}) - E(Y_{12k}) = \mu + \tau_1 + \gamma_1 + (\tau\gamma)_{11} - [\mu + \tau_1 + \gamma_2 + (\tau\gamma)_{12}]$ $\boxed{= \gamma_1 - \gamma_2 + (\tau\gamma)_{11} - (\tau\gamma)_{12}}$ |
| $Aa - Ba$ | $E(Y_{11k}) - E(Y_{21k}) = \mu + \tau_1 + \gamma_1 + (\tau\gamma)_{11} - [\mu + \tau_2 + \gamma_1 + (\tau\gamma)_{21}]$ $\boxed{= \tau_1 - \tau_2 + (\tau\gamma)_{11} - (\tau\gamma)_{12}}$ |
| $Aa - Bb$ | $E(Y_{11k}) - E(Y_{22k}) = \mu + \tau_1 + \gamma_1 + (\tau\gamma)_{11} - [\mu + \tau_2 + \gamma_2 + (\tau\gamma)_{22}]$ $\boxed{= \tau_1 - \tau_2 + \gamma_1 - \gamma_2 + (\tau\gamma)_{11} - (\tau\gamma)_{22}}$ |
| $Ab - Ba$ | $E(Y_{12k}) - E(Y_{21k}) = \mu + \tau_1 + \gamma_2 + (\tau\gamma)_{12} - [\mu + \tau_2 + \gamma_1 + (\tau\gamma)_{21}]$ $\boxed{= \tau_1 - \tau_2 - \gamma_1 + \gamma_2 + (\tau\gamma)_{12} - (\tau\gamma)_{21}}$ |
| $Ab - Bb$ | $E(Y_{12k}) - E(Y_{22k}) = \mu + \tau_1 + \gamma_2 + (\tau\gamma)_{12} - [\mu + \tau_2 + \gamma_2 + (\tau\gamma)_{22}]$ $\boxed{= \tau_1 - \tau_2 + (\tau\gamma)_{12} - (\tau\gamma)_{22}}$ |
| $Ba - Bb$ | $E(Y_{21k}) - E(Y_{22k}) = \mu + \tau_2 + \gamma_1 + (\tau\gamma)_{21} - [\mu + \tau_2 + \gamma_2 + (\tau\gamma)_{22}]$ $\boxed{= \gamma_1 - \gamma_2 + (\tau\gamma)_{21} - (\tau\gamma)_{22}}$ |

*Table 3 The above table shows the difference in expected value for the various combinations of treatment levels.*