AUC minus baseline contrast

We are fitting this model: $y_{i,j,g} = \alpha + \theta g + \beta_j + \gamma_j * g + \epsilon_{i,j}$ for $j = 1, \ldots, n$ and $g = 0, 1$ with $\beta_1 = \gamma_1 = 0$

- Calculate the expectation $E(y_{i,j,g})$ for different j and g.
- Let $\mu_{j,g} = E(y_{i,j,g})$. Then get the AUC minus baseline (baseline expectation is $\mu_{1,g}$) for $(\mu_{1,g}, \ldots, \mu_{n,g})$ when plotted against $(1, \ldots, n)$. AUC minus baseline is essentially the observed area under the $(\mu_{1,g}, \ldots, \mu_{n,g})$ - area assuming $\mu_{j,g} = \mu_{1,g}$ for all j. Let this quantity be AUCMB_g
- If the output of regression model looks like $\boldsymbol{\kappa} = (\alpha, \theta, \beta_2, \ldots, \beta_n, \gamma_2, \ldots, \gamma_n)$. Get $\boldsymbol{\ell}$ such that $\boldsymbol{\ell}^T\boldsymbol{\kappa} = \text{AUCMB}_1$ - AUCMB_0.

To calculate the area, you can use trapezoidal rule as the plot $(\mu_{1,g}, \ldots, \mu_{n,g})$ vs $(1, \ldots, n)$ will look like $n - 1$ trapezoids placed side by side.

The area of a trapezoid is $\frac{1}{2}(a + b)h$ with height h and the top and bottom sides are of length a and b.

Trapezoidal Rule

The Trapezoidal Rule is a numerical method for approximating definite integrals. Given a function $f(x)$, the definite integral $\int_a^b f(x)\,dx$ can be approximated using the formula:

$$\int_a^b f(x)\,dx \approx \frac{h}{2}\left[(f(a) + f(x_2)) + (f(x_2) + f(x_3)) + \ldots + (f(x_{n-1}) + f(b))\right]$$
$$= \frac{h}{2}\left[f(a) + 2f(x_2) + 2f(x_3) + \ldots + 2f(x_{n-1}) + f(b)\right]$$

where h is the width of each subinterval and n is the number of subintervals. The subinterval width h is given by:

$$h = \frac{b - a}{n - 1} \tag{1}$$

The points $x_2, x_3, \ldots, x_{n-1}$ divide the interval $[a, b]$ into $n - 1$ equal subintervals, assuming $x_1 = a$ and $x_n = b$.

Example

Let's consider an example. Suppose we want to approximate the integral $\int_0^1 x^2\,dx$ using the Trapezoidal Rule with 4 subintervals.

The width of each subinterval is:

$$h = \frac{1-0}{4} = 0.25 \tag{2}$$

The Trapezoidal Rule approximation is then:

$$\int_0^1 x^2\,dx \approx \frac{0.25}{2}\left[0 + 2(0.25^2) + 2(0.5^2) + 2(0.75^2) + 1\right] \tag{3}$$

Simplifying this expression will give the numerical approximation.

AUCMB$_g$

$\mu_{j,g} = \alpha + g\theta + \beta_j + g\gamma_j$.

We have the means at $(\mu_{1,g}, \ldots, \mu_{n,g})$ at time points $(1, \ldots, n)$. Thus, $a = 1, b = n$ Here, $h = (n-1)/(n-1) = 1$

Hence, AUC is $\int_1^n \mu_{t,g}dt \approx \frac{1}{2}[\mu_{1,g} + 2\mu_{2,g} + \ldots + 2\mu_{n-1,g} + \mu_{n,g}] = \frac{1}{2}[\alpha + g\theta + 2\sum_{j=2}^{n-1}\{\alpha + g\theta + \beta_j + g\gamma_j\} + \alpha + g\theta + \beta_n + g\gamma_n]$

AUCMB$_g$ is $\int_1^n \mu_{t,g}dt - \int_1^n \mu_{1,g}dt \approx \frac{1}{2}[\alpha + g\theta + 2\sum_{j=2}^{n-1}\{\alpha + g\theta + \beta_j + g\gamma_j\} + \alpha + g\theta + \beta_n + g\gamma_n] - \mu_{1,g}(n-1)$.

Thus, AUCMB$_1$ - AUCMB$_0$ = $\frac{1}{2}[\alpha + 1\theta + 2\sum_{j=2}^{n-1}\{\alpha + 1\theta + \beta_j + 1\gamma_j\} + \alpha + 1\theta + \beta_n + 1\gamma_n] - \mu_{1,1}(n-1) - \frac{1}{2}[\alpha + 0\theta + 2\sum_{j=2}^{n-1}\{\alpha + 0\theta + \beta_j + 0\gamma_j\} + \alpha + 1\theta + \beta_n + 0\gamma_n] + \mu_{1,0}(n-1)$.

This simplifies to $= \frac{1}{2}[\theta + 2\sum_{j=2}^{n-1}\{\theta + \gamma_j\} + \theta + \gamma_n] - (n-1)(\mu_{1,1} - \mu_{1,0}) = \frac{1}{2}[2(n-1)\theta + 2\sum_{j=2}^{n-1}\gamma_j + \gamma_n] - (n-1)(\theta) = (n-1)\theta - (n-1)\theta + \sum_{j=2}^{n-1}\gamma_j + \gamma_n/2 = \sum_{j=2}^{n-1}\gamma_j + \gamma_n/2$.

Hence $\boldsymbol{\ell} = (0, 0, 0, \ldots, 0, 1, \ldots, 1/2)$, only the last $n-1$ entries has non-zero values.

$$\begin{aligned}\boldsymbol{\ell} &= (0, 0, 0, \ldots, 0, 1, \ldots, 1/2)\\ \boldsymbol{\kappa} &= (\alpha, \theta, \beta_2, \ldots, \beta_n, \gamma_2, \ldots, \gamma_n)\end{aligned}$$

Then, $\boldsymbol{\ell}^T\boldsymbol{\kappa} = \sum_{j=2}^{n-1}\gamma_j + \gamma_n/2$ =AUCMB$_1$ - AUCMB$_0$.

1 General observation times

$$\int_a^b f(x)\,dx \approx \frac{x_2 - a}{2}\{f(a) + f(x_2)\} + \frac{x_3 - x_2}{2}\{f(x_2) + f(x_3)\} + \ldots + \frac{b - x_{n-1}}{2}\{f(x_{n-1}) + f(b)\}$$

If the observation times are $t_1, \ldots, t_n$, then AUC becomes $= \int_{t_1}^{t_n} \mu_{t,g} dt \approx \frac{t_2-t_1}{2}\{\mu_{t_1,g} + \mu_{t_2,g}\} + \frac{t_3-t_2}{2}\{\mu_{t_2,g} + \mu_{t_3,g}\} + \ldots + \frac{t_n-t_{n-1}}{2}\{\mu_{t_n,g} + \mu_{t_{n-1},g}\} = \frac{t_2-t_1}{2}\mu_{t_1,g} + \frac{t_3-t_1}{2}\mu_{t_2,g} + \frac{t_4-t_2}{2}\mu_{t_3,g} + \ldots + \frac{t_n-t_{n-2}}{2}\mu_{t_{n-1},g} + \frac{t_n-t_{n-1}}{2}\mu_{t_n,g}$.

Thus, $\mathrm{AUCMB}_g = \frac{t_2-t_1}{2}\mu_{t_1,g} + \frac{t_3-t_1}{2}\mu_{t_2,g} + \frac{t_4-t_2}{2}\mu_{t_3,g} + \ldots + \frac{t_n-t_{n-2}}{2}\mu_{t_{n-1},g} + \frac{t_n-t_{n-1}}{2}\mu_{t_n,g} - (t_n - t_1)\mu_{t_1,g}$.

Expressing the $\mu_{t,g}$'s in terms of the regression coefficients in the output, we can get the contrast for AUCMB_g and also for AUCMB_1-AUCMB_0.