Tukey's bi-weight

- Calculate probeset median $T_0 = median(X_i)$
- For each probe in a probeset, calculate

$$u_i = \frac{X_i - T_0}{c * S_n + \varepsilon}$$

  c is a tuning constant, S_n is a scale estimate (Median Absolute Deviation, MAD), ε is a small positive constant to ensure we are not dividing by 0 (typically, 0.0001)
- Update our estimate T_0 to obtain a better estimate of location

$$T^* = \frac{\sum w(u_i) * X_i}{\sum w(u_i)}$$

$$w(u_i) = \begin{cases} (1-u^2)^2, & |u| \leq 1 \\ 0, & |u| > 1 \end{cases}$$

  $w(u_i)$ - Tukey's biweight (M-estimator)

MAS5

- First, subtract background. Divide array in 16 rectangular zones. For each zone, background is the average of the lowest 2% probe cell intensities. Denote bZk - background of zone
- Calculate a weight. For a probe located at cell coordinates (X, Y) the distance from the probe to the center of each of the K zones is calculated dk(X, Y)
- $W_k(X,Y) = \frac{1}{d_k^2(X,Y) + Smooth}, Smooth = 100$
- Weighted background estimate for that probe

$$b(X,Y) = \frac{1}{\sum_{i=1}^{k} W_i(X,Y)} \sum_{i=1}^{k} W_i(X,Y) * bZk$$

- Previous steps are the "background correct" method = "mas" in expresso function

Pmcorrect method – to avoid negative values for PM-MM

- Calculate "specific background" $SB_i = Tukey's\ biweight(log_2(PM_{ij}) - log_2(MM_{ij}))$
- "Idealized mismatch value. Different scale values

$$IM_{ij} = \begin{cases} MM_{ij}, & if MM < PM \\ PM_{ij}/2^{SB_i}, & if MM \geq PM\ and\ SB > \tau \\ PM_{ij} \Big/ 2^{\frac{\tau}{1+\frac{\tau - SB_i}{scale\tau}}}, & if\ MM \geq PM\ and\ SB \geq \tau \end{cases}$$

- $\tau = 0, and\ scale\tau = 10$
- Calculate Tukey biweight (second time) estimate for each probeset as

$$2^{Tukey's\ biweight(log_2(PM_{ij} - IM_{ij}))}$$

Li & Wong method

- IxJ equations

- I θ_i array parameters, J ϕ_j gene parameters, I+J all parameters
- Assume ϕ_j is known, use to find best θ_i. Then, use θ_i estimates to estimate ϕ_j
- Iterative least squares procedure

RMA

- True signal follows some exponential distribution ~$\exp(\alpha)$, and the background follows some normal distribution ~$N(\mu, \sigma^2)$. True signal and background are independent. Estimate background (unclear, which procedure), then adjust PM values by BG.
- Carry out quantile normalization
- Take log2
- RMA expression summarization results from a **median polish**

$$log_2(Background\ correct\ PM)_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

- μ is the overall effect; α_i is the array effect; β_j is the probe effect. $\hat{\mu} + \alpha_i$ is the expression value for the probeset on array i
- How-to: Initialize all effects to 0. $m^{(0)} = 0$, $a_i^{(0)} = 0$, $b_j^{(0)} = 0$
- Find row medians, subtract them from all observations in the corresponding row values.
- Find the column medians, then subtract ("polish") them from the corresponding column values. Update the effects:
    - $\Delta m_a^{(n)} = median(a_i^{(n-1)} + \Delta a_i^{(n)})$
    - $\Delta m_b^{(n)} = median(b_j^{(n-1)})$
    - $m^{(n)} = m^{(n)} + \Delta m_a^{(n)} + \Delta m_b^{(n)}$
    - $where$
    - $a_i^{(n)} = a_i^{(n-1)} + \Delta a_i^{(n)} - \Delta m_a^{(n)}$
    - $b_j^{(n)} = b_j^{(n-1)} + \Delta b_j^{(n)} - \Delta m_b^{(n)}$
- p. 481, section 15.7.1 explains analysis of variance principles – translate them to median polish approach.
-