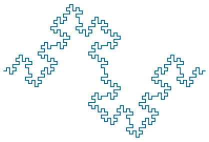
High-Throughput Sequencing Course

Statistical Interaction

Biostatistics and Bioinformatics

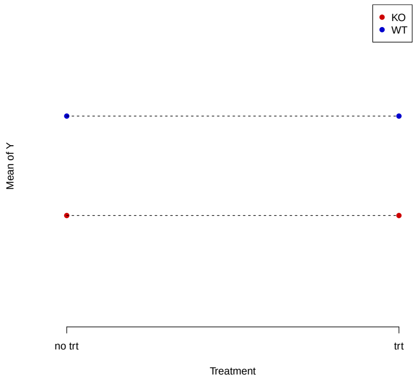


Summer 2019

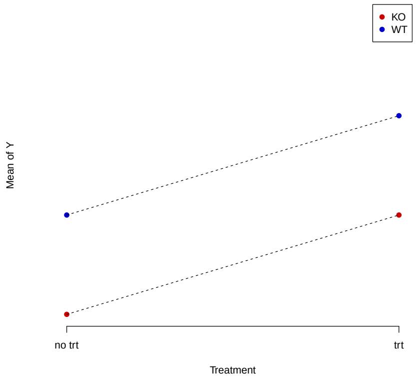
Duke University
School of Medicine

DUKE BIOSTATISTICS & BIOINFORMATICS

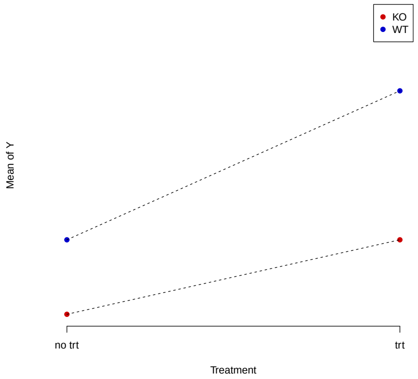
EXAMPLE 1: NO INTERACTION



EXAMPLE 2: NO INTERACTION



EXAMPLE 3: INTERACTION



MODEL INTERACTION

- Y denotes the gene expression
- Let x denote the treatment indicator
    - $x = 0$ if not treated or 1 if treated
- Let z denote the knock-out indicator
    - $z = 0$ is WT or 1 otherwise
- The expected value of Y given treatment indicator x and knock out indicator z is denoted by

$$\mu_{x,z} = E[Y|X = x, Z = z]$$

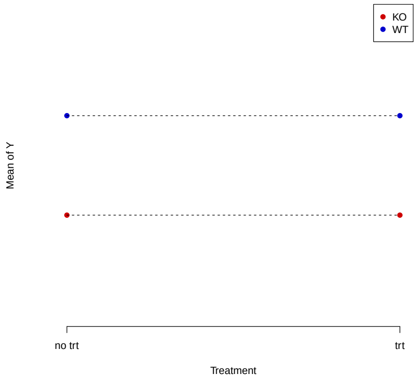
- The model will be

$$Y = \mu_{x,z} + \epsilon$$

  where ϵ is a the measurement error

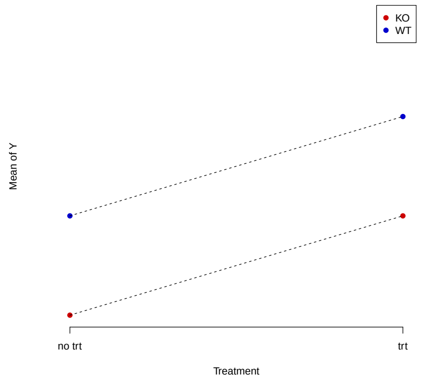
EXAMPLE 1: LINEAR MODEL FOR NO INTERACTION

$$Y = \beta_0 + \beta_1 z + \epsilon \ (\beta_2 = 0)$$



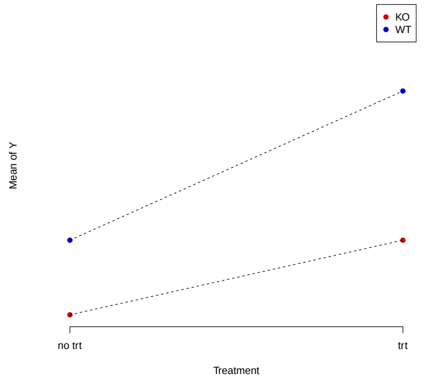
EXAMPLE 2: NO INTERACTION

$$Y = \beta_0 + \beta_1 z + \beta_2 x + \epsilon$$



EXAMPLE 3: INTERACTION

$$Y = \beta_0 + \beta_1 z + \beta_2 x + \beta_3 xz + \epsilon$$



INTERACTION EXAMPLES

- Example 1: What are the signs for β_0 and β_1?
- Example 2: What are the signs for β_0, β_1 and β_2?
- Example 2: What are the signs for $\beta_0, \beta_1, \beta_2$ and β_3?

INCORPORATING INTERACTIONS INTO THE NB MODEL

- Conditional on x_j and z_j, the observed number of reads mapped to gene i for sample j, K_{ij}, follows a negative binomial distribution with
  - Mean μ_{ij}
  - Dispersion parameter α_i (gene specific)
- Conditional on the treatment status of sample j ($x_j = 0$ or 1) and the temperature z_j, the expected value of K_{ij} is

$$\mu_{ij} = s_j \times q_{ij}$$

where

$$\log q_{ij} = \beta_{i0} + \beta_{i1}x_j + \beta_{i2}z_j + \beta_{i3}x_j z_j$$

- The normalization parameters are assumed to be sample (not gene) specific ($s_{ij} = s_j$)

2019 PILOT EXPERIMENT

- Condition (X): pH4 (=0) versus pH8 (=1)
- Strain (Z): WT (=0) versus sre1d (=1)
- Negative Binomial Mean Model

$$2^{\beta +_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ}$$

- Log 2 of the mean model

$$\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ$$

2019 PILOT EXPERIMENT

Condition	Genotype	Log 2 mean
0	0	β_0
1	0	$\beta_0 + \beta_1$
0	1	$\beta_0 + \beta_2$
1	1	$\beta_0 + \beta_1 + \beta_2 + \beta_3$

- DE (due to condition) when genotype is WT ($Z = 0$):

$$\begin{aligned} &\{\beta_0 + \beta_1(1) + \beta_2(0) + \beta_3(1)(0)\} \\ &-\{\beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(1)(0)\} \\ &= \beta_1 \end{aligned}$$

- DE (due to treatment) when genotype is sred1 ($Z = 1$):

$$\begin{aligned} &\{\beta_0 + \beta_1(1) + \beta_2(1) + \beta_3(1)(1)\} \\ &-\{\beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(1)(0)\}) \\ &= \beta_1 + \beta_2 + \beta_3 \end{aligned}$$

- Note that there is no interaction if $\beta_3 = 0$

2019 Pilot Experiment

- β_1 is log2 fold change (due to condition) when genotype is WT
- $\beta_1 + \beta_3$ is log2 fold change (due to condition) when genotype is sre1d
- 2^{β_1} is the fold change (due to condition) when genotype is WT
- $2^{\beta_1+\beta_3}$ is the fold change (due to condition) when genotype is sre1d
- The fold change of the two fold changes is

$$\frac{2^{\beta_1+\beta_3}}{2^{\beta_1}} = \frac{2^{\beta_1}2^{\beta_3}}{2^{\beta_1}} = 2^{\beta_3}.$$

When $\beta_3 = 0$ we have $2^0 = 1$