

# Differential Abundance Analysis (DAA)

1. Which methods are available?
2. Rough understanding of assumptions for selected methods
3. How do methods compare?
4. Problems with DAA
5. Perform 3 methods and make simple comparison

# What and Why?

- Statistical analysis of single bacterial abundances
- Test whether a gene has *differential abundance*

Nearing et al. (2021)									FDR Corr.	CoDa	Dev. For
Tool	Input	Norm.	Trans.	Distribution	Covariates	Random Effects	Hypothesis test				
ALDEx2	Counts	None	CLR	Dirichlet-multinomial	Yes*	No	Wilcoxon rank-sum	Yes	Yes	RNA-seq, 16S, MGS	
ANCOM-II	Counts	None	ALR	Non-parametric	Yes	Yes	Wilcoxon rank-sum	Yes	Yes	MGS	
corncob	Counts	None	None	Beta-binomial	Yes	No	Wald (default)	Yes	No	16S, MGS	
DESeq2	Counts	Modified RLE (default is RLE)	None	Negative binomial	Yes	No	Wald (default)	Yes	No	RNA-seq, 16S, MGS	
edgeR	Counts	RLE (default is TMM)	None	Negative binomial	Yes*	No	Exact	Yes	No	RNA-seq	
LEFse	Rarefied relative abundance	TSS	None	Non-parametric	Subclass factor only	No	Kruskal-Wallis	No	No	16S, MGS	
MaAsLin2	Counts	TSS	AST (default is log)	Normal (default)	Yes	Yes	Wald	Yes	No	MGS	
MaAsLin2 (rare)	Rarefied counts	TSS	AST (default is log)	Normal (default)	Yes	Yes	Wald	Yes	No	MGS	
metagenomeSeq	Counts	CSS	Log	Zero-inflated log-Normal	Yes	No	Moderated t	Yes	No	16S. MGS	
limma voom (TMM)	Counts	TMM	Log; Precision weighting	Normal (default)	Yes	Yes	Moderated t	Yes	No	RNA-seq	
limma voom (TMMwsp)	Counts	TMMwsp	Log; Precision weighting	Normal (default)	Yes	Yes	Moderated t	Yes	No	RNA-seq	
t-test (rare)	Rarefied Counts	None	None	Normal	No	No	Welch's t-test	Yes	No	N/A	
Wilcoxon (CLR)	CLR abundances	None	CLR	Non-parametric	No	No	Wilcoxon rank-sum	Yes	Yes	N/A	
Wilcoxon (rare)	Rarefied counts	None	None	Non-parametric	No	No	Wilcoxon rank-sum	Yes	No	N/A	

# Non-Parametric methods: Wilcoxon-Rank-Sum and Kruskall-Wallis Test

- Non-parametric
- Teststatistic based on ranks
- Applied to rarefied counts or log-ratio tranformed abundances
- Adjust p-values

<b>id</b>	<b>group</b>	<b>value</b>
1	A	12
2	A	17
3	A	16
4	A	14
5	A	13
6	B	17
7	B	22
8	B	15
9	B	14
10	B	16

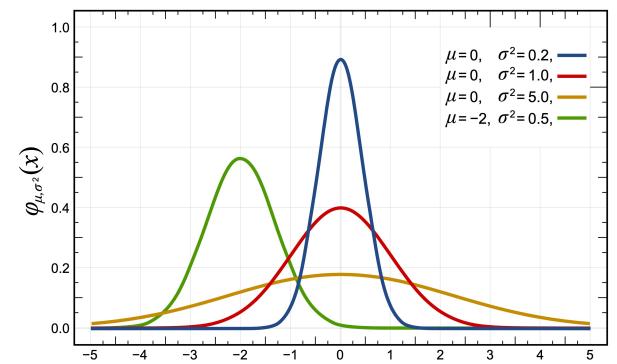


<b>id</b>	<b>group</b>	<b>value</b>	<b>rank</b>
1	A	12	1.0
5	A	13	2.0
4	A	14	3.5
9	B	14	3.5
8	B	15	5.0
3	A	16	6.5
10	B	16	6.5
2	A	17	8.5
6	B	17	8.5
7	B	22	10.0

# Parametric methods: T-Test and ANOVA

$$y_i \sim Normal(\mu, \sigma)$$

$$\mu = \alpha + \beta x_i$$



# Parametric methods: T-Test and ANOVA Example

$$y_i \sim Normal(\mu, \sigma)$$

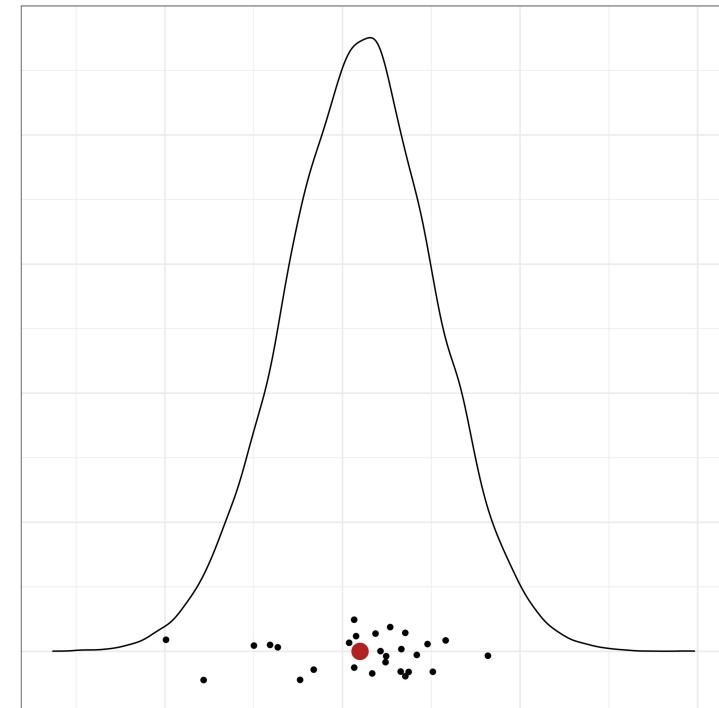
$$\mu = \alpha + \beta x_i$$



$$Akkermansia_i \sim Normal(\mu, \sigma)$$

$$\mu = \alpha$$

$$Akkermansia_i \sim Normal(4.2, 0.84)$$



# Parametric methods: T-Test and ANOVA Example

$$y_i \sim Normal(\mu, \sigma)$$

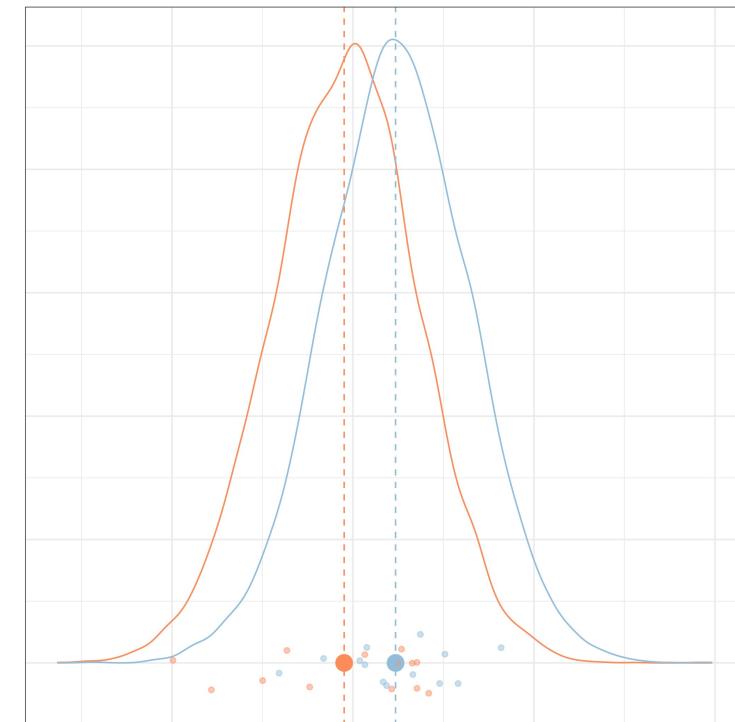
$$\mu = \alpha + \beta x_i$$



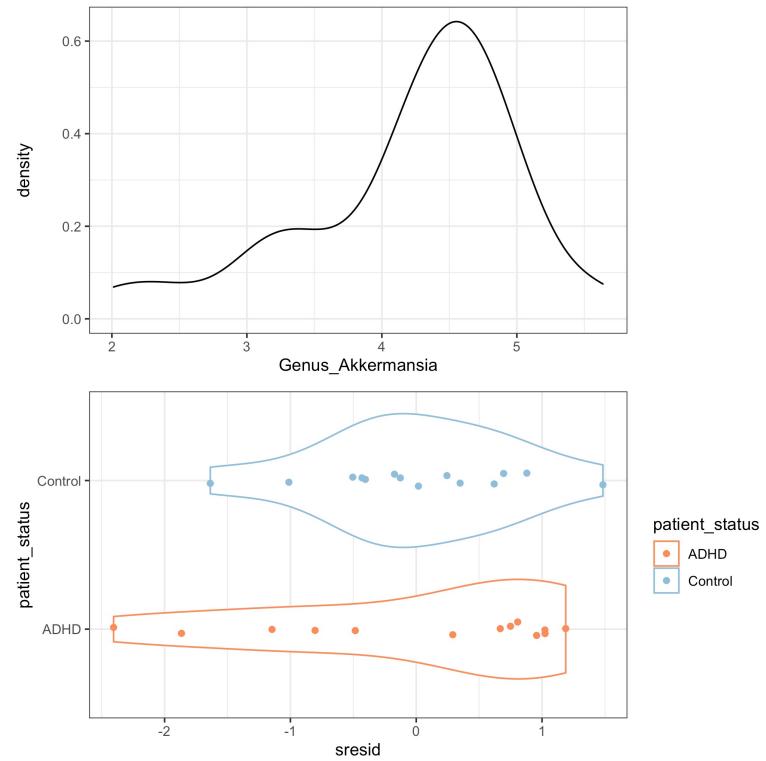
$$Akkermansia_i \sim Normal(\mu, \sigma)$$

$$\mu = \alpha + \beta_{patientstatus_i}$$

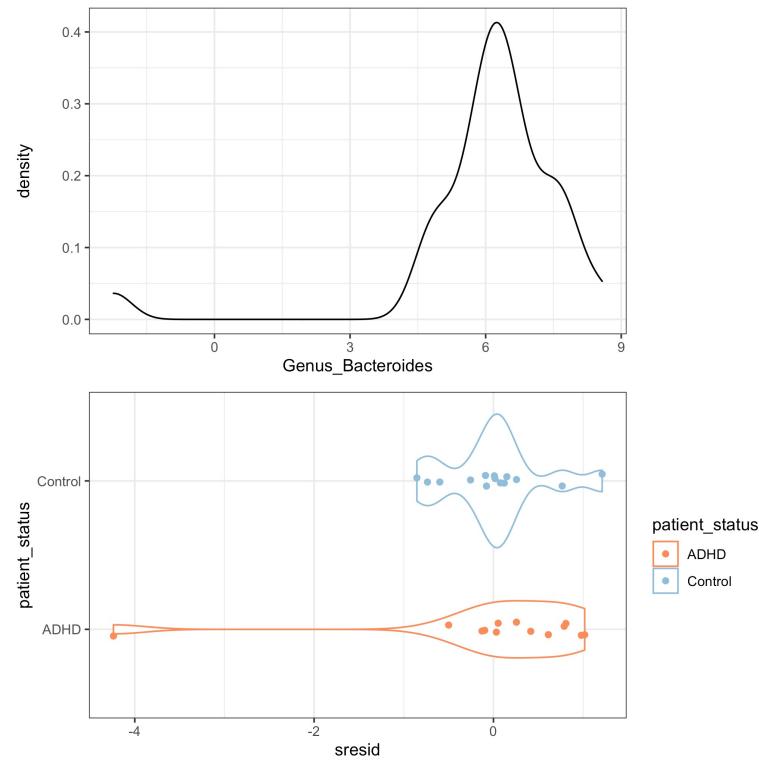
$$Akkermansia_i \sim Normal(3.90 + 0.57 * patientstatus_i, 0.80)$$



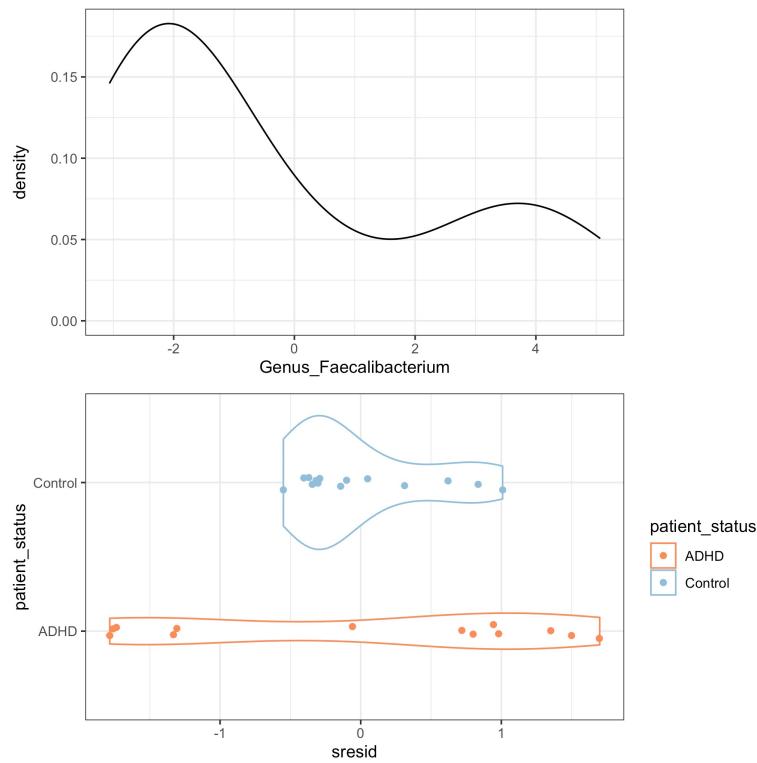
# Parametric methods: T-Test and ANOVA Example



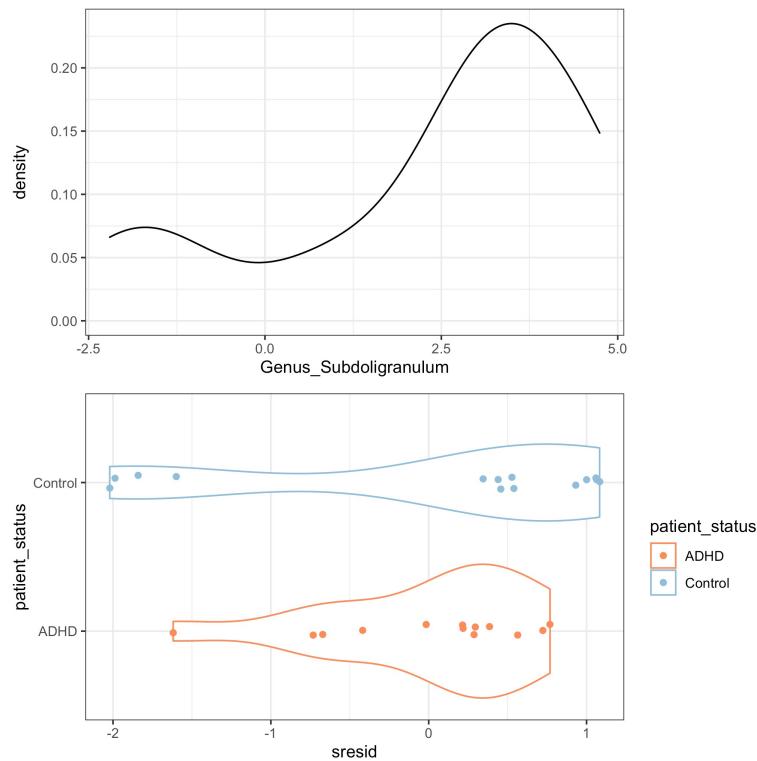
# Parametric methods: T-Test and ANOVA Example



# Parametric methods: T-Test and ANOVA Example



# Parametric methods: T-Test and ANOVA Example



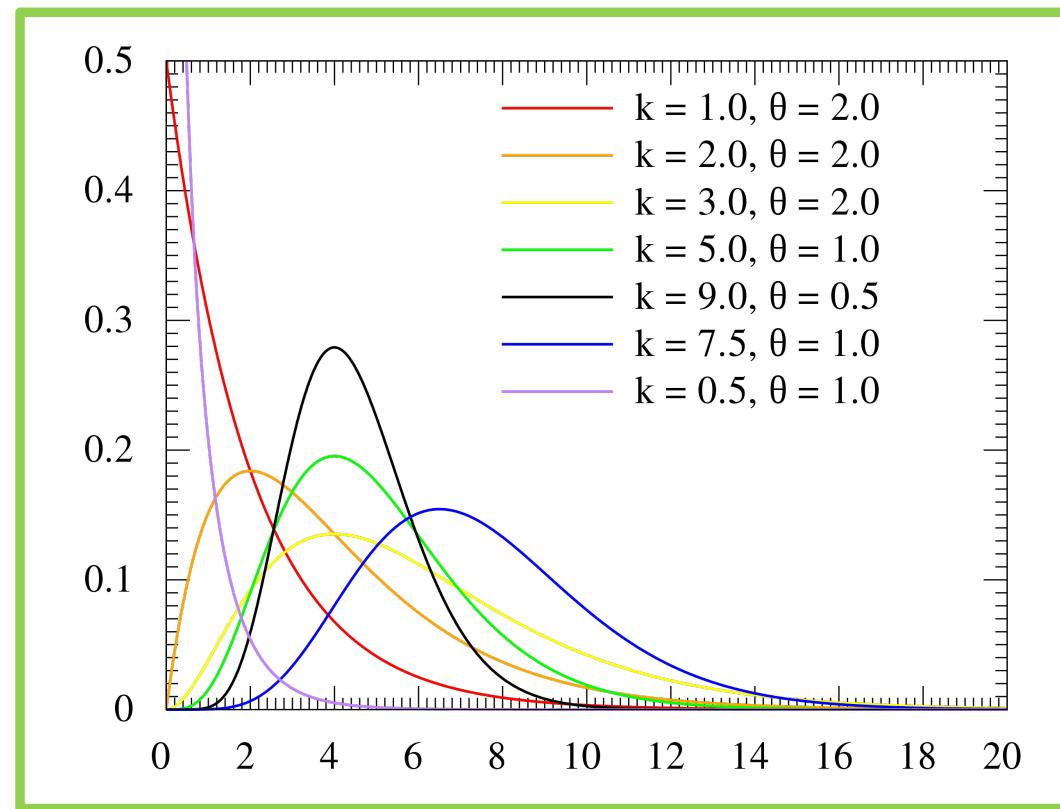
# Parametric methods: Gamma-Poisson Distribution

$$y_i \sim Poisson(\lambda_i)$$

$$\log(\lambda_i) = \alpha + \beta x_i$$

$$y_i \sim gamma(\mu, \theta)$$

$$\log(\mu) = \alpha + \beta x_i$$



# Parametric methods: Gamma-Poisson Distribution

$$y_i \sim \text{gamma}(\mu, \theta)$$

$$\log(\mu) = \alpha + \beta x_i$$

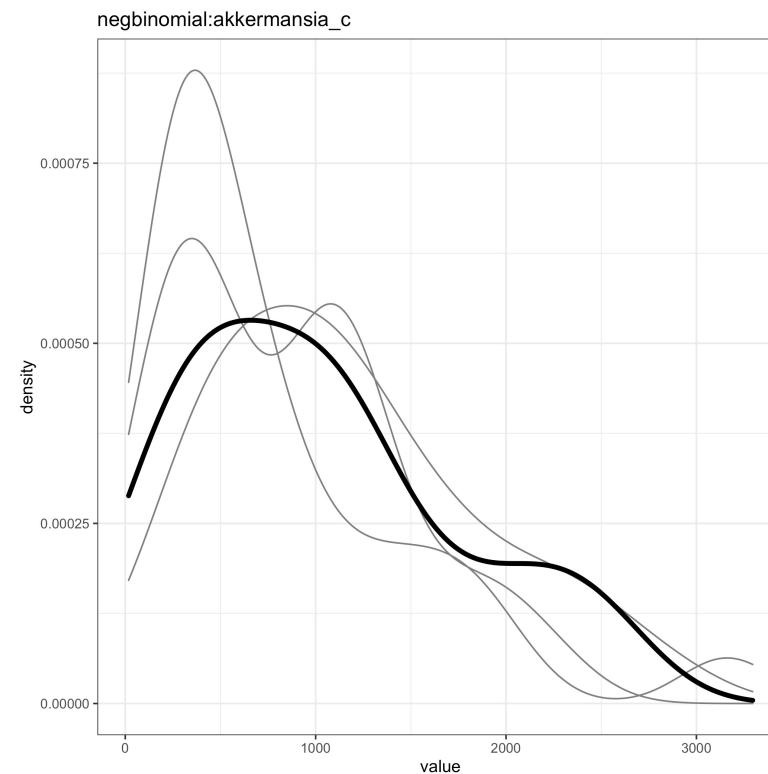


$$Akkermansia_i \sim \text{gamma}(\mu, \theta)$$

$$\log(\mu) = \alpha + \beta_{\text{patientstatus}_i}$$



Average difference = 454

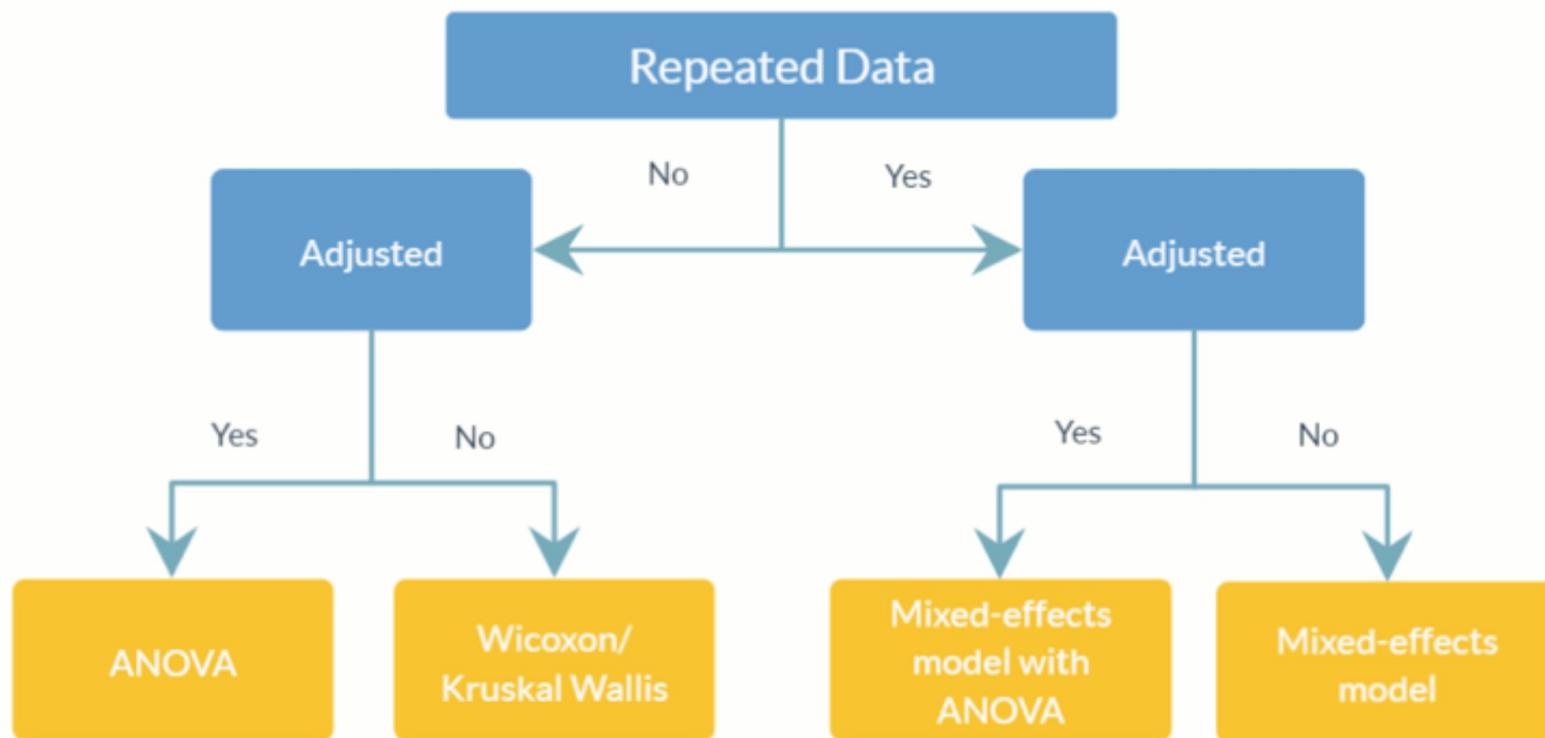


# Parametric methods: DESeq2

- Parametric model for count data
- Dependence of variance on the mean
- Uses shrinkage (partial pooling) to estimate
  - variance of each gene
  - Fold change

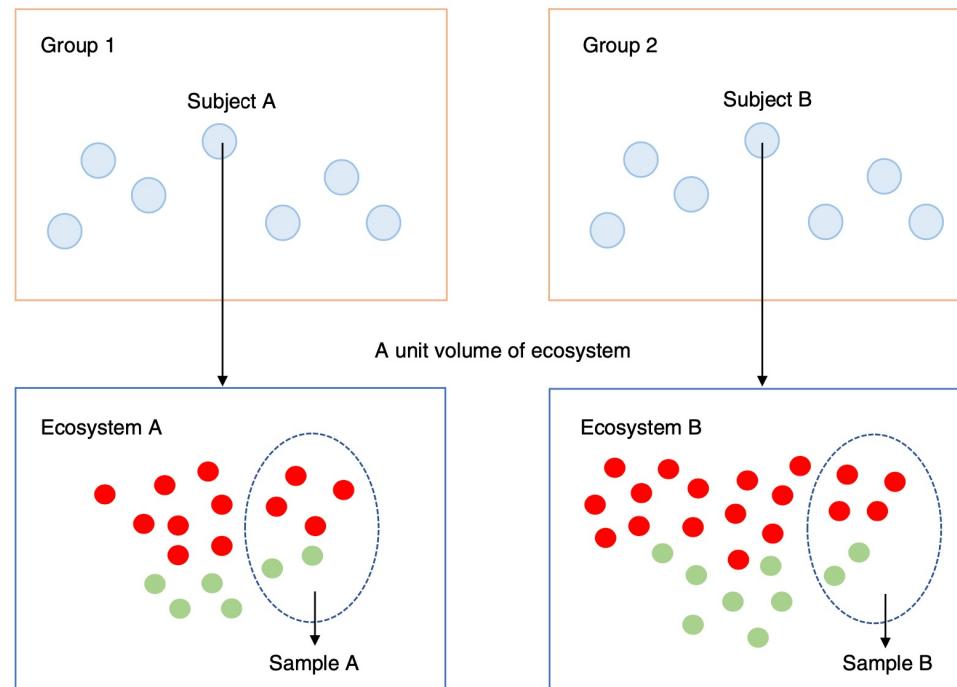
# Parametric methods: \*

## ANCOM



Mandal et al. (2015)

# Parametric methods: ANCOM-BC



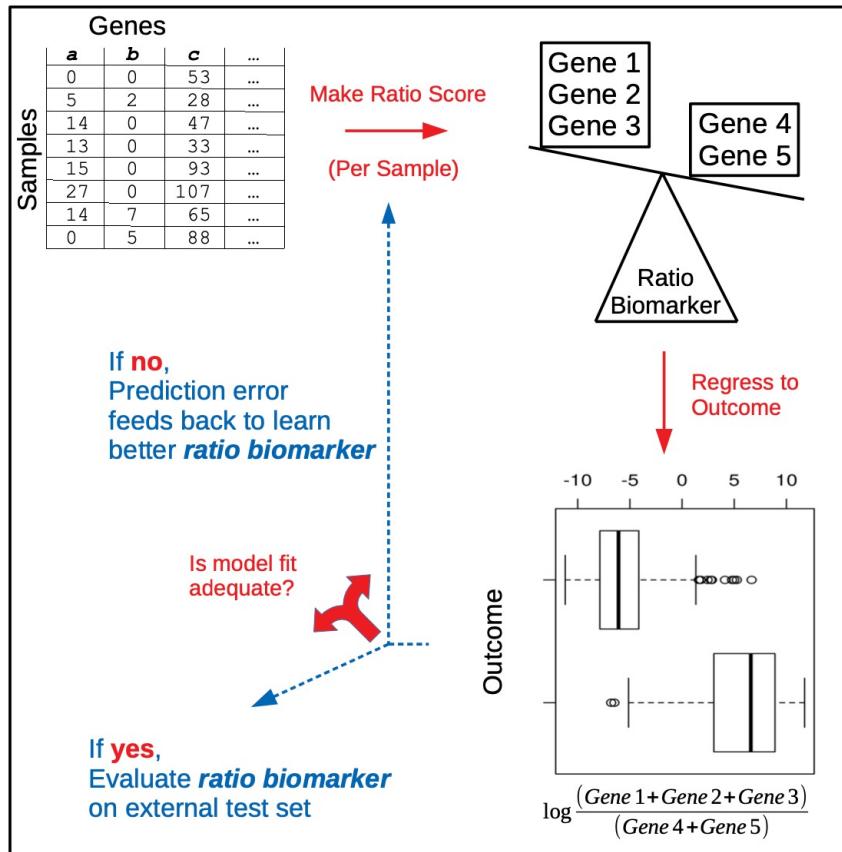
Huang Lin & Shyamal Das Peddada (2021)

Nearing et al. (2021)											
Tool	Input	Norm.	Trans.	Distribution	Covariates	Random Effects	Hypothesis test	FDR Corr.	CoDa	Dev. For	
ALDEx2	Counts	None	CLR	Dirichlet-multinomial	Yes*	No	Wilcoxon rank-sum	Yes	Yes	RNA-seq, 16S, MGS	
ANCOM-II	Counts	None	ALR	Non-parametric	Yes	Yes	Wilcoxon rank-sum	Yes	Yes	MGS	
corncob	Counts	None	None	Beta-binomial	Yes	No	Wald (default)	Yes	No	16S, MGS	
DESeq2	Counts	Modified RLE (default is RLE)	None	Negative binomial	Yes	No	Wald (default)	Yes	No	RNA-seq, 16S, MGS	
edgeR	Counts	RLE (default is TMM)	None	Negative binomial	Yes*	No	Exact	Yes	No	RNA-seq	
LEFse	Rarefied relative abundance	TSS	None	Non-parametric	Subclass factor only	No	Kruskal-Wallis	No	No	16S, MGS	
MaAsLin2	Counts	TSS	AST (default is log)	Normal (default)	Yes	Yes	Wald	Yes	No	MGS	
MaAsLin2 (rare)	Rarefied counts	TSS	AST (default is log)	Normal (default)	Yes	Yes	Wald	Yes	No	MGS	
metagenomeSeq	Counts	CSS	Log	Zero-inflated log-Normal	Yes	No	Moderated t	Yes	No	16S, MGS	
limma voom (TMM)	Counts	TMM	Log; Precision weighting	Normal (default)	Yes	Yes	Moderated t	Yes	No	RNA-seq	
limma voom (TMMwsp)	Counts	TMMwsp	Log; Precision weighting	Normal (default)	Yes	Yes	Moderated t	Yes	No	RNA-seq	
t-test (rare)	Rarefied Counts	None	None	Normal	No	No	Welch's t-test	Yes	No	N/A	
Wilcoxon (CLR)	CLR abundances	None	CLR	Non-parametric	No	No	Wilcoxon rank-sum	Yes	Yes	N/A	
Wilcoxon (rare)	Rarefied counts	None	None	Non-parametric	No	No	Wilcoxon rank-sum	Yes	No	N/A	

# Critique Quinn et al. (2021)

- Specific to approaches: unverifiable assumptions and unreliable constructs
- DA in general is reductionist: Ignores biological interdependence

# Recommendations Quinn et al. (2021)



Qinn et al. (2021)

# Gather experience

- Perform 3 of the methods used in Nearing et al (2021)
- Compare results among them

# References

1. Nearing JT, Comeau AM, Langille MGI (2021): Identifying biases and their potential solutions in human microbiome studies. *Microbiome* 9: 113.
2. Love MI, Huber W, Anders S (2014): Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550.
3. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD (2015): Analysis of composition of microbiomes: A novel method for studying microbial composition. *Microbial Ecology in Health & Disease* 26. <https://doi.org/10.3402/mehd.v26.27663>
4. Lin H, Peddada SD (2020): Analysis of compositions of microbiomes with bias correction. *Nat Commun* 11: 3514.
5. Quinn TP, Gordon-Rodriguez E, Erb I (2021): A Critique of Differential Abundance Analysis, and Advocacy for an Alternative. *arXiv:210407266 [q-bio, stat]*. Retrieved from <http://arxiv.org/abs/2104.07266>