Batch effect problem 1

2023-03-25

Table illustrating batch effects (outcomes of samples in the 5 batches)

batch 1	batch 2	batch 3	batch 4	batch 5
Normal	Normal	sTCC+CIS	mTCC	Biopsy
Normal	Normal	sTCC-CIS	mTCC	Biopsy
Normal	Normal	sTCC+CIS	mTCC	Biopsy
Normal	Normal	sTCC+CIS	mTCC	Biopsy
	sTCC-CIS	sTCC+CIS	mTCC	Biopsy
	sTCC-CIS	sTCC+CIS	mTCC	
	sTCC-CIS	sTCC+CIS	mTCC	
	sTCC-CIS	sTCC+CIS	mTCC	
	mTCC	sTCC+CIS	mTCC	
	sTCC-CIS	sTCC+CIS	mTCC	
	sTCC-CIS	sTCC+CIS	mTCC	
	sTCC-CIS	sTCC+CIS		
	sTCC-CIS	sTCC-CIS		
	sTCC-CIS	sTCC-CIS		
	sTCC-CIS	sTCC+CIS		
	sTCC-CIS	Biopsy		
	sTCC-CIS	Biopsy		
	sTCC-CIS	Biopsy		
		Biopsy		

How are variables distributed among 5 batches?

Batch 5 consists only of biopsy samples. Biopsy samples are mostly in batch 5 and only one is in batch 3. Normal samples are only in batches 1 and 2, batch 1 consists only from normal samples. In batch 4 there are only mTCC samples and there is only one mTCC sample in different batch (2). Samples with present carcinoma in situ (+CIS) were only in batch 3, and almost all -CIS samples were in batch 2. The problem with these data was that CIS condition was strongly connected to time of analysing samples (different batches were processed at different times). Different reasons for batch effect in outcome variable could be processing batches in different laboratories or different operator that can be connected to using different experiment protocols.