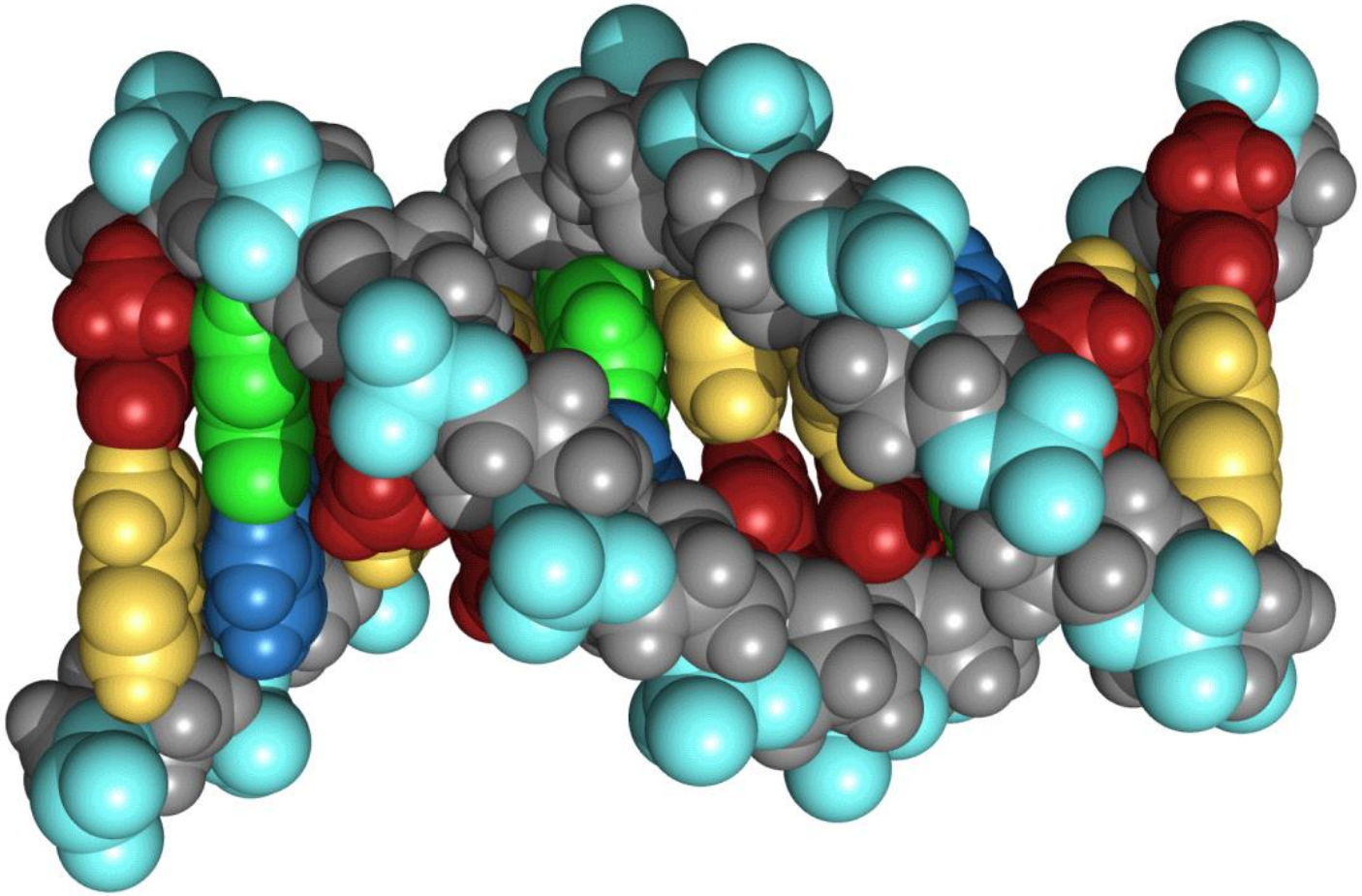
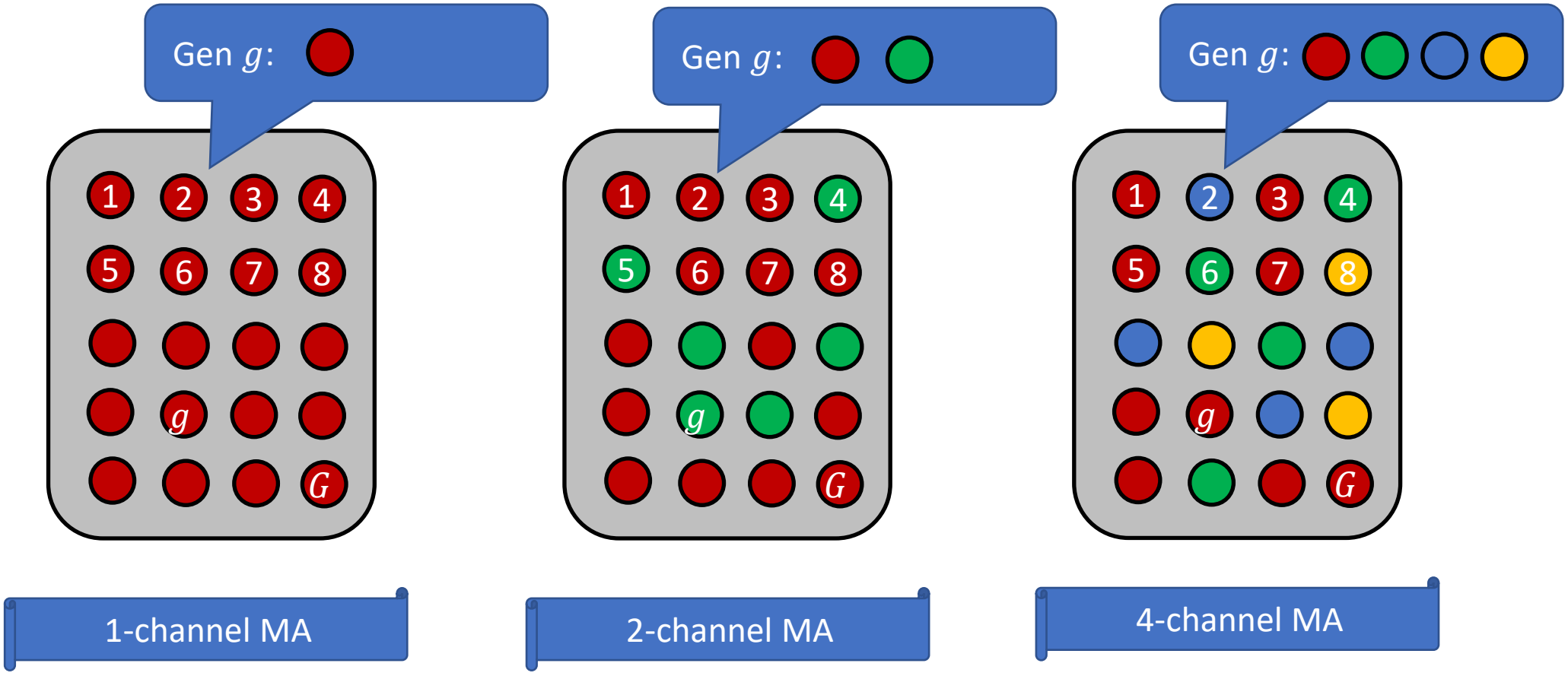
Statistiek 3 BIN

- Les 10

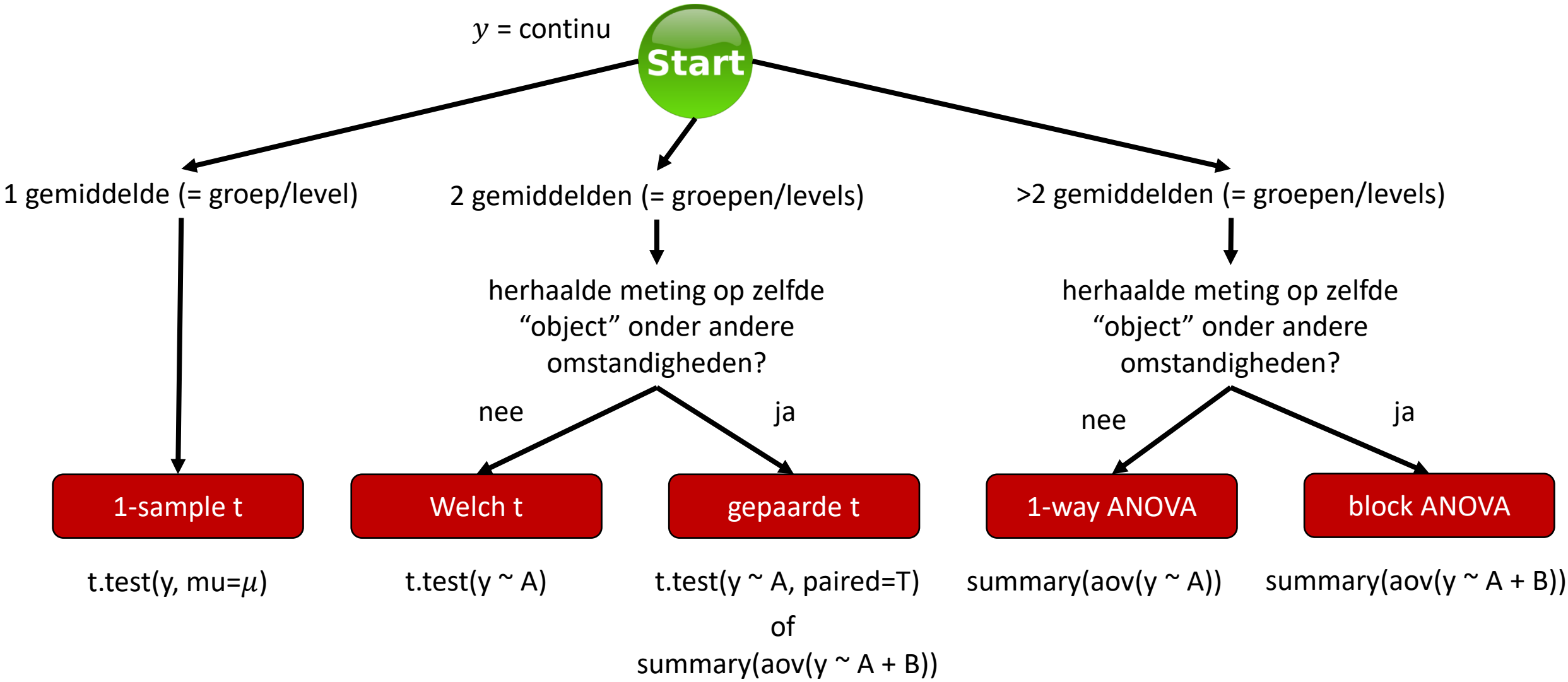


Typen microarray's

Er zijn verschillende typen microarray's:



Keuzeschema toetsen



Notatie

- y = ^{2}log intensiteitwaarde van microarray (MA)
- We bekijken G verschillende genen (= stippen) op MA: $g = 1, 2, \cdots, G$
- Factor A (bijv. Gezondheid) heeft a levels (groepen): $i = 1, 2, \cdots, a$

Dus: $(y_g)_i$ is de log intensiteit van gen g in de i^{de} groep van gezondheid;

per gen g voeren we een statistische toets uit om te zien of er verschil zit tussen de levels van factor A (d.w.z., is $p_{\mathrm{A}} < \alpha$?)

- Bij 1-channel MA's is elke MA een *replica* meting binnen groep i: $k = 1, 2, \cdots, n_i$
- Bij 2- of 4-channel MA's worden meerdere groepen i tegelijkertijd gemeten, dus een *herhaalde* meting m.b.v. *hetzelfde* MA: $j = 1, 2, \cdots, b$; nr. MA is dus eigenlijk een "ruisfactor" B met b levels

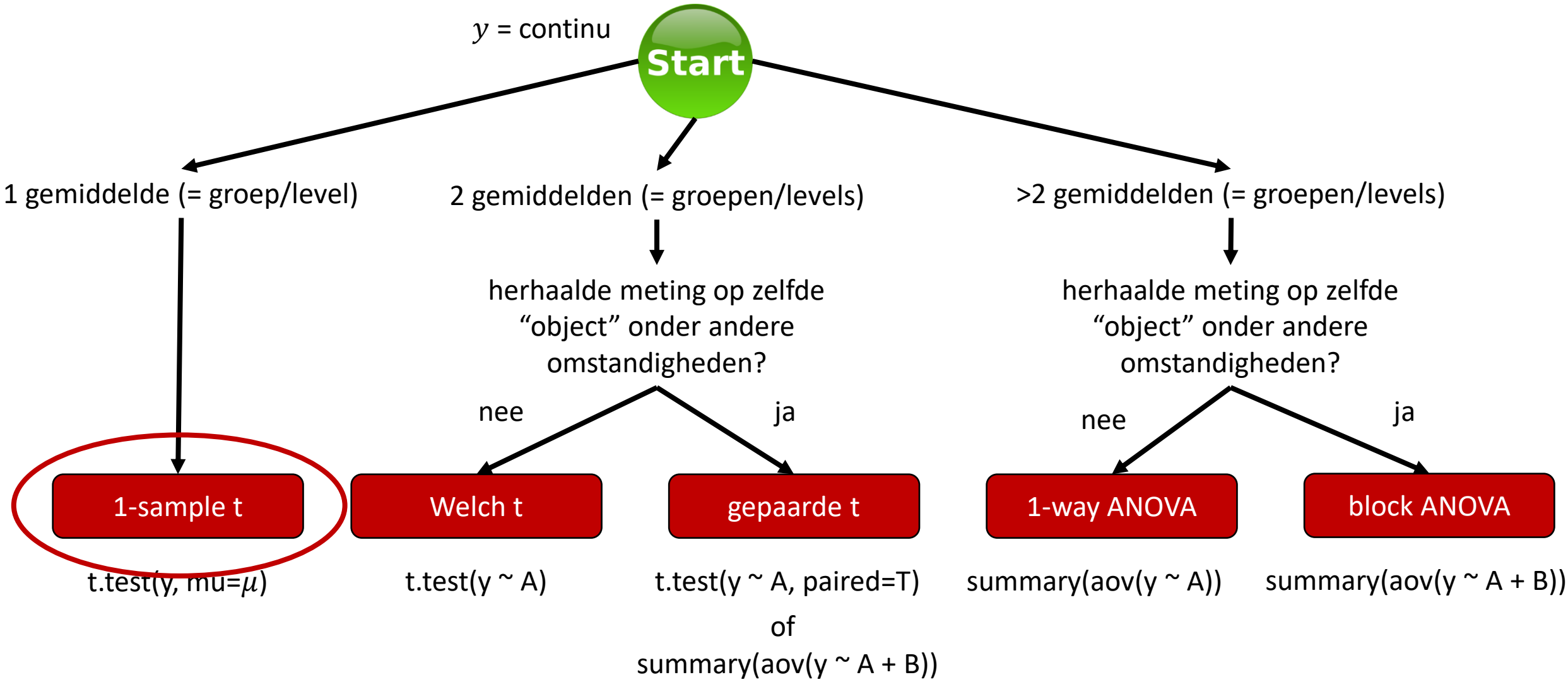
Dus: $(y_g)_{i,k}$ is de log intensiteit van gen g in groep i van k^e 1-channnel MA

Dus: $(y_g)_{i,j}$ is de log intensiteit van gen g in groep i van j^e multi-channel MA

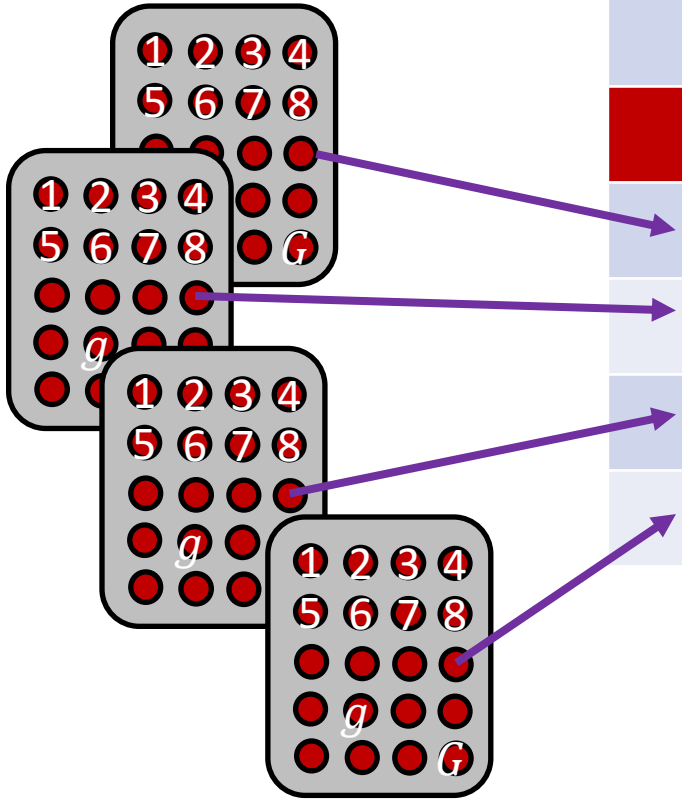
MA data format

		$i = 1$				$i = 2$			
		logR.1	logR.2	logR.3	⋯	logG.1	logG.2	logG.3	⋯
$g = 1$	Gen001	2.58	2.61	2.51	⋯	2.01	1.97	2.14	⋯
$g = 2$	Gen002	3.25	3.21	3.24	⋯	3.58	3.59	3.48	⋯
$g = 3$	Gen003	4.05	4.11	3.98	⋯	4.01	3.99	4.07	⋯
$g = 4$	Gen004	1.59	1.54	1.66	⋯	1.64	1.59	1.55	⋯
$g = 5$	Gen005	2.11	2.17	2.15	⋯	0.48	0.57	0.64	⋯
$g = 6$	Gen006	2.44	2.45	2.61	⋯	2.37	2.43	2.47	⋯
$g = 7$	Gen007	3.14	3.15	3.16	⋯	2.71	2.72	2.73	⋯
	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
	Gen$g$	2.71	2.79	2.75	⋯	3.14	3.19	3.08	⋯
	Gen$(g+1)$	3.11	3.08	3.07	⋯	3.08	3.21	3.17	⋯
	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
$g = G$	Gen$G$	0.08	0.14	0.17	⋯	0.45	0.48	0.34	⋯

Keuzeschema toetsen



1-channel experimenten (1)



$(y_g)_k$
Ziek
2.5
2.9
2.8
2.7

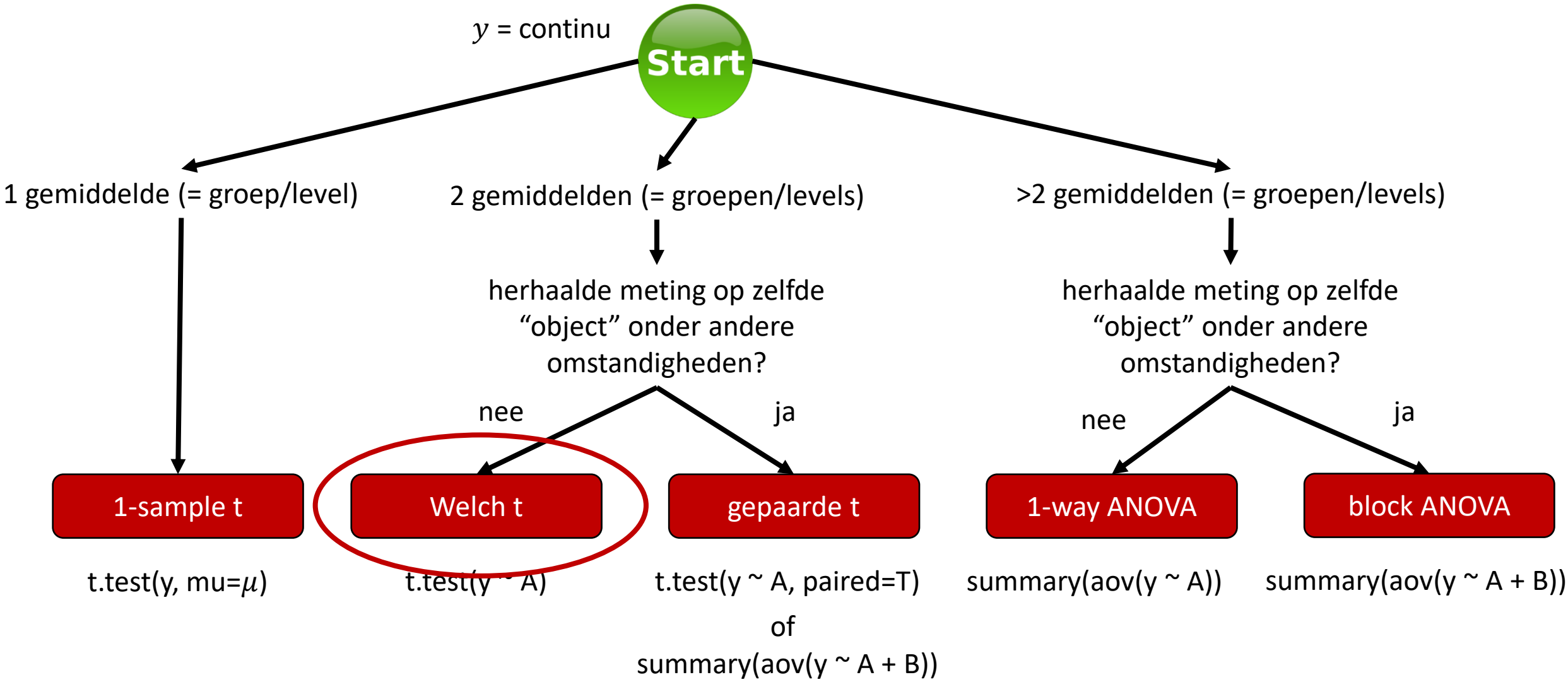
$\bar{y} = 2.0?$

4 1-channel MA's (= 4 replica's):
1-sample t

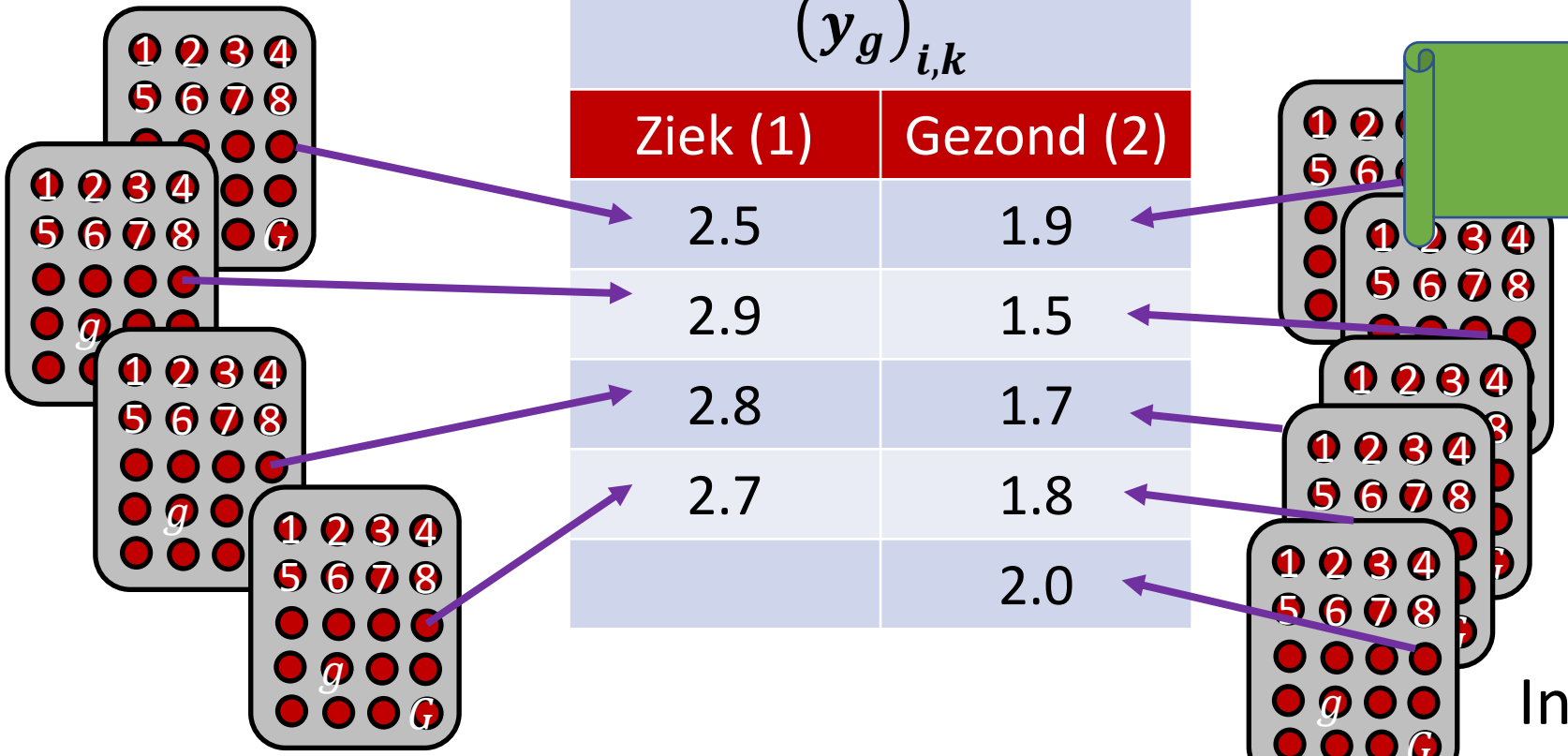
Z.1	Z.2	Z.3	Z.4
2.5	2.9	2.8	2.7

In R:
t.test(y_g, mu = μ)

Keuzeschema toetsen



1-channel experimenten (2)



$(y_g)_{i,k}$	
Ziek (1)	Gezond (2)
2.5	1.9
2.9	1.5
2.8	1.7
2.7	1.8
	2.0

$\bar{y}_1 = \bar{y}_2$?

4+5=9 1-channel MA's
(= 4+5 replica's):
Welch t

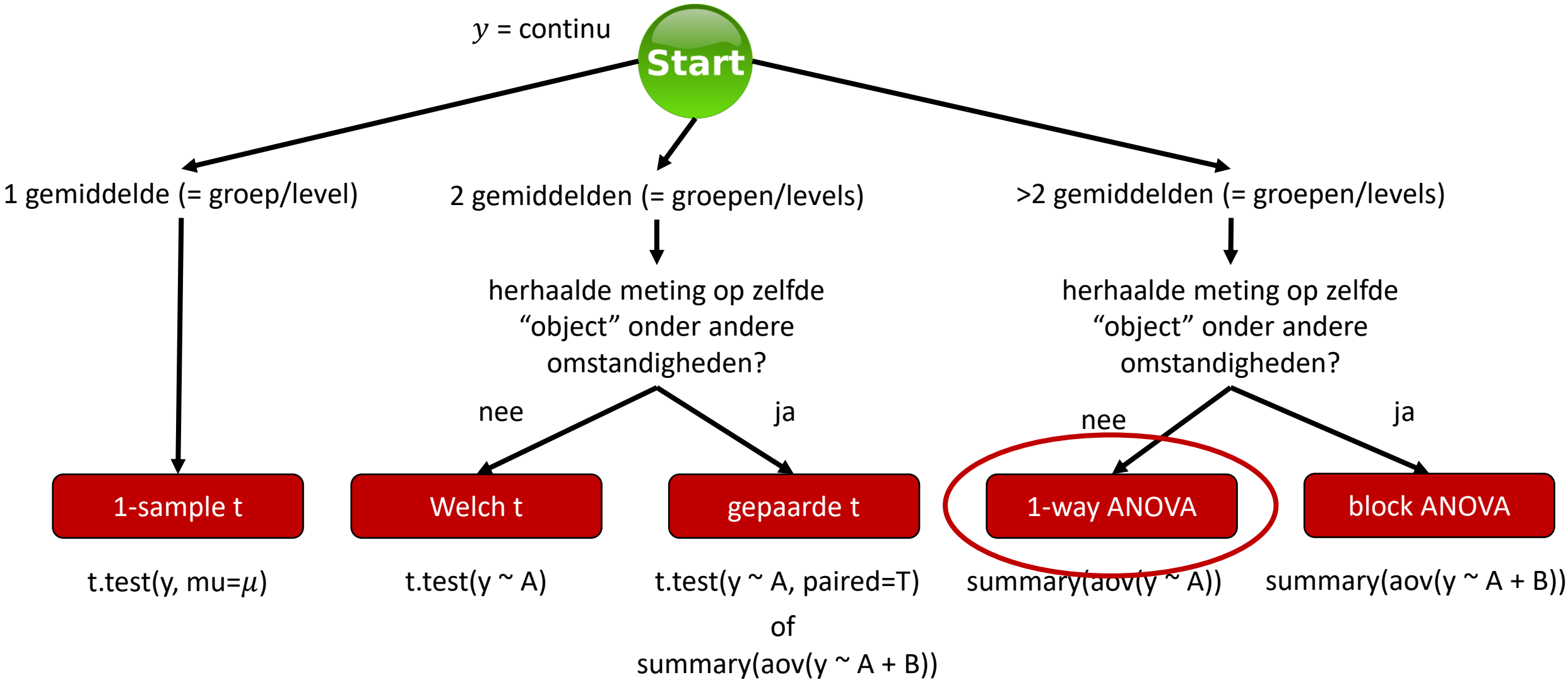
Z.1	Z.2	Z.3	Z.4	G.1	G.2	G.3	G.4	G.5
2.5	2.9	2.8	2.7	1.9	1.5	1.7	1.8	2.0

In R:

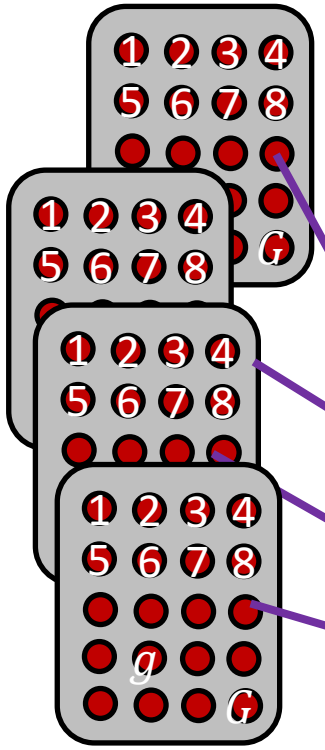
t.test(y_g ~ A)

A = [1, 1, 1, 1, 2, 2, 2, 2, 2]

Keuzeschema toetsen

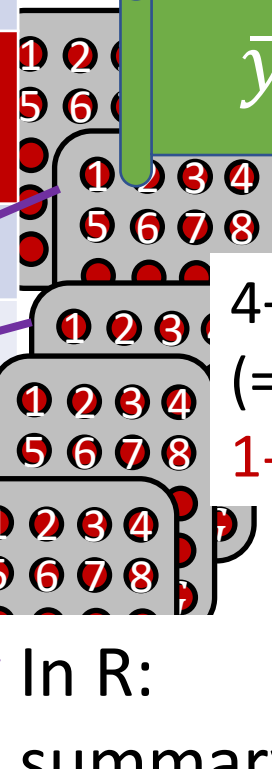


1-channel experimenten (3)



$(y_g)_{i,k}$

Stadium 1	Stadium 2	Stadium 3	Gezond (4)
2.5	2.9	4.0	1.9
2.9	3.4	4.1	1.5
2.8	3.2	3.9	1.7
2.7	3.1	3.8	1.8
	3.0		2.0



$\bar{y}_1 = \bar{y}_2 = \bar{y}_3 = \bar{y}_4$?

4+5+4+5=18 1-channel MA's
(= 4+5+4+5 replica's):
1-way ANOVA

S1.1	S1.2	S1.3	S1.4	S2.1	S2.2	S2.3	S2.4	S2.5
2.5	2.9	2.8	2.7	2.9	3.4	3.2	3.1	3.0

...

...

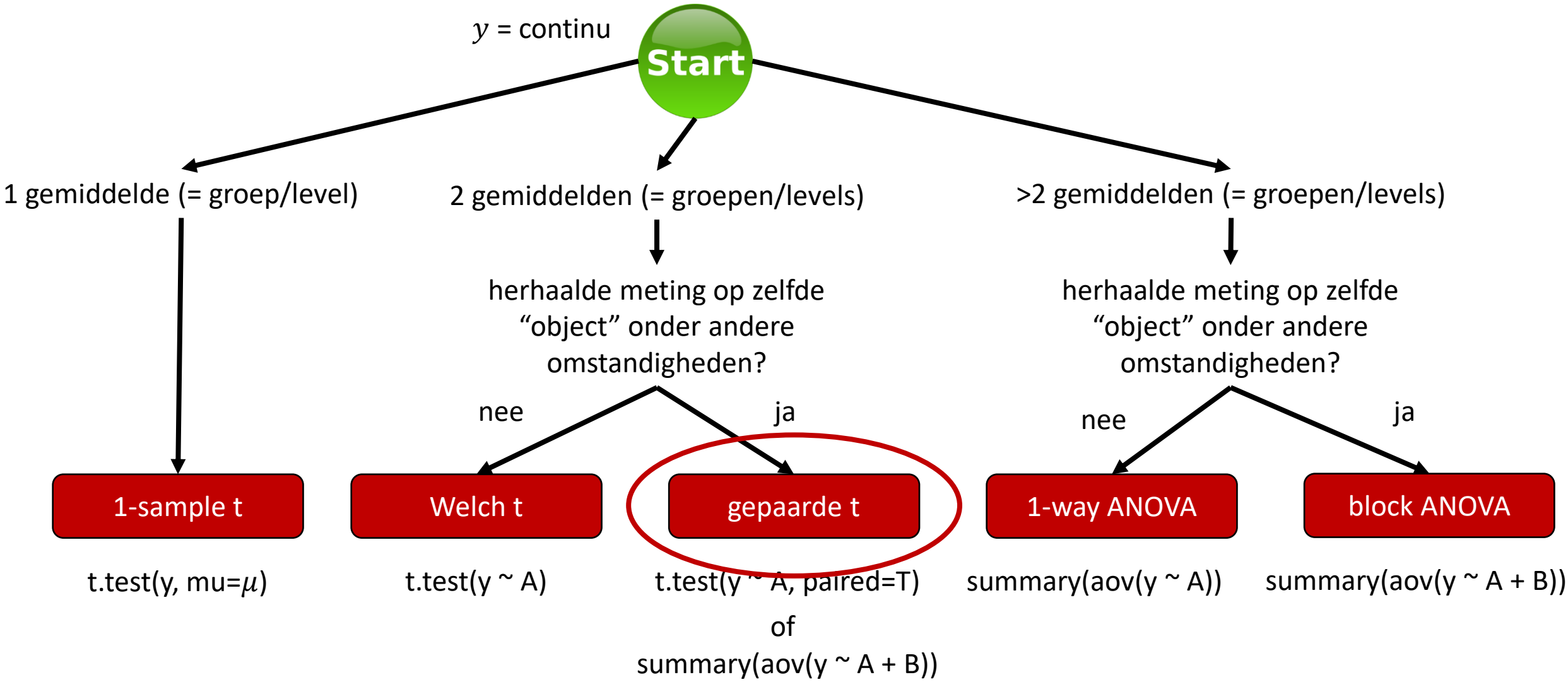
S3.1	S3.2	S3.3	S3.4	G.1	G.2	G.3	G.4	G.5
4.0	4.1	3.9	3.8	1.9	1.5	1.7	1.8	2.0

In R:

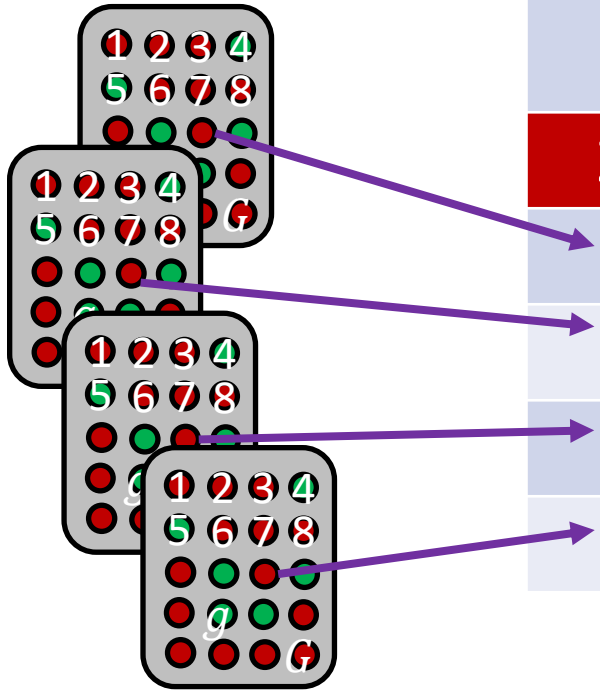
summary(aov(y_g ~ A))

A = [1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4]

Keuzeschema toetsen



2-channel experimenten (1)



$(y_g)_{i,j}$

Ziek (1)	Gezond (2)
2.5	1.5
2.9	2.0
2.8	1.8
2.7	1.7

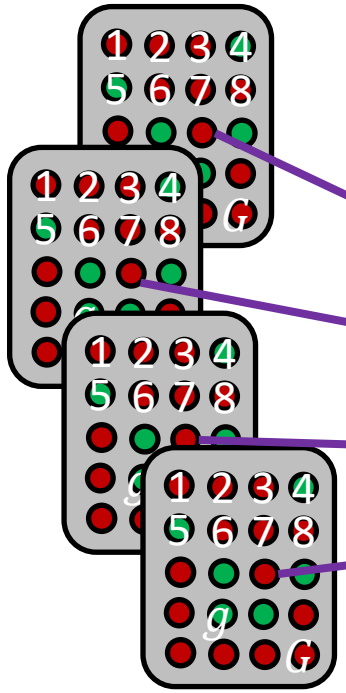
Z.1	Z.2	Z.3	Z.4	G.1	G.2	G.3	G.4
2.5	2.9	2.8	2.7	1.5	2.0	1.8	1.7

$\bar{y}_1 = \bar{y}_2$?

4 2-channel MA's
(= 4 herhaalde metingen):
gepaarde t

In R:
t.test(y_g ~ A, paired=T)
A = [1, 1, 1, 1, 2, 2, 2, 2]

2-channel experimenten (2)



$(y_g)_{i,j}$		$(M_g)_j$
Ziek (1)	Gezond (2)	M = R - G
2.5	1.5	1.0
2.9	2.0	0.9
2.8	1.8	1.0
2.7	1.7	1.0

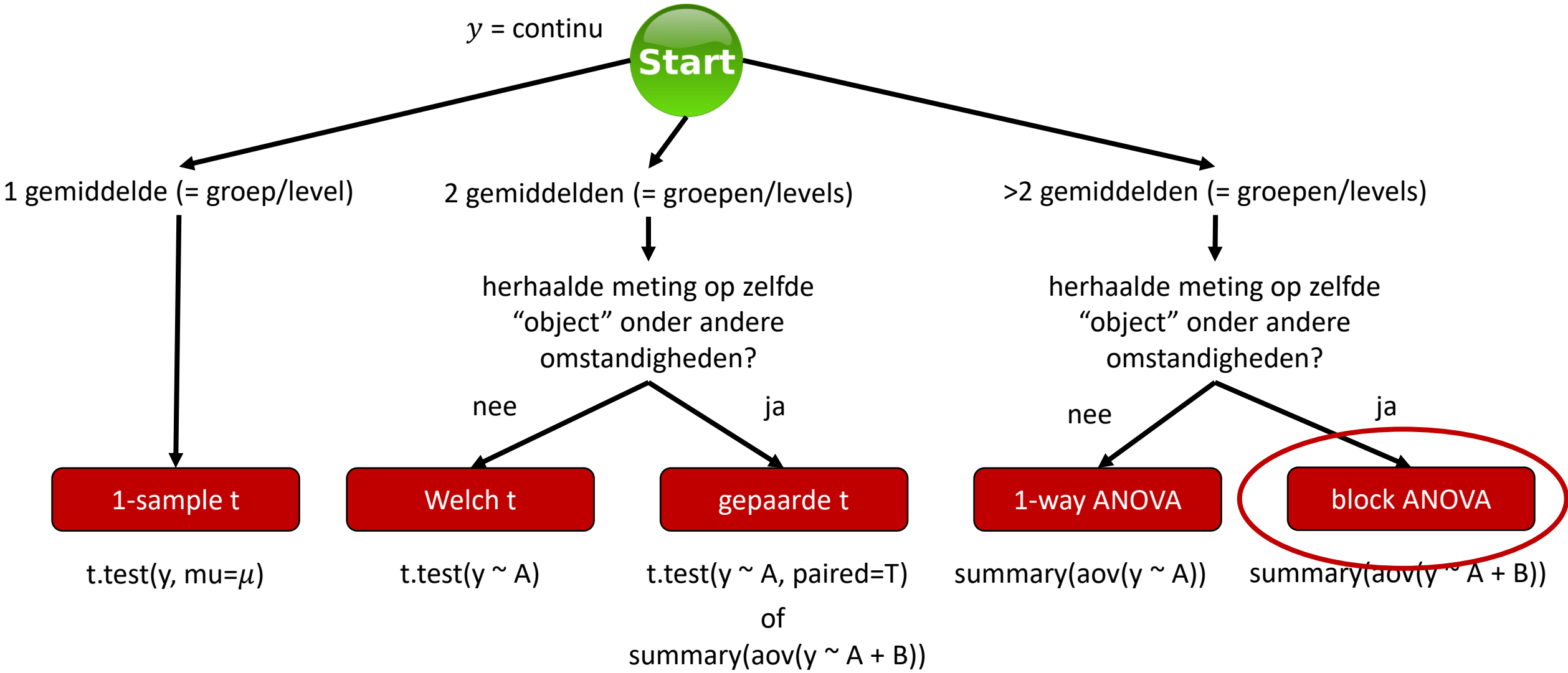
$\bar{y}_1 = \bar{y}_2$?

4 2-channel MA's
(= 4 herhaalde metingen):
1-sample t op logfold M

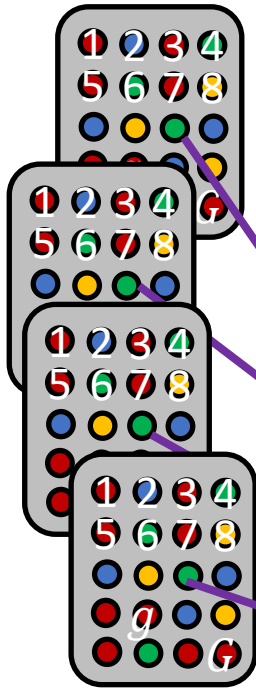
M.1	M.2	M.3	M.4
1.0	0.9	1.0	1.0

In R:
t.test(M_g, mu = 0)

Keuzeschema toetsen



4-channel experimenten (1)



$(y_g)_{i,k}$

Stadium 1	Stadium 2	Stadium 3	Gezond (4)
2.5	2.9	3.8	1.5
2.9	3.4	4.1	2.0
2.8	3.2	4.0	1.8
2.7	3.0	3.9	1.7

$$\bar{y}_1 = \bar{y}_2 = \bar{y}_3 = \bar{y}_4 ?$$

4 4-channel MA's
(= 4 herhaalde metingen):
block ANOVA

S1.1	S1.2	S1.3	S1.4	S2.1	S2.2	S2.3	S2.4
2.5	2.9	2.8	2.7	2.9	3.4	3.2	3.1

...

...

S3.1	S3.2	S3.3	S3.4	G.1	G.2	G.3	G.4
4.0	4.1	3.9	3.8	1.9	1.5	1.7	1.8

In R:

summary(aov(y_g ~ A + B))

A = [1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4]

B = [1,2,3,4,1,2,3,4,1,2,3,4,1,2,3,4]

Statistische modellen

- 1-way ANOVA (incl. 2-sample t-toets):

$$y_{ik} = \mu + \alpha_i + \varepsilon_{ik}$$

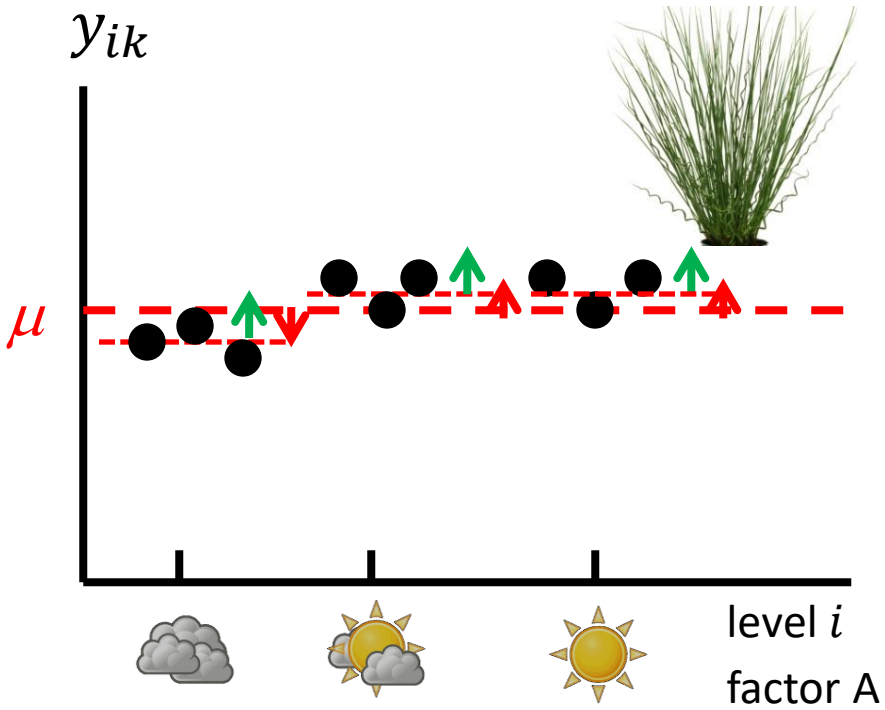
- block ANOVA (incl. gepaarde t-toets):

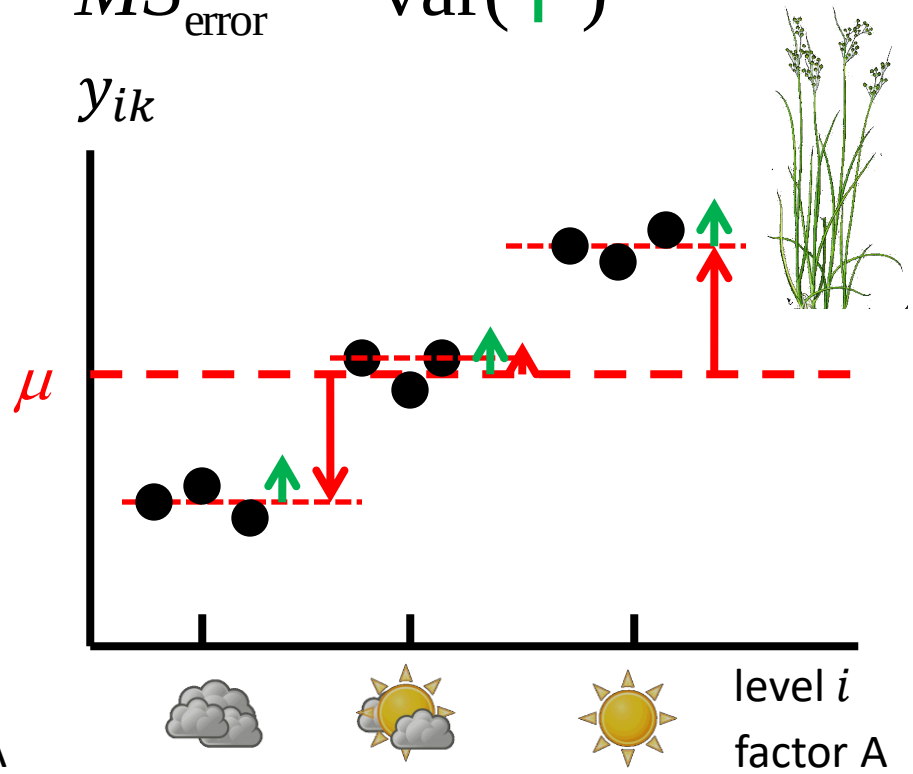
$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

1-way ANOVA: idee

Verhouding tussen spreiding (= variantie) *tussen* levels (groepen) en *binnen* levels:

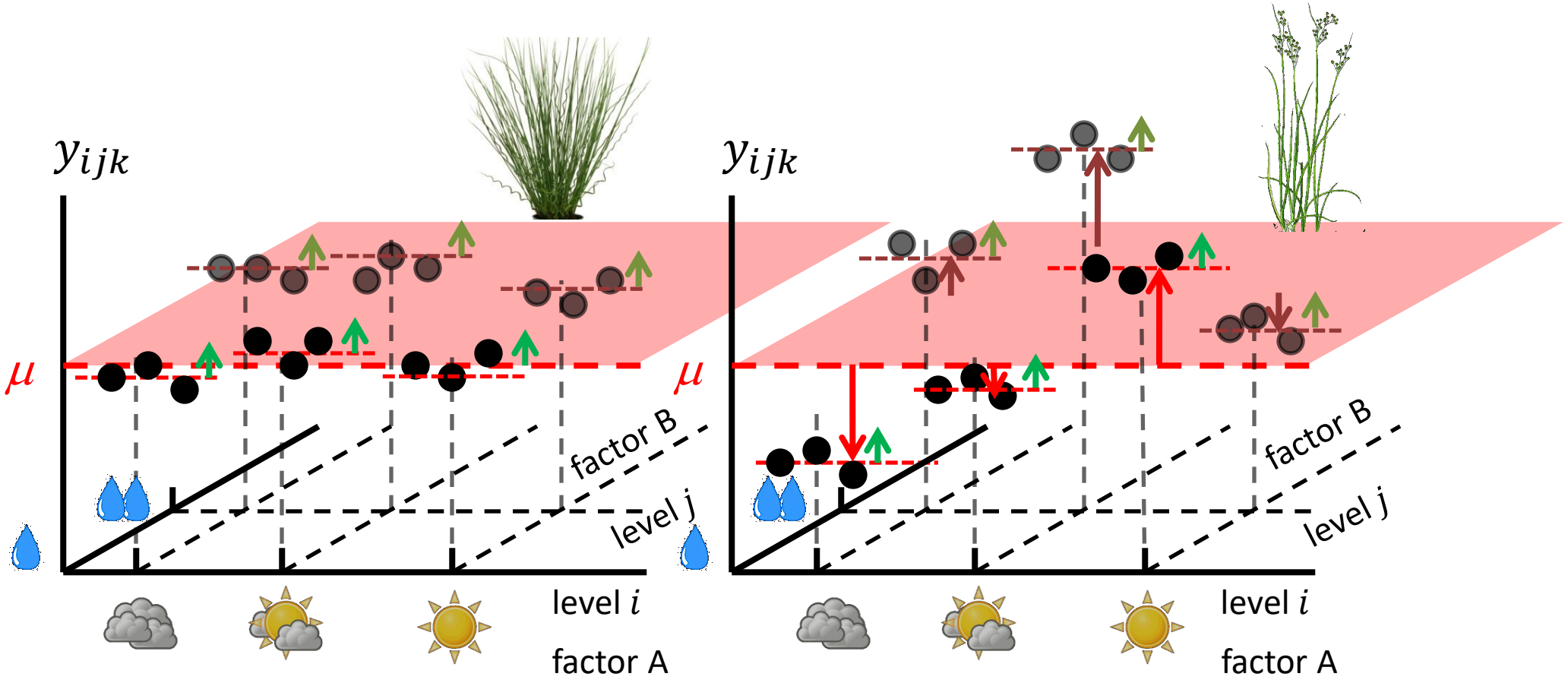
$$F = \frac{s^2_{\text{tussen}}}{s^2_{\text{binnen}}} = \frac{MS_{\text{tussen}}}{MS_{\text{binnen}}} = \frac{MS_{\text{A}}}{MS_{\text{error}}} = \frac{\text{var}(\uparrow)}{\text{var}(\uparrow)}$$





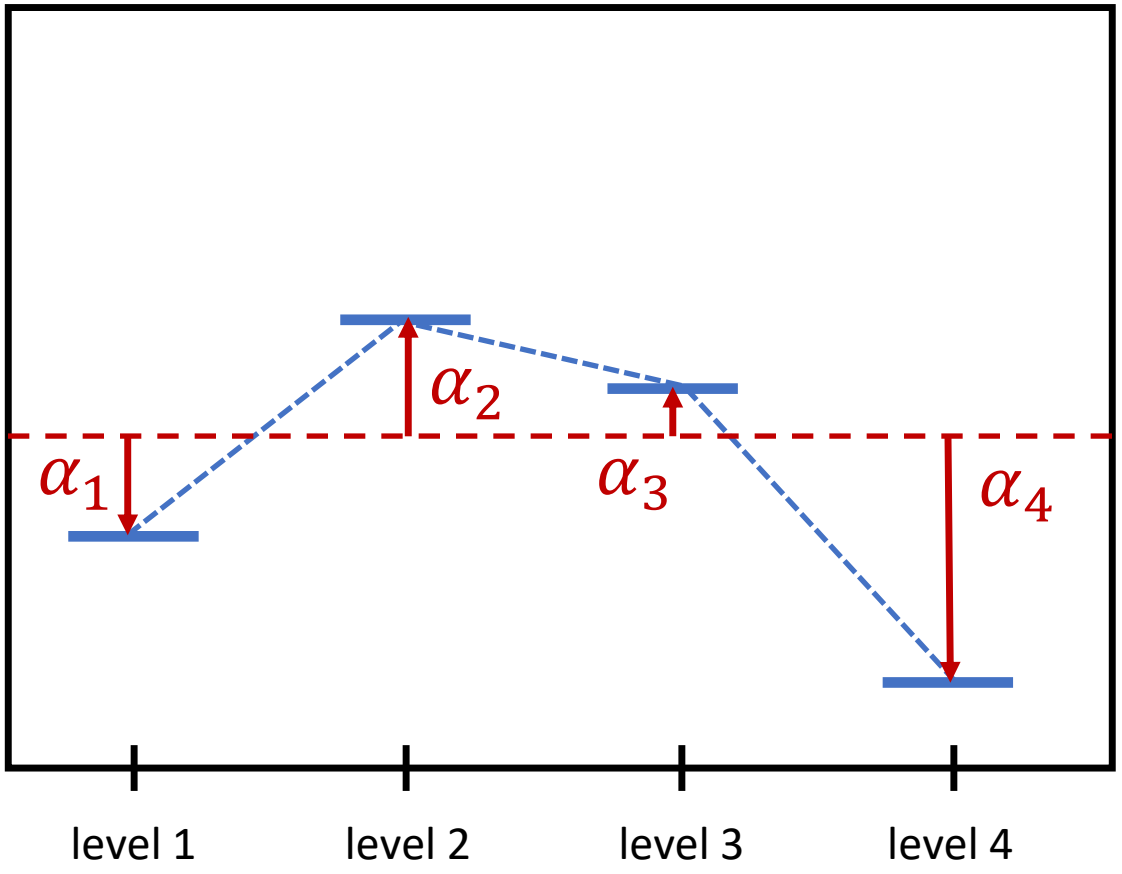
2-way ANOVA: idee

Verhouding tussen spreiding (= variantie) *tussen* levels (groep) en *binnen* levels (in 3D grafiek):



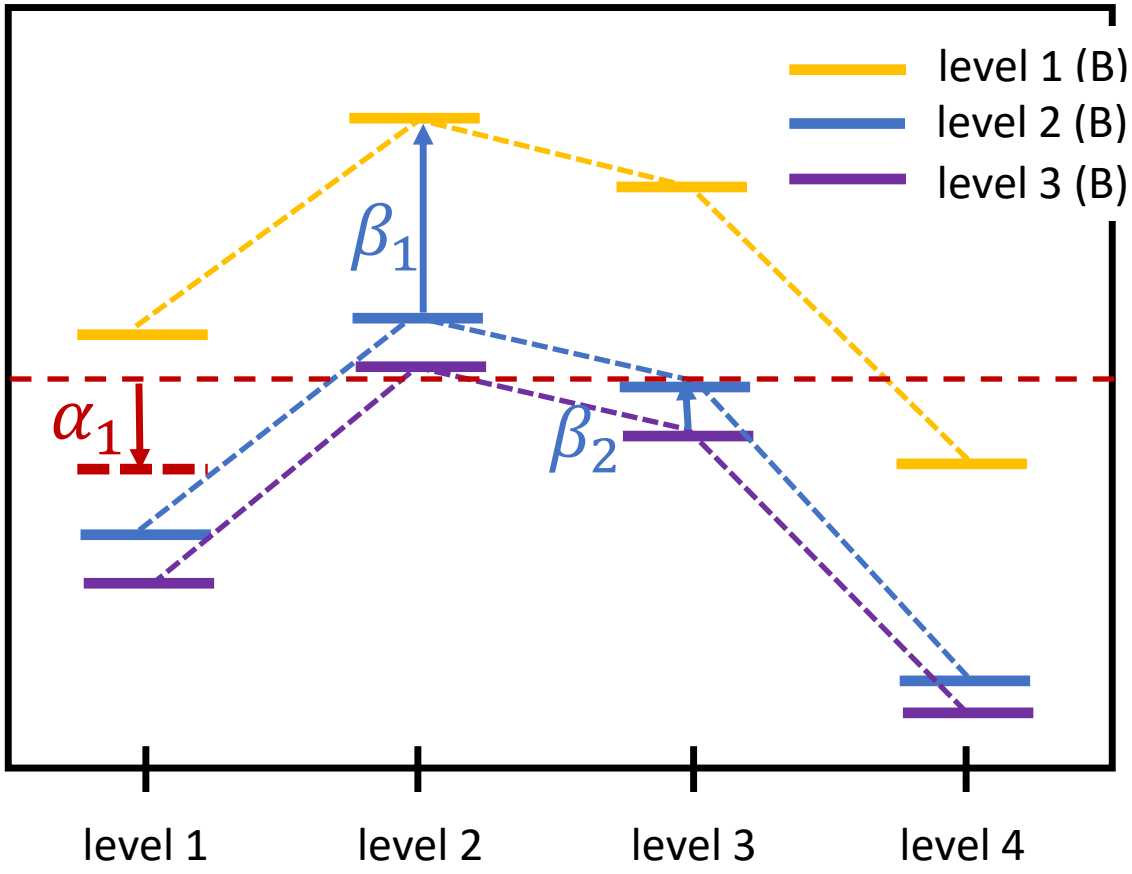
1-way en block ANOVA

$$y_{ik} = \mu + \alpha_i + \varepsilon_{ik}$$



factor A →

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$



factor A →

Overzicht

		Aantal levels (= groepen) van factor A om te vergelijken				
		1	2	3	4	> 4
Type MA	1-channel (R)	1-sample t	Welch t	1-way ANOVA		
		$(y_g)_k$	$(y_g)_{i,k} = \mu + \alpha_i + \varepsilon_{i,k}$			
		p: $\bar{y} = \mu$?	p: $\bar{y}_1 = \bar{y}_2$?	p_A: verschil tussen groepen?		
	2-channel (R, G)		gepaarde t			
			p: $\bar{y}_1 = \bar{y}_2$?			
	4-channel (R, G, B, Y)		gepaarde t	block ANOVA		
			$(y_g)_{i,j} = \mu + \alpha_i + \beta_j$			
			p of p_A: verschil tussen groepen?			