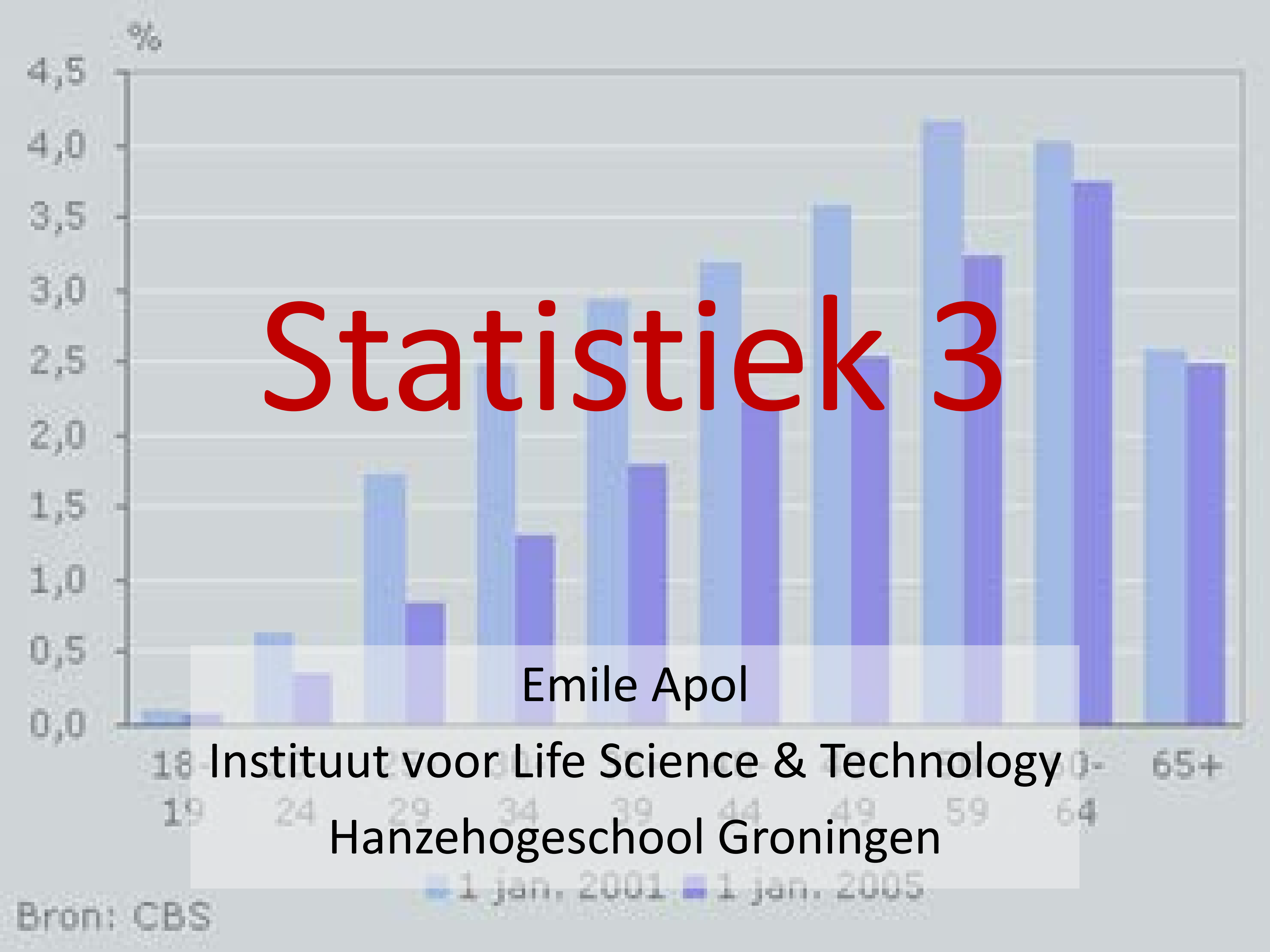
Statistiek 3

Emile Apol
Instituut voor Life Science & Technology
Hanzehogeschool Groningen

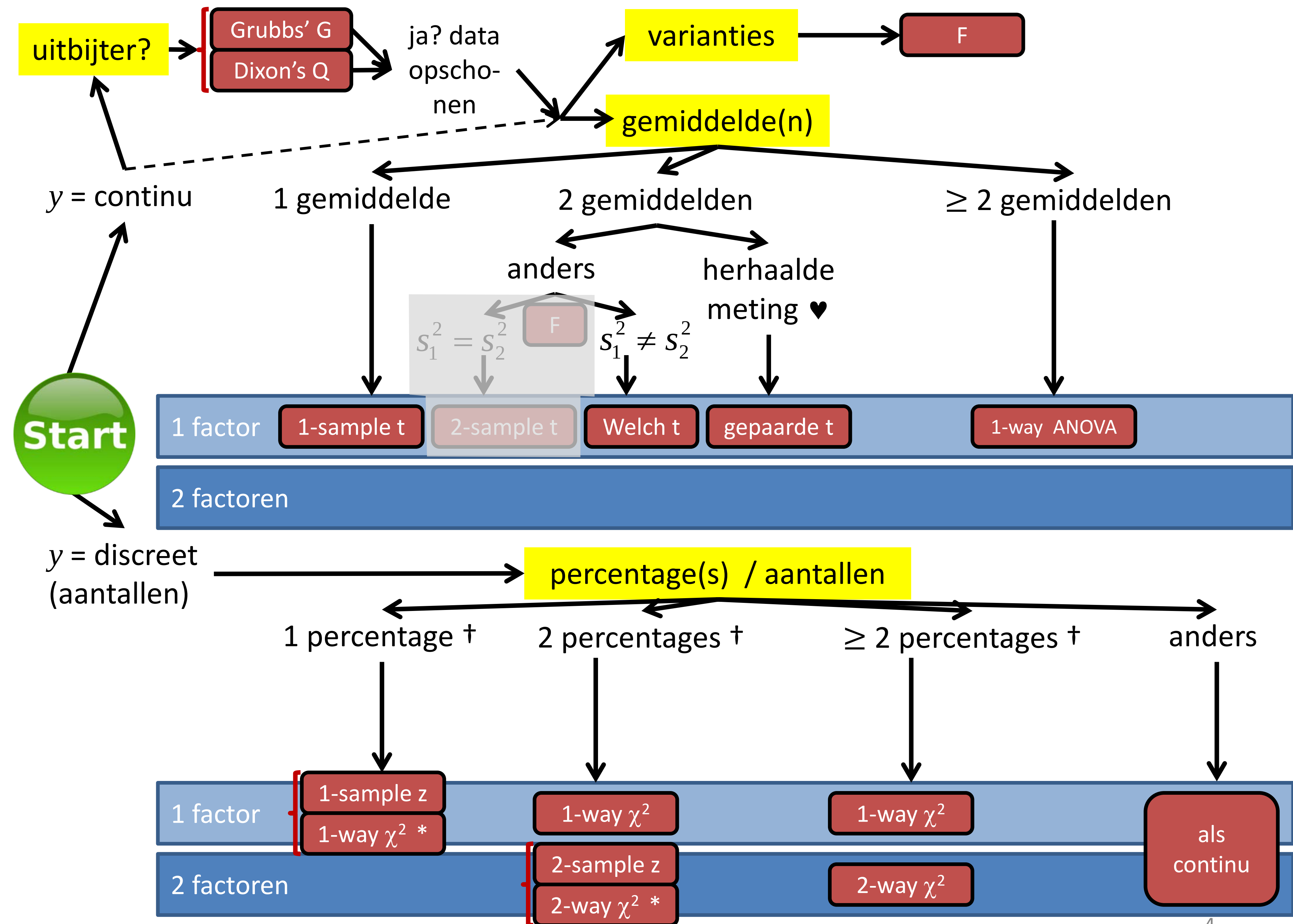


Verschiltoetsen in R

- 1-sample *t*-toets
  - Effectsterkte: Cohen's d_s
- Welch (2-sample) *t*-toets
  - Effectsterkte: Cohen's d_{av}
- Gepaarde *t*-toets
  - Effectsterkte: Cohen's d_{av}
- 1-way ANOVA
  - Effectsterkte: Eta-kwadraat: η^2

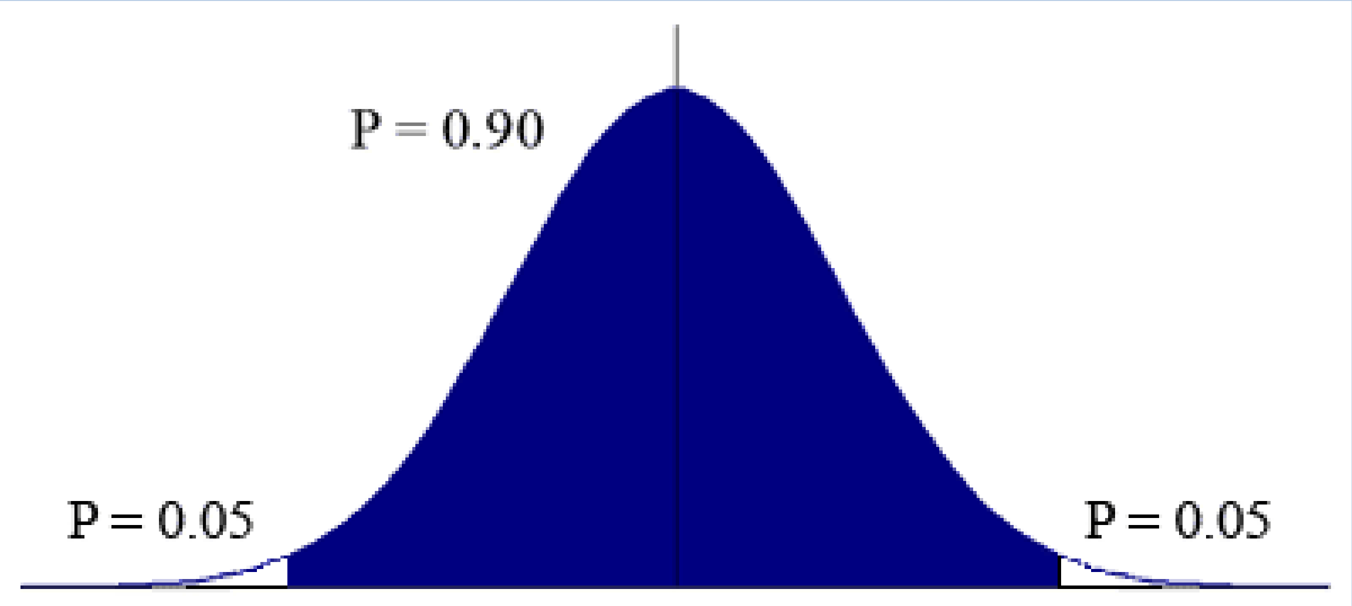
Stappenplan

1. Formuleer de vraag helder
2. Kies op basis van de soort data de juiste toets
3. Formuleer nul-hypothese H_0 ("alles is gelijk")
4. Formuleer op basis van je vraag (achtergrond informatie) de alternatieve hypothese H_1
   - 1-zijdig toetsen
   - 2-zijdig toetsen
5. Voer de toets uit
6. Formuleer de conclusie in woorden

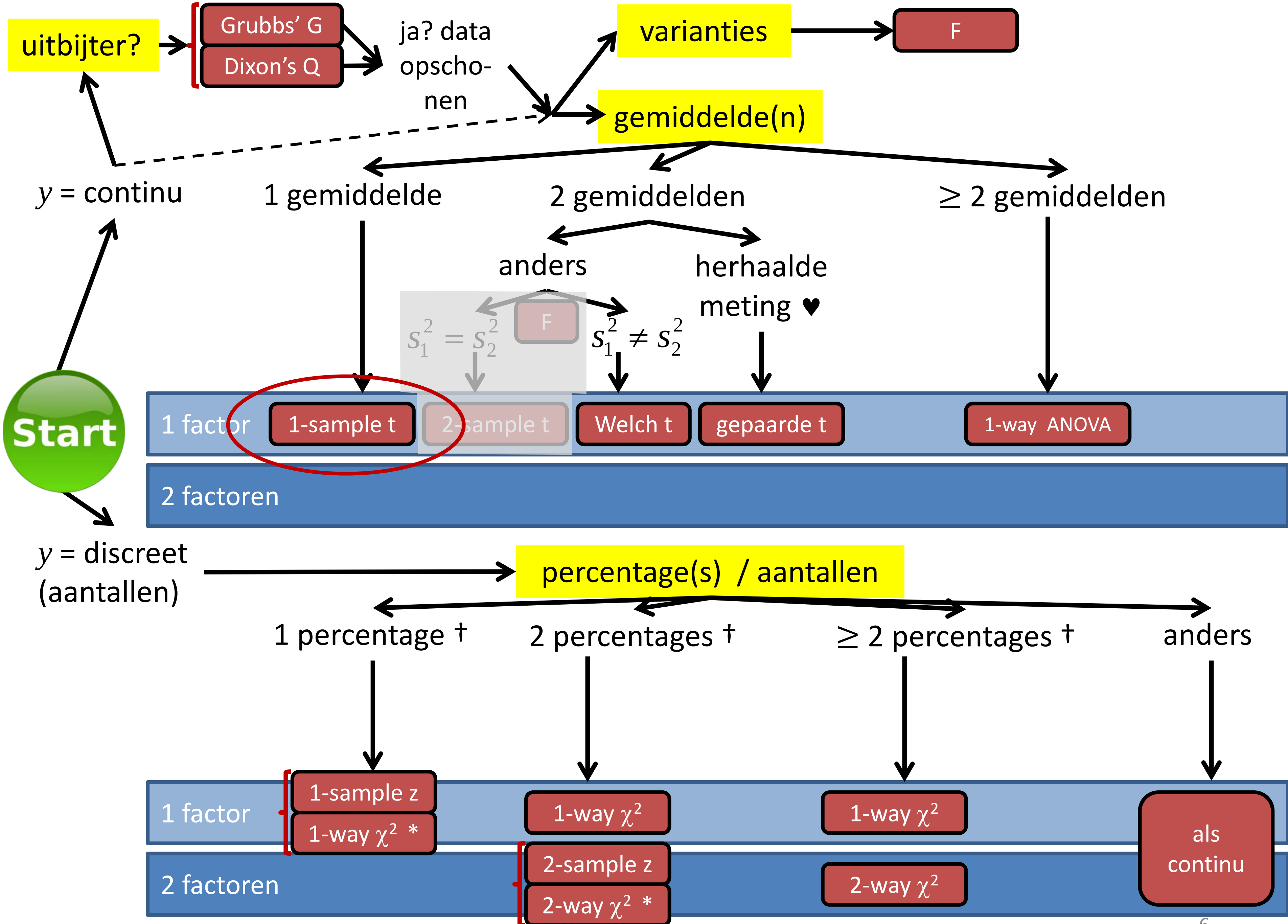


* alleen voor 2-zijdige vraag over percentages ♥ herhaalde meting op *zelfde* "object" onder *andere* omstandigheden † onafhankelijk

1-sample *t*-toets



Verschilt één experimenteel gemiddelde van een referentiewaarde?



* alleen voor 2-zijdige vraag over percentages ♥ herhaalde meting op *zelfde* "object" onder *andere* omstandigheden † onafhankelijk

1-sample t-toets

De formele t-toets:

$$t = \frac{\text{verschil tussen gemiddelden}}{\text{fout in verschil}}$$

$$\text{df} = n - 1$$

wordt

$$t = \frac{(\bar{y} - \mu)}{s_{\bar{y}}} = \frac{\bar{y} - \mu}{s/\sqrt{n}}$$

- Er is een significant *verschil* als:
  $|\, t \,| > t_{\text{krit.2}}$
- Het gemiddelde is *hoger* dan verwacht als:
  $t > t_{\text{krit.1}}$
- Het gemiddelde is *lager* dan verwacht als:
  $t < -t_{\text{krit.1}}$

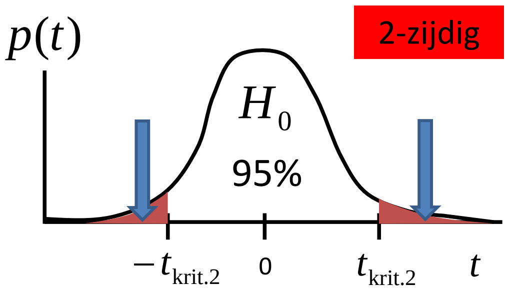
Grenswaarde (met $\alpha = 0.05$):

2-zijdig: $t_{\text{krit.2}} = \text{qt}\,(0.975, \text{df})$

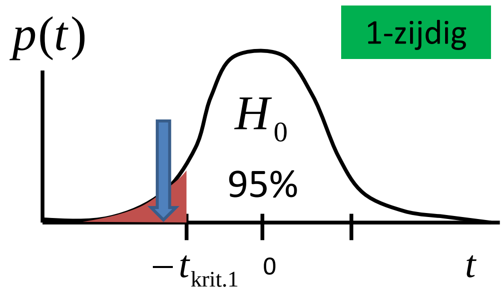
1-zijdig: $t_{\text{krit.1}} = \text{qt}\,(0.95, \text{df})$

Significant of niet?

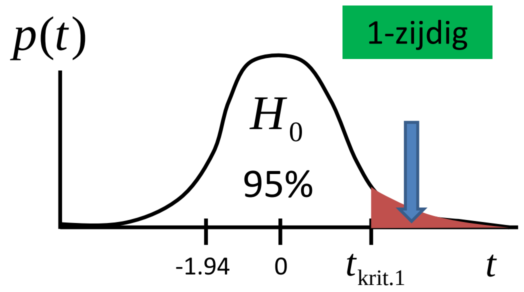
Een verschil is significant als



$$\left| t \right| > t_{\text{krit.2}}$$



$$t < -t_{\text{krit.1}}$$



$$t > t_{\text{krit.1}}$$

of (wat equivalent is)

$$p < \alpha$$

Geeft R

“Kans dat H_0 waar is”

Statistische en praktische significantie

- Statistische significantie:

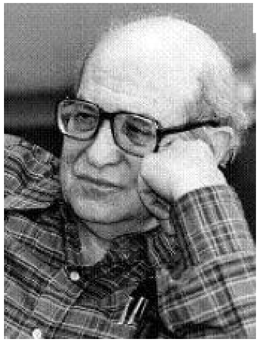
Is het verschil tussen gemeten gemiddelde en de referentiewaarde overtuigend? D.w.z. is het verschil (veel) groter dan de fout in dit verschil?

Significant als: $\left| t = \frac{\bar{y} - \mu}{s_{\bar{y}}} \right| > t_{\text{krit.2}}$

- Praktische significantie:

Is het verschil tussen gemeten gemiddelde en de referentiewaarde belangrijk? D.w.z. is het verschil (veel) groter dan de ruis is de data? Praktische significantie = effectsterkte: Cohen's d: $d_s = \frac{\bar{y} - \mu}{s}$

Effectsterkte: Cohen's d



Jacob Cohen

De effectsterkte bij een t-toets wordt meestal uitgedrukt in de waarde van Cohen's d:

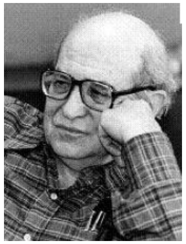
$$d_s = \frac{\bar{y} - \mu}{s}$$

- Cohen's d meet dus het verschil $\bar{y} - \mu$ in "eenheden" van de ruis = standaarddeviatie s
- Cohen's d is dimensieloos, $\pm$ onafhankelijk van n!
- Wat is "groot" verschil, wat is "klein" verschil? Hangt van de context af, maar algemene richtlijnen*:

Waarde Cohen's d	Interpretatie
0.2	Klein verschil
0.5	Matig verschil
0.8	Groot verschil

*) Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Effectsterkte: Cohen's d



Cohen's d: $d_s = \frac{\bar{y}-\mu}{s}$

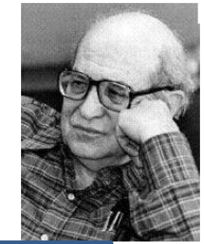
Berekening:

$$d_s = \frac{(\bar{y}-\mu)}{s} = \frac{t}{\sqrt{n}}$$

want

$$t = \frac{(\bar{y}-\mu)}{s/\sqrt{n}} = \frac{(\bar{y}-\mu)}{s} \cdot \sqrt{n} = d_s \cdot \sqrt{n}$$

Effectsterkte: Cohen's d



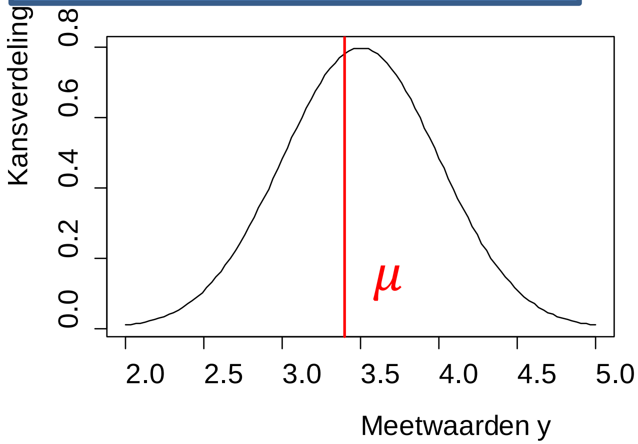
Cohen's d: $d_s = \frac{\bar{y} - \mu}{s}$

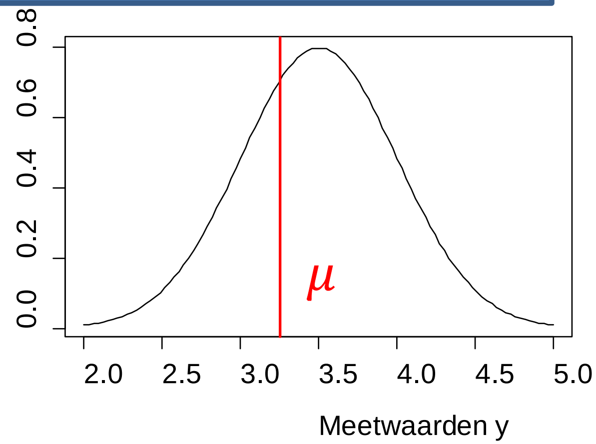
Voorbeeld: $\bar{y} = 3.50$, $s = 0.50$

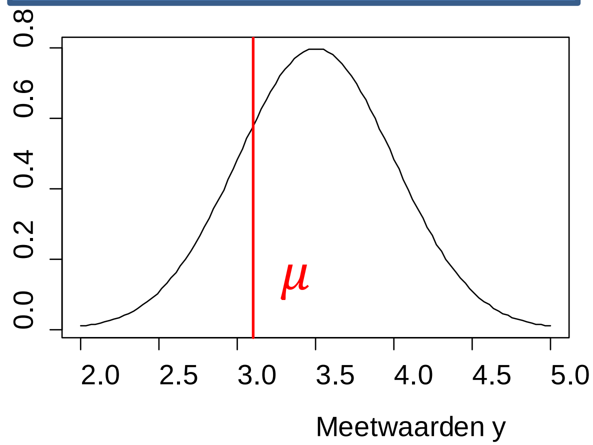
$d_s = 0.20$: klein verschil

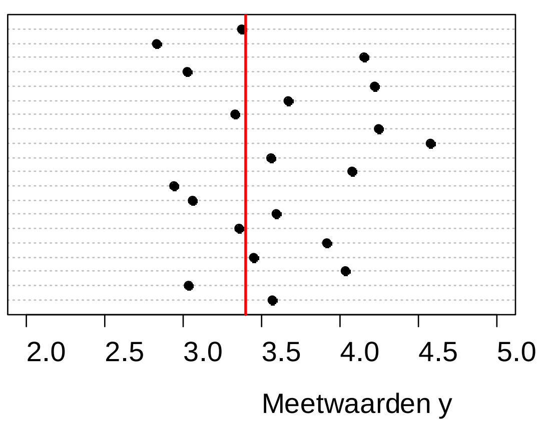
$d_s = 0.50$: matig verschil

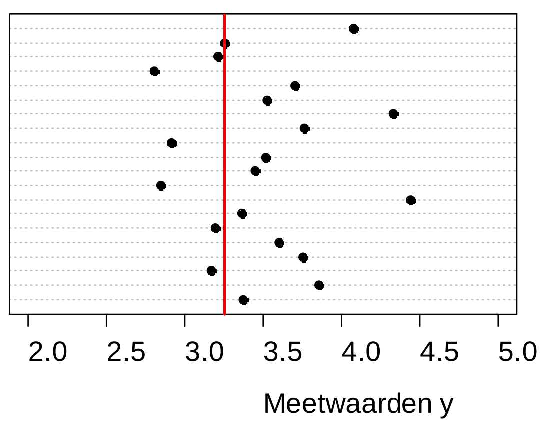
$d_s = 0.80$: groot verschil

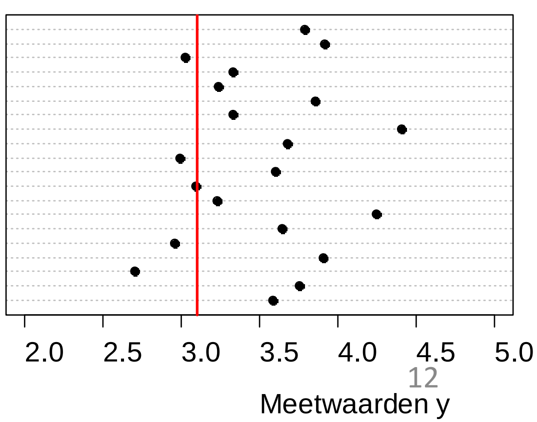




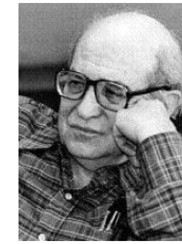






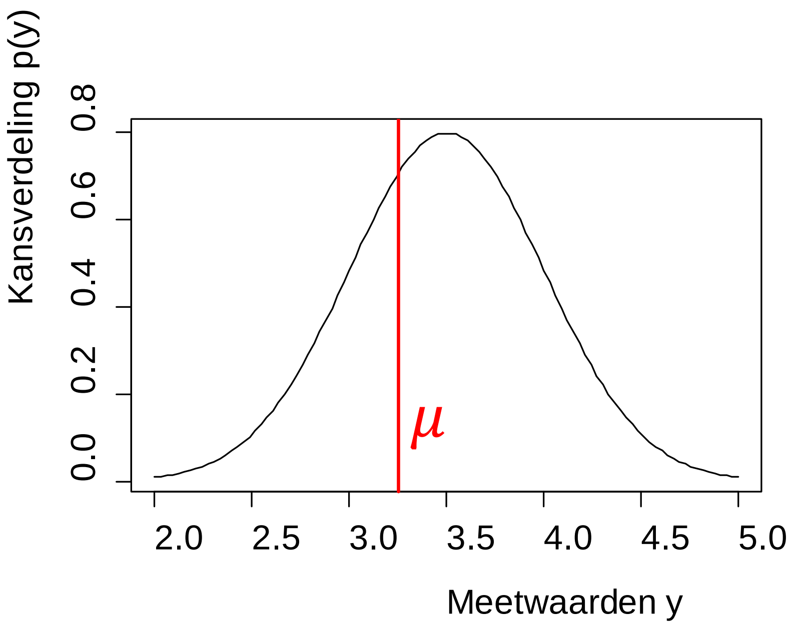


Effectsterkte: Cohen's d

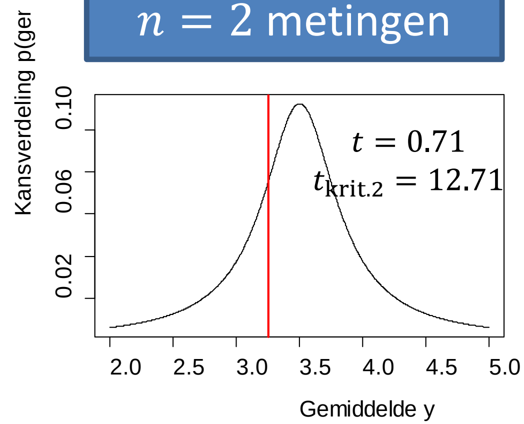


Cohen's d: $d_s = \frac{\bar{y} - \mu}{s}$

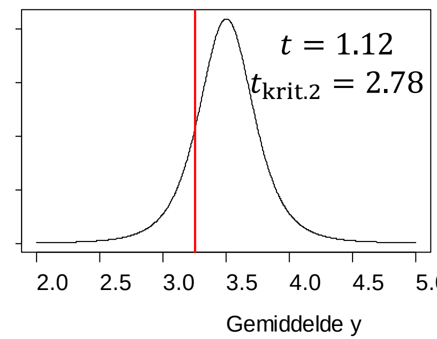
Voorbeeld: $\bar{y} = 3.50$, $s = 0.50$, $\mu = 3.25$



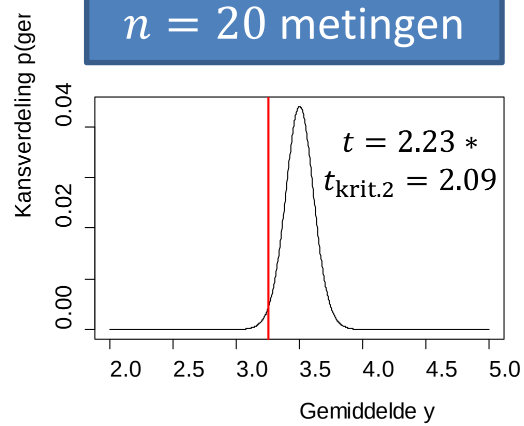
$n = 2$ metingen



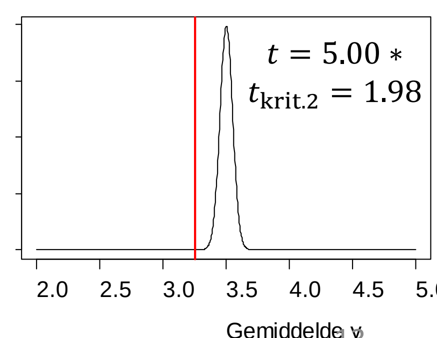
$n = 5$ metingen



$n = 20$ metingen



$n = 100$ metingen



Effectsterkte: $d_s = \frac{3.50 - 3.25}{0.50} = +0.50$,
dus een matig verschil!

Aantal metingen n en effectsterkte

- Om een "klein" verschil significant te noemen heb je veel metingen nodig…
- Om een "groot" verschil significant te noemen heb je minder metingen nodig…
- We hebben $t = d_s \cdot \sqrt{n}$, dus $t_{\text{krit.2}} = d_s \cdot \sqrt{n}$, dus

$$\text{qt}(0.975; n-1) = d_s * \text{sqrt}(n)$$

Waarde Cohen's d_s	Interpretatie	Minimum aantal metingen n voor significant verschil
0.2	Klein verschil	99
0.5	Matig verschil	18
0.8	Groot verschil	9

1-sample *t*-toets in R

- 1 vector y en waarde mu:

  ```
  t.test(y, mu = mu, ...)
  ```

- 1 kolom y uit dataframe M:

  ```
  t.test(M$y, mu = mu, ...)
  t.test(y, mu = mu, data = M, ...)
  ```

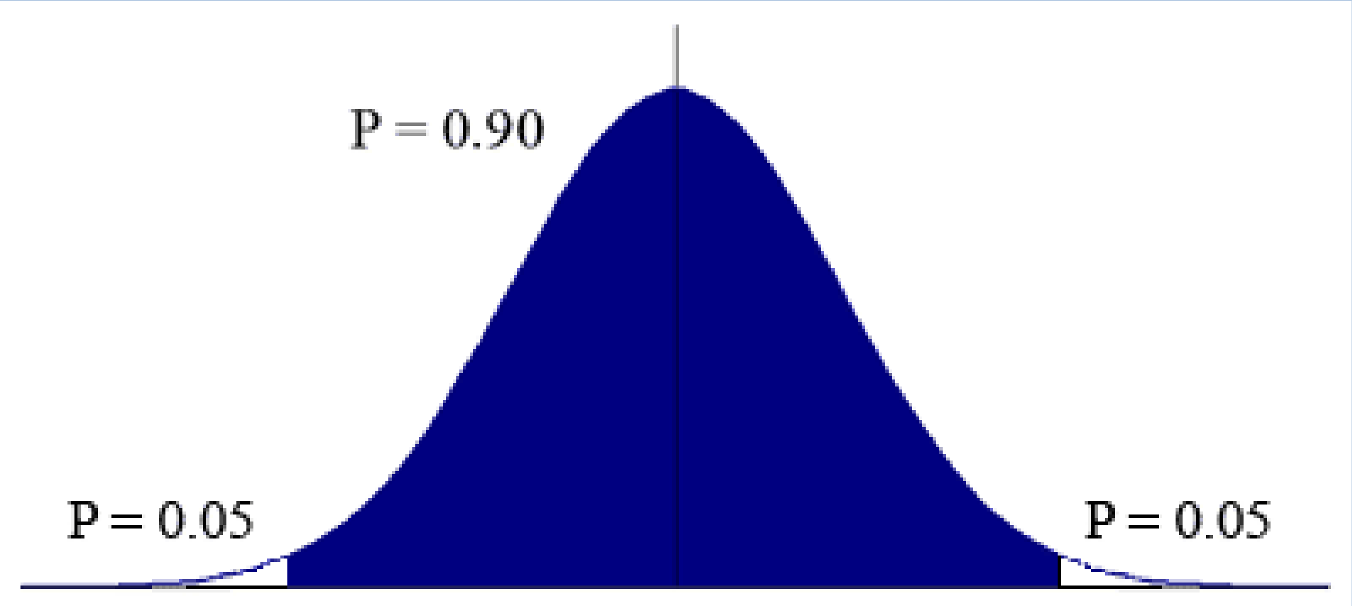
- 1- en 2-zijdig toetsen:
  - 2-zijdig: alternative="two.sided" (default)
  - 1-zijdig: alternative="less" or alternative="greater"

1-sample t-toets: samenvatting

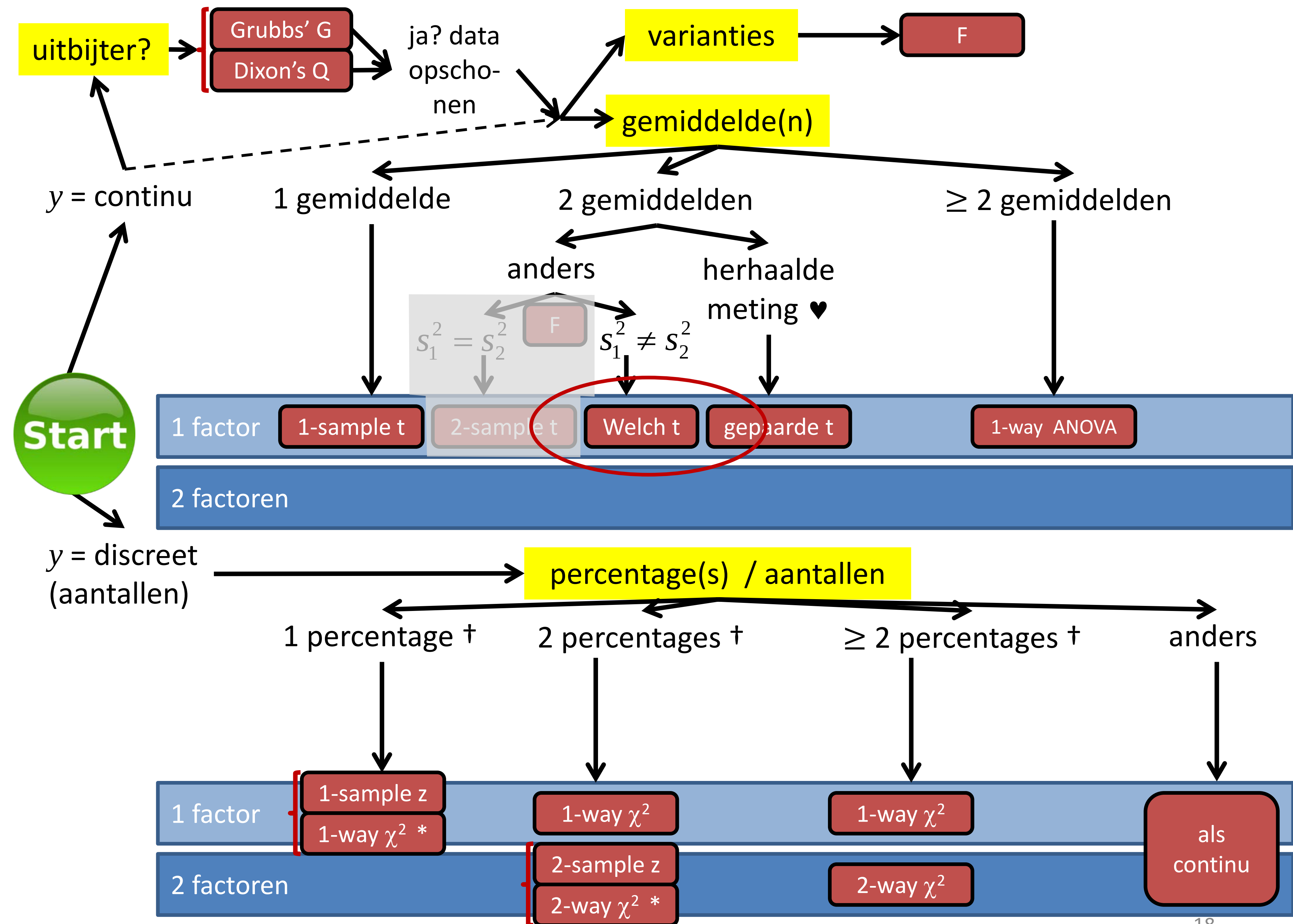
- Meetwaarden y continu, $a = 1$ groep
- $\mathrm{H_0}$: $\bar{y} = \mu$
- $\mathrm{H_1}$: $\bar{y} \neq \mu$ of $\bar{y} > \mu$ of $\bar{y} < \mu$
- $t = \frac{(\bar{y}-\mu)}{s_{\bar{y}}}$ $\mathrm{df} = n - 1$
- $t_{\text{krit.2}} = \mathrm{qt}\,(0.975, \mathrm{df})$ $t_{\text{krit.1}} = \mathrm{qt}\,(0.95, \mathrm{df})$
- $\mathrm{H_1}$ als: $p < \alpha$ of
  $|\,t\,| > t_{\text{krit.2}}$ of $t > t_{\text{krit.1}}$ of $t < -t_{\text{krit.1}}$
- Effectsterkte: $d_s = \frac{(\bar{y}-\mu)}{s} = \frac{t}{\sqrt{n}}$

Waarde Cohen's d: $\mid d_s \mid$	Interpretatie
0.2	Klein verschil
0.5	Matig verschil
0.8	Groot verschil

Welch (2-sample) *t*-toets



Verschillen twee experimentele gemiddelden van elkaar?



* alleen voor 2-zijdige vraag over percentages ♥ herhaalde meting op *zelfde* "object" onder *andere* omstandigheden † onafhankelijk

Ter info!

Afleiding t-toets 2 samples (1)

De t-waarde voor twee samples is het verschil in 2 gemiddelden gedeeld door de fout in dit verschil:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_{\bar{y}_1 - \bar{y}_2}}$$

Statistiek 1: bij de fout in een verschil $\bar{y}_1 - \bar{y}_2$ tellen de fouten in beide gemiddelden $\bar{y}_1$ en $\bar{y}_2$ *kwadratisch* op:

$$s_{\bar{y}_1 - \bar{y}_2}^2 = s_{\bar{y}_1}^2 + s_{\bar{y}_2}^2 = \left(\frac{s_1}{\sqrt{n_1}}\right)^2 + \left(\frac{s_2}{\sqrt{n_2}}\right)^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

dus

$$s_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

en

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Dit is de formule voor de Welch t-toets

Als beide varianties (standaarddeviaties) ongeveer gelijk zijn kunnen we ook de "gemiddelde" of "gepoolde" variantie s_p^2 gebruiken, waarbij

$$s_p^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 - 1 + n_2 - 1}$$

dus

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_p^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Dit is de formule voor de 2-sample t-toets

Ter info!

Afleiding t-toets 2 samples (2)

Voor de t-waarde voor de 2-sample t-toets, waarbij de varianties gelijk zijn ($s_1^2 = s_2^2 = s_p^2$):

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

is de kansverdeling *exact bekend*, namelijk een t-verdeling met

$$\mathrm{df} = n_1 + n_2 - 2$$

vrijheidsgraden. We kunnen hiervoor dus m.b.v. qt $(0.975, n_1 + n_2 - 2)$ exacte kritische grenzen $t_{\mathrm{krit.2}}$ uitrekenen, en dezelfde logica aanhouden als bij de 1-sample t-toets! Echter, we zouden eerst moeten toetsen of beide varianties s_1^2 en s_2^2 inderdaad wel ongeveer gelijk zijn, anders zijn de kritische grenzen namelijk niet goed… Dit toetsen kan via een zogenaamde *F*-toets (zie dictaat op BB en het statistiekboek Miller & Miller). Statistici waarschuwen echter dat dit niet de beste manier is!

Het is beter om *altijd* de Welch t-toets te gebruiken:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

waarbij je dus niet per se aanneemt dat beide varianties s_1^2 en s_2^2 gelijk zijn. De kansverdeling van deze t-waarde is niet exact bekend, maar is ongeveer een t-verdeling waarbij het aantal vrijheidsgraden df (dat je nodig hebt voor het berekenen van de kritische waarden) gegeven wordt door*

$$\mathrm{df} \approx \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 \Bigg/ \left(\frac{s_1^4}{n_1^2 \cdot (n_1 - 1)} + \frac{s_2^4}{n_2^2 \cdot (n_2 - 1)}\right)$$

De kritische waarden zijn nu te berekenen via qt $(0.975, \mathrm{df})$, en de logica is verder hetzelfde als bij de 1-sample t-toets.

* Welch, B. L. (1947) "The generalization of "Student's" problem when several different population variances are involved" *Biometrika* **34**, 28–35.

2-sample t-toets

Liever NIET gebruiken!

De formele t-toets:

$$t = \frac{\text{verschil tussen gemiddelden}}{\text{fout in verschil}}$$

$$\mathrm{df} = n_1 + n_2 - 2$$

$$s_p^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 - 1 + n_2 - 1}$$

wordt

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- Er is een significant *verschil* tussen $\bar{y}_1$ en $\bar{y}_2$ als:
  $$|\,t\,| > t_{\mathrm{krit.2}}$$
- Gemiddelde $\bar{y}_1$ is *hoger* dan $\bar{y}_2$ als:
  $$t > t_{\mathrm{krit.1}}$$
- Gemiddelde $\bar{y}_1$ is *lager* dan $\bar{y}_2$ als:
  $$t < -t_{\mathrm{krit.1}}$$

Grenswaarde (voor $\alpha = 0.05$):

2-zijdig: $t_{\mathrm{krit.2}} = \mathrm{qt}(0.975, \mathrm{df})$

1-zijdig: $t_{\mathrm{krit.1}} = \mathrm{qt}(0.95, \mathrm{df})$

Welch t-toets

De formele t-toets:

$$t = \boxed{\frac{\text{verschil tussen gemiddelden}}{\text{fout in verschil}}}$$

$$\text{df} \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^4}{n_1^2 \cdot (n_1 - 1)} + \frac{s_2^4}{n_2^2 \cdot (n_2 - 1)}\right)}$$

wordt

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Grenswaarde (voor $\alpha = 0.05$):

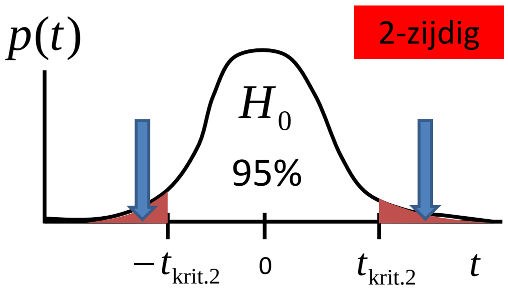
2-zijdig: $t_{\text{krit.2}} = \text{qt}\,(0.975, \text{df})$

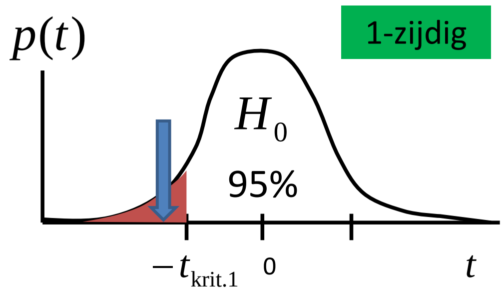
1-zijdig: $t_{\text{krit.1}} = \text{qt}(0.95, \text{df})$

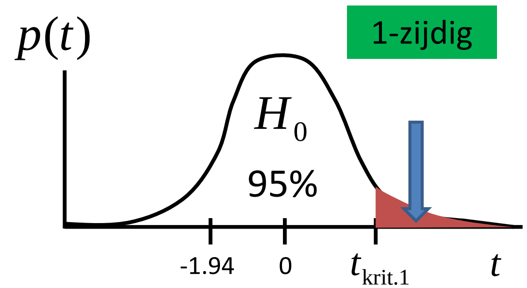
- Er is een significant *verschil* tussen $\bar{y}_1$ en $\bar{y}_2$ als:
  $$|\, t \,| > t_{\text{krit.2}}$$
- Gemiddelde $\bar{y}_1$ is *hoger* dan $\bar{y}_2$ als:
  $$t > t_{\text{krit.1}}$$
- Gemiddelde $\bar{y}_1$ is *lager* dan $\bar{y}_2$ als:
  $$t < -t_{\text{krit.1}}$$

Significant of niet?

Een verschil is significant als







$$|t| > t_{\text{krit.2}} \qquad t < -t_{\text{krit.1}} \qquad t > t_{\text{krit.1}}$$

of (wat equivalent is)

$$p < \alpha$$

Geeft R

"Kans dat H_0 waar is"

Statistische en praktische significantie

- Statistische significantie:

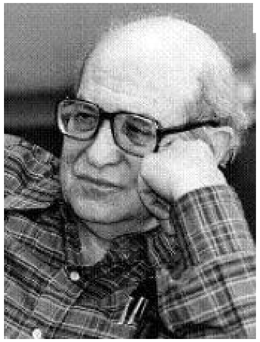
Is het verschil tussen gemeten gemiddelde en de referentiewaarde overtuigend? D.w.z. is het verschil (veel) groter dan de fout in dit verschil?

Significant als: $\left| t = \frac{\bar{y}_1 - \bar{y}_2}{s_{\bar{y}_1 - \bar{y}_2}} \right| > t_{\text{krit.2}}$

- Praktische significantie:

Is het verschil tussen gemeten gemiddelde en de referentiewaarde belangrijk? D.w.z. is het verschil (veel) groter dan de ruis is de data? Praktische significantie = effectsterkte: Cohen’s d: $d_{\text{av}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p}$

Effectsterkte: Cohen's d



Jacob Cohen

De effectsterkte bij een Welch t-toets wordt meestal uitgedrukt in de waarde van Cohen's d:

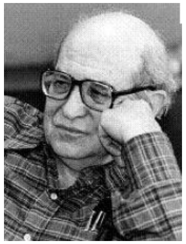
$$d_{\text{av}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p}$$

- Cohen's d meet dus het verschil $\bar{y}_1 - \bar{y}_2$ in "eenheden" van de gemiddelde ruis = gepoolde standaarddeviatie s_p
- Cohen's d is dimensieloos, $\pm$ onafhankelijk van n!
- Wat is "groot" verschil, wat is "klein" verschil? Hangt van de context af, maar algemene richtlijnen*:

Waarde Cohen's d	Interpretatie
0.2	Klein verschil
0.5	Matig verschil
0.8	Groot verschil

*) Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Effectsterkte: Cohen's d



Cohen's d: $d_{\text{av}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p}$

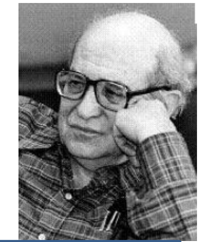
Berekening:

$$d_{\text{av}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p} = t \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

want

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{y}_1 - \bar{y}_2)}{s_p} \cdot \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = d_{\text{av}} \cdot \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

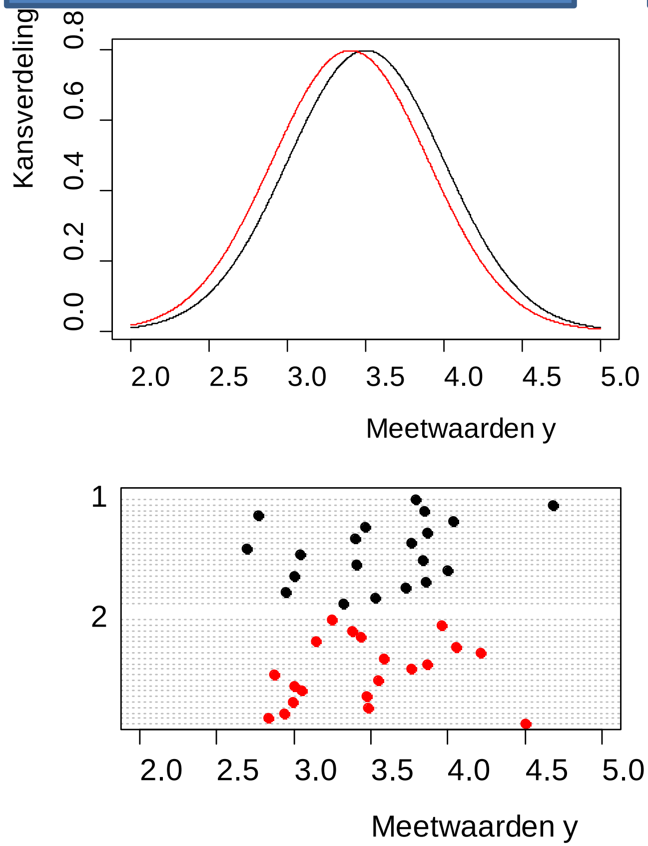
Effectsterkte: Cohen's d



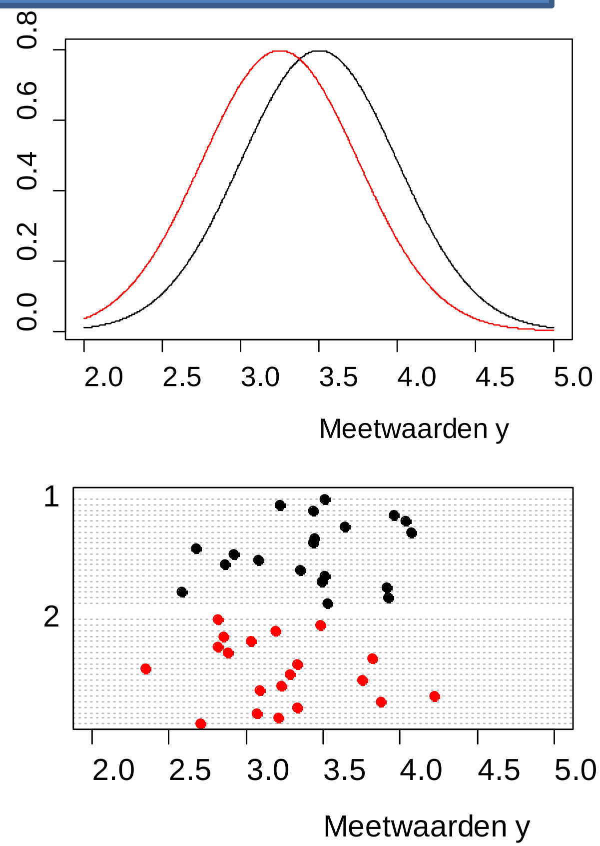
Cohen's d: $d_{\text{av}} = \dfrac{\bar{y}_1 - \bar{y}_2}{s_p}$

Voorbeeld: $\bar{y}_1 = 3.50$, $s_p = 0.50$

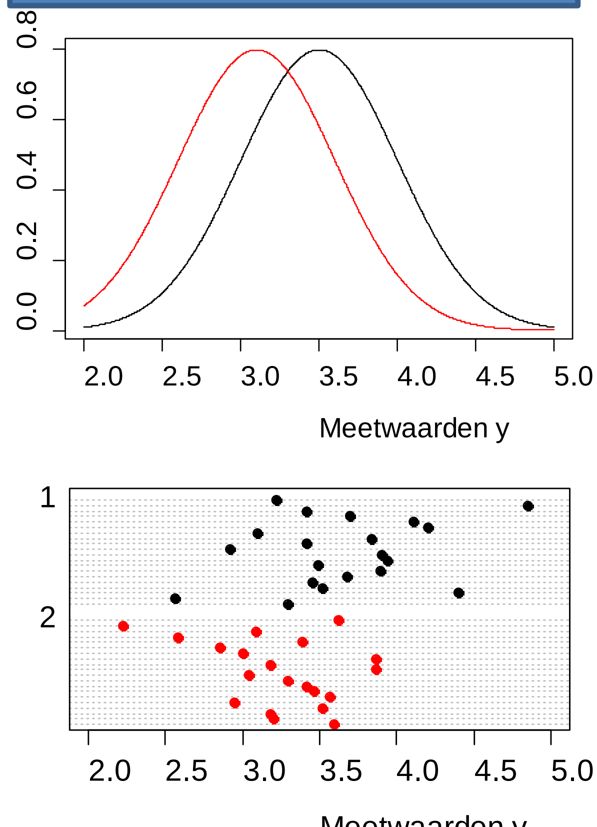
$d_{\text{av}} = 0.20$: klein verschil



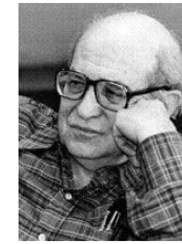
$d_{\text{av}} = 0.50$: matig verschil



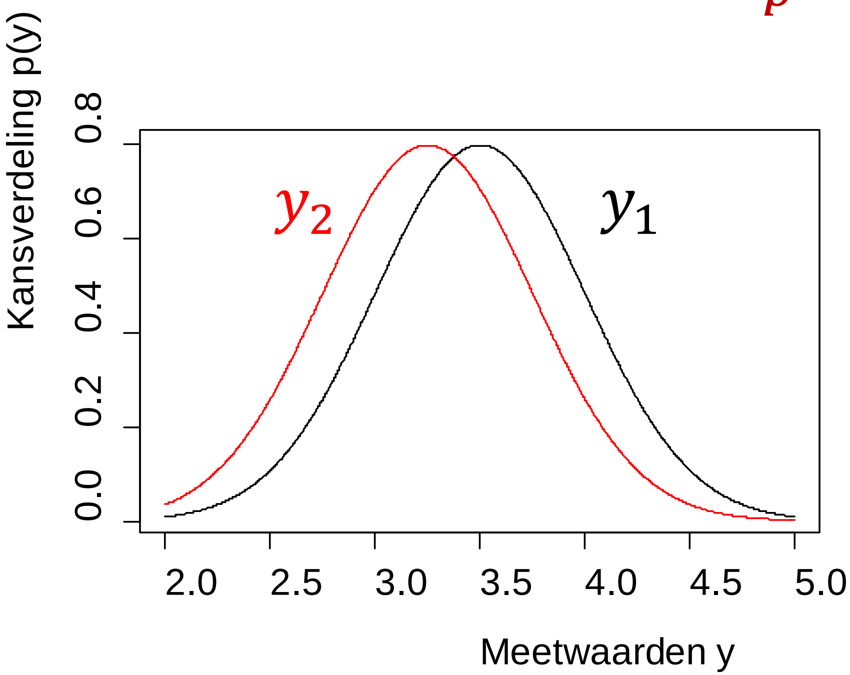
$d_{\text{av}} = 0.80$: groot verschil



Effectsterkte: Cohen's d

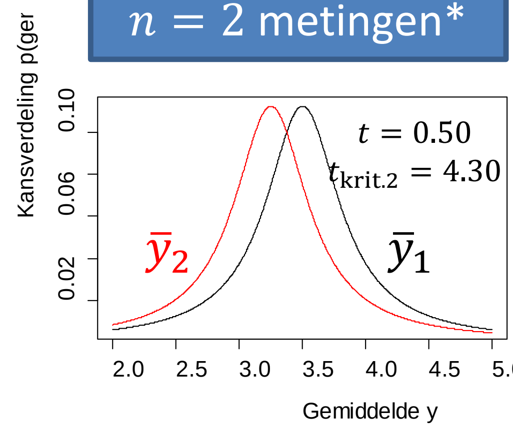


Cohen's d: $d_{\mathrm{av}} = \dfrac{\bar{y}_1 - \bar{y}_2}{s_p}$

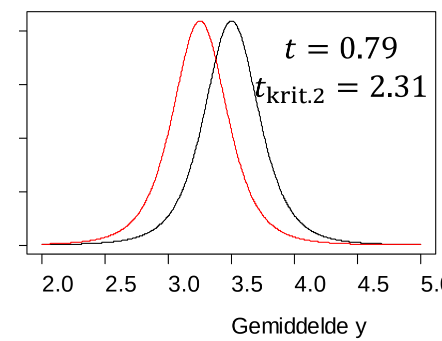


Voorbeeld: $\bar{y}_1 = 3.50$, $\bar{y}_2 = 3.25$, $s_p = 0.50$

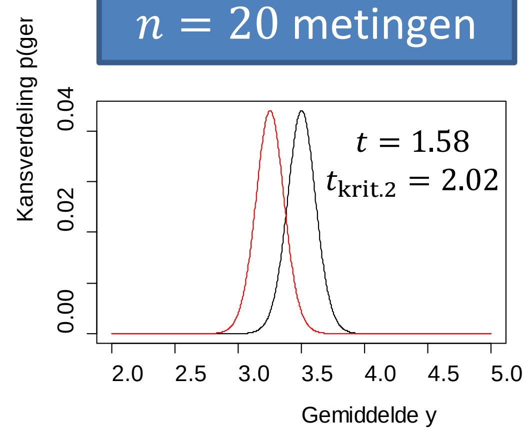
$n = 2$ metingen*



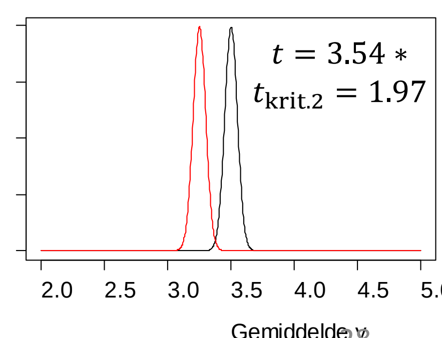
$n = 5$ metingen



$n = 20$ metingen



$n = 100$ metingen



Effectsterkte: $d_{\mathrm{av}} = \dfrac{3.50 - 3.25}{0.50} = +0.50$, dus een matig verschil!

* Per dataset n, totaal $2n$ metingen

Aantal metingen n en effectsterkte

- Om een "klein" verschil significant te noemen heb je veel metingen nodig…
- Om een "groot" verschil significant te noemen heb je minder metingen nodig…
- We hebben $t = d_{\mathrm{av}}/\sqrt{\frac{1}{n}+\frac{1}{n}}$, dus $t_{\mathrm{krit.2}} = d_{\mathrm{av}} \cdot \sqrt{n/2}$, dus

$$\mathrm{qt}(0.975; 2n-2) = d_{\mathrm{av}} * \mathrm{sqrt}(n/2)$$

Waarde Cohen's d_{av}	Interpretatie	Minimum aantal metingen n per set voor significant verschil
0.2	Klein verschil	194
0.5	Matig verschil	32
0.8	Groot verschil	14

Welch *t*-toets in R

- 2 vectoren y.1 en y.2:

  `t.test(y.1, y.2, …)`
- 2 kolommen y.1 en y.2 uit dataframe M:

  `t.test(M$y.1, M$y.2, …)`

  `t.test(y.1, y.2, data = M, …)`
- 1 vector met y-waarden, 1 factor g, via formule:

  `t.test(y ~ g, …)`
- 1 vector met y-waarden, 1 factor g uit dataframe M:

  `t.testt.test(M$y ~ M$g, …)`

  `t.test(y ~ g, data = M, …)`
- 1- en 2-zijdig toetsen*:
  - 2-zijdig: alternative="two.sided" (default)
  - 1-zijdig: alternative="less" or alternative="greater"

* Let op bij 1-zijdig: R nummert de levels van de factor g standaard *alfabetisch*!

Welch t-toets: samenvatting

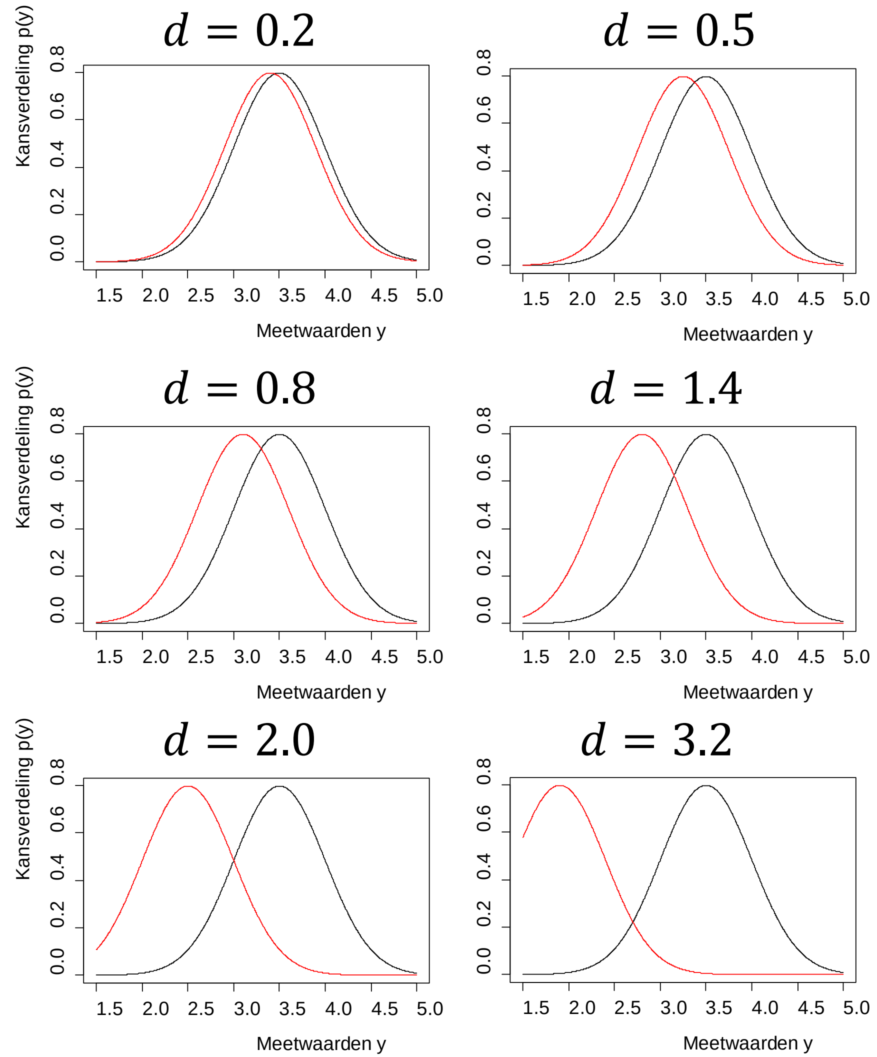
- Meetwaarden y continu, $a = 2$ groepen
- H_0:$\bar{y}_1 \neq \bar{y}_2$
- H_1: $\bar{y}_1 \neq \bar{y}_2$ of $\bar{y}_1 > \bar{y}_2$ of $\bar{y}_1 < \bar{y}_2$
- $t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$ $\mathrm{df} = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 \Big/ \left(\frac{s_1^4}{n_1^2 \cdot (n_1 - 1)} + \frac{s_2^4}{n_2^2 \cdot (n_2 - 1)}\right)$
- $t_{\mathrm{krit.2}} = \mathrm{qt}(0.975, \mathrm{df})$ $t_{\mathrm{krit.1}} = \mathrm{qt}(0.95, \mathrm{df})$
- H_1 als: $p < \alpha$

  $|\, t \,| > t_{\mathrm{krit.2}}$ of $t > t_{\mathrm{krit.1}}$ of $t < -t_{\mathrm{krit.1}}$
- Effectsterkte: $d_{\mathrm{av}} = \frac{(\bar{y}_1 - \bar{y}_2)}{s_p} = t \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

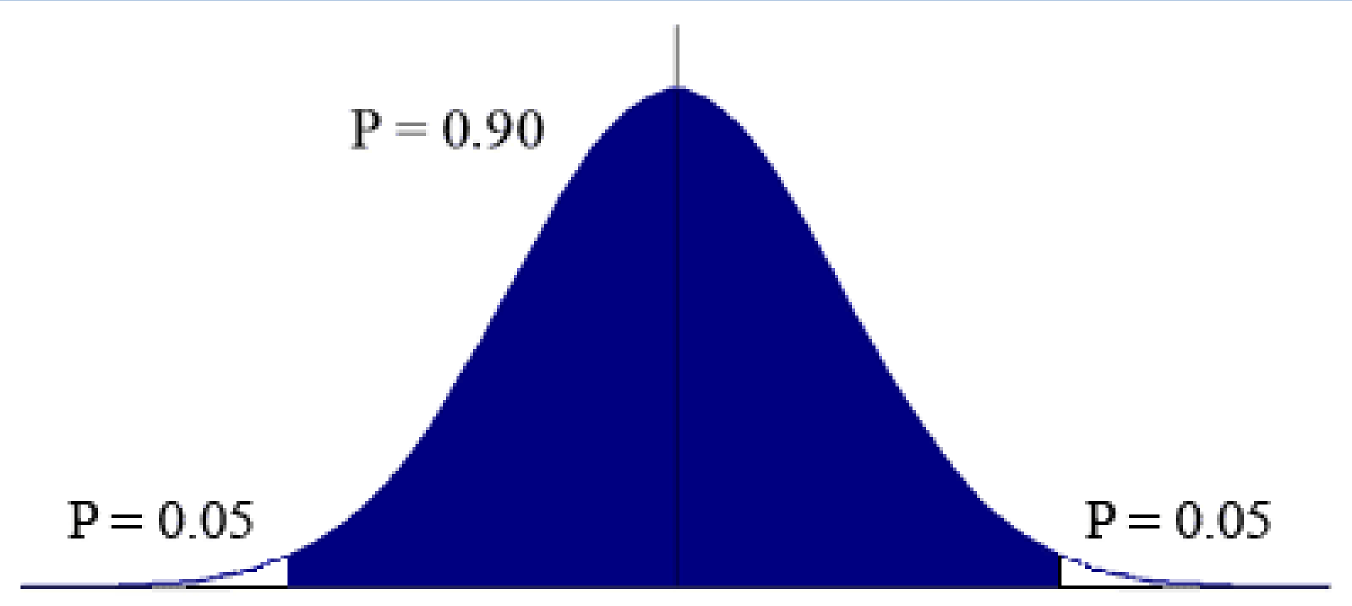
Waarde Cohen's d: $\mid d_{\mathrm{av}} \mid$	Interpretatie
0.2	Klein verschil
0.5	Matig verschil
0.8	Groot verschil

Aantal metingen voor significant verschil

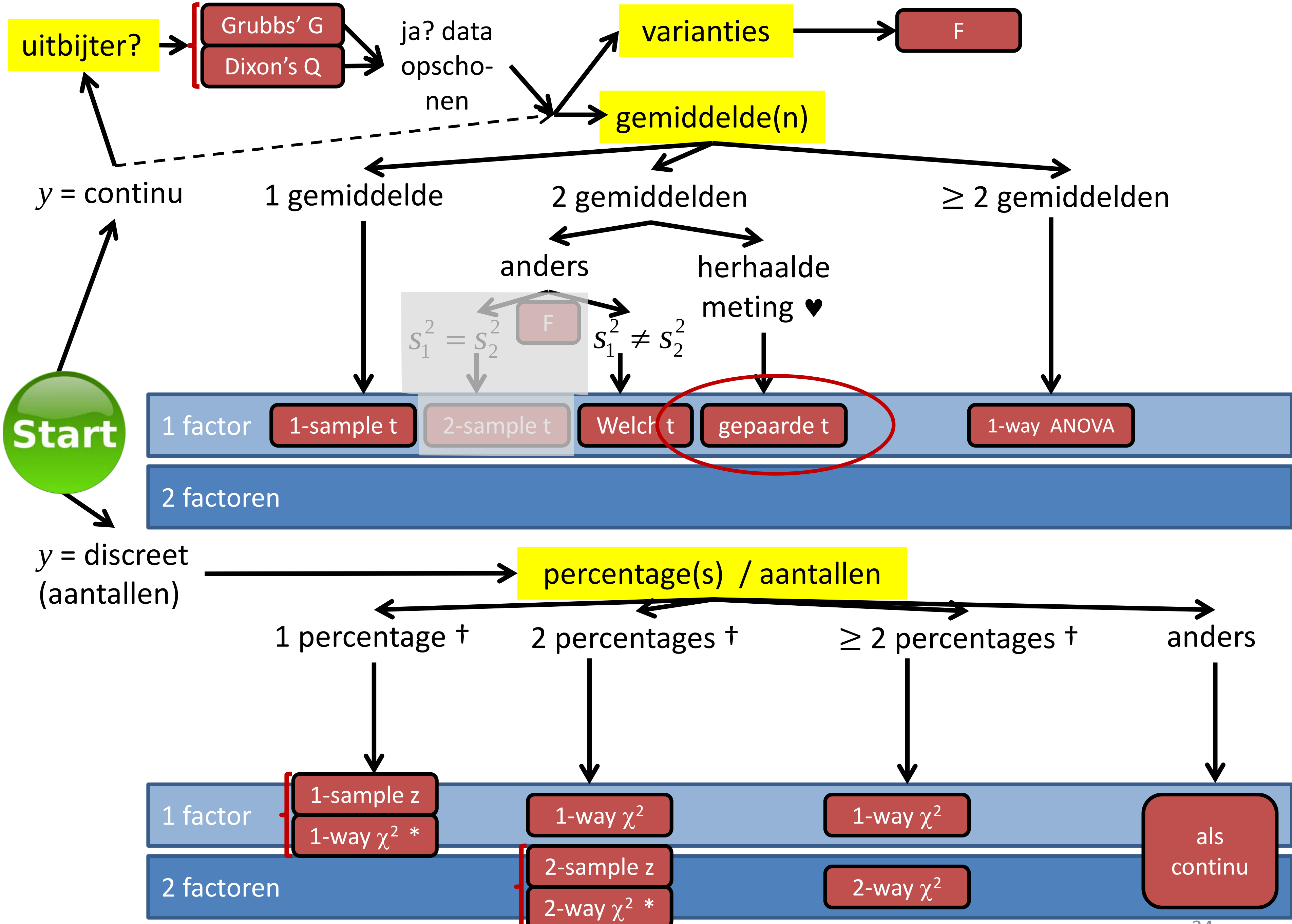
Cohen's d	1-sample t-toets: n metingen	Welch t-toets: n metingen per dataset
0.2	99	194
0.5	18	32
0.8	9	14
1.1	6	8
1.4	5	6
1.7	4	4
2.0	4	4
2.3	3	3
2.6	3	3
2.9	3	3
3.2	3	2
3.5	3	2



Gepaarde *t*-toets



Verschillen twee experimentele gemiddelden op basis van *herhaalde metingen* van elkaar?



* alleen voor 2-zijdige vraag over percentages ♥ herhaalde meting op *zelfde* "object" onder *andere* omstandigheden † onafhankelijk

Wel/niet gepaard?

Structuur in de verzamelde data:

- Is het een herhaalde meting van een eigenschap op hetzelfde "object" onder verschillende omstandigheden?
- Kan je niet zomaar datapunten op een andere plek in de dataset zetten?
- Spreiding per groep is niet normaal verdeeld?

Dan: gepaarde *t*-toets (Eng: paired *t*-test, repeated measures *t*-test)

Gepaarde t-toets

$$d_i = y_{1,i} - y_{2,i}$$

Als formele t-toets:

$$\boxed{t = \frac{\text{verschil tussen gemiddelden}}{\text{fout in verschil}}}$$

wordt

$$\text{df} = n_d - 1$$

$$d_i = y_{1,i} - y_{2,i}$$

verschil per paar

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n_d}}$$

aantal paren

- Er is een significant *verschil* tussen $\bar{y}_1$ en $\bar{y}_2$ als:
  $$|\,t\,| > t_{\text{krit.2}}$$
- Gemiddelde $\bar{y}_1$ is *hoger* dan $\bar{y}_2$ als:
  $$t > t_{\text{krit.1}}$$
- Gemiddelde $\bar{y}_1$ is *lager* dan $\bar{y}_2$ als:
  $$t < -t_{\text{krit.1}}$$

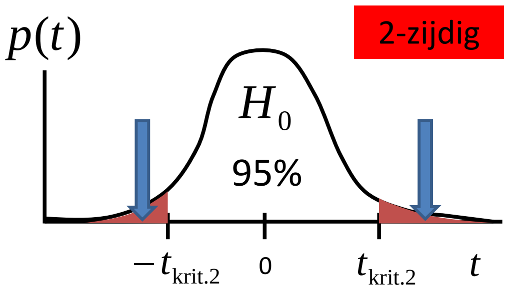
Grenswaarde (voor $\alpha = 0.05$):

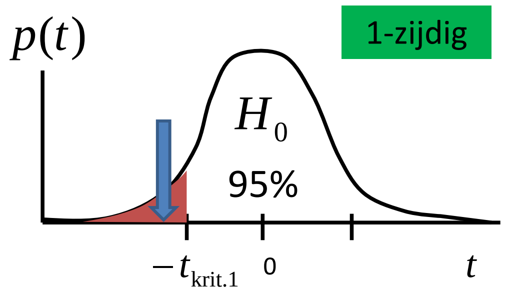
2-zijdig: $t_{\text{krit.2}} = \text{qt}\,(0.975, \text{df})$

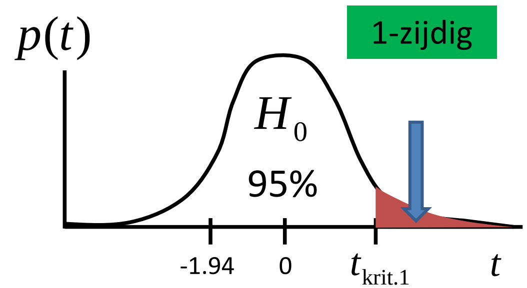
1-zijdig: $t_{\text{krit.1}} = \text{qt}\,(0.95, \text{df})$

Significant of niet?

Een verschil is significant als







$$|t| > t_{\text{krit.2}} \qquad t < -t_{\text{krit.1}} \qquad t > t_{\text{krit.1}}$$

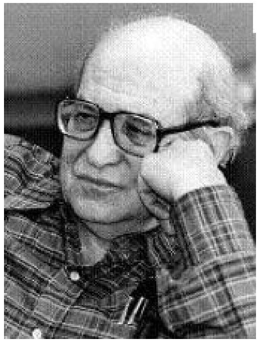
of (wat equivalent is)

$$p < \alpha$$

Geeft R

"Kans dat H_0 waar is"

Effectsterkte: Cohen's d



Jacob Cohen

De effectsterkte bij een gepaarde t-toets wordt meestal uitgedrukt in de waarde van Cohen's d:

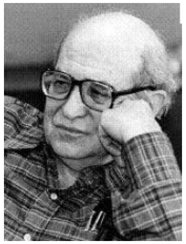
$$d_{\text{av}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p}$$

- Cohen's d meet dus het verschil $\bar{y}_1 - \bar{y}_2$ in "eenheden" van de gemiddelde ruis = gepoolde standaarddeviatie s_p
- Cohen's d is dimensieloos, $\pm$ onafhankelijk van n!
- Wat is "groot" verschil, wat is "klein" verschil? Hangt van de context af, maar algemene richtlijnen*:

Waarde Cohen's d	Interpretatie
0.2	Klein verschil
0.5	Matig verschil
0.8	Groot verschil

*) Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Effectsterkte: Cohen's d



Cohen's d: $d_{\mathrm{av}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p}$

Berekening:

$$d_{\mathrm{av}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p} = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

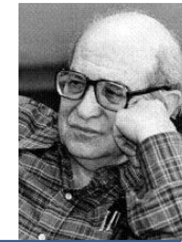
want t is gedefinieerd op basis van s_d, de standaarddeviatie van de verschillen d_i, en dus NIET bruikbaar om d_{av} te berekenen!

NB. Sommige auteurs definiëren ook een andere d-waarde:

$$d_z = \frac{\bar{y}_1 - \bar{y}_2}{s_d}$$

waarbij dus gedeeld wordt door de standaarddeviatie van de verschillen! Dit levert een veel grotere effectsterkte op. Wij kiezen als effectsterkte voor de waarde van d_{av} i.p.v. d_z…

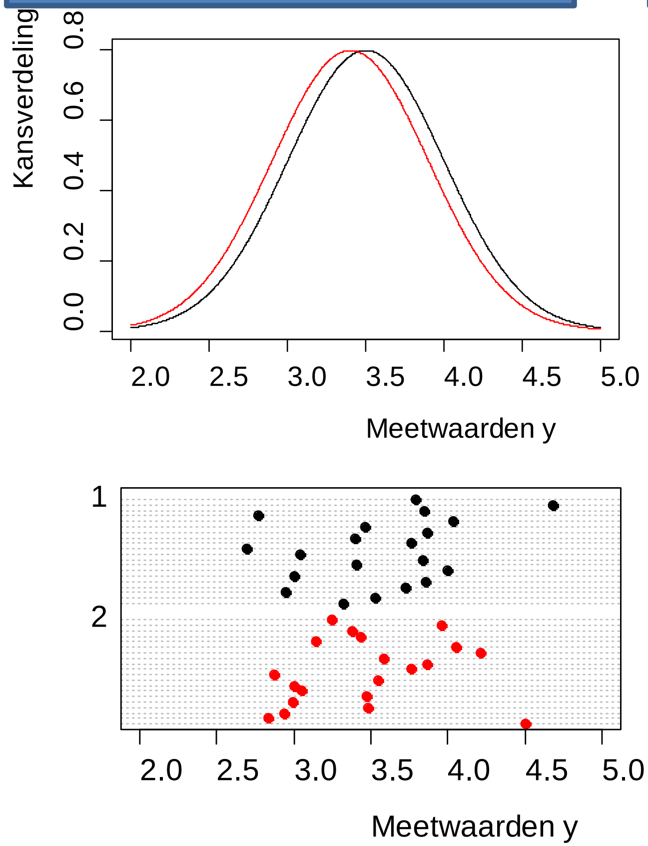
Effectsterkte: Cohen's d



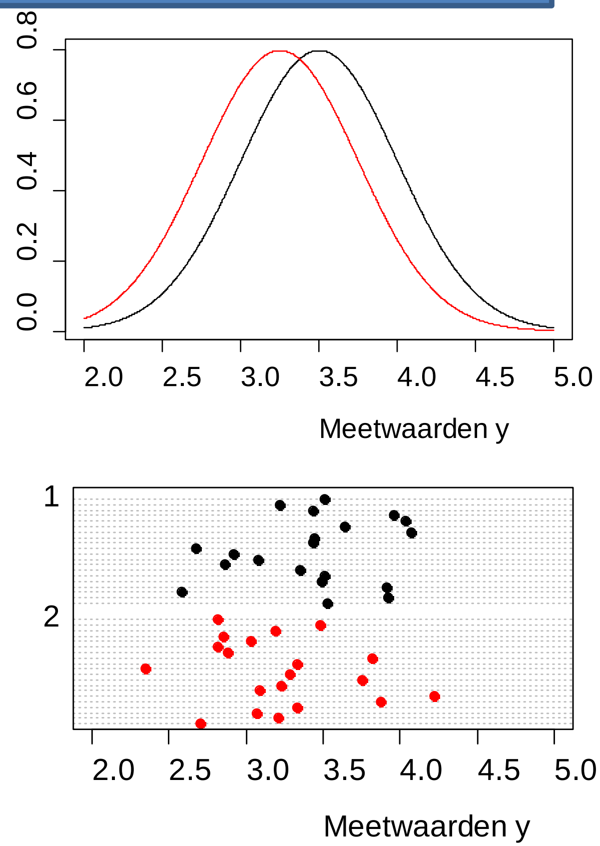
Cohen's d: $d_{\text{av}} = \dfrac{\bar{y}_1 - \bar{y}_2}{s_p}$

Voorbeeld: $\bar{y}_1 = 3.50$, $s_p = 0.50$

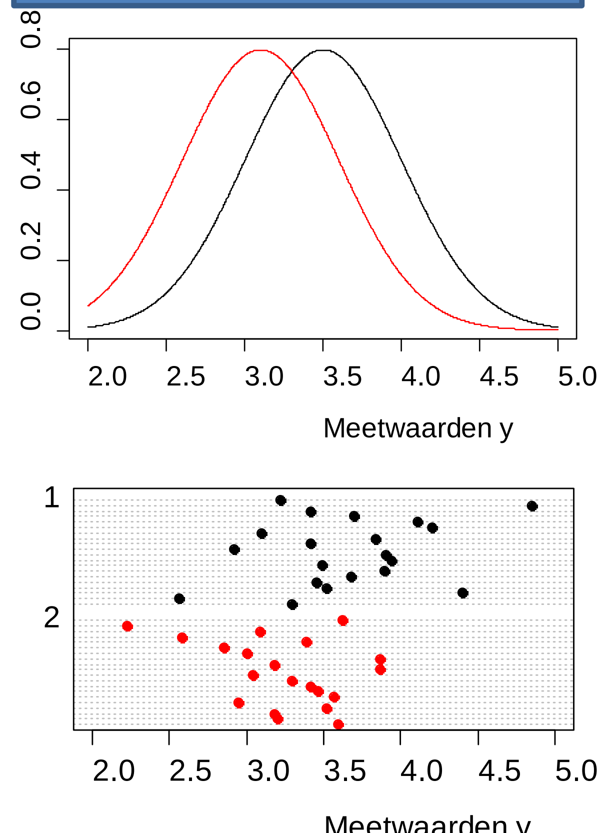
$d_{\text{av}} = 0.20$: klein verschil



$d_{\text{av}} = 0.50$: matig verschil



$d_{\text{av}} = 0.80$: groot verschil



Gepaarde t-toets in R

- 2 vectoren y.1 en y.2:

  t.test(y.1, y.2, paired=T, ...)
- 2 kolommen y.1 en y.2 uit dataframe M:

  t.test(M$y.1, M$y.2 , paired=T, ...)

  t.test(y.1, y.2, data = M , paired=T, ...)
- 1 vector met y-waarden, 1 factor g, via formule:

  t.test(y ~ g , paired=T, ...)
- 1 vector met y-waarden, 1 factor g uit dataframe M:

  t.test(M$y ~ M$g , paired=T, ...)

  t.test(y ~ g, data = M , paired=T, ...)
- 1- en 2-zijdig toetsen*:
  - 2-zijdig: alternative="two.sided" (default)
  - 1-zijdig: alternative="less" or alternative="greater"

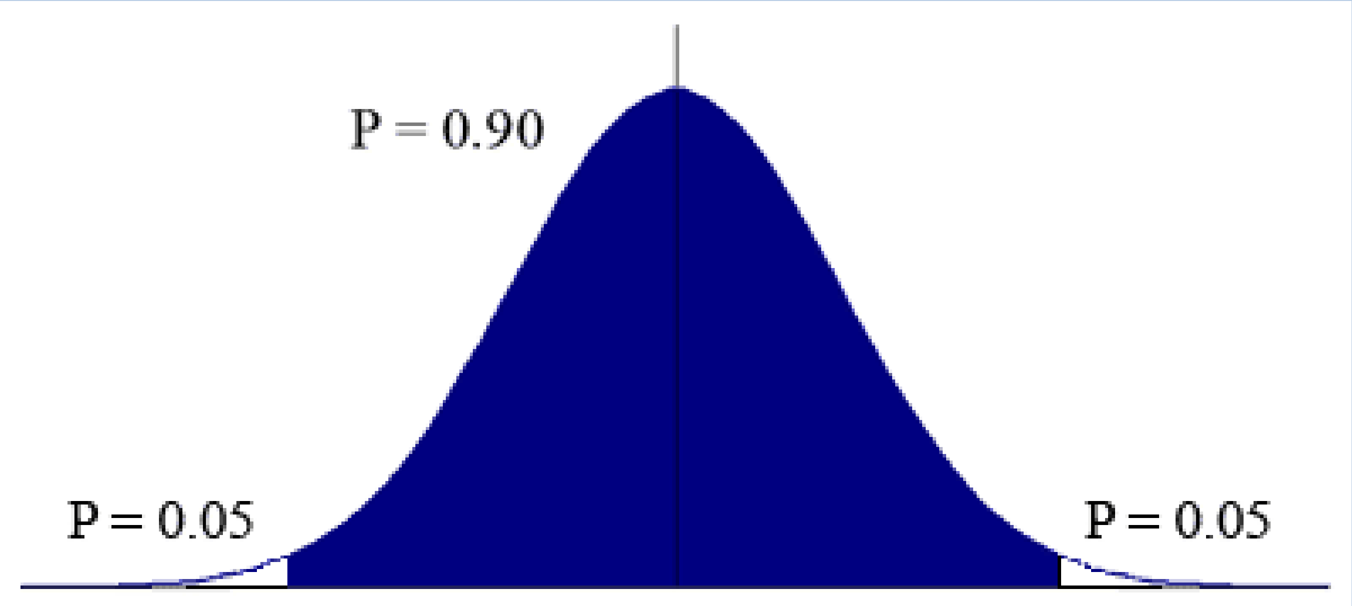
* Let op bij 1-zijdig: R nummert de levels van de factor g standaard *alfabetisch*!

Gepaarde t-toets: samenvatting

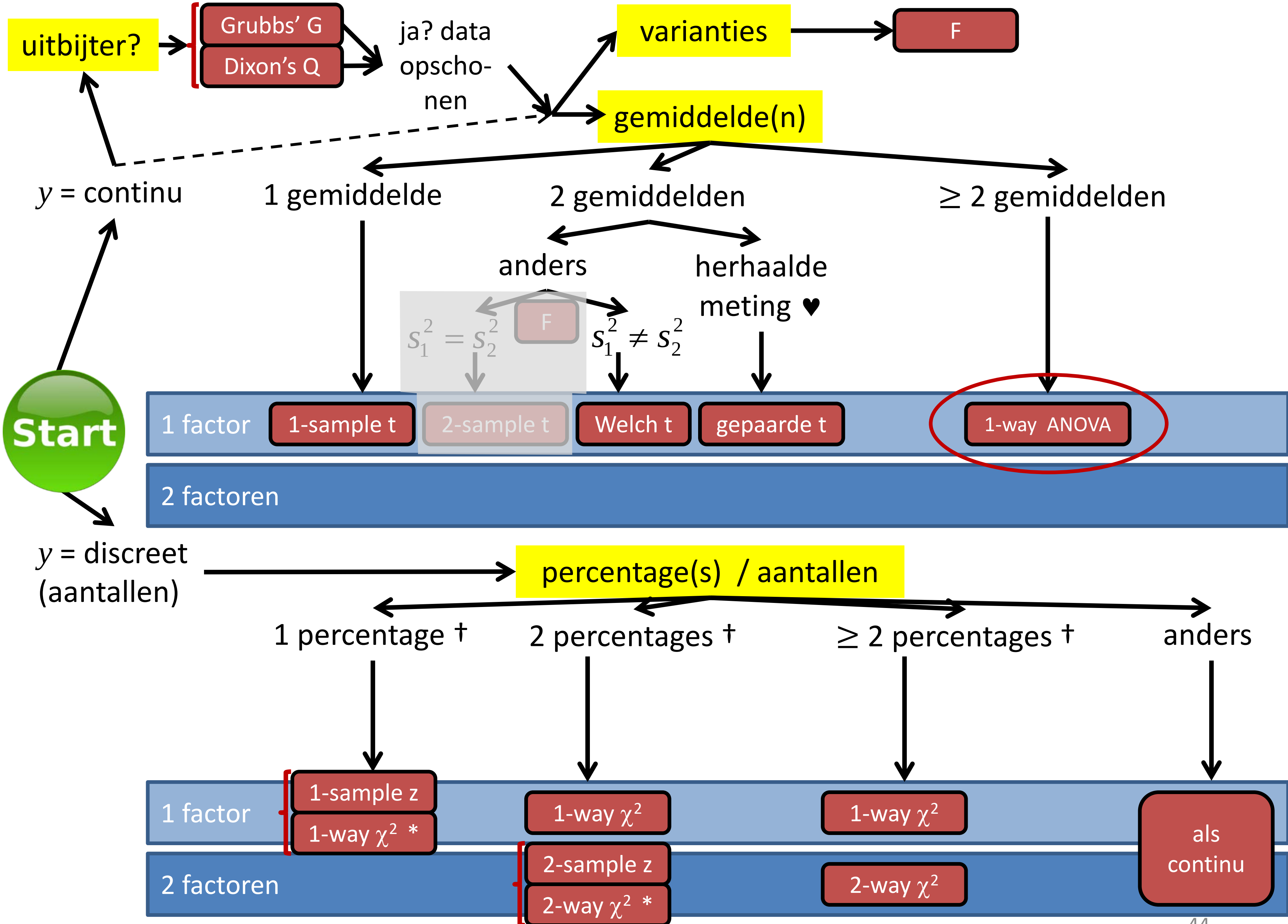
- Meetwaarden y continu, $a = 2$ groepen
- H_0:$\bar{y}_1 \neq \bar{y}_2$
- H_1: $\bar{y}_1 \neq \bar{y}_2$ of $\bar{y}_1 > \bar{y}_2$ of $\bar{y}_1 < \bar{y}_2$
- $t = \frac{(\bar{d}-0)}{s_d/\sqrt{n_d}}$ $\quad \mathrm{df} = n_d - 1 \quad d_i = y_{1,i} - y_{2,i}$
- $t_{\mathrm{krit.2}} = \mathrm{qt}(0.975, \mathrm{df})$ $\quad t_{\mathrm{krit.1}} = \mathrm{qt}(0.95, \mathrm{df})$
- H_1 als: $p < \alpha$
  $|\, t \,| > t_{\mathrm{krit.2}}$ of $t > t_{\mathrm{krit.1}}$ of $t < -t_{\mathrm{krit.1}}$
- Effectsterkte: $d_{\mathrm{av}} = \frac{(\bar{y}_1 - \bar{y}_2)}{s_p} = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$

Waarde Cohen's d: $\mid d_{\mathrm{av}} \mid$	Interpretatie
0.2	Klein verschil
0.5	Matig verschil
0.8	Groot verschil

1-way ANOVA

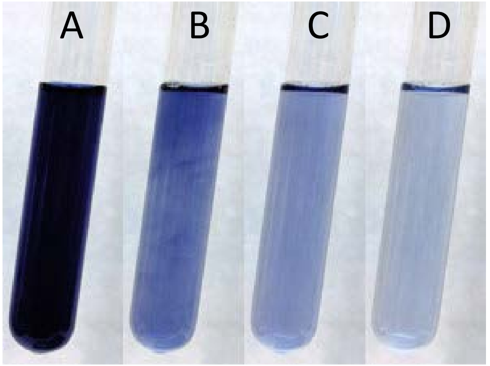


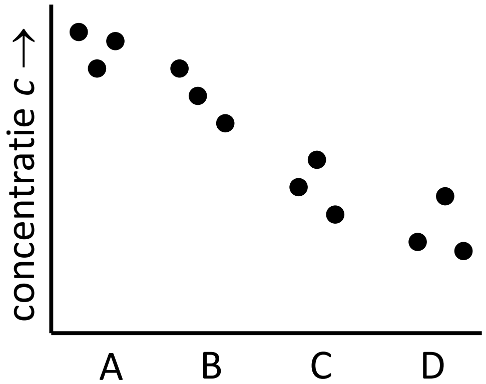
Verschillen meer dan twee experimentele gemiddelden van elkaar?



* alleen voor 2-zijdige vraag over percentages ♥ herhaalde meting op *zelfde* "object" onder *andere* omstandigheden † onafhankelijk

Meerdere groepen: t-toetsen?





Als je bijv. 4 groepen wilt vergelijken, zou je 6 verschillende t-toetsen kunnen doen:

A-B, A-C, A-D, B-C, B-D, C-D

met elk $\alpha = 0.05$ (= kans op verkeerde conclusie per toets)

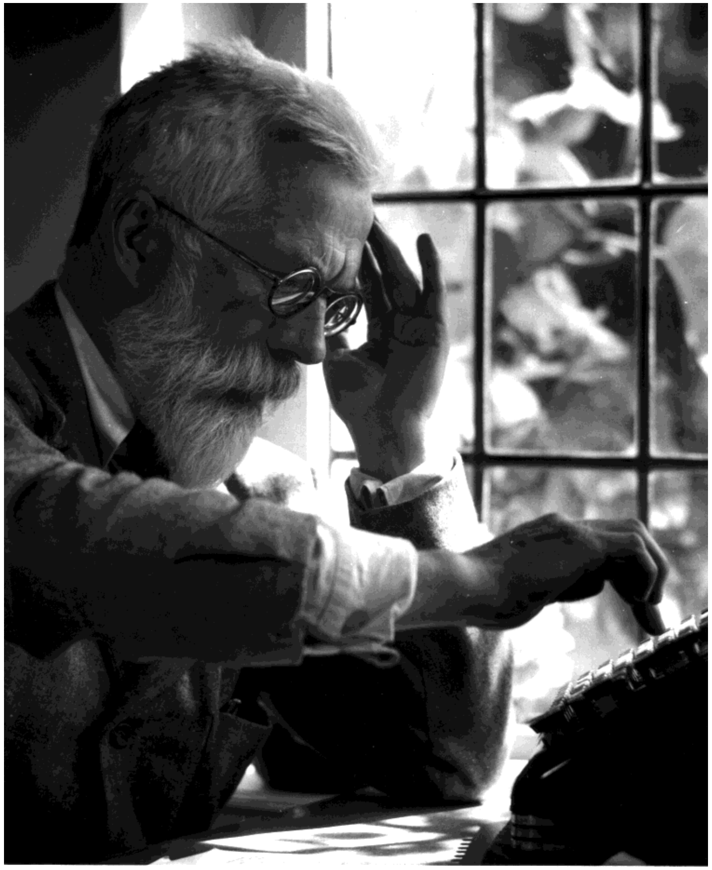
Maar: totale kans op verkeerde conclusie is:

$$\boxed{\text{kans} = 1 - 0.95^6 = 0.26}$$

Oplossing

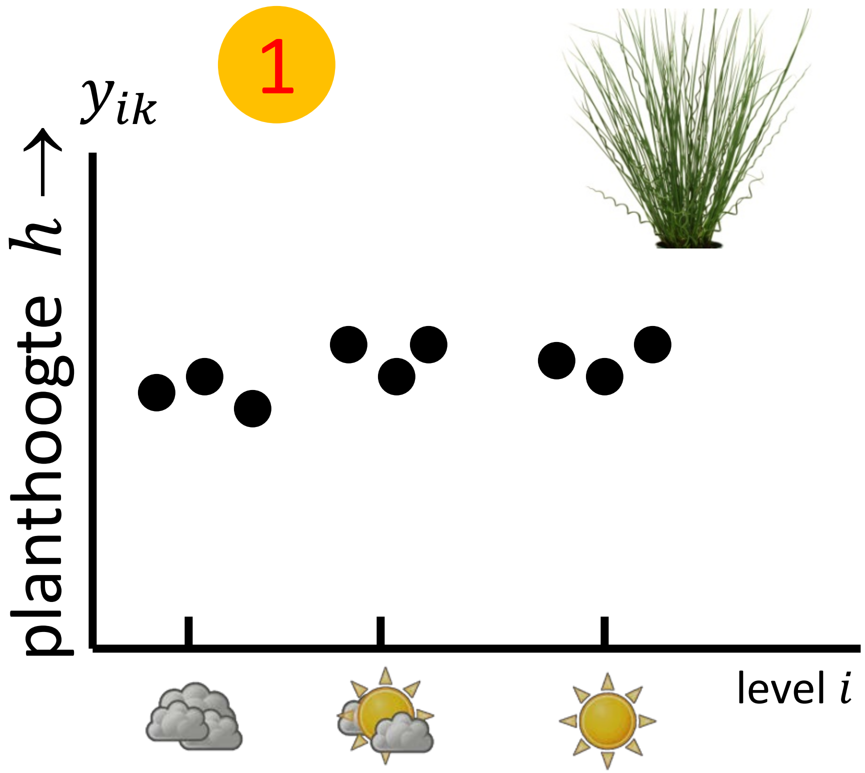
Analysis Of Variance (Variantieanalyse)

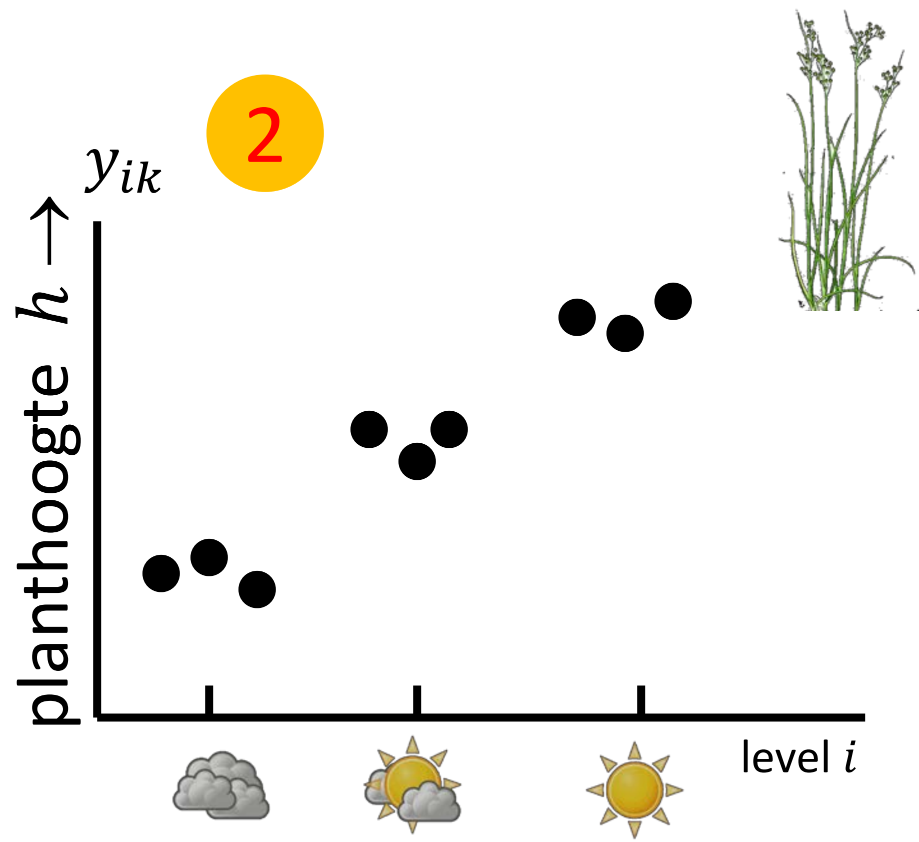
Ronald Fisher (1930)



1-way ANOVA: idee

Waarom vind je in situatie 1 dat er geen effect van licht is op de planthoogte, terwijl je dat in situatie 2 wel vindt?

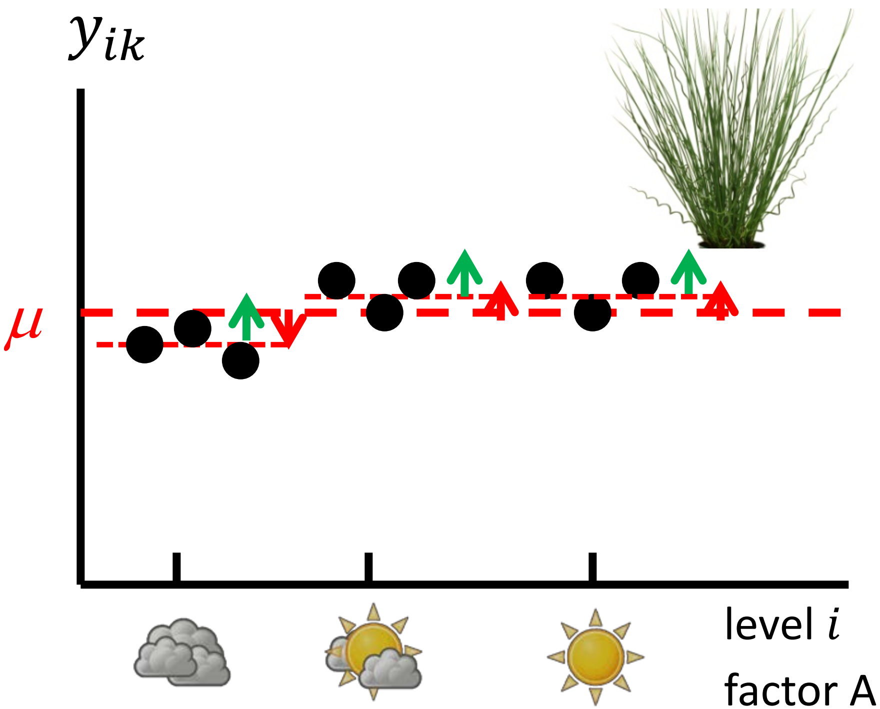


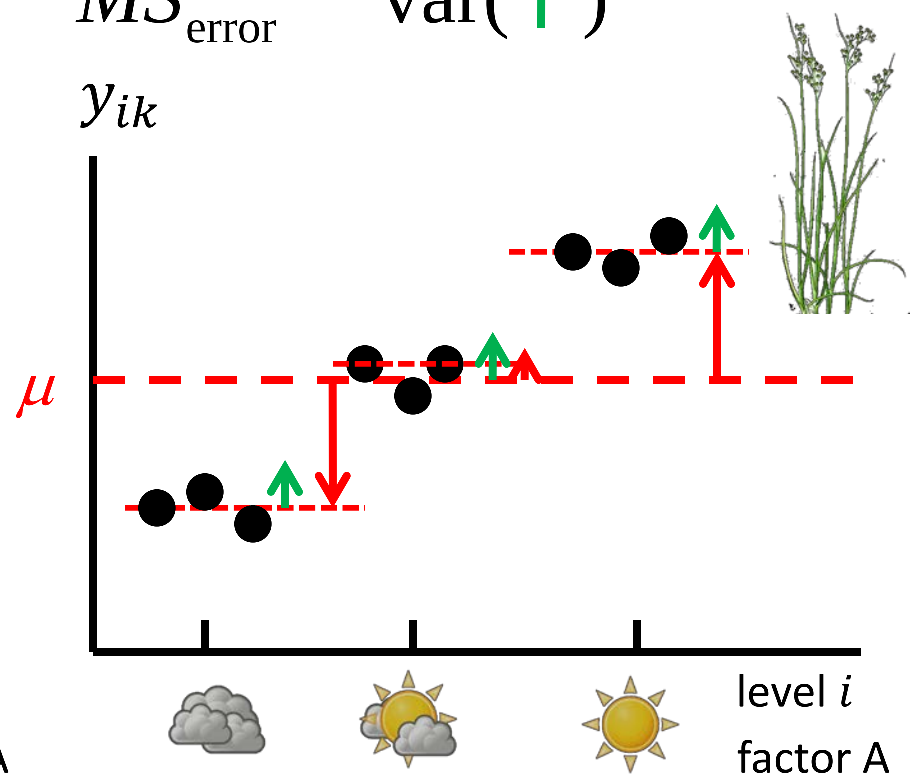


1-way ANOVA: idee

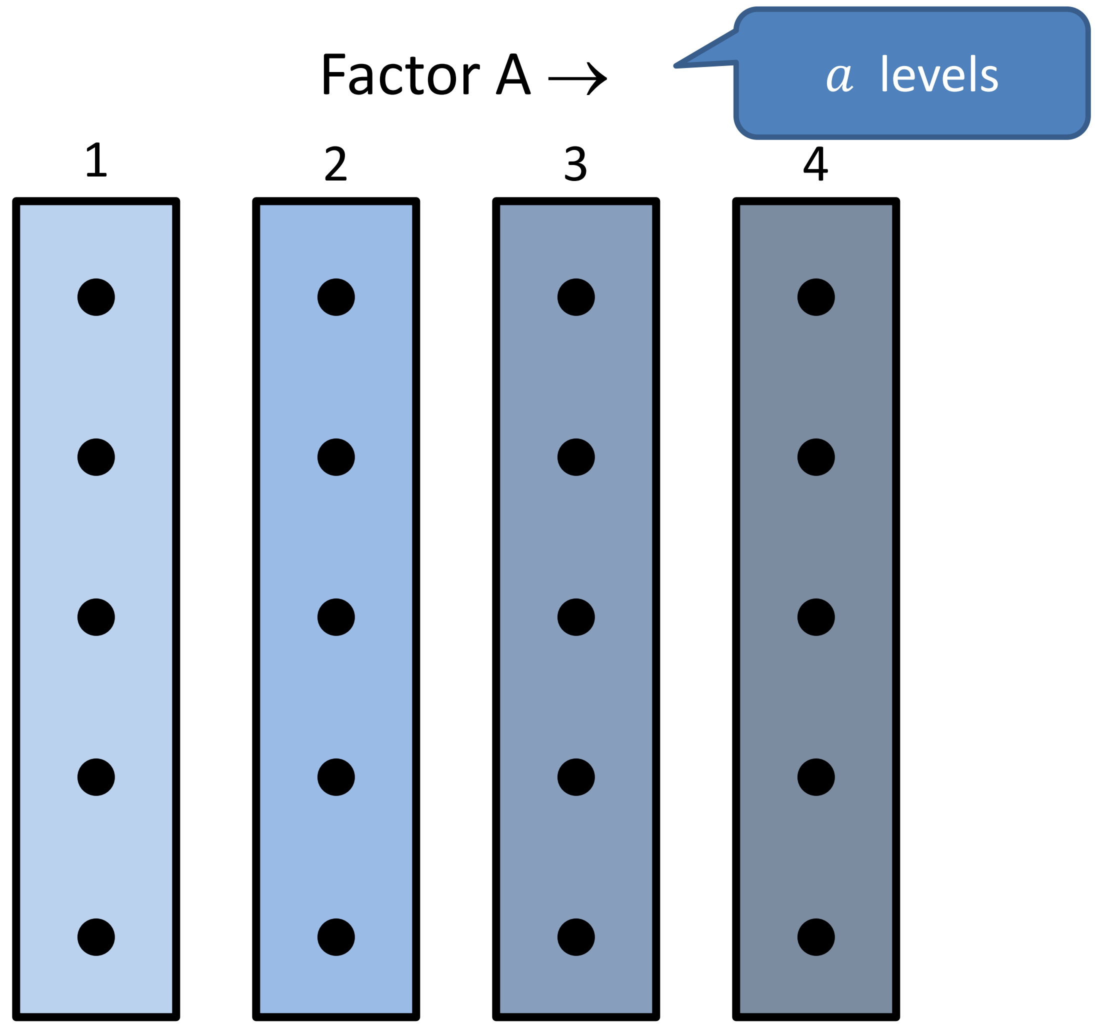
Verhouding tussen spreiding (= variantie) *tussen* levels (groepen) en *binnen* levels:

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





1-way ANOVA (fixed effects): design



1-way ANOVA: notatie

- Elke y-waarde behoort tot i^{e} level (groep) van factor en is k^{e} meting van dat level: y_{ik}
- Gemiddelde y per level i: $\overline{y}_{i\bullet}$
- Overall gemiddelde: $\overline{y}_{\bullet\bullet}$
- Aantal metingen per level ("cel"): $n_{i\bullet} = n_c$

Mag ook verschillen per level, maar liever niet!

1-way ANOVA: data

level i (factor A) $\rightarrow$

$\leftarrow$ meting k

y_{11}	y_{21}	y_{31}	y_{41}	y_{51}
y_{12}	y_{22}	y_{32}	y_{42}	y_{52}
y_{13}	y_{23}	y_{33}	y_{43}	y_{53}
y_{14}	y_{24}	y_{34}	y_{44}	y_{54}
y_{15}	y_{25}	y_{35}	y_{45}	y_{55}
y_{16}	y_{26}	y_{36}	y_{46}	y_{56}

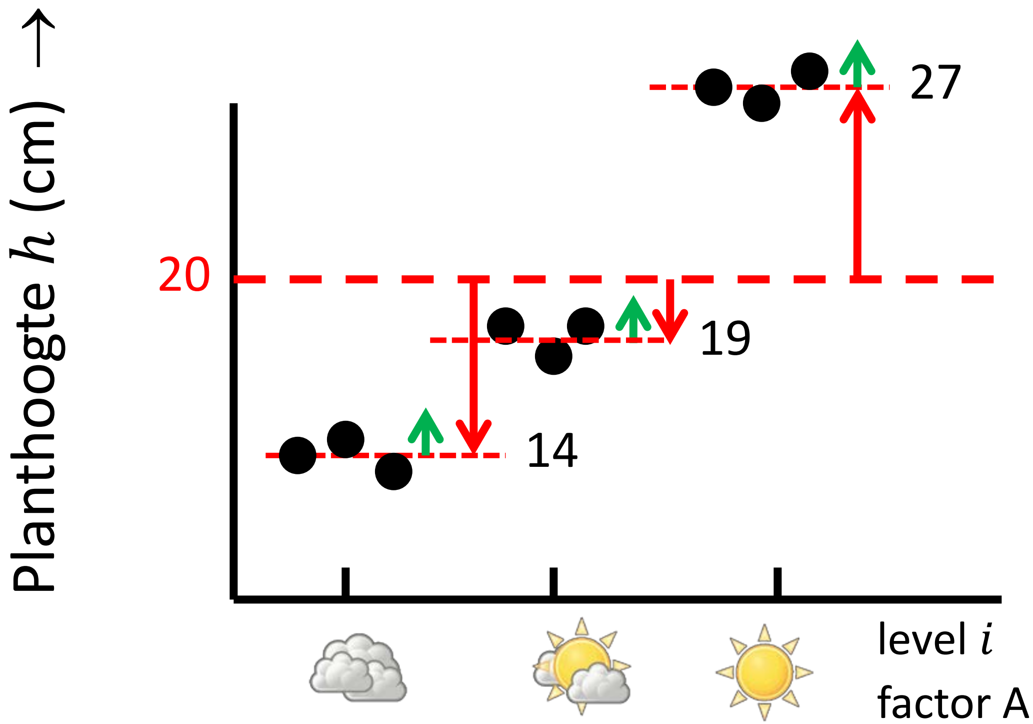
n_c

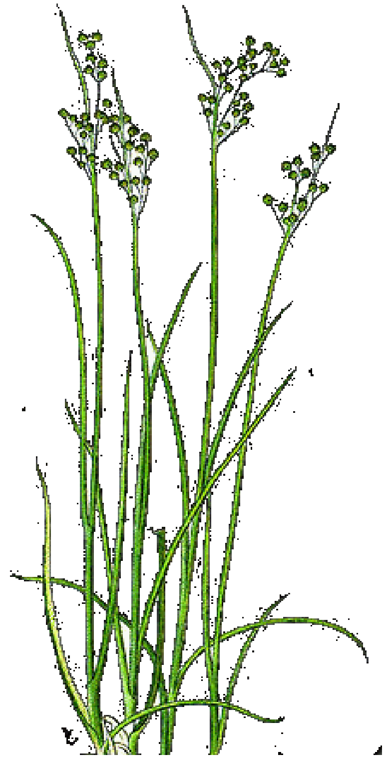
$\overline{y}_{2\bullet}$

$\overline{y}_{\bullet\bullet}$

Voorbeeld: planthoogte

Metingen van planthoogte in zon, halfschaduw en schaduw:

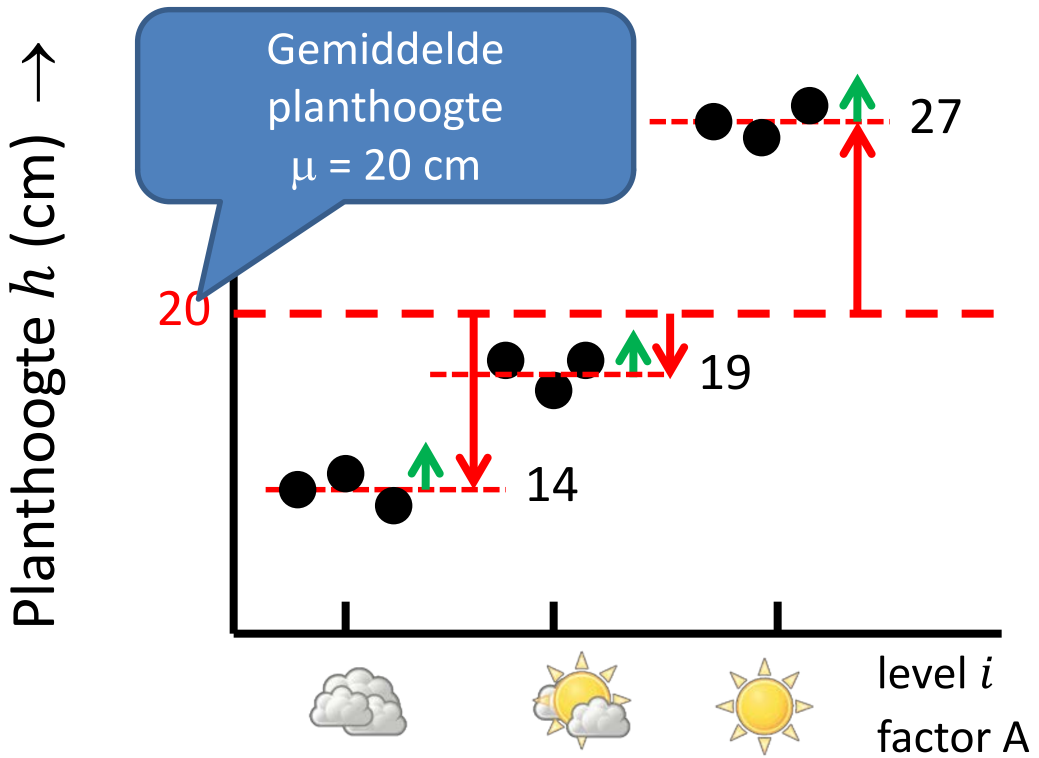


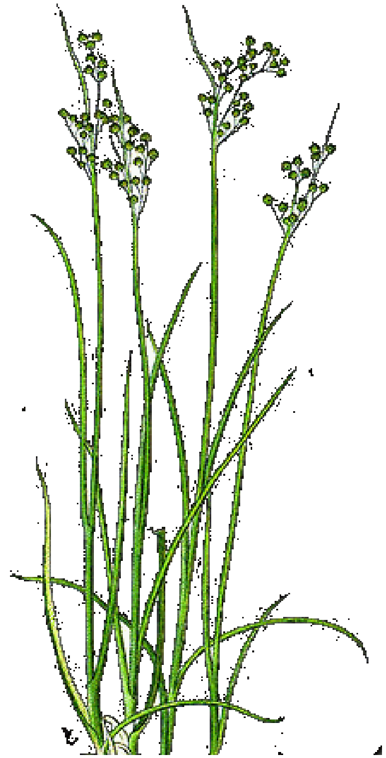


Platte rus

Voorbeeld: planthoogte

Metingen van planthoogte in zon, halfschaduw en schaduw:

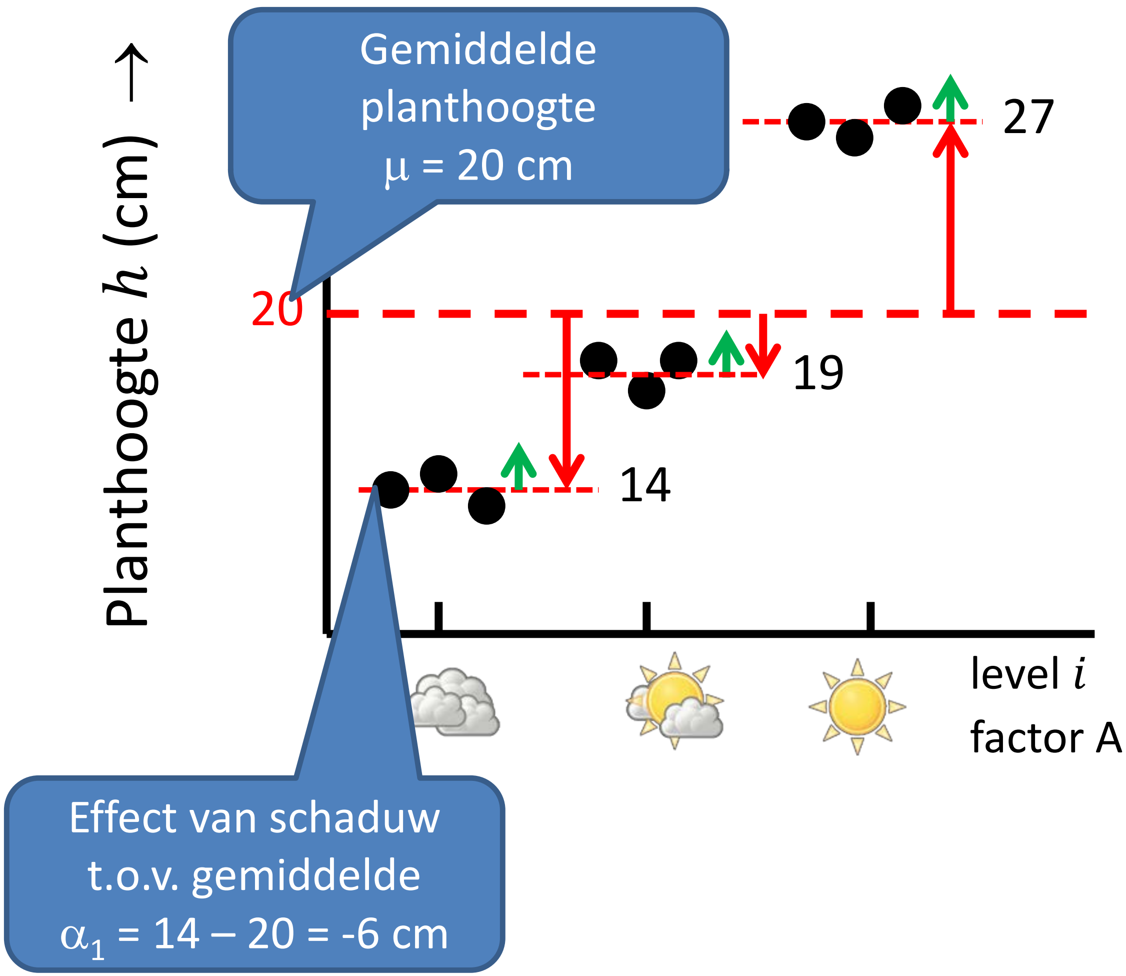




Platte rus

Voorbeeld: planthoogte

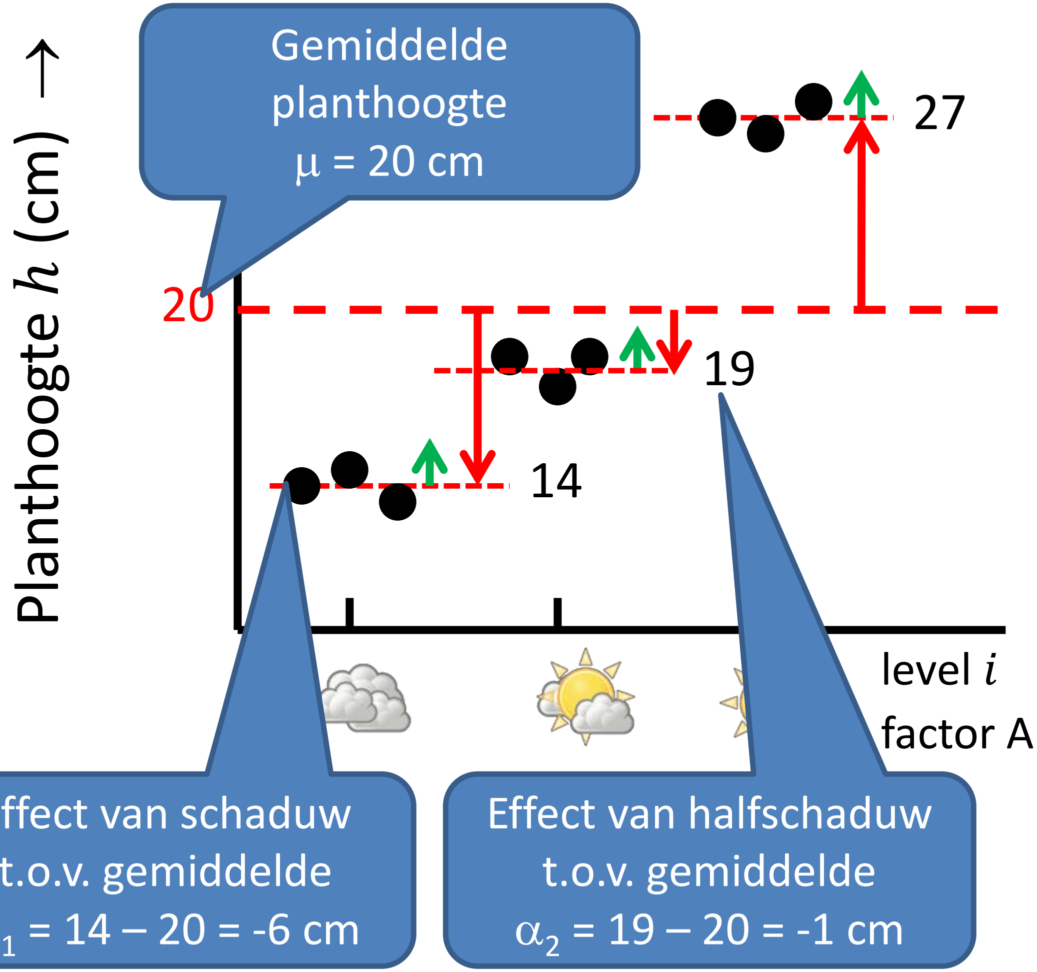
Metingen van planthoogte in zon, halfschaduw en schaduw:



Platte rus

Voorbeeld: planthoogte

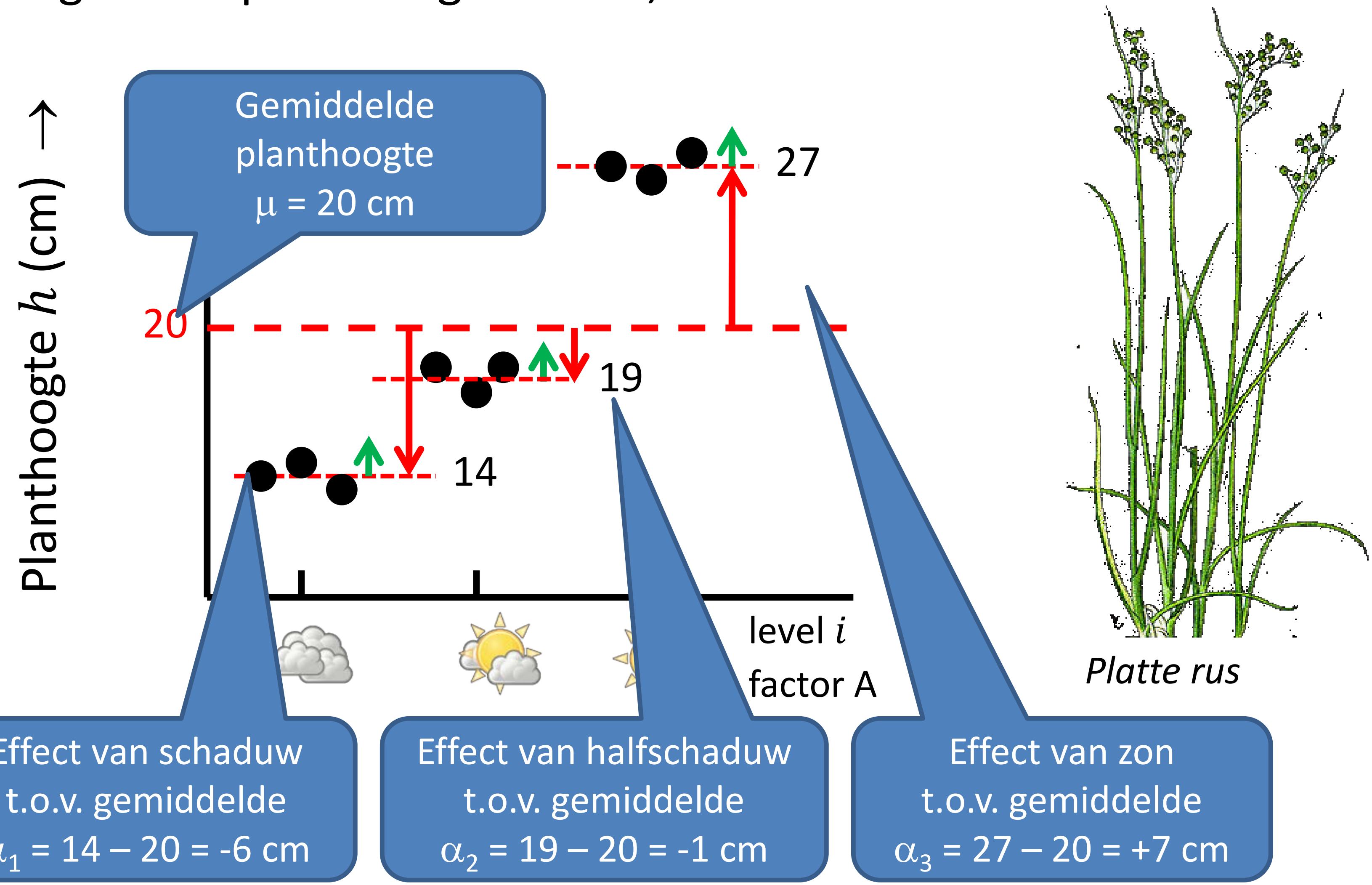
Metingen van planthoogte in zon, halfschaduw en schaduw:



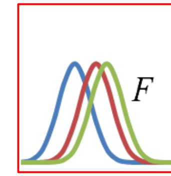
Platte rus

Voorbeeld: planthoogte

Metingen van planthoogte in zon, halfschaduw en schaduw:



1-way ANOVA: model



- Model: $y_{ik} = \mu + \alpha_i + \varepsilon_{ik}$

Effect van factor A

Error = ruis

$$y_{ik} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ik} - \bar{y}_{i.})$$

- *Sum of Squares* ("variatie") opdeling:

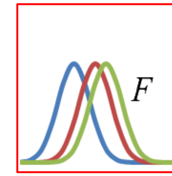
$$\underbrace{\sum_{i=1}^{a}\sum_{k=1}^{n_c}(y_{ik} - \bar{y}_{..})^2}_{SS_{\text{tot}}} = \underbrace{\sum_{i=1}^{a}\sum_{k=1}^{n_c}(\bar{y}_{i.} - \bar{y}_{..})^2}_{SS_{\text{A}}} + \underbrace{\sum_{i=1}^{a}\sum_{k=1}^{n_c}(y_{ik} - \bar{y}_{i.})^2}_{SS_{\text{err}}}$$

$$SS_{\text{tot}} = SS_{\text{A}} + SS_{\text{err}}$$

SS_{between}

SS_{within}

1-way ANOVA: hypothese



H_0: $\mu_1 = \mu_2 = \mu_3 = \cdots = \mu_a$

of

$$\alpha_1 = \alpha_2 = \alpha_3 = \cdots = \alpha_a = 0$$

alle levels hebben zelfde gemiddelde

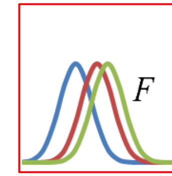
of: De factor A heeft GEEN effect op y

H_1: niet zo…

niet alle levels zelfde gemiddelde

of: De factor A heeft significant effect op y

1-way ANOVA: toets



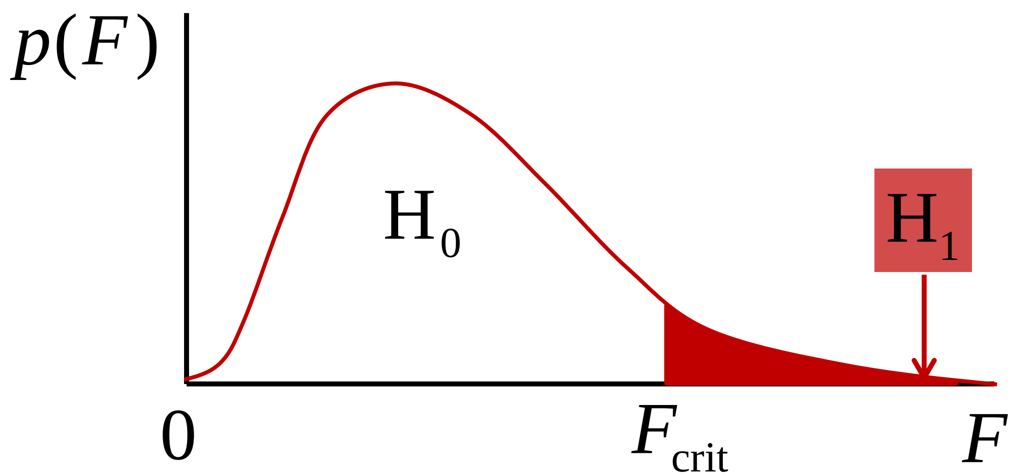
- Bereken MS's gebaseerd op SS_A en SS_err:

$$MS_\text{A} = \frac{SS_\text{A}}{a-1} \qquad MS_\text{err} = \frac{SS_\text{err}}{a(n_c-1)}$$

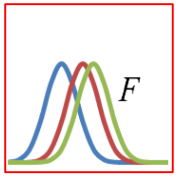
- Ratio MS_A / MS_err volgt een F- verdeling:

$$F = \frac{MS_\text{A}}{MS_\text{err}} = \frac{\text{var}(\uparrow)}{\text{var}(\uparrow)}$$

$$F_\text{crit} = F_{a-1,a(n_c-1)}$$

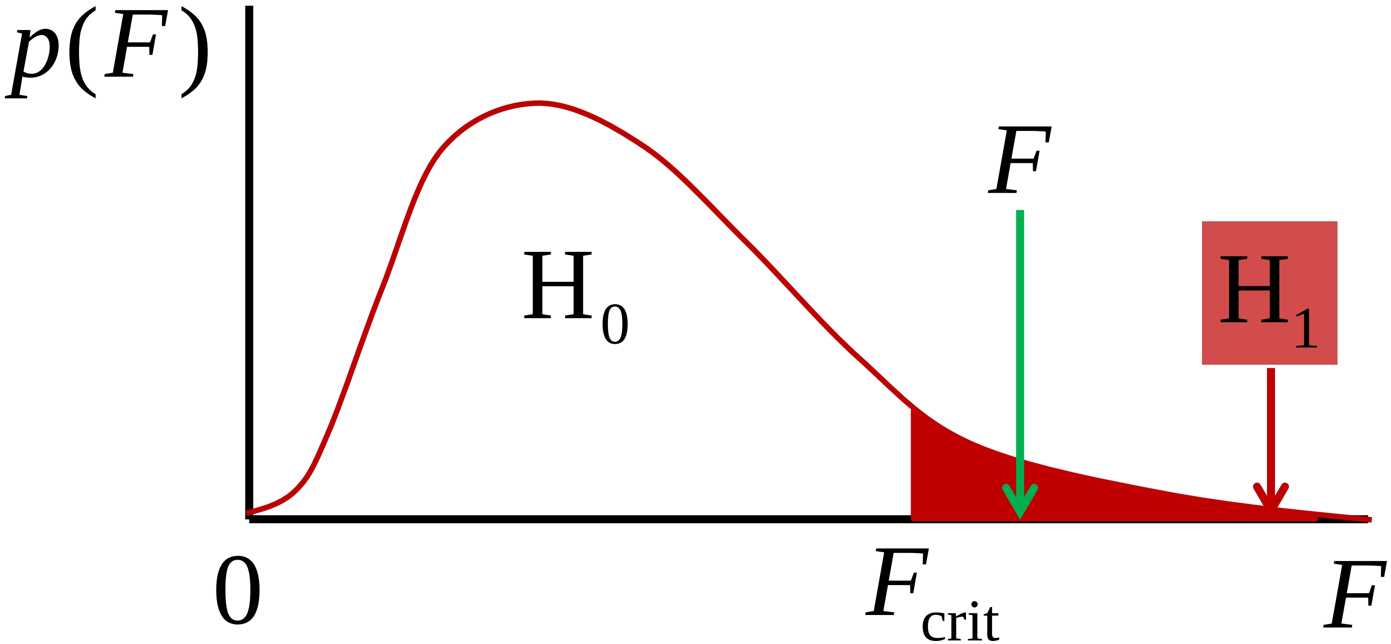


1-way ANOVA: toets

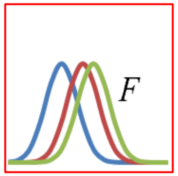


Significant verschil in gemiddelden (d.w.z. H_1 waar) als:

$$F > F_{\text{crit}}$$

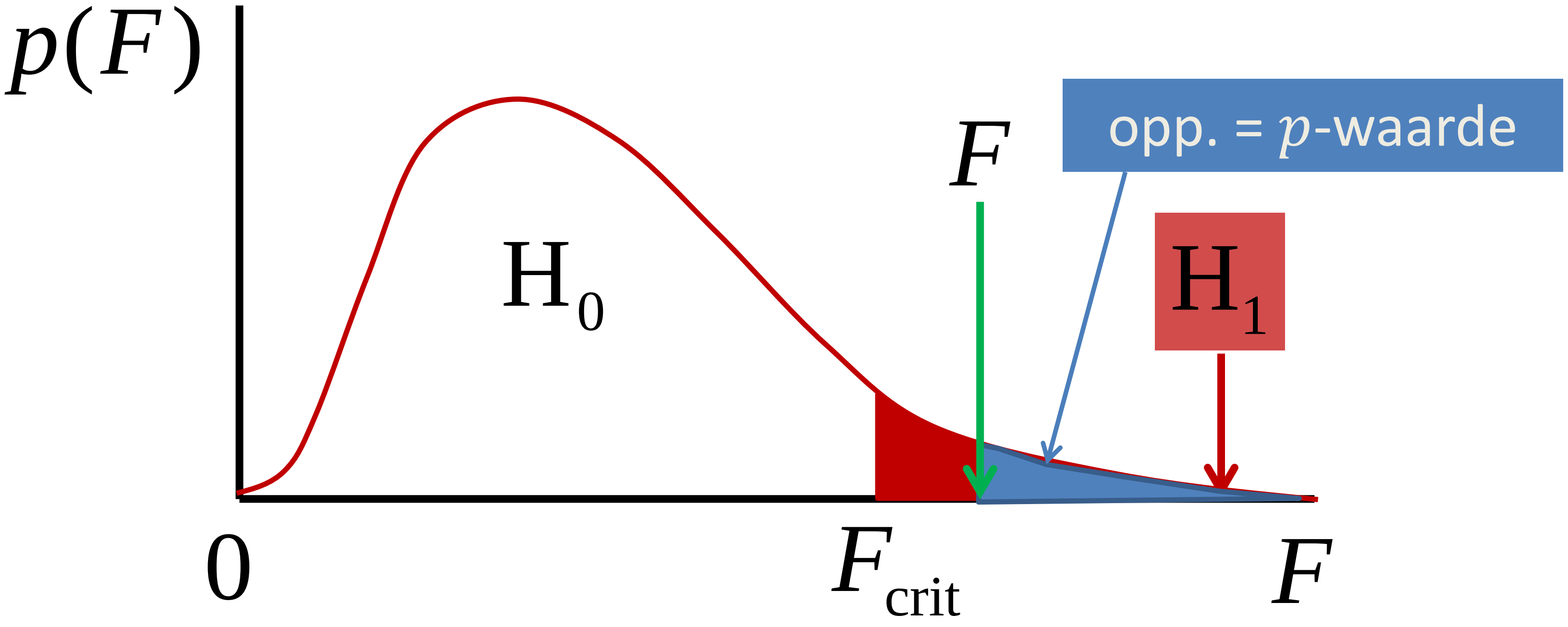


1-way ANOVA: toets



Significant verschil in gemiddelden (d.w.z. H_1 waar) als:

$$p < \alpha$$



p-waarden bij ANOVA

Significantie van factor A:

Zoals altijd in de statistiek:

- $p > \alpha$ geloof H_0
- $p < \alpha$ geloof H_1

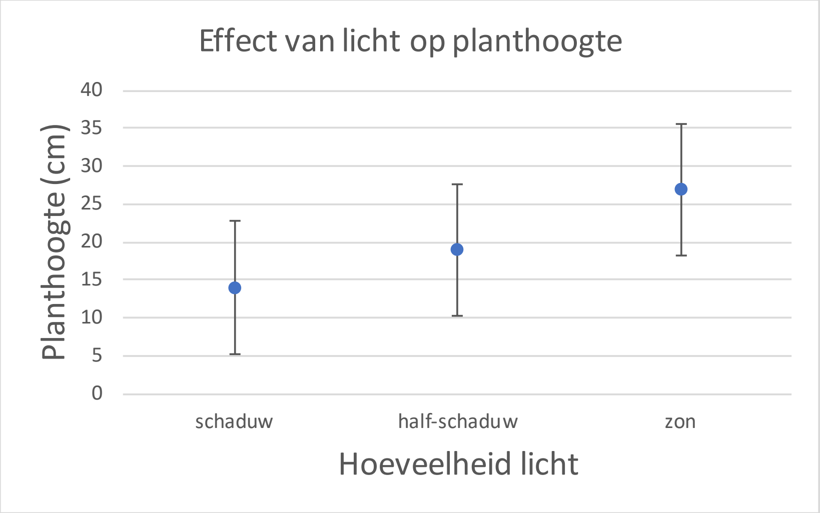
Bij de 1-way ANOVA geldt:

- H_0: géén verschil in gemiddelden, dus factor A heeft géén significant effect op y
- H_1: wel verschil in gemiddelden, dus factor A heeft wel significant effect op y

Voorbeeld: planthoogte

- Data:

Planthoogte (cm)		
Hoeveelheid licht		
schaduw	half-schaduw	zon
17.5	19.0	23.5
10.5	22.5	27.0
14.0	15.5	30.5



- 1-way ANOVA analyse:

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
schaduw	3	42	14	12.25		
half-schaduw	3	57	19	12.25		
zon	3	81	27	12.25		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	258	2	129	10.53061	0.0109	5.143253
Within Groups	73.5	6	12.25			
Total	331.5	8				

$p < 0.05$, dus H_1:
Licht heeft een significant effect op planthoogte

ANOVA: effectsterkte

- Sterkte verklarende effect van factor(en):
  - Regressie: R^2 (of R^2_{adj} bij model selectie)
  - ANOVA: eta-kwadraat = η^2

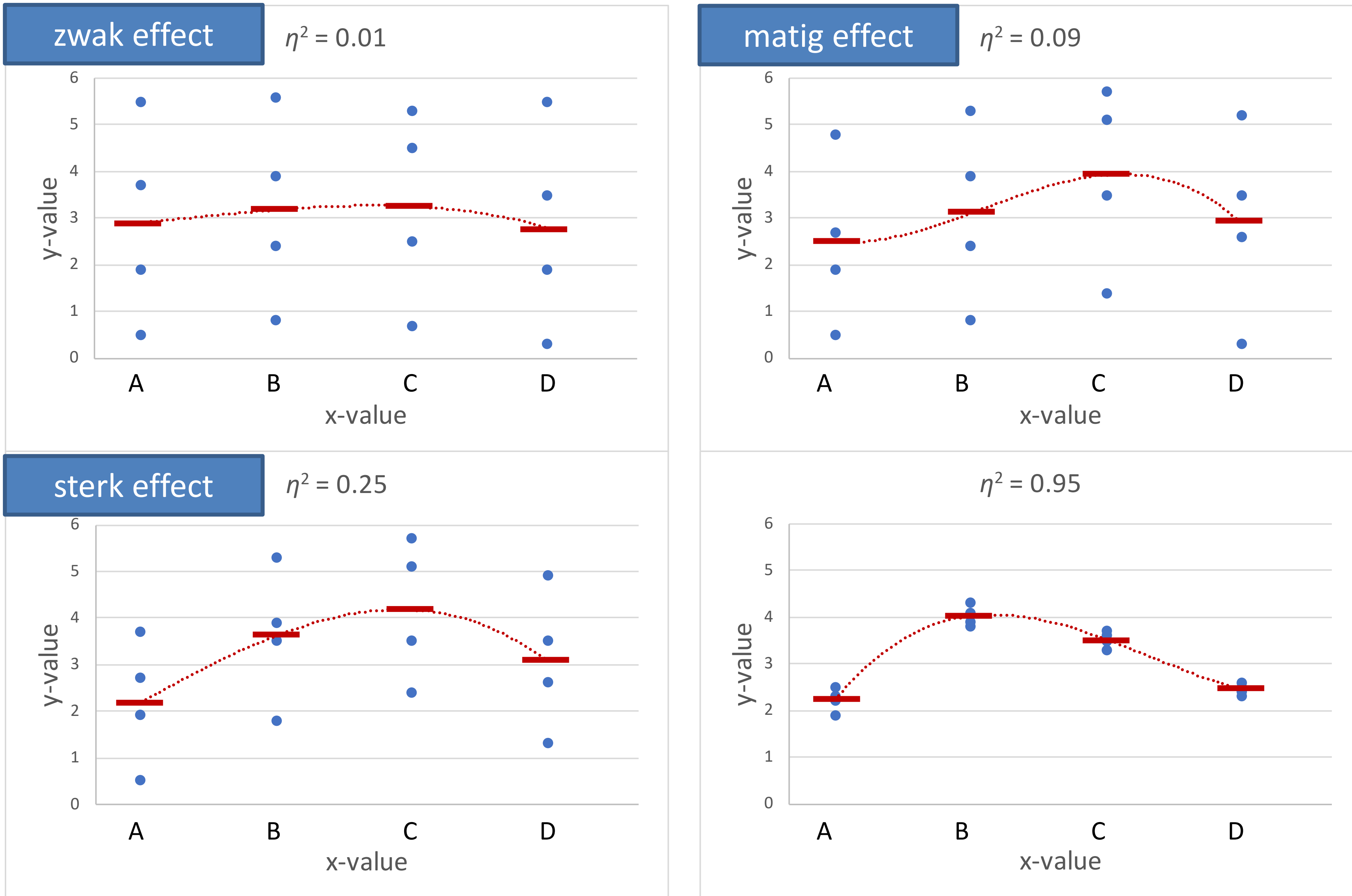
$$\eta^2 = \frac{SS_{\text{factor}}}{SS_{\text{tot}}}$$

Zelfde interpretatie als R^2 bij regressie: fractie door die factor/interactie verklaarde variantie

waarde η^2	betekenis *
0.01	zwak effect
0.09	gematigd effect
0.25	sterk effect

* Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.)

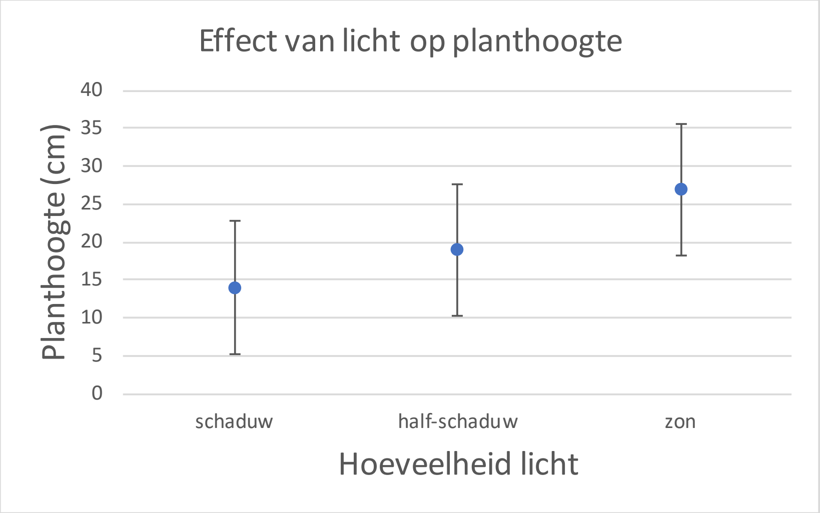
ANOVA: effectsterkte



Voorbeeld: planthoogte

- Data:

Planthoogte (cm)		
Hoeveelheid licht		
schaduw	half-schaduw	zon
17.5	19.0	23.5
10.5	22.5	27.0
14.0	15.5	30.5



- 1-way ANOVA analyse:

ANOVA							
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*eta.2*
Between Groups	258	2	129	10.53061	0.0109	5.143253	0.778
Within Groups	73.5	6	12.25				
Total	331.5	8					

$\eta^2 = \frac{258}{331.5} = 0.78$, dus 78% van alle variatie in planthoogte wordt "verklaard" door licht!

1-way ANOVA in R

- 1 vector met y-waarden, 1 factor g, via formule:

  summary(aov(y ~ g , ...))

- 1 vector met y-waarden, 1 factor g uit dataframe M:

  summary(aov(M$y ~ M$g , ...))

  summary(aov(y ~ g, data = M , ...))