Statistiek 2

Week 3: Toetsen met normale en t-verdeling

Toetsen Voorbeeld met normale verdeling met onbekende σ

Voorbeeld 9.11 Een visverwerker rapporteert dat zijn bedrijf dagelijks maximaal 200 VE's (Vervuilingseenheden) met het afvalwater loost. Vanuit de overheid wordt op 10 dagen het aantal VE's bepaald:

Dag	1	2	3	4	5	6	7	8	9	10
VE's	190	250	320	410	310	280	230	370	350	290

Deze gegevens lijken erop te wijzen dat de visverwerker vaak zijn limiet overschrijdt. Kan de overheid met deze gegevens aantonen dat het bedrijf teveel vervuilt, d.w.z. dat het bedrijf gemiddeld per dag meer dan 200 VE's loost?

Uitwerking 9.11 De aanname is dat de lozingen normaal verdeeld zijn met onbekende parameters μ en σ.

De geschatte gemiddelde waarde (steekproefgemiddelde) is $\bar{x} = \frac{190+250+\cdots+290}{10} = \frac{3000}{10} = 300$ VE's.

De steekproefvariantie is $s^2 = \frac{(190-300)^2+(250-300)^2+\cdots+(290-300)^2}{10-1} = \frac{40000}{9} = 4444{,}4$,
dus de steekproefstandaarddeviatie (geschatte standaarddeviatie) is $s = 66{,}67$ VE's.

Toetsen Voorbeeld met normale verdeling met onbekende σ

We hebben een normale verdeling aangenomen, maar σ is onbekend en moet worden geschat, bovendien is de steekproefgrootte ($n = 10$) kleiner dan 30, dus moeten we werken met de t-verdeling.

Het aantal vrijheidsgraden voor de t-verdeling is $n - 1 = 10 - 1 = 9$.
Als betrouwbaarheid kiezen we hier 99% (dus $\alpha = 0{,}01$). De t-waarde die hierbij hoort is

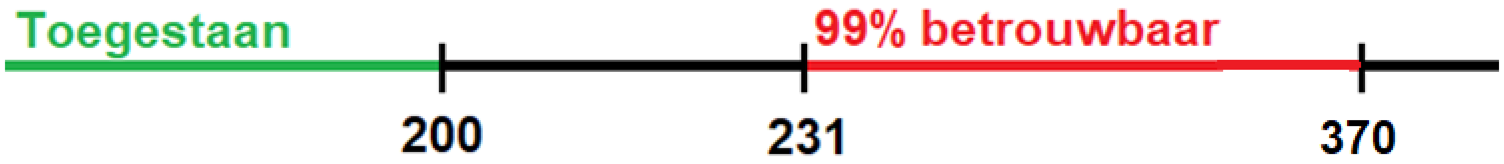
$$t = \mathrm{invT}(0.995\,, df = 9) = 3{,}2498.$$

Het 99% betrouwbaarheidsinterval voor de waarde van μ is dus

$$\left[\bar{x} - t\frac{s}{\sqrt{n}}, \bar{x} + t\frac{s}{\sqrt{n}}\right] = \left[300 - 3{,}2498\frac{66{,}67}{\sqrt{10}}, 300 + 3{,}2498\frac{66{,}67}{\sqrt{10}}\right] = [231{,}48\,, 369{,}53]$$

(Op sommige rekenmachines kun je met μ en σ in invT de grenzen direct uitrekenen (kansen 0,005 en 0,995))
Op basis van het steekproefgemiddelde kunnen we dus concluderen dat de werkelijke waarde van μ met 99% betrouwbaarheid in dit interval ligt.

De toegestane grenswaarde voor vervuiling is 200 en alle waarden in het betrouwbaarheidsinterval zijn te hoog, het gaat daarbij om minstens 99% van de mogelijkheden, dus kunnen we met 99% betrouwbaarheid concluderen dat het bedrijf teveel loost.



Toetsen **Voorbeeld met normale verdeling met onbekende σ**

Nu gaan we de probleemstelling op een andere manier formuleren, namelijk als een **hypothesetoets**.
De toets die we hiervoor formuleren is:

H$_0$: $\mu \leq 200$ (bedrijf heeft gelijk)
H$_1$: $\mu > 200$ (bedrijf heeft geen gelijk).

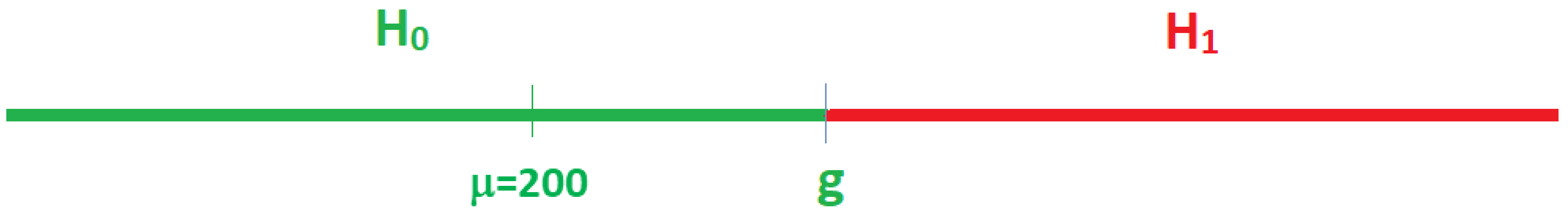
H$_0$ is de zogenaamde **nul-hypothese**, die zegt dat het bedrijf voldoet aan de milieueis.
H$_1$ heet de alternatieve hypothese, de overgebleven mogelijkheid, die zegt dat het bedrijf niet voldoet aan de milieueis.

De geschatte gemiddelde waarde $\bar{x} = 300$ doet vermoeden dat het bedrijf niet aan de milieueis van $\mu \leq 200$ voldoet, maar het is daarvoor niet voldoende om te zeggen dat $\bar{x} = 300 > 200$, dus het voldoet niet.
Het gaat namelijk maar om een steekproef en die kan toevallig hoog uitvallen. Het gaat om de nauwkeurigheid.

We gaan nu aantonen dat de kans dat we een steekproef van $\bar{x} = 300$ meten, terwijl het bedrijf toch aan de eis van $\mu \leq 200$ voldoet heel klein is en dat het dus onwaarschijnlijk is dat $\mu \leq 200$ geldt voor dit bedrijf.

Toetsen Voorbeeld met normale verdeling met onbekende σ

We gaan daarvoor een **kritieke waarde g** berekenen die groter is dan $\mu = 200$. Hierbij hoort een **acceptatiegebied $\overline{x} \leq g$** waarbinnen we H_0 accepteren als het gemeten steekproef-gemiddelde daarin valt, en een **kritiek gebied $\overline{x} > g$** waarin H_0 wordt verworpen en H_1 wordt geaccepteerd ($\overline{x} > g$) :



De kritieke waarde g kun je bepalen met de **fout van de eerste soort**, dat is de kans dat je H_0 ten onrechte verwerpt. In formule is dat: $\mathrm{P}(\overline{x} > g \mid H_0)$, ofwel $\mathrm{P}(\overline{x} > g \mid \mu \leq 200)$. Hierin betekent $\mid \mu \leq 200$: onder voorwaarde dat $\mu \leq 200$ (conditionele kans), m.a.w., je weet dat H_0 geldt.
We kiezen in dit geval $\alpha = 0{,}01$, d.w.z. een betrouwbaarheid van 99%, dus we moeten g zo bepalen dat

$$\mathrm{P}(\overline{x} > g? \mid \mu \leq 200) \leq \alpha = 0{,}01$$

Omdat we niet kunnen rekenen met een hele range aan μ- en g-waarden nemen we de worst case situatie:

$$\mathrm{P}(\overline{x} > g? \mid \mu = 200) = 0{,}01$$

Als $\underline{x}$ namelijk groter wordt dan g, of μ kleiner dan 200, dan is

$$\mathrm{P}(\overline{x} > g \mid \mu \leq 200) \leq \mathrm{P}(\overline{x} = g \mid \mu = 200) = 0{,}01$$

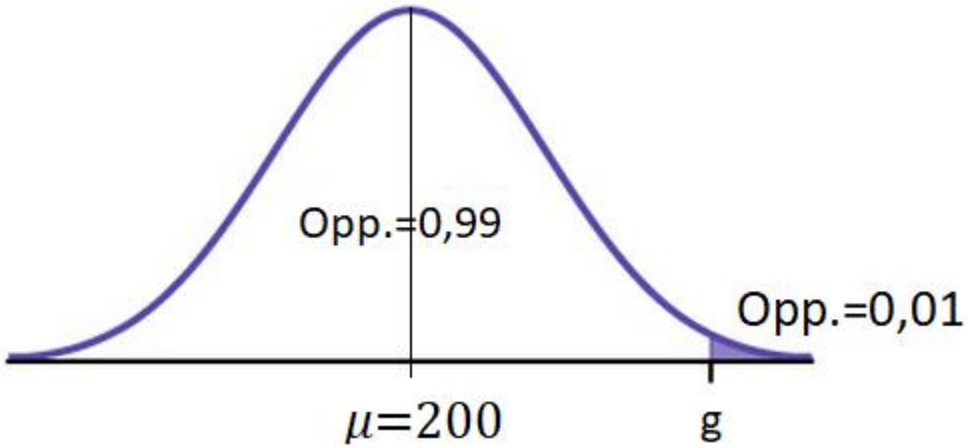
Toetsen Voorbeeld met normale verdeling met onbekende σ

We gaan dus g oplossen uit

$$\mathrm{P}(\overline{x} = g? \mid \mu = 200) = 0{,}01$$

$$\mathrm{tcdf}\left(t = \frac{g? - \mu}{\left(\frac{s}{\sqrt{n}}\right)}, 10^{10}, 10 - 1\right) = 0{,}01$$

$$\mathrm{tcdf}\left(\frac{g? - 200}{\left(\frac{66.67}{\sqrt{10}}\right)}, 10^{10}, 9\right) = 0{,}01$$



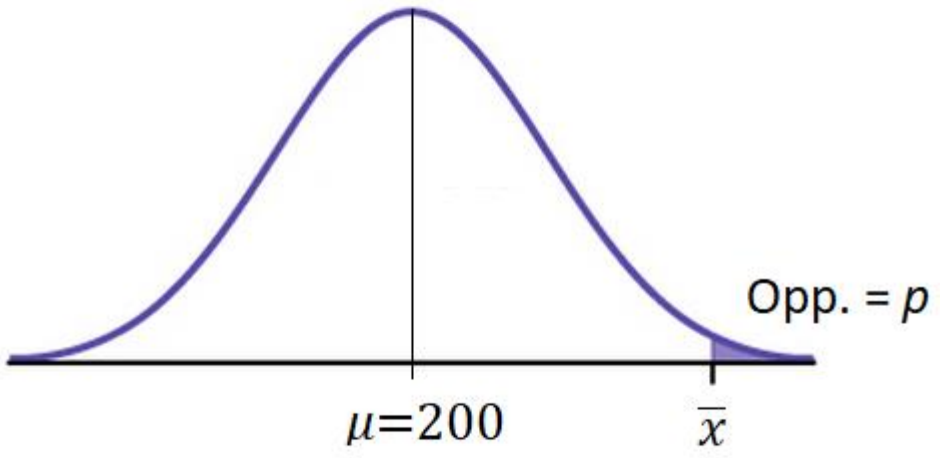
Met MATH -> SOLVER of Intersect vind je $g = 259{,}48$,

of gebruik $t = \mathrm{invT}(0.99\,, df = 9) = 2{,}8214$ en bereken $g = 200 + 2{,}82144 \cdot \frac{66.67}{\sqrt{10}} = 259{,}48$.

Het uitgerekende steekproefgemiddelde $\overline{x} = 300$ is groter dan de kritieke waarde $g = 259{,}48$, en ligt dus in het kritieke gebied, waardoor H_0 wordt verworpen.

Conclusie: Met een betrouwbaarheid van 99% kan worden gesteld dat de visverwerker op grond van de metingen ongelijk heeft en teveel vervuiling veroorzaakt.

Toetsen **Voorbeeld met normale verdeling met onbekende σ**



Een andere manier om de toets uit te voeren is door het **uitrekenen van de *p*-waarde.**

We berekenen dan niet eerst een kritiek gebied, maar direct de kans p op een fout van de eerste soort:

Ofwel, we berekenen de kans dat een steekproef $\underline{\bar{x}}$ minstens gelijk is aan de huidige meting $\underline{x} = 300$, terwijl H_0 geldt:

$$p = P(\underline{\bar{x}} \geq \bar{x} \mid H_0) = \text{tcdf}\left(t = \frac{300 - 200}{\left(\frac{66.67}{\sqrt{10}}\right)}, 10^{10}, 9\right) = 0{,}0005271$$

Deze kans (dat wij het fout hebben als we beweren dat H_1 geldt) is (veel) kleiner dan de toegelaten onbetrouwbaarheid $\alpha = 0{,}01$, dus we verwerpen H_0.

De conclusie is dus dat met een betrouwbaarheid van 99% kan worden gesteld dat de visverwerker op grond van de metingen ongelijk heeft en teveel vervuiling loost.
En eigenlijk kun je dit zelfs stellen met een betrouwbaarheid van 99,9%.

Toetsen Voorbeeld met normale verdeling en tweezijdige toetsing

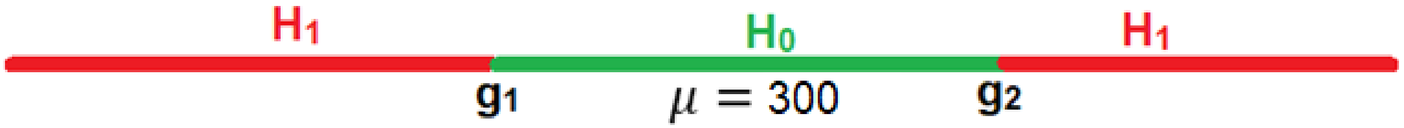
Voorbeeld 9.6 In een fabriek wordt tijdens productie 5 minuten gereserveerd voor montage van een bepaald onderdeel. Bekend is dat deze montagetijd normaal verdeeld is met $\mu = 300$ seconden en $\sigma = 15$ sec. Er wordt nieuwe apparatuur aangeschaft in de hoop dat de montageduur daardoor wordt verkort. Er worden 25 proefmontages gedaan, die een gemiddelde montagetijd van 292 sec. laten zien. Kan nu worden geconstateerd dat de montagetijd significant afwijkt van de vroegere situatie?
De gekozen nulhypothese is:

H_0: $\mu = 300$ (geen significante afwijking)
H_1: $\mu \neq 300$ (significante afwijking).

Er is hier gekozen voor **tweezijdige toetsing**, het gaat alleen om een afwijking, kleiner of groter maakt niet uit. De toetsingsvariabele is $\underline{\bar{x}}$, de gemiddelde tijd van de 25 proefmontages. Dit gemiddelde is normaal verdeeld met $\mu = 300$ seconden en $\sigma = \frac{15}{\sqrt{25}} = 3$ sec. We toetsen met betrouwbaarheid 98% (anders dan in het boek), dus linker- en rechteroverschrijdingskans zijn 0,01. Hierbij hoort een z-waarde (σ is bekend):

$$z = \mathrm{invNorm}(0.99) = 2{,}3263$$

Het acceptatie-interval is dus $[300 - 2{,}3263 * 3\,, 300 + 2{,}3263 * 3] = [293{,}0211\,, 306{,}9789]$ sec.
v



De gemeten waarde 292 sec ligt in het (tweedelige) kritieke gebied, buiten het acceptatie-interval.
H_0 wordt verworpen, H_1 geaccepteerd. **Conclusie**: De afwijking is significant bij betrouwbaarheidsniveau 98%.

Toetsen Voorbeeld met normale verdeling

Voorbeeld 9.6

Je kunt de toets ook weer uitvoeren door **berekening van de p-waarde**.

De gemeten montagetijd is $\overline{x} = 292$ sec. T.o.v. de eerdere waarde van 300 sec is deze waarde $300 - 292 = 8$ sec kleiner dan eerst. p is de kans dat de gemeten waarde minstens 8 sec. afwijkt, als de meting voldoet aan de normale verdeling met $\mu = 300$ sec en $\sigma = \frac{15}{\sqrt{25}} = 3$ sec, dus

$$p = P\left(\left|\underline{\overline{x}} - 300\right| > 8 \mid H_0\right) = 1 - \text{normalcdf}(300 - 8\,, 300 + 8\,, 300\,, 3) = 0{,}00766$$

Deze kans is kleiner dan de onbetrouwbaarheid $\alpha = 0{,}02$, dus H_0 wordt verworpen, H_1 geaccepteerd.

De conclusie is dat de afwijking is significant bij betrouwbaarheidsniveau 98%. Deze toets kijkt alleen of de waarde significant is veranderd, maar niet of het een verbetering is.

Toetsen Voorbeeld met normale verdeling

Je had in voorbeeld 9.6 in plaats van

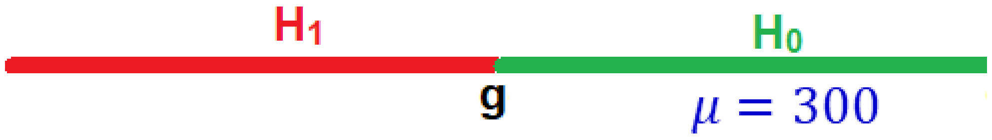
H_0: $\mu = 300$ (geen significante afwijking)
H_1: $\mu \neq 300$ (significante afwijking).

ook kunnen toetsen met

H_0: $\mu \geq 300$ (geen significante versnelling)
H_1: $\mu < 300$ (significante versnelling).

Er wordt dan **enkelzijdig getoetst**, want het gaat alleen om een verkleining.

Hierbij hoort een g-waarde van
$g = \text{invNorm}(0{,}02\,, \mu = 300, \sigma = 3) = 293{,}84$ sec.



Het acceptatie-interval is nu $[293{,}84\,, \infty]$ sec.
De gemeten waarde is 292 sec. Die ligt in het kritieke gebied. H_0 wordt verworpen, H_1 geaccepteerd.
Conclusie: Een significante versnelling van de montagetijd kan worden geconstateerd bij betrouwbaarheidsniveau 98%.

De conclusie is nu sterker door een herformulering van de hypothese (en ook eenvoudiger te berekenen).

Met de p-waarde: $p = \text{normalcdf}(-10^{10}, 292\,, 300\,, 3) = 0{,}00383$ is kleiner dan 0,02, dus H_0 verwerpen.

Toetsen Voorbeeld met normale verdeling

Je weet nu alleen vrij dat er een versnelling heeft plaatsgevonden, maar niet hoe groot die is. Je kunt de hypothese herformuleren om de uitspraak iets sterker te maken:

H_0: $\mu \geq 298{,}5$ (geen significante versnelling)
H_1: $\mu < 298{,}5$ (significante versnelling).

Bij deze enkelzijdige toets hoort een g-waarde van
$g = \text{invNorm}(0.02\,,\ \mu = 298.5\,, \sigma = 3) = 292{,}34$ sec.
Het acceptatie-interval is dus $[292{,}34\,, \infty]$ sec.

De gemeten waarde is 292 sec. Die ligt (net) buiten het acceptatie-interval.
H_0 wordt verworpen, H_1 geaccepteerd.
Conclusie: Met betrouwbaarheid 98% kan worden geconstateerd dat er minimaal 1,5 seconde versnelling optreedt ten opzichte van de eerdere montagetijd.

(Met de p-waarde: $p = \text{normalcdf}(-10^{10}, 292\,, 298.5\,, 3) = 0{,}01513$ is kleiner dan 0,02, dus H_0 verwerpen.)

De laatste conclusie is sterker dan de voorgaande twee. Dit laat zien dat de precieze formulering van de toets belangrijk is voor de conclusie die je kunt gaan trekken.

Toetsen Voorbeeld met normale verdeling

Kunnen we de conclusie nog sterker maken?

H_0: $\mu \geq 298$ (geen significante versnelling)
H_1: $\mu < 298$ (significante versnelling).

De p-waarde is nu:

$$p = \text{normalcdf}(-10^{10}, 292\,, 298\,, 3) = 0{,}02275$$

Deze waarde is niet kleiner dan 0,02, dus nu kunnen we H_0 niet verwerpen.

Conclusie: Bij een betrouwbaarheidsniveau van 98% en een steekproefgemiddelde van 292 kunnen we niet concluderen dat er gemiddeld een versnelling van 2 seconden gehaald kan worden. Ofwel: er kan op deze manier niet worden aangetoond dat een versnelling van 2 seconden gehaald wordt.

Om dit te concluderen zou je een lagere waarde van het steekproefgemiddelde nodig hebben, of dezelfde waarde maar bij een grotere steekproefomvang.

De toets

H_0: $\mu \geq 300$ (geen significante versnelling)
H_1: $\mu < 300$ (significante versnelling).

wordt in de praktijk vaak geformuleerd als

H_0: $\mu = 300$ (geen significante versnelling)
H_1: $\mu < 300$ (significante versnelling).

Formeel klopt dit niet omdat H_1 nu niet de ontkenning van H_0 is, maar uit de formulering is wel duidelijk wat er eigenlijk wordt bedoeld.

Op het tentamen moet je een toets op twee manieren kunnen uitrekenen, door het bepalen van een kritiek gebied en door het berekenen van een overschrijdingskans.

Wat je eigenlijk wilt aantonen ("Het proces wordt versneld") wordt normaal gesproken als H_1 geformuleerd en niet als

H_0: $\mu < 300$ (significante versnelling).

H_1: $\mu \geq 300$ (geen significante versnelling)

omdat "verwerpen van H_0" sterker is dat het "niet kunnen verwerpen van H_0" (op deze manier), terwijl dat op een andere manier misschien wel zou kunnen.