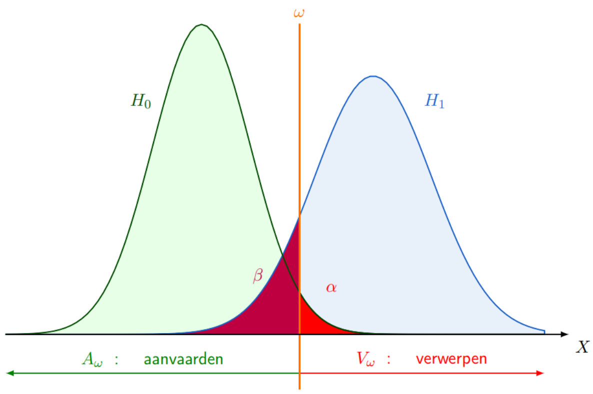
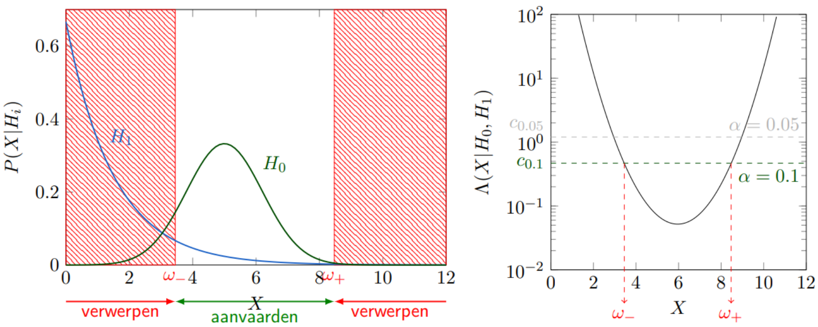
H8: beslissingen nemen

8.1 testen van hypothesen

beslissingen nemen	Beschrijvende statistiek vat experiment samen met minimum verlies aan info > we willen de experimentele data gebruiken om een beslissing te maken > dit is een hypothese of theorie die al dan niet in overeenstemming is met experiment kwantitatief: hypothese/theorie opstellen > testen > aanvaarden of verwerpen We kunnen ook meerdere overlappende en/of tegenstrijdige hypothesen opstellen
8.1.1 de nulhypothese	
nulhypothese H_0	= hypothese geformuleerd waarbij we ervan uitgaan van een situatie waarbij er geen effect, geen verschil,… tussen theorie en experiment verwacht wordt ie: een veronderstelling wanneer men een bepaald effect, verschil of relatie onderzoekt > De nulhypothese stelt dat er geen significant verschil, geen effect of geen relatie bestaat in de populatie waaruit de steekproef is genomen
alternatieve hypothese H_1	= hypothese die overwogen wordt als tegenhanger van de nulhypothese ie: er is wel een effect/verschil/relatie
eenvoudige hypothese	een hypothese is eenvoudig als ze één bewering bevat
samengestelde hypothese	een hypothese is samengesteld als ze meerdere beweringen omvat
8.1.2 types van fouten	
teststatistiek X	= één variabele X als een functie van onze waarnemingen > wssheidsverdelingen van X onder de nulhypothese H_0 en een alternatieve hypothese H_1 worden beschreven door $P(X\|H_0)$ en $P(X\|H_1)$
beslissingsregel ω	= waarde die gebied van X in twee verdeeld > in een *aanvaard gebied* A_ω en *verworpen gebied* V_ω 
types fouten	bij het maken van een beslissing kan men echter een fout maken > fout type I of fout type II
def: type I fout	= onterecht verwerpen van de nulhypothese H_0 > hoe vaak een type I fout gebeurt wordt beschreven door de *significantie* α vd test $$\int_{V_\omega} P(X\|H_0)dX = \alpha$$
significantie van een test	Bij een goede test moet α klein zijn (minder dan 5% of 1%) > hoe kleiner α, hoe groter de significantie van de test > een test is significant op het α niveau als de geïntegreerde wssheid om een correcte hypothese te verwerpen kleiner of gelijk is aan: $$\int_{V_\omega} P(X\|H_0)dX \leq \alpha$$ Merk op dat meer significante tests de minder significante omvatten. > Dus een test die significant is op het 1% niveau is ook significant op het 5% niveau (vermits 1 ≤ 5).

<table>
<tr><td>def: type II fout</td><td>= aanvaarden van de nulhypothese H_0 terwijl de alternatieve hypothese H_1 de waarheid geeft:
$$\int_{A_\omega} P(X|H_1)dX = \beta$$
> de power van de test 1-β geeft hoe vaak een type II fout niet gebeurt
<table>
<tr><td>Beslissing o.b.v. steekproef</td><td colspan="2">Waarheid</td></tr>
<tr><td></td><td>H_0</td><td>H_1</td></tr>
<tr><td>H_0 aanvaarden</td><td>✓
$1-\alpha$</td><td>type-II fout
β</td></tr>
<tr><td>H_0 verwerpen</td><td>type-I fout
α</td><td>✓
$1-\beta$</td></tr>
</table></td></tr>
<tr><td>goede test</td><td>= één waarvoor zowel α als β klein zijn
ie: kleine significantie en grote power
> hiervoor moeten $P(X|H_0)$ en $P(X|H_1)$ sterk verschillend zijn</td></tr>
<tr><td colspan="2">8.1.3 Neyman-Pearson test</td></tr>
<tr><td>Neyman-Pearson test</td><td>= een test waarvoor de H_0 en H_1 eenvoudig zijn
> zorgt voor kleine α en β</td></tr>
<tr><td>Verwerpingsgebied kiezen</td><td>ipv één ω te kiezen, kunnen we verschillende verwerpingsgebieden opstellen
> de definities van significantie en power:
$$\begin{cases} \int_{V_\alpha} P(X|H_0)dX = \alpha \\ \int_{V_\alpha} P(X|H_1)dX = 1-\beta \end{cases}$$
om β klein te maken voor een bepaalde α, moet het optimale verwerpingsgebied V gekozen worden dat de tweede integraal zo groot mogelijk is
> deze kan herschreven worden als:
$$\int_{V_\alpha} \frac{P(X|H_1)}{P(X|H_0)} P(X|H_0)dX = 1-\beta$$
is maximaal indien het gebied V_α de waarden X omvat die de grootste waarden hebben voor de likelihood ratio: $\frac{P(X|H_1)}{P(X|H_0)}$
het optimale verwerpingsgebied omvat alle waarden voor X die:
$$\Lambda(X|H_0,H_1) \equiv \frac{P(X|H_1)}{P(X|H_0)} \geq c_\alpha$$
waarvoor c_α gekozen wordt zodat $\int_{V_\omega} P(X|H_0)dX = \alpha$ voldoende is voor α

</td></tr>
<tr><td>st: verwerpen van H_0</td><td>voor een bepaalde waarde x vd teststatistiek X wordt H_0 met sign. α verworpen indien:
$$\Lambda(x) \geq c_\alpha.$$</td></tr>
</table>

8.2 een test interpreteren	
8.2.1 bewijzen is enkel verwerpen	
tegendeel verwerpen	Stel je test een ESP via 100 tests > je verwacht 10 correcte antwoorden door puur gokwerk > een testpersoon haalt 99 op 100 correcte antwoorden foute conclusie: resultaten zijn consistent met bestaan van ESP juiste conclusie: resultaten zijn niet consistent met hypothese van puur toeval ie: je kan enkel hypotheses verwerpen
st: tegendeel formuleren	Als je het bestaan van iets wil aantonen > dan moet je de hypothese van het tegendeel formuleren nl: dat er niet zo'n effect bestaat > deze is de nulhypothese H_0
hypotheses bewijzen	In de statistiek kan je nooit op een betekenisvolle manier hypotheses bewijzen > je kan enkel hypotheses verwerpen
8.2.2 aanvaarden is nooit bewijzen	
bewijzen van een stelling	Als de test van H_0 slaagt > dan heb je NIET bewezen dat de nulhypothese vals is dan heb je bewezen dat ze niet correct is > nl: het kan zijn dat het effect bestaat, maar te klein om in je experiment waar te nemen
st: effect bewijzen	Je kan nooit bewijzen dat een effect er niet is > het beste dat je kan doen is een limiet zetten op een bepaald confidentieniveau
st: significant resultaat	een resultaat is significant als de wssheid dat het toevallig voorkomt onder de nulhypothese klein is
8.2.3 beslissing vs confidentieniveaus	
verschil schatting-confidentie-beslissing	schattingstheorie: 'Wat is de waarde vd parameter θ die het meest met mijn meting overeenkomt?' confidentie: 'Wat is het interval van waarden $[\theta_-,\theta_+]$ van een parameter θ dat ik met mijn data kan bepalen, zodat de echte waarde θ_0 vd parameter met een confidentieniveau 1-α bevat?' > conclusie: 'Dankzij mijn metingen heb ik (1-α).100% vertrouwen dat de echte waarde vd parameter binnen het confidentie-interval ligt' > probleem: we kunnen θ_0 nooit echt weten > opl: 'stel als nulhypothese H_0 dat de waarde vd parameter θ_0 is > komt dan mijn meting met een significantie α overeen met H_0?'

a) tests voor het gemiddelde van een normale verdeling

vb: gebruik van significantie

stel we willen gemm. ve steekproefgemiddelde tegen een verwachte waarde μ_0 testen
> drie mogelijke tests:

(a) (Eenvoudige) tweezijdig – $H_0 : \mu = \mu_0$,

(b) Eenzijdig rechts – $H_0 : \mu \leq \mu_0$,

(c) Eenzijdig links – $H_0 : \mu \geq \mu_0$.

<u>indien standaardafwijking gekend is:</u>

dan vinden we: (zie paragraaf 7.4)

$$\bar{X}_N \sim \mathcal{N}\left(\mu; \frac{\sigma}{\sqrt{N}}\right)$$

die we gebruiken om 3 uitkomsten te vinden:

(a) $H_0 : \mu = \mu_0$ verwerpen met een significantie α indien :

$$p\left(Z > \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{N}}\right) < \alpha$$

(b) $H_0 : \mu \leq \mu_0$ verwerpen met een significantie α indien :

$$\mathcal{P}_{H_0}(\bar{X}_N > \bar{x}_N) = p\left(Z > \frac{\bar{x}_N - \mu_0}{\sigma/\sqrt{N}}\right) < \alpha$$

(c) $H_0 : \mu \geq \mu_0$ verwerpen met een significantie α indien :

$$\mathcal{P}_{H_0}(\bar{X}_N < \bar{x}_N) = p\left(Z < \frac{\bar{x}_N - \mu_0}{\sigma/\sqrt{N}}\right) < \alpha$$

<u>indien de standaardafwijking niet gekend is</u>

> resolutie s_N moet geschat worden
> Gauss verdeling is vervangen door Student t-verdeling met N-1 vrijheidsgraden
> drie tests worden:

(a) $H_0 : \mu = \mu_0$ verwerpen met een significantie α indien :

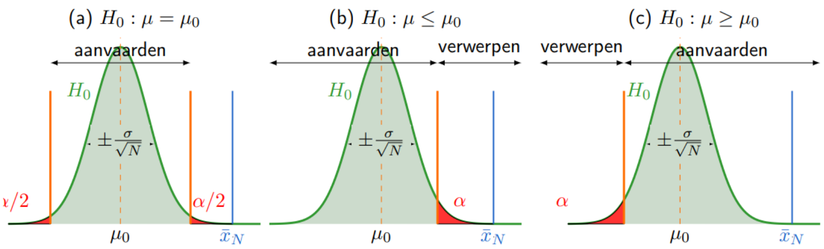
$$p\left(T > \frac{|\bar{x} - \mu_0|}{s_N/\sqrt{N}}\right) < \alpha$$

(b) $H_0 : \mu \leq \mu_0$ verwerpen met een significantie α indien :

$$\mathcal{P}_{H_0}(\bar{X}_N > \bar{x}_N) = p\left(T > \frac{\bar{x}_N - \mu_0}{s_N/\sqrt{N}}\right) < \alpha$$

(c) $H_0 : \mu \geq \mu_0$ verwerpen met een significantie α indien :

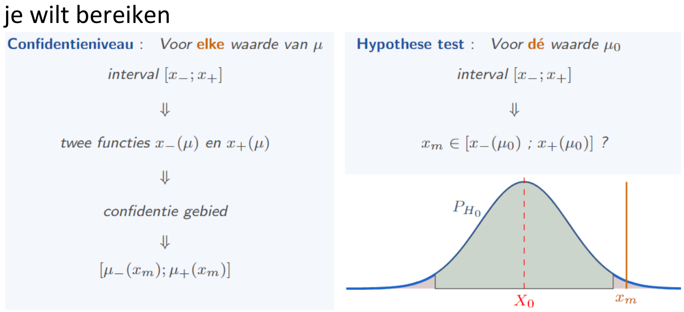
$$\mathcal{P}_{H_0}(\bar{X}_N < \bar{x}_N) = p\left(T < \frac{\bar{x}_N - \mu_0}{s_N/\sqrt{N}}\right) < \alpha$$



8.2.4 het bepalen van steekproefgrootte

steekproefgrootte bepalen

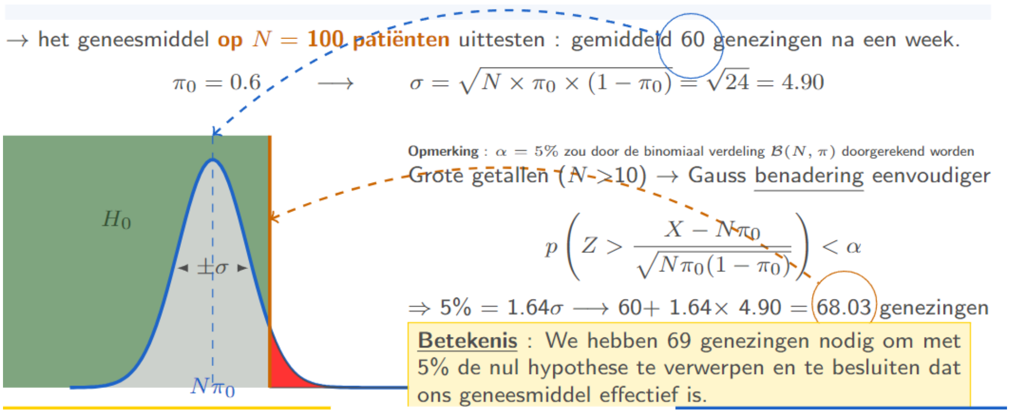
steekproefgegevens verzamelen is duur en tijdrovend
> bepaal op voorhand het minimum aantal benodigde gegevens
> afh. van de significantie die je wilt bereiken



<table>
<tr><td colspan="2">voor rest van paragraaf 8.2.4 bekijken we de binomiale statistiek met een voorbeeld
nl: nieuw geneesmiddel tegen verkoudheid die getest moet worden
> belangrijk: evenwichtige representatieve steekproefpopulatie</td></tr>
<tr><td colspan="2"><u>a) is een behandeling effectief</u></td></tr>
<tr><td>effectiviteit vd behandeling</td><td>p_0=0.60 zal spontaan genezen na een week zonder het geneesmiddel
> en we willen bvb een significantie α=0.05

Stel dat de wssheid dat een patiënt genezen is na een week gelijk is aan P
> de nulhypothese die we willen testen is de wssheid voor spontane genezing:
$$H_0 : P \leq p_0$$
als we testen op N=100 verwachten we via de hypothese dat $p_0N=N_0$=60 genezen
> de standaardafwijking vd binomiale verdeling is gegeven door
$$\sigma = \sqrt{N\, p_0(1-p_0)} = \sqrt{24} = 4.90$$
Om de 5% significantie limiet te bepalen kunnen we de binomiale verdeling doorrekenen
> betere oplossing: gebruik de Gauss benadering
> voor grote getallen verschilt de binomiale verd. weinig van een Gauss
> voor Gauss is 5% limiet dan 1.64σ
ie: 60+1.64 x 4.9 = 68.03 genezingen
> we hebben 69 genezingen nodig om met 5% significantie de nulhypothese te verwerpen
> waaruit we dan besluiten dat ons geneesmiddel effectief is

We kunnen echter N arbitrair kiezen
> stel dat we een 100% zekere behandeling willen
> dan zullen op N patiënten alle N genezen
> de wssheid dat dit gebeurt onder de nulhypothese is:
$$P(X=N; p_0 = 0.6, N) = 0.6^N$$
voor een 5% significantie eisen we dus dat:
$$0.6^N \leq 0.05$$
hetgeen ons N=6 geeft
> als we dus testen op 6 personen en alle 6 genezen dan kunnen we met 5% significantie zeggen dat het medicijn werkt
> 6 personen testen is veel gemakkelijker dan 100 personen testen


</td></tr>
</table>

b) hypothese op de verwachte efficiëntie

hypothese verwachte efficiëntie

stel dat we een genezingskans van 70% willen (ipv 60% daarnet)
> en we willen 2 mogelijkheden onderzoeken:

- nulhypothese dat het middel niets doet
  > maar door toevallige fluctuaties het aantal genezingen in de steekproef consistent is met p_1=0.70
- middel werkt wel degelijk
  > maar door toevallige fluctuaties het aantal genezingen kleiner is dan 70%

> we willen een significantie van 5% en een 99% power ie: α=0.05 β = 0.01

1: aantal genezingen N_g vanaf waar we met 5% significantie de nulhypothese verwerpen is:

$$p_0 N + 1.64\sqrt{N p_0 (1-p_0)} \leq N_g$$

2: als genezingskans werkelijk p_1=0.70 of meer is
> dan is de tweede conditie dat het aantal genezingen met 1-β=0.99 wssheid groter moet zijn dan de 1% CL onderlimiet
> onderlimiet correspondeert met 2.33σ voor een Gauss
> alternatieve hypothese met 99% power verworpen voor:

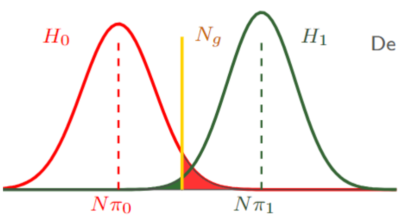
$$N_g \leq p_1 N - 2.33\sqrt{N p_1 (1-p_1)}$$

de twee samen geeft:

$$p_0 N + 1.64\sqrt{N p_0 (1-p_0)} \leq p_1 N - 2.33\sqrt{N p_1 (1-p_1)}$$

wat kan opgelost worden als N≥351, dan is N_g=266

conclusie: als we 351 patiënten het geneesmiddel toedienen en eisen dat er na een week 266 of meer genezen zijn
> dan is de kans dat we effectief een geneesmiddel missen 1%
\het middel toch niet effectief is 5%



De twee ongelijkheden samen geven :

$$0.6N + 1.64\sqrt{0.24N} \leq 0.7N - 2.33\sqrt{0.21N}$$

$$\Downarrow$$

$$\boxed{N \geq 351} \qquad (N_g = 226)$$

8.2.5 is er een signaal?

Fysische processen zijn vaker Poisson-verdeeld
> we willen weten of een piek significant is
ie: komt het overeen met ruis/toevallige fluctuatie of niet?

vb: radioactief verval

Meet de γ-straling afkomstig van radioactief verval van een isotoop
> we verwachten 54 detecties van achtergrond en meten in werkelijkheid 87 detecties
> hebben we een bewijs voor het signaal?

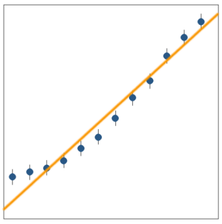
H_0: wat is de wssheid voor een dergelijke fluctuatie van de achtergrond

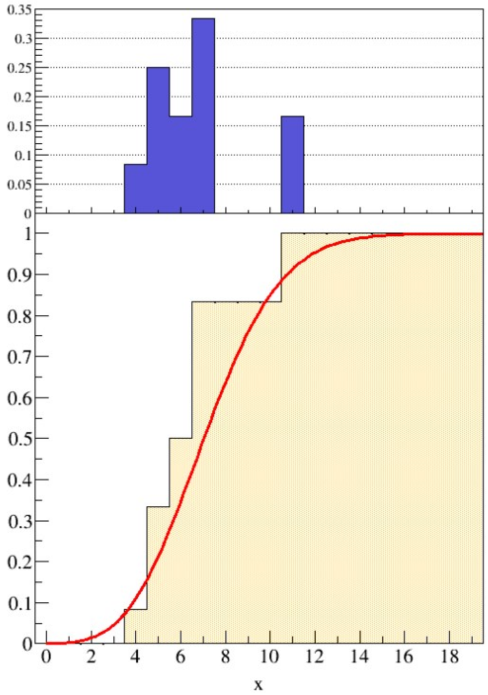
Voor een Poissonverdeling vinden we $\sigma = \sqrt{54} = 7.35$ tellen
> aantal tellen is groot genoeg om Gauss benadering te gebruiken
> de afwijking die we waarnemen is:

$$\frac{(87-54)}{7.35} = 4.5\sigma$$

>> we mogen aannemen dat er een signaal is

vb: radioactief verval	Stel nu dat het berekenen van de gemiddelde achtergrond niet volledig zeker is ie: er is een fout op de gemiddelde achtergrond, bvb +-3 > voeg deze bij de Poisson fout die we hebben op de waargenomen tellen > om de opp. onder de piek te bepalen moet men 2 getallen aftrekken > de variantie op het resultaat is dan: $54+3^2 = 63$ > effect is nu maar 4.2σ is Stel daarbovenop dat we de energie van de γ-straling niet kennen > we weten niet waar de piek zich zou moeten bevinden bvb: je meet energie met resolutie 3keV en zoekt een piek in gebied van 1MeV breed > verm. kans op voorkomen van fluctuaties met aantal energiebins > verhoogt de wssheid voor het voorkomen ve fluctuatie met factor 333 > confidentieniveau voor nieuwe ontdekking verkleint: $$\mathrm{CL}_{\text{ontdekking}} = 1 - \frac{3}{1000} \times (1 - \mathrm{CL})$$
8.4 Goodness of fit	
We bestuderen de nulhypothese H_0 met teststatistiek X in een verwerpingsgebied V_ω op significantie niveau α > alternatieve hypothese H_1 omvat alle andere mogelijke hypothesen > kan nooit uitgedrukt worden > type II fout en power van de test 1-β blijven ongedefinieerd	
Goodness of fit GOF	= Test om vast te stellen of de ware verdeling van waarnemingen tot een bepaalde familie van theoretische verdelingen behoort
8.3.1 van teststatistiek naar P-waarde	
def: P-waarde vd test	Indien H_0 correct is, dan verwacht men met wssheid P gegevens die minstens zover van H_0 zullen liggen als de gemeten gegevens > P is de *P-waarde van de test* voor de combinatie van steekproef en nulhypothese
P-waarde	stel we hebben X discrete waarnemingen en teststatistiek ω=ω(X) als functie van X > we hebben een waarde $\omega_0=\omega(X_0)$ voor de dataset X_0 > de P-waarde wordt dan gegeven door: $$P_X = \sum_{X:\omega>\omega_0} P(X\|H_0)$$ indien X continu: $$P_X = \int_{X:\omega>\omega_0} P(X\|H_0)$$ >> lastig te berekenen, dus wordt in praktijk niet vaak gebruikt
8.3.2 χ²-test	
werking χ²-test	stel we hebben een verzameling metingen (x,y) met elke y een fout σ > we hebben een functie f(x) die voor elke x de ideale waarde y geeft > definieer de χ² waarde als: $$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - f(x_i)}{\sigma_i}\right]^2$$ als de y-metingen gecorreleerd zijn, dan moeten we de volledige covariantiematrix gebruiken: $$\chi^2 = (\tilde{\mathbf{y}} - \tilde{\mathbf{f}})\mathbf{V}^{-1}(\mathbf{y} - \mathbf{f})$$ met **V** de covariantiematrix voor de y_i metingen > als f(x) de gegevens goed beschrijft, dan zal het verschil tss de metingen en echte waarden 1σ bedragen > waarde van χ² is dan van orde N De wssheidsverdeling voor χ² is gegeven door de vgl: $$P(\chi^2; N) = \frac{2^{-N/2}}{\Gamma(N/2)} \chi^{2(\frac{N}{2}-1)} e^{-\chi^2/2}$$ om te bepalen hoe goed de overeenkomst is tss theorie en experiment wordt de P-waarde geëvalueerd: $$p_{\chi^2} = \int_{\chi^2}^{\infty} P(\chi^2_*; N) d\chi^2_*$$

<u>a) interpretatie</u>	
kwaliteit van een fit	1: χ^2 groot > p_{χ^2} is klein, aka slechte fit > fouten werden onderschat / f(x) beschrijft gegevens niet goed 2: χ^2 klein > p_{χ^2} groot, aka goede fit 3: χ^2 te klein > fouten werden overschat of gegevens werden speciaal geselecteerd
<u>b) na een fit</u>	
st: opletten bij fit	Als de punten gebruikt zijn om de functie te vinden (bvb via kleinste kwadraten) > dan zal χ^2 kleiner zijn dan verwacht omdat je die zo klein mogelijk hebt gemaakt
> effect van deze stelling	We kunnen dit effect in rekening brengen nl: voor N termen in de χ^2som en je hebt m parameters aangepast om de som te minimaliseren > dan is de relevante verdeling nog altijd de χ^2 verdeling met n=N-m vrijheidsgraden: P(χ^2;n)
> voorbeeld	Een fit van een rechte lijn doorheen 20 punten geeft een χ^2 van 36.3. Een parabool geeft een χ^2 van 20.1 en een kubische fit een χ^2 van 17.6. • De rechte lijn heeft 20-2=18 vrijheidsgraden. De tabel geeft dan een waarschijnlijkheid om meer dan 34.8 te vinden als kleiner dan 1%, dus de waarschijnlijkheid om 36.3 te vinden is nog kleiner. Het is hoogst onwaarschijnlijk dat een dergelijk grote waarde bij toeval ontstaat en dus zeer onwaarschijnlijk dat de gegevens door een rechte lijn beschreven worden. • De χ^2 voor de parabool ligt braaf beneden de 10% drempel voor 17 vrijheidsgraden. De gegevens worden goed beschreven door een parabool en er is geen reden om een ingewikkelder functie als een derdegraadspolynoom te gebruiken.
toepassing: histogram	histogram ve steekproef van waarden > y_i is het aantal waarden in bin i, i=1,…,n met centrale waarde x_i > $f(x_i)$ is voorspeld aantal evenementen in elk interval > fouten gegeven door Poisson statistiek en χ^2 is: $$\chi^2 = \sum_j^{N_b} \frac{(n_j - f_j)^2}{f_j}$$ het aantal vrijheidsgraden is nu aantal intervallen min aantal gefitte parameters >> gevolg: - deze methode enkel gebruiken als de grootte van de bins vastligt, niet subjectief - er moeten minstens 5 elementen in elke bin zitten voor bruikbare χ^2
<u>c) limiet naar grote n</u>	
limiet naar grote n	voor grote n nadert $\sqrt{2\chi^2}$ naar een Gauss met gemiddelde $\sqrt{2n-1}$ en standaard afwijking 1. vb: Als je het 5% niveau voor 30 vrijheidsgraden wil vinden > dan haal je uit de tabel voor Gaussverdelingen dat de 95% limiet op 1.645 σ ligt. > de χ^2 limiet wordt dan $$\sqrt{2\chi^2} = \sqrt{2 \times 30 - 1} + 1.645 = 43.39$$ wat dichtbij de werkelijke waarde 43.77 ligt
8.3.3 run test	
voorbeeld runs	bekijk de figuur met een gefitte rechte met een redelijke χ^2 > toch zien we dat de fit niet echt goed is en dat een hogere-graad polynoom beter zou zijn bekijk elk punt en bepaal of deze boven (B) of onder (O) de rechte ligt: BBBOOOOOOBBB > geeft 2 runs van B's en 1 run van O's > in betere fits zouden punten constant boven en onder de grafiek moeten springen ie: er zouden veel meer runs moeten zijn 

st: kans op runs	Stel er zijn N_B punten boven en N_O punten onder de lijn, dus $N_B+N_O=N$ > de wssheid om r runs te hebben, met r even: $$P_r = 2\frac{C^{N_B-1}_{\frac{r}{2}-1} \times C^{N_O-1}_{\frac{r}{2}-1}}{C^{N}_{N_B}}$$ Indien r oneven is wordt de waarschijnlijkheid: $$P_r = \frac{C^{N_B-1}_{r-\frac{3}{2}} \times C^{N_O-1}_{r-\frac{1}{2}} + C^{N_O-1}_{r-\frac{3}{2}} \times C^{N_B-1}_{r-\frac{1}{2}}}{C^{N}_{N_B}}$$ >> deze twee geven ons de run distributie
gemiddelde en standaard afwijking van runs	$$<r>= 1 + 2\frac{N_B N_O}{N} \qquad V(r) = 2\frac{N_B N_O(2N_B N_O - 1)}{N^2(N-1)}$$ voor r>10 kan de run distributie goed benaderd worden door een Gauss
effectiviteit van run tests	run test is veel minder krachtig dan χ^2-test > echter: χ^2 negeert het teken vd afwijkingen, run test kijkt enkel naar de tekens > run test kan nuttige bijkomende info geven
8.3.4 Kolmogorov test	
situatie Kolmogorov test	= alternatief in geval dat steekproef te klein is voor een nuttige χ^2-test
werking Kolmogorov test	neem de waarden en rangschik ze in stijgende orde > plot de cumulatieve distributie $cd_N(x)$ gedeeld door aantal waarden N > geeft trapfunctie van 0 naar 1 die 1/N toeneemt bij elk punt > teken dan de cumulatieve verdeling: $$F(X) = \int_{-\infty}^{X} f(X')dX'$$ voor wssheidsverdeling f(X) hetgeen een continue functie zal zijn De grootste afstand tss de twee curven is dan: $$D = \max(\lvert \mathrm{cd}_N(x) - F(X) \rvert)$$ definieer $d=D\sqrt{N}$ > als de overeenkomst tss theorie en experiment goed is, is d klein 
gebrek Kolmogorov test	test is enkel te gebruiken als de verdeling op voorhand gekend is

8.4 vergelijking van twee steekproeven	
vergelijken van steekproeven	= bekijken of twee steekproeven compatibel zijn vb: groep van mensen met een medicijn en een controlegroep als de 2 metingen steekproeven zijn uit dezelfde ouderverdeling > dan zullen ze verschillen van elkaar door statistische fluctuaties
8.4.1 twee Gaussische steekproeven met bekende σ: Gauss	
gelijkheid via Gauss	We hebben twee waarden X en Y met standaardafw. σ_X en σ_y > deze zijn compatibel als X-Y compatibel is met 0 > eenvoudig uit te werken via: $$V(X-Y)=\sigma_X^2+\sigma_Y^2$$ >> vraag herleid tot hoeveel σ=V(X-Y) afwijkt van $\mu_0=0$ > zie daarvoor vorige secties
8.4.2 twee Gaussische steekproeven met onbekende σ: Student	
gelijkheid via Student	veronderstel dat X en Y afkomstig zijn van Gauss verdelingen > we kennen de standaardafw. niet dus voor meervoudige metingen: $$\hat{\sigma}_X = s_X = \sqrt{\frac{\sum(x_i-\bar{x})^2}{N_X-1}} \qquad \hat{\sigma}_Y = s_Y = \sqrt{\frac{\sum(y_i-\bar{y})^2}{N_Y-1}}$$ aangezien we een schatting voor σ gebruiken moeten we een t-distributie gebruiken > t testen die opgebouwd is uit een Gauss gedeeld door een χ^2 - onder de nulhypothese is $\mu_1=\mu_2$ en is het verschil: $$\frac{(\bar{x}-\bar{y})}{\sqrt{\frac{\sigma_X^2}{N_X}+\frac{\sigma_Y^2}{N_Y}}}$$ > Gaussisch verdeel met gemm. 0 en standaardafw. 1. Dit geeft de teller - De som $$\frac{(N_X-1)s_X^2}{\sigma_X^2}+\frac{(N_Y-1)s_Y^2}{\sigma_Y^2}$$ is een χ^2 met N_X+N_Y-2 vrijheidsgraden zoals je kan zien uit vergelijking 8.12. Twee vrijheidsgrader gaan verloren omdat de twee gemiddelden onbekend zijn. >> hieruit halen we t: $$t=\frac{(\bar{x}-\bar{y})}{S\sqrt{\frac{1}{N_X}+\frac{1}{N_Y}}}$$ met: $$S^2=\frac{(N_X-1)s_X^2+(N_Y-1)s_Y^2}{N_X+N_Y-2}$$ en is verdeeld volgens een t-distributie met vrijheidsgraden $n=N_X+N_Y-2$
> betekenis S	= samengestelde schatter van de standaardafwijking > schatter op basis van twee steekproeven met een aangepaste gewichtsfactor dan is de term $S\sqrt{(1/N_X)+(1/N_Y)}$ analoog aan de standaardfout op gemiddelde σ/N dat gebruikt wordt als σ gekend is
> voordeel methode	deze methode is heel robuust > zelfs als centraal Limiet Theorema niet geldt, aangezien er te weinig metingen zijn, en verdelingen zijn niet helemaal Gaussisch zal dit nog steeds betrouwbare waarden geven

8.4.3 overeenstemmende en gecorreleerde steekproeven

gecorreleerde steekproeven vergelijken

in plaats van de individuele gemiddelden te bekijken, bekijk het gecorreleerde gemiddelde:

$$\sum_i (x_i - y_i)$$

de verdeling $d_i \equiv (x_i - y_i)$ heeft als variantie:

$$\sigma_d^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$$

> de variantie verkleint indien beide steekproeven gecorreleerd zijn

vb: overeenstemmende paren

Bekijk de topsnelheid van fietsers met en zonder een nieuw onderdeel:

Fietser	A	B	C	D	E	F	G
Topsnelheid (km/h) zonder onderdeel	29	30	42	34	37	45	32
Topsnelheid (km/h) met onderdeel	36	26	46	36	40	51	35
Verschil	7	-4	4	2	3	6	3

dan:

☐ Als je de correlatie negeert zijn de gemiddelden 35.6 ± 2.3 zonder, en 38.6 ± 3.1 met het nieuwe onderdeel. Het verschil van 3 km/h is niet significant want het is kleiner dan de fouten.

☐ Kijk nu naar de verschillen. Het gemiddelde daar is 3.0. De geschatte standaardafwijking s is 3.6, zodat de fout op het geschatte gemiddelde gelijk is aan 3.6/ $\sqrt{7} = 1.36$. t is dan 3.0/1.36 = 2.2. Dit is significant op het 5% niveau volgens de Student t verdeling (eenzijdige test, 6 vrijheidsgraden, $t_c = 1.943$).

8.4.4 vergelijking van twee varianties: Fisher-Snedecor

def: vergelijken van varianties

Om de geschatte varianties van de twee steekproeven te vergelijken vorm je de verhouding:

$$F = \frac{\hat{V}_1}{\hat{V}_2}$$

waar $\hat{V} = \sum(x_i - \bar{x})^2/(N_x - 1)$ voor elke steekproef.

>> zal bij 1 liggen als de twee oudervarianties gelijk zijn

st: Fisher distributie

Dergelijke grootheid F is verdeeld volgens de Fisher distributie:

$$P(F) = \frac{\Gamma(\frac{f_1+f_2}{2})}{\Gamma(\frac{f_1}{2})\Gamma(\frac{f_2}{2})}\sqrt{f_1^{f_1} f_2^{f_2}}\frac{F^{\frac{f_1}{2}-1}}{(f_2 + f_1 F)^{\frac{f_1+f_2}{2}}}$$

met $f_i = N_i - 1$ het aantal vrijheidsgraden van twee χ^2 verdelingen.

st: F voor grote getallen

Voor grote getallen kan bewezen worden dat:

$$Z = \frac{1}{2}\log F$$

een verdeling heeft die redelijk goed overeenstemt met een Gauss met gemiddelde $\mu = \frac{1}{2}\left(f_2^{-1} - f_1^{-1}\right)$ en variantie $\sigma^2 = \frac{1}{2}\left(f_2^{-1} + f_1^{-1}\right)$.

<table>
<tr><td colspan="2">8.4.5 het algemene geval</td></tr>
<tr><td>algemeen geval</td><td>geen veronderstellingen over de verdeling
> vraag niet of de gemiddelden of varianties gelijk zijn
> enkel of de verdelingen hetzelfde zijn</td></tr>
<tr><td colspan="2"><u>a) run test</u></td></tr>
<tr><td>run test in voor vgl van verdelingen</td><td>orden beide steekproeven tezamen om een rij van x'en en y's te krijgen
> als x en y uit dezelfde verdeling getrokken zijn
dan zullen ze helemaal door elkaar gemengd zijn
> runs van opeenvolgende waarden uit dezelfde steekproef zal kort zijn
> er zullen veel runs voorkomen

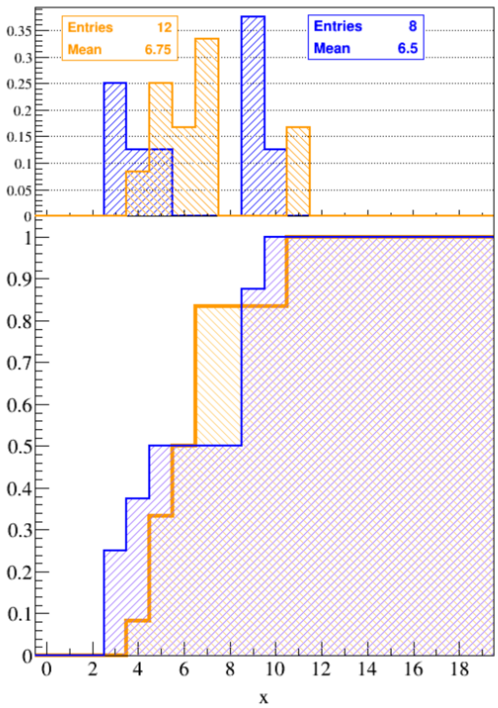
ie: hoe meer runs, hoe beter ze overeenkomen</td></tr>
<tr><td>situatie run test</td><td>enkel als N_x~=N_y, anders krijg je altijd lange runs</td></tr>
<tr><td colspan="2"><u>b) kolmorogov test</u></td></tr>
<tr><td>Kolmorogov test</td><td>plot twee cumulatieve distributies voor x en y afzonderlijk, gedeeld door toepasselijk totaal
> twee trapfuncties van 0 naar 1 die stijgen met $1/N_x$ en $1/N_y$ bij elk meetpunt
> bepaal grootst absolute verschil tss de twee verdelingen en noem dit:

$$D = \max(|\mathrm{cd}(x) - \mathrm{cd}(y)|).$$

deze waarde wordt dan geschaald en levert:

$$d = D\sqrt{\frac{N_x N_y}{N_x + N_y}}$$

hetgeen je vergelijkt met de tabellen voor een bepaalde significantie
> is d te groot, dan zijn steekproeven afkomstig van verschillende ouderdistributie


</td></tr>
</table>