Note til styrkefunktionen

Først er det vigtigt at gøre sig klart, at når man laver statistiske test, så kan man begå to forskellige typer af fejl:

- Type 1 fejl: *At forkaste H_0 når den egentlig var sand, og dermed skulle have været accepteret.* Sandsynligheden for at begå denne fejl er direkte forbundet med testets signifikansniveau. Hvis man (sædvanligvis) tester på 95% niveau, så er risikoen for at begå en *type 1* fejl, alt andet lige større end hvis man tester på eks. 99% niveau. Hvis man derfor kun er bekymret over type 1 fejl, så kan man sige, at så bør man blot sætte signifikansniveauet meget lavt, eks. 0,005 – da man så kun gennemsnitligt i 1 ud af hver 200 tests vil komme til at begå en type 1 fejl – altså at forkaste en sand H_0, men her kommer type 2 fejl ind i billedet.
- Type 2 fejl: *At undlade at forkaste H_0 når den egentlig var falsk – altså at acceptere H_0 på trods af at H_1 var sand.* Der er et grundlæggende trade-off mellem de to typer af fejl. Hvis man vælger at reducere sandsynligheden for at begå type 1 fejl, fx ved at teste på 99% niveau i stedet for fx 95% niveau, så vil man da samtidig automatisk øge sandsynligheden for at begå en type 2 fejl, altså vil man være mere tilbøjelig til at lade tvivl tale til fordel for nulhypotesen, da man tester på 99% niveau, og dermed mere ofte komme til at opretholde nulhypotesen til trods for at alternativhypotesen faktisk var sand.

- Sandsynligheden for at begå en type 1 fejl er lig med testets signifikansniveau α.
- Sandsynligheden for at begå en type 2 fejl hænger sammen med temaet for denne note – testets styrkefunktion.

Hvad er styrkefunktionen?

Styrkefunktionen angiver sandsynligheden for at forkaste H_0 givet at denne rent faktisk *er* falsk, dvs. givet at H_1 (alternativhypotesen) er den rigtige.

Den hypotese vi tester er følgende: H_0: $\theta = \theta_0$ overfor et givet alternativ (kan både være mindre end større end eller forskellig fra θ_0). Det vi ønsker er at få en vurdering af, er sandsynligheden for at parameteren rent faktisk er θ_1 og ikke θ_0 som H_0 angiver. Denne størrelse, der netop er testets styrke, må naturligvis afhænge af værdien af θ_1. Vi skriver derfor testets styrkefunktion som $\eta(\theta_1)$. Testets styrke må alt andet lige være svagere, når man vælger en θ_1 tæt på θ_0 end når man vælger en θ_1 længere væk fra θ_0.

Hvis et tests styrke over for et givet alternativ (i form af parameteren θ_1) er en høj værdi (tæt på 1), så vil det med andre ord sige, at sandsynligheden for at forkaste H_0 givet at den rent faktisk bør forkastes er stor, og dermed at sandsynligheden for at begå type 2 fejl er relativt lille. Sandsynligheden for at begå en type 2 fejl er givet som:

$$1 - \eta(\theta_1)$$

Derfor er et test (over for et givet alternativ) et "bedre" test hvis det har en høj styrke end et test der har en lav styrke. Med andre ord et stærkere test – deraf navnet styrkefunktionen.

<u>Beregninger på styrkefunktionen</u>

Man beregner et test styrke over for et givet alternativ ved at udregne sandsynligheden for at havne i forkastelsesområdet R, givet at θ_1 er den sande værdi af θ. Dvs.

$$\eta(\theta_1) = P(X \in R \mid \theta = \theta_1)$$

<u>Eksempel (binomialfordelingen)</u>[1]:

Antag at vi har en binomialfordelt stokastisk variabel X med antalsparameteren 133 og en sandsynlighedsparameter θ. Altså

$$X \sim \text{bin}(n{=}133, \theta)$$

Vi har observeret en værdi af X til 88. Spørgsmålet er om den sandsynlighedsparameteren kan tænkes at være 0,65? Altså ønsker vi at teste:

$$\boxed{H_0\!: \theta = 0{,}65 \qquad H_1\!: \theta \neq 0{,}65}$$

Signifikanssandsynligheden for dette test er[2]

$p = 2 \cdot \min\{P(X \geq 88), P(X \leq 88)\} = 2 \cdot \min\{0{,}4278 \,;\, 0{,}6423\} = \underline{0{,}8556}$ (eksakt beregning)

[1] Dette eksempel er hentet fra Eksamensopgaven sommer 2000 Opgave 2 spørgsmål 3
[2] jf. lærebogens side 318.

Vha. normalfordelingsapproksimationen fås:

$$p = 2 \cdot \min\left\{1 - \Phi\left(\frac{88 - \frac{1}{2} - 133 \cdot 0{,}65}{\sqrt{133 \cdot 0{,}65 \cdot (1 - 0{,}65)}}\right) ;\ \Phi\left(\frac{88 + \frac{1}{2} - 133 \cdot 0{,}65}{\sqrt{133 \cdot 0{,}65 \cdot (1 - 0{,}65)}}\right)\right\}$$

$$p = 2 \cdot \min\{1 - \Phi(0{,}1909) ; \Phi(0{,}3727)\} = 2 \cdot \min\{0{,}4243 ; 0{,}6453\} = \underline{\underline{0{,}8486}}$$

Dermed må hypotesen om at parameteren θ er lig med 0,65 altså klart accepteres.

Vi skal nu beregne dette tests styrke i forhold til et alternativ om at $\theta = 0{,}50$.

Vi skal altså beregne sandsynligheden for at havne i forkastelsesområdet givet at $\theta = 0{,}50$.
Vi beregner derfor først acceptområdet (konfidensintervallet) for hypotesen H_0 (på signifikansniveau 0,05). Der er tale om et approksimativt konfidensinterval i normalfordelingen:

$$n \cdot p \pm u_{1-\alpha/2} \cdot \sqrt{n \cdot p \cdot (1 - p)} = 133 \cdot 0{,}65 \pm 1{,}96 \cdot \sqrt{133 \cdot 0{,}65 \cdot 0{,}35} = [75{,}67 ; 97{,}23]$$

Nu skal vi da finde sandsynligheden for at X *ikke* ligger i dette interval givet at den sande $\theta = 0{,}50$ – først beregnes den vha. normalfordelingsapproksimationen[3]:

$$\eta(0{,}50) = P(X \notin [75{,}67 ; 97{,}23] | \theta = 0{,}50) = \Phi\left(\frac{75{,}67 - 133 \cdot 0{,}50}{\sqrt{133 \cdot 0{,}5 \cdot 0{,}5}}\right) + \left(1 - \Phi\left(\frac{97{,}23 - 133 \cdot 0{,}50}{\sqrt{133 \cdot 0{,}5 \cdot 0{,}5}}\right)\right)$$

$$= \Phi(1{,}59) + 1 - \Phi(5{,}33) = \Phi(1{,}59) + \Phi(-5{,}33) = 0{,}9441 + 0{,}0000 = \underline{\underline{0{,}9441}}$$

Og dernæst eksakt i binomialfordelingen:

$$\eta(0{,}50) = P(X \notin [75{,}67 ; 97{,}23] | \theta = 0{,}50) = P(X \le 75{,}67) + 1 - P(X \le 97{,}23) =$$

$$P(X \le 75) + 1 - P(X \le 97) = \underline{\underline{0{,}9409}}$$

(, idet $X \sim \text{bin}(n = 133, \theta = 0{,}50)$)

Altså har testet en styrke på 94,1%. Ergo er sandsynligheden for at begå en fejl af type 2, altså at acceptere en hypotese om at $\theta = 0{,}65$, hvis den sande værdi rent faktisk er 0,50 altså omkring 5,9%.

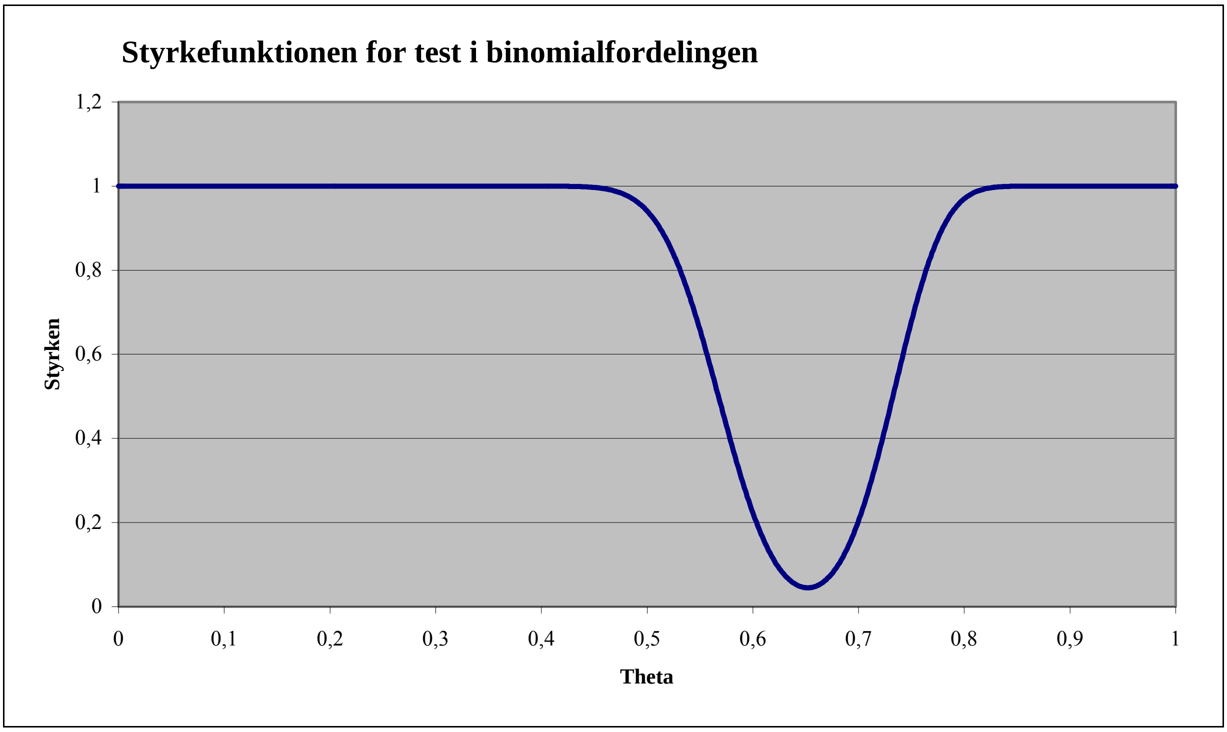
[3] Bemærk, at der her i approksimationen ikke er tillagt/fratrukket ½ som vi ellers plejer at gøre når vi skal approksimere en diskret fordeling i U-fordelingen. Grunden til dette er at de grænser der indsættes ikke er heltallige observationer fra den diskrete fordeling, men i stedet approksimativt udregnede værdier i *normalfordelingen*.

Sandsynligheden for at begå en fejl af type 2 er altså ikke særlig stor i dette tilfælde, hvilket hænger sammen med at der er relativ stor forskel på de to opstille parameterværdier – 0,65 og 0,50.

Hvis man i stedet beregner styrken for et alternativ om at $\theta = 0{,}60$, så fås følgende:

$$\eta(0{,}60) = P\left(X \notin [75{,}67\,;97{,}23] \middle| \theta = 0{,}60\right) = P(X \le 75{,}67) + 1 - P(X \le 97{,}23) =$$
$$P(X \le 75) + 1 - P(X \le 97) = \underline{\underline{0{,}2233}}$$

Altså får vi et svagere test med en ret stor risiko for at begå en type 2 fejl over for det givne alternativ, at $\theta = 0{,}60$. Tilsvarende kan man for alle $\theta \in [0;1]$ beregne testets styrke. Dette er vist i figuren nedenfor:



Eksempel (sammenligning af to normalfordelinger)[4]

Dette eksempel viser hvordan man beregner styrken for et test der sammenligner middelværdierne i to normalfordelinger.

[4] Dette eksempel bygger på Eksamensopgaven 2001 sommer opgave 2.

Vi har givet to normalfordelinger:

Gruppe	Gennemsnit	Empirisk varians
1 (n = 50 obs.)	37,89	$12{,}233^2$
2 (m = 83 obs.)	28,62	$14{,}078^2$

For at kunne udføre testet for om middelværdierne er ens, må vi først undersøge om varianserne er ens, dette testes ved følgende V-test:

H_0: $\sigma_1^2 = \sigma_2^2$ vs. H_1: $\sigma_1^2 \neq \sigma_2^2$

$$V = \frac{\max\{s_1^2, s_2^2\}}{\min\{s_1^2, s_2^2\}} \sim F(obsantal_{tæller} - 1, obsantal_{nævner} - 1)$$

$$v^* = \frac{14{,}078^2}{12{,}233^2} = 1{,}32 \sim F(83-1\,,50-1)$$

$$p = 2 \cdot P(V > v^*) = 0{,}2892$$

Ergo kan varianserne antages at være ens, og dermed kan vi udregne den poolede varians:

$$s_{pooled}^2 = \frac{(n-1)\cdot s_1^2 + (m-1)\cdot s_2^2}{n+m-2} = 13{,}42^2$$

Nu kan vi bruge t-testet for at teste om middelværdierne er ens:

H_0: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$

$$T = \frac{\overline{X}_1 - \overline{X}_2}{s_{pooled} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim T(n+m-2)$$

$$t = \frac{37{,}89 - 28{,}62}{13{,}42 \cdot \sqrt{\frac{1}{50} + \frac{1}{83}}} = 3{,}86 \sim t(50+83-2)$$

$$p = 2 \cdot P(T > |t|) = 0{,}0002 < \alpha = 0{,}05$$

I opgaveteksten tester man ensidet med alternativet H_1: $\mu_1 > \mu_2$, og da bliver signifikanssandsynligheden lig med

$$p = P(T > t) = 0{,}0001 < \alpha = 0{,}05$$

Ergo er middelværdierne *ikke* ens – middelværdien for gruppe 1 *er* signifikant større end middelværdien for gruppe 2.

Vi skal nu beregne testets styrke over for et alternativ om, at forskellen i middelværdier er lig med 12, hvor vi før testede om forskellen i middelværdien var lig med 0 – altså at de to middelværdier var ens.

Vi starter med at antage at de i de foregående beregninger indgående varianser kan betragtes som kendte med deres estimater som værdier[5].

Vi konstruerer nu en stokastisk variabel X_D som differensen mellem X_1 og X_2. Husk at fordelingen af gennemsnittene[6] er givet ved: $\overline{X}_i \sim Nf\left(\mu_i, \frac{\sigma_i^2}{n_i}\right)$. Dermed må der også gælde:

$$\overline{X}_D = \left(\overline{X}_1 - \overline{X}_2\right) \sim Nf\left(\mu = \mu_1 - \mu_2\ , \sigma^2 = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

Idet variansen kan betragtes som kendt, udregnes denne:

$$\sigma^2 = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} = \frac{12{,}233^2}{50} + \frac{14{,}078^2}{83} = 2{,}402^2$$

Dvs. at under hypotesen om at middelværdierne er ens, dvs. H_0: $\mu_D = 0$ gælder, at

$$\overline{X}_D \sim Nf\left(0\,;2{,}402^2\right)$$

Et 95% konfidensinterval (acceptområdet) for μ_D under det ensidede alternativ, at $\mu_D > 0$, findes ved at finde 95% fraktilen i den til $\overline{X}_D$ hørende normalfordeling:

$$\overline{X}_{D;0,95} = \sigma U_{0,95} + \mu = 2{,}402 \cdot 1{,}645 + 0 = 3{,}95$$

Dermed bliver 95% konfidensintervallet:

$$\mu_D \in \left]-\infty\,;3{,}95\right]$$

Nu skal vi beregne styrken for dette *ensidede* test under en alternativ hypotese om at $\mu_D = 12$:

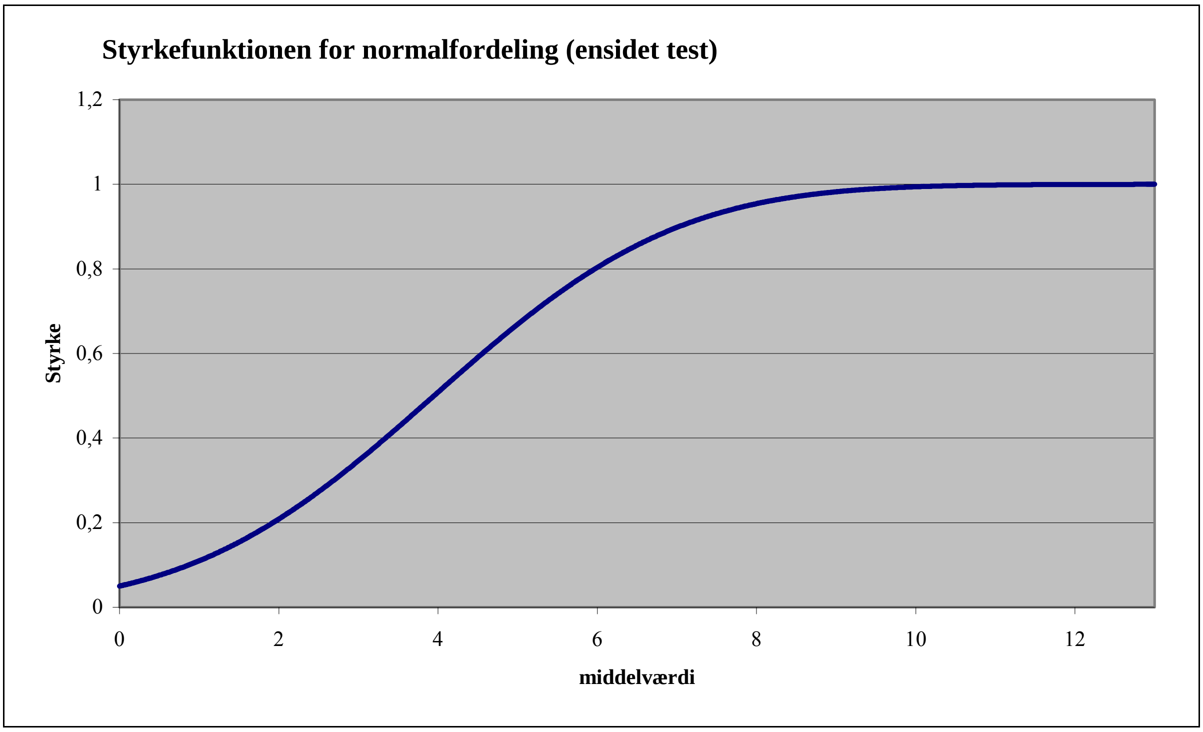
$$\eta(12) = P\left(\overline{X}_D \notin \left]-\infty\,;3{,}95\right] \middle| \mu_D = 12\right) = P\left(\overline{X}_D \in \left[3{,}95\,;\infty\right[\middle| \mu_D = 12\right) = P\left(\overline{X}_D \geq 3{,}95 \middle| \mu_D = 12\right)$$

$$= 1 - \Phi\left(\frac{3{,}95 - 12}{2{,}402}\right) = 1 - \Phi(-3{,}35) = 1 - 0{,}0004 = \underline{\underline{0{,}9996}}$$

[5] Dette er en normal antagelse at gøre, da vi ellers ikke længere kan beregne styrken i U-fordelingen.
[6] Jf. Sætning 6.2 i lærebogen.

Dette test har således en relativ høj styrke, og risikoen for at begå en type 2 fejl, altså risikoen for at acceptere hypotesen om ens middelværdier på trods af at den sande værdi af μ_D, dvs. forskellen mellem middelværdierne i de to normalfordelinger, rent faktisk er 12, er således så lav som 0,0004.

Styrken for det ensidede test kan beregnes for alle (positive) værdier af μ_D og dette er illustreret i figuren nedenfor:



Hvis i stedet man laver testet to-sidet, så skal man naturligvis blot slå 2½% og 97½% fraktilerne op i $\overline{X}_D$'s normalfordeling, og intervallet mellem disse to grænser er acceptområdet. Dermed bliver sandsynligheden for at havne i forkastelsesområdet en toleddet størrelse.

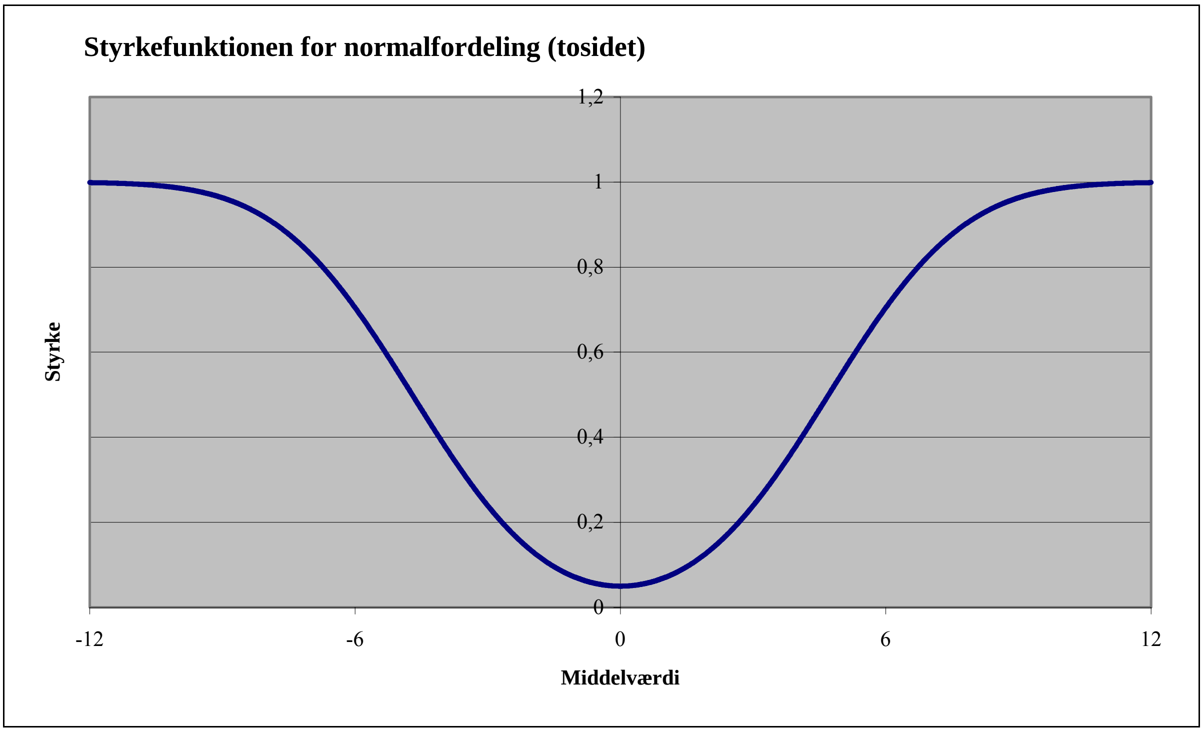
95% konfidensintervallet for μ_D bliver i dette tilfælde [-4,71 ; 4,71].

Dermed bliver styrken over for alternativet at $\mu_D = 12$ i dette tilfælde:

$$\eta(12) = P\left(\overline{X}_D \notin [-4{,}71\,;4{,}71] \,|\, \mu_D = 12\right) = P\left(\overline{X}_D \in \left]-\infty\,;-4{,}71\right] \cup \left[4{,}71\,;\infty\right[\,|\, \mu_D = 12\right)$$
$$= 1{,}75 \cdot 10^{-12} + 0{,}9988 = \underline{\underline{0{,}9988}}$$

Dermed bliver styrken for det dobbeltsidede test altså en smule svagere end for det enkeltsidede test.

Tilsvarende kan styrken også beregnes for alle værdier i det dobbeltsidede tilfælde, hvilket er vist i figuren nedenfor.



Eksempel (poissonfordelingen):

Vi betragter nu en poissonproces, hvor X angiver antallet af patienter der ankommer til en skadestue i døgnet. I forbindelse med en undersøgelse som skadestuen tidligere har gennemført har man estimeret at der gennemsnitligt ankommer 80 patienter i døgnet. Ledelsen af sygehuset har hørt i krogene, af der ikke længere er den samme mængde patienter på skadestuen, og udfører derfor en enkelt dag en stikprøve, hvor det viser sig at der ankommer 67 patienter til skadestuen. Ledelsen mener hermed at have konkluderet at besøgsmængden på skadestuen er faldet og at der derfor skal skæres i personalet.

Skadestuens personale er oprørte og har nu bedt de seje politstuderende om at vurdere om der rent faktisk er sket en signifikant reduktion i antallet af besøgene eller om det blot kan tænkes at være udslag af tilfældigheder. Dette gøres ved at teste en hypotese om at den sande λ er 80. Dvs.

H_0: $\lambda = 80$	vs.	H_1: $\lambda < 80$[7]

$X \sim \mathrm{Ps}(\lambda = 80)$

Signifikanssandsynligheden for dette test er:

$p = P(X \leq 67) = 0{,}0782 > \alpha = 0{,}05$

Altså må vi glæde personalet på skadestuen med, at vi på 5% niveau ikke kan afvise at det faktum at der den pågældende dag blot var 67 patienter, alene skyldes tilfældige udsving fra et gennemsnitligt antal patienter på 80 pr. dag.

Dette meddeler personalet på skadestuen til sygehusledelsen, der dog straks bemærker at der i den opstillede testprocedure jo er en risiko for at man kommer til at acceptere H_0 selv om den rent faktisk er falsk. Dette tygger de ansatte i skadestuen så lidt på og beder så atter en gang de politstuderende om hjælp[8].

En af de kvikke studerende finder hurtigt ud af, at det der spørges til reelt er styrken af det test der tidligere blev udført. Man beslutter sig derfor for at beregne styrken af testet over for et alternativ om at $\lambda = 67$.

Denne er givet som sandsynligheden for at X tilhører forkastelsesområdet for H_0: $\lambda = 80$, givet at $\lambda = 67$ er den sande værdi. Derfor beregner vi først konfidensintervallet for det tidligere udførte test. Idet vi er i en diskret fordeling kan vi ikke præcis ramme et 95% konfidensinterval. Det tætteste på vi kommer er et 95,1% konfidensinterval, der er intervallet $\lambda \in [66; \infty[$. Styrken η er derfor givet ved:

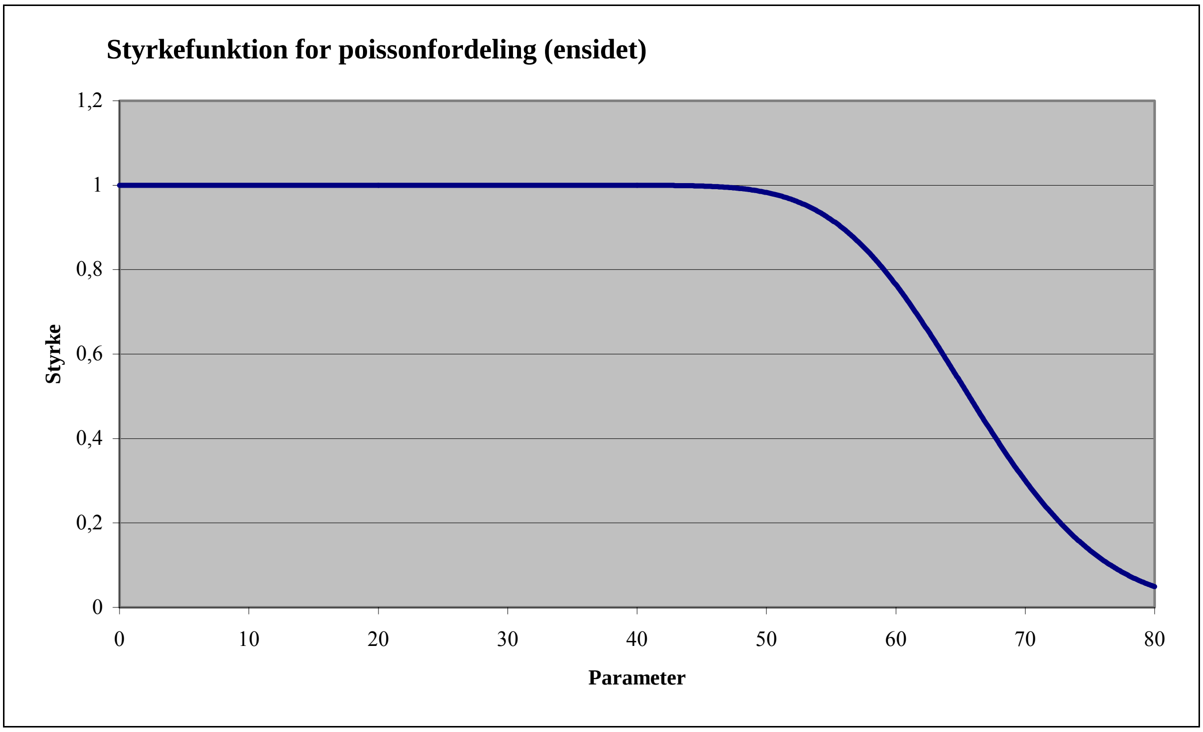
$\eta(67) = \mathrm{P}(X \notin [66; \infty[\mid \lambda = 67) = \mathrm{P}(X \in]0 ; 65] \mid \lambda = 67) = 0{,}4351$

Dvs. at sandsynligheden for at acceptere nulhypotesen om at $\lambda = 80$ givet at en alternativ værdi $\lambda = 67$ faktisk er den sande værdi af λ ret beset er ganske stor, nemlig $1 - \eta(67) = 0{,}5649$.

Derfor må man trods alt give sygehusledelsen ret i at der er en ganske stor risiko for at komme til at acceptere en hypotese om at $\lambda = 80$ selvom den sande værdi faktisk er $\lambda = 67$.

---

[7] Da man kun er interesseret i at forkaste den pågældende hypotese, såfremt man da påviser at ankomstraten er *større* end 60.



[8] Hvad skulle vi dog gøre uden dem… ☺