

Vejledning i brug af Monte Carlo program (2007)

Udarbejdet af Sune Boye.

Programmet kommer som et tilføjelsesprogram (en Add-in fil), der skal installeres, før det kan køres. Programfilen hedder 'mcc.xla' og den skal lægges i en vilkårlig mappe.

Når programmet skal installeres, er det vigtigt ikke at åbne 'mcc.xla' filen. Åben derimod et blankt Excel ark og i menuen vælges 'Funktioner-> Tilføjelsesprogrammer...' ('Tools -> Add-Ins...'). Tryk på 'Gennemse...' ('Browse...'), find dernæst 'mcc.xla' filen, vælg den og tryk 'OK'. Det er nu muligt at sætte flueben ud for tilføjelsesprogrammet (Add-In programmet) 'Mcc' i 'Tilføjelsesprogrammer' listen. For at køre programmet skal tillægspakkerne 'Analysis ToolPak' og 'Analysis ToolPak – VBA' også være installerede. De er allerede indbygget i Excel, men de skal aktiveres. Det gøres ved at sætte flueben ud for de to pakker i 'Tilføjelsesprogrammer' listen. Luk derefter 'Tilføjelsesprogrammer' vinduet ved at trykke 'OK'. Under menupunktet 'Funktioner' er det nu muligt at vælge 'Monte Carlo program...'. Hvis programmet skal opdateres eller slettes fra Excel, slet da filen 'mcc.xla' fra computeren, og i Excel vælg 'Funktioner-> Tilføjelsesprogrammer...'. I 'Tilføjelsesprogrammer' listen trykkes på 'Mcc' og det accepteres, at menupunktet slettes.

Programmet er forholdsvis ukompliceret at køre, og i det følgende gennemgås et simpelt eksempel for at komme godt i gang. Dernæst gives en mere detaljeret gennemgang af betingelserne for programdata. Og til sidst forklares programmet fra en mere teknisk synsvinkel.

1 Godt i gang

Monte Carlo programmet lægger ganske få begrænsninger på Excel-regnearkets formatering. En undtagelse er navnene på de enkelte ark – der må ikke være mellemrum i navnet. Dvs. et navn som 'Ark 1' skal omdøbes til fx 'Ark_1'. I dette eksempel benyttes to ark, og de kaldes 'Data' og 'Beregning'. I 'Data'-arket indsættes en søjle med tal. På figur 1 kan det ses, at området 'B4:B14' er benyttet. Dette søjleområde får overskriften 'Forventet værdi'. Idet det antages, at disse talværdier varierer stokastisk, indtastes yderligere en søjle med 'Best Case værdi' og 'Worst Case værdi'. På figur 1 ses det, at søjlerne 'D4:D14' og 'F4:F14' er benyttet. Der er ingen begrænsninger på, i hvilke ark data indtastes. For at illustrere, at rådata kan indtastes i et vilkårligt ark, sættes yderligere rådata ind i 'Beregning'-arket. På figur 2 kan det ses, at tallene er indtastet i cellerne 'A2:C2'. Gå nu tilbage til 'Data'-arket og lav referencer til 'Beregning'-arkets 'A2:C2'. Se figur 3. Her er referencerne lavet i cellerne 'B16', 'D16' og 'F16'. Til sidst laves en beregning i 'Beregning'-arket. I celle 'B10' indsættes formlen '=SUM(Data!B4:B14)+Beregning!A2'. Se figur 2. Således består regnearket af to ark, hvor det ene, 'Data'-arket, indeholder både rådata og referencer til rådata, og det andet indeholder rådata og en formelcelle.

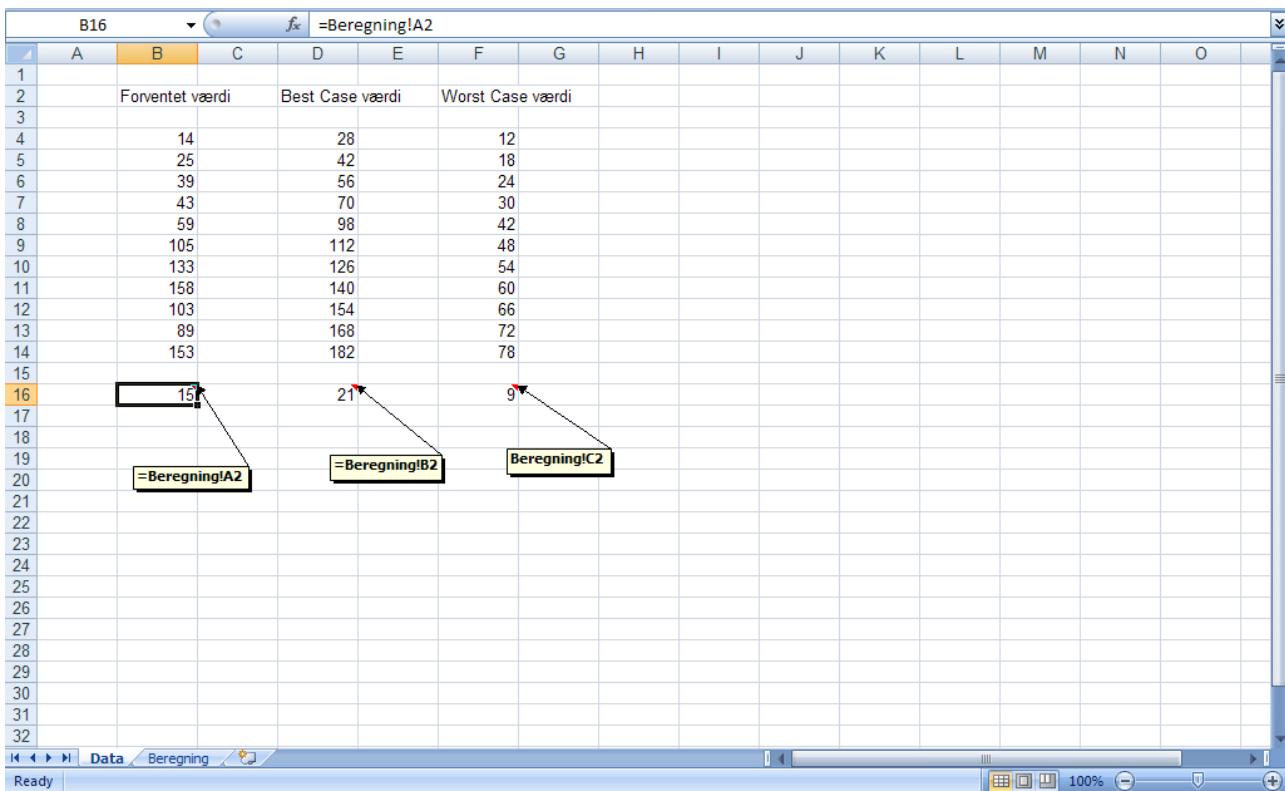
Formålet med at køre programmet i dette eksempel er at illustrere variationen i formelcellen, 'Beregning!B10', når der antages stokastisk variation i rådata. Når rådata varierer, vil formelcellen ligeledes variere, men på forhånd kan det være svært at sige præcis hvordan. Programmet giver et billede af denne variation.

A1	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1														
2	Forventet værdi		Best Case værdi		Worst Case værdi									
3														
4	14		28		12									
5	25		42		18									
6	39		56		24									
7	43		70		30									
8	59		98		42									
9	105		112		48									
10	133		126		54									
11	158		140		60									
12	103		154		66									
13	89		168		72									
14	153		182		78									
15														
16	15		21		9									
17														
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														
28														
29														
30														
31														
32														

Figur 1. Rådata i 'Data'-arket

B10	B	C	D	E	F	G	H	I	J	K	L	M	N	O
		=SUM(Data!B4:B14)+Data!B16												
2	15	21	9											
3														
4														
5														
6														
7														
8														
9														
10		936												
11														
12														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														
28														
29														
30														
31														
32														
33														

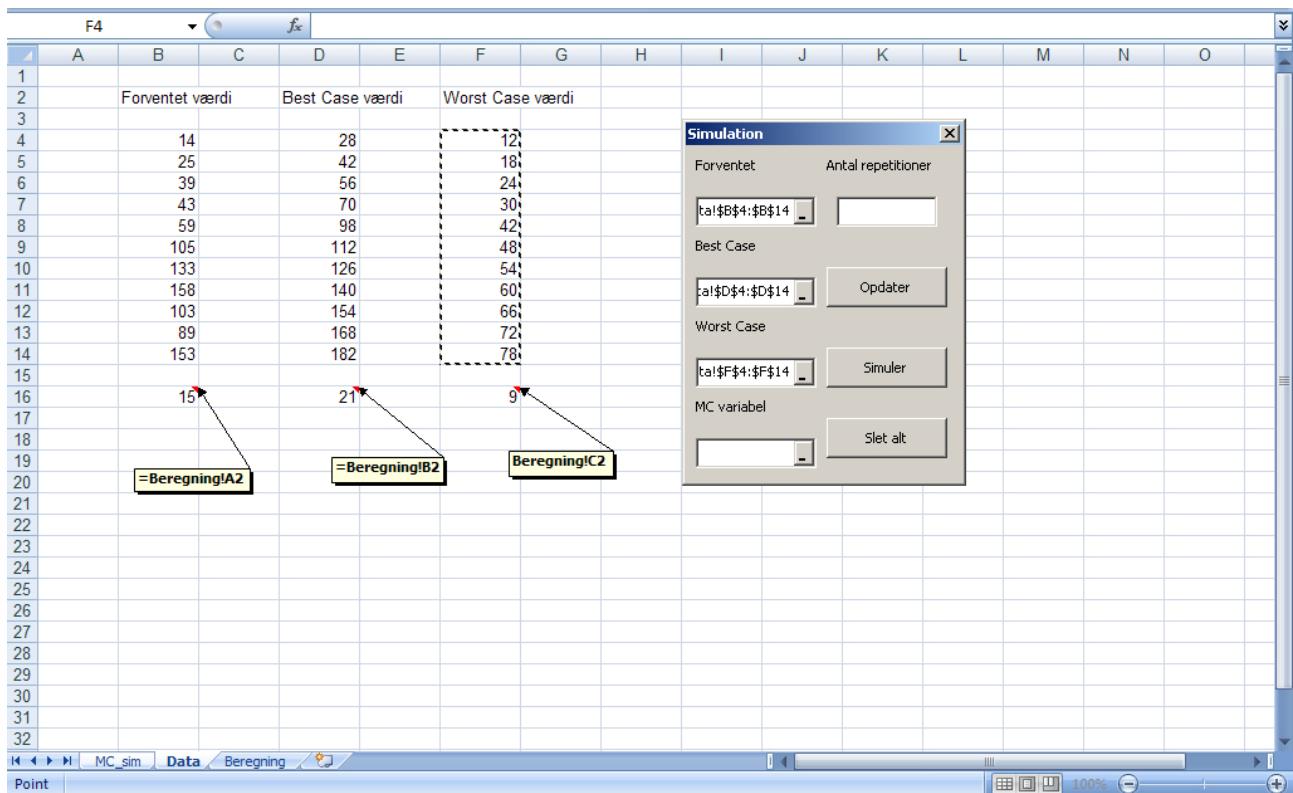
Figur 2. Rådata og formel i 'Beregning'-arket



Figur 3. Referencer i 'Data'-arket

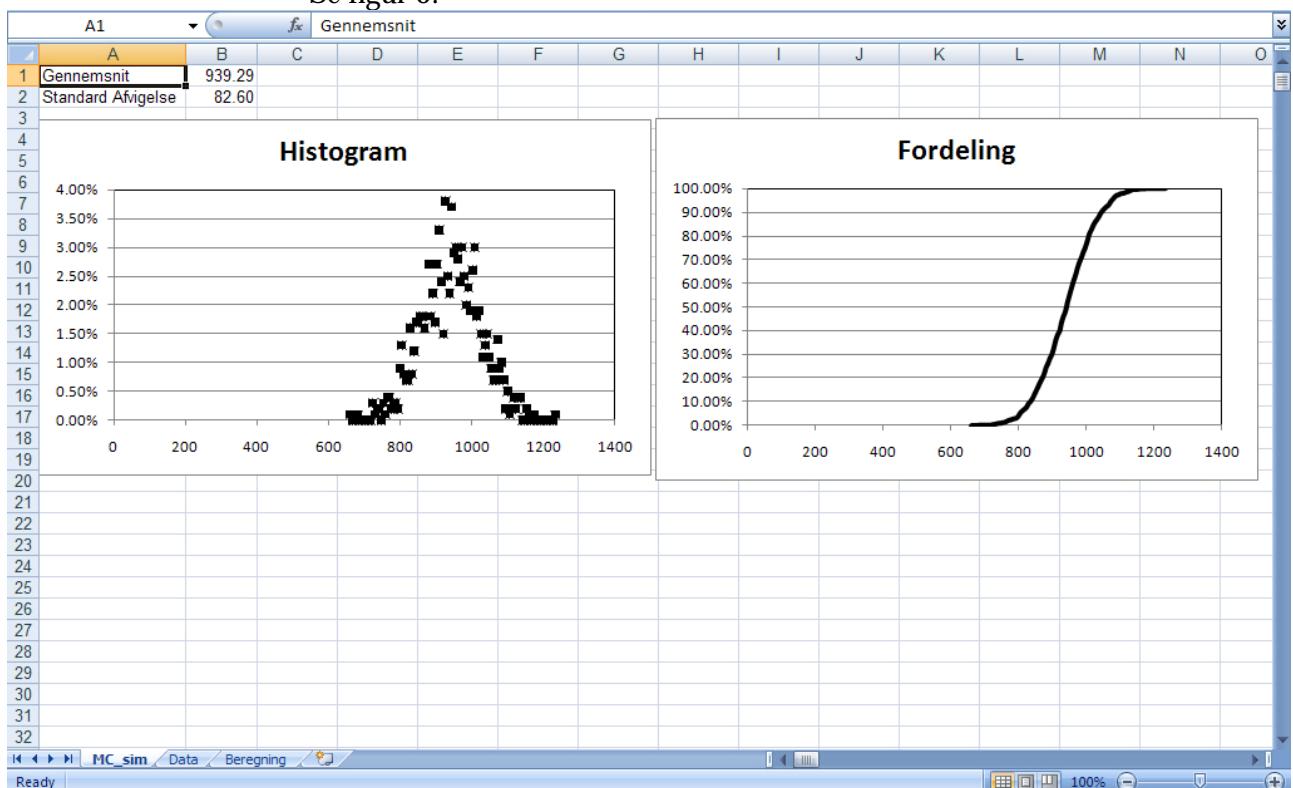
En af styrkerne ved programmet er, at det kan håndtere referencer til data. Når programmet køres, indsættes celler fra 'Data'-arket inklusiv referencerne til 'Beregning'-arket. Det er en fordel, når man arbejder med større projekter. I praksis kan det anbefales, at 'Data'-arket fungerer som en opsummering af alle variable i et datasæt og derfor udelukkende består af referencer til andre ark i det fulde regneark.

Programmet sættes i gang 'Funktioner -> Monte Carlo program...'. Bemærk at der dannes et nyt ark, 'MC_sim', og i dette ark præsenteres simulationsresultaterne, når programmet er kørt. En brugerform åbnes, der består af en række referencebokse og tre knapper. Først skal der indsættes værdier for 'Forventet', 'Best Case' og 'Worst Case'. Det gøres ved at vælge områderne 'Data!\$B\$4:\$B\$14', 'Data!\$D\$4:\$D\$14' og 'Data!\$F\$4:\$F\$14' i de respektive bokse. Se figur 4. Dernæst trykkes på knappen 'Opdater'. Derefter indsættes områderne 'Data!\$B\$16', 'Data!\$D\$16' og 'Data!\$F\$16' og der opdateres igen med tryk på knappen 'Opdater'. Så vælges et antal gentagelser i boksen 'Antal repetitioner', fx 1000, og til sidst vælges den interessante variabel, formelcellen fra 'Beregning'-arket, i 'MC variabel' boksen. Programmet køres ved at trykke på knappen 'Simuler'. Når programmet er færdig med at regne, lukkes det ved at trykke på krydset i toppen brugerformen.



Figur 4. Rådata indsættes i programmet

Resultaterne vises i 'MC_sim'-arket og består af to figurer og to værdier i cellerne 'B1' og 'B2'. Figurerne viser henholdsvis fordelingen af den interessante variabel og et histogram, og i celle 'B1' vises gennemsnittet af alle de simulerede værdier af variablen, mens celle 'B2' viser standardafvigelsen. Se figur 5.



Figur 5. Resultatet vises, når programmet er kørt

2 Betingelser for data og regnark

Som nævnt tidligere er der ikke mange begrænsninger på formateringen af data i regnarket. Men hvis betingelserne ikke overholdes, vil programmet melde fejl. Nedenfor beskrives først krav til formatering og dernæst gives gode råd til kørsel af programmet.

Når der er peget på data i de tre bokse i programmet, altså forventet, worst og best case værdier, skal der trykkes på knappen 'Opdater'. Men for at programmet skal levere resultater, skal brugeren sørge for, at data er indtastet i korrekte blokke. Det betyder først og fremmest, at blokken med forventede værdier har samme antal celler som blokken med worst og best case værdier. En blok må gerne bestå kun af en enkelt celle, så længe forventet, worst og best case hver især kun består af én celle. Med andre ord skal der altid angives både worst og best case værdi for hver forventet værdi. Desuden skal data altid indtastes i søjleblokke lige som i eksemplet ovenfor.

En, flere eller alle rækker i en datablok må gerne være begrænset af sin egen forventede værdi. Det betyder, at enten best case værdien er lig med den forventede værdi, eller også er worst case værdien lig med den forventede værdi eller også er både worst og best case værdien lig den forventede værdi. I sidstnævnte tilfælde vil der ikke være nogen stokastisk variation og disse celler vil ikke påvirke variationen i den interessante celle. Hvis en celle i datablokken er formateret således, at enten worst eller best case værdien er lig den forventede værdi, så er den stokastiske variabel enten opad eller nedad begrænset. Det betyder rent praktisk, at den simulerede værdi aldrig antager værdier enten højere eller lavere end den begrænsende værdi. Worst og best case værdien kan altså uden problemer være den samme som den forventede værdi, men to regler skal følges. Først og fremmest skal alle celler udfyldes med en værdi. Programmet accepterer ikke tomme celler. Et eksempel er, at den forventede værdi er 0, og at den er nedadtil begrænset således, at worst case værdien også er 0. Brugeren må ikke undlade at skrive et '0' i den specifikke celle. Selvfølgelig skal 0'et også skrives i cellen med den forventede værdi. Den anden regel er beslægtet. Hvis worst case værdien er lig den forventede værdi, må brugeren ikke lave en formel i worst case cellen, altså skrive at cellen med worst case værdien er lig med cellen med den forventede værdi. Overordnet set gælder, at alle rádata skal indtastes som rádata - formler og huller accepteres ikke.

En datablok består således af en sammenhængende søjle af én eller flere celler, og hver gang der trykkes på knappen 'Opdater', skal der altså være 3 datablokke med det samme antal rækker. Det er vigtigt, at blokken består af kun 1 søjle, og at rækkerne hænger sammen, men der er ingen begrænsning på højden af søjlen, dvs. antallet af rækker. Derudover viste eksemplet ovenfor, at programmet ikke kun tager rádata. Programmet tager også referencer til rádata som input, og reglen er således, at datablokkene enten består af rádata eller af referencer til rádata. Det er imidlertid en god idé altid at samle rádata og ikke blande rádata og referencer sammen. Hvis de forventede værdier i et dataark er indtastet som rádata, skal både worst og best case værdierne også indtastes som rádata, og ligeført hvis de forventede

værdier er angivet som reference til et ark med rådata, så skal worst og best case værdierne også være angivet som referencer. Det er også vigtigt, at celler med rådata ikke indeholder formler. Det skyldes hovedsageligt, at når programmet er færdig med at køre, så overskrives cellen med rådata med sin egen oprindelige værdi. Hvis fx cellen har indholdet ' $=20*0,10$ ', før programmet køres, vil indholdet ændres til '2' efter programmet er kørt. Rådata består udelukkende af tal.

Der skal således tre datablokke (forventet, best og worst case blokke) til at specificere en gruppe stokastiske variable. Selvom programmet kan håndtere mange af disse grupper af stokastiske variable, anbefales det, at der i praksis ikke benyttes mere end fyra grupper; dvs. at brugeren ikke opdaterer mere end fyra gange.

Brugeren anbefales altid at gemme sit regneark inden programmet startes. Programmet er ung, og der kan opstå fejl. Hvis der opstår fejl anbefales det simpelthen at starte forfra med backuppen af regnearket. Desuden kan det anbefales at køre programmet med få repetitioner, når det køres første gang afhængigt af antallet af varierende celler. Når programmet køres på et regneark, hvor mange blokke varierer stokastisk, kan det tage adskillige minutter at simulere, hvis der er valgt mange repetitioner, fx 10000. Det anbefales at begynde med et par hundrede repetitioner først og derefter sætte antallet op. Bemærk at hver gang der skal køres en simulationsopgave, så skal alle datablokke opdateres igen. En god regel er altid at lukke programmet på krydset i toppen af brugerformen, når det har regnet færdigt. Når programmet startes på ny, slettes al den tidligere information, grafer og resultater. Hvis der skal køres flere forskellige scenarier, så kopier 'MC_sim'-arket over i et separat regneark.

Hvis brugeren finder en fejl e.l. i regnearket, mens programmet er åbnet, og allerede har opdateret programmet, benyttes knappen 'Slet alt'. Et tryk på denne knap sletter ikke 'MC-sim'-arket, men bringer al information tilbage til før programmet blev opdateret. Hvis programmet skal lukkes før simulationen er fuldendt, bruges altid først 'Slet alt'-knappen, før der trykkes på krydset i toppen af brugerformen.

Til sidst kan to ting anbefales. Først og fremmest kan det anbefales i tilfælde med store regneark at samle al rådatainformation i et forudsætningsark – f.eks. i et referenceark. Det er meget nemmere at opdatere programmet, når datablokkene er samlet i et regneark. Det er ikke noget krav, at rådata eller referencer til rådata samles i et ark, da programmet sagtens kan opdatere data fra forskellige ark, men af praktiske årsager anbefales det, at brugeren samler rådatainformation i et ark. Den anden bemærkning drejer sig specifikt om worst og best case værdierne. Programmet benytter altid normalfordelingen til at beskrive den stokastiske variation. Normalfordelingen er symmetrisk omkring middelværdien, og derfor er programmet på nuværende tidspunkt ikke klargjort til skæve fordelinger. Eftersom variationen udregnes udelukkende på baggrund af værdierne, der er indtastet som forventet, best og worst case, bør best og worst case ligge omkring den forventede værdi nogenlunde symmetrisk, dvs. afstanden bør være lige stor mellem best og forventet værdi og forventet og worst værdi. Dette gælder selvfolgelig ikke, når variationen begrænses som beskrevet ovenfor. Her beregnes de simulerede værdier en smule anderledes.

3 Den tekniske vinkel

Før de endelige resultater kan ses i 'MC_sim'-arket udfører programmet en lang række operationer, hvor langt de fleste sørger for, at Excel læser de rigtige værdier, printer de rigtige værdier osv. Disse operationer beskrives ikke her. I stedet gennemgås de statistiske forudsætninger, som programmet opererer ud fra.

Al variation i de stokastiske variable antages at kunne beskrives med normalfordelingen. Det er grundforudsætningen for programmet (programmet kan potentielt udbygges til at omfatte flere fordelinger). Normalfordelingen tager to parametre, en middelværdi og en varians. Derfor er udfordringen at finde en middelværdi og en varians, når de tre værdier, som er programmets input, er en forventet, en best og en worst case værdi.

Middelværdien findes direkte som den indtastede forventede værdi. Variansen er mere kompliceret. Den baserer sig på en vigtig antagelse, om hvad det vil sige at være best case og worst case. Hvis best case værdien er større end den forventede værdi, så er antagelsen, at der er 97,5 % sandsynlighed for at en realisering af den stokastiske variabel er lavere end best case værdien. Hvis den er lavere end den forventede værdi, er antagelsen, at der er 2,5 % sandsynlighed for, at en realisering af den stokastiske værdi er lavere end best case værdien. Så benyttes to hjælpemidler fra Excel, en normalfordelingsfunktion og en målsøgningsfunktion. Først udregnes forskellen mellem den forventede værdi og best case værdien. Når denne værdi indsættes som standardafvigelse, og den forventede værdi indsættes som middelværdi er det muligt at benytte normalfordelingsfunktionen i Excel. Derefter benyttes målsøgningsfunktionen på normalfordelingsfunktionen, således at et mål på enten 2,5 % eller 97,5 % nås, når forskellen på forventet og best case værdi tillades at kunne ændres. På den måde findes en standardafvigelse ud fra best case værdien, og på fuldstændig tilsvarende måde findes en standardafvigelse på baggrund af worst case værdien. Den endelige standardafvigelse, der benyttes i programmet er blot gennemsnittet af disse to afvigelser beregnet på baggrund af worst case og best case værdien. Men netop fordi, normalfordelingen er symmetrisk anbefales det, at worst og best case værdierne befinner sig symmetrisk rundt om den forventede værdi. På den måde opnås de pæneste resultater.