Vedlegg 1 – Statistisk grunnlag

1.1 Hva er usikkerhet

I prosjektsammenheng sier man ofte overordnet om usikkerhet at det er gapet mellom den viten og kontroll som eksisterer i prosjektet, og den viten og kontroll man skulle hatt for å være sikre på å oppnå et optimalt resultat. I forbindelse med kostnadsoverslag er det mer korrekt å si at usikkerhet er knyttet til ukjente størrelser, som enten ikke kan måles, eller avhenger av hendelser som ennå ikke har inntruffet. Det er for eksempel umulig å vite på forhånd eksakt hvor store mengder stein som vil bli sprengt ut av tunnelen, eller å forutsi hvordan værforholdene vil bli under byggeperioden.

1.2 Stokastiske kostnadsoverslag

Et kostnadsoverslag som tar utgangspunkt i usikre verdier, kalles et stokastisk kostnadsoverslag. Hva dette er, og hva det innebærer er forklart i detalj i de neste delkapitlene.

Det motsatte er deterministiske overslag, der man sier at et element i kostnadsoverslaget kommer til å koste et eksakt kronebeløp. Fordelen med stokastiske kostnadsoverslag er at de som regel gir et mye riktigere bildet av kostnadene og usikkerhetene knyttet til disse enn deterministiske overslag. Ulempen er at dette stiller mye større krav til oppbyggingen av overslaget og til de som gjør jobben med dette.

Anslagsmetoden som er beskrevet i denne håndboken, er en metode for å lage stokastiske kostnadsoverslag. Men samtidig er det påkrevd at man har en viss kjennskap til en del grunnprinsipper fra statistikk og sannsynlighetsteori. Hvis man ikke kan disse er det stor sannsynlighet for at man ender opp med å bygge inn feil i overslagene.

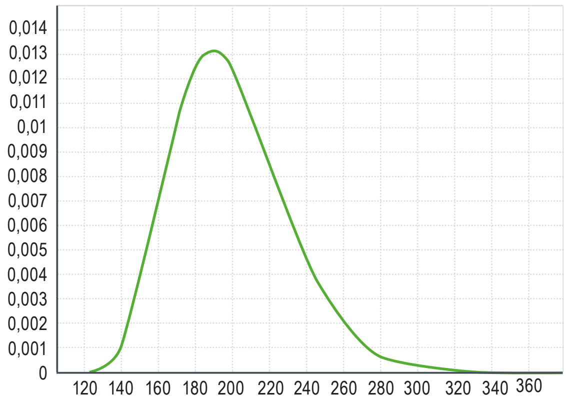
1.3 Usikre størrelser

For de aller fleste vil den mest naturlige måten å angi usikkerhet på være å angi et spenn. Man sier at en størrelse er større enn X, men mindre enn Y. Dette fungerer greit i det daglige, hvis man snakker om hvor mye en kartong melk koster eller tilsvarende. Men for kostnadsoverslag som skal være beslutningsgrunnlag for investeringer i million- og milliardklassen er ikke dette godt nok. Man kan si at kostnaden av en rundkjøring vil ligge mellom X og Y, men det er klart for de fleste at det er ikke like sannsynlig at kostnaden havner på Y som et sted mellom X og Y. For å kunne utrykke både spenn og sannsynlig- heter for usikre verdier er det mest hensiktsmessig å bruke en sannsynlighetsfordeling.

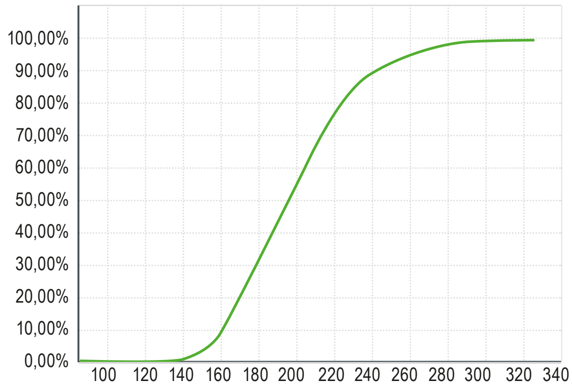
Sannsynlighetsfordelinger

En sannsynlighetsfordeling angir den relative sannsynligheten for at en usikker størrelse skal bli en viss verdi. Sannsynlighetsfordelinger kan enten være diskrete eller kontinuerlige. En terning har en diskret sannsynlighetsfordeling, det er seks bestemte utfall som er mulig, mens stort sett alt som har med penger og mengder å gjøre er kontinuerlige størrelser.

Det er to måter å tegne opp en sannsynlighetsfordeling på. Dette er vist i hhv. figur 1 Sannsynlighets-vurdering, og figur 2 Kumulativ sannsynlighetsvurdering. Figur 1 viser en sannsynlighetsvurdering der sannsynligheten for at den usikre størrelsen havner mellom to verdier. Siden det er umulig å lese av sannsynligheten direkte i kurven til denne type sannsynlighetsvurdering, benytter man seg vanligvis heller av en kumulativ sannsynlighetsfordeling, eller S-kurve, som vist i figur 2, for å kunne lese ut sannsynlighetene for at en usikker størrelse ikke blir større enn en gitt verdi rett fra kurven.



Figur 1: Sannsynlighetsfordeling



Figur 2: Kumulativ sannsynlighetsfordeling

I Anslag angis alle usikre størrelser ved hjelp av sannsynlighetsfordelinger av typen gammafordeling. Dette er en fordeling som er svært godt egnet til å modellere de typer usikre størrelser man håndterer i Anslag.

Angivelse av usikre størrelser

Gammafordelingen har en matematisk formulering, men det er noe man slipper å forholde seg til ved bruk av Anslag. Når man legger inn tall på postene skjer dette i form av trippeloverslag. I teorien kan man velge hvilke som helst tre punkt på kurven, og det vil være nok for å entydig definere gammafordelingen. I praksis har man valgt å bruke 10 prosent-kvantilet, mest sannsynlig verdi og 90 prosent-kvantilet. Dette er av psykologiske årsaker. Hvis man ber noen om å gi et overslag over hvor dyrt eller billig noe kan bli, klarer de i praksis aldri å anslå ekstremverdiene. Det har vist seg at 10/90 er mer korrekt plassering for verdiene man kommer opp med.

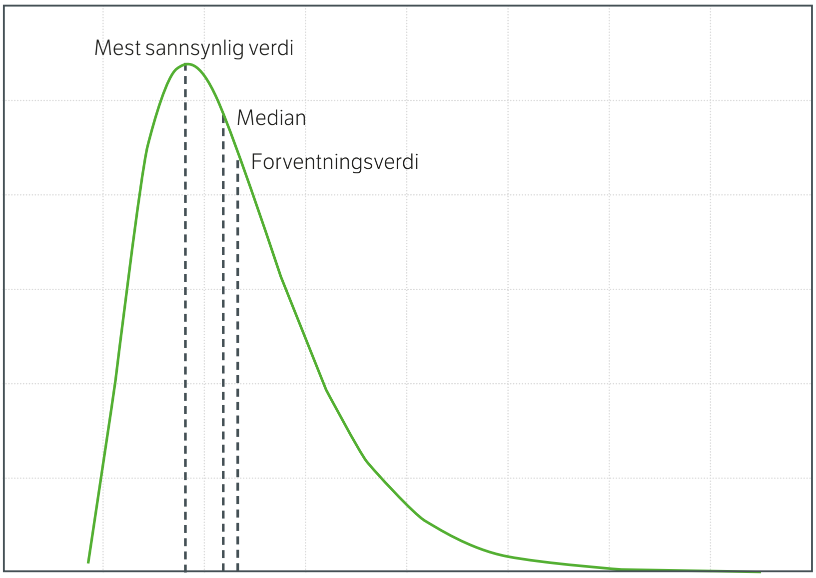
Mest sannsynlig verdi er toppunktet på en sannsynlighetsfordeling. Dette er den verdien som er mest sannsynlig at vil forekomme. Det er også den verdien folk vil anslå hvis man bare ber om en verdi for hvor mye noe vil koste.

Sannsynlighetsfordelingers anatomi
Skjevhet
Sannsynlighetsfordelinger som er usymmetriske sier man er skjeve. En fordeling der toppunktet ligger mot venstre og med en lengre hale mot høyre, sier man er høyreskjev.

Gjennomsnitt - Mål på sentraltendens

Et gjennomsnitt er en verdi som er typisk, eller representativ, for et sett med data. Siden slik typiske verdier tenderer til å ligge sentralt innen et sett med data ordnet etter størrelse, kalles gjennomsnitt også mål på sentraltendens. Flere ulike typer gjennomsnitt kan defineres, og i Anslag er det tre ulike man forholder seg til; mest sannsynlig verdi, forventningsverdi og median.

For en symmetrisk sannsynlighetsfordeling vil disse verdiene tre være sammenfallende. Mens de vil være ulike for en skjev fordeling. Figur 3 viser forholdet mellom disse for en høyreskjev fordeling.



Figur 3: Sammenhengen mellom de tre målene for sentraltendens.

Mest sannsynlig verdi

Den mest sannsynlige verdi er toppunktet i en sannsynlighetsfordeling og, som navnet tilsier, den enkeltverdien det er størst sannsynlighet at vil forekomme. I Anslag benytter man denne i forbindelse med angivelse av trippelansalg, men er ikke noe man forholder seg til for resultatkurven.

Forventningsverdi

Forventningsverdien er tyngdepunktet i en sannsynlighetsfordeling. Den er summen av alle tenkelige utfall, hvor hver av dem er vektet med sine respektive sannsynligheter.

Hvis man snakker om statistikk i stedet for sannsynlighetsregning, så er det tilsvarende begrepet aritmetisk middelverdi, eller med andre ord, hva de fleste forbinder med begrepet gjennomsnitt.

Når man jobber i Anslag så er det forventningsverdien man forholder seg til som resultatene på postene.

Median – P50

Medianen er det punktet i en sannsynlighetsfordeling der halvparten av arealet under kurven ligger til venstre og den andre halvparten av arealet ligger til høyre. Det vil med andre ord si at medianen er identisk med 50 prosent-kvantilet, eller P50.

Prosentkvantiler og P-verdier

Et X-prosentkvantil er den verdien av den stokastiske variable som er slik at det er X prosent sannsynlig at den ikke blir overskredet. 85 prosent-kvantilet angir for eksempel den verdien som det er 85 prosent sannsynlig at ikke overskrides. I praksis benevner en prosentkvantilene som P85, P50 og tilsvarende.

Mål på usikkerhet

Varians

Varians er det forventede kvadratavviket fra forventningsverdien. Variansen til usikre størrelser er ikke noe man forholder seg til direkte i Anslag. Definisjonen er kun tatt med her siden standardavviket er avledet fra den.

Standardavvik - σ

Standardavviket er matematisk definert som kvadratroten av varians. Den har samme størrelsesorden som forventningsverdien og er derfor det vanligste målet på usikkerhet, enten direkte eller avledet til relativt standardavvik. For en normalfordeling er det 68,2 prosent sannsynlighet at verdien vil ligge innenfor +/- et standardavvik. Figur 4 illustrerer dette.

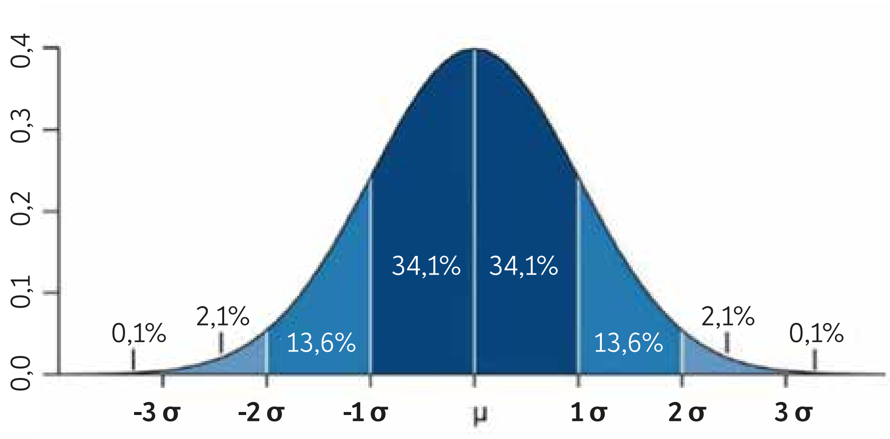
Relativt standardavvik

Relativt standardavvik er lik standardavvik delt på forventningsverdien og oppgis i prosent. Når folk snakker om størrelsen på usikkerheten i et anslag er det som regel dette tallet de refererer til.

Estimering av usikre størrelser

Hvis man sitter på store mengder erfaringsdata er det mulig å tilpasse en sannsynlighetsfordeling til de dataene man har. Dette bygger på det man kaller frekventativ statistikk. Problemet med de usikre størrelsene man forholder seg til i vegprosjekter, er at man svært sjelden har så mye erfaringsdata at dette er mulig. Løsningen da er å benytte bayesiansk statistikk. Her benytter man det man sitter på av erfaringsdata i kombinasjon med subjektive ekspertvurderinger.

Subjektive ekspertvurderinger vil si at man får personer som har lang erfaring innenfor sitt fagfelt, til å anslå verdien av de usikre størrelsene ut ifra sin erfaring og de dataene som foreligger. Det er som regel gunstig å få flere eksperter til å vurdere de samme størrelsene. I Anslag gjøres dette fortrinnsvis i en gruppeprosess. Hvordan dette gjøres i praksis er beskrevet i kapittel 6.



Figur 4: Standardavvik og sannsynligheter (Kilde: Wikipedia)

1.4 Regning med usikre størrelser

Regning med usikre eller stokastiske størrelser er en god del mer komplisert enn regning med deterministiske størrelser. Selve matematikken i dette trenger man i liten grad å forholde seg til, bergningene tar Anslag-programmet seg av. Men det er visse forhold ved bruk av stokastiske overslag som gir føringer for hvordan overslagene bygges opp.

Stokastisk uavhengighet

Når man jobber med stokastiske kostnadsoverslag er det viktig at man har klart for seg prinsippet om stokastisk uavhengighet og hva det innebærer.

To verdier sies å være stokastisk uavhengige hvis utfallet av den ene ikke betyr noe for utfallet av den andre. Det vil for eksempel si at hvis en post i overslaget skulle vise seg i virkeligheten å havne rundt sitt øvre anslag, så skal det fortsatt være (teoretisk) mulig for samtlige andre poster å ende opp på sine nedre anslag. Det er klart at for at dette skal være mulig må man ha urelaterte årsaksforhold. Om en post får en kostnad større enn forventet av en gitt årsak, kan det ikke også føre til at en eller flere andre poster får kostnader større eller mindre enn forventet.

Samvariasjon

Hvis man har to eller flere poster i overslaget som helt eller delvis har samme årsaksforhold, så sier vi at de har en samvariasjon. Det er to mulige måter å håndtere samvariasjon på i Anslag for å sikre korrekt overslagresultat.

Den tradisjonelle måten man har håndtert samvariasjon på i Anslag er å skille ut de felles årsaksforholdene i usikkerhetsfaktorer. For eksempel har man som regel alltid en egen faktor for marked. Det betyr at når man estimerer verdiene for kalkylepostene, så ser man bort fra eventuelle prissvingninger som skyldes markedsforhold. Disse behandles i den egne usikkerhetsfaktoren ved siden av.

Den andre måten å håndtere samvariasjon på er å eksplisitt modellere samvariasjon. Man angir da for eksempel hvor mange prosent post A er samvariert med post B. Dette prosenttallet kan ses på som hvor mange prosent av årsakene til en posts kostnad er felles med årsakene til en annen posts kostnad.

For enkelthets skyld opererer man i Anslag kun med middels (50 %) eller full (100 %) når man angir samvariasjon. Vedlegg 6 gir et eksempel på bruk av samvariasjon i et anslag."

Summering av usikre størrelser

Når man summerer usikre størrelser så er det to egenskaper ved sumfordelingen man bør vite om. Usikkerheten, eller spennet i en sumfordeling, er mindre enn summen av de summerte størrelsenes usikkerheter, forutsatt at de ikke er 100 prosent samvarierte. Dette skyldes at de størrelsene man summerer til en viss grad nuller hverandre ut. Den ene verdien kan bli høy samtidig som den andre kan bli lav.

Det andre punktet som er verdt å merke seg er at en sumfordeling alltid vil bli mindre skjev enn fordelingene som summeres. Jo flere poster som summeres, jo mer symmetrisk vil sumkurven blir.

Sentralgrenseteoremet sier at summen av et stort antall stokastiske verdier med tilfeldige fordelinger tenderer mot å være normalfordelt. Begrensningen er at verdiene er uavhengige variabler, og at ingen av verdiene er svært dominerende i forhold til de andre.

På grunn av dette kan man forvente at resultatkurven i Anslag vil være tilnærmet normalfordelt. Men det skal et veldig høyt antall poster til før man får bortimot 100 prosent normalfordeling. Dette er årsaken til at det alltid er en viss forskjell mellom P50 og forventningsverdien.

Beregning - Monte Carlo-simulering

Det finnes to prinsipielle måter å beregne stokastiske kostnadsoverslag på: ved bruk av analytiske metoder eller ved bruk av Monte Carlo-simulering.

Ved bruk av analytiske metoder beskrives kostnadene for hvert element ved et matematisk uttrykk. Hver av kostnadene regnes så i hop til et nytt uttrykk som gir fordelingen av mulige utfall. Å modellere og regne en kostnadsanalyse ved bruk av eksakte matematiske sannsynlighetsfordelinger og beregningsmetoder er både meget tidkrevende og krever betydelige matematiske ferdigheter, men det er mulig å benytte lineære tilnærmingsformler som gir svar som er tilstrekkelig nøyaktig. Forrige versjon av Anslag-programmet baserte seg bruk av en slik metodikk.

Monte Carlo-simulering er mer en «rå-makt»-metode. I stedet for å regne gjennom kalkylen en gang regner man gjennom den fra noen hundre til flere tusen ganger. Det vil si: man får naturligvis en datamaskin til å gjøre det for seg. For hver gjennomregning trekkes en tilfeldig verdi ut fra sannsynlighetsfordelingene og benyttes i beregningen. Med andre ord: for hver gjennomregning «kaster» datamaskinen en «terning» for å finne ut hvilken verdi den skal bruke for en gitt kostnad. Datamaskinen vil for hver gjennomregning ta vare på de tallene man er interessert i. Etter at simuleringen er kjørt ferdig vil den da ha et statistisk grunnlag for å utarbeide en sannsynlighetsfordeling for sluttsummen og andre tall man måtte ønske å få ut.