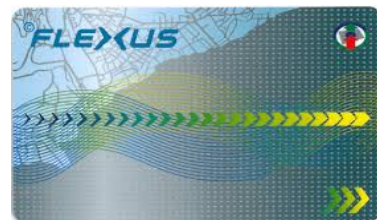




Estimering av IT-utvikling

Magne Jørgensen
Simula Research Laboratory
Universitetet i Oslo
Scientia



Temaer

- Grunnleggende om estimering
 - Hvorfor og når bommer vi på estimatene?
 - Kundens bidrag til underestimering
 - Vet du hva et estimat er?
 - Kan du addere estimater?
- Hvordan bør vi estimere og hva bør vi unngå?

Vi estimerer tid og arbeid hele tiden ...

- Det å "planlegge seg selv" inn i framtiden er en kjerne-egenskap hos mennesker (homo praedico).
- Du har allerede estimert tidsbruk flere ganger i dag! Ca. 10% av tankene er prediksjoner.
- "No estimate"-bevegelsen skaper et feil bilde av hva de er for og i mot:
 - Basert på en snever definisjon av hva som menes med estimat
 - Eksempel fra beskrivelsene hos dere: " Vi estimerer ikke men henter inn en chunk med oppgaver som magefølelsen sier at vi kan klare å bli ferdige med i løpet av en fire ukers sprint. Blir vi ikke ferdige tas oppgavene med over i neste sprint."
- Gode estimater av arbeid er ofte (men ikke alltid!) viktig for lønnsomhetsvurderinger, tilbudsskriving, nyttevurderinger, planlegging, styring av prosjekt (burndown, Earned Value), budsjettering, ressursallokering m.m.
- Noen ganger er estimater skadelige og noen ganger bruker vi altfor mye tid på å utarbeide dem.

[**simula** . research laboratory]

Hvor gode er IT-bransjen til å estimere?

- Stort sett så er vi ikke så aller verst.
- Trolig er gjennomsnittlig overskridelse på 20-30%. Ingen eller kun liten bedring over tid.
- Andelen store overskridelser ("black swans", mer enn dobling av kostnader) er 10-20%.
- Det er forskjell på de "skumle" og de "normale" prosjektene. Vi bør bli bedre til å anslå hvor skummelt et prosjekt er og legge mer estimeringsressurser i disse.
- Ca 10% av alle prosjekter blir aldri fullført eller fullfører uten å levere særlig nytte.
- Underveis-estimater (sprint m.m.) og forvaltningsjobber estimeres stort sett nokså nøyaktig. Disse har, i motsetning til prosjektestimater, en tendens til å være for høye! Mao, ingen generell tendens til overoptimisme.

[**simula** . research laboratory]

Oppgave: Angi to enkle metoder der du alltid vil treffe 100% på estimatene?

Hint 1: Hvordan kan man sørge for at for høye estimater blir korrekte?

Hint 2: Spesifikasjoner er aldri fullstendige. Hvilken innvirkning kan de å øke fleksibiliteten på spesifikasjonen ha på estimeringsnøyaktigheten?

Metode 1: Estimer skyhøyt, bruk opp all tiden til finpuss og kvalitetsøkning.

Metode 2: Legg inn så mye fleksibilitet i spesifikasjonen av hva som skal leveres at prosjektet kan stoppe når det vil og likevel være ferdig.

Hører du om noen som alltid treffer med estimatene kan du anta at de bruker en av disse metodene. Særlig Metode 1 er uheldig for kundene.

[**simula** . research laboratory]

Når er det vanskeligst å estimere realistisk?

- Vi gjør ting som er grunnleggende forskjellig fra det vi har gjort før
- Mange grensesnitt til andre systemer og/eller mange interessenter
- Involverer stor grad en prosessendring
- Problemene som skal løses er komplekse
- Urealisme i ambisjoner
- Situasjoner som stimulerer til over-optimisme
- Lav kompetanse til kunde og/eller leverandør – eller stor usikkerhet i grad av kompetanse (indikatorer: lite relevant erfaring, eller tidligere høy fiaskorate)
- Manglende oppfølging, støtte, og ledelsesforankring hos kunde eller leverandør
- Dårlig kommunikasjon, manglende opptatthet av ulik kultur
- Prisfokuserte anbudsprosesser ("winner's curse" og "adverse selection")
- Uegnete utviklingsmodeller (særlig relatert til oppdeling av leveranse)

[**simula** . research laboratory]

NOEN GRUNNER TIL AT VI ESTIMERER FEIL

[**simula** . research laboratory]



Hjelper det å være oppmerksom på vurderingsfeil? McKinsey-rapport fra 2010



- Undersøkelse av 1048 strategiske vurderinger og beslutninger i ulike firma
- De beste 25% firmaene mhp å unngå og redusere vurderings og beslutningsskjøvheter hadde et overskudd syv (!) ganger høyere enn de dårligste 25%.
- Det å unngå og redusere vurderings og beslutningsskjøvheter var seks (!) ganger viktigere for overskuddet enn graden av eller detaljeringen av analyse forut for beslutningene.
- Lovallo, Dan, and Olivier Sibony. "The case for behavioral strategy." *McKinsey Quarterly* 2 (2010): 30-43. www.edpiccolino.com/workspace/articles/mckinsey-the-case-for-behavioral-strategy.pdf

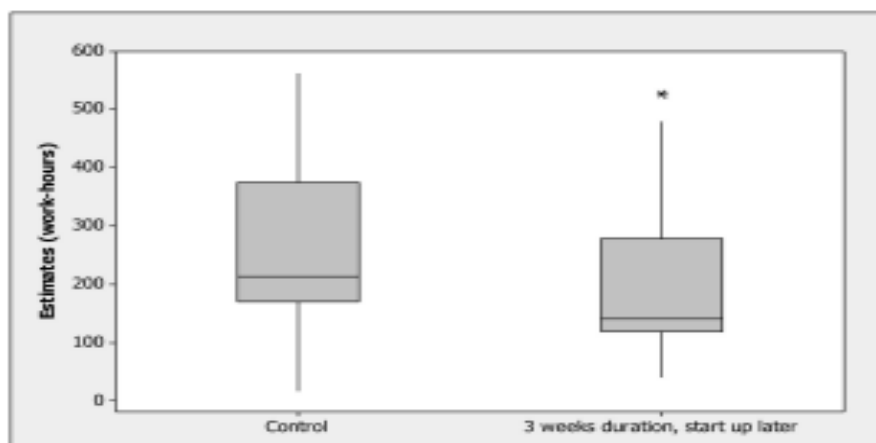
[**simula** . research laboratory]

Overvurdering av egne evner og undervurdering av risiko!



[[simula](#) . research laboratory]

Ønsketenkning: Manglende skille mellom ”håper dette tar ...” og ”dette vil ta ...”



[[simula](#) . research laboratory]

Illusjon av kontroll når vi visualiserer fremtiden

- Planlegging (visualisering av fremtidsscenarier) skaper ofte overoptimisme.
- Overvurdering av hvor mye vi kan kontrollere/håndtere er en viktig kilde til denne overoptimismen ("illusion of control")
- Analogi-basert estimering (se tilbake på historiske data) gir av den grunn ofte mer realisme

[**simula** . research laboratory]

Ankereffekter

Eksperiment:

- HIGH (LOW) group: *"The customer has indicated that he believes that **1000 (50)** work-hours is a reasonable effort estimate for the specified system. However, the customer knows very little about the implications of his specification on the development effort and you shall not let the customer's expectations impact your estimate. Your task is to provide a realistic effort estimate of a system that meets the requirements specification and has a sufficient quality."*
- Deltakere: Erfarne systemutviklere.
- Alle (HIGH, LOW, CONTROL) fikk samme kravspesifikasjon.

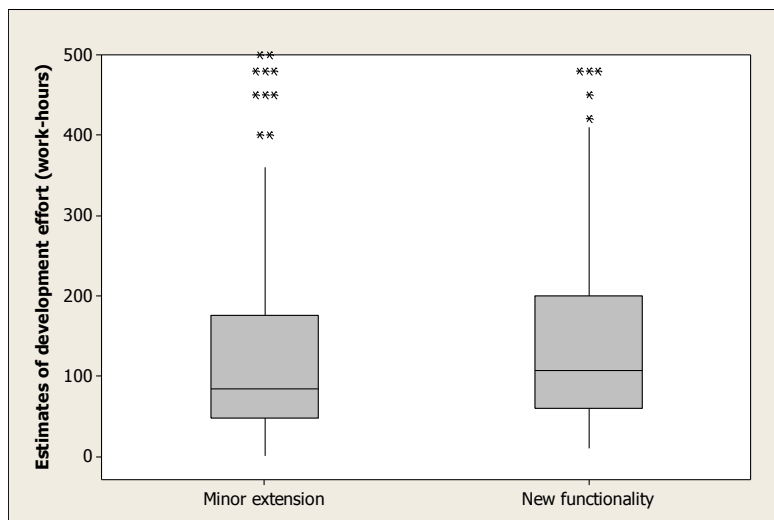
[**simula** . research laboratory]

12

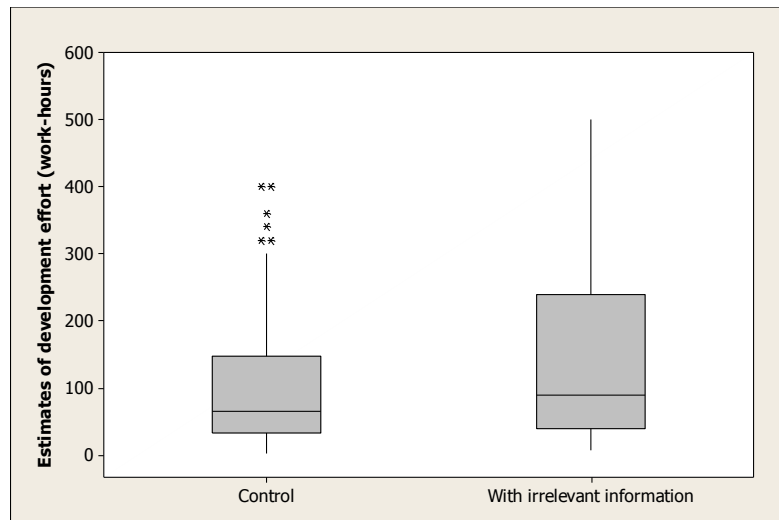
Ankereffekter

- Resultater:
 - HIGH group gjennomsnitt: 555 timeverk
 - CONTROL group (uten forventinger) gjennomsnitt: 456 timeverk
 - LOW group gjennomsnitt: 99 timeverk!!!
- Ingen av utviklerne opplevde at de hadde blitt mye påvirket. De fleste mente at de ikke var påvirket i det hele tatt av kundens forventninger.
- Opplæring i og oppmerksomhet rundt ankereffekter reduserer men fjerner ikke effekten! Eneste som virker er å unngå ankere.

Ladede ord kan fungere som anker



Irrelevant informasjon påvirker estimatet



[**simula** . research laboratory]

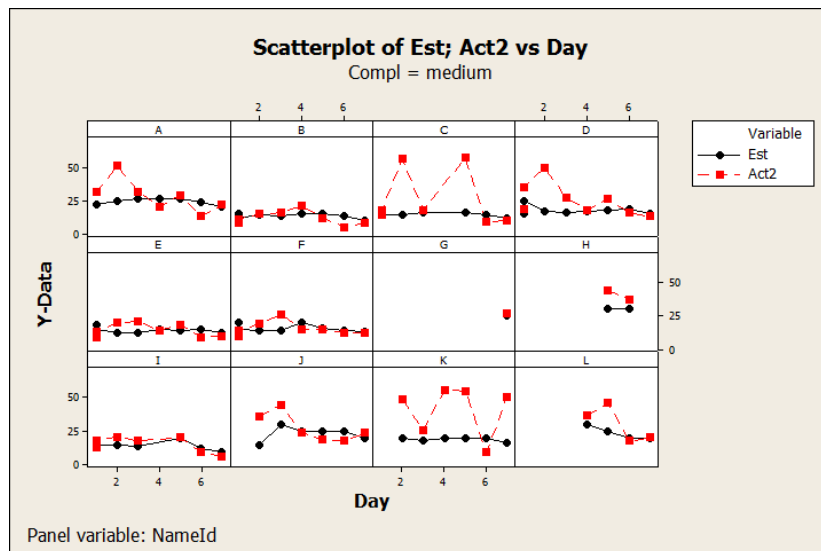
Gjør det forskjell om vi estimerer i timeverk eller dagsverk?

- Helt fersk studie på dette nedenfor:

	Most likely effort (mean)
Estimates in workdays, $n=29$	177 work-hours (std. dev. 254)
Estimates in work-hours, $n=19$	72 work-hours (std. dev. 57)
Difference between Groups	105 work-hours (59% decrease)
T-test of difference (p -value)	0.02 ¹

[**simula** . research laboratory]

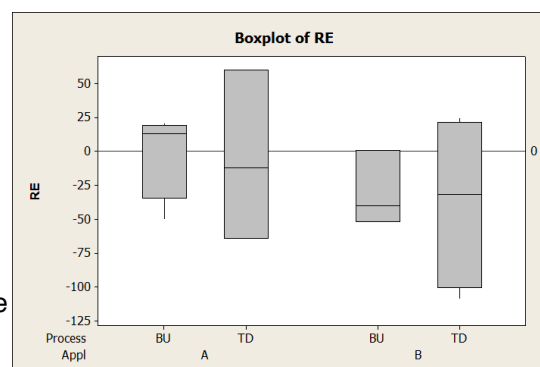
Estimat synes typisk å gjenspeile “Hvis ingen ting går galt”



[[simula](#) . research laboratory]

Top-down eller bottom-up estimering

- Ja takk, begge deler.
- Studie viste at dersom gode analogier ble funnet, så var top-down den beste.
- I gjennomsnitt var imidlertid bottom-up best.
- Overraskende funn: Estimeringsteamene var relativt dårlige i å finne og bruke lignende prosjekter.
- Kan være stort potensiale for opplæring og bruk av relativ estimering på prosjektnivå – a la relativ estimering på user story nivå i agile.



[[simula](#) . research laboratory]

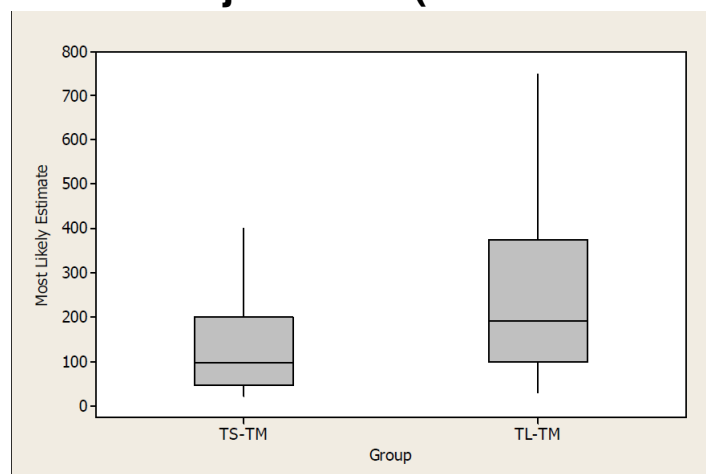
Relativ estimering: Lønner det seg å hoppe etter Wirkola?

SVAR: JA



[**simula** . research laboratory]

Asimilasjonseffekt (Sekvenseffekt)



TS-TM: Estimerte først et lite system (ca. 25 tv), så det middels store.

TL-TM: Estimerte først et relativt sett stort system (ca. 500 tv) så det middels store.

[**simula** . research laboratory]

Viktighet av hvilken vei vi sammenligner: Kan A være mer lik B, enn B er lik A?

- Deltagere: Ca. 100 utviklere fra Ukraina. Stor oppgave B, liten oppgave A.
- Reell forskjell i arbeidsmengde (B-A) var på minst 100 timeverk.
- Alternativ 1: Hvor mye større er B enn A? (A er referanse)
 - Svar (gjennomsnitt): 80 timeverk
- Alternativ 2: Hvor mye mindre er A enn B? (B er referanse)
 - Svar (gjennomsnitt): 40 timeverk
- Alternativ 3: Estimer A, så B (A blir trolig referanse for B)
 - Svar (gjennomsnitt): Diff. på 77 timeverk
- Alternativ 4: Estimer B, så A (B blir trolig referanse for A)
 - Svar (gjennomsnitt): Diff. på 68 timeverk

Mao, forskjellen blir større når vi ser en stor oppgave i forhold til en liten, enn motsatt. I tillegg er det en stor assimileringseffekt, dvs ting blir likere når de sammenlignes.

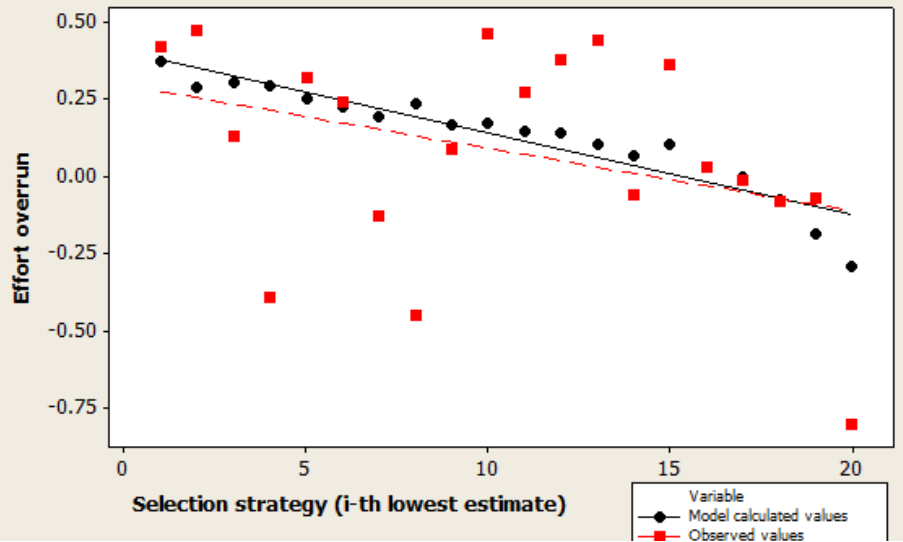
[[simula](#) . research laboratory]

Konsekvenser for relativ estimering (som f eks ved bruk av story points eller analogi-basert estimering)

- **Sammenligning gjør ting stort sett likere enn de er**
 - Bruk middels store oppgaver som “baseline” i Story Point-basert estimering
 - Bruk prosjekter av samme størrelse som referanse i tidligfase-estimering
- **Lite kunnskap om det du skal estimere gjør at nytt prosjekt ligner på referansen**
 - Neglisjering av egenskaper som finnes hos referansen, men ikke hos det du skal estimere

[[simula](#) . research laboratory]

The winner's curse



[simula . research laboratory]

Kunder velger over-optimistiske, mindre kompetente leverandører

- "Winner's curse"
 - Vektlegging av pris + mange tilbydere: "Vi vinner nesten bare når vi har vært over-optimistiske".
 - Kan gjøre gode leverandører dårlige
- Manipulerende informasjon
 - Budsjettinformasjon (Lånekassens IT-system hevdes å ha blitt estimert av stortingspolitikere, som bestemte budsjettet).
 - Bruk av ladede ord ("lite system")
- "Adverse selection"
 - Jo mindre kompetanse hos leverandør, jo lavere estimat og pris
 - Jo mindre kompetanse hos kunde, jo mer vektlegging av lav pris
 - Mye av prosjektproblemene og overskridelse skyldes prisfokus i valg av leverandør.

[simula . research laboratory]

24

Fastpris er dårlig både for budsjettkontroll og nytte (andel vellykkede prosjekter)

	Per time	"Smidig"	Risikodeling	Fastpris
Nytte	59%	29%	22%	0%
Kvalitet	24%	43%	22%	22%
Budsjett	31%	71%	22%	33%
Tid	29%	43%	44%	11%
Effektivitet	19%	29%	33%	0%
Andel	37%	14%	41%	18%

Table 10: Project success and contract type

Success dimensions	Increase in success rate (percentage points) for projects applying per-hour rather than fixed-price contracts
Client benefit (n=49)	34%
Functionality (n=51)	5%
Technical quality (n=51)	11%
Budget control (n=50)	14%
Delivery on time (n=50)	3%
Work efficiency (n=49)	10%

Kategoriserte smidig som per time og risikodeling som fastpris.

VET DU HVA ET ESTIMAT ER?

[**simula** . research laboratory]

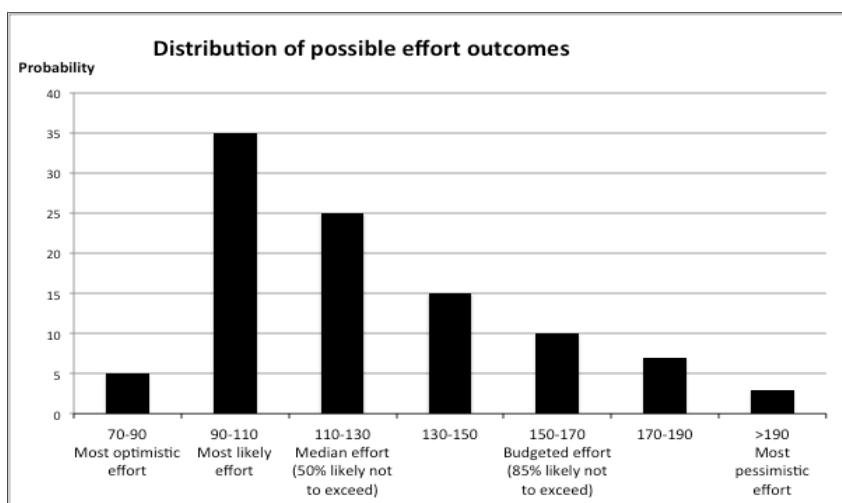
Estimat arbeid/kostnad brukes om alt dette:

- Arbeidsmengde gitt at ingen vesentlige ting går galt og vi får jobbet uforstyrret (ideelle timer)
- Mest sannsynlig arbeidsmengde
- p50-estimat (like sannsynlig å bruke mer som å bruke mindre)
- Mest sannsynlig arbeidsmengde + tillegg for usikkerhet
- Planlagt bruk av timeverk
- Antall timeverk lagt til grunn for pris
- Tilbudet til kunden
- Det som "føles" riktig ... (magefølelse, intuisjon, ...)
- Vet ikke helt hva jeg/vi har ment

VI TRENGER Å BLI BEDRE PÅ Å VITE OG KOMMUNISERE ESTIMATER!

[**simula** . research laboratory]

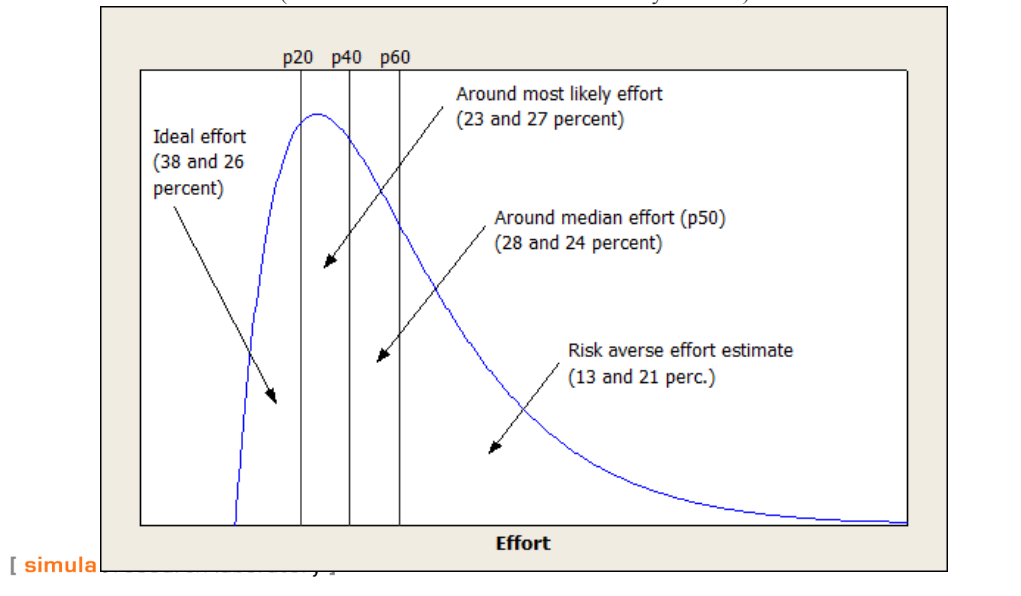
Den eneste presise måten å snakke om estimater på er (så vidt jeg vet) sannsynlighetsbasert



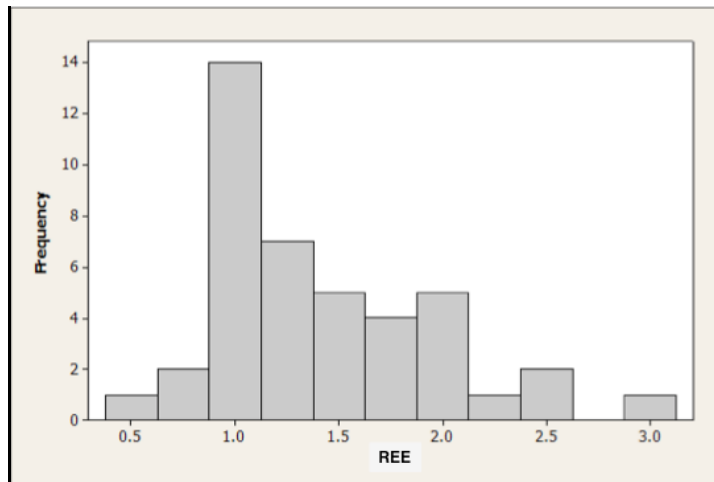
[**simula** . research laboratory]

Resultater fra en undersøkelse tyder på stort sprik i hva man mener med estimat ...

(44 firma som estimerte de samme to systemene)

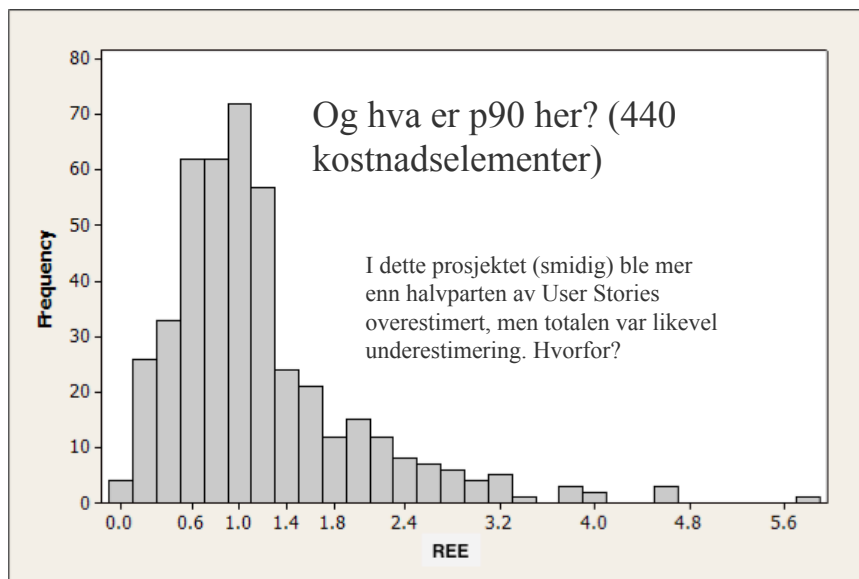


Eksempel på fordeling av estimeringsfeil for 42 prosjekter (1.0 er 100% treff, 1.5 = 50% overskridelse, etc.)



Oppgave: Hva er nødvendig påslag for å finne p90?

[simula . research laboratory]

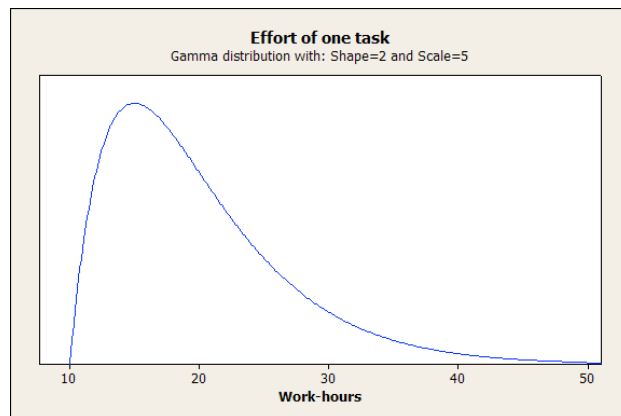


[**simula** . research laboratory]

KAN DU ADDERE ESTIMATER?

[**simula** . research laboratory]

Mest sannsynlig arbeidsmengde for en oppgave er 15 timeverk. Hva er totalestimatet for 10 oppgaver?



Svar: ca. 200 timeverk. Kun 1% sannsynlig å bruke 150 eller færre timeverk.

[**simula** . research laboratory]

Addisjon av estimater er ikke enkelt

- Å legge sammen estimater er (statistisk sett) å legge sammen "stokastiske variable"
- Summen av mest sannsynlig arbeidsmengde for kostnadselementer gir, for "høyre-skjeve" fordelinger (som de vi så tidligere her), et estimat det er SVÆRT SANNSYNLIG å overskride!
- Feilaktig summering er ofte en årsak til for lave estimater.
- **NB: Du kan bare adderer "aritmetisk middel", dvs forventningsverdiene**

[**simula** . research laboratory]

Matematikken bak WBS + PERT

Du kan kun addere gjennomsnitt (aritmetisk middel) of varians når du har stokastiske variable som ikke er symmetriske.

$$E[X + c] = E[X] + c$$

$$E[X + Y] = E[X] + E[Y]$$

$$E[aX] = a E[X]$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

Sentralgrenseverditeoremet: Under visse betingelser, blant annet tilstrekkelig mange verdier, gir summen av skjeve fordelinger en normalfordeling.

[**simula** . research laboratory]

Hvordan finne “mean” (forventingsverdi)

- Bestem “mest sannsynlig”
- Bestem p10 og p90 (mer om det senere)
- Beregn forventingsverdi som:
 - $(p10 + 4 \times m + p90)/6$
- Beregn standardavvik som (NB: Fakoren 2.65 er **ikke** standard, men mer korrekt)
 - $(p90 - p10)/2.65$

[**simula** . research laboratory]

Hvordan finne pX-estimer?

- Vanlig metode er å spørre “Hva er minimum/mest optimistisk/ best case ...?”
- Våre undersøkelser viser at dette medfører at intervallene blir for smale (60-70% hit rate) og at vi totalt ignorerer om vi skal være 98%, 90%, 70% eller 60% sikre (alle gir samme intervallbredde).
- Bedre metode:
 - Se tilbake på estimeringsfeil for lignende oppgaver.
 - Motvasjon: Å se tilbake gir økt realisme. Vi har et mer realistisk forhold til frekvenser. Historisk feil på lignende oppgaver er en god (den beste) indikator på estimatusikkerhet.

[**simula** . research laboratory]

Øvelse: Historisk estimeringsfeil for Knowit

Faktisk arbeidsmengde	Frekvens
Minst 50% lavere enn estimert	
Minst 25% lavere enn estimert	
Ca. likt eller lavere enn estimert	
Mindre enn 25% overskridelse	
Mindre enn 50% overskridelse	
Mindre enn 75% overskridelse	
Mindre enn 100% overskridelse	
Mindre enn 200% overskridelse	

$$\text{Median} = X \times MS$$

$$\text{Aritmetisk middel} = Y \times MS$$

$$p90 = X \times MS$$

[**simula** . research laboratory]

38

Hvem er gode til å estimere? Indikatorer

- Lengden på erfaring?
 - Svært dårlig indikator
- Erfaring fra lignende prosjekter?
 - Absolutt, men husk at ekspertise er smalere enn vi typisk antar
- Den beste utvikleren?
 - Ikke alltid. De dyktigste utviklerne tenderer til å underestimere dersom de estimerer arbeidet for “noviser”
 - “Utsidesyn” kan være bedre (ikke-teknologer er ofte forbløffende gode på å estimere, uten selv å kunne lage systemet)

Indikatorer:

- Den som er mest sikker på estimatet?
 - Dårlig indikator. Ofte det motsatte.
- De som har best “treff-record”?
 - God, men ikke veldig sikker indikator.
- Personlighetstype?
 - Ikke mye hjelp. Ikke en gang tester på optimisme forbedrer valget av eksperter i særlig grad.
- Best estimeringsprosess
 - Ja, til en viss grad. Strukturering, sjekklister og refleksjon rundt egne ekspertvurderinger hjelper.

Hvordan bedre estimeringsevnen?

- Sørg for å "aktivere" relevant erfaring:
 - Strukturering av informasjonen og prosesser.
 - God tilgang på relevante historiske data. Eksempel: Velocity i smidige prosjekter.
- Sørg for å fjerne irrelevant og villedende informasjon
 - Irrelevant informasjon kan føre til feil analogier/assosiasjoner/aktivering
 - Intuisjonen kan finne på å vekte all tilgjengelig informasjon noe, selv om vi vet den ikke er relevant
- Skap estimeringssituasjoner der realisme er ENESTE mål
 - Skreddersy situasjonen slik at angivelse av mest sannsynlig kostnad er eneste mål.
 - Unngå at noen som er "forelsket i prosjektet" eller har vært utsatt for estimeringsankere er med på estimeringsarbeidet.
 - Vit hva du estimerer og hva du skal bruke det til. Pris (mål = vinne anbudsrunder), budsjett (mål = kostnadskontroll), plan (mål = prosjektstyring) eller mest sannsynlig kostnad (mål = realisme).

Hvordan bedre estimeringsevnen?

- Dersom tidligere erfaringer ikke er særlig relevante bør grad av analyse økes
 - Intuisjonen er da ekstra lite å stole
 - Øk bruken av estimeringsregler/modeller
 - Innhent ekspertise utenfra (se hvem er eksperter).
- Sørg for feedback og gjør årsaksanalyser både ved nøyaktige og unøyaktige estimater.
 - Lær deg å gjenkjenne situasjoner der du kan stole på din og andres intuisjon.
- Kombiner metoder (triangulering)
 - Både top-down (analogi) og bottom-up (WBS)
 - Ekspert + enkel modell

SLUTTEN ER HER

[**simula** . research laboratory]