

14. september 2020

STK3405

Obligatorisk oppgave 2 av 2

Innleveringsfrist

Torsdag 5. november 2020, klokken 14:30 i Canvas (www.canvas.uio.no).

Instruksjoner

Du velger selv om du skriver besvarelsen for hånd og scanner besvarelsen eller om du skriver løsningen direkte inn på datamaskin (for eksempel ved bruk av L^AT_EX). Besvarelsen skal leveres som én PDF-fil. Scannede ark må være godt lesbare. Besvarelsen skal inneholde navn, emne og obliqnummer.

Det forventes at man har en klar og ryddig besvarelse med tydelige begrunnelser. Husk å inkludere alle relevante plott og figurer. Studenter som ikke får sin opprinnelige besvarelse godkjent, men som har gjort et reelt forsøk på å løse oppgavene, vil få én mulighet til å levere en revidert besvarelse. Samarbeid og alle slags hjelpemidler er tillatt, men den innleverte besvarelsen skal være skrevet av deg og reflektere din forståelse av stoffet. Er vi i tvil om du virkelig har forstått det du har levert inn, kan vi be deg om en muntlig redegjørelse.

I oppgaver der du blir bedt om å programmere må du legge ved programkoden og levere den sammen med resten av besvarelsen. Det er viktig at programkoden du leverer inneholder et kjøreeksempel, slik at det er lett å se hvilket resultat programmet gir.

Søknad om utsettelse av innleveringsfrist

Hvis du blir syk eller av andre grunner trenger å søke om utsettelse av innleveringsfristen, må du ta kontakt med studieadministrasjonen ved Matematisk institutt (e-post: studieinfo@math.uio.no) i god tid før innleveringsfristen.

For å få adgang til avsluttende eksamen i dette emnet, må man bestå alle obligatoriske oppgaver i ett og samme semester.

For fullstendige retningslinjer for innlevering av obligatoriske oppgaver, se her:

www.uio.no/studier/admin/obligatoriske-aktiviteter/mn-math-oblig.html

LYKKE TIL!

Spesielt om det obligatoriske oppgavesettet i STK3405

I denne oppgaven skal vi benytte simuleringsprogrammet RISCUETM. Dette kan lastes gratis ned fra www.riscue.org. Velg plattform (Windows, Mac eller Unix) ved å klikke på ikonene på venstre side, og følg instruksjonene på websiden¹. For å få godkjent besvarelsen, må du ha minst 65% riktig.

Oppgave 1. En fordeling som benyttes mye i risikoanalyse er *lognormalfordelingen*. En stokastisk variabel, Y , sies å være lognormalt fordelt med parametre μ og σ hvis $X = \ln(Y)$ er normalfordelt med forventning μ og standardavvik σ . Vi skriver i så fall $Y \sim \text{lognorm}(\mu, \sigma)$.

(a) Anta at $U \sim N(0, 1)$, dvs. at U er standard normalfordelt, og la $Y = \exp(\sigma U + \mu)$. Forklar kort hvorfor det da følger at $Y \sim \text{lognorm}(\mu, \sigma)$.

(b) Anta at $Y \sim \text{lognorm}(\mu, \sigma)$, og la y_α være slik at:

$$P(Y \leq y_\alpha) = \alpha\%, \text{ der } 0 \leq \alpha \leq 100.$$

Med denne notasjonen er altså f.eks. y_{10} lik 10%-percentilen i $\text{lognorm}(\mu, \sigma)$ -fordelingen. Vis at:

$$y_\alpha = \exp(\sigma u_\alpha + \mu),$$

der u_α er α %-percentilen i standard normalfordelingen.

(c) Vis at:

$$\frac{y_{90}}{y_{50}} = \frac{y_{50}}{y_{10}}$$

(d) Anta at $y_{10} = 5.0$ og at $y_{90} = 20.0$. Finn den tilhørende verdien av y_{50} . Bestem også verdiene på μ og σ for denne fordelingen. Du kan her benytte at $u_{90} = 1.28155$.

[Løsning: $\mu = 2.30259$, og $\sigma = 0.54087$.]

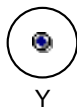
(e) Benytt RISCUETM til å lage en simuleringsmodell bestående av kun én node. Kall denne noden Y . (Se Figure ?? under.) Som funksjon i denne noden benyttes:

$$\text{EXP}(\text{SUM}(\text{PRODUCT}(\text{NORMAL}()); 0.54087); 2.30259) \quad (1)$$

Kjør en simulering på denne modellen. Benytt f.eks. 10000 iterasjoner. Lag et *S-kurve*-plot og et tetthetsplot for resultatnoden. Kontrollér at 10%-percentilen og 90%-percentilen for resultatnoden blir (omtrent) det ønskede (dvs. henholdsvis 5.0 og 20.0). Lag også et plot som viser hvordan disse percentilverdiene konvergerer. Gi en kort kommentar til resultatene.

Vi skal nå se hvordan vi kan konstruere en tidsrekkemodell der alle variablene er lognormalt fordelte. For å konstruere denne starter vi først med en rekke

¹Hvis du har en Mac med OS 10.8 (Mountain Lion) eller nyere, må du skru av sperren som hindrer installering av programmer som ikke er forhåndsgodkjent av Apple før du installerer programmet. Se detaljert beskrivelse av hvordan du går frem på www.riscue.org.



Figur 1: Simuleringsmodell med kun én node.

$\{W_n\}$ av uavhengige standard normalfordelte variable. Basert på denne rekken danner vi så en ny rekke $\{U_n\}$ som følger:

$$U_1 = W_1, \quad U_n = a \cdot U_{n-1} + b \cdot W_n, \quad n = 2, 3, \dots$$

Vi bemerker at siden U_{n-1} inngår i uttrykket for U_n , $n = 2, 3, \dots$, så blir variablene i denne rekken stokastisk avhengige.

(f) Bestem konstantene a og b slik at $U_n \sim N(0, 1)$ og $\text{Corr}(U_n, U_{n-1}) = c$, $n = 2, 3, \dots$, der c er en passende konstant slik at $|c| \leq 1$.

[Løsning: $a = c$ og $b = \sqrt{1 - c^2}$.]

(g) Vi skal nå benytte RISCUETM til å lage en simuleringsmodell bygget opp som i punktet over. Før du starter å lage denne modellen, kan det være lurt å lukke alle åpne vinduer i RISCUETM. Start deretter med å opprette en ny tom modell ved å benytte meny-kommandoen: *File>New*. Du får da opp et blankt vindu. Velg deretter meny-kommandoen *File>Save As...*, og gi modellfilen navnet *STK3405_oblig2*.

For at vi på en enkel måte skal kunne variere korrelasjonen c , skal vi la denne være representert i modellen som en parameter. En slik parameter opprettes ved først å åpne parametervinduet ved å benytte meny-kommandoen: *Model>Parameter table...* Deretter oppretter vi en parameter ved å benytte meny-kommandoen: *Model>Add Parameter*. Til slutt fyller vi inn informasjon i de ulike kolumnene slik at vi ender opp som vist i Figure ??.

Name	ID	Def. value	Values	Multivalued	Info
Tidskorrelasjon	0	0.5	0.1; 0.2; ... ; 1.0	<input type="checkbox"/>	Korrelasjon mellom påhverandre følgende ledd

Figur 2: Parameter table.

Tip: Vær spesielt nøye med hva du taster inn i kolonnen *Values*. Som skilletegn mellom tallene skal det benyttes semikolon. Mellom 0.2 og 1.0 skal det stå: *semikolon, blank, punktum, punktum, punktum, blank, semikolon, blank*.

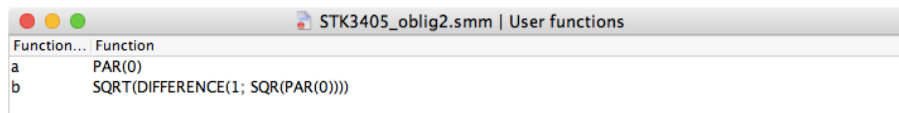
Tip: Legg merke til at det i tittelfeltet på vinduet med parametertabellen står:

STK3405_oblig2.smm | Parameter data

Første delen av tittelfeltet angir altså navnet på modellfilen du jobber med i øyeblikket. Hvis man jobber med flere modellfiler samtidig, vil meny-kommandoen *Model>Parameter table...* alltid åpne parametertabellen som hører til det modellvinduet som ligger øverst. I slike situasjoner kan det være lurt å forvise seg om at man faktisk redigerer den *riktige* parametertabellen, og det gjør man altså ved å sjekke hva som står i tittelfeltet.

Bemerk at verdien i kolonnen *ID column*, dvs. 0, blir fylt inn automatisk. Denne verdien benyttes når vi skal benytte parameteren i formler. Vi kan da bruke formelen $\text{PAR}(0)$, der argumentet er nettopp denne ID-verdien. Avkrysningsfeltet i kolonnen *Multivalued* skal foreløpig ikke stå avkrysset, hvilket betyr at modellen skal benytte parameterverdien som står i kolonnen *Def. value*, dvs. 0.5 i simuleringene.

For å gjøre modellen vår litt lettere å lese, skal vi også lage noen brukerdefinerte funksjoner. Vi åpner derfor vinduet med slike funksjoner ved å benytte meny-kommandoen *User>User Functions*. Deretter oppretter vi to slike funksjoner ved å benytte meny-kommandoen *User>Add User Function*, og fyller inn detaljene slik at vi til slutt ender opp som i Figure ??.



Figur 3: User function table.

Tip: Legg merke til at det i tittelfeltet på vinduet med brukerdefinerte funksjoner står:

STK3405_oblig2.smm | User functions

Akkurat som for vinduet med parametertabellen, angir første delen av tittelfeltet navnet på modellfilen du jobber med i øyeblikket. Hvis man jobber med flere modellfiler samtidig, er det igjen viktig å forvise seg om at man har åpnet riktig brukerfunksjonstabell. Dette styres igjen av hvilket modellvindu som ligger øverst når man åpner vinduet med brukerfunksjonstabellen.

Vi legger merke til at siden $\text{PAR}(0)$ refererer til korrelasjonen, dvs. c , så blir de to brukerdefinerte funksjonene, kalt henholdsvis **a** og **b**, nettopp det vi fant som uttrykk for konstantene a og b over.

Vi kan nå lage vår tidsrekkemodell som en kjede av $n = 12$ noder som illustrert i Figure ??.



Figur 4: A time series model.

For node **U01** benyttes følgende formel:

$$\text{NORMAL} () \tag{2}$$

n	10%	90%
1	5.0	8.2
2	8.0	11.0
3	10.0	12.5
4	10.0	12.5
5	10.0	12.5
6	8.0	11.0
7	6.5	10.0
8	5.0	8.2
9	3.3	7.1
10	2.5	5.5
11	2.1	4.6
12	1.7	3.1

Tabell 1: Percentiler for Y_1, \dots, Y_{12} .

For alle de øvrige nodene benyttes følgende formel:

$$\text{SUM}(\text{PRODUCT}(\mathbf{a}()); \text{INPUT}()); \text{PRODUCT}(\mathbf{b}()); \text{NORMAL}()) \quad (3)$$

Tip: For å spare tid er det lurt å taste inn formelen (??) i node **U02**, og deretter kopiere denne formelen over til nodene **U03**, ..., **U12**. For å gjøre dette, går du frem på følgende vis. Legg først formel (??) inn i **U02**. Høyre-klikk så på node **U02**, og velg *Copy formula* fra pop-up menyen. Velg deretter alle de etterfølgende nodene **U03**, ..., **U12** ved å dra en seleksjonsfirkant rundt dem (hold museknappen nede mens du flytter markøren slik at du drar uten firkant som omslutter alle nodene). Mens nodene **U03**, ..., **U12** er valgt, høyre-klikker du så på én av dem (det er likegyldig hvilken node du klikker på). Velg så *Paste formula* fra menyen.

Kjør en simulering på denne modellen. Lag en korrelasjonstabell for resultat-nodene og kontrollér at korrelasjonen mellom to påhverandre følgende ledd er omtrent 0.5. Kommentér resultatene.

(h) Vi skal nå transformere tidsrekken fra punktet over til en tidsrekke, $\{Y_n\}$ der alle variablene er lognormalt fordelte men ikke nødvendigvis identisk fordelte. Siden vi vet at $U_n \sim N(0, 1)$, $n = 1, 2, \dots, 12$, kan dette oppnås ved å la:

$$Y_n = \exp(\sigma_n \cdot U_n + \mu_n), \quad n = 1, 2, \dots, 12.$$

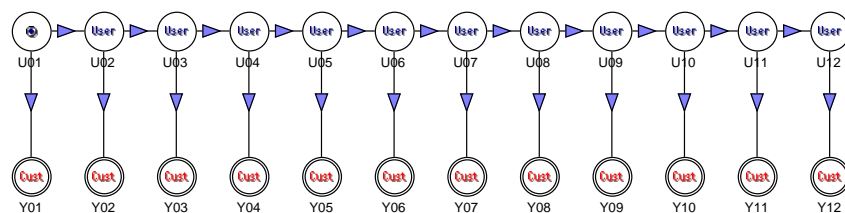
der σ_n og μ_n er beregnet basert på 10%- og 90%-percentilene til Y_n , $n = 1, 2, \dots, 12$, gitt i Table ??.

Beregn σ_n og μ_n , $n = 1, 2, \dots, 12$, basert på informasjonen i Table ?? ved å benytte resultatet fra punkt (b).

Ta så utgangspunkt i modellen i Figure ??, og lag en ny rad med noder under den første. kall de nye nodene henholdsvis **Y01**, ..., **Y12**. Trekk så piler fra **U01** til **Y01**, fra **U02** til **Y02**, osv. til du får et diagram som vist i Figure ??.

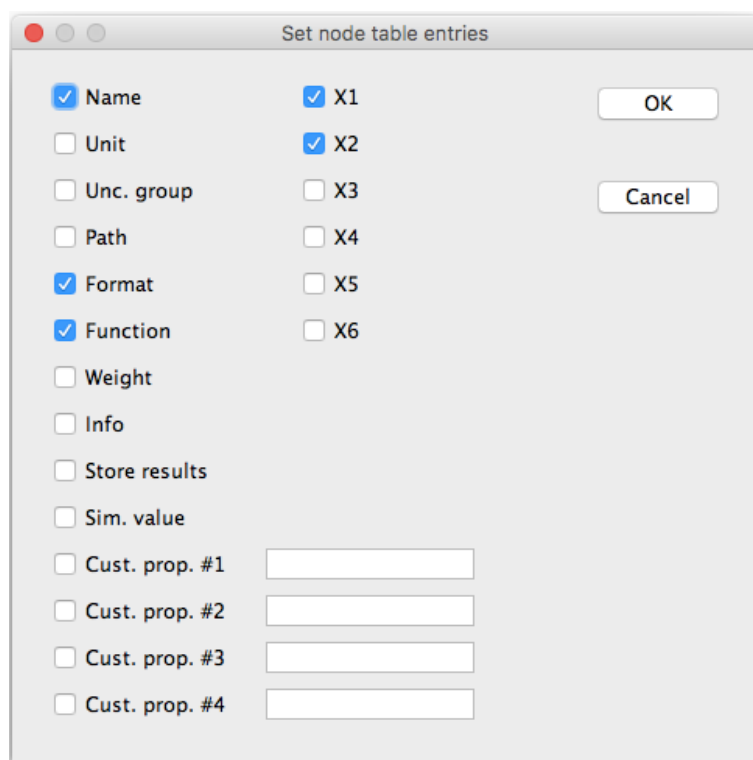
For alle de nye nodene benyttes en formel av form:

$$\text{EXP}(\text{SUM}(\text{PRODUCT}(\mathbf{X2}()); \text{INPUT}()); \mathbf{X1}()) \quad (4)$$



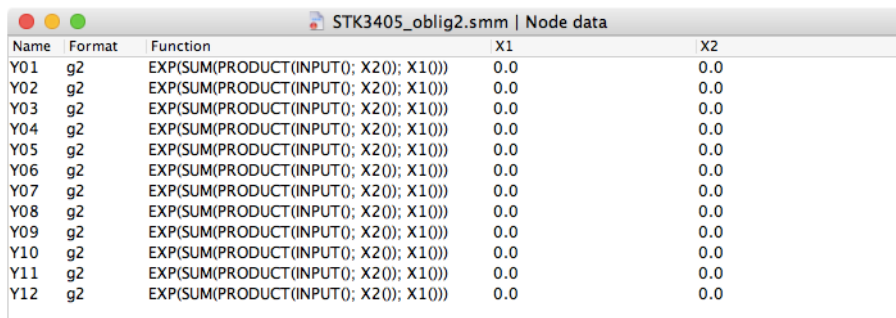
Figur 5: Two linked timeseries models.

der $X1()$ og $X2()$ er funksjoner som refererer til verdiene for henholdsvis μ_n og σ_n , beregnet basert på Table ???. Returverdiene til disse funksjonene må legges inn i hver enkelt node som følger. Start med å velge meny-kommandoen *Model>Set Node Table Entries....* Du skal da få opp en dialogboks som den som vises i Figure ??. Sørg så for at avkrysningsboksene er satt opp vist i figuren. Pass spesielt på at $X1$ og $X2$ er avkryssset.



Figur 6: The *Set Node Table Entries...* dialog box.

Klikk så på **OK**-knappen for å lukke dialogboksen. Deretter selekterer du nodene $Y01, \dots, Y12$, and velger meny-kommandoen *Model>Node Table*. Du får da opp en tabell med data for nodene $Y01, \dots, Y12$. I kolonnene merket



Name	Format	Function	X1	X2
Y01	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y02	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y03	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y04	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y05	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y06	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y07	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y08	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y09	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y10	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y11	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y12	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0

Figur 7: Node table.

X1 og X2 kan du så taste inn verdiene for μ_n og σ_n for de 12 nodene.

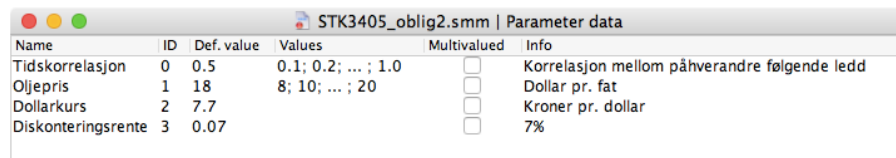
Etter at alle verdiene er tastet inn, kan du kjøre en simulering på denne modellen. Lag en korrelasjonstabell for resultatnodene for de nye nodene. Synes du det ser ut til at korrelasjonen mellom to påhverandre følgende noder fortsatt er omtrent 0.5? Kommentér resultatene.

Tip: For å unngå å få så mange resultatnoder, kan du med fordel la være å lagre resultater for de gamle nodene, dvs. **U01**, \dots , **U12**. Dette gjøres ved å selektere alle disse nodene, høyre-klikke på én av dem, og velge *Don't store results* fra menyen som dukker opp.

(i) Tidsrekken $\{Y_n\}$ representerer årlig produksjon (angitt i millioner fat) fra et oljefelt med "økonomisk levetid" på 12 år. For å finne inntekten (angitt i millioner norske kroner) fra feltet i denne perioden, må vi multiplisere med oljeprisen (angitt i dollar pr. fat) og dollarkurs (angitt i kroner pr. dollar). Både oljepris og dollarkurs er naturligvis usikre størrelser. Det er imidlertid vanlig å legge disse størrelsene inn i modellene som parametre slik at man kan foreta sensitivitetsanalyser.

I tillegg vil man typisk "neddiskontere" inntekter som kommer i fremtiden med en passende fast rentesats. Resultatet av dette omtales gjerne som "nåverdien" av inntektene.

Opprett parametre for oljepris, dollarkurs og diskonteringsrente som vist i Figure ??.



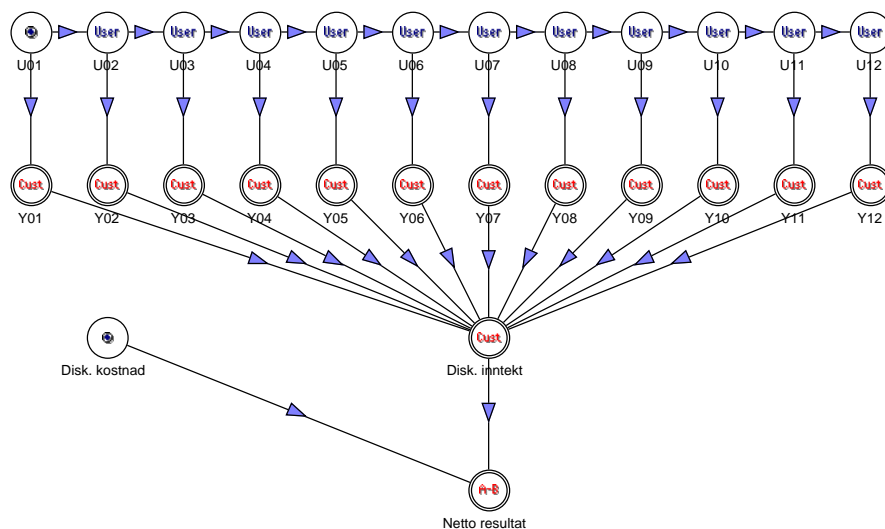
Name	ID	Def. value	Values	Multivalued	Info
Tidskorrelasjon	0	0.5	0.1; 0.2; ...; 1.0	<input type="checkbox"/>	Korrelasjon mellom påhverandre følgende ledd
Oljepris	1	18	8; 10; ...; 20	<input type="checkbox"/>	Dollar pr. fat
Dollarkurs	2	7.7		<input type="checkbox"/>	Kroner pr. dollar
Diskonteringsrente	3	0.07		<input type="checkbox"/>	7%

Figur 8: Parameter table.

Vi føyer så til tre nye noder i modellen: **Disk. inntekt**, **Disk. kostnad** og **Netto resultat**. Fra hver av nodene **Y01**, \dots , **Y12** trekker vi piler ned til **Disk. inntekt**. Endelig trekker vi piler fra **Disk. inntekt** og **Disk. kostnad**

til **Netto resultat**. Resultatet skal være et diagram som det som er illustrert i Figure ??.

VIKTIG! Opprett pila fra **Disk. inntekt** til **Netto resultat** før du oppretter pila fra **Disk. kostnad** til **Netto resultat**.



Figur 9: The complete model.

I noden **Disk. inntekt** benyttes følgende formel:

$$\text{PRODUCT}(\text{INPUTNPV}(\text{PAR}(3)); \text{PAR}(1); \text{PAR}(2)). \quad (5)$$

I noden **Disk. kostnad** benyttes følgende formel:

$$\text{NORMAL}(6000; 1000). \quad (6)$$

Til slutt benyttes følgende formel i noden **Netto resultat**:

$$\text{INPUTDIFF}(). \quad (7)$$

Sørg også for at den eneste noden som det lagres resultater for undersimuleringene, er **Netto resultat**. Gi en kort beskrivelse av hvordan formel (??) fungerer.

Kjør en simulering på denne modellen. Lag et *S-curve-plot* og et *tetthetsplot* for resultatnoden (kun for noden **Netto resultat**). Prøv å anslå ut fra S-kurven hvor stor sannsynligheten er for at prosjektet gir et netto resultat som er lavere enn en milliard.

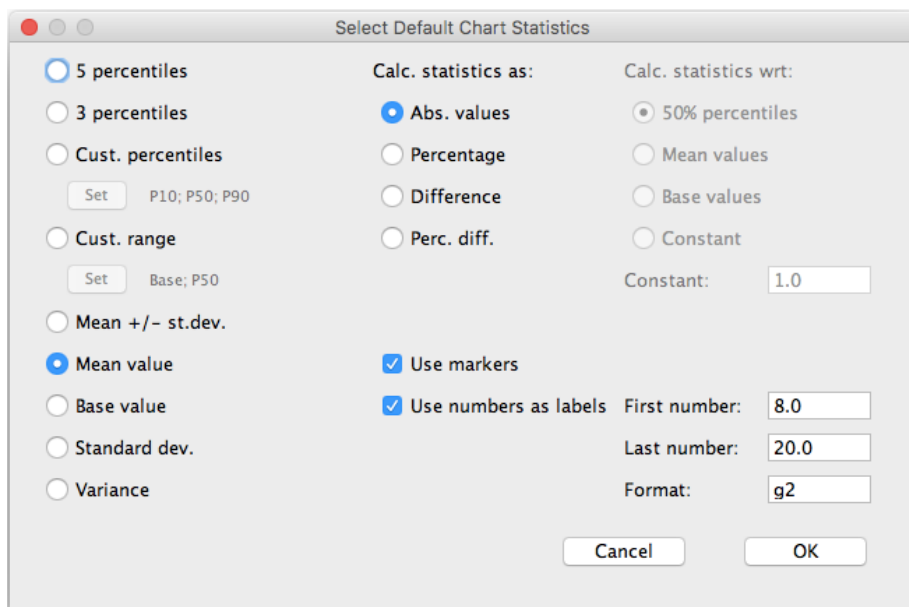
(j) Vi skal i dette punktet undersøke hvor følsomme resultatene er for variasjoner i tids-korrelasjonen, dvs. parameteren med ID = 0. Vi går derfor inn i parametervinduet igjen og krysser av for *Multivalued* for denne parameteren.

Kjør en simulering med dette oppsettet. Du skal da få opp et vindu med 10 resultatnoder, én resultatnode for hver korrelasjonsverdi, fra 0.1 opp til

1.0. Velg alle resultatnodene, og lag et felles S-kurveplot. På hvilken måte påvirker korrelasjonen fordelingen for noden **Netto resultat**? Kommentér dette resultatet.

(k) Vi skal til slutt prøve å finne ut hvor høy oljeprisen må være for at prosjektet skal ha positivt forventet netto resultat. For å undersøke dette krysser vi av for *Multivalued* for oljeprisparameteren, dvs. parameteren med ID = 1. Pass på at ingen andre parametre er krysset av som *Multivalued*.

Kjør en simulering med dette oppsettet. Du skal da få opp et vindu med 7 resultatnoder, én resultatnode for hver oljeprisverdi, dvs. for verdiene 8, 10, 12, ..., 20 (dollar pr. fat). Benytt så meny-kommandoen *Statistics>Set Chart Statistics...* Du skal da få opp en dialogboks der du kan gjøre en del valg. Sørg for at oppsettet blir nøyaktig som vist i Figure ?? under.



Figur 10: Dialogboks for *Chart Statistics...*

Velg så alle resultatnodene, og lag et linjeplot. Hvis du har gjort alt riktig, skal dette plottet vise forventet netto resultat som funksjon av oljeprisen. Hvor stor oljeprisen må være for at prosjektet skal ha positivt forventet netto resultat, kan nå leses av dette plottet som skjæringspunktet mellom kurven og null-linjen (grå stiplet vannrett linje). Kommentér resultatet.